UQAR **Université du Québec à Rimouski**

INFLUENCE DES CONTRAINTES ENVIRONNEMNENTALES, DES INTERACTIONS ET DE LA DISTRIBUTION DES TRAITS SUR LA RÉPARTITION GÉOGRAPHIQUE DES ESPÈCES

Thèse présentée

dans le cadre du programme de doctorat en Biologie

en vue de l'obtention du grade de Philosophiae Doctor

PAR

©KÉVIN CAZELLES

Septembre 2016

Composition du jury :

[Prénom Nom], président du jury, [Université d'attache]

[Dominique Gravel], directeur de recherche, [Université du Québec à Rimouski]

[Nicolas Mouquet], codirecteur de recherche, [Université de Montpellier]

[Nicolas Gotelli], examinateur externe, [Université d'attache]

Dépôt initial le [13 septembre 2016]   Dépôt final le [-]

UNIVERSITÉ DU QUÉBEC À RIMOUSKI

Service de la bibliothèque

Avertissement

[A JASCREB]

REMERCIEMENTS

La liste est longue des personnes qui ont soutenu plus ou moins directement.

TABLE DES MATIÈRES

LISTE DES ABRÉVIATIONS

SDM Modèle de distribution d'espèce (en référence terme anglais : *Species Distribution Model*)

TIB Théorie insulaire de la biogéographie

TTIB Théorie trophique de la biogéogrpahi des île (en référence terme anglais : *Species Distribution Model*)

RCP Representative Concentration Pathway

LISTE DES SYMBOLES

SYMBOLE 1 Ceci est la définition du symbole 1.

SYMBOLE 2 Ceci est la définition du symbole 2.

SYMBOLE 3 Ceci est la définition du symbole 3.

INTRODUCTION GÉNÉRALE

ARTICLE 1

A PROPOS DES INTERACTIONS BIOTIQUES ET DES CONTRAINTES ENVIRONNMNENTALES A L'ECHELLE BIOGEOGRAPHIQUE

1.1 Résumé

En 1967, Robert MacArthur et Edward Osborne Wilson publient leur théorie de la biogéographie des îles. Comme expliquer dans l'introduction, la TIB est toujours le support de nombreux travaux qui ont été repris plus récemment, en 2010, dans un livre édité par J. Losos et R. Ricklefs : *The Theory of Island Biogeography Revisited.* Ce livre souligne l'importance des travaux de R. MacArthur et E. O. Wilson et fait l'inventaire des questions qui restent à explorer. Parmi ces interrogations, on trouve celle qui porte sur le rôle des relations trophiques dans la théorie, développée au sixième chapitre par Robert D. Holt.

C'est précisement sur ce sujet que portent les travaux de D. Gravel et collègue présentés dans l'article *Trophic Theory of Island Biogeography* publiée dans *Ecology Letters* en 2011. Dans cet article, les auteurs montrent comment les résulats de la théorie classique sont modifiés par la prise en compte des liens écologiques unissant proies et prédateurs. Cet article est également le point de départ de mon premier artcle de thèse. L'objectif fixé était de 1- généraliser à tous types d'interaction le travail de Gravel et collègues et 2- introduire les contraintes environnementales afin de comprendre dans quelle mesure les prédicitons de la théorie classique étaient affectées.

Pour y parvenir, la clé de mon travail a été de considérer les espèces non pas une à une, mais de les considérer en assemblage. J'ai alors été capable de bâtir des probabilités de survie qui étaient dépendantes du réseau écologique présent sur l'île. De même, les pobabilités de colonisation des espèces du continent ont été reliées aux conditions environemntales de

l'îles. Après avoir montré et développé comment le modèle a été construit et donné des prédicitons simples, nous nous sommmes intéressés à des scénarios portant sur 10 espèces et pour des types d'interactions différents : mutualistes, prédation et compétitions le long de gradients environnemtaux. Ce qui apparait ressort de mos simulations est un portrait des impacts potentielles des interactions sur la distributions des espèces. Dependemment de leur nature et de leur nombre, les interactions peuvent changer drastiquement la biodiversité attendue dans le cadre de le théorie classique. Cela pourrait avoir des conséquences majeures sur nos prévisions de richesse spécifique dans le contexte actuel des changements globaux.

Le travail réalisé a donné lieu à un article intitulé "*On the integration of biotic interaction and environmental constraints at the biogeographical scale*". Il fut accepté pour publication au printemps 2015 dans le journal *Ecography*. La conception de l'article est le résultat de nombreux échanges entre les quatre auteurs de l'article. J'ai dévelopé le modèle et l'ensemble des scripts pour aboutir aux résultats finaux. Dominique Gravel a supervisé l'ensemble des étapes et est devenu le dernier auteur. David Mouillot et Nicolas Mouquet ont grandement contribué à la rédaction du manuscrit. Aussi sur le forum...

1.2 Titre

On the integration of biotic interaction and environmental constraints at the biogeographical scale.

1.3 Auteurs

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

1.4 Résumé en anglais

Biogeography is primarily concerned with the spatial distribution of biodiversity, including performing scenarios in a changing environment. The efforts deployed to develop species distribution models have resulted in predictive tools, but have mostly remained correlative and have largely ignored biotic interactions. Here we build upon the theory of island biogeography as a first approximation to the assembly dynamics of local communities embedded within a metacommunity context. We include all types of interactions and introduce environmental constraints on colonization and extinction dynamics. We develop a probabilistic framework based on Markov chains and derive probabilities for the realization of species assemblages, rather than single species occurrences. We consider the expected distribution of species richness under different types of ecological interactions. We also illustrate the potential of our framework by studying the interplay between different ecological requirements, interactions and the distribution of biodiversity along an environmental gradient. Our framework supports the idea that the future research in biogeography requires a coherent integration of several ecological concepts into a single theory in order to perform conceptual and methodological innovations, such as the switch from single-species distribution to community distribution.

1.5 Introduction

Biogeography is concerned with the description of the distribution of biodiversity and understanding its underlying processes. The discipline is central to the simulation of future scenarios of biodiversity under climate change (Thuiller et al., 2013). The extensive development of statistical models of species distributions based on actual ranges and environmental data have provided valuable knowledge and predictions (Kearney and Porter, 2004), but often remain purely correlative. There is now consensus that future developments in biogeography will require solving critical limitations of species distribution models (Kissling et al., 2012)

and incorporating explicitly biotic interactions and dispersal (Lavergne et al., 2010). This effort must be supported by theory in order to guide model development, maintain tractability and manage complexity. Developing a mechanistic theory of species distribution will require an integration of three fundamental principles and their interplay (Thuiller et al., 2013): 1) how local and regional dynamics are linked, 2) how species interact with the abiotic environment and 3) how they are embedded in a network of biotic interactions. Each of these principles are discussed in detail below.

A cornerstone of biogeography is the recognition of the contribution of regional-scale processes such as disturbances, historical contingencies (e.g. macro evolutionary history or glaciations) and dispersal limitations to local community dynamics (Ricklefs, 1987). The metacommunity concept has been proposed as a simple framework to link different spatial scales in ecology (Leibold et al., 2004). It emphasizes reciprocal feedbacks between local scale processes, such as competitive interactions and local adaptation, and regional scale processes such as dispersal, gene flow, and speciation. A central concept of metacommunity ecology is the idea that local communities are highly dynamic owing to colonization events and local interaction, resulting in a spatial mosaic of assemblages sampled non-randomly from the regional species pool. As the concept matures there are new themes emerging, such as the investigation of evolution in metacommunities (Urban et al., 2008), and spatial food webs (Massol et al., 2011; Gravel et al., 2011). The field provides remarkable concepts and tools to build an integrated theory for biogeography.

Species distribution is also constrained by physiological requirements, which is at the core of the niche concept (Peterson et al., 2011). The niche is usually defined as a N-dimensional environmental and resource hyper-volume within which a species is able to maintain a viable population over the long term (Chase and Leibold, 2003). Recent developments refined this definition based on demography and metapopulation dynamics (Holt and Barfield, 2009). The abiotic niche, often referred as the Grinnelian niche, has been central to the development of species distribution models (SDMs, Jeschke and Strayer, 2008). De-

spite all of its criticisms, SDMs remain remarkably popular and operational for conservation ecology (Guisan et al., 2013). Recent attempts to improve the quantification of the niche include the addition of experimental assessments of the fundamental physiological constraints, as well as dispersal and proxies of biotic interactions (Boulangeat et al., 2012). The search for the most adequate set of environmental variables explaining diversity should be continued despite criticisms of the actual SDMs, and most of all must constitute a central principle of a general theory for biogeography.

Finally, species are not isolated, they are embedded within complex networks of ecological interactions. While interactions define community ecology, they are less informative for biogeography (Peterson, 2003). Theory predicts that interactions in small community modules (2-4 species) should influence range limits (Gilman et al., 2010), but there is no extension to highly diverse communities. It has been hypothesized that factors determining distribution are hierarchical, such that climate would govern the distribution at the regional scale while biotic interactions would be more important at the local scale (Araújo and Rozenfeld, 2014). However an increasing number of studies emphasizes the role of local interactions as a major factor influencing geographical ranges (Jabot and Bascompte, 2012; Gotelli et al., 2010). The representation of interactions in a network is a convenient method to summarize the type and strength of interactions among species, their organization (Proulx et al., 2005) and their consequences on dynamics (Allesina and Tang, 2012). Food webs were first considered in the development of a trophic theory of biogeography (Gravel et al., 2011), where it was shown that a diversity of interactions enhance persistence. Networks are however more than food webs and are rarely made of a single type of interaction (Kéfi et al., 2012). Mutualism, competition and indirect effects (Wootton, 1994), for instance, also impact local environmental suitability (Godsoe and Harmon, 2012). Tools and knowledge acquired through the study of local ecological networks, such as the community matrix and metrics of structure (Allesina and Tang, 2012), must now be incorporated into a theory for biogeography.

These three principles should be all mixed together to provide an integrated assessment

of their relative contribution to species distribution. To do so, the theory of Island Biogeography (hereafter referred as TIB) (MacArthur and Wilson, 1967; Warren et al., 2015) is a convenient starting point. The TIB describes variations of species richness among islands as a dynamic equilibrium between two opposite processes, colonization and extinction, directly linked with island characteristics. The TIB is a metaphor that goes beyond the intrinsic interest of islands; it serves as a first approximation to understanding the assembly of local communities embedded in a metacommunity context with straightforward species flux. The simplicity of the model and the relevance of its predictions demonstrate after more than 50 years since its publication it is still a useful tool in ecology and conservation (Cook et al., 2002; Warren et al., 2015). The TIB emphasizes the role of regional processes to local community assembly. Indeed it can be regarded as the simplest representation of metacommunity dynamics (Leibold et al., 2004). Furthermore, the model is easily expandable. Following (Holt and Barfield, 2009), Gravel et al. (2011) introduced trophic interactions in the TIB (hereafter the trophic TIB, TTIB;). Species interactions were found to be a key factor to understand species distributions, as the probability of finding any species in a locality increases with the generality of its diet and decreases with trophic rank.

We propose to generalizes of the TIB by integrating the three principles described above. The TIB already explicitly includes the effect of regional processes (colonization and extinction dynamics) on local community assembly, and the TTIB includes predator-prey interactions. We extend this framework to all potential interactions, thus resulting in a general model of metacommunity dynamics, akin to the Lotka-Volterra equations for local community dynamics. We also incorporate abiotic constraints on colonization and extinction dynamics. Hence we integrate the ingredients we believe are essential to model biodiversity distribution at the biogeographical scale. With this model in hand we then describe species distribution along environmental gradients. We use the mathematical formalism of Markov Chains (Kemeny and Snell, 1983; Black and McKane, 2012) to derive expected assemblages and co-distribution at both the local and the regional scale. We illustrate how the interplay between biotic interactions and environmental requirements can affect the distribution of bio-

diversity over environmental gradients. Our results support the idea that the future research in biogeography require a consistent integration of several ecological concepts into a single framework to build promising approaches such as the switch from single-species distributions to community distributions.

1.6 The model

1.6.1 A simple probabilistic biogeographical model

The challenge of adding species interactions within the classical model of the TIB is gaining generality without losing simplicity. Following MacArthur and Wilson's theory, we model the dynamics of occurrence probability of a species i in a local community. Species occurrence is the result of a balance between colonization and extinction dynamics, which occur at rates c_i and e_i respectively,. Local species richness is given by the sum of occurrence probabilities over all species of the regional species pool P, here simply defined as the set of all species whose propagules (as defined in Simberloff and Wilson, 1969) can land on the island considered. The model thereby takes into account local (extinction) and regional (colonization) processes. More precisely, the dynamics of occurrence probability of species i, p_i, follows:

$$\frac{dp_i}{dt} = c_i(1 - p_i) - e_i p_i \tag{1.1}$$

Here, c_i and e_i are constant and a property of species i. In this widespread version of the TIB, also called the linear version of the TIB (Schoener, 2010), the equilibrium occurrence probability of a species i is given by $p_{i,eq} = \frac{c_i}{e_i + c_i}$. Also, species are assumed to be independent, therefore, the richness S_{eq} is given by the sum of the P different $p_{i,eq}$. The linear TIB can be modified to include trophic interactions (after Gravel et al., 2011) and we propose to extend it to all types of interactions. To reach that goal, the first step is to find a way to derive

the expected species composition at any time. This composition can actually be depicted at any time by a vector of P zeroes and ones indicating, respectively, presences and absences of each species considered, these combinations will be referred as assemblages. Following Mac-Arthur and Wilson, we use a stochastic modelling approach to describe the dynamics of assemblages. The simplest scenario is the one species case. Here there are only two assemblages for the locality: one with species i present and the other without. Let X_i be a random variable describing the occurrence of species i. When species i is present in the locality, X_i is 1, when it is absent X_i is 0; X_i is then a Bernoulli variable. We define this random variable at any time t which describes a stochastic process we denote $\mathbf{X_{i,t>0}}$. The occurrence probability of species i at time $t + dt$ (dt being a very small time step) is then given as follows:

$$
\begin{aligned}
\mathbb{P}(X_{i,t+dt} = 1) &= \mathbb{P}(X_{i,t+dt} = 1|X_{i,t} = 1)\mathbb{P}(X_{i,t} = 1) \\
&\quad + \mathbb{P}(X_{i,t+dt} = 1|X_{i,t} = 0)\mathbb{P}(X_{i,t} = 0)
\end{aligned}
\tag{1.2}
$$

$\mathbb{P}(X_{i,t+dt}|X_{i,t})$ is the conditional probability describing $X_{i,t+dt}$ stating $X_{i,t}$. As $X_{i,t+dt}$ solely depends on $X_{i,t}$ (not on other earlier time steps) we have a discrete-time Markov chain. In this process, species i will be present in a locality at time $t + dt$ if it was already present at time t and persisted (meaning it did not go extinct, with probability $(1 - e_i dt)$, or if it was absent and colonized the community from the mainland (with probability $c_i dt$). Note that dt is small enough to get $0 < c_i dt < 1$ and $0 < e_i dt < 1$. Hence, equation (1.2) becomes:

$$
\mathbb{P}(X_{i,t+dt} = 1) = c_i dt \mathbb{P}(X_{i,t} = 0) + (1 - e_i dt)\mathbb{P}(X_{i,t} = 1)
\tag{1.3}
$$

This equation leads to (1.1) when dt tends to zero. This formulation keeps the simplicity of the original MacArthur and Wilson model, but can also more generally be used to consider the probability of any given assemblage. $\mathbb{P}(X_{i,t+dt}|X_{i,t})$ defines the rules to switch from one assemblage to one another during the interval dt. There are P occurrence probabilities we gather within $\mathbf{Y_{t>0}} = (\mathbf{X_{1,t>0}}, \mathbf{X_{2,t>0}}, ..., \mathbf{X_{P,t>0}})$ which leads to the description of 2^P assem-

blages depicted by a given collection of zeros and ones. The conditional probabilities provide the transition from one local assemblage k to any other l during dt. For any species i there are only four possible cases: at time t either species i is locally absent and colonizes the locality (I_1) or not (I_2) during dt, either species i is present and goes extinct (I_3) or survives (I_4) during dt. The conditional probabilities between two communities states (l and k) can then be simply derived from these four probabilities:

$$\mathbb{P}(\mathbf{Y_{t+dt}} = \text{"state k"}|\mathbf{Y_t} = \text{"state l"}) \quad = \quad \prod_{i_1 \in I_1} c_{i_1} dt \prod_{i_2 \in I_2} (1 - c_{i_2} dt)$$
$$\prod_{i_3 \in I_3} e_{i_3} dt \prod_{i_4 \in I_4} (1 - e_{i_4} dt) \qquad (1.4)$$

We now apply the complete probability formula as defined in (1.2) to get the probability of observing one assemblage at $t + dt$ given its state at t. This is where the main benefit of Markov chain models is: it allows us to derivate exact solutions for the probabilities for assemblages, instead of a set of independent occurrence probabilities for each species. This approach is promising for building joint species distribution models (see Discussion). This property will be fully explored in the next section to include interactions.

Consider as an example a pool of two species ($P = 2$) for which we find four assemblages: at any time t, a locality can contain either two species ($X_{1,t} = 1, X_{2,t} = 1$), only one species ($X_{1,t} = 1, X_{2,t} = 0$) and ($X_{1,t} = 0, X_{2,t} = 1$), or none of them ($X_{1,t} = 0, X_2 = 0$). The transition from one local assemblage to another is then easily obtained. Table 1 presents these conditional probabilities (application of (1.4)). This is actually the transition matrix of a Markov chain we solve (by calculating one eigen value, see below). To illustrate the dynamics expected in TIB from our assemblage point of view, we simulate the model as follows: $c_1 = c_2 = 0.15$, $e_1 = e_2 = 0.05$, $\mathbb{P}(X_{1,0} = 0, X_{2,0} = 0) = 0.6$ and $\mathbb{P}(X_{1,0} = 1, X_{2,0} = 0) = 0.4$, so species 2 is absent at time $t = 0$. Just as for the single species situation, the probabilities of observing each community tend to an equilibrium (Fig.1, panel A). By summing the previous probabilities where a given species (1 or 2) is present (the conditional probabilities) we get

its overall occurrence probability (marginal probability, Fig.1-B). Finally, we can calculate the expected number of species in a locality (Fig.1-C), in agreement with the TIB. Interestingly, this calculation is often achieved in the other way. Firstly, the presence probability of all species are computed: $\mathbb{P}(X_i) = \frac{c_i}{c_i + e_i}$. Then the richness is obtained under the assumption that species are independent and so $P(X_i, X_j) = \mathbb{P}(X_i)\mathbb{P}(X_j)$. We show below that occurrence probabilities of each assemblage is a key to introduce interactions among species.

1.6.2 Integrating biotic interactions

We start by representing the interaction network by a community matrix $\mathbf{A}$ of P species that we incorporate into the Markovian TIB chain model. The elements $\alpha_{i,j}$ of $\mathbf{A}$ quantify the effect of species j on the dynamics of species i. We first consider that interactions could alter both the colonization and the extinction probabilities (Gravel et al., 2011). When $\alpha_{i,j}$ is negative, the colonization probability of species i decreases and/or its extinction probability increases when j is found locally. Inversely, when $\alpha_{i,j}$ is positive, the colonization probability increases and/or the extinction probability decreases. Note that diagonal elements provide the extinction probability per time unit when no other species is present.

$(\mathbf{X_{1,t}}, \mathbf{X_{2,t}})$	$(\mathbf{X_{1,t+dt}}, \mathbf{X_{2,t+dt}})$
	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$
$(0,0)$	$(1-c_1 dt)(1-c_2 dt)$	$(1-c_1 dt)c_2 dt$	$c_1 dt(1-c_2 dt)$	$c_1 dt c_2 dt$
$(0,1)$	$(1-c_1 dt)e_2 dt$	$(1-c_1 dt)(1-e_2 dt)$	$c_1 dt e_2 dt$	$c_1 dt(1-e_2 dt)$
$(1,0)$	$e_1 dt(1-c_2 dt)$	$e_1 dt c_2 dt$	$(1-e_1 dt)(1-c_2 dt)$	$(1-e_1 dt)c_2 dt$
$(1,1)$	$e_1 dt e_2 dt$	$e_1 dt(1-e_2 dt)$	$(1-e_1 dt)e_2 dt$	$(1-e_1 dt)(1-e_2 dt)$

Table 1: **Conditional probabilities between potential assemblages.** At any time t we calculate all the possible conditional probabilities between the four potential assemblages for a two species regional pool. These probabilities are derived by multiplying probabilities of single species events defined in equation (1.4). By doing so, we build the transition matirx of our markov chain where species are assumed to be independent. We release this hypothesis by linking extinction coefficients and species assemblages.

The elements of the community matrix $\mathbf{A}$ represent the pairwise effects of ecological interactions on transition probabilities. To account for the cumulative effects of local interactions on transition probabilities, we make colonization and extinction probabilities community dependent. As explained above, at a time t, the $\mathbf{Y_t}$ vector gives the local assemblages. We calculate the sum of interactions at any time and for each species as $\mathbf{v} = \mathbf{AY_t}^T$ (where ^{T} denotes the transpose operator). Our approach can be interpreted as a spatial analogue to the generalized Lotka-Volterra model because it takes into account the impact of the whole network of interactions on each species dynamics and can deal with any type of interaction. We denote the coefficients of $\mathbf{v}$ by v_i, they are species-specific parameters (weighted by parameter d_i) of two species-specific functions: f_i and g_i, respectively, standing for extinction and colonization probabilities for species i. Note that at this stage we do not define any specific function relating interactions to colonization (f_i) and extinction probabilities (g_i), to keep the description of the model general (see below for some proposed functions). At each time step, the local community composition impacts: i) the colonization probability of species present in the regional pool but absent from the local community, and ii) the extinction probability of species present on the local community.

If we expand the two species example (labelled 1 and 2, Table 1), according to the general model, we define two f functions (f_1 and f_2) linking interaction and extinction and two g functions linking extinction and colonization (g_1, g_2). At this stage, to reduce the model's complexity, we consider that interactions solely impact extinction probabilities. This assumption is reasonable if we consider that local interactions impact mostly demography (possibly leading to extinction) and that colonization success solely depends on the first propagule (interactions occur after arrivals). Therefore g_1 and g_2 are constant functions, respectively,, returning c_1 and c_2. The functions f are assumed to have a sigmoid shape (1.5). There are many reasons such a function is of interest: 1) we get a clear link with the basic extinction probability, i.e. e_i for an interaction strength of 0; 2) we define both a minimum and a maximum extinction probability; 3) the first interactions to occur are the most influential (Gravel et al. (2011) considered that at least one interaction was required to persist, which is very

similar).

$$
\begin{aligned}
f_i(\mathbf{v}) &= f(\mathbf{v}, (e_i, e_{i,min}, e_{i,max}, d_i)) \\
&= e_{i,min} + \frac{1}{\frac{1}{e_{i,max}-e_{i,min}} + \left(\frac{1}{e_i-e_{i,min}} - \frac{1}{e_{i,max}-e_{i,min}}\right)\exp(d_i * v_i)} \quad (1.5)\\
g_i(\mathbf{v}) &= c_i \quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad (1.6)
\end{aligned}
$$

To illustrate how interactions modify occurrence probabilities, we simulate the model for two networks: A_1 where all interactions are negative and A_2 where they are all positive. We consider null diagonal elements for both networks. Consequently, there is no difference with the model without interaction when one species is alone in the locality. Simulation results are presented at Figure 2. Panel A presents the functions f_1 and f_2 we chose for our two species example. For networks A_1 and A_2, we show how interactions alter the probabilities of observing different assemblages (respectively, Fig.2-B and Fig.2-C). The assemblage with both species present (solid red lines) is the most affected by interactions, switching from an occurrence probability of 0.2 (for negative interactions) to 0.8 (for positive interactions). Positive interactions enhance, as expected, co-occurrence while negative interactions prevent species from being found on the same island. Consequently, occurrence probabilities of single species states are lower in A_2 than in A_1. According to a defined network, occurrence probabilities of the different assemblages are then modified, which affect the expected species richness (Fig.2-D).

1.6.3 Integrating environmental gradients

We now introduce the effect of abiotic conditions, such as climatic variables, on transition probabilities. We denote the vector of n environmental conditions by $\mathbf{w}$: $\mathbf{w} = (w_1, w_2, ...w_n)$. We first assume that physiological constraints can affect both colonization and extinction probabilities through the functions f_i and g_i (affecting, respectively,, extinction and colonization rates). Again the model in its general formulation does not presume any shape for these
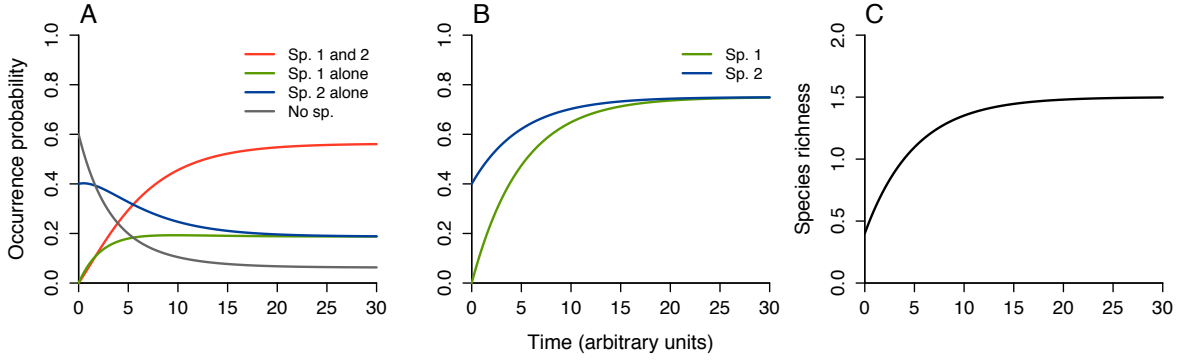
Figure 1: **Dynamics of the community assembly.** As a direct development of MacArthur and Wilson model, we simulate for two species, the dynamics of the four communities states with different initial conditions associated (A). By summing every states where one given species is present we get the occurrence probability of two considered species (B). Finally by summing the four states probabilities weighted by their species richness, we get the classical model of MacArthur and Wilson (C). The calculation of (B) and (C) does not require species being independent while classical approaches focus on (B) to derive (A) and (C) under this assumption of independence.

functions. We now have all the ingredients of an integrated model of biogeography as the transition probabilities at a location depend on 1) species-specific colonization and existence probabilities, 2) the network of interactions, 3) local community composition, and 4) local environmental conditions. In the general formulation of the model, functions f_i and g_i are functions of multiple variables ($\mathbf{v}$ and $\mathbf{w}$).

At any time t, for a regional pool of P species among which interactions are summarized by the community matrix $\mathbf{A}$, in an environment characterized by $\mathbf{w}$, we can derive all transition probabilities. These constitute a transition matrix of a Markov chain that we denote $\mathbf{M}(\mathbf{v}, \mathbf{w})$. Its elements, $\mu_{k,l}(\mathbf{v}, \mathbf{w})$, give the probability the locality in assemblage k turns into assemblage l (left side of equation (1.4)):

$$\mu_{k,l}(\mathbf{w}, \mathbf{v}) = \prod_{i_1 \in I_1} g_{i_1}(\mathbf{v}, \mathbf{w})dt \prod_{i_2 \in I_2} (1 - g_{i_2}(\mathbf{v}, \mathbf{w})dt)$$
$$\prod_{i_3 \in I_3} f_{i_3}(\mathbf{v}, \mathbf{w})dt \prod_{i_4 \in I_4} (1 - f_{i_4}(\mathbf{v}, \mathbf{w})dt) \tag{1.7}$$

Note that the dimension of $\mathbf{M(w)}$ will increase as a power of the number of species P and thus can rapidly becomes large. Let $\mathbf{C_t}$ be the line vector of the probability of observing each assemblage, defined by: $\mathbf{C_t} = (\mathbb{P}(\mathbf{Y_t} = \text{"state 1"}), \mathbb{P}(\mathbf{Y_t} = \text{"state 2"}), ..., \mathbb{P}(\mathbf{Y_t} = \text{"state } 2^{P}\text{"}))$. The Markov Chain formalism defines the probability of the future community composition at time $t + dt$ as $\mathbf{C_{t+dt}} = \mathbf{C_t M}$. $\mathbf{C_t}$ asymptotically reaches the $\mathbf{C_{eq}}$ after a certain number of time steps. $\mathbf{C_{eq}}$ is given by the normalized left eigenvector associated to the first left eigenvalue.

$$\lim_{\substack{l \to +\infty \\ l \in \mathbb{N}}} \mathbf{C_0 M}^l = \mathbf{C_{eq}} \tag{1.8}$$

$\mathbf{C_{eq}}$ contains the probability of all assemblages at the equilibrium. The occurrence probability of a given species, is provided by the sum of all probabilities of assemblage where that species is present. The richness at the equilibrium S_{eq} is the sum of $\mathbf{C_{eq}}$ elements weighted by the number of species found in the associated assemblages.

For the sake of illustration, we further reduce the complexity of our model. We have previously removed the interactions ($\mathbf{v}$) from colonization (g) functions; we now state that extinction does not depend on environmental variables and so we remove the abiotic environment ($\mathbf{w}$) from extinction functions (f). This can be interpreted as the effects of the abiotic environment on extinction rate being included within e_i (i.e. extinction rate without interaction). Furthermore, we assume solely one environmental variable and a Gaussian shape for g_i functions (1.9). A simple function with a clear optimum and very low colonization for extreme environment values is.

$$g_i(w_1) = g(w_1, (c_i, h_i, r_i)) \quad = \quad c_i * exp\left(-\left(\frac{w_1 - h_i}{r_i}\right)^2\right) \tag{1.9}$$

This enables us to define an environmental optimum (h_i), a colonization probability per time unit (c_i) and also suitable range (r_i) for each species. Figure 3 presents the interplay between the three components of the integrated biogeographical model. The chosen functions for the environment-colonization relationship are illustrated in Panel A. For the two previous

networks (A_1 and A_2; illustrated in Fig 2) we now compute the probabilities of observing the different assemblages at equilibrium, along the environmental gradient (Panel B and C). When interactions are negative (network A_1), species repulse each other and rarely co-occur, whatever the environment is. Most of their occurrence follow their abiotic niche (blue and green lines) as they are barely found together. Inversely, when interactions are positive (for A_2 network) they often co-occur where their abiotic niches overlap, thereby decreasing the probability of an empty community (Panel D, solid grey line). Finally, we present how interactions modify the resulting community composition along the environmental gradient (Panel D). Species richness is constrained by the distribution of abiotic niches and the sign of the interactions. As expected, the role of interactions is strongest when abiotic niches largely overlap.

1.7 Exploring the model

In our exploration, we choose a regional pool P of 10 species to keep the number of assemblages reasonable ($2^{10} = 1024$) and to numerically compute the exact solution of the equilibrium distribution $\mathbf{C_{eq}}$. We consider four types of interaction matrices $\mathbf{A}$. The first situation corresponds to the classical MacArthur and Wilson model, where the $\mathbf{A}$ matrix is null (no interactions). For the three other scenarios we generate random matrices with fixed connectance (number of existing links divided by the number of potential links). The coefficients within $\mathbf{A}$ are drawn uniformly within [0, 1] and the sign of the interaction is determined by the action of one species on another, for instance, a predator has a negative impact on its prey leading to a negative α coefficient; in return, a prey has a positive effect on its predators. The intensity of the interaction is then determined by the d coefficient of extinction functions (see equation (1.8)). We assume that the distribution of the links are given by the niche model (Williams and Martinez, 2000). This model is simple and provides relevant random food webs with the same number of positive and negative interactions. For the two last scenarios, we keep the rules to distribute the links, but turn all the coefficients in $\mathbf{A}$ positive to
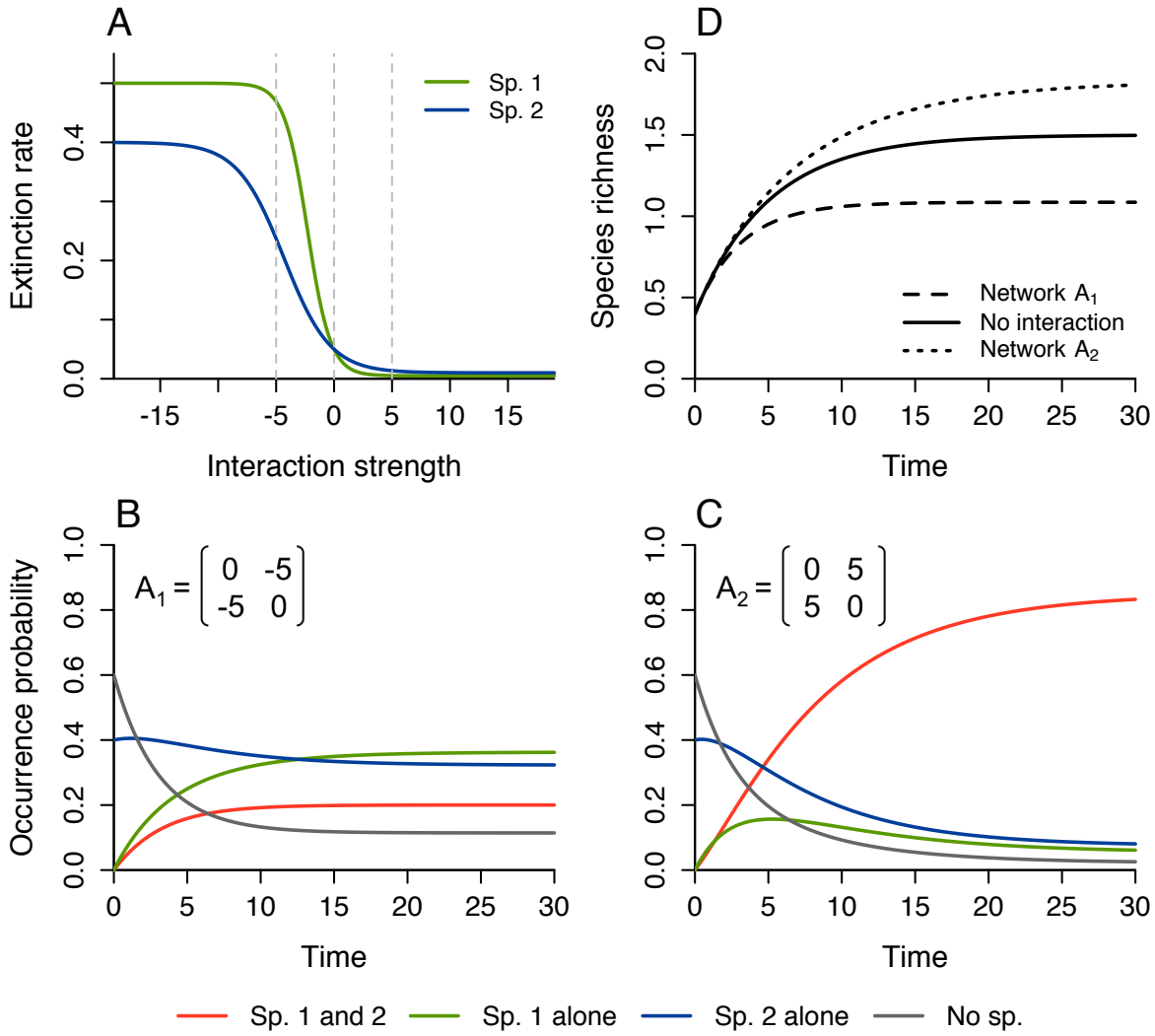
Figure 2: **Effects of biotic interactions on colonization-extinction dynamics.** For any species i, the extinction probability e_i is related to the strength of the interaction as shown in (A). The intersections of extinction curves with the grey dotted lines indicate the potential values of e_i according to the different biotic context (A_1, A_2 and without interaction). We set the other parameters as follows: $c_1 = c_2 = 0.15$, $\mathbb{P}(X_{1,0} = 1, X_{2,0} = 0) = 0.4$, $\mathbb{P}(X_{1,0} = 0, X_{2,0} = 0) = 0.6$, $\mathbb{P}(X_{1,0} = 0, X_{2,0} = 1) = \mathbb{P}(X_{1,0} = 1, X_{1,0} = 1) = 0$. We then simulate the model for two simple networks A_1 and A_2 and present community assembly dynamics associated ((B) and (C)). Finally we compare the expected species richness on the locality (D) for our two networks and for the case without interaction which corresponds to the widespread linear version of the TIB.

generate a mutualism network, or negative for competition networks. Although these basic structures with exclusive interaction types are not realistic, they facilitate comparison among results. Hence, the scenarios simply differ by the sign distribution within the matrix **A**: (i) no interaction **A** is null, (ii) predation mixes both signs "+/-", (iii) mutualism only "+", (iv)- competition, only "-". With these scenarios in hands, we 1) present the assemblages probabilities associated with a given level of species richness and 2) we look at the species richness expected along an environmental gradient. For all figures presented hereafter we used 1000 randomly-generated **A** matrices.

Assemblage probabilities

First, we illustrate how interactions affect richness of species assemblages. To do so, we build the Markov chains for all the 1000 **A** matrices generated (connectance set to 0.2) and we calculate the vector **C$_{eq}$**. This is a vector of 1024 occurrence probabilities (as we consider 10 species). Then we sum all the probabilities that correspond to assemblages of the same richness. We do so for three values of d coefficient (0.1, 1 and 10); that is, we look at how the strength of interaction affect community richness predictions. Figure 4 presents the results of such investigation, with Panels A to C corresponding to the results for the three different values of the d parameter.

As expected, positive interactions increase local species richness by diminishing extinction probabilities, while negative interactions weaken large communities (see the contrast between blue and red symbols on Fig.4). This is stressed as interaction strengths increase, that is for increasing values of d. Indeed, when d is low, there is almost no difference among scenarios because interactions do not impact strongly colonization and extinction dynamics; occurring species can be regarded as mostly independent. All scenarios converge to the classical TIB scenario (no-interaction, grey symbols), the resulting species richness distribution is binomial (here for all species $p_{i,eq} = 0.5$ as $c_i = e_i = 10^{-5}$). Differences between interaction types increase with d. Species rich mutualistic communities are more likely to

occur since positive interactions tends to promote co-occurrence. Therefore species occurrence can be dramatically affected by the strngth of interactions: for $d = 10$ (Panel C in Fig. 4), the species richness is 9.46 for positive interactions (red symbols), 2.24 for the negative ones (blue symbols) and 5 without interactions. When positive and negative interactions are mixed (our predation scenario, green symbols on Fig.4), it seems that the negative effect of predators on their prey prevails and so predation reduces species richness, but less than for competitive networks.

As we introduce variability through the use of randomly-generated matrices, we also compute the standard deviation associated with occurrence probabilities. The variability is provided as the coloured vertical bars found in Fig. 4 which stand for 50% of the total standard deviation. Clearly, variability increases with (i) the strength of interaction and (ii) the occurrence probability. Although this can simply reflect the variability of values found in **A** matrices, this could potentially be caused by the variability of the location of non-zero values in **A** matrices; that is, the structure of the networks we use.

1.7.1   Biodiversity distribution over environmental gradients

In this section, we introduce an environmental gradient to emphasize the interplay between interactions species-specific requirements along an environmental gradient. Our environmental gradient takes values from 0 to 30, for each of them we calculate the expected species richness associated to all scenario. To do so, we start by computing the colonization functions (g_i functions): species optima h_i are drawn from a uniform distribution from the range $[10, 20]$ and the widths of the abiotic niches are kept constant for all the simulations $r_i = 5$. Then we build the Markov chains for the different values of the environmental gradient and for the different **A** matrices. Again, we derive the vector $\mathbf{C_{eq}}$ and we sum its elements, i.e. occurrence probabilities of assemblage community, weighted by the species richness to which they refer. We repeat the procedure for an increasing value of connectance of **A** matrices: from 0 to 0.4. For this section, the parameter d is set to 10, also extinction paramaters
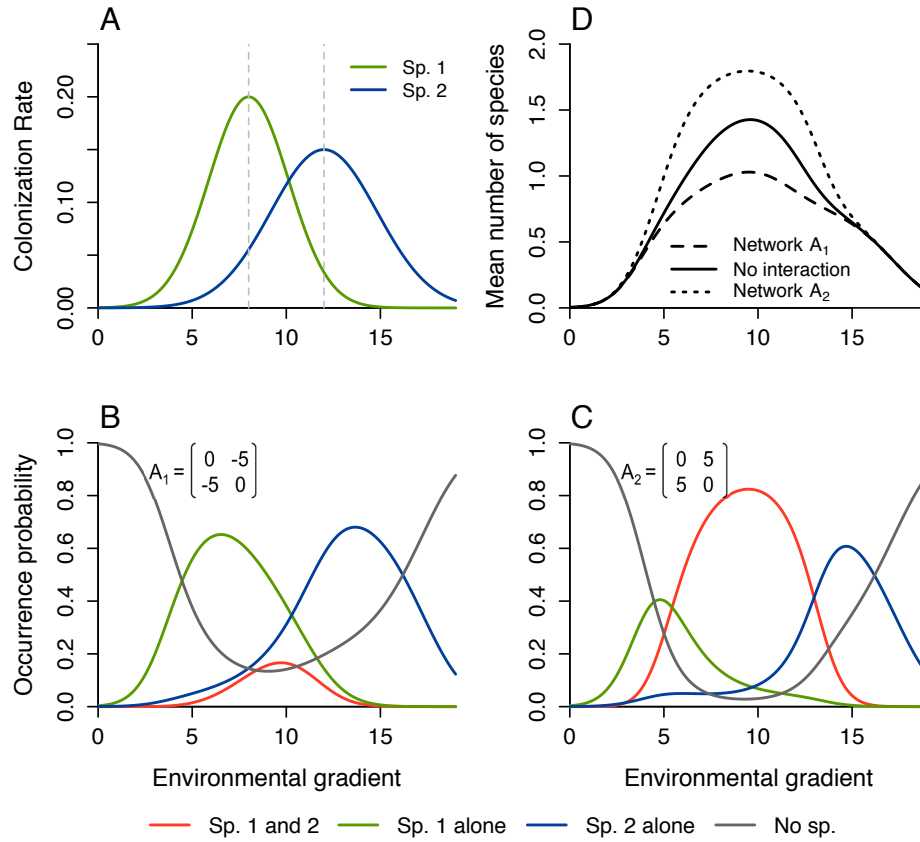
Figure 3: **Equilibrium for interacting species along an environmental gradient.** The colonization probability of species i, c_i, is related to the environment variable **w** according to species-specific requirements (A). The intersection of the colonization curve of species i with the grey dotted lines represents the value of c_i associated with its environmental optimum h_i. We compute equilibrium occurrence probabilities for the different assemblages along the environmental gradient, for the networks A_1 with negative interactions (B) and A_2 with positive one (C). We calculate the expected species richness on the locality for the two networks and without interaction (D).

are set as follows: $e_i = 10^{-5}$, $e_{i,min} = 10^{-3}e_i$, $e_{i,max} = 10^3 e_i$ and $c_i = 10^{-5}$. Like so we obtain the profile of species richness we report on Figure 5.

For all scenarios, the richness is maximal at the center of the environmental gradient (Fig. 5). This is due to the distribution of species optima in the range $[10, 20]$. Also this is the range of environmental values for which the effect of interaction are the most important. Indeed, the higher the colonization probabilities, the higher interactions occur, therefore,

interactions strongly impact species richness for favourable abiotic conditions. We also find that changes in species richness increase with connectance, as depicted by the colour of the solid lines for the three panels of Fig. 5: from black (without interaction) to the lightest blue (connectance set to 0.4).

Species richness is inversely related to connectance when interactions are negative (Panel A in Fig. 5). Moreover, when abiotic conditions are favourable, the number of species expected tends to 1. At the centre of the gradient, even though colonization probabilities are maximal, many species colonize but likely go extinct because of competition. We expect the locality to be most often occupied by species that are not affected by competition. Alternatively, in the case of positive interactions (Panel B in Fig. 5), the expected species richness is strongly enhanced by interactions even for low connectance. The expected species richness tends to reach the total number of species from the most favourable to semi-harsh abiotic conditions. As the connectance increases the Gaussian shape of the richness profile turns into a hat shape, which has one major consequence: from favourable to semi-harsh conditions, the species richness is maintained thanks to positive interaction, but it also quickly collapses as the environment becomes slightly harsher.

Finally, when positive and negative interactions are mixed, the higher the connectance, the flatter the richness profile (Panel B in Fig. 5). The expected species richness declines as connectance increases but far less than it does for negative interactions only. We think this is caused by the colonization of numerous prey that promote the survival of predators which in turn prevent assemblages to be as large as they can be without interaction (as predators reduce the persistence of prey). Conversely, from harsh to intermediate environmental conditions, mixed sign interactions positively affect the species richness. We explain this as the consequence of the benefit predators take from the preys presence. Assemblages with few predators, promoted by positive effect of the prey on their predators, may be relatively stable. Since colonization is low, this assemblage may enhance species richness over time but they may also collapse as soon as an extra predator colonizes the island.
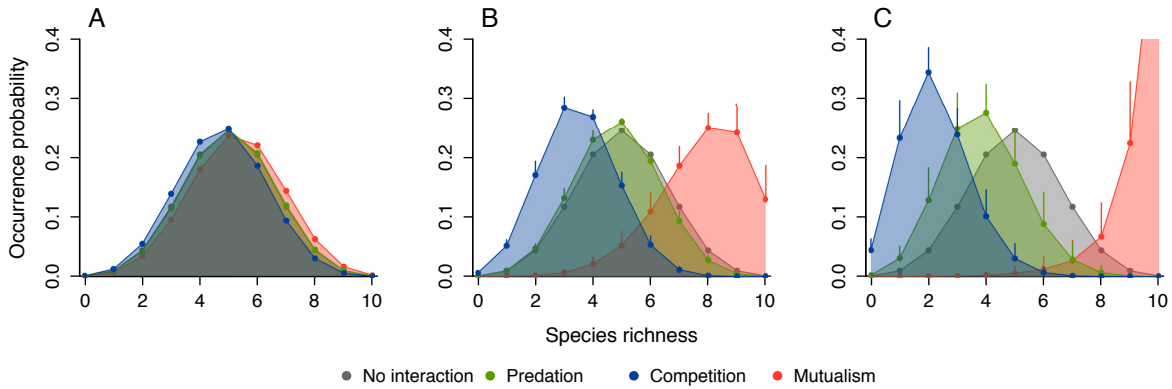
Figure 4: **Probabilities of species richness for different types of interaction**. We compute expected species richness at the equilibrium with the following set of parameters: $e_i = 10^{-5}$, $e_{i,min} = 10^{-3}e_i$, $e_{i,max} = 10^3 e_i$ and $c_i = 10^{-5}$. We do so for three different interaction weights: $d = 0.1$ (A), $d = 1$ (B), $d = 10$ (C). In each panel, the four colours stand for the following types of networks: no interaction (grey), predation (green), competition (blue) and mutualism (red). Probabilities associated to coloured points are the means calculated for 1000 interaction matrices randomly-generated according to the niche model (Williams and Martinez, 2000) with a connectance set to 0.2. Additionally, vertical bars represent 50% of the standard deviations associated to these means. To facilitate comparisons among panels, we do not represent the occurrence probability of the 10 assemblages community in panel C for mutualism, which is 0.66 (the standard deviation associated is 0.33).

1.8   Discussion

Understanding how colonization-extinction dynamics influence species distribution and community structure remains a major challenge in biogeography (Wiens, 2011; Jabot and Bascompte, 2012; Godsoe and Harmon, 2012). Here, we build upon the simplicity of the Theory of Island Biogeography (TIB) to integrate crucial ecological processes, namely biotic and abiotic dimensions of the niche. Using the formalism of Markov chains, we derive an exact general solution for the occurrence probabilities of all possible assemblages that we calculate numerically (up to 10 species). Our approach is in stark contrast to the classic TIB (MacArthur and Wilson, 1967) where environmental gradients were not introduced and the co-occurrence among species was not modelled, despite empirical evidence of their impact (Diamond and Gilpin, 1982). By taking these constraints together we reveal how they in-

terplay and affect species richness. We believe our approach offers new perspectives on the theory of biogeography and will support the development of species distribution models with the addition of species interactions.

In our model, we introduce the effect of biotic interactions as an ecological process affecting colonization/extinction probabilities. This has already been considered in many ways in the literature. For instance, more than forty years ago, Levins and Culver introduced extinction and migration rates affected by competition and showed analytically how it reduces co-occurrence (Levins and Culver, 1971). More recently, Jabot and Bascompte introduced production of eggs and seeds affected by interaction in an individual-based, meta-community framework and, hence, highlighted the potential effects of interactions on local diversity (Jabot and Bascompte, 2012). Also, Calcagno and colleagues demonstrated that tuning extinction and colonization rates based on the trophic relationships among species could explain the limited length of food chain (Calcagno et al., 2011). In contrast with previous studies, our approach is fully rooted on the TIB which yields well-defined null predictions (adding neither interaction nor environmental gradients), focuses on assemblages, and allows the investigation of the impact of any kind of network, including mixed interactions.

Networks are convenient representations of the structure of ecological communities to study persistence and resilience (Thébault and Fontaine, 2010). A strength of our model is that it not only takes all direct interactions into account, but also indirect ones (Wootton, 1994). For instance, in a linear trophic chain of three species, the occurrence of the top predator depends not only on the presence of its prey but also on the species at the bottom of the chain (Gravel et al., 2011). This means that the distribution of the top predator will be influenced not only by its own abiotic requirements, but also by those of its prey and the species at the bottom of the chain. The signature of such indirect interactions should be common in co-occurrence networks. This property comes from the assumption that interactions change extinction rates and the Markov chain formalism employed. Our formalism therefore provides a tool, similar to the general Lotka-Volterra equations for the local scale, that could
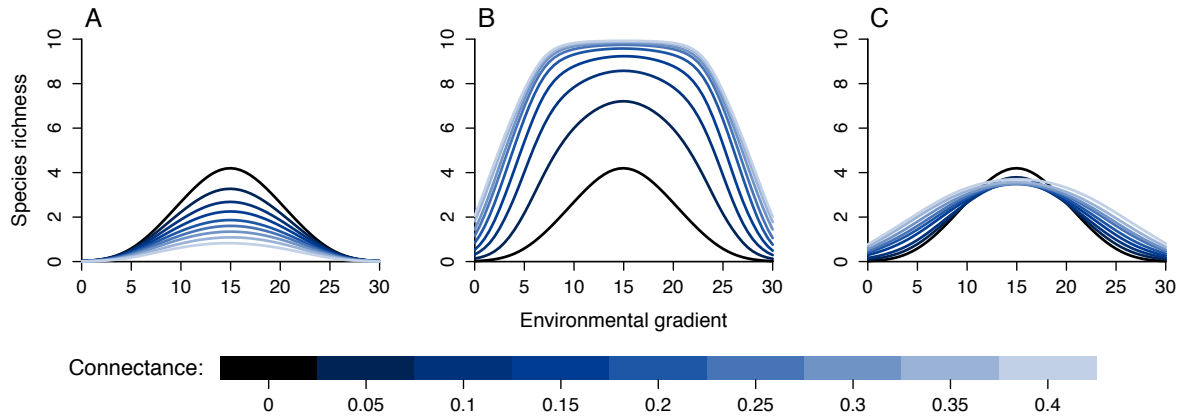
Figure 5: **Biodiversity distribution along environmental and connectance gradients.** We compute the expected species richness along an environmental gradient for competition (A), mutualism (B) and predation (C). We do so for different values of connectance depicted by the shades of blue. Species richness profile associated with the scenario without interaction is provided in each panel by the darkest solid line (connectance set to 0). Abiotic niches do have the same range for all species ($r_i = 5$) and the optima are randomly drawn in the interval [10, 20]. The interaction weight (d) is set to 10. The extinction parameters are set as follows: $e_i = 10^{-5}$, $e_{i,min} = 10^{-3}e_i$, $e_{i,max} = 10^3 e_i$ and $c_i = 10^{-5}$.

be used to study the emergence of indirect interactions in networks at the large spatial scale.

The challenge of developing joint species distribution models (Pollock et al., 2014; Pellissier et al., 2013) have recently motivated researchers to investigate co-occurrence (Araújo and Rozenfeld, 2014; Veech, 2013). Our framework helps to disentangle the two main processes by which non-random species associations (co-occurrence) can arise. First, two species not interacting with each other could be non-randomly co-distributed because of similar or antagonistic ecological requirements. As we introduced an abiotic constraint on the colonization probability, some assemblages will be more likely than others on a given environment simply because some species are favoured and others filtered out. We thus expect to find a signature of the covariance in species response to the environment on these assemblage probabilities. Secondly, non-random co-distribution will arise from ecological interactions. We considered an additive impact of all ecological interactions a species is experiencing from the community. Species interact in various ways, but at the end all interactions do impact de-

mography by definition. This reality enters the model by either enhancing of decreasing extinction probabilities. In other words, the occurrence of a single species is derived from the expectation of observing all other species in the community.

Our framework therefore provides a formalism to investigate the relationship between co-occurrence networks (Araújo et al., 2011) and interaction networks. There is a significant amount of information contained in the data of co-occurrence, which is overlooked by most current methods of community analysis. Standard species distribution models are fitted to univariate presence/absence data, neglecting the information contained in the distribution of associated taxa. Multivariate statistics summarize the spatial structure of ecological communities, but they are essentially limited to the description of co-occurrence, they are not meant to predict species distributions conditional on other species. Most analyses of co-occurrence aggregate pairwise observations into a single index for the whole community, thereby missing substantial information pertaining to the consequences of biotic interactions (Boulangeat et al., 2012). This situation is not surprising given there is no general theory for co-occurrence. Current hypotheses are mostly limited to negative interactions, leading to negative co-occurrence (repulsion), or positive interactions, leading to positive co-occurrence (attraction). Many theoretical achievements are required to study co-occurrence for more complex assemblages, mixing positive, negative and antagonistic interactions. In addition, the impact of indirect interactions emerging in interaction networks on species distribution is ignored. Our approach provides a formal framework to overcome these limitations as we calculate assemblage probability at biogeographical scale and then derive co-occurrence. It also allows the decomposition of the strength of pairwise associations between abiotic and biotic drivers, opening the way for novel statistical developments of species distribution models taking into account this multi-occurrence information. We propose that studying the role of biotic factors at large scale requires us to introduce them as assemblages instead of adding species as factors which likely leads to non-equivocal conclusion (Araújo and Luoto, 2007). In addition, our approach is not limited to species pairs, the assemblage probabilities provide a valuable tool to the co-occurrence of groups of species such as motifs (Stouffer et al., 2007).

The importance of interactions across different scales is still debated (McGill, 2010; Araújo and Rozenfeld, 2014). A common assumption is that interactions are negligible at large spatial scales, based on the rational that abiotic filters primarily determine the composition of assemblages (Pearson and Dawson, 2003). This argument persists even though theoretical (Gravel et al., 2011) and empirical (Gotelli et al., 2010) evidence suggest the opposite. The key issue to solve this debate is thus to know how interactions can influence species assemblages with increasing spatial scale. Although the TIB still provide insights into the assembly of natural communities, the success of recent approaches integrating interactions strongly support their relevance at large scales. Indeed the addition of network structure (Pellissier et al., 2013) or correlation between species (Pollock et al., 2014) as proxies for interactions have adequately improved forecast accuracy. Here we do not solve this fundamental issue, however our model illustrates how species distribution at large scale will be impacted by the kind of interaction, their numbers and their distribution.

Although our framework is not readily applicable to real datasets, it nonetheless provides a theoretical foundation for the derivation of new statistical modelling approaches. We propose a different perspective which is rooted on theory, in contrast with what is usually done with phenomenological model representing the structure of the data (Thuiller et al., 2013). There are nonetheless significant challenges to apply our framework to empirical data. First, we must find a way to deal with large numbers of species. At present, given n species in the regional pool, we compute an eigen vector of 2^n probabilities from a $2^n * 2^n$ transition matrix of a Markov chain. Moreover, in its current formulation, it requires us to evaluate a very large amount of data including a description of network of the same species across time and space to get accurate estimations. Solving this issue will requires a rational to reduce the number of species considered. This could be achieved either by inference of the relevant interactions, or alternatively by pooling species into groups. A systematic and rigorous method to build meaningful groups of interacting species from proxies such as traits and phylogenies remains to be developed, but there are nonetheless promising avenues (Baskerville et al., 2011). The relatively small number (from 3 to 7) of dimensions to ecological networks, i.e.

the number of trait-axes required to properly infer interactions (Eklöf et al., 2013), supports its feasibility. A second challenge is to account for spatial structure that constrains population flux. Despite the theoretical developments, applied approaches to model species distribution struggle to introduce it efficiently (Boulangeat et al., 2012). The island-mainland approximation remains elegant but might be too simple for applied situations. One solution may be to identify source and sink localities, (Boulangeat et al., 2012). This requires us to consider i) species abundances and ii) spatial structures which would strongly increase the complexity of the model. One first step forward could be to apply the Levins model rather than the island-mainland model as Levins and Culver did to study the impact of competition (Levins and Culver, 1971)).

Despite our call for a new integrated theory of biogeography, we acknowledge the limitations of the framework. Recent studies aimed at integrating population dynamics, for instance, using approximations from the metabolic theory of ecology. This is a hopeful direction to assess local extinction risk, accounting for network structure, body size and abundance (Schneider et al., 2012). Beyond body size, other functional traits (McGill et al., 2006) could help us to escape from species singularity toward more general rules. Hence, approximating food web structure could be done using traits (Gravel et al., 2013) and energetic requirements can be easily quantified through body size and local temperature constraints (Brown et al., 2004). Moreover considering changes in traits over time may be a key to introduce evolutionary processes. This would help us to release one strong assumption of our work: no speciation processes are taken into account. Although it might not matter for short time periods, having a pool of species unchanged becomes a major issue for time scales that exceed by far the lifespans of species we consider. Further, our framework could be applied to investigate diversification dynamics on remote areas, with a particular emphasis on the effect of ecological interactions on adaptive radiations. Despite the complexity of such model, it would very likely provide valuable insights on the future of biodiversity under current global change.

Since the seminal work of Davis et al. (Davis et al., 1998), there is growing evidence that the response of species to climate change must be studied at the community scale (Suttle et al., 2007). Even though species respond individually to climate change, they are constrained by complex direct and indirect biotic interactions emerging from large scale organization (Lavergne et al., 2010). The study of Cahill and colleagues (Cahill et al., 2013) has revealed the difficulties to link climate changes and species extinction. Even when the climate is expected to drive local extinctions, it actually implies a chain of perturbations amidst which biotic factors prevail (*e.g.* loss of prey (Durance and Ormerod, 2010)). For instance, species contributing to the persistence of plant-pollinator networks are paradoxically the most vulnerable to extinction (Saavedra et al., 2011), highlighting the risk of extinction cascades. As Säterberg et al. expressed, "the species to be the first to go extinct is not the one whose mortality rate is increased but instead some other species in the food web", thereby suggesting that perturbations which affect species differently also spread over the network making extinction difficult to predict (Säterberg et al., 2013). Although this is fully understandable as species interact, this makes forecasting of future species distributions more complicated. Therefore the challenge of proposing biodiversity scenarios to global change requires new approaches integrating ecological processes over time and spatial scales, and to disentangle their relative contribution (Lavergne et al., 2010). We think that the assemblage-based approach we propose here is a promising perspective to introduce interactions in biogeographical models.

1.9 Acknowledgment

ARTICLE 2

THEORIE DE CO-OCCURRENCE DES ESPECES DANS LES RÉSEAUX D'INTERACTION

2.1 Résumé en français du deuxième article

Dans le chapitre précédent, j'ai mis en évidence l'impact potentiel des interactions sur la présence locale des espèces dans le cadre de la théorie de la biogéographie des îles. Nous avons ainsi mis en évidence un effet des interactions sur la répartition géographique des espèces. Nous avons proposés une solution efficace pour essayer d'inclure ces interactions : ne pas traiter les espèce une à une mais les utiliser sous forme d'assemblage. Cette approche a déjà une place importante dans la litérature en biogéographie sous la forme des diverses approches de modélisation à l'échelle des communautés. Dans ces approches, il existe toujours une étape au cours de laquelle les différentes espèces sont assemblées en vue d'améliorer les prédictions de biodiversité. Seulement, ces méthodes ne font pas de lien entre les interactions et ces assemblages qui demeurent des groupes d'espèce trouvés fréquemment ensemble. Pour montrer l'importance des intéractions dans les modèles en biogéographie, nous nous sommes intéressés à leur impact sur une mesure majeure en biogéographie : la co-occurrence.

La co-occurence de deux espèces est simplement le nombre total de sites où les espèces sont présentes ensemble rapporté au nombre totale de sites étudiés. Pour pouvoir allez plus loin, nous avons défini une mesure de co-occurence sous l'hypothèse d'indépendance de deux espèces. Cela signigfie simplement que nous prenons l'occurrence respective des deux espèces et que nous les mulitplions pour obetnir notre seconde mesure. Grâce à la comparaison entre ces deux valeurs nous avons pu illustrer, dans l'article présenté ci-dessous, cinq grands principes relatifs à la co-occurence des espèces en interaction :

1. **Les interactions directes entre deux espèces affectent leur probabilité de co-**

occurence. C'est une transposition directe du chapitre précédent sur la mesure de co-occurrence : s'il existe un lien entre deux espèces, leur probabilité d'être présente simultannément dans une localité diffère de la probabilité attendue si elle se recontrait aléatoirement.

2. **Les interactions indirectes modifient leur probabilité de co-occurence**. Malgrè qu'il n'existe pas d'interaction directe entre deux espèces, ces dernières peuvent néanmoins être liées par une ou plusieurs autres espèces, l'interaction est dite indirecte. Si les conséquence des interactions directes se propagent propager à travers le réseau via ces relations indirectes, il est alors possible que répartition d'une espèce soit affectée par une autre espèce avec laquelle aucune interaction directe n'est constatée.

3. **L'effet des interactions sur la co-occurrence n'est pas symétrique**. Il n'existe a priori aucune raison pour que ces effets soient symétriques. Néanmoins en utilisant la mesure de co-occurrence telle que décite ci-dessus, nous la considérons comme telle. Nous montrons alors comment les probabilités conditionnelles peuvent prendre en compte l'asymétrie des effets des interactions.

4. **La force d'association entre deux espèces diminue avec la longueur du plus court chemin entre deux espèces**. Plus les espèces sont éloignées dans le réseau, moins les conséquences des interactions indirectes seront perceptibles, nous illustrons donc que les effets des interactions directes diminuent lors de leur propagations dans le réseau.

5. **La force d'une association avec une autre espèce diminue avec le nombre d'inetraction qu'elle entretient.** Si une espèce a de nombreux liens dans le réseau (par exemple, un prédateur généraliste), alors celle-ci sera moins dépendante d'une espèce en particulier et de fait la relation qu'elle entretien avec les expèces se rapprochera de la co-coorrence sous hypothèse d'indépendance.

Pour ce second papier, le contexte est particulier : Dominique Gravel a été invité à un numéro spécial de *Theoritical Ecology* sur les réseaux écologique. Dominique Gravel m'a alors proposé de travailler sur le prolongement de la reflexion mené au premier chapitre et de

l'appliquer sur les données de co-occurrence. J'ai alors conceptualisé un modèle probabiliste pour tenter de comprendre comment les interactions peuvent affecter la co-occurrence. Je me suis occupé de toute la partie modèle et des figures. Dominique Gravel a écrit la majeure partie de l'introduction et de la discussion. La reflexion mené ayant été entammé par Dominique Gravel et Miguel B. Araújo, ce chercheur est devenu second auteur et à participer activement à la rédaction. Enfin, Nicolas Mouquet a participé substentiellement à la rédaction du manuscript.

2.2   Titre

A Theory for species co-occurrence in interaction networks

2.3   Auteurs

Kévin Cazelles, Miguel Araújo, Nicolas Mouquet et Dominique Gravel

2.4   Résumé en anglais

The study of species co-occurrences has been central in community ecology since the foundation of the discipline. Co-occurrence data are, nevertheless, a neglected source of information to model species distributions and biogeographers are still debating about the impact of biotic interactions on species distributions across geographical scales. We argue that a theory of species co-occurrence in ecological networks is needed to better inform interpretation of co-occurrence data, to formulate hypotheses for different community assembly mechanisms, and to extend the analysis of species distributions currently focused on the relationship between occurrences and abiotic factors. The main objective of this paper is to provide the first building blocks of a general theory for species co-occurrences. We formalize the problem with definitions of the different probabilities that are studied in the context of co-occurrence analyses. We analyse three species interactions modules and conduct multi-species simulations in order to document five principles influencing the associations between species within an ecological network: i) direct interactions impact pairwise co-occurrence; ii) indirect interactions impact pairwise co-occurrence; iii) pairwise co-occurrence rarely are symmetric; iv) the strength of an association decreases with the length of the shortest path between two species; v) the strength of an association decreases with the number of interactions a species is experiencing. Our analyses reveal the difficulty of the interpretation of species

interactions from co-occurrence data. We discuss whether the inference of the structure of interaction networks is feasible from co-occurrence data. We also argue that species distributions models could benefit from incorporating conditional probabilities of interactions within the models as an attempt to take into account the contribution of biotic interactions to shaping individual distributions of species.

2.5 Mot-clef

Co-occurence, Ecological networks, Biogeography, Indirect interactions, Null models

2.6 Introduction

Understanding of the processes driving the assembly of communities has been a central theme of ecology since the foundation of the discipline. How do we start from a regional species pool to assemble a structured community? Why are some species associated with each other? The work of Diamond (1975) pioneered the analysis of species co-occurrence in geographical space and, together with the controversy triggered by Connor and Simberloff (1979), it stimulated the development of a new field of research in numerical ecology (Stone and Roberts, 1990; Gotelli and Graves, 1996; Legendre and Legendre, 2012). The foundational work on species co-occurrences also led to the development of a rich array of methodological tools designed to test null hypotheses in ecology. Even if null models could be achieved numerically (*e.g.*, Araújo et al., 2011), typically they are based on permutations of distribution data. Null models have been used to infer the role of biotic interactions between pairs of species on their individual distributions. Studying the different drivers of species co-occurrence is not only of theoretical interest for improving understanding of the mechanisms of community assembly. It is also instrumental in predictive ecology, because a considerable amount of information is contained in species distributions data.

Despite its historical importance for community ecology, co-occurrence data remain a neglected source of information in models of species distributions. Biogeographers are still debating the impact of biotic interactions on species distributions (Guisan and Thuiller 2005; Gotelli et al. 2010; Kissling et al. 2012; Pellissier et al. 2013). The distribution of a species is thought to be first influenced by its physiological tolerance to environmental conditions, but also by interactions with other species (Hutchinson, 1957; MacArthur, 1972; Peterson et al., 2011; Boulangeat et al., 2012). The question of whether such interactions leave imprints in the distributions of individual species at biogeographical scales is still open to debate (*e.g.* Davis et al. 1998), but recent empirical Gotelli et al. 2010, modeling (*e.g.*, Araújo and Luoto, 2007), and theoretical (Araújo et al., 2011) evidence invites the interpretation that this might indeed be the case.

The overwhelming majority of species distributions modelling applications, nonetheless, neglect information contained in joint distributions. Even multivariate analysis of community data (*e.g.* redundancy analysis - (Legendre and Legendre, 2012) do not use co-occurrence in geographical space to condition individual species response to environmental variation. There has been a recent rise of interest however in joint species distribution modelling (Clark et al., 2014; Harris, 2015; Pollock et al., 2014). These methods estimate the distribution of all species from a pool simultaneously and allow to condition the presence of a species on all other ones. However, estimated relationships are inferred from co-occurrence in environmental space rather than geographical space. That is, joint responses to the environment are inferred rather than biotic interactions themselves (Baselga and Araújo, 2009). JSDMs are, nonetheless, a first step towards developing a next generation of models accounting for the impact of biotic interactions on the distributions of species. They are, however, purely empirically driven and carry no specific hypotheses about how interactions can affect distributions. An exception is the recent attempt to model the effects of predator-prey dynamics on distributions and abundances using a meta- community framework coupled with phenomenological species distributions models (Fordham et al., 2013). The problem with such approaches is that data to parametrize interactions mechanistically are generally lacking

(Morales-Castilla et al., 2015); therefore, they are hardly applied in most circumstances. It follows that we are faced with at least two major problems: i) understanding of the ecological interactions underlying the distributions of species is limited; and ii) knowledge of interactions is typically limited to net interactions, mixing both direct and indirect interactions. A theory of species co-occurrences in ecological networks is, therefore, needed to help interpret co-occurrence data, to formulate hypotheses for different community assembly mechanisms, and to extend the analysis of species distributions currently focused on the relationship between occurrences and abiotic factors.

The analysis of species co-occurrences starts with a matrix representing the presence and absence of each species over a set of sites. There are two aspects to the quantitative study of co-occurrence. The first is the choice of the metric used to quantify the strength of associations (relationships between species occurrences) between pairs of species. The simplest measure of species co-occurrence is the number of species combinations, as defined by Pielou and Pielou (1968). A second index is the count of checkerboards Diamond (1975): "In such a pattern, two or more ecologically similar species have mutually exclusive but interdigitating distributions in an archipelago, each island supporting only one species" (p. 32). Another popular index of co-occurrence is the C-score (Stone and Roberts, 1990). This index is similar to the count of checkerboards; it measures the average association or repulsion between pairs of species.

The second aspect of the analysis of species co-occurrence is the formulation of a null model. The controversy generated in (Connor and Simberloff, 1979) was partly (and rightly) based on the absence of a valid null hypothesis in Diamond's analysis. Subsequent debates were mostly concerned with the formulation of the null hypothesis (*e.g.*, Diamond and Gilpin, 1982). Thanks to the theoretical work of Gotelli and Graves (1996), there is now a clear understanding of the different null models that can be constructed from the community matrix. New indices are constantly proposed, such as in (Boulangeat et al., 2012; Veech, 2013); see also Table 2 in (Ulrich and Gotelli, 2013) for a description of 15 indices for co-occurrence

analysis. A promising avenue is the one proposed by Araújo et al. (2011) for the study of the matrix of species co-occurrence with tools borrowed from network theory.

Surprisingly, there is currently no theory for co-occurrence in multi-species communities. The basic hypotheses are that pairwise negative interactions result in repulsion, while pairwise positive interactions result in attraction. Attraction and repulsion are assessed by a comparison of the number of co-occurrence events to the number expected under a totally independent distribution. Similar environmental requirements between species could also result in attraction, even in the absence of interactions, if the sampling is conducted across heterogeneous environmental conditions. This theory is limited to pairwise and symmetric interactions; there is nothing for antagonistic and indirect interactions. Food web ecologists were among the first to recognize the important effect of indirect interactions on abundance (Wootton, 1994). For instance, plant and carnivore abundances are expected to correlate across a productivity gradient (Hairston et al., 1960; Oksanen et al., 1981) because of top-down control on the herbivore population. Similarly, the propagation of indirect interactions has been studied in more complex interaction networks (Yodzis, 1988). Indirect interactions could reverse the net interaction in a surprising way, such that predator-prey abundances could be positively related (Montoya et al., 2009). Empirical analysis of co-occurrence for several taxa has shown that they are usually asymmetric (Araújo et al. 2011), such that a species distribution tended to be nested within the distribution of other (*e.g.* predator-prey distributions Holt and Barfield, 2009; Gravel et al., 2011). In such a case, even if the co-distribution signature is quite understood, available methods will likely fail at using this piece of information to improve forecasts.

The main objective of this paper is to provide the first building blocks of a general theory of species co-occurrences. We formalize the proposed theory with definitions of different quantities that are studied in the context of co-occurrence analyses. Herewith, we analyse three species interactions modules in order to document five principles influencing the association between pairs of species from an ecological network: i) direct interactions impact

pairwise co-occurrence; ii) indirect interactions impact pairwise co-occurrence; iii) pairwise co-occurrence does not have to be symmetric; iv) the strength of an association decreases with the length of the shortest path between two species; v) strength of an association decreases with the number of interactions a species is experiencing. We base our mathematical argument on a general model of species distributions that is free of any assumption about how the ecological interactions operate. Finally we extend our analysis with simulations of multi-species networks in order to analyse how these mechanisms scale up in species rich communities.

2.7 Definitions

We start with definitions to formalize the quantities that can be computed from species distribution data and be used in the context of co-occurrence analyses. Let X_i be the random variable representing the presence of species i. $X_i = 1$ when species i is present, $X_i = 0$ otherwise. Then $X_{i,t>0}$ is the random process associated, giving the value that $X_{i,t}$ takes at any time t. Let $p_{i,t}$ standing for the probability $\mathbb{P}(X_{i,t} = 1)$. Also, to illustrate the defintions, we derive the quantities for a simple presence/absence dataset (see Table 2).

The **marginal occurrence probability** $\mathbb{P}(X_{i,\infty} = 1) = p_i^*$ represents the occurrence probability of species i when the system is at equilibrium, in the sense of the classical theory of island Biogeography MacArthur and Wilson (1967). As we assume so for all species, we drop the $*$ and the ∞ for the sake of clarity. The marginal occurrence probability is the sum of the occurrence of the species across all possible set of species in the data. In other words, it corresponds to the sum of the column of the site $\times$ species table, divided by the total number of sites N. Marginal occurrence probabilities for species in Table 2 are: $p_1 = 0.6$, $p_2 = 0.6$ and $p_3 = 0.4$.

The **observed co-occurrence** between species i and j is the joint probability $p_{i,j} = \mathbb{P}(X_i = 1 \cap X_j = 1)$. It represents the number of sites where the two species are found together,

across all possible set of species in the data (in other words, it is a marginal probability with respect to other species), divided by N. In our dataset, for instance, we have $p_{1,2} = 0.3$ and $p_{1,3} = 0.2$.

The **conditional co-occurrence** between species i and j is $p_{i|j} = \mathbb{P}(X_i = 1|X_j = 1)$. It represents the probability of observing species i, knowing that species j is already present. This quantity is close to the measure of association between two species because it is independent of the marginal occurrence probability of both species. The problem is that, as soon as there are other species present, the conditional co-occurrence as expressed here is marginalized over the set of all other species from the community K. For instance, for three species, we have: $p_{1|2} = \mathbb{P}(X_1 = 1|X_2 = 1, X_3 = 1) + \mathbb{P}(X_1 = 1|X_2 = 1, X_3 = 0)$. It, therefore, includes both the effect of *direct* and *indirect* associations between species, e.g. the direct association of species 1 with species 2 or the indirect association of species 3 with 1 via its effect on 2. Consequently, the measure of pairwise association should be: $p_{i|j,\overline{K}} = \mathbb{P}(X_i = 1|X_j = 1, X_K = 0)$, where the horizontal bar over K denotes absence of all other species. We name this the **fundamental conditional co-occurrence**. For instance, in Table 2, we get $p_{1|2} = \frac{p_{1,2}}{p_2} = 0.5$ and $p_{1|2,\overline{3}} = \frac{p_{1,2,\overline{3}}}{p_{2,\overline{3}}} = \frac{0.2}{0.3} = 0.67$.

Following the same logic, we define the **fundamental occurrence** as $p_{i|\overline{K}} = \mathbb{P}(X_i = 1|X_K = 0)$. The fundamental occurrence is conceptually equivalent to the fundamental niche of Hutchinson (1957) and represents the probability of observing a species in the absence of biotic interactions, i.e., when all other species are absent. By analogy, the marginal occurrence should be interpreted as the realized distribution. For species 1 in Table 2 we calculate $p_{1|\overline{23}} = \frac{p_{1,\overline{2},\overline{3}}}{p_{\overline{2},\overline{3}}} = \frac{0.2}{0.3} = 0.67$.

Finally, we define the **independent co-occurrence** as $p_{i,j;IND} = \mathbb{P}(X_i = 1)\mathbb{P}(X_j = 1)$. It represents the co-occurrence between any pairs of species expected in absence of any association between them. In ecological terms, it would represent the co-occurrence when ecological interactions and habitat filtering do not impact species distribution. It also represents the null model against which observed co-occurrence is usually compared to. Note the independent

co-occurrence is different from the one expected under a neutral model (Stephen P. Hubbell, 2001). Firstly because strong competitive interactions in the neutral model forces repulsion and, secondly, because dispersal limitation also causes spatial aggregation and thus a non-random distribution of co-occurrence (Bell, 2005). In our example, we obtain, for instance, $p_{1,2;IND} = 0.36$ and $p_{2,3;IND} = 0.24$.

Direct association between two species

We start with the analysis of a two species situation, labeled species 1 and species 2, in order to understand direct associations between species pairs. A third species, 3, will be introduced in the next section to study indirect associations. The model we develop is general, as we do not specify the type of ecological interactions involved. It therefore accounts for all possible mechanisms from which an association between a pair of species could arise, such as trophic interactions involving energy fluxes, non-consumptive interactions, parasitism, direct interference, territoriality, space pre-emption, niche construction, etc. The impact of predator-prey interactions in a metapopulation setting with colonization and extinction dynamics will be considered for the multi-species simulations.

As we are willing to understand the role played by interactions in co-occurrence, we start by defining marginal co-occurrence probabilities of our two species by a decomposition into conditionnal co-occurrences. By the formula of total probability we have:

$$
\begin{aligned}
p_1 &= \mathbb{P}(X_1 = 1 \cap X_2 = 1) + \mathbb{P}(X_1 = 1 \cap X_2 = 0) \\
&= \mathbb{P}(X_1 = 1 | X_2 = 1)\mathbb{P}(X_2 = 1) \\
&\quad + \mathbb{P}(X_1 = 1 | X_2 = 0)\mathbb{P}(X_2 = 0)
\end{aligned}
\tag{2.1}
$$

We do the same for species 2. Using the notation described above, (2.1) could be

rewritten as:

$$\begin{cases} p_1 = p_{1|2}p_2 + p_{1|\bar{2}}(1 - p_2) \\ p_2 = p_{2|1}p_1 + p_{2|\bar{1}}(1 - p_1) \end{cases} \tag{2.2}$$

where the vertical bar denotes the absence of a species. By solving the latter system, we get:

$$\begin{cases} p_1 = \dfrac{p_{1|\bar{2}} + p_{2|\bar{1}}(p_{1|2} - p_{1|\bar{2}})}{1 - (p_{2|1} - p_{2|\bar{1}})(p_{1|2} - p_{1|\bar{2}})} \\ p_2 = \dfrac{p_{2|\bar{1}} + p_{1|\bar{2}}(p_{2|2} - p_{2|\bar{1}})}{1 - (p_{2|1} - p_{2|\bar{1}})(p_{1|2} - p_{1|\bar{2}})} \end{cases} \tag{2.3}$$

When species are independent, we have $p_{1|\bar{2}} = p_{1|2} = p_1$ and $p_{2|\bar{1}} = p_{2|1} = p_2$, then we logically find (2.1) again. Then, we can deduce the following interpretation of the impact of **direct interactions** on co-occurrence:

i if species 1 cannot persist in absence of 2 (e.g., a parasite requiring its host), then $p_{1|\bar{2}} \to 0$, therefore $p_1 \to p_{1|2}p_2$

ii if species 1 depends strongly on 2 thereby perfectly tracking its distribution 2, the $p_{1|\bar{2}} \to 0$ and $p_{1|2} \to 1$, and therefore $p_1 \to p_2$

iii if species 2 excludes 1, then $p_{1|2} \to 0$ and $p_{2|1} \to 0$, so $p_1 = \frac{p_{1|\bar{2}} - p_{2|\bar{1}}p_{1|\bar{2}}}{1 - p_{2|\bar{1}}p_{1|\bar{2}}}$ and $p_2 = \frac{p_{2|\bar{1}} - p_{2|\bar{1}}p_{1|\bar{2}}}{1 - p_{2|\bar{1}}p_{1|\bar{2}}}$. Therefore, if $p_{1|\bar{2}} \to 1$, then $p_1 \to 1$ and $p_2 \to 0$.

Co-occurrence in three-species modules

Now, we consider the co-occurrence between three species. We start with a general derivation of co-occurrence and then interpret the results for particular modules in order to reveal fundamental principles underling co-occurrence in ecological networks. Our solution provides insights to decipher the solution of species-rich networks since the three-node

connected subgraphs are fundamental building blocks of larger networks (Milo et al. 2002; Stouffer et al. 2007; Stouffer and Bascompte 2010). We use the same approach as in (2.1) and get the subsequent equation:

$$
\begin{aligned}
p_1 \;=\;& \mathbb{P}(X_1 = 1 \cap X_2 = 1 \cap X_3 = 1) + \mathbb{P}(X_1 = 1 \cap X_2 = 0 \cap X_3 = 1) \\
& + \mathbb{P}(X_1 = 1 \cap X_2 = 1 \cap X_3 = 0) + \mathbb{P}(X_1 = 1 \cap X_2 = 0 \cap X_3 = 0)
\end{aligned}
\tag{2.4}
$$

As $\{X_3 = 1, X_3 = 0\}$ forms a partition we get:

$$
p_1 \;=\; \mathbb{P}(X_1 = 1 | X_3 = 1)p_3 + \mathbb{P}(X_1 = 1 | X_3 = 0)(1 - p_3)
\tag{2.5}
$$

This equation is analogous to the two-species interactions equation but enables the study of networks involving three species interactions, with species 2 being hidden by marginalization. We split the three species problem in two distinct two-interactions species problems. Firstly, we solve the equation for sites without species 3 and get:

$$
p_{1|\bar{3}} = \mathbb{P}(X_1 = 1 | X_3 = 0) \;=\; \frac{p_{1|\overline{23}} + p_{2|\overline{13}}(p_{1|2\bar{3}} - p_{1|\overline{23}})}{1 - (p_{2|1\bar{3}} - p_{2|\overline{13}})(p_{1|2\bar{3}} - p_{1|\overline{23}})}
\tag{2.6}
$$

which is similar to equation (2.3) but with an explicit absence of species 3. We do similarly for the conditional occurrence of 1 on species 3:

$$
p_{1|3} = \mathbb{P}(X_1 = 1 | X_3 = 1) \;=\; \frac{p_{1|\bar{2}3} + p_{2|\bar{1}3}(p_{1|23} - p_{1|\bar{2}3})}{1 - (p_{2|13} - p_{2|\bar{1}3})(p_{1|23} - p_{1|\bar{2}3})}
\tag{2.7}
$$

Doing so, we get the following set of equations describing the marginal occurrence probabilities for the three species:

$$
\begin{cases}
p_1 = p_{1|3}p_3 + p_{1|\bar{3}}(1 - p_3) \\
p_2 = p_{2|3}p_3 + p_{2|\bar{3}}(1 - p_3) \\
p_3 = p_{3|2}p_2 + p_{3|\bar{2}}(1 - p_2)
\end{cases}
\tag{2.8}
$$

Note that we could have chosen a different set of equations depending on the way we split the problem, for instance, we could have started by considering the occurrence of species 1 given the occurrence of species 2 instead of species 3. Now, we solve the above linear system of three equations with three unknowns and find that:

$$
\begin{cases}
p_1 = \dfrac{p_{1|\bar{3}} + p_{3|\bar{2}}(p_{1|3} - p_{1|\bar{3}}) + (p_{3|2} - p_{3|\bar{2}})(p_{1|3}p_{2|\bar{3}} - p_{1|\bar{3}}p_{2|3})}{1 - (p_{2|3} - p_{2|\bar{3}})(p_{3|2} - p_{3|\bar{2}})} \\[4mm]
p_2 = \dfrac{p_{2|\bar{3}} + p_{3|\bar{2}}(p_{2|3} - p_{3|\bar{2}})}{1 - (p_{2|3} - p_{2|\bar{3}})(p_{3|2} - p_{3|\bar{2}})} \\[4mm]
p_3 = \dfrac{p_{3|\bar{2}} + p_{2|\bar{3}}(p_{3|2} - p_{3|\bar{2}})}{1 - (p_{2|3} - p_{2|\bar{3}})(p_{3|2} - p_{3|\bar{2}})}
\end{cases}
\tag{2.9}
$$

Conditional probabilities of the right-hand sides can all be derived as we did for $p_{1|3}$ in equation (2.7).

Community modules

We now interpret these equations with examples of well-studied food web modules in community ecology: 1) linear food chain, 2) exploitative competition and 3) apparent competition. To do so, we consider matrices of direct associations representing the conditional co-occurrence probabilities among all pairs of species (see Table 3).

We are interested by the *observed co-occurrence* because this is the quantity that is

easily measurable from species distributions data, thus being the one that is typically studied. We consider that the marginal occurrence is also a known quantity and, therefore, we examine the effect of particular conditional co-occurrence arrangements on observed co-occurrences. We will not provide derivations for each module, but focus on particular pairs to illustrate two of the five principles.

Sites	Species 1	Species 2	Species3
1	0	1	1
2	0	1	1
3	1	1	0
4	1	0	1
5	0	0	0
6	1	1	1
7	0	1	0
8	1	0	0
9	1	0	0
10	1	1	0

Table 2: Presence/absence dataset for three species and 10 sites.

Indirect interactions . The comparison between the observed co-occurrence and the conditional co-occurrence reveals the role of indirect interactions on species associations. Based on (2.9) and (2.6) we get the association between species i and k:

$$
\begin{aligned}
p_{i,k} &= p_i - p_{i,\bar{k}}(1 - p_k) \\
p_{i,k} &= p_i - \frac{p_{i|\overline{jk}} + p_{j|\overline{ik}}(p_{i|j\bar{k}} - p_{i|\overline{jk}})}{1 - (p_{j|i\bar{k}} - p_{j|\overline{ik}})(p_{i|j\bar{k}} - p_{i|\overline{jk}})}(1 - p_k)
\end{aligned}
\tag{2.10}
$$

Therefore the observed co-occurrence between species i and k depends on their respective interaction with species j ($p_{j|\bar{i}\bar{k}}$, $p_{j|i\bar{k}}$ and $p_{j|\bar{i}k}$). The conditional co-occurrence between two species could be null, but their observed co-occurrence be non-independent because of a shared interaction. This principle is best illustrated by the co-occurrence between a carnivore and a plant (species 3 and 1, respectively) in a linear food chain. In this situation, according to Table 3, we find that the observed co-occurrence between the plant and the carnivore is:

$$p_{1,3} = p_1 - \frac{p_{1|\overline{23}}}{1 - p_{2|1\bar{3}}(p_{1|2\bar{3}} - p_{1|\overline{23}})}(1 - p_3) \tag{2.11}$$

It is clear from this equation that there is a significant association between the carnivore and the plant, despite the conditional co-occurrence of the two species being totally independent. The indirect association gets stronger with the strenght of both conditional co-occurrence.

Similar observations could be made by studying the observed co-occurrence between consumers (species 2 and 3) in the exploitative competition module:

$$p_{2,3} = p_2 - \frac{p_{1|\overline{23}}p_{2|1\bar{3}}}{1 - (p_{1|2\bar{3}} - p_{1|\overline{23}})p_{2|1\bar{3}}}(1 - p_3) \tag{2.12}$$

And between resources in the apparent competition module (species 1 and 2):

$$p_{1,2} = p_1 - \frac{p_{1|\overline{23}}}{1 - p_{3|1\bar{2}}(p_{1|\bar{2}3} - p_{1|\overline{23}})}(1 - p_2) \tag{2.13}$$

Associations do not have to be symmetrical. Many studies of co-occurrence assume pairwise associations to be symmetrical (but see Araújo et al., 2011; Boulangeat et al., 2012). The reason is simple, usually the observed co-occurrence is compared to the independent co-occurrence. These two metrics of association are perfectly symmetrical. This information is

providing us an inappropriate interpretation of the effect of interactions on species distribution. If we consider for instance the association between the two consumers (species 2 and 3) competing for a single resource (species 1), we have the observed co- occurrence at (2.12), which is symmetrical by definition. The proportion of the area occupied by species 2 where species 3 is also present is not however equivalent to the proportion of the areas occupied by species 3. Rephrasing the problem, we find that using (2.7) and (2.12), $p_{2,3}/p_2$ is not equal to $p_{2,3}/p_3$. One species could have a stronger impact on the distribution of the other one. Predator distribution for instance tends to be nested within the distribution of the prey (Gravel et al., 2011), and consequently the predator has a high conditional co-occurrence with the prey, and alternatively the prey has a low conditional co-occurrence with the predator.

General case			Linear chain		
$p_{1|\overline{23}}$	$p_{1|2\overline{3}}$	$p_{1|\overline{2}3}$	$p_{1|\overline{23}}$	$p_{1|2\overline{3}}$	$p_{1|\overline{2}3}$
$p_{2|1\overline{3}}$	$p_{2|\overline{13}}$	$p_{2|\overline{1}3}$	$p_{2|1\overline{3}}$	0	0
$p_{3|1\overline{2}}$	$p_{3|\overline{1}2}$	$p_{3|\overline{12}}$	0	0	0

Exploitative competition			Apparent competition		
$p_{1|\overline{23}}$	$p_{1|2\overline{3}}$	$p_{1|\overline{2}3}$	$p_{1|\overline{23}}$	$p_{1|2\overline{3}}$	$p_{1|\overline{2}3}$
$p_{2|1\overline{3}}$	0	0	$p_{2|\overline{13}}$	$p_{2|\overline{13}}$	$p_{2|\overline{1}3}$
$p_{3|1\overline{2}}$	0	0	$p_{3|1\overline{2}}$	$p_{3|\overline{1}2}$	0

Table 3: Direct associations between pairs of species for different modules. Entries indicate the fundamental conditional probabilities of occurence of species i given the presence of species j and the absence of species k. *Linear chain*: 1 is the resource, 3 the top predator ; *Exploitative competition*: 2 and 3 are the consumers; *Apparent competition*: 1 and 2 are the resources. When $p_{i|j\overline{k}} = 0$, it means that species i cannot be found without k. When two species i and j do not interact directly, if the absence of species k do not impact species i survival then : $p_{i|j\overline{k}}=p_{i|\overline{jk}}$. For apparent competition, if species 1 and 2 are interchangeable for species 3 then : $p_{3|1\overline{2}} = p_{3|\overline{1}2}$.

Multi-species simulations

Now we move to multi-species simulations of more complex networks to reveal the last two principles of our theory. To do so, we run simulations of the model of trophic island biogeography developped by Gravel et al. (2011). The model describes the occurrence of a S species regional network. Species stochastically colonize islands with probability c and go extinct with probability e, as in the original model of MacArthur and Wilson (1967). Interactions are introduced with three additional assumptions: i) a consumer species could colonize an island only if it has at least one prey present (for simplicity, we consider producers to be resident permanently on the island); ii) a consumer species goes extinct if it loses its last prey species and iii) the presence of at least one predator species increases the extinction probability by e_d. The consequence of these assumptions is a sequential build-up of the food web on the island, starting with low trophic level species with a general diet. Small and isolated islands promote selection in favor of the most generalist species. The predictions converge to the classic island biogeography theory for highly connected regional food webs and large and connected islands (details in Gravel et al., 2011).

As mentioned above, there is a strong dependence of the predator occurrence on the presence of its preys. Alternatively, when e_d is sufficiently large, the preys will tend to avoid locations with the predator present. We consequently expect a strong signature of the network of interactions on the co-occurrence matrix. We are however concerned that indirect associations could emerge, as exemplified with the analysis of three species modules above, and thereby mask the signal of conditional co-occurrences.

We simulated complex networks from 5 to 100 species using the niche model of food web structure (Williams and Martinez, 2000). The diversity of primary producers was fixed at 2, and their niche position was drawn randomly between 0 and 1 according to a uniform distribution. We fixed connectance at $C = 0.1$. Colonization probability was set at $c = 0.1$, baseline extinction probability at $e = 0.2$ and predator-dependent additional extinction

probability at $e_d = 0.2$. Simulations were run for 10^7 time steps to evaluate the conditional occurrence probabilities, and 100 replicated networks were simulated for each level of species richness.
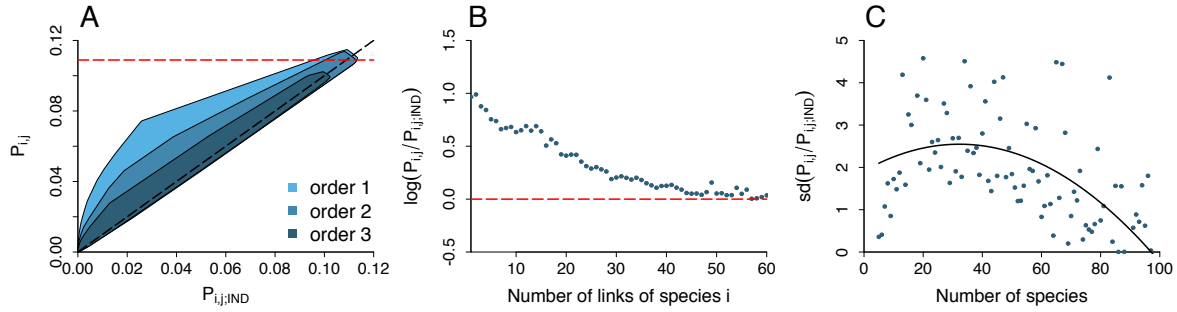

Figure 6: Co-occurrence in multi-species networks. (A) The disparity between observed co-occurrence ($P_{i,j}$) and independent co-occurrence ($P_{i,j;\mathrm{IND}}$) decreases with the path length between nodes (species). The enveloppes are drawn around the 5% and 95% quantiles of all of the data, from 100 replicated simulations for every species richness value (5 to 100 species). (B) The strength of co-occurrence ($log(P_{i,j}/P_{i,j;\mathrm{IND}})$) decreases with the number of interactions of a species i (*i.e.* the degree of a node). Points represent the mean for a particular degree of node value (1 to 60). The solid line represents the overall trends and the grey envelopped refelcts the variance associated. At least 3000 values were used for each point. (C) The standard deviation of the strength of association ($sd(P_{i,j}/P_{i,j;\mathrm{IND}})$) and thus the variance decreases with species richness. Taken together, (B) and (C) imply that species distributions converge to independence with increasing species richness.

Distance decay of observed co-occurrence. The distribution of observed co-occurrence is illustrated for pairs of species separated by different path lengths at Figure 1A. The observed co-occurrence is presented as a function of the expected co-occurrence under the hypothesis of independent distributions. The strongest associations (given by the distance between the observed and the independent co-occurrence) are observed among pairs of species directly interacting with each other. The variance of the distribution reduces from direct to first order indirect interactions, and from first-order to higher interactions. We conclude that indirect non- independent co-occurrences are possible in complex networks, but their magnitude decreases as the number of links between two nodes decreases. This result is similar to the

observation of a distance decay of indirect interactions in food webs (Berlow et al., 2009).

Strength of co-occurrence decreases with degree and species richness. We performed simulations with a gradient of species richness and observed that the variance of observed co-occurrence also decreases with the degree of a species, i.e. the number of direct interactions a species is experiencing (Fig. 1B). We illustrated the relationship between the degree of a species and the observed co-occurrence for pairs of species with a direct association (Fig. 1C). This phenomenon has the consequence that the strength of observed co-occurrence reduces with species richness. The niche model has a constant connectance (Williams and Martinez, 2000), which has for consequence an increase of the degree with species richness. We find that the strength of co-occurrence decreases with the degree. This result is straightforward to interpret: the more diverse are the interactions, the weaker the impact of each pairwise direct interaction on the species distribution. Again, this result is similar to the observation of a scaling relationship between pairwise interactions and food web diversity (Berlow et al., 2009).

Discussion

We first develop a probabilistic species distribution model constrained by biotic interactions using conditional probabilities of co-occurrence. We then illustrate five general principles underlying the impact of ecological interactions on co-occurrence and that should be considered for the formulation of a general theory of species co-occurrence. Two of them have been widely noted before: **i)** direct interactions affect species distributions and generate deviations in co-occurrences from that expected if distributions of species were independent from each other; **ii)** the effect of direct associations is often asymmetric, as envisioned in trophic metacommunity ecology (Holt and Barfield, 2009). We also illustrate principles that have been overlooked in most studies of co-occurrence (Araújo et al., 2011); **iii)** indirect interactions generate deviate co-occurrence from expectation under independence assumption;

iv) the strength of indirect associations decreases with the length of the shortest path distance between species pairs in a network; while **v)** also decreasing with the number of interactions a species is experiencing. We started with the analysis of three species modules to document these principles and then showed their applicability in multi-species networks. We find that the above principles also apply in larger networks, but that the strength of pairwise associations weakens as the number of species increases.

Our results have considerable implications for interpretation of co-occurrence data. Firstly, they demonstrate the considerable variety of mechanisms causing pairwise associations. Such variety of mechanisms makes interpretation aggregated indices of co-occurrence, such as the C-score, very difficult (see also Araújo and Rozenfeld, 2014). Previous studies already made the argument that positive and negative interactions could balance each other (Boulangeat et al., 2012) and consequently associations should be studied on a pairwise basis (Veech, 2013). At least, some measure of the variability of the associations is required, and at best metrics such as network analyses (Araújo et al., 2011) should be used to characterize their complex structure. But most importantly, our analyses reveal the difficulty to infer species interactions from co-occurrence matrices. Associations are not symmetric and, therefore, indices that are capable of dealing with them are required. Null model testing is not sufficient; significance is assessed from the difference between observed co-occurrence and co-occurrence expected under independent distributions and is, consequently, symmetric. In addition, statistically significant associations cannot be interpreted as evidence of direct interactions. Our results also show that indirect interactions, and not only second order interactions, contribute to generate apparent non-independent co-occurrence. These indirect associations could be of any kind and are impossible to detect solely based on knowledge of direct interactions.

Null models of species associations should, thus, be used only to reveal the structure of co-occurrence data. The lack of an association between a pair of species is no unequivocal evidence of absence of direct interactions. It must be interpreted as the absence of a net effect

in the spatial co-occurrence arising from pairwise interaction alone. For instance, in the case species A is competing with species B and species C, and B with C, it is possible that A and C could be independently co-occurring if there is a strong indirect positive interaction A-C arising from the A-B and B-C direct interactions. Null model testing is consequently subject to important type I (false interpretation of a significant association) and type II errors (false interpretation of an absence of association). The problem itself does not come from the statistical method per se, the description of co-occurrence in the data will be right provided that the technique is adequate, but from the interpretation of the null model analysis.

Should we, therefore, abandon joint species distributions modelling (JSDM) and all of the information contained in co-distribution data? While our results might lead to such an interpretation, there is still some value in species co-occurrence data that could be used in distribution models. The appropriate use of JSDMs is to remove biases in the evaluation of species-specific relationship with the environment. Accounting for joint distribution will contribute to the evaluation of the conditional distribution of a species when all other species are absent. In other words, they should be used to improve the evaluation of the fundamental niche. The JSDMs will, however, fail to predict the right occurrence probability of a species for communities that have no analogue to the training dataset. JSDMs are using only the net associations between pairs of species and are not meant to recover the direct pairwise conditional co-occurrences. For instance, a JSDM evaluated for a plant, an herbivore and a carnivore will provide the correct description of the joint distribution of all three species, but will be of limited use to predict the distribution of the plant and the herbivore if the carnivore disappears from the system. Further developments are, consequently, required to solve the issue and account for both direct and indirect interactions. One possible solution would be to constrain JSDMs with a prior expectation of the underlying structure of direct interactions. It is also valuable to ask whether the inference of the structure of interaction networks is feasible from the observation of co-occurrences (as they result from many ecological processes). There is growing interest in inferring ecological network structure from alternative sources of information (Gravel et al., 2013; Morales-Castilla et al., 2015). This problem is challenging

because of the multiple influences on co-occurrence. Our analysis of three species modules with conditional probabilities revealed it is feasible numerically, to obtain an estimate of all pairwise conditional probabilities when accounting for higher order interactions. Known quantities are the marginal probabilities and observed co-occurrence. The parameters to be evaluated are all fundamental conditional probabilities, representing the direct associations between pairs of species (the $p_{i|j,\overline{K}}$). This is a $S \times S$ problem to solve and thus requires a significant amount of data. It might, however, be solved with large datasets where the number of sites N is much larger than S. There might also be methods to reduce the dimension of the problem because usually only a small fraction of potential interactions are met in a network (corresponding to the connectance C). While a net interaction network is likely to be fully connected ($S \times S$ links), the direct interaction network has still only a fraction C of these links realized. Bayesian approach with latent variables could even further help reducing the dimension of the problem (*e.g.* Rohr et al., 2010; Ovaskainen et al., 2010). In such methods, latent variables are evaluated for each species to represent the underlying structure of the ecological network. It was found that between two and four parameters per species would be required to successfully represent more than 80% of interactions in a predator-prey network (Rohr et al., 2010). This approach could, therefore, be used to represent the underlying structure of direct interactions and to evaluate numerically the non-null conditional probabilities. Note that these pairwise direct interactions should be interpreted specifically with reference to spatial dynamics because they would still represent phenomenologically the consequences of interactions, not the mechanisms of interactions. The next step in the development of a theory of species co-occurrence (and of species distribution) is the addition of environmental constraints. Our approach assumed a homogeneous environment, mainly for tractability of equations. We acknowledge that non-independent co-occurrence could also arise because of shared environmental requirements. The addition of environmental constraints would be easy to implement in our framework by simply making the conditional probability in absence in absence of interactions a function of the environment. Every quantity we derive after would be conditional on the environment. What would be more challenging but, nonetheless, feasi-

ble numerically, would be to make the direct interaction itself a function of the environment. There is now growing evidence that ecological interactions are context dependent (Chamberlain et al., 2014; Poisot et al., 2012). We view this integration as the next step to the derivation of a theory-driven species distribution model taking into account biotic interactions (Thuiller et al., 2013).

2.8 ACKNOWLEDGMENTS

CONCLUSION GÉNÉRALE

RÉFÉRENCES

Allesina, S., Tang, S., mar 2012. Stability criteria for complex ecosystems. Nature 483 (7388), 205–208.
URL http://www.ncbi.nlm.nih.gov/pubmed/22343894 6

Araújo, M. B., Luoto, M., nov 2007. The importance of biotic interactions for modelling species distributions under climate change. Global Ecology and Biogeography 16 (6), 743–753.
URL http://doi.wiley.com/10.1111/j.1466-8238.2007.00359.x 25, 34

Araújo, M. B., Rozenfeld, A., dec 2014. The geographic scaling of biotic interactions. Ecography 37 (5), 406–415.
URL http://doi.wiley.com/10.1111/j.1600-0587.2013.00643.x 6, 24, 26, 49

Araújo, M. B., Rozenfeld, A., Rahbek, C., Marquet, P. A., dec 2011. Using species co-occurrence networks to assess the impacts of climate change. Ecography 34 (6), 897–908.
URL http://doi.wiley.com/10.1111/j.1600-0587.2011.06919.x 25, 33, 34, 36, 44, 48, 49

Baselga, A., Araújo, M. B., apr 2009. Individualistic vs community modelling of species distributions under climate change. Ecography 32 (1), 55–65.
URL http://doi.wiley.com/10.1111/j.1600-0587.2009.05856.x 34

Baskerville, E. B., Dobson, A. P., Bedford, T., Allesina, S., Anderson, T. M., Pascual, M., dec 2011. Spatial guilds in the Serengeti food web revealed by a Bayesian group model. PLoS computational biology 7 (12), e1002321.
URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3248389{&}tool=pmcentrez{&}rendertype=abstract 26

Bell, G., jul 2005. The co-distribution of species in relation to the neutral theory of community ecology. Ecology 86 (7), 1757–1770.
URL http://www.esajournals.org/doi/abs/10.1890/04-1028 39

Berlow, E. L., Dunne, J. A., Martinez, N. D., Stark, P. B., Williams, R. J., Brose, U., 2009. Simple prediction of interaction strengths in complex food webs. Proceedings of the National Academy of Sciences of the United States of America 106 (1), 187–191. 48

Black, A. J., McKane, A. J., jun 2012. Stochastic formulation of ecological models and their applications. Trends in ecology & evolution 27 (6), 337–345.
URL http://www.ncbi.nlm.nih.gov/pubmed/22406194 7

Boulangeat, I., Gravel, D., Thuiller, W., jun 2012. Accounting for dispersal and biotic interactions to disentangle the drivers of species distributions and their abundances. Ecology

letters 15 (6), 584–593.

URL http://www.ncbi.nlm.nih.gov/pubmed/22462813 6, 25, 27, 34, 35, 44, 49

Brown, J. H., Gillooly, J. F., Allen, A. P., Savage, V. M., West, G. B., jul 2004. Toward a metabolic theory of ecology. Ecology 85 (7), 1771–1789.

URL http://www.esajournals.org/doi/abs/10.1890/03-9000 27

Cahill, A. E., Aiello-Lammens, M. E., Fisher-Reid, M. C., Hua, X., Karanewsky, C. J., Ryu, H. Y., Sbeglia, G. C., Spagnolo, F., Waldron, J. B., Warsi, O., Wiens, J. J., jan 2013. How does climate change cause extinction? Proceedings. Biological sciences / The Royal Society 280 (1750), 20121890.

URL http://www.ncbi.nlm.nih.gov/pubmed/23075836http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3574421{&}tool=pmcentrez{&}rendertype=abstract 28

Calcagno, V., Massol, F., Mouquet, N., Jarne, P., David, P., oct 2011. Constraints on food chain length arising from regional metacommunity dynamics. Proceedings. Biological sciences / The Royal Society 278 (1721), 3042–3049.

URL http://www.ncbi.nlm.nih.gov/pubmed/21367786 23

Chamberlain, S. A., Bronstein, J. L., Rudgers, J. A., jul 2014. How context dependent are species interactions? Ecology Letters 17 (7), 881–890.

URL http://www.ncbi.nlm.nih.gov/pubmed/24735225http://doi.wiley.com/10.1111/ele.12279 52

Chase, J. M., Leibold, M. A., 2003. Ecological niches : linking classical and contemporary approaches. The University of Chicago Press, Chicago.

URL http://www.americanbanker.com/issues/179{_}124/which-city-is-the-next-big-fintech-hub-new-york-stakes-its-claim-1068345-1.html$\delimiter"026E30F$nhttp://www.ncbi.nlm.nih.gov/pubmed/15003161$\delimiter"026E30F$nhttp://cid.oxfordjournals.org/lookup/doi/10.1093/cid/cir991$\delimiter"026E30F$nhttp://www.scielo.cl/pd 5

Clark, J. S., Gelfand, A. E., Woodall, C. W., Zhu, K., jul 2014. More than the sum of the parts: forest climate response from joint species distribution models. Ecological Applications 24 (5), 990–999.

URL http://www.ncbi.nlm.nih.gov/pubmed/25154092http://doi.wiley.com/10.1890/13-1015.1 34

Connor, E. F., Simberloff, D., dec 1979. The Assembly of Species Communities: Chance or Competition? Ecology 60 (6), 1132.

URL http://www.jstor.org/stable/1936961?origin=crossref 33, 35

Cook, W. M., Lane, K. T., Foster, B. L., Holt, R. D., sep 2002. Island theory, matrix effects and species richness patterns in habitat fragments. Ecology Letters 5 (5), 619–623.

URL http://doi.wiley.com/10.1046/j.1461-0248.2002.00366.x 7

Davis, A. J., Jenkinson, L. S., Lawton, J. H., Shorrocks, B., Wood, S., feb 1998. Making-mistakes when predicting shifts in species range in response to global warming. Nature 391 (6669), 783–786.
URL www.blight.comhttp://www.ncbi.nlm.nih.gov/pubmed/280836http://www.ncbi.nlm.nih.gov/pubmed/9486646http://www.nature.com/doifinder/10.1038/35842 28, 34

Diamond, J. M., 1975. Assembly of species communities. In: Cody, M. L., Diamond, J. M. (Eds.), Ecology and Evolution of Communities. Harvard University Press, Cambridge, Massachusetts, USA., pp. 342–444. 33, 35

Diamond, J. M., Gilpin, M. E., jan 1982. Examination of the "null" model of connor and simberloff for species co-occurrences on Islands. Oecologia 52 (1), 64–74.
URL http://link.springer.com/10.1007/BF00349013 22, 35

Durance, I., Ormerod, S. J., dec 2010. Evidence for the role of climate in the local extinction of a cool-water triclad. Journal of the North American Benthological Society 29 (4), 1367–1378.
URL http://www.bioone.org/doi/abs/10.1899/09-159.1 28

Eklöf, A., Jacob, U., Kopp, J., Bosch, J., Castro-Urgal, R., Chacoff, N. P., Dalsgaard, B., de Sassi, C., Galetti, M., Guimarães, P. R., Lomáscolo, S. B., Martín González, A. M., Pizo, M. A., Rader, R., Rodrigo, A., Tylianakis, J. M., Vázquez, D. P., Allesina, S., may 2013. The dimensionality of ecological networks. Ecology letters 16 (5), 577–583.
URL http://www.ncbi.nlm.nih.gov/pubmed/23438174 27

Fordham, D. A., Akçakaya, H. R., Brook, B. W., Rodríguez, A., Alves, P. C., Civantos, E., Triviño, M., Watts, M. J., Araújo, M. B., jul 2013. Adapted conservation measures are required to save the Iberian lynx in a changing climate. Nature Climate Change 3 (10), 899–903.
URL http://www.nature.com/doifinder/10.1038/nclimate1954 34

Gilman, S. E., Urban, M. C., Tewksbury, J., Gilchrist, G. W., Holt, R. D., jun 2010. A framework for community interactions under climate change. Trends in ecology & evolution 25 (6), 325–331.
URL http://www.ncbi.nlm.nih.gov/pubmed/20392517 6

Godsoe, W., Harmon, L. J., sep 2012. How do species interactions affect species distribution models? Ecography 35 (9), 811–820.
URL http://doi.wiley.com/10.1111/j.1600-0587.2011.07103.x 6, 22

Gotelli, N. J., Graves, G. R., 1996. Null models in ecology. Vol. 14. Smithsonian Institution Press, Washington, D.C.
URL http://books.google.com/books?id=vjsBAAAACAAJ{&}printsec=frontcover 33, 35

Gotelli, N. J., Graves, G. R., Rahbek, C., mar 2010. Macroecological signals of species interactions in the Danish avifauna. Proceedings of the National Academy of Sciences 107 (11), 5030–5035.
URL http://www.pnas.org/content/107/11/5030.abstract$\delimiter"026E30F$nfile://c/DocumentsandSettings/Cristina/Meusdocumentos/MyDropbox/MeuDocumentos/Papers/2010/Gotellietal2010PNAS.pdfhttp://www.pnas.org/cgi/doi/10.1073/pnas.0914089107 6, 26, 34

Gravel, D., Massol, F., Canard, E., Mouillot, D., Mouquet, N., 2011. Trophic theory of island biogeography. Ecology Letters 14 (10), 1010–1016. 5, 6, 7, 8, 11, 12, 23, 26, 36, 45, 46

Gravel, D., Poisot, T., Albouy, C., Velez, L., Mouillot, D., 2013. Inferring food web structure from predator-prey body size relationships. Methods in Ecology and Evolution 4 (11), 1083–1090.
URL http://timotheepoisot.fr/reprints/gravel{_}mee{_}niche.pdf 27, 50

Guisan, A., Thuiller, W., sep 2005. Predicting species distribution: offering more than simple habitat models. Ecology Letters 8 (9), 993–1009.
URL http://doi.wiley.com/10.1111/j.1461-0248.2005.00792.x 34

Guisan, A., Tingley, R., Baumgartner, J. B., Naujokaitis-Lewis, I., Sutcliffe, P. R., Tulloch, A. I. T., Regan, T. J., Brotons, L., McDonald-Madden, E., Mantyka-Pringle, C., Martin, T. G., Rhodes, J. R., Maggini, R., Setterfield, S. a., Elith, J., Schwartz, M. W., Wintle, B. a., Broennimann, O., Austin, M., Ferrier, S., Kearney, M. R., Possingham, H. P., Buckley, Y. M., oct 2013. Predicting species distributions for conservation decisions. Ecology letters 16, 1424–1435.
URL http://www.ncbi.nlm.nih.gov/pubmed/24134332 6

Hairston, N. G., Smith, F. E., Slobodkin, L. B., jan 1960. Community Structure, Population Control, and Competition. The American Naturalist 94 (879), 421.
URL http://www.journals.uchicago.edu/doi/abs/10.1086/282146 36

Harris, D. J., 2015. Estimating species interactions from observational data with Markov networks. bioRix preprint. 34

Holt, R. D., Barfield, M., apr 2009. Trophic interactions and range limits: the diverse roles of predation. Proceedings. Biological sciences / The Royal Society 276 (1661), 1435–1442.
URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2677217{&}tool=pmcentrez{&}rendertype=abstract 5, 7, 36, 48

Hutchinson, G. E., 1957. Concluding remarks. Cold Spring Harbor Symposium. Quantitative Biology 22, 415–427. 34

Jabot, F., Bascompte, J., mar 2012. Bitrophic interactions shape biodiversity in space. Proceedings of the National Academy of Sciences of the United States of America 109 (12), 4521–4526.
URL http://www.ncbi.nlm.nih.gov/pubmed/22393006http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3311373{&}tool=pmcentrez{&}rendertype=abstract 6, 22, 23

Jeschke, J. M., Strayer, D. L., jan 2008. Usefulness of bioclimatic models for studying climate change and invasive species. Annals of the New York Academy of Sciences 1134, 1–24.
URL http://www.ncbi.nlm.nih.gov/pubmed/18566088 5

Kearney, M., Porter, W. P., nov 2004. MAPPING THE FUNDAMENTAL NICHE: PHYSIOLOGY, CLIMATE, AND THE DISTRIBUTION OF A NOCTURNAL LIZARD. Ecology 85 (11), 3119–3131.
URL http://www.esajournals.org/doi/abs/10.1890/03-0820http://doi.wiley.com/10.1890/03-0820 4

Kéfi, S., Berlow, E. L., Wieters, E. A., Navarrete, S. A., Petchey, O. L., Wood, S. A., Boit, A., Joppa, L. N., Lafferty, K. D., Williams, R. J., Martinez, N. D., Menge, B. A., Blanchette, C. A., Iles, A. C., Brose, U., apr 2012. More than a meal… integrating non-feeding interactions into food webs. Ecology Letters 15 (4), 291–300.
URL http://www.ncbi.nlm.nih.gov/pubmed/22313549http://doi.wiley.com/10.1111/j.1461-0248.2011.01732.x 6

Kemeny, J. G., Snell, J. L., 1983. Finite Markov Chains: With a New Appendix "Generalization of a Fundamental Matrix". Undergraduate Texts in Mathematics. Springer New York.
URL https://books.google.ca/books?id=0bTK5uWzbYwC 7

Kissling, W. D., Dormann, C. F., Groeneveld, J., Hickler, T., Kühn, I., McInerny, G. J., Montoya, J. M., Römermann, C., Schiffers, K., Schurr, F. M., Singer, A., Svenning, J.-C., Zimmermann, N. E., O'Hara, R. B., dec 2012. Towards novel approaches to modelling biotic interactions in multispecies assemblages at large spatial extents. Journal of Biogeography 39 (12), 2163–2178.
URL http://doi.wiley.com/10.1111/j.1365-2699.2011.02663.x 4, 34

Lavergne, S., Mouquet, N., Thuiller, W., Ronce, O., dec 2010. Biodiversity and Climate Change: Integrating Evolutionary and Ecological Responses of Species and Communities. Annual Review of Ecology, Evolution, and Systematics 41 (1), 321–350.
URL http://www.annualreviews.org/doi/abs/10.1146/annurev-ecolsys-102209-144628 5, 28

Legendre, P., Legendre, L. F. J., 2012. Numerical Ecology. Developments in Environmental Modelling. Elsevier Science.
URL https://books.google.ca/books?id=6ZBOA-iDviQC 33, 34

Leibold, M. a., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J. M., Hoopes, M. F., Holt, R. D., Shurin, J. B., Law, R., Tilman, D., Loreau, M., Gonzalez, A., jun 2004. The metacommunity concept: a framework for multi-scale community ecology. Ecology Letters 7 (7), 601–613.
URL http://doi.wiley.com/10.1111/j.1461-0248.2004.00608.x 5, 7

Levins, R., Culver, D., jun 1971. Regional Coexistence of Species and Competition between Rare Species. Proceedings of the National Academy of Sciences 68 (6), 1246–1248.
URL papers://aa15ed4a-8b41-4036-84a6-41087bba0cd6/Paper/p811http://www.pnas.org/cgi/doi/10.1073/pnas.68.6.1246 23, 27

MacArthur, R. H., 1972. Geographical Ecology: Patterns in the Distribution of Species. Biology / [Princeton University Press]. Princeton University Press.
URL https://books.google.co.in/books?id=3NAYEKc--WAC 34

MacArthur, R. H., Wilson, E. O., 1967. Theory of Island Biogeography. Vol. 1 of Princeton Landmarks in Biology. Princeton University Press, Princeton, NJ. 7, 22, 37, 46

Massol, F., Gravel, D., Mouquet, N., Cadotte, M. W., Fukami, T., Leibold, M. A., mar 2011. Linking community and ecosystem dynamics through spatial ecology. Ecology letters 14 (3), 313–323.
URL http://www.ncbi.nlm.nih.gov/pubmed/21272182 5

McGill, B. J., apr 2010. Matters of Scale. Science 328 (5978), 575–576.
URL http://www.ncbi.nlm.nih.gov/pubmed/20431001http://www.sciencemag.org/cgi/doi/10.1126/science.1188528 26

McGill, B. J., Enquist, B. J., Weiher, E., Westoby, M., apr 2006. Rebuilding community ecology from functional traits. Trends in ecology & evolution 21 (4), 178–185.
URL http://www.ncbi.nlm.nih.gov/pubmed/16701083 27

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., Alon, U., oct 2002. Network motifs: simple building blocks of complex networks. Science (New York, N.Y.) 298 (5594), 824–7.
URL http://www.ncbi.nlm.nih.gov/pubmed/12399590 41

Montoya, J., Woodward, G., Emmerson, M. C., Solé, R. V., sep 2009. Press perturbations and indirect effects in real food webs. Ecology 90 (9), 2426–2433.
URL http://www.ncbi.nlm.nih.gov/pubmed/19769121http://doi.wiley.com/10.1890/08-0657.1 36

Morales-Castilla, I., Matias, M. G., Gravel, D., Araújo, M. B., 2015. Inferring biotic interactions from proxies. Trends in Ecology & Evolution 30 (6), 347–356.
URL http://linkinghub.elsevier.com/retrieve/pii/S0169534715000774 35, 50

Oksanen, L., Fretwell, S. D., Arruda, J., Niemela, P., aug 1981. Exploitation Ecosystems in Gradients of Primary Productivity. The American Naturalist 118 (2), 240–261.
URL http://www.journals.uchicago.edu/doi/10.1086/283817 36

Ovaskainen, O., Hottola, J., Shtonen, J., 2010. Modeling species co-occurrence by multivariate logistic regression generates new hypotheses on fungal interactions. Ecology 91 (9), 2514–2521. 51

Pearson, R. G., Dawson, T. P., sep 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? Global Ecology and Biogeography 12 (5), 361–371.
URL http://doi.wiley.com/10.1046/j.1466-822X.2003.00042.x 26

Pellissier, L., Rohr, R. P., Ndiribe, C., Pradervand, J.-N., Salamin, N., Guisan, A., Wisz, M., nov 2013. Combining food web and species distribution models for improved community projections. Ecology and evolution 3 (13), 4572–4583.
URL http://doi.wiley.com/10.1002/ece3.843http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3856755{&}tool=pmcentrez{&}rendertype=abstract 24, 26, 34

Peterson, A. T., dec 2003. Predicting the Geography of Species' Invasions via Ecological Niche Modeling. The Quarterly Review of Biology 78 (4), 419–433.
URL http://www.ncbi.nlm.nih.gov/pubmed/14737826http://www.jstor.org/stable/10.1086/378926 6

Peterson, A. T., Soberon, J., Pearson, R. G., Martinez-Meyer, E., 2011. Ecological Niches and Geographic Distributions. Vol. 2011. Princeton University Press, Princeton, NJ.
URL http://books.google.com/books?id=Q{_}h9FlvgM6wC{&}pgis=1 5, 34

Pielou, D. P., Pielou, E. C., nov 1968. Association among species of infrequent occurrence: the insect and spider fauna of Polyporus betulinus (Bulliard) Fries. Journal of theoretical biology 21 (2), 202–16.
URL http://www.ncbi.nlm.nih.gov/pubmed/5700435 35

Poisot, T., Canard, E., Mouillot, D., Mouquet, N., Gravel, D., Jordan, F., dec 2012. The dissimilarity of species interaction networks. Ecology letters 15 (12), 1353–61.
URL http://www.ncbi.nlm.nih.gov/pubmed/22994257 52

Pollock, L. J., Tingley, R., Morris, W. K., Golding, N., O'Hara, R. B., Parris, K. M., Vesk, P. A., McCarthy, M. A., may 2014. Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). Methods in Ecology and Evolution 5 (5), 397–406.
URL http://doi.wiley.com/10.1111/2041-210X.12180 24, 26, 34

Proulx, S. R., Promislow, D. E. L., Phillips, P. C., jun 2005. Network thinking in ecology and evolution. Trends in ecology & evolution 20 (6), 345–353.
URL http://www.ncbi.nlm.nih.gov/pubmed/16701391 6

Ricklefs, R. E., jan 1987. Community Diversity: Relative Roles of Local and Regional Processes. Science 235 (4785), 167–171.
URL http://www.sciencemag.org/cgi/doi/10.1126/science.235.4785.167 5

Rohr, R. P., Scherer, H., Kehrli, P., Mazza, C., Bersier, L.-F., aug 2010. Modeling food webs: exploring unexplained structure using latent traits. The American naturalist 176 (2), 170–7.
URL http://www.ncbi.nlm.nih.gov/pubmed/20540609 51

Saavedra, S., Stouffer, D. B., Uzzi, B., Bascompte, J., oct 2011. Strong contributors to network persistence are the most vulnerable to extinction. Nature 478 (7368), 233–235.
URL http://www.ncbi.nlm.nih.gov/pubmed/21918515 28

Säterberg, T., Sellman, S., Ebenman, B., jul 2013. High frequency of functional extinctions in ecological networks. Nature 499 (7459), 468–70.
URL http://www.ncbi.nlm.nih.gov/pubmed/23831648 28

Schneider, F. D., Scheu, S., Brose, U., may 2012. Body mass constraints on feeding rates determine the consequences of predator loss. Ecology letters 15 (5), 436–443.
URL http://www.ncbi.nlm.nih.gov/pubmed/22380767http://doi.wiley.com/10.1111/ele.12333 27

Schoener, T. W., 2010. The MacArthur-Wilson Equilibrium Model. In: Losos, J. B., Ricklefs, R. E. (Eds.), The Theory of Island Biogeography Revisited. Princeton University Press, Princeton, NJ, Ch. 2, pp. 52–87. 8

Simberloff, D. S., Wilson, E. O., mar 1969. Experimental Zoogeography of Islands: The Colonization of Empty Islands. Ecology 50 (2), 278–296.
URL http://doi.wiley.com/10.2307/1934856 8

Stephen P. Hubbell, 2001. The unified neutral theory of Biodiversity and Biogeography. Princeton University Press, Princeton, NJ. 39

Stone, L., Roberts, A., nov 1990. The checkerboard score and species distributions. Oecologia 85 (1), 74–79.
URL http://link.springer.com/10.1007/BF00317345 33, 35

Stouffer, D. B., Bascompte, J., feb 2010. Understanding food-web persistence from local to global scales. Ecology Letters 13 (2), 154–161.
URL http://www.ncbi.nlm.nih.gov/pubmed/19968697 41

Stouffer, D. B., Camacho, J., Jiang, W., Amaral, L. A. N., aug 2007. Evidence for the existence of a robust pattern of prey selection in food webs. Proceedings. Biological sciences / The Royal Society 274 (1621), 1931–1940.
URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2275185{&}tool=pmcentrez{&}rendertype=abstract 25, 41

Suttle, K. B., Thomsen, M. A., Power, M. E., feb 2007. Species interactions reverse grassland responses to changing climate. Science 315 (5812), 640–2.
URL http://www.ncbi.nlm.nih.gov/pubmed/17272720 28

Thébault, E., Fontaine, C., aug 2010. Stability of ecological communities and the architecture of mutualistic and trophic networks. Science 329 (5993), 853–856.
URL http://www.ncbi.nlm.nih.gov/pubmed/20705861 23

Thuiller, W., Münkemüller, T., Lavergne, S., Mouillot, D., Mouquet, N., Schiffers, K., Gravel, D., may 2013. A road map for integrating eco-evolutionary processes into biodiversity models. Ecology Letters 16, 94–105.
URL http://doi.wiley.com/10.1111/ele.12104 4, 5, 26, 52

Ulrich, W., Gotelli, N. J., 2013. Pattern detection in null model analysis. Oikos 122 (1), 2–18. 35

Urban, M. C., Leibold, M. A., Amarasekare, P., De Meester, L., Gomulkiewicz, R., Hochberg, M. E., Klausmeier, C. A., Loeuille, N., de Mazancourt, C., Norberg, J., Pantel, J. H., Strauss, S. Y., Vellend, M., Wade, M. J., jun 2008. The evolutionary ecology of metacommunities. Trends in ecology & evolution 23 (6), 311–317.
URL http://www.ncbi.nlm.nih.gov/pubmed/18439707 5

Veech, J. A., feb 2013. A probabilistic model for analysing species co-occurrence. Global Ecology and Biogeography 22 (2), 252–260.
URL http://doi.wiley.com/10.1111/j.1466-8238.2012.00789.x 24, 35, 49

Warren, B. H., Simberloff, D., Ricklefs, R. E., Aguilée, R., Condamine, F. L., Gravel, D., Morlon, H., Mouquet, N., Rosindell, J., Casquet, J., Conti, E., Cornuault, J., Fernández-Palacios, J. M., Hengl, T., Norder, S. J., Rijsdijk, K. F., Sanmartín, I., Strasberg, D., Triantis, K. a., Valente, L. M., Whittaker, R. J., Gillespie, R. G., Emerson, B. C., Thébaud, C., feb 2015. Islands as model systems in ecology and evolution: prospects fifty years after MacArthur-Wilson. Ecology Letters 18 (2), 200–217.
URL http://doi.wiley.com/10.1111/ele.12398 7

Wiens, J. J., aug 2011. The niche, biogeography and species interactions. Philosophical transactions of the Royal Society of London. Series B, Biological sciences 366 (1576), 2336–2350.
URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3130432{&}tool=pmcentrez{&}rendertype=abstract 22

Williams, R. J., Martinez, N. D., mar 2000. Simple rules yield complex food webs. Nature 404 (6774), 180–183.
URL http://www.ncbi.nlm.nih.gov/pubmed/10724169 16, 22, 46, 48

Wootton, J. T., jan 1994. Predicting Direct and Indirect Effects: An Integrated Approach Using Experiments and Path Analysis. Ecology 75 (1), 151.
URL http://www.jstor.org/stable/1939391?origin=crossref 6, 23, 36

Yodzis, P., apr 1988. The Indeterminacy of Ecological Interactions as Perceived Through Perturbation Experiments. Ecology 69 (2), 508–515.
URL http://www.jstor.org/stable/1940449?origin=crossref 36