

# Predicted interactions to iteratively update species distribution models

Last revision: *January 21, 2021*

**Abstract:** Species distribution models are characterized by three aspects of a species occurrence: its biotic environment, its abiotic environment, and its mobility range. Nevertheless, most distribution models do not address the biotic interactions that potentially shape a species range, thus measuring only how adequate an environment is for a given population. In this chapter we discuss how we can work towards better species distribution models without ignoring our knowledge about ecological networks and communities' assemblage. We suggest this can be done in seven steps: 1. careful selection of species to be modelled; 2. performance of habitat suitability models; 3. prediction of a metaweb which contains the selected species; 4. prediction of local networks' structure and assessment of their feasibility; 5. adjustment of the probabilities of occurrence based on the probabilities of realization of networks; and 6. iteration of this process until the differences between probabilities tends to zero. We highlight how this can be done with the help of machine learning techniques both to predict local networks and to update the results of grinellian habitat suitability models. Furthermore, we point to promising directions on the development of these techniques and the main challenges ecologists might face in the near future.

## Authors

Gracielle Higino <sup>1,\*</sup>  
Francis Banville <sup>2,3,4</sup>  
Gabriel Dansereau <sup>2,4</sup>  
Timothée Poisot <sup>2,4</sup>

## Affiliations

<sup>1</sup> Universidade Federal de Goiás  
<sup>2</sup> Université de Montréal  
<sup>3</sup> Université de Sherbrooke  
<sup>4</sup> Québec Centre for Biodiversity Sciences

## Correspondance

\* [graciellehigino@gmail.com](mailto:graciellehigino@gmail.com)

# 1 Introduction

2 The occurrence of a species in a given location is an encrypted message that travels  
3 through time. It carries the species' evolutionary history, long migration journeys,  
4 effects of other species we do not even know that exist, and ultimately the ele-  
5 ments that shape its, yet unknown, future. Ecologists have been trying to decode  
6 this message with progressively more powerful tools, since their own field notes  
7 to highly complex computational algorithms, such as habitat suitability models.  
8 These models were born as an attempt to model species' distribution based on  
9 their niche, considering their occurrences as sample points of suitable abiotic vari-  
10 ables and their absences as sample points of unsuitable variables. However, these  
11 observations (environmental variables and geographic location) only unveils part  
12 of the mystery, and the missing link are ecological interactions.

13 Species distribution models can be untangled in three aspects of a species occur-  
14 rence: its biotic environment - the connections it makes with other species -, its  
15 abiotic environment - the connection it makes with non-living resources -, and its  
16 mobility range - how far it can go (fig. 1)(Peterson et al. 2012). The biotic en-  
17 vironment act on these models as potential and realized interactions, constrained  
18 or enabled by abiotic factors, geographical conformation and migratory ability.  
19 Nevertheless, most distribution models do not address the biotic interactions that  
20 potentially shape a species range (or rather do so indirectly), thus measuring only  
21 how adequate an environment is for a given population. Because of that, they are  
22 denominated Habitat Suitability Models (hereafter HSMs).

23 [Figure 1 about here.]

24 The use of HSMs is very convenient because environmental variables and geo-  
25 graphic limits are not (highly) dynamic variables from the evolutionary point of

view. Because the climate (used to) change at a very slow pace, as well as species' niche, we could expect to find the same pool of species that are able to live in a certain region, even if populations fluctuated at a smaller temporal scale. This is because the cumulative effect of small scale variation on climate, population dynamics and habitat suitability itself results in macroecological outcomes such as combinations of extinction and cladogenesis, which lead to biodiversity distribution at continental scales. Also, abiotic variables are not under the influence of the focal species, which make them statistically safe, and their relationship with the species' niche is assumed to be static in space and time, which adds generalization to the model. The biotic space, on the other hand, is usually highly dynamic and variable, and it can be stochastic at very small scales to predictable structures at large scales. Additionally, because ecological networks are the cumulative result of local events (Poisot and Stouffer 2016; Guimarães 2020), its properties can vary with environmental factors and species evolutionary history (Martín González et al. 2015; Dalsgaard et al. 2013).

There is a big ecological and evolutionary leap between local dynamics of species and the biogeographical processes that are the primary assumptions to the habitat suitability and species distribution models. However, because ecological networks are very informative and aggregate populations' dynamics through scales, it is conceptually important to include them in HSMs. In fact, it has been shown that HSMs are more efficient when ecological interactions are accounted for (either directly or indirectly) (Wisz et al. 2013; Cazelles et al. 2016). Some strategies have been adopted by the scientific community to accomplish that and are shortly reviewed later in this paper. Correlative approaches assume that the co-occurrence of related species accounts for interaction, while mechanistic models try to refine this assumption by species traits and phenology. Currently, the scenario of habitat suitability models accounting for the biotic environment is either too generalistic

(correlative approaches) or too precise (mechanistic), in the sense that they only work when we have a good amount of information about that specific species. However, empirical data on ecological interactions are scarce, and, on the other hand, we cannot just assume that two species will always interact when they co-occur. How could we find balance and go further?

The good news is that ecologists have been developing techniques to predict and forecast the ecologically realistic number of links (MacDonald, Banville, and Poisot 2020), the nature of ecological interactions (Elmasri et al. 2020), and networks' properties with good accuracy. These techniques can mitigate the large and biased eltonian shortfall that we have now (Poisot et al. 2020; Hortal et al. 2015). In this context, we can envision an integrative approach of species distribution modelling combined with network prediction resulting in a more realistic, yet generalist, model where the predicted networks update the probabilities of occurrence computed by an HSM. In this paper we invite you to envision better species distribution models, which do not ignore our knowledge about ecological networks and communities assemblage. Here we suggest this can be done with the help of machine learning techniques both to predict local networks and to update the results of grinellian HSMs. We point to promising directions on the development of these techniques and main challenges ecologists might face in the near future.

[Figure 2 about here.]

## **HSMs: the mechanics, innovations and drawbacks**

Habitat suitability models aim at finding relationships between the occurrence of species and their environment (Guisan and Zimmermann 2000; Guisan, Thuiller, and Zimmermann 2017). The reader might have encountered different termi-

nologies in this area, such as species distribution modelling and ecological niche modelling, that supposedly have the same objectives. The terminology is a matter of debate on the scientific community, and here we chose to distinguish habitat suitability models from species distribution models.

Ecological niche models and habitat suitability models focus on the the area of A of the BAM diagram fig. 1 where a species can occur, which means they calculate the fundamental niche of species (Peterson et al. 2012). This can be achieved by finding the relationship between environmental conditions and the presence or absence of a certain species. This relationship can be static or dynamic in space, and only makes sense when calculated for the area inside M. Therefore, this means that they will find suitable areas inside the area 2, where species really are, but can also find suitable areas in 5, where the species probably are not because of biotic unsuitability. Species distribution models, on the other hand, should aim at modelling 2 ( $B \cap A \cap M$ ), which means considering biotic constraints (Peterson et al. 2012). Although they rarely do so, Guisan and Zimmermann (2000) argue that the observation data used as input on these models carries these information: when we use the physiological limits of a species as a variable to be correlated to the environment, we are modelling its fundamental niche, while using observational field occurrence data implies that we are modelling the realized niche (thus, the species distribution) because these data implicitly accounts for biotic limitations (Guisan and Zimmermann 2000). However, because the biotic constraints are not explicitly considered in the models, it is possible that the predicted distribution area reaches places with completely different communities. Because we did not consider previous knowledge about species interactions, how can we interpret this result? Given these differences between SDMs and HSMs, many statistical approaches can be used to model both of them. Here we focus on the most innovative algorithms to find the species' suitable habitats, to further

develop ideas on how to integrate biotic constraints.

Habitat suitability models are built over five steps: conceptualization, data sampling, calibration, evaluation and prediction (Guisan, Thuiller, and Zimmermann 2017). The conceptualization is related to the core ideas of niche, as illustrated in fig. 1, but also to the main goals of the investigation. What processes are important to the patterns we are dealing with, what is the frequency in which they repeat in space (determining the grain and extent of the environmental variables) and what are the direct and indirect predictors are examples of questions that should be asked at this step. The answers to these questions will determine how the data should be sampled (the spatial scale of the variables) and the appropriate selection of the variables (Guisan and Zimmermann 2000). Then the actual building of the mathematical model enters the stage, with iterative adjustments to capture its variations. The data is separated in training validation sets, and then the model is used to predict and project probabilities in space that can be interpreted as habitat suitability.

While statistical inference is a telescope that allows us to investigate the past of biodiversity structure, machine learning algorithms are some of the time machines we use to assess the past in order to predict the future. They have the power to learn from the data, identifying structure and generating predictions (Olden, Lawler, and Poff 2008). One sound advantage of these models compared to statistical inference is that they allow us to explore “mischievous” data, embracing the multiple frequency distributions we find in nature. Many of them have become popular among ecologists, such as MaxEnt (Phillips, Anderson, and Schapire 2006) and Random Forest (Breiman 2001), while making a good match with the continuously growing amount of biodiversity data. Although these models call for cautious usage for their complexity (and hard to interpret outputs) (Rangel and Loyola 2012), it also demands some ecological knowledge about the

relationships being modelled and careful parameterisation of terms. These very characteristics are the ones that allow us to explore nuances of biodiversity relationships, such as the role of ecological interactions in shaping species distribution - i.e. putting the B back in the BAM model.

However, these time machines need a past to learn from, and the lack of information about species interactions makes this past quite blurry. Ecological interactions occur across taxonomic and geographic scales (Guimarães 2020), and the task to collect these information through large extents and in fine grain (both temporal and spatial) is daunting. In addition, it is known that networks vary with climate, in space and with phylogenetic distance [Poisot, Stouffer, and Gravel (2014); CITATIONS]. Despite the dynamic nature of networks' structures (that are predictable outcomes of common processes), methods such as stacked (S-SDMs) and joint Species Distribution Models (JSDMs) are very common to address the influence of one species over another [CITATIONS]. These methods assume that, if a pair of species are known to interact, they will always do so once they co-occur [CITATIONS]. For example, in order to do a stacked SDM, one can model the assembly as a single unit, use the predicted distribution of each species and pile them up together, or model species at the same time, but as separate units (Ferrier and Guisan 2006). On the other hand, JSDMs use a hierarchical regression model to calculate the probability of occurrence of all species together in response to environmental variables (Pollock et al. 2014). Both approaches can be considered great advances regarding previous methods that simply added the distribution of one or more other species as limiting factors in the model [CITATIONS]. S-SDMs and JSDMs try to incorporate another level of complexity by modelling assembly rules as an intermediate step towards more realistic HSMs (see Zurell et al. (2020) for a comparison between both methods). Nevertheless, our time machines can take us even further and incorporate actual species interaction and predicted



network properties in species distribution models.

## Going further

Spatially structured co-occurrence can be the result of many historical and geographical processes, such as environmental heterogeneity and the similarity of environmental preferences between species (Bar-Massada and Belmaker 2017). HSMs that account for co-occurrence as a proxy for interactions might wrongly capture these processes as biotic interactions (Blanchet, Cazelles, and Gravel 2020). An important analysis to be made before going further is at which scale ecological interactions influence the distribution of species. Although empirical evidence exists that ecological interactions are important processes on population arrangements in space [CITATIONS] and networks variation can be detected in macroecological scales [CITATIONS], their precise relationship with spatial scale remains unclear. A second challenge is to deal with highly dynamic properties of ecological interactions. How can we account for dependency between local processes in macroecological scales? We argue that Artificial Intelligence (AI) tools can help us identify structure on the available data and predict interactions and networks for communities we did not access yet.

## The Eltonian Noise paradox

The Eltonian Noise Hypothesis states that the effect of ecological interactions on species distribution is captured at coarse resolution analyses as a covariation with environmental factors (Soberón and Nakamura 2009). This logic reverts as an assumption that always that two given species co-occur, they will interact, which has been demonstrated to be not true (Blanchet, Cazelles, and Gravel 2020). In fact, the contrary is likely to be true: the nature of interactions happening at

a given location affects co-occurrence (Bullock et al. 2000; Chesson and Kuang 2008; Godsoe and Harmon 2012; Svenning et al. 2014; Godsoe et al. 2017), and they vary in space and time due to climate change, phylogenetic diversity (addition or extinction of clades in the community, for example)(Davies and Buckley 2011), population density (Carnicer, Jordano, and Melián 2009), and many other factors. Additionally, analyses of range expansion rates - a common and very important application of HSMs - are intrinsically connected to interspecific interactions at the border of the ranges: the expansion tends to be slower when generalists predators are present or when mutualists are absent (Svenning et al. 2014). On the other hand, range preservation is also associated with ecological interactions, once connected species can be protected of climate change and invasion (Dunne, Williams, and Martinez 2002; Memmott, Waser, and Price 2004; Ramos-Jiliberto et al. 2012).

This hypothesis also seems to find no support when we investigate the betadiversity of links in ecological networks. In parasite-hosts systems in Eurasia, it is possible to detect a distinct difference in interactions between shared species in macroecological scale, suggesting an “unpacking” of species when biodiversity is expected to be lower [CHAPTER 1]. In addition, it is likely that species with larger ranges of distribution and those that are more generalists would co-occur with a greater number of other species (Dáttilo et al. 2020), while dispersal capacity of competitive species modulate their aggregation in space and the effect of interactions on their range limits (Godsoe et al. 2017).

Certainly, the exact macro scale effect of interspecies interactions might depend on the nature of these interactions, the system being studied and the combination of geographical and environmental factors. Younger communities may be more affected by environmental limitations, once they are dominated by generalist species, while older metacommunities are probably affected in different ways

210 in the centre of the distribution, at the edge of ranges and in sink and source  
211 communities. In fact, there is a growing number of evidence demonstrating that  
212 biotic interactions are important factors modulating range shift and expansion  
213 due to rapid climate change (where evolutionary processes are ignored, assuming  
214 that species will not have time to adapt) (Bullock et al. 2000; Hellmann, Prior,  
215 and Pelini 2012; Afkhami, McIntyre, and Strauss 2014; Godsoe et al. 2017; Sirén  
216 and Morelli 2020).

217 Modelling species distribution considering the Eltonian Noise Hypothesis to be  
218 true can be very useful to understand general structures of biodiversity. At coarse  
219 resolutions, with dubious quality and sparse data, and when the space for error  
220 is large, these models are probably the best we can do. However, no species is  
221 an island and we should go further whenever possible, investigating whether and  
222 how ecological interactions exert important pressures on species ranges and spatial  
223 arrangements. This will help us avoid spurious inferences, especially when these  
224 models aim at conservation strategies. Fortunately, the data we need to do that  
225 has been increasingly available (König et al. 2019), and technologies to deal with  
226 them (with their biases and gaps) are following.

## 227 **Non-stationarity of interactions and networks in space, time and across** 228 **resolutions**

229 Ecological interactions are dynamic entities of ecosystems that vary in space and  
230 time. The reason for this fluctuability is a combination of factors that shape the  
231 probability of interactions: environmental changes that affect the metabolism of  
232 individuals and the abundance of populations (Rall et al. 2012; Poisot, Cirtwill, et  
233 al. 2016; Muola et al. 2010), changes in habitat (Tylianakis and Morris 2017), and  
234 the phylogenetic structure of communities (Coelho, Rodrigues, and Rangel 2017).  
235 Notably, the variation in traits distribution (including phenological mismatches)

are important drivers of variation in species interactions [CITATIONS].

Ecological networks reflect these dynamics, and, as a result, varies in space not always in the same fashion that species richness do (Poisot, Stouffer, and Gravel 2014). The non-stationarity of ecological networks and their relationship with the environment can be measured and have been demonstrated in the last few years. Dalsgaard et al. (2013) found that modularity can be reduced and nestedness increased due to climate change speed (Dalsgaard et al. 2013). The phylogenetic diversity of networks is affected by dispersion and speciation rates (Coelho, Rodrigues, and Rangel 2017; Sebastián-González et al. 2015; Trøjelsgaard and Olesen 2013), characteristics highly correlated to both traits and geographical structure of the environment. These examples illustrate how networks can vary both in space (due to environmental filtering of species traits, for example) and time (due to evolutionary changes and phenology).

Interestingly, because ecological networks are the cumulative pool of local interactions, it is possible that network structure varies across scales (Galiana et al. 2018). For example, increased species richness can promote species packing and specialization of interactions at local scale even for generalist species [CITATIONS; CHAP 1]. Moreover, the local combinations of species and interactions in a community can result in very different regional pictures, since interactions can affect one another. For example Sanders and Van Veen (2012) observed that the persistence of two species of wasps separated by four trophic links are conditioned to one another because they regulate the population of competitive aphids (Sanders and Van Veen 2012). The replacement of one species in this network can result in the local extinction of another, even if they are not directly linked, and this dynamic could be easily misinterpreted in regional scales as an environmental constraint.

All these peculiarities about ecological interactions and networks can be faced

as challenges or opportunities for the next years. Once we understand how and where interactions are important to the accuracy of species distribution models, we can use the dynamics described here as refinements on our models. On the other hand, some of these mechanisms act at a scale so small that stochasticity takes the stage and can be accounted for in the error term of most models. This trade-off reinforces the need to carefully adjust distribution models according to the system being studied.

## **Why we can go further**

A combination of global efforts and technology can make us take another step in combining species occurrence data and ecological interactions. In the last decades we have witnessed an increase in biodiversity data, including citizen-science projects (Pocock et al. 2015; Callaghan et al. 2019) and organized dedicated databases mostly accessed by specialists, such as mangal and GBIF (Poisot, Baiser, et al. 2016; “GBIF,” n.d.). These data frequently describe the occurrence and the taxonomic identity of a species, but they also capture real-time interactions (Roy et al. 2016; Ryan et al. 2018). Moreover, both occurrence and interaction data share common characteristics that can be learned and subsequently inferred. Because species occurrence are subject to their evolutionary history that shaped their ecological niche, one can roughly infer where it could occur only based on known occurrence point locations (that is how HSMs work, as described earlier in this paper). This is connected to how ecological networks also show signs of common ancestry and preserve a basic structure (Riva and Stouffer 2016; Dallas and Poisot 2018). The ever-growing number of available data and detectable structure on ecological networks are the main ingredients we need to develop mathematical models capable of learning from what we already know to predict what we do not know yet.

## **Filling in eltonian gaps with mathematics: predicting interactions, networks, and doing it across scales**

A complete assessment of ecological interactions is even more difficult than sufficient sampling of biodiversity. The number of interactions sampled will always be lower than the number of possible interactions, mainly due to the existence of forbidden links (Jordano 2016). This lack of information, known as the Eltonian Shortfall, is aggravated by biases and differences in sampling methods (Hortal et al. 2015). Nevertheless, in the same way as the knowledge about the natural history of organisms help us validate HSMs, understanding a species' behaviour, traits and phylogenetic relationships can help us identify if the lack of links sampled is due to insufficient effort or natural mismatch between species. Additionally, interactions vary in space and time, which adds to the uncertainty of an unregistered interaction. Hence, rather than assuming interactions to be binary variables, we can think of them as probabilities (Poisot, Cirtwill, et al. 2016). In fact, assuming interactions as probabilistic events is more of an opportunity than an obstacle. Once we understand what is the basic mix than can result in a connection between ecological units (from individuals to ecosystems), we can use probabilities to estimate the likelihood of our ideas even when data is lacking. Usually, this basic mix is composed by abundance, co-occurrence in space and time (conditional to a given environment), and traits matching [Poisot, Stouffer, and Gravel (2014); CITATIONS]. The regional pool of interactions also plays an important role on interactions realization through indirect effect (Sanders and Van Veen 2012; Poisot, Stouffer, and Gravel 2014), but they also add another layer of complexity in predictive models.

When we do not have information about any of these properties, interactions tend to behave as stochastic elements in our model, while the more information we have, the more our equations will look like a niche model. Maybe one of the most neutral approaches to predicting interactions is assuming that two species will always

316 interact once they co-occur. On possible step forward is to account for spatial  
 317 structure of species distribution: interactions are more or less likely to occur  
 318 according to the location of the population inside the species range (Svenning  
 319 et al. 2014; Godsoe et al. 2017; Bar-Massada and Belmaker 2017). Moving  
 320 away from stochasticity a little bit, Canard et al. (2012) illustrates how we can  
 321 build neutral networks based on species abundance and richness, and found that  
 322 the emergent properties of these networks are compatible with empirical ones.  
 323 With this method, it is possible to identify neutral forbidden links (Canard et al.  
 324 2012), which unfolds many other questions that can be answered without previous  
 325 empirical data. For example, we can investigate whether abundance fluctuation  
 326 is the main driver of betadiversity of links between networks, or compare how  
 327 networks are different from the neutral model when only the traits vary. It is  
 328 important to notice, however, that abundance also plays a role on the nature of  
 329 the interaction. Mutualistic interactions, for example, can turn into parasitism or  
 330 competition when one of the players is very abundant (Wolin and Lawlor 1984).  
 331 An intermediate step could be to find detectable interactions in a species pool.  
 332 Xiao Fu et al. (2019) proposed a combination of a Poisson N-mixture model and  
 333 collaborative filtering to predict potential links under imperfect detection. This  
 334 machine learning model allowed them to successfully infer interactions based on  
 335 a few observed occurrences sampled in the field (Xiao Fu et al. 2019). This ap-  
 336 proach deals with interactions as countable units, ignoring evolutive and ecological  
 337 mechanisms that could lead to variation in the probability of their realization.  
 338 However, it can be a good exploratory method to investigate the completeness of  
 339 interaction sampling.

340 On our way to more mechanistic models, we can start looking for evolutionary  
 341 clues on our encrypted message. First, we can assume that similar species will  
 342 interact in similar ways with their “pairs” due to phylogenetic inertia (Gómez,

Verdú, and Perfectti 2010; Peralta 2016). If predators  $A$  and  $B$  are closely related, they are likely to compete for the same set of preys. In this sense, we can use machine learning algorithms to find in the “potential interactions” pool those that are more ecologically likely to occur. For example, Desjardins-Proulx et al. (2017) showed that the  $K$  nearest neighbour (KNN) and the Random Forest (RF) algorithms are complementary when it comes to identifying the probability of a predator to interact with a certain prey and binary occurrence of interactions (RF predicts either interaction or non-interaction) (Desjardins-Proulx et al. 2017). According to the authors, the usefulness of these methods depend on the kind of data we have in hands: KNNs need previous information about interactions to learn from, while RFs should use a set of traits. Another example of link prediction based on trait matching is Dallas, Park, and Drake (2017), where the authors identified cryptic associations between hosts and parasites based on a Bayesian model. Similar to the approach on Xiao Fu et al. (2019), this model also allow us to investigate the completeness of interaction sampling, but now based on traits matching and not only on occurrence observations. Pichler et al. provide an extensive methodological comparison of Machine Learning models to predict interactions based on traits matching. Nonetheless, these techniques isolate interactions from their biotic and abiotic environments, and this should be kept in mind whenever predictions are made.

Another step further would be implementing evolutionary models in the relationship between links and phylogeny. Elmasri et al. (2020) point out that recent approaches to link prediction using phylogenetic similarity assume a linear or fixed diversification rate. They proceed to develop a phylogenetic matching model of predicted interactions using the early-burst tree scaling model, where the traits diversification tree is transformed, adjusted by a single parameter, either to early or late diversification (Harmon et al. 2010; Elmasri et al. 2020). The authors also



370 accounted for uncertainty in unobserved interactions, which improved the accu-  
371 racy of the predictions. The adjustment of the early-burst parameter can be based  
372 on ecological reasoning. For example, food webs with species clustered in trophic  
373 levels (specialized predators, for example) seems to have a trait evolution with a  
374 high diversification rate early in the tree, while networks with many omnivorous  
375 species are thought to have a late diversification of traits (Ingram, Harmon, and  
376 Shurin 2012). These predictions are highly mechanistic, accounting for evolution  
377 and ecology of interactions.

378 Connected to the Eltonian Shortfall described at the beginning of this session,  
379 there are global gaps and biases on network sampling (Poisot et al. 2020). Be-  
380 cause ecological networks are made of nodes and edges, and both can be treated  
381 as probabilities (addressing the occurrence of a node and the realization of an  
382 interaction), we can investigate the chances of different networks to emerge from  
383 a set of species co-occurring in a given location (Poisot, Cirtwill, et al. 2016).  
384 Recently, MacDonald, Banville, and Poisot (2020) demonstrated how we can esti-  
385 mate the number of possible links in a network based on the number of species it  
386 has, which take us closer to predict networks themselves since we have much more  
387 information about species than we have on interactions (MacDonald, Banville,  
388 and Poisot 2020). Knowing the number of possible links given the species rich-  
389 ness in a given location allow us to calculate the probability distribution for each  
390 network property, which can be further connected to what we know about the  
391 relationship between networks' properties and species spatial distribution

392 One question that is frequently stressed by macroecologists is how to connect local  
393 processes, such as interactions, to large scale biodiversity structures. The sam-  
394 pling of both interactions and environmental variables are usually made in oppos-  
395 ing ends of spatial resolution, which makes them incomparable from a statistical  
396 point of view (Peterson et al. 2012). Also, because of the nature of the sampling,

397 it is hard to transform those variables to make them vary in compatible scales.  
 398 How can we deal with this problem when predicting interactions and networks  
 399 for macroscales? In fact, the basic mix components that leads to the occurrence  
 400 of an interaction varies with the spatial resolution in which it is assessed. Conse-  
 401 quently, the distribution of probabilities for interactions and network properties  
 402 would change, possibly leading to different interpretation of biodiversity distri-  
 403 bution. However, the different ways in which networks properties vary through  
 404 scales have also been investigated, and simulations can help us understand that  
 405 more deeply (Poisot, Stouffer, and Gravel 2014; Peralta 2016; Guimarães 2020).  
 406 *Because* interactions vary more in space than species occurrence (while adding to  
 407 this information) (Poisot et al. 2017), they can represent biodiversity with a more  
 408 useful macroecological variance to interplay with environmental variables. More-  
 409 over, because interaction predictions accounts only for *potential interactions*, they  
 410 are easily translated into regional networks (that also represent potential inter-  
 411 actions). Finally, some properties of networks are correlated to the phylogenetic  
 412 diversity of communities [CHAP 1], which also represents a level of organization  
 413 compatible with the geographical scale of environmental variables.

414 **Updating probabilities of occurrences with network probabilities** At  
 415 this point the reader may be wondering how probabilistic networks and species  
 416 distribution models fit together. Some efforts in this sense have already been  
 417 made, and in this session we will quickly review them while pointing to future  
 418 directions. We will not consider, however, the majority of papers that aim at  
 419 discussing the integration of biotic interactions on SDMs because their definitions  
 420 of ecological interactions are not compatible with our view. Usually, these discus-  
 421 sions consider the mere addition of the occurrence of a competitor or mutualist  
 422 species as a predictor variable on the model to be sufficient to account for ecologi-

cal interactions [Guisan et al. (2006); Wisz et al. (2013); CITATIONS]. Although this practice have demonstrated to improve the accuracy of distribution models [CITATIONS], the ecological interpretation is still not convincing. First, as we discussed earlier, the realization of an interaction is conditional to many factors, including abundance and other interactions, and it's strength and signal can vary in space and time. Second, adding a competitor as an equivalent of an absence of the focal species just informs the model about environmental conditions that are supposedly not suitable for the establishment of a population, not that an ecological interaction is taking place there in such a manner that it will influence the survival of a pair of individuals.

As reviewed by Cabral, Valente, and Hartig (2017), many models in ecology have considered competition and facilitation in range shifts, but the assessment of trophic interactions remains insufficient. An example of good theoretical exercise was done in Godsoe and Harmon (2012), where authors calculated the probability that a given site will be environmentally suitable in the presence of a competitor, based on their population dynamic models (Godsoe and Harmon 2012). Although it is close to assuming a binary nature for interactions, it does account for dynamics and simulate results for different abundances, thus approximating to a probabilistic approach. There are only a few attempts on the literature that tries to build mechanistic species distribution models accounting for interspecies interaction (independent of its nature) as a probabilistic entity (Cabral, Valente, and Hartig 2017). A first step in this direction was taken in Gravel et al. (2011), where authors demonstrated how interactions can be included in the Theory of Island Biogeography (MacArthur and Wilson 1967). This experiment provided a starting framework for the next generation of species distribution models to include migration dynamics and trophic levels.

In Cazelles et al. (2016), the authors elaborate on Gravel et al. (2011) model

450 and include environmental constraints in a metacommunities dynamic model in a  
451 biogeographical scale. Their model add some necessary complexity by evaluating  
452 frames of communities in different times, calculating the probability of change  
453 in community composition due to migration, interactions and environmental re-  
454 quirements. This is a very useful model to understand spatial dynamics of species  
455 richness, and could be used as a validation strategy for HSMs with only two time  
456 frames: the current known distribution and the potential distribution inferred by  
457 the HSM. Because the probability of occurrence of one species is derived from the  
458 probability of occurrence of all other species (Cazelles et al. 2016), we could infer  
459 the probability of occurrence of a focal species based on its potential distribu-  
460 tion and local species pools. This strategy would require to predict interactions  
461 between species that did not previously co-occur.

462 In this sense, an interesting approach was elaborated in Staniczenko et al. (2017),  
463 where the authors address interactions with Bayesian Networks models that up-  
464 date the probability of occurrence resulted from a HSM. Bayesian Networks (BN)  
465 can be derived from empirical data or inferred from macroecological data, but  
466 they should only represent directional interactions. This was pointed as a limita-  
467 tion for the use of this workflow, as interactions are often retroactive [CITATION],  
468 but it showed good efficacy when the research subject allows this simplification  
469 (Staniczenko et al. 2017). In these networks, the effects of direct and indirect in-  
470 teractions are spread among the species, and the nodes represent the probabilities  
471 of occurrence. This allows that multiple species are considered at the same time  
472 and to derive the network from the assemblage resulted of a HSM.

## Synthesis

We envision a combination of all techniques described here, which would allow a more reliable estimation of the area where a species would probably occur (illustrated as the area 2 in fig. 1 and in fig. 3). This combination would involve three steps, as delineated by Staniczenko et al. (2017), with an initial HSM to assess the potential distribution of one or more focal species, and the use of machine learning algorithms to update the probability of occurrence from information on potential interactions. In this sense, there are two possible venues that could be followed: (i) calculate the probability of interaction from the local species pool, as suggest Cazelles et al. (2016), and use it to update the probabilities from the HSM, or (ii) calculate the probabilities of connections, then the probability of interactions where the chance of connection is higher than a threshold, and finally using this probability as a prior on a species distribution model. We propose a workflow that follows the later path, with four steps described below.

[Figure 3 about here.]

### (i) Potential distribution of focal species

A traditional habitat suitability model should be used as a first step to assess the M+A space of the species (fig. 1). The choice of the best model will depend on a few criteria such as the nature and quality of your data, and the processes you want to investigate, as described in Guisan, Thuiller, and Zimmermann (2017) and Peterson et al. (2012). A HSM will indicate where there are suitable habitats for the species, which may include areas where the species was never sampled before. These areas, the “error” of the model, should be the object of assessment on the next steps since it is there where the unknown information about interactions is. For each sample cell where the species is predicted to occur, the value of known

species richness can be used to calculate potential networks with the addition of the focal species.

## **(ii) Potential networks based on species richness**

The minimum and maximum number of links that are possible depends on how many species share the same space (MacDonald, Banville, and Poisot 2020). Based on that, we can calculate the degree distributions that are possible based on Maximum Entropy (HALP francisbanville?)

## **(iii) Potential interactions where the probability of a degree $\geq 1$ is not null**

For the cells where the focal species has at least one possible network where its degree is  $\geq 1$ , we can proceed to investigate whether this link is ecologically relevant. This can be done by sampling a subset of species that occur and that could be important to that focal species (either a competitor, a prey, a predator or a facilitator). Some variables that can be important in this process are the phylogenetic or the functional diversity, since they can help us subset the species pool by evolutionary distance or ecological redundancy. Then the probability of interaction can be calculated for the focal species and the subset of species following the methods we discussed above, depending on the nature of the data available at a macroecological scale. This analysis should result on a spatial distribution of interaction probabilities for each cell where the species were predicted to occur but was not sampled before.

#### (iv) Probability of occurrence given the probability of interactions

For those cells where we now have the probability of occurrence and interaction, the probability of occurrence given the probability of interaction is a new function  $h(x)$ , which is a combination between a function of probability of occurrence  $f(x)$  and a function of probability of interaction  $g(x)$  eq. 1.

$$h(x) = d(x) \left[ \int f(x) - \left( \int f(x) - g(x) \right) - \left( \int g(x) - f(x) \right) \right] \quad (1)$$

TODO

Updating the probabilities of occurrence based on the probabilities of interaction is a way to integrate biotic variables on species distribution models as a separate step. We believe that this pathway can simplify this intricate problem because it disentangles the complex calculation of ecological interactions from the abiotic suitability models. Both sides can be run in parallel, and in some cases the probability of interaction does not need to be calculated every time a model is run.

- What don't we have yet?
- What are the roads never taken?
- Where should we invest in order to achieve the best SDMs we can?
- What are the most promising areas for development?

## References

Afkhami, Michelle E., Patrick J. McIntyre, and Sharon Y. Strauss. 2014. "Mutualist-Mediated Effects on Species' Range Limits Across Large Geographic Scales." *Ecology Letters* 17 (10): 1265–73. <https://doi.org/10.1111/ele.12332>.

- Bar-Massada, Avi, and Jonathan Belmaker. 2017. “Non-Stationarity in the Co-Occurrence Patterns of Species Across Environmental Gradients.” *Journal of Ecology* 105 (2): 391–99. <https://doi.org/10.1111/1365-2745.12713>.
- Blanchet, F. Guillaume, Kevin Cazelles, and Dominique Gravel. 2020. “Co-Occurrence Is Not Evidence of Ecological Interactions.” *Ecology Letters* 23 (7): 1050–63. <https://doi.org/10.1111/ele.13525>.
- Breiman, Leo. 2001. “Random Forests.” *Machine Learning* 45 (1): 5–32. <https://doi.org/10.1023/A:1010933404324>.
- Bullock, James M., Rebecca J. Edwards, Peter D. Carey, and Rob J. Rose. 2000. “Geographical Separation of Two *Ulex* Species at Three Spatial Scales: Does Competition Limit Species’ Ranges?” *Ecography* 23 (2): 257–71. <https://doi.org/10.1111/j.1600-0587.2000.tb00281.x>.
- Cabral, Juliano Sarmento, Luis Valente, and Florian Hartig. 2017. “Mechanistic Simulation Models in Macroecology and Biogeography: State-of-Art and Prospects.” *Ecography* 40 (2): 267–80. <https://doi.org/10.1111/ecog.02480>.
- Callaghan, Corey T., Jodi J. L. Rowley, William K. Cornwell, Alistair G. B. Poore, and Richard E. Major. 2019. “Improving Big Citizen Science Data: Moving Beyond Haphazard Sampling.” *PLOS Biology* 17 (6): e3000357. <https://doi.org/10.1371/journal.pbio.3000357>.
- Canard, Elsa, Nicolas Mouquet, Lucile Marescot, Kevin J. Gaston, Dominique Gravel, and David Mouillot. 2012. “Emergence of Structural Patterns in Neutral Trophic Networks.” *PLOS ONE* 7 (8): e38295. <https://doi.org/10.1371/journal.pone.0038295>.
- Carnicer, Jofre, Pedro Jordano, and Carlos J. Melián. 2009. “The Temporal Dynamics of Resource Use by Frugivorous Birds: A Network Approach.” *Ecology*



90 (7): 1958–70.

Cazelles, Kévin, Nicolas Mouquet, David Mouillot, and Dominique Gravel. 2016.

“On the Integration of Biotic Interaction and Environmental Constraints at the Biogeographical Scale.” *Ecography* 39 (10): 921–31. <https://doi.org/10.1111/ecog.01714>.

Chesson, Peter, and Jessica J. Kuang. 2008. “The Interaction Between Predation and Competition.” *Nature* 456 (7219): 235–38. <https://doi.org/10.1038/nature07248>.

Coelho, Marco Túlio Pacheco, João Fabrício Mota Rodrigues, and Thiago F Rangel. 2017. “Neutral Biogeography of Phylogenetically Structured Interaction Networks.” *Ecography* 40 (12): 1467–74.

Dallas, Tad, Andrew W. Park, and John M. Drake. 2017. “Predicting Cryptic Links in Host-Parasite Networks.” *PLOS Computational Biology* 13 (5): e1005557. <https://doi.org/10.1371/journal.pcbi.1005557>.

Dallas, Tad, and Timothée Poisot. 2018. “Compositional Turnover in Host and Parasite Communities Does Not Change Network Structure.” *Ecography* 41 (9): 1534–42. <https://doi.org/10.1111/ecog.03514>.

Dalsgaard, Bo, Kristian Trøjelsgaard, Ana M Martín González, David Nogués-Bravo, Jeff Ollerton, Theodora Petanidou, Brody Sandel, et al. 2013. “Historical Climate-Change Influences Modularity and Nestedness of Pollination Networks.” *Ecography* 36 (12): 1331–40.

Davies, T Jonathan, and Lauren B Buckley. 2011. “Phylogenetic Diversity as a Window into the Evolutionary and Biogeographic Histories of Present-Day Richness Gradients for Mammals.” *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 366 (1576): 2414–25.

Dáttilo, Wesley, Nathalia Barrozo-Chávez, Andrés Lira-Noriega, Roger Guevara,

Fabricio Villalobos, Diego Santiago-Alarcon, Frederico Siqueira Neves, Thiago Izzo, and Sérgio Pontes Ribeiro. 2020. “Species-Level Drivers of Mammalian Ectoparasite Faunas.” *Journal of Animal Ecology* 89 (8): 1754–65. <https://doi.org/10.1111/1365-2656.13216>.

Desjardins-Proulx, Philippe, Idaline Laigle, Timothée Poisot, and Dominique Gravel. 2017. “Ecological Interactions and the Netflix Problem.” *PeerJ* 5 (August): e3644. <https://doi.org/10.7717/peerj.3644>.

Dunne, Jennifer A., Richard J. Williams, and Neo D. Martinez. 2002. “Network Structure and Biodiversity Loss in Food Webs: Robustness Increases with Connectance.” *Ecology Letters* 5 (4): 558–67. <https://doi.org/10.1046/j.1461-0248.2002.00354.x>.

Elmasri, Mohamad, Maxwell J. Farrell, T. Jonathan Davies, and David A. Stephens. 2020. “A Hierarchical Bayesian Model for Predicting Ecological Interactions Using Scaled Evolutionary Relationships.” *Annals of Applied Statistics* 14 (1): 221–40. <https://doi.org/10.1214/19-AOAS1296>.

Ferrier, Simon, and Antoine Guisan. 2006. “Spatial Modelling of Biodiversity at the Community Level.” *Journal of Applied Ecology* 43 (3): 393–404. <https://doi.org/10.1111/j.1365-2664.2006.01149.x>.

Galiana, Nuria, Miguel Lurgi, Bernat Claramunt-López, Marie-Josée Fortin, Shawn Leroux, Kevin Cazelles, Dominique Gravel, and José M. Montoya. 2018. “The Spatial Scaling of Species Interaction Networks.” *Nature Ecology & Evolution* 2 (5): 782–90. <https://doi.org/10.1038/s41559-018-0517-3>.

“GBIF.” n.d. <https://www.gbif.org/>.

Godsoe, William, and Luke J. Harmon. 2012. “How Do Species Interactions Affect Species Distribution Models?” *Ecography* 35 (9): 811–20. <https://doi.org/10.1111/j.1600-0587.2011.07103.x>.

- 619 Godsoe, William, Jill Jankowski, Robert D. Holt, and Dominique Gravel. 2017.  
620 “Integrating Biogeography with Contemporary Niche Theory.” *Trends in Ecology and Evolution* 32 (7): 488–99. [https://doi.org/10.1016/j.tree.2017.](https://doi.org/10.1016/j.tree.2017.03.008)  
621 [03.008](https://doi.org/10.1016/j.tree.2017.03.008).  
622
- 623 Gómez, José M., Miguel Verdú, and Francisco Perfectti. 2010. “Ecological Inter-  
624 actions Are Evolutionarily Conserved Across the Entire Tree of Life.” *Nature*  
625 465 (7300): 918–21. <https://doi.org/10.1038/nature09113>.
- 626 Gravel, Dominique, François Massol, Elsa Canard, David Mouillot, and Nicolas  
627 Mouquet. 2011. “Trophic Theory of Island Biogeography.” *Ecology Letters* 14  
628 (10): 1010–16. <https://doi.org/10.1111/j.1461-0248.2011.01667.x>.
- 629 Guimarães, Paulo R. 2020. “The Structure of Ecological Networks Across Levels of  
630 Organization.” *Annual Review of Ecology, Evolution, and Systematics* 51 (1):  
631 433–60. <https://doi.org/10.1146/annurev-ecolsys-012220-120819>.
- 632 Guisan, Antoine, Anthony Lehmann, Simon Ferrier, Mike Austin, Jacob Mc. C.  
633 Overton, Richard Aspinall, and Trevor Hastie. 2006. “Making Better Biogeo-  
634 graphical Predictions of Species’ Distributions.” *Journal of Applied Ecology*  
635 43 (3): 386–92. <https://doi.org/10.1111/j.1365-2664.2006.01164.x>.
- 636 Guisan, Antoine, Wilfried Thuiller, and Niklaus E. Zimmermann. 2017. *Habitat*  
637 *Suitability and Distribution Models: With Applications in R*. Ecology, Biodi-  
638 versity and Conservation. Cambridge: Cambridge University Press. [https:](https://doi.org/10.1017/9781139028271)  
639 [//doi.org/10.1017/9781139028271](https://doi.org/10.1017/9781139028271).
- 640 Guisan, Antoine, and Niklaus E. Zimmermann. 2000. “Predictive Habitat Dis-  
641 tribution Models in Ecology.” *Ecological Modelling* 135 (2): 147–86. [https:](https://doi.org/10.1016/S0304-3800(00)00354-9)  
642 [//doi.org/10.1016/S0304-3800\(00\)00354-9](https://doi.org/10.1016/S0304-3800(00)00354-9).
- 643 Harmon, Luke J., Jonathan B. Losos, T. Jonathan Davies, Rosemary G. Gillespie,  
644 John L. Gittleman, W. Bryan Jennings, Kenneth H. Kozak, et al. 2010. “Early

Bursts of Body Size and Shape Evolution Are Rare in Comparative Data.”  
*Evolution* 64 (8): 2385–96. <https://doi.org/10.1111/j.1558-5646.2010.01025.x>.

Hellmann, Jessica J., Kirsten M. Prior, and Shannon L. Pelini. 2012. “The  
Influence of Species Interactions on Geographic Range Change Under Climate  
Change.” *Annals of the New York Academy of Sciences* 1249 (February): 18–  
28. <https://doi.org/10.1111/j.1749-6632.2011.06410.x>.

Hortal, Joaquín, Francesco de Bello, José Alexandre F. Diniz-Filho, Thomas M.  
Lewinsohn, Jorge M. Lobo, and Richard J. Ladle. 2015. “Seven Shortfalls  
That Beset Large-Scale Knowledge of Biodiversity.” *Annual Review of Ecology,  
Evolution, and Systematics* 46 (1): 523–49. <https://doi.org/10.1146/annurev-ecolsys-112414-054400>.

Ingram, T., L. J. Harmon, and J. B. Shurin. 2012. “When Should We Expect Early  
Bursts of Trait Evolution in Comparative Data? Predictions from an Evolutionary  
Food Web Model.” *Journal of Evolutionary Biology* 25 (9): 1902–10.  
<https://doi.org/10.1111/j.1420-9101.2012.02566.x>.

Jordano, Pedro. 2016. “Sampling Networks of Ecological Interactions.” *Functional Ecology* 30 (12): 1883–93. <https://doi.org/10.1111/1365-2435.12763>.

Jorge Soberón. 2007. “Grinnellian and Eltonian Niches and Geographic Distributions  
of Species.” *Ecology Letters* 10 (12): 1115–23. <https://doi.org/10.1111/j.1461-0248.2007.01107.x>.

König, Christian, Patrick Weigelt, Julian Schrader, Amanda Taylor, Jens Kattge,  
and Holger Kreft. 2019. “Biodiversity Data Integrationthe Significance of  
Data Resolution and Domain.” *PLOS Biology* 17 (3): e3000183. <https://doi.org/10.1371/journal.pbio.3000183>.

- 671 MacArthur, Robert H., and Edward O. Wilson. 1967. *The Theory of Island*  
672 *Biogeography*. Princeton University Press.
- 673 MacDonald, Arthur Andrew Meahan, Francis Banville, and Timothée Poisot.  
674 2020. “Revisiting the Links-Species Scaling Relationship in Food Webs.” *Pat-*  
675 *terns* 1 (0). <https://doi.org/10.1016/j.patter.2020.100079>.
- 676 Martín González, Ana M., Bo Dalsgaard, David Nogués-Bravo, Catherine H.  
677 Graham, Matthias Schleuning, Pietro K. Maruyama, Stefan Abrahamczyk,  
678 et al. 2015. “The Macroecology of Phylogenetically Structured Hummingbird-  
679 Plant Networks.” *Global Ecology and Biogeography* 24 (11): 1212–24. <https://doi.org/10.1111/geb.12355>.  
680
- 681 Memmott, J., N. M. Waser, and M. V. Price. 2004. “Tolerance of Pollination Net-  
682 works to Species Extinctions.” *Proceedings of the Royal Society B: Biological*  
683 *Sciences* 271 (1557): 2605–11. <https://doi.org/10.1098/rspb.2004.2909>.
- 684 Muola, Anne, Pia Mutikainen, Marianna Lilley, Liisa Laukkanen, Juha Pekka  
685 Salminen, and Roosa Leimu. 2010. “Associations of Plant Fitness, Leaf Chem-  
686 istry, and Damage Suggest Selection Mosaic in Plant-Herbivore Interactions.”  
687 *Ecology*. <https://doi.org/10.1890/09-0589.1>.
- 688 Olden, Julian, Joshua Lawler, and N. Poff. 2008. “Machine Learning Methods  
689 Without Tears: A Primer for Ecologists.” *The Quarterly Review of Biology* 83  
690 (July): 171–93. <https://doi.org/10.1086/587826>.
- 691 Peralta, Guadalupe. 2016. “Merging Evolutionary History into Species Interac-  
692 tion Networks.” *Functional Ecology* 30 (12): 1917–25. <https://doi.org/10.1111/1365-2435.12669>.  
693
- 694 Peterson, A. Townsend, Jorge Soberón, Richard G. Pearson, Robert P. Anderson,  
695 Enrique Martínez-Meyer, Miguel Nakamura, and Miguel Bastos Araújo. 2012.  
696 *Ecological Niches and Geographic Distributions*. *Choice Reviews Online*. Vol.

697 49. <https://doi.org/10.5860/CHOICE.49-6266>.

698 Phillips, Steven J., Robert P. Anderson, and Robert E. Schapire. 2006. “Maxi-  
699 mum Entropy Modeling of Species Geographic Distributions.” *Ecological Mod-*  
700 *elling* 190 (3): 231–59. [https://doi.org/10.1016/j.ecolmodel.2005.03.](https://doi.org/10.1016/j.ecolmodel.2005.03.026)  
701 [026](https://doi.org/10.1016/j.ecolmodel.2005.03.026).

702 Pichler, Maximilian, Virginie Boreux, Alexandra-Maria Klein, Matthias Schle-  
703 uning, and Florian Hartig. “Machine Learning Algorithms to Infer Trait-  
704 Matching and Predict Species Interactions in Ecological Networks.” *Meth-*  
705 *ods in Ecology and Evolution* 11 (2): 281–93. [https://doi.org/10.1111/](https://doi.org/10.1111/2041-210X.13329)  
706 [2041-210X.13329](https://doi.org/10.1111/2041-210X.13329).

707 Pocock, Michael J. O., Helen E. Roy, Chris D. Preston, and David B. Roy. 2015.  
708 “The Biological Records Centre: A Pioneer of Citizen Science.” *Biological*  
709 *Journal of the Linnean Society* 115 (3): 475–93. [https://doi.org/10.1111/](https://doi.org/10.1111/bij.12548)  
710 [bij.12548](https://doi.org/10.1111/bij.12548).

711 Poisot, Timothée, Benjamin Baiser, Jennifer A Dunne, Sonia Kéfi, François Mas-  
712 sol, Nicolas Mouquet, Tamara N Romanuk, Daniel B Stouffer, Spencer A  
713 Wood, and Dominique Gravel. 2016. “Mangal - Making Ecological Network  
714 Analysis Simple.” *Ecography* 39 (4): 384–90.

715 Poisot, Timothée, Gabriel Bergeron, Kevin Cazelles, Tad Dallas, Dominique Gravel,  
716 Andrew Macdonald, Benjamin Mercier, Clément Violet, and Steve Vissault.  
717 2020. “Environmental Biases in the Study of Ecological Networks at the  
718 Planetary Scale.” *bioRxiv*, January. [https://doi.org/10.1101/2020.01.](https://doi.org/10.1101/2020.01.27.921429)  
719 [27.921429](https://doi.org/10.1101/2020.01.27.921429).

720 Poisot, Timothée, Alyssa R Cirtwill, Kévin Cazelles, Dominique Gravel, Marie  
721 Josée Fortin, and Daniel B Stouffer. 2016. “The Structure of Probabilistic  
722 Networks.” *Methods in Ecology and Evolution* 7 (3): 303–12. [https://doi.](https://doi.org/10.1111/2041-210X.12548)

723 [org/10.1111/2041-210X.12468](https://doi.org/10.1111/2041-210X.12468).

724 Poisot, Timothée, Cynthia Guéveneux-Julien, Marie Josée Fortin, Dominique  
725 Gravel, and Pierre Legendre. 2017. “Hosts, Parasites and Their Interactions  
726 Respond to Different Climatic Variables.” *Global Ecology and Biogeography*  
727 26 (8): 942–51. <https://doi.org/10.1111/geb.12602>.

728 Poisot, Timothée, and Daniel Stouffer. 2016. “How Ecological Networks Evolve.”  
729 *bioRxiv*, no. Jablonski 2008: 071993. <https://doi.org/10.1101/071993>.

730 Poisot, Timothée, Daniel B. Stouffer, and Dominique Gravel. 2014. “Beyond  
731 Species: Why Ecological Interaction Networks Vary Through Space and Time.”  
732 *Oikos* 124 (3): 243–51. <https://doi.org/10.1111/oik.01719>.

733 Pollock, Laura J., Reid Tingley, William K. Morris, Nick Golding, Robert B.  
734 O’Hara, Kirsten M. Parris, Peter A. Vesk, and Michael A. McCarthy. 2014.  
735 “Understanding Co-Occurrence by Modelling Species Simultaneously with a  
736 Joint Species Distribution Model (JSDM).” *Methods in Ecology and Evolution*  
737 5 (5): 397–406. <https://doi.org/10.1111/2041-210X.12180>.

738 Rall, B. C., U. Brose, M. Hartvig, G. Kalinkat, F. Schwarzmuller, O. Vucic-Pestic,  
739 and O. L. Petchey. 2012. “Universal Temperature and Body-Mass Scaling of  
740 Feeding Rates.” *Philosophical Transactions of the Royal Society B: Biological*  
741 *Sciences* 367 (1605): 2923–34. <https://doi.org/10.1098/rstb.2012.0242>.

742 Ramos-Jiliberto, Rodrigo, Fernanda S. Valdovinos, Pablo Moisset de Espanés,  
743 and José D. Flores. 2012. “Topological Plasticity Increases Robustness of  
744 Mutualistic Networks.” *Journal of Animal Ecology* 81 (4): 896–904. <https://doi.org/10.1111/j.1365-2656.2012.01960.x>.

746 Rangel, Thiago Fernando, and Rafael Dias Loyola. 2012. “Labeling Ecological  
747 Niche Models.” *Natureza & Conservação* 10 (2): 119–26. <https://doi.org/10.4322/natcon.2012.030>.

- 749 Riva, Giulio V. Dalla, and Daniel B. Stouffer. 2016. “Exploring the Evolutionary  
750 Signature of Food Webs’ Backbones Using Functional Traits.” *Oikos* 125 (4):  
751 446–56. <https://doi.org/10.1111/oik.02305>.
- 752 Roy, Helen E., Elizabeth Baxter, Aoine Saunders, and Michael J. O. Pocock. 2016.  
753 “Focal Plant Observations as a Standardised Method for Pollinator Monitoring:  
754 Opportunities and Limitations for Mass Participation Citizen Science.” *PLOS*  
755 *ONE* 11 (3): e0150794. <https://doi.org/10.1371/journal.pone.0150794>.
- 756 Ryan, S. F., N. L. Adamson, A. Aktipis, L. K. Andersen, R. Austin, L. Barnes,  
757 M. R. Beasley, et al. 2018. “The Role of Citizen Science in Addressing Grand  
758 Challenges in Food and Agriculture Research.” *Proceedings of the Royal So-*  
759 *ciety B: Biological Sciences* 285 (1891). [https://doi.org/10.1098/rspb.](https://doi.org/10.1098/rspb.2018.1977)  
760 [2018.1977](https://doi.org/10.1098/rspb.2018.1977).
- 761 Sanders, Dirk, and F. J. Frank Van Veen. 2012. “Indirect Commensalism Promotes  
762 Persistence of Secondary Consumer Species.” *Biology Letters* 8 (6): 960–63.  
763 <https://doi.org/10.1098/rsbl.2012.0572>.
- 764 Sebastián-González, Esther, Bo Dalsgaard, Brody Sandel, and Paulo R Guimarães.  
765 2015. “Macroecological Trends in Nestedness and Modularity of Seed-Dispersal  
766 Networks: Human Impact Matters.” *Global Ecology and Biogeography* 24 (3):  
767 293–303.
- 768 Sirén, Alexej P. K., and Toni Lyn Morelli. 2020. “Interactive Range-Limit Theory  
769 (iRLT): An Extension for Predicting Range Shifts.” *Journal of Animal Ecology*  
770 89 (4): 940–54. <https://doi.org/10.1111/1365-2656.13150>.
- 771 Soberón, Jorge, and Miguel Nakamura. 2009. “Niches and Distributional Ar-  
772 eas: Concepts, Methods, and Assumptions.” *Proceedings of the National*  
773 *Academy of Sciences* 106 (Supplement 2): 19644–50. [https://doi.org/10.](https://doi.org/10.1073/pnas.0901637106)  
774 [1073/pnas.0901637106](https://doi.org/10.1073/pnas.0901637106).



- 775 Staniczenko, Phillip P. A., Prabu Sivasubramaniam, K. Blake Suttle, and Richard  
 776 G. Pearson. 2017. “Linking Macroecology and Community Ecology: Refin-  
 777 ing Predictions of Species Distributions Using Biotic Interaction Networks.”  
 778 *Ecology Letters* 20 (6): 693–707. <https://doi.org/10.1111/ele.12770>.
- 779 Svenning, Jens Christian, Dominique Gravel, Robert D. Holt, Frank M. Schurr,  
 780 Wilfried Thuiller, Tamara Münkemüller, Katja H. Schiffrers, et al. 2014. “The  
 781 Influence of Interspecific Interactions on Species Range Expansion Rates.”  
 782 *Ecography* 37 (12): 1198–1209. [https://doi.org/10.1111/j.1600-0587.](https://doi.org/10.1111/j.1600-0587.2013.00574.x)  
 783 [2013.00574.x](https://doi.org/10.1111/j.1600-0587.2013.00574.x).
- 784 Trøjelsgaard, Kristian, and Jens M Olesen. 2013. “Macroecology of Pollination  
 785 Networks.” *Global Ecology and Biogeography* 22 (2): 149–62.
- 786 Tylianakis, Jason M., and Rebecca J. Morris. 2017. “Ecological Networks Across  
 787 Environmental Gradients.” *Annual Review of Ecology, Evolution, and System-*  
 788 *atics* 48 (1): annurev-ecolsys-110316-022821. [https://doi.org/10.1146/](https://doi.org/10.1146/annurev-ecolsys-110316-022821)  
 789 [annurev-ecolsys-110316-022821](https://doi.org/10.1146/annurev-ecolsys-110316-022821).
- 790 Wisz, Mary Susanne, Julien Pottier, W Daniel Kissling, Loïc Pellissier, Jonathan  
 791 Lenoir, Christian F Damgaard, Carsten F Dormann, et al. 2013. “The Role  
 792 of Biotic Interactions in Shaping Distributions and Realised Assemblages of  
 793 Species: Implications for Species Distribution Modelling.” *Biological Reviews*  
 794 *of the Cambridge Philosophical Society* 88 (1): 15–30. [https://doi.org/10.](https://doi.org/10.1111/j.1469-185X.2012.00235.x)  
 795 [1111/j.1469-185X.2012.00235.x](https://doi.org/10.1111/j.1469-185X.2012.00235.x).
- 796 Wolin, Carole L., and Lawrence R. Lawlor. 1984. “Models of Facultative Mutu-  
 797 alism: Density Effects.” *The American Naturalist* 124 (6): 843–62. [https:](https://doi.org/10.1086/284320)  
 798 [//doi.org/10.1086/284320](https://doi.org/10.1086/284320).
- 799 Xiao Fu, Eugene Seo, Justin Clarke, and Rebecca A. Hutchinson. 2019. “Link  
 800 Prediction Under Imperfect Detection: Collaborative Filtering for Ecological

801 Networks.” *IEEE Transactions on Knowledge & Data Engineering*, no. 01  
802 (December). <https://doi.org/10.1109/TKDE.2019.2962031>.

803 Zurell, Damaris, Niklaus E. Zimmermann, Helge Gross, Andri Baltensweiler,  
804 Thomas Sattler, and Rafael O. Wüest. 2020. “Testing Species Assemblage  
805 Predictions from Stacked and Joint Species Distribution Models.” *Journal of*  
806 *Biogeography* 47 (1): 101–13. <https://doi.org/10.1111/jbi.13608>.

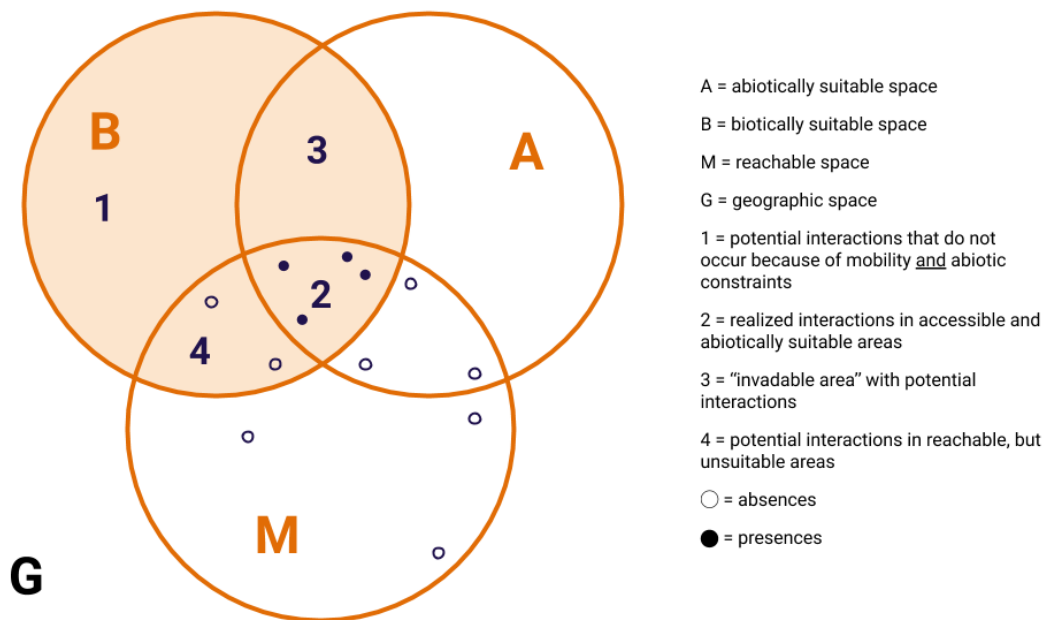


Figure 1: The “BAM diagram,” adapted from (Jorge Soberón 2007). Open circles are absences and solid circles are observed presences. Big circles correspond to the theoretical space of a species, regarding its biotic interactions (the B), the abiotically suitable space (the A) and the geographic area accessible to it (the M). These three aspects represent real points of occurrence on the real geographic space (the G). Ecological interactions act over this model in four ways: in (1), there are potential interactions that are never realized because of geographical and environmental constraints; in (2) interactions are realized on accessible, abiotically suitable areas; the space (3) is where the species could eventually go and establish new interactions, while (4) is the area where the occurrence of the species is limited only by abiotic factors.

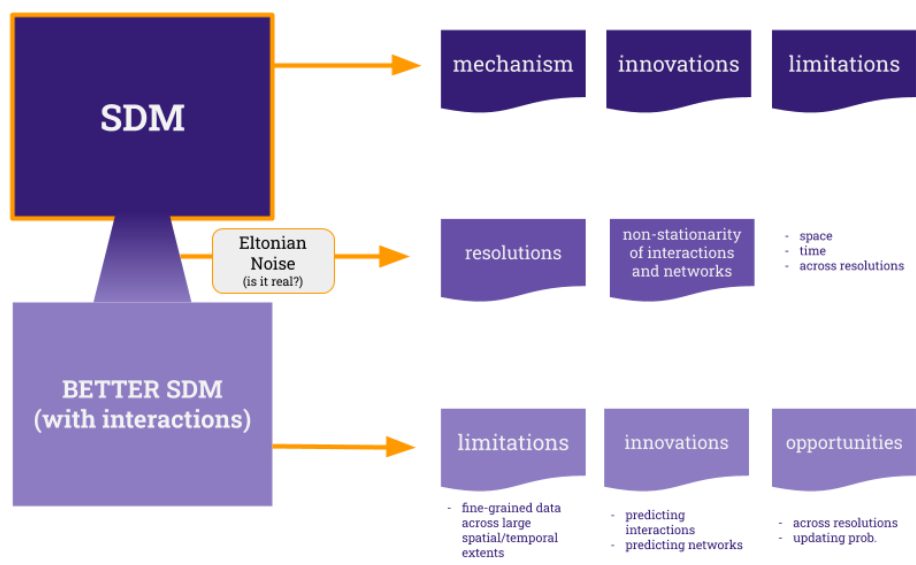


Figure 2: TODO

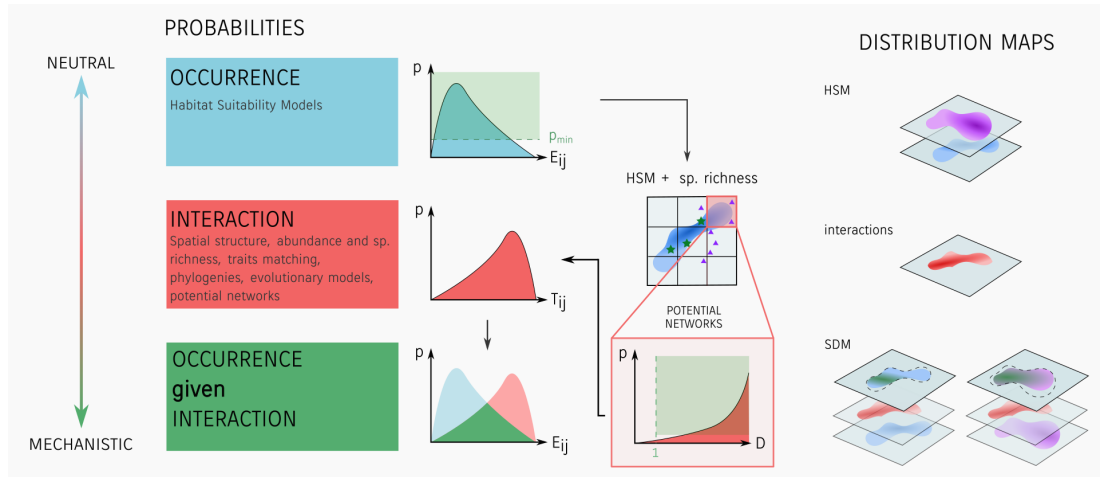


Figure 3: In the proposed workflow, probabilities of occurrence would be updated by the probabilities of new interactions where the Habitat Suitability Model suggests potential occurrence. For the sample sites where a species is predicted to occur, but did not occur before, the probability of this species to belong to a network where it has at least one link would be calculated based on the species richness of those cells. For the cells where the probability of having at least one link is higher than 0 (or a given threshold, depending on the system), the probability of interaction with important clades would then be assessed, and finally this probability updates the probability of occurrence.