



**INFLUENCE DES CONTRAINTES ENVIRONNEMENTALES, 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

Mois 2016

Composition du jury :

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

[Prénom Nom], directeur de recherche, [Université d'attache]

[Prénom Nom], codirecteur de recherche, [Université d'attache]

[Prénom Nom], examinateur externe, [Université d'attache]

Dépôt initial le [date mois année]

Dépôt final le [date mois année]

UNIVERSITÉ DU QUÉBEC À RIMOUSKI

Service de la bibliothèque

Avertissement

La diffusion de ce mémoire ou de cette thèse se fait dans le respect des droits de son auteur, qui a signé le formulaire « *Autorisation de reproduire et de diffuser un rapport, un mémoire ou une thèse* ». En signant ce formulaire, l'auteur concède à l'Université du Québec à Rimouski une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de son travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, l'auteur autorise l'Université du Québec à Rimouski à reproduire, diffuser, prêter, distribuer ou vendre des copies de son travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de la part de l'auteur à ses droits moraux ni à ses droits de propriété intellectuelle. Sauf entente contraire, l'auteur conserve la liberté de diffuser et de commercialiser ou non ce travail dont il possède un exemplaire.

[A JASCREB]

TABLE DES MATIÈRES

TABLE DES MATIÈRES	ix
LISTE DES ABRÉVIATIONS	xiii
LISTE DES SYMBOLES	xv
INTRODUCTION GÉNÉRALE	1
Des îles et des espèces	1
En suivant Wallace	1
En suivant MacArthur et Wilson	3
Quelles informations renferment les distributions d'espèces ?	6
Enjeux de la connaissance de la répartition géographique des espèces	7
Travail théorique et modélisation	9
Les processus qui façonnent les aires de répartition	13
Biogéographie historique	13
Capacités de dispersion	14
Contraintes abiotiques et niche écologique	15
Réseaux d'interactions : interdépendance des espèces	18
Echanges d'informations génétiques et processus micro-evolutifs	20
L'invasion européenne du frelon asiatique	21
Cadre théorique de la thèse	23
Le rôle des interactions dans la distribution des espèces	32
Importance des interactions dans la distribution	32
Un problème d'échelle ?	34
Vers une biogéographie énergétique	38
ARTICLE 1	
A PROPOS DES INTERACTIONS BIOTIQUES ET DES CONTRAINTES ENVIRONNEMENTALES A L'ECHELLE BIOGEOGRAPHIQUE	51
1.1 RESUMÉ	51
1.2 TITRE	53

1.3	AUTHORS	53
1.4	ABSTRACT	53
1.5	Introduction	54
1.6	The model	57
1.6.1	A simple probabilistic biogeographical model	57
1.6.2	Integrating biotic interactions	60
1.6.3	Integrating environmental gradients	62
1.7	Exploring the model	64
1.7.1	Biodiversity distribution over environmental gradients	66
1.8	Discussion	68
1.9	Acknowledgment	73
1.10	Tables	80
1.11	Figures	80

ARTICLE 2

THEORIE DE CO-OCCURRENCE DES ESPECES DANS LES RÉSEAUX D'INTERACTION 85

2.1	Résumé en français du deuxième article	85
2.2	TITLE	88
2.3	AUTHORS	88
2.4	ABSTRACT	88
2.5	KEYWORDS	89
2.6	Introduction	89
2.7	Definitions	93
2.8	ACKNOWLEDGMENTS	106
2.9	FIGURES	107
2.10	Context	108
2.11	Material and Methods	111
2.11.1	Datasets	111
2.11.2	Measures of co-occurrence	111
2.12	Results	112

2.12.1 Discussion (~ 4000 char)	113
2.12.2 Box 1	115
2.12.3 Tables	117
2.13 Figures	118
Reference	122
2.14 Material and methods	122
2.14.1 Datasets	122
2.14.2 Building metawebs	123
2.14.3 Co-occurrence measurement	124
2.14.4 Hypergeometric distribution	124
2.14.5 GLM and RF	125
2.15 Supporting Tables	126
2.16 Supporting Figures	133
Supporting References	142
ARTICLE 3	
ISLAND THEORY OF BIOGEOGRAPHY : FROM SPECIES TO FONCTIONNAL TRAITS	143
CONCLUSION GÉNÉRALE	145
Que peut-on prédire ?	145
3.1 Catégorie	145
3.2 Quelles type de prédictions pouvons nous faire ?	145
3.2.1 Une question d'échelle	145
3.2.2 Des classes d'espèces ?	147
3.2.3 Prédire des communautés	147
Vers une biogéographie intégrative	148
3.2.4 Les données	148
3.2.5 L'abstraction des espèces	148
Traits fonctionnels	148
des prédictions fiables ?	150

Les dangers d'aller trop vite	150
Un monde en changement : entre espoir et illusion	151
Une érosion de la biodiversité affolantes	151
Un monde biaisé ?	151
3.2.6 Avons-nous des espoirs vains ?	151
3.2.7 DEB	152

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 biogéographie des îles (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

La biogéographie est l'étude de la répartition géographiques des espèces. Aujourd'hui, ce terme est souvent remplacé par celui de macroécologie. Outre la distinction historique, ce mot met en avant l'importance du rapport des espèces à leur environnement (biotique ou abiotique) plutôt que la dimension évolutive pourtant tout aussi importante. C'est pour garder à l'esprit la richesse des facteurs qui dessinent les aires de répartition que je garde le terme de biogéographie, discipline dont je dresse un portrait dans la présente introduction. J'y aborde aussi bien la complexité de la compréhension de la distribution spatiale des espèces que les cadres théoriques associés. Chemin faisant, je discute de l'importance du lien qu'il existe entre les interactions écologiques et la répartition des espèces ; cette réflexion est l'essence même de ma thèse.

Des îles et des espèces

En suivant Wallace

Dans l'introduction de son livre *Island Life* paru en 1881, le célèbre naturaliste Alfred Russel Wallace nous rapporte deux faits étonnants qui justifient pleinement l'examen attentif de la répartition géographique des espèces (Wallace, 1881). Premièrement, le biogéographe démontre, à l'aide de nombreux exemples, que l'éloignement entre deux régions du monde n'est pas suffisant pour conclure quand à l'éloignement de leur composition faunistique et floristique. Ainsi, la comparaison des avifaunes de l'île japonaise d'Hokkaido et de l'Angleterre, séparées par des milliers de kilomètres, révèle une proximité des paysages ornithologiques très supérieure à celle constatée dans l'analyse comparée des oiseaux des îles indonésiennes de Bali et de Lombok pourtant distantes de quelques kilomètres seulement. Deuxièmement, en s'appuyant sur les différences des faunes brésiliennes et africaines, Wallace souligne la faiblesse du pouvoir prédictif des variables climatiques pour décrire les compositions fau-

nistiques présentes sous des latitudes similaires. Ces constatations soulignent l'utilité de croiser les informations des distributions à la lumière d'une analyse taxonomique pour y apporter du sens. Dans le cadre de la théorie de l'évolution¹, encore toute jeune en 1881, cette analyse taxonomique est une analyse historique : Wallace montre que la compréhension d'un problème spatial, celui des aires de répartition de groupes d'espèces, n'est possible que par une compréhension temporelle, celle de l'histoire des espèces. Cette idée est clairement énoncée dans la suite de son introduction :

« Many years study of this class of subjects has convinced me that there is no short and easy method of dealing with them ; because they are, in their very nature, the visible outcome and residual product of the whole past history of the earth. »

Tout au long de son livre, Wallace démontre que la connaissance à l'échelle mondiale de la distribution des êtres vivants permet d'associer les différentes îles aux grands ensembles régionaux biologiques (que nous appelons aujourd'hui écozones) sur la base des ressemblances biologiques des espèces qui témoignent du lien temporel unissant les différentes zones géographiques de la Terre. Ce travail de caractérisation d'ensemble géographiques conduit notamment Wallace, dans un article de 1860 (Wallace, 1860), à tracer la ligne éponyme séparant l'écozone indomalaise de l'écozone australienne (séparant les îles de Bali et Lombok mentionnée au paragraphe précédent). La connaissance apportée à la géographie par l'histoire est saisissante et les exemples de Wallace deviennent autant d'arguments en faveur de la théorie de l'évolution. Le discours de Wallace porte sur des processus à des échelles spatiales et temporelles très grandes², ce qui apporte certes un éclairage substan-

1. Wallace a publié en 1858 un article *On the Tendency of Varieties to Depart Indefinitely From the Original Type* qui témoigne très clairement que ses idées sur les variations temporelles des espèces étaient très proches de celles de Charles Darwin à qui il avait d'ailleurs envoyé son manuscrit (Wallace, 1858).

2. L'âge de la terre est très débattu à l'époque. Bien que l'ensemble des savants s'accordent pour aller bien au-delà des 6000 ans bibliques, il n'y a alors pas de consensus. Wallace affirme à la page 212 du chapitre 10 de *Island Life* que la vie se développait il y a au moins 500 millions d'années (Wallace, 1881), ce qui est

tiel qui se double cependant d'un obstacle épistémologique majeur : si l'explication ultime de la présence d'une espèce en un point donné est le produit d'une série de contingences historiques, quelles peuvent être les fondations d'une théorie de la biogéographie ? Ce n'est qu'au XX^{ème} siècle que des réponses convaincantes émergeront.

En suivant MacArthur et Wilson

Parmi les visions les plus importantes de la biogéographie ce trouve celle contenue dans le livre publié en 1967 *The Theory of Island Biogeography*, produit de la fructueuse rencontre du mathématicien et biologiste Robert MacArthur et du myrmécologue Edward Wilson³. A partir d'un grand nombre de données sur les faunes insulaires de diverses régions du monde, ces auteurs ont construit un cadre théorique puissant pour expliquer les variations du nombre d'espèce trouvé sur ces îles (MacArthur and Wilson, 1967). Leur démarche théorique permet de lier des observations à un modèle mathématique donnant une explication simple et convaincante des variations étudiées. Ils font ainsi basculer la discipline dans une ère nouvelle, ce dont les auteurs étaient conscients, en atteste le premier paragraphe du dernier chapitre de leur livre :

« Biogeography has long remained in a natural history phase, accumulating information about the distribution of species and higher taxa and the taxonomic composition of biotas. Interpretative reasoning has been largely directed to the solution of special problems connected with the histories of individuals taxa and biotas. Without doubt this descriptive activity will continue to be of fundamental

audacieux pour l'époque mais bien en-dessous de l'âge des plus anciens fossiles découverts à ce jour qui est estimé à 3.4 milliards d'années (Wacey et al., 2011).

3. Cet actuel professeur émérite à l'université d'Harvard est reconnu pour ses apports en biologie et en sociologie, il est notamment l'auteur de 32 livres. C'est pour son immense connaissance des fourmis que j'ai choisi l'adjectif de myrmécologue.

importance to the science, one of the most physically adventurous of all scientific enterprises and, in the richness of the detail it unfolds, esthetically pleasing. But biogeography is also in a position to enter an equally interesting experimental and theoretical phase. »

Dans cet extrait, MacArthur et Wilson affirment que l'étude de la distribution des espèces doit sortir du royaume des contingences historiques pour devenir un objet de science au sens d'être manipulé aussi bien expérimentalement que par l'abstraction mathématique. La validation expérimentale de la théorie a d'ailleurs été menée par Wilson et son étudiant au doctorat de l'époque Daniel Simberloff devenu depuis un célèbre écologue (Daniel S Simberloff and Edward O Wilson, 1969). Le travail d'abstraction mathématique a été conduit par MacArthur dans le livre de 1967 et prolongé dans les annexes de son livre de 1972 (MacArthur, 1972). Ces auteurs proposent une explication de la variation spécifique des îles fondée sur deux processus opposés : la colonisation d'espèce depuis le continent qui augmente le nombre d'espèce sur l'île et un processus d'extinction locale qui diminue ce nombre. C'est en reliant ces processus aux propriétés physiques de l'île (aire et isolation) et en interprétant la richesse spécifique des îles en terme d'équilibre entre ces deux processus que les auteurs parviennent à expliquer de manière convaincante les relations observées entre richesse spécifique, taille de l'île et isolement. Dans le troisième temps de cette introduction, je reviens amplement sur cette théorie nommée théorie de la biogéographie des îles que je noterai TIB dans la suite.

Le paradigme de la TIB est un legs qui a eu un impact considérable sur les développements théoriques en écologie (Warren et al., 2015). Au centre du projet de la TIB se loge la volonté de mettre l'espèce au cœur de la biogéographie afin de permettre à la discipline de s'enrichir des mécanismes biologiques qui sont un moteur essentiel de la variation dans la distribution des espèces. L'intérêt de leur *biogéographie de l'espèce* (terme donné à l'avant-dernière phrase de leur livre) est dans l'affirmation qu'il faut regarder les contraintes conjointes de l'évolution (qui met un certain nombre d'espèces avec des caractéristiques propres en présence)

et du contexte écologique qui détermine les conditions d'extinction. Cette intrication de l'écologie et de l'évolution est bien inscrite dans la pensée de MacArthur et Wilson même si la puissance de leur vision réside dans le fait de les occulter en partie.

Près de 50 ans après la parution de leur livre, une des clef en biologie semble être la compréhension des rétroactions qu'il existe entre écologie et évolution dans les variations spatiales et temporelles de la biodiversité. Je reprend ci-dessous trois aphorismes cités par Schoener (2011a) concernant les liens entre biologie, écologie et évolution :

1. « Nothing in biology makes sense except in the light of evolution. » (Dobzhansky, 1973)
2. « Nothing in evolutionary biology makes sense except in the light of ecology. » (Grant and Grant, 2008)
3. « Nothing in evolution or ecology makes sense except in the light of the other. » (F. Pelletier et al., 2009)

La chronologie de ces citations est un indice de la reconnaissance actuel du besoin (de la nécessité ?) de croiser écologie et évolution. Un parallèle avec les sciences humaines me semble possible dans lequel l'écologie serait à la biologie ce que la géographie est aux sciences humaines et l'évolution serait à la biologie ce que l'histoire est aux sciences humaines. Nous pouvons certes étudier l'une sans l'autre, mais le dialogue entre les deux disciplines est indispensable. En son absence, les disciplines avancent en trainant avec elles des hypothèses fortes sur l'autre qui finiront éventuellement par nuire à une compréhension plus profonde de la biologie. Par exemple, supposer que les variations démographiques ont des origines purement écologiques devient problématique si les variations génétiques sont suffisantes pour expliquer qu'une partie importante de cette variation comme cela l'a été montré sur une population de mutons de Soay (Pelletier et al., 2007). Je ne cherche pas à nier l'utilité des savoirs acquis de manière autonome par un champ disciplinaire, j'insiste simplement sur l'importance de mettre ces connaissances en commun dans une synthèse indispensable pour

décrypter l'information contenue dans les distributions d'espèce.

Quelles informations renferment les distributions d'espèces ?

Cette question est non seulement une invitation à découvrir les raisons de la présence de tel ou tel organisme en un lieu donné du globe, mais elle suggère aussi que certaines informations ne peuvent pas être obtenues par l'analyse de la répartition géographique des espèces. Les auteurs mentionnés dans les paragraphes précédents y ont apporté des éléments de réponse cruciaux : Wallace a montré que les distributions géographiques reflétaient en partie les liens de parenté entre les espèces, quant à MacArthur et Wilson, ils ont suggérés que ces distributions étaient le résultat de processus écologiques dynamiques. Examiner les aires de répartition, en détailler la géométrie exacte et les variations spatio-temporelle, faire des recoupements entre les répartitions géographiques de différentes espèces ou encore avec la distribution de variables abiotiques sont des démarches fondamentales pour en apprécier les mécanismes sous-jacents.

Dans son ouvrage de 1972, MacArthur discute de l'ensemble de ces mécanismes, il considère aussi bien le rôle que peuvent jouer les variables climatiques que celui des interactions écologiques. En plus des exemples concrets amenés pour illustrer ses propos, MacArthur développe des modèles mathématiques pour prolonger la discussion. Au chapitre 2, il formalise l'impact de la compétition sur la coexistence des espèces aboutissant ainsi sur un principe de ségrégation spatiale des espèces liées par ce type de relation : deux compétiteurs ne peuvent pas co-occure (résider durablement au même endroit) sauf éventuellement sur zone très restreinte de leur distribution (MacArthur, 1972). Toujours dans cet ouvrage, MacArthur évoque la distribution en damier (*checkerboard*) que peuvent générer des espèces en compétition. La discussion sur ce type de distribution sera approfondie par Jared Diamond (Diamond, 1975) dont les travaux déclencheront un débat important sur la détermination de modèle nul de co-occurrence (Connor and Simberloff, 1979) et sur laquelle ma thèse apporte quelques éléments nouveaux.

L'étude sur un grand nombre d'espèce de leurs limites spatiales permet d'y déceler des généralités quant aux mécanismes qui les déterminent (MacArthur, 1972). L'examen des variations spatio-temporelles apporte une information très utile sur l'importance relative des divers mécanismes. Le contexte des changements climatiques est une bonne illustration de ce principe car les bouleversements actuels des répartitions géographiques permettent en effet de pointer le rôle majeur de certains mécanismes évolutifs auparavant sous-estimés (Lavergne et al., 2010). Enfin, l'examen des distributions doit aussi être un examen des co-distributions, il faut s'intéresser à l'information de sous ensembles d'espèces et notamment les espèces en interaction afin de tester si la biologie laisse des empreintes dans la géométrie des aires de répartition. Par exemple, dans ma thèse je propose de regarder l'intersection des l'aire associée à un ensemble de proies pour savoir ce qu'elle nous apprend sur la distribution de leur prédateur.

Enjeux de la connaissance de la répartition géographique des espèces

Les observations et la compréhension des causes profondes de la géométrie et la dynamique des aires de répartitions des espèces ont déjà amené à des découvertes majeures en écologie et en évolution. La phase d'expérimentation et de théorisation de la biogéographie décrite par MacArthur et Wilson se poursuit et se tourne vers un objectif très ambitieux : faire de la biogéographie une discipline prédictive, pourvoyeuse de prédictions fiables sur les aires de répartitions futures de n'importe quelle espèce. Ce problème est d'autant plus présent dans la littérature récente que nous sommes dans un contexte où ces aires sont profondément bouleversées. En biogéographie, les changements climatiques ont en effet canalisés l'attention des chercheurs qui constatent avec stupeur l'ampleur à laquelle la biodiversité mondiale est affectée (Koh, 2004, Bellard et al. (2012)). La volonté d'anticiper la localisation future des espèces a également engendré des efforts conséquents pour développer des outils statistiques essentiellement centrés sur la corrélation entre les variables abiotiques et les données de présence (d'occurrence) des espèces (Elith et al., 2006).

En ciblant l'étude de la distribution de certaines espèces, la biogéographie rencontrent des enjeux socio-économiques majeurs. Ainsi, pour un pays comme la France, la restriction des zones favorables à la croissance de la vigne envisagée à l'aide des scénarios de changements climatique (Hannah et al., 2013) pourrait conduire à des pertes économiques importantes et un bouleversement identitaire des grandes régions viticoles. De plus, détecter aujourd'hui un potentiel viticole futur dans des régions où cette production n'existe pas peut conduire à des augmentations drastiques du prix des terres agricoles. En guise de second exemple, je pose la question suivant : où seront les érablières de demain ? La réponse est donnée par la détermination de la répartition future des aires favorables à la croissance de l'érable à sucre (*Acer saccharum*) et de sa capacité à les atteindre afin de s'y établir. Je termine avec un troisième exemple : la perte des pollinisateurs et notamment des abeilles. Pas moins de quatre grandes classes de facteurs d'origine anthropique les mettent en danger : les changements climatiques, le changement dans l'utilisation des terres⁴, l'apparition de nouveaux pathogènes (dont l'acarien parasite *Varroa destructa* vecteur de nombreux virus) et l'arrivée d'espèces invasives (comme le frelon asiatique) (Vanbergen, 2013). Le défi actuel est de prédire la distribution future des pollinisateurs en intégrant ces multiples aspects et leurs interactions. De plus, dans le cas des espèces invasives, il faut comprendre comment une espèce peut sortir de son aire de répartition naturelle et en établir une nouvelle.

Actuellement, les outils de prédictions des aires de répartition future reposent essentiellement sur les scénarios de changements climatiques. La démarche est cohérente : la connaissance basée sur les corrélations de variables climatiques permet d'établir une relation climat-présence. En utilisant les résultats des climatologues qui dérivent les variations climatiques associées à des scénarios d'émission de gaz à effet de serre par les activités humaine, les biogéographes établissent les probabilités de présence des espèces dans les conditions climatiques futures. Cependant, la relation climat-présence n'est qu'une facette du lien qui

4. Les changements dans l'utilisations des terres sont accompagnés, entre autres, de l'utilisation parfois massive de pesticides de la famille des néonicotinoïdes affaiblissant les colonies.

unissent les espèces à l'espace et chaque nouvelle invasion nous montre à quel point il est difficile de prédire les aires de répartition. Ces problèmes de qualité de prédictions sont le reflet de lacunes théoriques qui amènent plusieurs chercheurs à se positionner en faveur d'un renouvellement des fondations théoriques pour édifier une biogéographie plus intégrative (M. V. Lomolino, 2000, Beck et al. (2012), Thuiller et al. (2013)). Bien sur, ces appels soulèvent des défis importants dont on ne peut qu'espérer qu'ils soient relevés au plus vite pour faire face à l'urgence.

Travail théorique et modélisation

Avant d'énumérer, avec des exemples concrets, l'ensemble des forces qui régissent la répartition géographique d'une espèce, je souligne dans cette partie l'importance du travail de théorie et de modélisation qui joue un rôle prépondérant dans ma thèse.

Rassembler et intégrer des faits

Le travail de théorie est avant tout la mise en cohésion d'un certain nombre de faits, d'observations. Dans la TIB, par exemple, MacArthur et Wilson proposent une explication cohérente de l'augmentation de la richesse spécifique dans les îles de plus grande taille. Trois principes encadrent la construction d'une théorie scientifique :

1. la théorie doit pouvoir être testées (par une expérience ou par la récolte de données),
2. la théorie doit être falsifiable : la théorie demeure valide tant qu'elle n'est pas prouvée fausse, tant qu'une théorie alternative ne la supprime pas,
3. la théorie doit être parcimonieuse, ne pas invoquer de multiples processus sans raison (c'est-à-dire sans une augmentation du nombre de faits expliqués), c'est un principe qui

est aussi connu sous le nom de Rasoir d'Ockham.

Une boutade, dont je ne suis pas capable de me souvenir son auteur, énonce que les physiciens expliquent 95% de l'univers avec 5 règles alors que les économistes expliquent 5% des phénomènes qu'ils étudient avec 95 règles⁵. Le problème n'est pas tant de dénigrer une discipline mais de constater d'un côté la puissance prédictive d'une théorie mature et de l'autre, les problèmes posés par une théorie lacunaire. En biogéographie, j'ai le sentiment que les théories manquent de maturité, la TIB donne certes une vision cohérente de la richesse spécifique insulaire mais c'est une théorie peu précise : prédire un nombre d'espèce n'aide que partiellement à comprendre le monde qui nous entoure. Pour faire un peu de prospective, une théorie qui donnerait des prédictions sur la topologie des réseaux et la composition en masse des espèces présentes supplanterait la TIB car elle expliquerait davantage de faits au prix probable d'une complexité supérieure.

Des modèles pour explorer et tester la théorie

Le terme de modèle signifie simplement que l'objet en question a des propriétés bien connues. Un organisme modèle, par exemple, est un organisme souvent facile à élever et manipuler pour lequel beaucoup de connaissances ont été acquises, il sert souvent d'unité empirique pour un ou plusieurs groupes de recherche. Les modèles statistiques sont des outils pour tester des relations basées sur des hypothèses issues de théories. De même, pour un travail de modélisation mathématique, la description du modèle est contenue dans une série d'équations dérivée d'une théorie. A travers les modèles, quelle qu'en soit leur nature on explore et on teste une théorie que l'on a éventuellement participé à établir.

Les modèles sont souvent décrits comme une simplification de la réalité : comment,

5. Une variante indique que les économistes ont prédit 12 des trois dernières crises économiques. Je pense que pour ce qui est de nos capacités de prédictions, la biogéographie est plus proche de l'économie que la de la physique..

en effet, prétendre que les mécanismes biologiques décelés chez *Arabidopsis Thaliana*⁶ sont les mêmes à l'oeuvre pour l'ensemble des plantes à fleurs ? pour combien de systèmes proie-prédateur le modèle de Lotka-Volterra est-il pertinent ? Les limites des modèles doivent être reconnues mais il ne faut pas nier l'apport de ces derniers. Les modèles sont autant de chance pour explorer une ou plusieurs prédiction d'une théorie. Le choix du modèle employé est lié à l'histoire du chercheur qui l'utilise, à ses propensions mentales à utiliser avec succès telle ou telle démarche scientifique, c'est ce que rappelle Kevin McCann dans la préface de son livre *Food Webs* (McCann, 2011) :

« It just so happens that some people find it easier to think about things in terms of x's and y's, and other in terms rabbits of and lynx. »

En d'autres termes, certaines personnes ont plus de facilités pour penser en termes d'abstraction mathématique alors que d'autres font meilleur usage de manipulations plus concrètes. Je suis plutôt dans la première catégorie de personne, je pense que les mathématiques sont un cadre de penser très puissant comme l'indique le grand écologue Robert May (May, 2004) :

« The virtue of mathematics in such a contexte is that it forces clarity and precision upon the conjecture, thus enabling meaningful comparison between the consequences of basics assumptions and the empirical facts. Here mathematics is seen in its quintessence : no more, but no less, than a way to think clealy. »

Dans ma thèse, j'ai essayé d'utiliser les mathématiques pour développer des modèles dont le point de départ a été une réflexion collective autour du rôle que pouvaient jouer les interactions dans la répartition géographique des espèces. J'ai alors établi un cadre théorique

6. Il s'agit de la plante modèle par excellence dont le génome fut le premier à être séquencé chez les plantes (*Arabidopsis Genome Initiative*, 2000).

avec lequel j'ai dérivé des prédictions dont certaines semblent être vérifiées dans les données empiriques.

Nouvelles prédictions

Après l'établissement d'une théorie expliquant un certain nombre de faits et pour laquelle un certain nombre de tests ont été réalisés, le raisonnement fondé sur celle-ci peut conduire à la production de nouvelles prédictions dont la vérification la renforceront. En revanche, l'apparition des faits expérimentaux en désaccord avec cette théorie demanderont des réponses qui se traduiront soit par une meilleure compréhension de la limite d'application de la théorie soit par l'émergence d'une théorie nouvelle qui expliquera ces faits nouveaux tout en couvrant le rayon de compréhension de la théorie précédente. Ces dernières années, la physique nous a donné deux exemples très probants du pouvoir de l'imagination avec la vérification expérimentale de théories énoncées bien avant que les outils permettant de la vérifier existent. En 2012, c'est la détection du Boson de Higgs dont l'existence fut prédite en 1964⁷. Cette année, c'est la détection des ondes gravitationnelles soit 100 ans après qu'Einstein en ait prédit l'existence (Waldrop, 2016). En biogéographie, une théorie devrait être capable, par exemple, de dresser des cartes d'invasibilité à l'échelle mondiale pour l'ensemble des espèces. Je pense que nous en sommes encore loin, néanmoins, le chemin pour y parvenir passe par une connaissance approfondie de l'ensemble des mécanismes qui interviennent dans le tracé des aires de répartition, c'est-à-dire connaître leur nature, la portée de leur action mais aussi leurs interactions et leurs importances relatives.

7. Pour plus de détail, je réfère le lecteur au bulletin du CERN disponible en ligne <http://cds.cern.ch/journal/CERNBulletin/2012/28/News%20Articles/1459456?ln=fr>

Les processus qui façonnent les aires de répartition

Biogéographie historique

Il s'agit de la compréhension des impacts sur les êtres vivants des événements de grande amplitude temporelle (allant de quelques milliers d'années à plusieurs millions d'années). C'est dans l'étude de la proximité des taxons mais aussi des fossiles éventuels que l'on peut déchiffrer les mouvements de colonisation des différentes branches de l'arbre du vivant. Pour prendre un exemple de phénomène de très grande amplitude, on peut faire appel à la théorie de la dérive des continents établie par Alfred Lothar Wegener⁸ qui implique que des groupes éventuellement proches il y a des millions d'années ont été séparés et ont donné naissance à des lignées différentes. Aujourd'hui, nous sommes capables de retracer ces liens de parenté à l'aide de phylogénies moléculaires qui sont des outils très efficaces pour estimer le temps que sépare différents. Ainsi, par la comparaison des génomes mitochondriaux, il a été montré récemment que les lémuriens (primates malgaches) ont été séparés de toute autre lignée de primates il y a 60 millions d'années environs (Finstermeier et al., 2013). Cette séparation questionne bien sûr la série d'événements qui ont conduit à l'isolation de ce groupe de singes à Madagascar et à la construction des communautés que nous y observons actuellement (Razafindratsima et al., 2013).

Les processus de grandes amplitudes temporelles sont cependant dominés par leur composante historique (et donc non reproductible) et prédire des phénomènes tel que l'extinction des dinosaures est, dans le meilleur des cas, très compliqué. Néanmoins, dans les mouvements de grandes amplitudes se manifestent des processus qui sont en permanence à l'œuvre. Ainsi, l'étude de la diversification des bousiers entreprise par Joachim Hortal et collègues (Hortal et al., 2011) montre que la dernière glaciation a laissé des empreintes encore visibles dans

8. La similarité des fossiles trouvés sur des continents très éloignés a été un des arguments en faveur de cette théorie.

la carte de répartition de la diversité de ce groupe : la limite de la thermocline 0°C durant le dernier maximum glaciaire (il ya 21000 ans environs) sépare les zones de fortes diversifiées en bousier des autres. De plus, ils montrent que la diversité phylogénique des espèces nordiques, c'est-à-dire plus tolérantes au froid, est un sous-ensemble phylogénétique bien identifié, par conséquent peu de branches des bousiers sont à l'origine des colonisations nordiques. Ainsi, après une contraction de la zone favorable au développement des bousiers, les mouvements de colonisation ont marqué à la fois la carte de répartition de la richesse spécifique de ce groupe mais aussi la carte de la répartition des différentes branches de l'arbre phylogénétique des bousiers européens (Hortal et al., 2011).

Capacités de dispersion

La remonté nordique des bousiers depuis le dernier maximum glaciaire est le résultat d'événements de dispersion individuel. Au cours de leur vie, les bousiers parcourent de grandes distances à la recherche de nourriture, on peut imaginer qu'au fil des génération, si les conditions environnementales le permettent, certains individus établissent des populations de plus en plus nordiques. Ce qui est vrai pour ce groupe d'espèce mobile l'est aussi pour des espèces sessiles comme les plantes qui possèdent également des capacités de dispersion du fait de la dissémination de leurs semences par des mécanismes très diversifiés. Ce rapport à l'espace des différents organismes est une forme de diffusion : des mouvements stochastiques conduisent à une augmentation de la répartition (c'est une question de probabilité), mais cette diffusion n'est pas totalement libre.

Plusieurs types de contrainte limitent l'élargissement de l'aire de répartition d'une espèce. Pour les espèces terrestres, les mers et les océans sont des obstacles majeurs à la colonisation de nouveaux territoires. A l'échelle régionale, les rivières, les hauts reliefs peuvent fortement limiter la dispersion d'une espèce. De même, pour les plantes dont la stratégie de dissémination est l'anémochorie, la vitesse et la direction des vents sont des facteurs primordiaux pour comprendre la propagation de l'espèces. Enfin, à l'échelle du paysage, il existe

très souvent une mosaïque d'habitats plus ou moins favorables à la dispersion d'une espèce. Toutes ces possibilités sont complexes à intégrer et c'est en partie pour cela que la théorie en biogéographie a été fondée sur les îles, les flux de colonisation y étant relativement faciles à identifier : de la côte la plus proche vers l'île.

Dans l'expérience historique de Simberloff et Wilson qui valida la TIB, les chercheurs ont éradiqué la faune de six îlots de mangrove rouge dans la Baie de Floride et ils ont alors observé qu'en une année, la richesse spécifique en insecte était similaire à celle constatée avant de commencer l'expérience (Daniel S. Simberloff and Edward O. Wilson, 1969). Ainsi, les événements de colonisation, bien qu'individuel, peuvent être assez fréquents pour conduire rapidement à l'établissement de populations et même d'une communauté locale d'insecte. A l'échelle d'un continent, malgré les divers obstacles physiques, il est très probable qu'une espèce donnée puisse, en un temps plus ou moins long, atteindre n'importe quelle zone du continent. Cependant, le plus souvent, les aires de répartition des espèces sont limitées à une portion du continent. Pour comprendre ces restrictions, il faut invoquer les performances des espèces dans des conditions environnementales données.

Contraintes abiotiques et niche écologique

Dans le chapitre 6 de son livre de 1972 *Geographical Ecology* MacArthur (1972) illustre l'importance des contraintes climatiques avec l'exemple de l'aire de répartition du cactus Saguaro (*Cereus giganteus* en 1972 mais aujourd'hui *Carnegiea gigantea*). Ce résident des hauteurs du désert de Sonora (bordé à l'ouest par l'océan pacifique) est sensible au gel et ne peut guère résister à une exposition de quelques dizaines d'heures au gel. Cette contrainte physiologique explique bien les limites nord et est de sa répartition. Pour la limite sud, il semblerait que l'abondance des pluies hivernales qu'il y trouve lui soit défavorables. En s'appuyant sur les conditions climatiques actuelles dans lesquelles le cactus se développe, des résultats récents prédisent que dans le cadre des changements climatiques, *Carnegiea gigantea* trouvera refuge à des altitudes supérieures mais que ce mouvement pourrait être

entravé par l'augmentation de la fréquence des feux (Springer et al., 2015).

Cette démarche de croisement de la limite des aires de répartition avec des variables climatiques est une forme répandue de la détermination de la niche écologique d'une espèce. Ce concept de niche est très débattu en écologie et son caractère élastique s'accompagne d'un certain nombre de problèmes de définition.⁹ Afin d'éviter ces problèmes, je parlerai de la niche au sens de Joseph Grinnell qui en tentant d'expliquer la restriction de la répartition du Moqueur de Californie écrit :

« An explanation of this restricted distribution is probably to be found in the close adjustment of the bird in various physiological and psychological respects to a narrow range of environmental conditions. »

Dans ses travaux, Grinnell montre que la présence du Moqueur de Californie est fortement corrélée à des conditions de températures et d'humidité assez élevées (Grinnell, 1917). Ainsi la niche écologique au sens de Grinnell est un ensemble de conditions environnementales dans laquelle une espèce donnée est trouvée. Si on ne se restreint pas aux observations *in situ* et que l'on détermine l'ensemble des conditions d'existence possibles, alors on caractérise une niche écologique théorique appelée niche fondamentale. Cette caractérisation expérimentale a été poussée à son paroxysme dans l'article de Michael Kearney et Waren Porter sur le gecko nocturne australien *Heteronotia binoei* (Kearney and Porter, 2004). Ils ont montrés qu'en combinant des mesures physiologiques (dont le taux métaboliques au repos, la température cumulée nécessaire au bon développement des oeufs et des mesures de températures caractéristiques) avec des données climatiques, ils obtenaient une bonne concordance des probabilités d'occurrence et des observations, justifiant ainsi la démarche

9. En 1957, Hutchinson propose de voir la niche écologique comme un hyperespace (un espace d'un grand nombre de dimension) dans lequel une espèce peut se développer. Le problème est de savoir quelles sont les dimensions et notamment si les autres espèces sont parmi ces dimensions. Pour essayer d'avoir une définition plus claire de la niche écologique, certains auteurs proposent de parler de la niche comme un espace où le taux de croissance net de l'espèce est supérieur à 0 (Chase and Leibold, 2003). En dépit de l'aspect plus quantitatif de cette définition, un problème subsiste, celui de trouver une méthode générale pour trouver cet espace.

prédictive en s'appuyant sur des scénarios de changements climatiques pour aller essayer de comprendre les répartitions futures. De manière générale, cette méthode est la recherche de facteurs abiotiques limitant le développement d'une espèce et donc sa répartition géographique. Au niveau du Panama, par exemple, Engelbrecht et al. (2007) ont montrés que les distributions locales et régionales de 48 espèces d'arbres s'expliquent par la sensibilité à la sécheresse, donc à une variation dans la disponibilité d'une ressource. Ces corrélations convaincantes fondent les modèles de distributions d'espèces (SDM en référence au terme anglais utilisé souvent dans le reste de la thèse) qui sont des solutions techniques (statistiques) pour l'application de la méthode générale que je viens de décrire (Elith et al., 2006, Elith and Leathwick (2009)).

L'engouement actuel autour de ces modèles est lié à l'espoir de pouvoir faire des prédictions fiables sur la géographie de la biodiversité mondiale de demain dans un contexte de changement climatique. Cette démarche s'est appliquée avec succès à différents cas, par exemple en 2009, Tingley et collègues ont ainsi montré que sur 53 espèces d'oiseaux étudiés dans la Sierra Nevada, 48 ont colonisé de nouveaux sites où les conditions de température et de précipitations leur étaient plus favorables (Tingley et al., 2009). Une autre justification de l'utilisation abondante des SDMs est la relative facilité de leur mise en application grâce à l'abondance des données climatiques et des données d'occurrence et au partage des implémentations numériques de ces méthodes statistiques. Pour le premier type de données, WorldClim illustre cette facilité d'accès en proposant des données à l'échelle mondiale gratuitement téléchargeables (voir <http://worldclim.org>, Hijmans et al. (2005)). Pour les données d'occurrence, plusieurs initiatives offrent des données gratuites dont les plus exhaustives sont celles disponibles sur le portail de données sur la biodiversité à l'échelle mondiale GBIF (*Global Biodiversity Information Facility*, voir <http://www.gbif.org>) qui présentent cependant des biais liés à l'inégalité d'échantillonnage des régions du globe (Beck et al., 2014). Enfin pour ce qui est le partage des implémentations des SDM, on peut évoquer le logiciel libre R (R Core Team, 2015) qui a des paquets dédiés à l'utilisation des SDMs et qui sont largement utilisés dans la communauté scientifique.

Un des principaux problèmes posés par l'utilisation massive de ces approches est le manque est la faible remise en question des hypothèses sur lesquelles elles reposent. Le message délivré par les SDMs doit être pris comme une potentialité : étant donné les conditions actuelles dans lesquelles une espèce est trouvée et sachant les variations climatiques donnés par les modèles climatologiques, s'il n'existe pas d'obstacle majeur au mouvement de l'espèce en question, alors il est probable que celle-ci se déplace en suivant les conditions climatiques qui sont similaires à celles dans laquelle elle est actuellement trouvée, ce qui nous permet de savoir ou sera l'espèce demain. Ce message est délivré en supposant que 1- une forme d'équilibre de la distribution des espèces et des conditions climatiques actuelles et 2- que les espèces sont indépendantes (Jeschke and Strayer, 2008). Ces deux hypothèses sont très fortes et demandent un examen approfondi. Etant donné que ma thèse porte sur la seconde, je propose de la discuter dans le paragraphe suivant.

Réseaux d'interactions : interdépendance des espèces

Au chapitre 6 de son livre *Geographical Ecology*, MacArthur parle précisément du rôle que peut avoir la compétition dans la distribution des espèces (MacArthur, 1972). Il reprend l'exemple donné par Brown en 1971 de l'exclusion compétitive de deux espèces de de tamias, *Eutamias dorsalis* et *E. umbrinus*, dans les forêts d'altitude (au-dessus des déserts) de pins et de genévriers (*pinyon-juniper woodland*) du sud ouest des Etats-Unis. L'article de Brown montre bien comment une différence comportementale peut engendrer une séparation des distributions locales. Ainsi, l'agressivité de *Eutamias dorsalis* lui est favorable dans les forêts clairsemées de basse-altitude où son compétiteur doit dépenser beaucoup d'énergie pour lui échapper en se réfugiant dans un arbre, elle devient pénalisante lorsque l'abondance des arbres augmente car cela facilite la fuite de *E. umbrinus* (Brown, 1971). La ségrégation locale des deux espèces reflète donc bien une interaction biotique, il y donc a une information comportementale dans ces aires de répartition.

Au-delà de la compétition, l'écologie des réseaux nous montre aujourd'hui la difficulté

de concevoir les espèces comme étant des entités indépendantes, elles sont reliées par des relations de natures très diverses. Les relations trophiques sont les plus évidentes, il existe cependant une myriade d'interactions non trophiques qui affectent aussi la démographie des espèces (voir Kéfi et al. (2012) pour une réflexion sur le sujet et une classification de ces interactions). De plus, aucun argument théorique ne justifie actuellement la primauté d'un type d'interaction sur les autres. Récemment, les interactions trophiques et non-trophiques ont été exhaustivement analysées pour 104 espèces des écosystèmes intertidaux rocheux de la partie centrale de la côte chilienne révélant ainsi que les interactions non-trophiques y étaient globalement plus abondantes et concentrées sur les bas niveau trophiques (Kéfi et al., 2015).

L'écologie des réseaux est traversée de débats dont le plus important est sans doute celui de la relation qu'il existe entre la diversité spécifique d'un écosystème et sa stabilité (May, 1973, McCann (2000)). Autour de cette question, l'écologie s'est considérablement enrichi en terme d'outils mathématiques. Une preuve récente réside dans la mise en évidence par Stefano Allesina et Si Tang du caractère déstabilisant des interactions de compétition et de mutualisme et du rôle stabilisant des relations trophiques (Allesina and Tang, 2012). Ce résultat est en effet la mise en application directe d'un résultat mathématique récent établi par Terence Tao et Vam Vu (Tao et al., 2010). Les réseaux contiennent de nombreuses informations sur l'écologie des population et à mon avis, ils doivent être placés au centre d'une théorie intégrative de la biogéographie. Cette idée n'est pas seulement la mienne, MacArthur et Wilson l'ont suggérée au dernier paragraphe de leur théorie de la biogéographie avec ces mots (MacArthur and Wilson, 1967) :

« In short, biogeography appears to us to have developed to the extent that it can be reformulated in terms of the first principles of population ecology and genetics.

>

Et pour appuyer cette phrase dans son entièreté, je développe un certain nombre d'idées relatives à l'importance des échanges génétiques.

Echanges d'informations génétiques et processus micro-evolutifs

La vie, telle que nous la connaissons, pérennise l'information accumulée au cours du temps via à un support moléculaire, l'ADN. J'ai déjà évoqué que les informations véhiculées par cette molécule pouvaient permettre d'établir des relations de parenté entre les espèces. Cette possibilité est rendue possible par les mécanismes qui la modifient. L'information génétique d'un individu est un ensemble de bases dont la séquence renferme l'ensemble de l'information pour assurer le développement de l'individu. Néanmoins, le code génétique de certaines cellules de l'individu peut être modifiées (des mutations) et si ces cellules sont celles qui seront transmises à la descendance, alors ces modifications peuvent être transmises à la génération suivante. Sous certaines conditions, la mutation peut rester dans la population, c'est le moteur de la variation à l'échelle populationnelle du code génétique. Bien loin d'être une combinaison précise de paires de bases, l'ADN d'une espèce est en effet un ensemble de possibilités, un ensemble de versions du code possible mais contraint par certaines règles. Pour schématiser, les échanges de gènes doivent rester possibles entre individus d'une même espèce. A l'échelle des populations, tant que les échanges d'informations sont importants, la compatibilité est assurée mais lorsque ces échanges diminuent ou même cessent, les supports d'information peuvent alors diverger au point d'empêcher les échanges, ce qui conduit à la distinction de deux espèces. Bien que cette vision soit très simplifiée, elle permet de comprendre que l'ADN de deux espèces puissent refléter leur lien de parenté qu'il permet l'établissement d'une phylogénie moléculaire.

Cela étant dit, les causes de la divergence de l'ADN sont multiples et ce qui m'intéresse ici, ce sont que ces variations puissent engendrer un différentiel démographique positive dans un milieu nouvellement exploré par une population alors que cette même variation dans un autre milieu ne l'était pas. La vitesse des mécanismes mis en jeu semble bien plus rapide au point que ceux-ci puissent jouer des rôles prépondérants dans la réponse des espèces aux changements climatiques (Lavergne et al., 2010). En 2009, Joan Balanyá et collègues publient un article dans lequel ils comparent la composition génétique de la mouche *Drosophila*

subobscura entre des échantillons contemporains et des échantillons prélevé 24 années auparavant en Europe et en Amérique (où elle a été introduite accidentellement). Leurs résultats montrent que dans les zones de réchauffement climatique avéré, il y a aussi un changement de la composition génotypique avec une plus grande importance des génomes adaptés aux températures chaudes (Balanyá et al., 2006).

Les preuves récentes de l'impact des variations génétiques rapides sur la démographie des espèces populations poussent les chercheurs à se demander si négliger ces processus dans les travaux de dynamique des populations est une hypothèse raisonnable (F Pelletier et al., 2009, Post and Palkovacs (2009), Schoener (2011b)). Takehito Yoshida et collègues ont montré en 2003 que la réponse des algues vertes unicellulaires *Chlorella vulgaris* aux rotifères *Brachionus calyciflorus* conduit à un changement dans la fréquence et la phase des cycles de la dynamiques proie-prédateur (Yoshida et al., 2003). En 2009, une étude basée sur un suivi de plus de 20 ans d'une population de moutons Soay sur l'île d'Hirta dans l'archipel de Saint-Kilda (au nord-est de l'Écosse), Fanie Pelletier et collègues établissent les variations dans la taille corporelle des ovins, d'origine génétique, et les variations dans leur survie et leur reproduction, ils démontrent alors que les facteurs génétiques peuvent contribuer jusqu'à 20% dans la croissance de la population certaine année. Les conséquences des dynamiques eco-evolutives et l'intégration des flux d'information génétique sont certainement capitaux pour comprendre la biodiversité de demain (Sexton et al., 2009, Lavergne et al. (2010)). Nous sommes face à un enjeu appliqué capital et pourtant nos connaissances fondamentales restent insuffisantes. Pour illustrer ces lacunes et l'urgence dans laquelle nous nous trouvons, je discute d'un exemple concret : l'invasion européenne du frelon asiatique.

L'invasion européenne du frelon asiatique

Vespa velutina est une espèce présente depuis le nord-est de l'Inde jusqu'à l'est de la Chine et de la péninsule et de l'indochinoise à l'archipel indonésien (Villemant et al., 2006). Dix sous-espèces sous identifiées dont *Vespa velutina nigrithorax* qui a été observé pour la

première fois en France en 2004 dans le Lot-et-Garonne chez un producteur de bonsaï qui importe régulièrement des poteries du Yunnan (Villemant et al., 2006). Ce frelon généraliste se nourrit notamment des abeilles domestiques et les conséquences sur les récoltes de miel sont désastreuses et ce même dans les zones d'origine où l'abeille asiatique (*Apis cerana*) est pourtant capable de tuer efficacement le frelon. Pour ce faire, les abeilles forment une boule autour du frelon et battent des ailes pour augmenter la température corporelle de leur prédateur, ce qui conduit à la mort de ce dernier. L'abeille européenne (*Apis mellifera*) est capable d'utiliser la même stratégie de défense mais avec une efficacité moindre (Villemant et al., 2006). Ce frelon représente un danger pour l'entomofaune européenne et aussi menace l'apiculture qui s'ajoute aux nombreuses autres que connaît actuellement le secteur (Vanbergen, 2013). Plusieurs éléments sont remarquables dans ce cas d'invasion : c'est un cas unique (première colonisation avec succès d'une nouvelle espèce frelon en France), la rapidité de propagation de ce frelon, le besoin urgent d'anticiper sa répartition dans les prochaines années pour mettre le plus rapidement en place les mesures d'éradication.

Après son arrivée en 2004, le frelon s'étendait déjà en 2006 largement sur l'Aquitaine avec une aire de répartition française constituée d'une bande de 300km du nord au sud et de 150 km d'est en ouest (Villemant et al., 2006) et cela malgré l'éradication systématique des nids détectés. Alors que 2 nids étaient observés en 2004, 1636 nids ont été observés en 2009 et en 2013 près des trois quarts des départements étaient affectés (Robinet et al., 2016). Des travaux récents tentent de caractériser les conditions climatiques favorables au développement de l'espèce (Villemant et al., 2011) et révèlent alors qu'une large partie de l'Europe occidentale est une zone de développement probable. Un autre phénomène intéressant lié à cette invasion est que dans la même période de la colonisation européenne, le frelon est arrivé la Corée du Sud où sa propagation est cependant bien moins rapide (Villemant et al., 2011). L'explication plausible de la différence de succès de la même espèce est une différence dans la composition en espèce proche des deux régions : en Europe, il n'y a qu'une espèce de frelon *Vespa crabro*, alors qu'il y en a de six en Corée du Sud dont *Vesp mandarinia* qui est une meilleur compétitrice (Villemant et al., 2011). Cette nécessité de faire appel à la compo-

sition biologique pour comprendre les raisons d'un changement d'aire de répartition est ce qui donne tout l'intérêt du travail théorique mené durant ma thèse.

Cadre théorique de la thèse

Les développements entrepris durant ma thèse sont des tentatives d'encrage des interactions écologiques dans la TIB. Je vais maintenant revenir sur cette théorie plus en détail pour expliquer pourquoi elle a marqué durablement l'écologie. Je signale d'ailleurs que ces idées étaient partagées par d'autres écologues et qu'il y a, à ma connaissance, deux autres découvertes indépendantes des idées qui ont conduit à la théorie. La première découverte est attribuée au spécialiste des lépidoptères Eugene Gordon Munroe qui a formulé dès 1948, des idées similaires dans 5 des 555 pages de sa dissertation de graduation (Brown and Lomolino, 1989, Lomolino and Brown (2009)). La seconde est celle de Richard Levins et Harold Heatwole qui publie en 1963, soit la même année que l'article fondateur de la TIB, l'idée d'un équilibre de la richesse spécifique régit par les mêmes processus que ceux décrits par MacArthur et Wilson (Levins and Heatwole, 1963). Néanmoins, ce sont sans aucun doute MacArthur et Wilson qui ont marqués les écologues par l'ensemble des développements présentés dans leur livre de 1967, *The Theory of Island Biogeography* (MacArthur et al., 1967).

Une vision puissante de la dynamique des distributions d'espèces

Dans la préface de l'ouvrage de 1967, MacArthur et Wilson doutent que les idées proposées résisteraient longtemps à l'essor de la biogéographie expérimentale dont ils furent des acteurs de premier plan :

« We do not seriously believe that that the particular formulations advanced in in the chapters to follow will fit for very long the exacting results of future empirical invesitgation. »

Et pourtant près de 50 ans après la parution de ce livre, leurs travaux sont le fondement de nombreux développements récents, en témoigne le livre paru en 2010 *The Theory of Island Biogeography Revisited* (Losos and Ricklefs, 2010) et l'article de perspectives publié récemment par Ben Warren et collègues dans *Ecology Letters* (Warren et al., 2015). L'idée majeure de la TIB est simple et puissante : étant donné une île colonisable par un ensemble d'espèces depuis un continent voisin, la diversité locale résulte de la balance entre 1- des événements de colonisation depuis le continent et 2- des extinctions locales. La TIB est une métaphore, le cas simple d'un territoire isolé (l'île) où les flux d'individus depuis le pool d'espèces régionales (le continent) sont facilement représentables. Le modèle peut donc être étendu à de nombreux cas où un territoire isolé est colonisé par les organismes à proximité, par exemple après un incendie ou une fragmentation de l'habitat (Cook et al., 2002). Au chapitre 5 de son livre de 1972, MacArthur prend notamment l'exemple des îlots de paramo (végétation andine située au-dessus des forêts mais en-dessous des neiges éternelles). De manière générale, le modèle est acceptable est très adaptable au prix d'un certain nombre d'hypothèse notamment une certaine rigidité du réservoir d'espèces régional (au moins en nombre d'espèce) et une absence de rétroaction dans la communauté locale sur celui-ci.

Il y a une forme de hasard et de nécessité qui fait écho à l'oeuvre de Jaques Monod (Monod, 1970). Ce prix Nobel de médecine présente les mutations au niveau de l'ADN comme une source de hasard dont la persistance n'est rendu possible que dans un cadre physico-chimico-évolutifs précis, la nécessité. Dans les travaux de MacArthur et Wilson, l'événement de colonisation peut être interprété comme un pourvoyeur de stochasticité alors que les contraintes écologiques régissent l'organisation des communautés. Outre le fait que la prédiction de la colonisation ne peut se faire qu'en terme de fréquence, le caractère stochastique de cette dernière donne une dimension historique aux assemblages insulaires. L'arrivée d'une espèce est en fait un tirage aléatoire (éventuellement pondéré par les capacités respectives de dispersion) dans un réservoir régional d'une singularité historique car l'espèce en question à une histoire évolutive propre et des singularités physiologiques qui en découlent. A son arrivée sur l'île, son éventuelle insertion est déterminée par la rencontre des singula-

rités de l'espèce et du contexte biotique et abiotique de l'île. Les espèces installées sur une île ont ainsi été passées au crible des contraintes écologiques, de cette forme de nécessité qui est renouvelée à chaque nouvelle insertion. C'est ainsi que l'on peut décrire le moteur de la reconfiguration perpétuelle des réseaux écologiques locaux. Une telle dynamique peut être également analysée comme une imbrication de deux échelles de processus : régionalement, le réservoir d'espèce est façonné par une histoire évolutive de grande amplitude lié à des processus climatiques eux aussi de grande échelle, alors que les événements insulaires relèvent de processus de plus courte portée (Ricklefs, 1987).

Enfin, la TIB, bien que cela soit rarement souligné, fait l'hypothèse de l'équivalence écologique des espèces considérées : il n'y a ni plantes ni animaux, ni proies ou prédateurs, simplement des espèces qui compte pour un. Étant donné les exemples choisis par les auteurs on peut néanmoins penser que la théorie est développée pour des groupes d'espèce au rôle écologique similaire et phylogénétiquement proches. Ainsi, le premier exemple donné est pour l'herpétofaune (amphibiens et reptiles) et non sur un inventaire exhaustif de toutes les espèces de l'île (MacArthur and Wilson, 1967). Cette hypothèse est à relier aux objectifs des auteurs notamment celui d'expliquer les relations constatées entre la taille des îles et leur richesse spécifique, pour y arriver réduire les espèces à deux caractéristiques est suffisant et convenant. La démarche peut néanmoins être perçue comme antithétique pour des auteurs qui cherchent à formuler une « biogéographie de l'espèce » et de surcroît quand on connaît la qualité de ces deux naturalistes (Lomolino and Brown, 2009). Cependant, la forme d'équivalence amenée par MacArthur et Wilson ne nie la diversité et la complexité, elle est plutôt une abstraction nécessaire pour capturer les processus essentiels, pour aller au-delà des singularités des êtres vivants, vers des généralisations (Lomolino and Brown, 2009).

Le modèle mathématique et les prédictions de la TIB

Je ne rentre pas ici dans les détails mathématiques du modèle, ils sont néanmoins abordés dans le premier chapitre et aussi dans les deux annexes de la thèse¹⁰. J'écris ci-dessous l'équation qui résume à elle seule le paradigme livré par la TIB : les P espèces d'un continent colonisent l'île avec un taux individuel c , ce qui en augmente la richesse spécifique S mais augmente les risques d'extinctions dont le taux par espèce est noté e :

$$\frac{dS}{dt} = c(P - S) - eS \quad (1)$$

La dynamique ainsi engendrée conduit S jusqu'à un équilibre S_{eq} pour lequel les variations temporelles s'annulent, qui est donné par :

$$S_{eq} = P \frac{c}{c + e} \quad (2)$$

Cet équilibre est une prédiction très importante de la théorie, c'est même le point de départ des développements mathématiques dans le livre de 1967 (MacArthur and Wilson, 1967). L'existence d'un tel équilibre a été validée par l'expérience de défaunation de Simerloff et Wilson mentionnée plus haut (Daniel S Simberloff and Edward O Wilson, 1969). Une seconde prédiction de la TIB est la variation de cet équilibre avec les caractéristiques de l'île. Dès leurs article de 1963, MacArthur et Wilson présentent la taille de l'île comme un un facteur affectant le taux d'extinction : plus l'île est grande, moins le risque d'extinction est grand

10. La première annexe est un article de vulgarisation qui aborde de manière didactique la formulation la plus simple du modèle. La seconde annexe est aborde des aspects plus techniques qui ont été l'objet d'un article dont je suis co-auteur.

(MacArthur and Wilson, 1963). De même, ils supposent que l'isolement de l'île en affecte le flux de migrants : plus l'île est isolée moins les événements de colonisation sont fréquents. J'ai résumé la vision classique de la TIB sur la figure :figMW@cref

:figMW@cref

:figMW@cref

:figMW@cref

?? en y ajoutant les graphiques de l'article de 1963. Cette prédiction de la théorie en est aussi l'origine : MacArthur et Wilson expliquent avec ces mécanismes que les îles de plus grandes tailles est plus d'espèces mais aussi que des exceptions liées à l'isolement puisse exister. Cette relation est d'ailleurs présentée dès le début du chapitre 2 de la TIB avec l'augmentation linéaire du nombre d'espèce de l'herpetofaune avec le logarithme de la surface des îles de l'ouest des Caraïbes.

De manière plus générale, la TIB fournit une explication à la relation aire-espèce qui est un des objets les plus discutés de l'écologie (M. Lomolino, 2000). Il s'agit de la courbe d'augmentation de la richesse spécifique (S) avec la surface d'échantillonnage (A). La question soulevée par l'étude de ces courbes porte sur la nature des mécanismes qui régissent les variations régionales. La TIB propose une explication à cette relation et supporte une courbe de la forme $S = CA^z$ avec les observations présentées (MacArthur and Wilson, 1967). La relation aire-espèce est surtout connue pour ses applications dans le domaine de la conservation¹¹. Elle permet d'estimer la taille qu'une zone de protection doit avoir pour atteindre un objectif de sauvegarde chiffré en nombre d'espèce (Neigel, 2003, Desmet and Cowling (2004)). La relation peut être aussi utilisée dans le sens inverse pour apprécier les taux d'ex-

11. Récemment Wilson a répondu à une entrevue dans laquelle il se base sur cette relation pour indiquer la proportion de la Terre qu'il faudrait épargner afin de maximiser la sauvegarde des espèces sans pour autant empêcher le développement humain <http://www.nytimes.com/2016/03/13/opinion/sunday/the-global-solution-to-extinction.html>.

tion liés à une dégradation d'habitat (He and Hubbell, 2011).

L'importance de la TIB dans des développements théoriques plus récents

La théorie des métapopulations

Bien que ne représentant que cinq pour-cents des terres émergées, ce sont bien les observations de la faune des îles qui ont mené à une vision paradigmatique de la biogéographie. L'importance des îles s'expliquent par leur relative abondance, leur disparité, leur diversité, la relative simplicité des assemblages biologiques qu'on y trouve et aussi, comme je l'ai évoqué précédemment, par la clarté des flux de migrations (Simberloff, 1974). Cette dernière propriété est souvent absente pour des populations continentales¹². La théorie des métapopulations s'intéresse justement aux populations reliées entre elles par des flux de migrations (Hanski, 2010). Le premier modèle de métapopulations a été proposé par Levins¹³ lors d'une réflexion sur le contrôle démographique des ravageurs dans les cultures (Levins, 1969). Pour un ravageur donné, les îlots de culture sont autant de patchs où une population peut se maintenir et disperser dans les autres patchs alentour. Levins montre alors que les mesures de la lutte biologique doivent être conduites à large échelle pour en augmenter les probabilités de succès, c'est-à-dire d'extinction régional du ravageur (Levins, 1969). Le modèle est simple et très proche de celui de la TIB : l'évolution de la proportion p est aussi gouvernée par des événements de colonisation c et d'extinction e :

$$\frac{dp}{dt} = cp(1 - p) - ep \quad (3)$$

12. Les îles sont cependant souvent dans des archipels où la lecture de ces flux n'est pas si simple.

13. Richard Levins qui avec Heatwole est un des co-découvreurs des idées de la TIB.

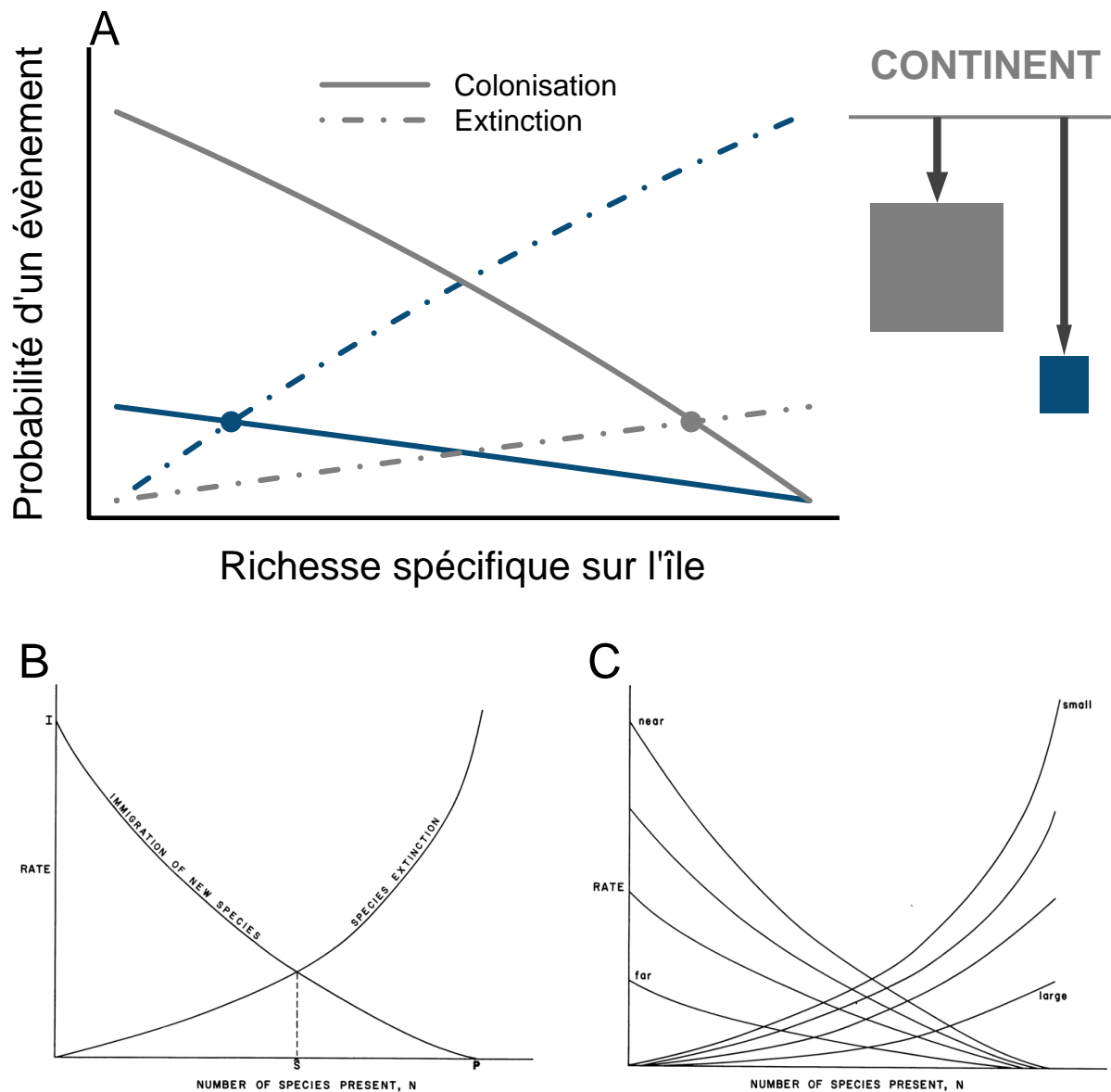


Figure 1: **La Théorie de la biogéographie des Îles.** (A) illustre l'évolution des taux de colonisation et d'extinction est présentée pour deux îles aux caractéristiques différentes. Les tailles relatives des îles et les distances qui les séparent du continent sont schématisées sur la droite, les couleurs associent les îles à leurs courbes respectives. Le réservoir d'espèce régionale (P) est constitué de 100 espèces, les taux de colonisation et d'extinction sont exprimés en terme de probabilité d'évènement (de colonisation ou d'extinction). Les points marquent les intersections entre les courbes d'extinction et de colonisation c'est-à-dire lorsque ces processus s'équilibrent. L'abscisse de ces point indique les richesses spécifiques de l'île à l'équilibre S_{eq} . (B) et (C) sont respectivement les figures 4 et 5 extraites de l'article de 1963 de MacArthur et Wilson qui livre essentiellement le même message que celui illustré en (A) (MacArthur and Wilson, 1963). La forme convexe des courbes de 1963 sont justifiées par des facteurs biologiques qui ne sont pas intégrés dans l'équation qui confère une forme concave aux courbes comme vu en (A).

La différence fondamentale avec la TIB est que la migration dépend de la proportion de patches occupés : plus elle est importante plus la migration est importante. Parmi les démonstrations il y a les travaux menés notamment par Ikkha Hanski sur les population du Mélitée du plantain (*Melitaea cinxia*) au sud-ouest de la Finland (Hanski, 1998). En plus de données un cadre de penser plus réaliste en terme de configuration spatiale, les dynamiques populationnelles associées sont bien comprises et mènent à des risques d'extinction mieux évalués (Hanski, 1998). C'est aussi un cadre apropié pour insérer l'étude des flux génétiques liés à l'arrangement spatial des populations. Ainsi, toujours sur ces mêmes populations de papillon, Ilik Saccheri et collègues montrent qu'en ajoutant le degrés d'hétérozygotie, ils obtiennent des prédictions précises quant l'extinction locale des populations (Saccheri et al., 1998). Les travaux théoriques autour du concept de metapopulations proposent un certain nombre de paradigmes qui permettent d'évaluer le rôle que joue les processus de colonisation et d'extinction dans les variations spatio-temporelles de la démographie d'une espèce (Leibold et al., 2004). La prépondérance de ces mécanismes qui font la force de la TIB et de la théorie des metapopulations a été poussée à son paroxysme dans la théorie neutre de la biogéographie.

La théorie neutre de la biogéographie et le débat qu'elle soulève

La théorie neutre postule l'équivalence écologique entre les différents individus d'espèces éventuellement différentes et décrit les dynamiques populationnelles reposant sur les différences d'abondance relative à l'échelle régionale et locale. Ainsi, en 1997, dans l'article fondateur de la théorie neutre, Stephen Hubbell décrit un modèle dans lequel le remplacement d'un individu mort dans une communauté locale est le résultat d'un tirage aléatoire : le nouvel individu peut soit être recruté localement et la probabilité que l'individu soit d'une espèce donnée dépend de l'abondance relative de cette dernière dans la communauté locale soit le nouvel individu peut-être un immigrant dont l'identité de l'espèce à laquelle il appartient est liée à l'abondance à l'échelle régionale de celle-ci (Hubbell, 1997). En plus des exemples données

dans l'article de 1997, Hubbell montre de manière convaincante que dans la forêt tropicale du Panama, à la suite d'un chablis, le recrutement de l'arbre n'est pas prévisible par ses caractéristiques et que le recrutement est similaire à la composition alentour (Hubbell, 1999). La dynamique engendrée est appelée la dérive écologique, elle dominée par la stochasticité qui conduit presque certainement à l'extinction presque certaine de toutes les espèces, ce qui est contrebalancé par l'apparition d'espèces nouvelles (Hubbell, 2010, Ricklefs (2003)).

La théorie neutre partage beaucoup de caractéristiques avec la TIB : en plus des principes fondamentaux d'extinction et de colonisation et du principe d'équivalence écologique, elle implique l'imbrication des échelles régionales et locales. Comme le fait remarquer Hubbell en 2010 dans le chapitre qu'il écrit dans *The Theory of Island Biogeography Revisited*, la théorie neutre place l'équivalence écologique au niveau des individus et non plus au niveau des espèces (Hubbell, 2010). Une conséquence directe revendiquée par Hubbell est que cette hypothèse explique la forme convexe des courbes de colonisation et d'extinction décrites par MacArthur et Wilson mais qu'elle n'explique pas leur modèle (voir :figMW@cref

:figMW@cref

:figMW@cref

:figMW@cref

?? et Hubbell (2010)). Le principe d'équivalence et la place importante que semble jouer le hasard dans cette théorie a soulevé de très vif débat avec des démonstrations à charge contre la véracité de la théorie (voir par exemple McGill and Collins (2003) et Ricklefs (2003)). A mon sens, l'équivalence écologique doit, comme dans le cas de la TIB, être prise pour une abstraction de la singularité des espèces, une simplification de la diversité des systèmes biologiques, pour isoler une portion restreinte des phénomènes qui la modifient pour en évaluer finalement le pouvoir explicatif. Bien qu'un certain nombre de cas d'études permettent de rejeter cette théorie (McGill and Collins, 2003, John et al. (2007)), les défenseurs de la théorie neutre affirment qu'elle est tout aussi utile quand une étude en démontre

la fausseté (Rosindell et al., 2012). La théorie neutre peut en effet être présentée comme une jauge qui mesure sur l'importance des processus de différenciation de niches (Wennekes et al., 2012). Ainsi pour certaines communautés la dérive écologique est plus importante que dans d'autres et du point de vue de formalisme des solutions ont déjà été proposées pour dresser un continuum de la théorie neutre vers la théorie de la niche écologique (Gravel et al., 2006). Malgré les possibilités offertes par ces deux théories, elles occultent largement les interactions écologiques qui sont factuelles ; si les observations donnent crédit à ces théories, une théorie intégrative de la biogéographie doit expliquer pourquoi.

Le rôle des interactions dans la distribution des espèces

Ma thèse a pour objectif de trouver des leviers pour comprendre comment les interactions peuvent affecter la répartition géographique des espèces et de comprendre où chercher les traces qu'elles pourraient éventuellement laisser dans les données d'occurrence des espèces. Comme je l'ai mentionné plus haut cette idée est très ancienne, Wallace le remarque dans son livre publié en 1881 :

« Both competition and predation appear now to be much more important in biogeography than people had formerly guessed » (Wallace (1881) :28)

Le problème de ces relations écologiques est leur spécificité, l'unicité de chacune d'entre elles, dont découle nos difficultés pour les prévoir bien que des travaux récents explorent des pistes prometteuses pour les prédire notamment sur la base de relations allométriques entre proie et prédateur (Gravel et al., 2013). Au point de vue théorique et à l'examen des chapitres du dernier livre de MacArthur (MacArthur, 1972), on peut dire que l'intégration des interactions est une étape clef pour aller vers une biogéographie intégrative et c'est dans cette direction que j'ai mené ma thèse en essayant d'apporter quelques pistes de réflexion.

Importance des interactions dans la distribution

Dans la théorie de la biogéographie des îles, les interactions sont en fait omniprésentes car elles sont une des composantes principales du processus d'extinction. Cependant dans la formulation du modèle, elles ne sont jamais mentionnées explicitement, cachées dans le taux d'extinction e . Comme je le montre à la figure :figMW@cref

:figMW@cref

:figMW@cref

:figMW@cref

?? , la différence dans l'allure des courbes dessinées par MacArthur et Wilson et celles obtenues en supposant un taux d'immigration et de colonisation sont différentes. D'après les auteurs, l'immigration devient plus difficile lorsque les espèces s'accumulent sur l'île et les extinctions sont de plus en plus fréquentes dues à l'intensification des interactions. Pour parler en terme de réseau d'interaction, l'accumulation d'espèces sur l'île sature le réseau local et rend difficile l'intégration d'une nouvelle espèce et le rend de plus en plus instable. Une interprétation en terme de communauté de la TIB est tout à fait possible mais les liens entre les espèces ne sont pas formulés mathématiquement en 1967.

Depuis les années 60, la littérature théorique n'a cessé de discuter le rôle joué par les interactions intra- et inter-spécifiques dans la distribution spatiale des espèces. Il est reconnu que l'interdépendance des espèces conditionne, l'aspect favorable de l'environnement au sens large (biotique et abiotique). En 2009, Robert Holt et Michael Barfield discutent de l'impact de la prédation sur la répartition d'espèces en compétition insistant alors sur le rôle majeur des interactions dans le dessin des aires de répartition (Holt and Barfield, 2009). En 2012, William Godsoe et Luke Harmon Godsoe introduisent les interactions dans un modèle simple de distribution d'espèce et montre comment la probabilité de présence d'une espèce peut être affectée par la distribution d'une seconde et concluent que cela doit affecter vraisemblable-

ment la qualité de prédictions des SDMs (Godsoe and Harmon, 2012). Ils remettent alors en question ces derniers dont triomphe à la fin du siècle dernier avait relégué les interactions écologiques au second plan en démontrant que la corrélation avec les variables climatiques étaient peut-être suffisante, au moins en première approximation pour expliquer les aires de répartitions (Pearson and Dawson, 2003). Pourtant, dès 1998, le travail précurseur d'Andrew Davis et collègues montrent que les interactions pourraient affecter nos prédictions (Davis et al., 1998) remanant ainsi largement en cause l'hypothèse d'indépendance des espèces (Jeschke and Strayer, 2008). L'expérience dont les résultats sont publiés en 1998 est une analyse d'abondance de trois espèces de drosophile le long d'un gradient de température. Les comparaisons d'abondance sont menées pour toutes les combinaisons possibles de ces trois mouches (assemblages à 1, 2 ou 3 espèces) mais aussi en présence ou en absence d'un parasitoïde. La démonstration est sans appel, la compétition et parasitisme affectent drastiquement la survie le long du gradient de température, les interactions affectent probablement les réponses aux changements climatiques.

Plus récemment, on constate une grande motivation pour intégrer les relations écologiques dans les modèles de distribution d'espèces (Kissling et al., 2012, Guisan and Rahbek (2011)). Une méthodologie récente appelée JSdM intègre par exemple les corrélations dans la présence des espèces pour améliorer les prédictions (Pollock et al., 2014). Néanmoins, ces efforts se heurtent à un manque de maturité des modèles et théories qui cherchent à rassembler distribution et interactions. Parmi les travaux récents, Franck Jabot et Jordi Bascompte ont rassemblé métacommunautés et écologie des réseaux soulignant l'importance des relations écologiques dans la répartition géographique des espèces (Jabot and Bascompte, 2012). De même, Dominique Gravel et collègues introduisent en 2011 l'interdépendance proie-prédateur dans le modèle de la TIB menant aux prémices d'une théorie trophique de la biogéographie des îles (Gravel et al., 2011) préfigurée par Holt (Holt, 2009).

C'est dans la lignée de ces développements théoriques récents que s'inscrit mon premier chapitre de thèse. J'y ai montré comment l'intégration du concept de réseau écologique

dans la TIB était possible tout en ajoutant la reconnaissance de performances plus ou moins importantes des espèces dans un contexte abiotique donné (niche écologique). Pour y arriver, j'ai montré l'intérêt de ne pas considérer des espèces une à une mais bien sous forme d'assemblage. Du point de vue technique, mon travail montre aussi qu'un retour au processus stochastique tels que ceux présentés en 1967 est une démarche puissante pour ajouter des processus dans le modèle original.

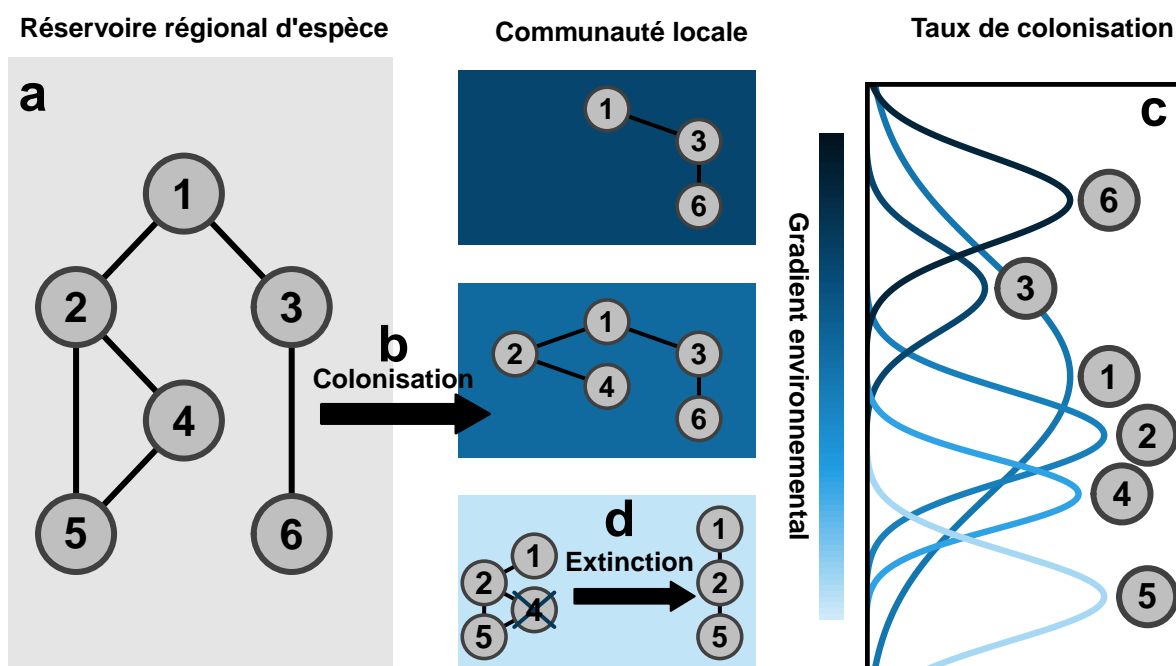


Figure 2: **Intégration des interactions et des contraintes abiotiques dans la TIB.** Pour intégrer les interactions j'ai considéré non pas un ensemble d'espèce indépendant mais un des espèces au sein d'un réseau décrit à l'échelle régionale (a). Comme dans la TIB ces espèces peuvent être colonisées l'île (b), mais dans le modèle que j'ai développé, les taux de colonisation varient avec le long d'un gradient environnemental (c). Enfin les interactions influencent les taux d'extinction locaux (d).

Un problème d'échelle ?

En repartant de l'exemple classique de la ségrégation spatiale des tamias *Eutamias dorsalis* et *E. umbrinus* (Brown, 1971), j'ai précédemment mis en évidence qu'une informa-

tion sur les interactions contenus dans l'analyse des aires de répartition. Il y a cependant deux caractéristiques importantes qui peuvent faire obstacle à l'abondance de ce type de lecture : la singularité de l'interaction et son caractère locale. Je reviens un peu plus bas sur la première propriété et m'arrête ici sur la seconde. Une idée forte relative aux interactions est leur rôle majeur à l'échelle locale qui a des conséquences de moins en moins perceptible au fur et à mesure que l'on considère des échelles spatiales de plus en plus grande (voir l'unique figure de McGill (2010)). Du point de vue théorique, c'est tout à fait ce qui est décrit dans la TIB car c'est à l'échelle locale que les interactions influencent l'extinction. Néanmoins, ces conséquences locales sont présentes sur l'ensemble de la distribution de l'espèce, il est alors pertinent de se demander pourquoi nous ne sommes pas capables de détecter les interactions en examinant les distributions d'espèces. En fait, nous avons des preuves que cela est possible dans certains cas. En 2010, Nicholas Gotelli et collègues divisent l'avifaune danoise en différentes catégories fondées sur la similarité écologique et démontrent que les espèces d'une même catégorie sont très souvent significativement ségréguées (Gotelli et al., 2010). De même, en 2007, Risto Heikkinen et collègues avaient obtenu des performances accrues de leurs modèles statistiques par l'utilisation de la répartition de six espèces de pics pour expliquer la présence de quatre espèces de hiboux (Heikkinen et al., 2007). Dans cette même étude, le signal est plus fort quand les données sont sur de 10x10km que 40x40km en faveur d'une dépendance à l'échelle, récemment supportée par d'autres travaux (Belmaker et al., 2015). Ce qui est remarquable dans les travaux de Gotelli et de Heikkinen est que l'utilisation d'une connaissance biologique et écologique a permis de révéler une trace des interactions dans la distribution d'espèces.

La dépendance spatiale de la détection des interactions est facile à comprendre : en examinant des données de présence à des échelles spatiales de plus en plus large, le nombre d'espèces s'accumule (c'est le principe de la relation aire-espèce) menant à la dégradation de l'information potentielle contenue dans des différences plus locales. Cela signifie que l'information nécessaire pour déceler des empreintes laissées par les interactions sera fournie par des données à l'échelle relativement fine, cela ne permet pas de conclure sur le rayon d'action

de ces interactions. Pour dépasser la question spatiale, il fait aussi envisagée l'impact de la nature des interactions sur la répartition géographique. Ainsi, en 2014, Miguel Araújo et Alejandro Rozenfeld ont prouvé théoriquement que les interactions positives (mutualisme) se propageaient davantage que les interactions négatives (Araújo and Rozenfeld, 2014), la nature de la relation qui unie des espèces peut donc influencer la perte d'information contenue dans les aire de répartition. Suite à mes travaux sur l'intégrations des interactions, je me suis penché sur un autre aspect qui peut influencer la perte d'information dans les données de présence : l'abondance des interactions. Au chapitre 2, je montre que les interactions directes et indirectes affectent les données de distributions mais aussi que l'abondance des interactions rend difficile de distinguer la co-occurrence d'espèces en interactions d'une co-occurrence aléatoire. Ce qui est encore plus intéressant, c'est que j'ai accumulé un certain nombre d'indices dans des données de présence et d'absence réelle qui semblent confirmer nos prédictions. Je discute de ces résultats dans le troisième chapitre de ma thèse mon troisième chapitre.

En constatant que l'abondance des interactions peut justifier l'hypothèse d'indépendance des espèces, je soulève le même paradoxe que celui relevé par MacArthur dans son oeuvre de 1972 (MacArthur, 1972) :

« A few decades ago it was fashionable for ecologists to study communities in the arctic on the grounds that these would be very simple communities and hence easy to understand. Many excellent ecologists still follow this belief, but there are others who feel that it may be easier to understand the extremely complex communities. This sounds paradoxical : How can a more complex community be easier to understand ? A possible answer might be that complex community has strong interactions among species so that the lives of the separate species are less independent than in a simple community. Where there is greater interdependence, patterns may be more conspicuous. »

Encore une fois, je déplace le problème car si l'interdépendance est importante pour

des systèmes simples, le problème est de prédire dans ces systèmes le sont. Autrement dit il serait peut-être pertinent de situer les prédictions en biogéographie au niveau du réseau écologique. Le problème d'échelle n'est plus seulement spatial et temporel il est aussi un problème d'échelle biologique : individus, population, communauté ou réseaux ?

Vers une biogéographie énergétique

Le problème d'échelle biologique est aussi un problème de catégorisation des espèces. J'ai suggéré que les prédictions étaient plus faciles pour des espèces généralistes que pour des espèces spécialistes. Malheureusement le spectre est très large et plutôt balancé avec un continuum entre des espèces hyperspécialistes et d'autres très généralistes (Timothée Poisot et al., 2015). On peut néanmoins espérer que la réduction des espèces à un certain nombre de traits (McGill et al., 2006, T. Poisot et al. (2015)) et des réseaux à un certain nombre de propriétés puissent permettre des généralisations utiles dans notre compréhension de leur distribution. Il m'apparaît aujourd'hui urgent que le bon niveau de détail dans nos descriptions des systèmes écologiques soit trouvé afin de renforcer les fondements théoriques de la dynamique des aires de répartition.

Une piste prometteuse pour prolonger la recherche des propriétés est me semble-t-il de s'appuyer sur la nature profonde des espèces : des systèmes énergétiques qui se perpétuent. La lecture de la théorie de la dynamique du budget énergétique de Sebastian Kooijman (Kooijman, 2000) m'a été très profitable pour cerner les possibilités offertes par une telle approche. Si, comme il est montré par Kooijman, il est possible de dériver de manière précise un grand nombre de propriétés énergétiques des espèces sur leur masse et leur forme, alors les espoirs sont grands de pouvoir trouver des règles d'assemblages fiables des communautés et donc de comprendre d'un point de vue mécaniste les extinctions locales. Ce sont les mêmes espoirs que ceux nourris par la théorie métabolique de l'écologie qui rassemble des relations entre la taille des espèces et différentes de leurs propriétés (Brown et al., 2004) qui montrent en somme qu'il est possible d'aller au-delà de l'espèce (T. Poisot et al., 2015). Mes réflexions

sur l'intersection entre la TIB et une vision énergétique de l'écologie sont présentées au chapitre 4 de ma thèse, dans un chapitre qui se veut aussi comme une ouverture vers les projets de recherche que j'aimerais mener dans un futur proche.

Allesina, S., Tang, S., 2012. Stability criteria for complex ecosystems. *Nature* 483, 205–208. doi :[10.1038/nature10832](https://doi.org/10.1038/nature10832)

Arabidopsis Genome Initiative, 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815. doi :[10.1038/35048692](https://doi.org/10.1038/35048692)

Araújo, M.B., Rozenfeld, A., 2014. The geographic scaling of biotic interactions. *Ecography* 37, 406–415. doi :[10.1111/j.1600-0587.2013.00643.x](https://doi.org/10.1111/j.1600-0587.2013.00643.x)

Balanyá, J., Oller, J.M., Huey, R.B., Gilchrist, G.W., Serra, L., 2006. Global genetic change tracks global climate warming in *Drosophila subobscura*. *Science* (New York, N.Y.) 313, 1773–5. doi :[10.1126/science.1131002](https://doi.org/10.1126/science.1131002)

Beck, J., Ballesteros-Mejia, L., Buchmann, C.M., Dengler, J., Fritz, S.A., Gruber, B., Hof, C., Jansen, F., Knapp, S., Kreft, H., Schneider, A.-K., Winter, M., Dormann, C.F., 2012. What's on the horizon for macroecology? *Ecography* 35, 001–011. doi :[10.1111/j.1600-0587.2012.07364.x](https://doi.org/10.1111/j.1600-0587.2012.07364.x)

Beck, J., Böller, M., Erhardt, A., Schwanghart, W., 2014. Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. *Ecological Informatics* 19, 10–15. doi :[10.1016/j.ecoinf.2013.11.002](https://doi.org/10.1016/j.ecoinf.2013.11.002)

Bellard, C., Bertelsmeier, C., Leadley, P., Thuiller, W., Courchamp, F., 2012. Impacts of climate change on the future of biodiversity. *Ecology letters* 15, 365–377. doi :[10.1111/j.1461-0248.2011.01736.x](https://doi.org/10.1111/j.1461-0248.2011.01736.x)

Belmaker, J., Zarnetske, P., Tuanmu, M.-N., Zonneveld, S., Record, S., Strecker, A., Beaudrot, L., 2015. Empirical evidence for the scale dependence of biotic interactions. *Global*

Ecology and Biogeography 24, 750–761. doi :[10.1111/geb.12311](https://doi.org/10.1111/geb.12311)

Brown, J.H., 1971. Mechanisms of Competitive Exclusion Between Two Species of Chipmunks. *Ecology* 52, 305–311. doi :[10.2307/1934589](https://doi.org/10.2307/1934589)

Brown, J.H., Gillooly, J.F., Allen, A.P., Savage, V.M., West, G.B., 2004. Toward a metabolic theory of ecology. *Ecology* 85, 1771–1789. doi :[10.1890/03-9000](https://doi.org/10.1890/03-9000)

Brown, J.H., Lomolino, M.V., 1989. Independent Discovery of the Equilibrium Theory of Island Biogeography. *Ecology* 70, 1954–1957. doi :[10.2307/1938125](https://doi.org/10.2307/1938125)

Chase, J.M., Leibold, M.A., 2003. Ecological niches : linking classical and contemporary approaches. doi :[10.1007/s13398-014-0173-7.2](https://doi.org/10.1007/s13398-014-0173-7.2)

Connor, E.F., Simberloff, D., 1979. The Assembly of Species Communities : Chance or Competition ? *Ecology* 60, 1132. doi :[10.2307/1936961](https://doi.org/10.2307/1936961)

Cook, W.M., Lane, K.T., Foster, B.L., Holt, R.D., 2002. Island theory, matrix effects and species richness patterns in habitat fragments. *Ecology Letters* 5, 619–623. doi :[10.1046/j.1461-0248.2002.00366.x](https://doi.org/10.1046/j.1461-0248.2002.00366.x)

Davis, A.J., Jenkinson, L.S., Lawton, J.H., Shorrocks, B., Wood, S., 1998. Making mistakes when predicting shifts in species range in response to global warming. *Nature* 391, 783–786. doi :[10.1038/35842](https://doi.org/10.1038/35842)

Desmet, P., Cowling, R., 2004. Using the species-area relationship to set baseline targets for conservation. *Ecology And Society* 9, 1–39.

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.

Dobzhansky, T., 1973. Nothing in Biology Makes Sense except in the Light of Evolu-

tion. *The American Biology Teacher* 35, 125–129. doi :[10.2307/4444260](https://doi.org/10.2307/4444260)

Elith, J., H. Graham, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann, F., R. Leathwick, J., Lehmann, A., Li, J., G. Lohmann, L., A. Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. M. Overton, J., Townsend Peterson, A., J. Phillips, S., Richardson, K., Scachetti-Pereira, R., E. Schapire, R., Soberón, J., Williams, S., S. Wisz, M., E. Zimmermann, N., 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29, 129–151. doi :[10.1111/j.2006.0906-7590.04596.x](https://doi.org/10.1111/j.2006.0906-7590.04596.x)

Elith, J., Leathwick, J.R., 2009. Species Distribution Models : Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics* 40, 677–697. doi :[10.1146/annurev.ecolsys.110308.120159](https://doi.org/10.1146/annurev.ecolsys.110308.120159)

Engelbrecht, B.M.J., Comita, L.S., Condit, R., Kursar, T. a, Tyree, M.T., Turner, B.L., Hubbell, S.P., 2007. Drought sensitivity shapes species distribution patterns in tropical forests. *Nature* 447, 80–82. doi :[10.1038/nature05747](https://doi.org/10.1038/nature05747)

Finstermeier, K., Zinner, D., Brameier, M., Meyer, M., Kreuz, E., Hofreiter, M., Roos, C., 2013. A Mitogenomic Phylogeny of Living Primates. *PLoS ONE* 8, 1–10. doi :[10.1371/journal.pone.0069504](https://doi.org/10.1371/journal.pone.0069504)

Godsoe, W., Harmon, L.J., 2012. How do species interactions affect species distribution models ? *Ecography* 35, 811–820. doi :[10.1111/j.1600-0587.2011.07103.x](https://doi.org/10.1111/j.1600-0587.2011.07103.x)

Gotelli, N.J., Graves, G.R., Rahbek, C., 2010. Macroecological signals of species interactions in the Danish avifauna. *Proceedings of the National Academy of Sciences* 107, 5030–5035. doi :[10.1073/pnas.0914089107](https://doi.org/10.1073/pnas.0914089107)

Grant, P.R., Grant, B.R., 2008. *How and Why Species Multiply : The Radiation of Darwin's Finches*, Princeton series in evolutionary biology. Princeton University Press.

Gravel, D., Canham, C.D., Beaudet, M., Messier, C., 2006. Reconciling niche and neu-

trality : the continuum hypothesis. *Ecology letters* 9, 399–409. doi : [10.1111/j.1461-0248.2006.00884.x](https://doi.org/10.1111/j.1461-0248.2006.00884.x)

Gravel, D., Massol, F., Canard, E., Mouillot, D., Mouquet, N., 2011. Trophic theory of island biogeography. *Ecology Letters* 14, 1010–1016. doi : [10.1111/j.1461-0248.2011.01667.x](https://doi.org/10.1111/j.1461-0248.2011.01667.x)

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, 1083–1090. doi : [10.1111/2041-210X.12103](https://doi.org/10.1111/2041-210X.12103)

Grinnell, J., 1917. The Niche-Relationships of the California Thrasher. *The Auk* 34, 427–433. doi : [10.2307/4072271](https://doi.org/10.2307/4072271)

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

Hannah, L., Roehrdanz, P.R., Ikegami, M., Shepard, A.V., Shaw, M.R., Tabor, G., Zhi, L., Marquet, P.a., Hijmans, R.J., 2013. Climate change, wine, and conservation. *Proceedings of the National Academy of Sciences* 110, 6907–6912. doi : [10.1073/pnas.1210127110](https://doi.org/10.1073/pnas.1210127110)

Hanski, I., 2010. The Theories of Island Biogeography and Metapopulation Dynamics, in : *The Theory of Island Biogeography Revisited*. Princeton University Press, Princeton, NJ, p. 476.

Hanski, I., 1998. Metapopulation dynamics. *Nature reviews* 396, 41–49. doi : [10.1016/0169-5347\(89\)90061-X](https://doi.org/10.1016/0169-5347(89)90061-X)

He, F., Hubbell, S.P., 2011. Species-area relationships always overestimate extinction rates from habitat loss. *Nature* 473, 368–71. doi : [10.1038/nature09985](https://doi.org/10.1038/nature09985)

Heikkinen, R.K., Luoto, M., Virkkala, R., Pearson, R.G., Körber, J.-H., 2007. Biotic interactions improve prediction of boreal bird distributions at macro-scales. *Global Ecology*

and Biogeography 16, 754–763. doi :[10.1111/j.1466-8238.2007.00345.x](https://doi.org/10.1111/j.1466-8238.2007.00345.x)

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25, 1965–1978. doi :[10.1002/joc.1276](https://doi.org/10.1002/joc.1276)

Holt, R.D., 2009. Bringing the Hutchinsonian niche into the 21st century : ecological and evolutionary perspectives. *Proceedings of the National Academy of Sciences of the United States of America* 106 Suppl, 19659–65. doi :[10.1073/pnas.0905137106](https://doi.org/10.1073/pnas.0905137106)

Holt, R.D., Barfield, M., 2009. Trophic interactions and range limits : the diverse roles of predation. *Proceedings. Biological sciences / The Royal Society* 276, 1435–1442. doi :[10.1098/rspb.2008.1536](https://doi.org/10.1098/rspb.2008.1536)

Hortal, J., Diniz-Filho, J.A.F., Bini, L.M., Rodríguez, M.Á., Baselga, A., Nogués-Bravo, D., Rangel, T.F., Hawkins, B.A., Lobo, J.M., 2011. Ice age climate, evolutionary constraints and diversity patterns of European dung beetles. *Ecology Letters* 14, 741–748. doi :[10.1111/j.1461-0248.2011.01634.x](https://doi.org/10.1111/j.1461-0248.2011.01634.x)

Hubbell, S.P., 2010. Neutral Theory and the Theory of Island Biogeography, in : Losos, J.B., Ricklefs, R.E. (Eds.), *The Theory of Island Biogeography Revisited*. Princeton University Press, Princeton, NJ, p. 479.

Hubbell, S.P., 1999. Light-Gap Disturbances, Recruitment Limitation, and Tree Diversity in a Neotropical Forest. *Science* 283, 554–557. doi :[10.1126/science.283.5401.554](https://doi.org/10.1126/science.283.5401.554)

Hubbell, S.P., 1997. A unified theory of biogeography and relative species abundance and its application to tropical rain forests and coral reefs. *Coral Reefs* 16, S9–S21. doi :[10.1007/s0033800502](https://doi.org/10.1007/s0033800502)

Jabot, F., Bascompte, J., 2012. Biotrophic interactions shape biodiversity in space. *Proceedings of the National Academy of Sciences of the United States of America* 109, 4521–

4526. doi :[10.1073/pnas.1107004109](https://doi.org/10.1073/pnas.1107004109)

Jeschke, J.M., Strayer, D.L., 2008. Usefulness of bioclimatic models for studying climate change and invasive species. *Annals of the New York Academy of Sciences* 1134, 1–24. doi :[10.1196/annals.1439.002](https://doi.org/10.1196/annals.1439.002)

John, R., Dalling, J.W., Harms, K.E., Yavitt, J.B., Stallard, R.F., Mirabello, M., Hubbell, S.P., Valencia, R., Navarrete, H., Vallejo, M., Foster, R.B., 2007. Soil nutrients influence spatial distributions of tropical tree species. *Proceedings of the National Academy of Sciences* 104, 864–869. doi :[10.1073/pnas.0604666104](https://doi.org/10.1073/pnas.0604666104)

Kearney, M., Porter, W.P., 2004. MAPPING THE FUNDAMENTAL NICHE : PHYSIOLOGY, CLIMATE, AND THE DISTRIBUTION OF A NOCTURNAL LIZARD. *Ecology* 85, 3119–3131. doi :[10.1890/03-0820](https://doi.org/10.1890/03-0820)

Kéfi, S., Berlow, E.L., Wieters, E.A., Joppa, L.N., Wood, S.A., Brose, U., Navarrete, S.A., 2015. Network structure beyond food webs : mapping non-trophic and trophic interactions on Chilean rocky shores. *Ecology* 96, 291–303. doi :[10.1890/13-1424.1](https://doi.org/10.1890/13-1424.1)

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., 2012. More than a meal... integrating non-feeding interactions into food webs. *Ecology Letters* 15, 291–300. doi :[10.1111/j.1461-0248.2011.01732.x](https://doi.org/10.1111/j.1461-0248.2011.01732.x)

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., 2012. Towards novel approaches to modelling biotic interactions in multispecies assemblages at large spatial extents. *Journal of Biogeography* 39, 2163–2178. doi :[10.1111/j.1365-2699.2011.02663.x](https://doi.org/10.1111/j.1365-2699.2011.02663.x)

Koh, L.P., 2004. Species Coextinctions and the Biodiversity Crisis. *Science* 305, 1632–

1634. doi :[10.1126/science.1101101](https://doi.org/10.1126/science.1101101)

Kooijman, S.A.L.M., 2000. *Dynamic Energy and Mass Budgets in Biological Systems*. Cambridge University Press, Cambridge. doi :[10.1017/CBO9780511565403](https://doi.org/10.1017/CBO9780511565403)

Lavergne, S., Mouquet, N., Thuiller, W., Ronce, O., 2010. Biodiversity and Climate Change : Integrating Evolutionary and Ecological Responses of Species and Communities. *Annual Review of Ecology, Evolution, and Systematics* 41, 321–350. doi :[10.1146/annurev-ecolsys-102209-144628](https://doi.org/10.1146/annurev-ecolsys-102209-144628)

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., 2004. The meta-community concept : a framework for multi-scale community ecology. *Ecology Letters* 7, 601–613. doi :[10.1111/j.1461-0248.2004.00608.x](https://doi.org/10.1111/j.1461-0248.2004.00608.x)

Levins, R., 1969. Some Demographic and Genetic Consequences of Environmental Heterogeneity for Biological Control. *Bulletin of the Entomological Society of America* 15, 237–240. doi :[10.1093/besa/15.3.237](https://doi.org/10.1093/besa/15.3.237)

Levins, R., Heatwole, H., 1963. On the distribution of organisms on islands. *Caribbean Journal of Science* 3, 173–177.

Lomolino, M., 2000. Ecology's most general, yet protean pattern : the species area relationship. *Journal of Biogeography* 27, 17–26.

Lomolino, M.V., 2000. A call for a new paradigm of island biogeography. *Global Ecology and Biogeography* 9, 1–6. doi :[10.1046/j.1365-2699.2000.00185.x](https://doi.org/10.1046/j.1365-2699.2000.00185.x)

Lomolino, M.V., Brown, J.H., 2009. The reticulating phylogeny of island biogeography theory. *Q. Rev. Biol.* 84, 357–390. doi :[10.1017/CBO9781107415324.004](https://doi.org/10.1017/CBO9781107415324.004)

Losos, J.B., Ricklefs, R.E., 2010. *The Theory of Island Biogeography Revisited*. Prin-

ceton University Press, Princeton, NJ.

MacArthur, R.H., 1972. *Geographical Ecology : Patterns in the Distribution of Species*, Biology / [princeton university press]. Princeton University Press.

MacArthur, R.H., Wilson, E.O., 1967. *Theory of Island Biogeography*, Princeton landmarks in biology. Princeton University Press, Princeton, NJ.

MacArthur, R.H., Wilson, E.O., 1963. An equilibrium theory of insular zoogeography. *Evolution* 17, 373–387.

MacArthur, R.H., Wilson, E.O., MacArthur, W., 1967. The theory of island biogeography. doi :[10.2307/1796430](https://doi.org/10.2307/1796430)

May, R.M., 2004. Uses and abuses of mathematics in biology. *Science* (New York, N.Y.) 303, 790–3. doi :[10.1126/science.1094442](https://doi.org/10.1126/science.1094442)

May, R.M., 1973. Stability and complexity in model ecosystems. *Monographs in population biology* 6, 1–235. doi :[10.1109/TSMC.1978.4309856](https://doi.org/10.1109/TSMC.1978.4309856)

McCann, K.S., 2011. *Food Webs*, Monographs in population biology. Princeton University Press.

McCann, K.S., 2000. The diversity-stability debate. *Nature* 405, 228–33. doi :[10.1038/35012234](https://doi.org/10.1038/35012234)

McGill, B., Collins, C., 2003. A unified theory for macroecology based on spatial patterns of abundance. *Evolutionary Ecology Research* 5, 469–492. doi :[10.1038/nature01569.1](https://doi.org/10.1038/nature01569.1).

McGill, B.J., 2010. Matters of Scale. *Science* 328, 575–576. doi :[10.1126/science.1188528](https://doi.org/10.1126/science.1188528)

McGill, B.J., Enquist, B.J., Weiher, E., Westoby, M., 2006. Rebuilding community eco-

logy from functional traits. *Trends in ecology & evolution* 21, 178–185. doi : [10.1016/j.tree.2006.02.002](https://doi.org/10.1016/j.tree.2006.02.002)

Monod, J., 1970. *Le hasard et la nécessité* : Editions Du Seuil.

Neigel, J., 2003. Species-area relationships and marine conservation. *Ecological Applications* 13, 138–145. doi : [10.1890/1051-0761\(2003\)013\[0138:SARAMC\]2.0.CO;2](https://doi.org/10.1890/1051-0761(2003)013[0138:SARAMC]2.0.CO;2)

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

Pelletier, F., Clutton-Brock, T., Pemberton, J., Tuljapurkar, S., Coulson, T., 2007. The evolutionary demography of ecological change : Linking trait variation and population growth. *Science* 315, 1571–1574. doi : [10.1126/science.1139024](https://doi.org/10.1126/science.1139024)

Pelletier, F., Garant, D., Hendry, a P., 2009. Eco-evolutionary dynamics. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 364, 1483–9. doi : [10.1098/rstb.2009.0027](https://doi.org/10.1098/rstb.2009.0027)

Pelletier, F., Garant, D., Hendry, A., 2009. Eco-evolutionary dynamics. *Philosophical Transactions of the Royal Society B : Biological Sciences* 364, 1483–1489. doi : [10.1098/rstb.2009.0027](https://doi.org/10.1098/rstb.2009.0027)

Poisot, T., Kéfi, S., Morand, S., Stanko, M., Marquet, P.A., Hochberg, M.E., 2015. A continuum of specialists and generalists in empirical communities. *PLoS ONE* 10, 1–12. doi : [10.1371/journal.pone.0114674](https://doi.org/10.1371/journal.pone.0114674)

Poisot, T., Stouffer, D.B., Gravel, D., 2015. Beyond species : why ecological interactions vary through space and time. *Oikos* 124, 243–251. doi : [10.1101/001677](https://doi.org/10.1101/001677)

Pollock, L.J., Tingley, R., Morris, W.K., Golding, N., O'Hara, R.B., Parris, K.M., Vesk, P.A., McCarthy, M.A., 2014. Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Methods in Ecology and Evolution*

5, 397–406. doi :[10.1111/2041-210X.12180](https://doi.org/10.1111/2041-210X.12180)

Post, D.M., Palkovacs, E.P., 2009. Eco-evolutionary feedbacks in community and ecosystem ecology : interactions between the ecological theatre and the evolutionary play. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 364, 1629–40. doi :[10.1098/rstb.2009.0012](https://doi.org/10.1098/rstb.2009.0012)

R Core Team, 2015. R : A Language and Environment for Statistical Computing.

Razafindratsima, O.H., Mehtani, S., Dunham, A.E., 2013. Extinctions, traits and phylogenetic community structure : Insights from primate assemblages in Madagascar. *Ecography* 36, 047–056. doi :[10.1111/j.1600-0587.2011.07409.x](https://doi.org/10.1111/j.1600-0587.2011.07409.x)

Ricklefs, R.E., 2003. A comment on Hubbell 's zero-sum ecological drift model. *Oikos* 1001, 185–192.

Ricklefs, R.E., 1987. Community diversity : relative roles of local and regional processes. *Science* 235, 167–171. doi :[10.1126/science.235.4785.167](https://doi.org/10.1126/science.235.4785.167)

Robinet, C., Suppo, C., Darrouzet, E., 2016. Rapid spread of the invasive yellow-legged hornet in France : the role of human-mediated dispersal and the effects of control measures. *Journal of Applied Ecology*. doi :[10.1111/1365-2664.12724](https://doi.org/10.1111/1365-2664.12724)

Rosindell, J., Hubbell, S.P., He, F., Harmon, L.J., Etienne, R.S., 2012. The case for ecological neutral theory. *Trends in Ecology and Evolution* 27, 203–208. doi :[10.1016/j.tree.2012.01.004](https://doi.org/10.1016/j.tree.2012.01.004)

Saccheri, I., Kuussaari, M., Kankare, M., Vikman, P., Fortelius, W., Hanski, I., 1998. Inbreeding and extinction in a butterfly metapopulation. *Nature* 392, 491–494. doi :[Doi 10.1038/33136](https://doi.org/10.1038/33136)

Schoener, T.W., 2011a. The Newest Synthesis : Understanding Ecological Dynamics. *Science* 331, 426–429. doi :[10.1126/science.1193954](https://doi.org/10.1126/science.1193954)

Schoener, T.W., 2011b. The newest synthesis : understanding the interplay of evolution-

nary and ecological dynamics. *Science* (New York, N.Y.) 331, 426–9. doi :[10.1126/science.1193954](https://doi.org/10.1126/science.1193954)

Sexton, J.P., McIntyre, P.J., Angert, A.L., Rice, K.J., 2009. Evolution and Ecology of Species Range Limits. *Annual Review of Ecology, Evolution, and Systematics* 40, 415–436. doi :[10.1146/annurev.ecolsys.110308.120317](https://doi.org/10.1146/annurev.ecolsys.110308.120317)

Simberloff, D.S., 1974. Equilibrium Theory of Island Biogeography and Ecology. *Annual Review of Ecology and Systematics* 5, 161–182. doi :[10.1146/annurev.es.05.110174.001113](https://doi.org/10.1146/annurev.es.05.110174.001113)

Simberloff, D.S., Wilson, E.O., 1969. Experimental Zoogeography of Islands : The Colonization of Empty Islands. *Ecology* 50, 278–296. doi :[10.2307/1934856](https://doi.org/10.2307/1934856)

Simberloff, D.S., Wilson, E.O., 1969. Experimental zoogeography of islands : a model for insular colonization. *Ecology* 50, 296–314. doi :[10.2307/1934856](https://doi.org/10.2307/1934856)

Springer, A., Swann, D., Crimmins, M., 2015. Climate change impacts on high elevation saguaro range expansion. *Journal of Arid Environments* 116, 57–62. doi :[10.1016/j.jaridenv.2015.02.004](https://doi.org/10.1016/j.jaridenv.2015.02.004)

Tao, T., Vu, V., Krishnapur, M., 2010. Random matrices : Universality of ESDs and the circular law. *The Annals of Probability* 38, 2023–2065. doi :[10.1214/10-AOP534](https://doi.org/10.1214/10-AOP534)

Thuiller, W., Münkemüller, T., Lavergne, S., Mouillot, D., Mouquet, N., Schiffrers, K., Gravel, D., 2013. A road map for integrating eco-evolutionary processes into biodiversity models. *Ecology Letters* 16, 94–105. doi :[10.1111/ele.12104](https://doi.org/10.1111/ele.12104)

Tingley, M.W., Monahan, W.B., Beissinger, S.R., Moritz, C., 2009. Birds track their Grinnellian niche through a century of climate change. *Proceedings of the National Academy of Sciences of the United States of America* 106 Suppl, 19637–43. doi :[10.1073/pnas.0901562106](https://doi.org/10.1073/pnas.0901562106)

Vanbergen, A.J., 2013. Threats to an ecosystem service : Pressures on pollinators. *Frontiers in Ecology and the Environment* 11, 251–259. doi :[10.1890/120126](https://doi.org/10.1890/120126)

Villemant, C., Barbet-Massin, M., Perrard, A., Muller, F., Gargominy, O., Jiguet, F., Rome, Q., 2011. Predicting the invasion risk by the alien bee-hawking Yellow-legged hornet

Vespa velutina nigrithorax across Europe and other continents with niche models. *Biological Conservation* 144, 2142–2150. doi :[10.1016/j.biocon.2011.04.009](https://doi.org/10.1016/j.biocon.2011.04.009)

Villemant, C., Haxaire, J., Streito, J., 2006. Premier bilan de l’invasion de *Vespa velutina* Lepeletier en France (Hymenoptera, Vespidae). *Bulletin de la Société entomologique de France* 111, 535–538.

Wacey, D., Kilburn, M.R., Saunders, M., Cliff, J., Brasier, M.D., 2011. Microfossils of sulphur-metabolizing cells in 3.4-billion-year-old rocks of Western Australia. *Nature Geoscience* 4, 698–702. doi :[10.1038/ngeo1238](https://doi.org/10.1038/ngeo1238)

Waldrop, M.M., 2016. The hundred-year quest for gravitational waves — in pictures. *Nature*. doi :[10.1038/nature.2016.19340](https://doi.org/10.1038/nature.2016.19340)

Wallace, A.R., 1881. *Island Life : Or, The Phenomena and Causes of Insular Faunas and Floras, Including a Revision and Attempted Solution of the Problem of Geological Climates*. Harper & brothers.

Wallace, A.R., 1860. On the Zoological Geography of the Malay Archipelago. *Journal of the Proceedings of the Linnean Society of London. Zoology* 4, 172–184. doi :[10.1111/j.1096-3642.1860.tb00090.x](https://doi.org/10.1111/j.1096-3642.1860.tb00090.x)

Wallace, A.R., 1858. On the Tendency of Varieties to depart indefinitely from the Original Type. *Proceedings of the Linnean Society Of London* 3, 53–62.

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., Rijdsdijk, K.F., Sanmartín, I., Strasberg, D., Triantis, K.a., Valente, L.M., Whittaker, R.J., Gillespie, R.G., Emerson, B.C., Thébaud, C., 2015. Islands as model systems in ecology and evolution : prospects fifty years after MacArthur-Wilson. *Ecology Letters* 18, 200–217. doi :[10.1111/ele.12398](https://doi.org/10.1111/ele.12398)

Wennekes, P.L., Rosindell, J., Etienne, R.S., 2012. The Neutral—Niche Debate : A

Philosophical Perspective. *Acta Biotheoretica* 60, 257–271. doi :[10.1007/s10441-012-9144-6](https://doi.org/10.1007/s10441-012-9144-6)

Yoshida, T., Jones, L.E., Ellner, S.P., Fussmann, G.F., Hairston, N.G., 2003. Rapid evolution drives ecological dynamics in a predator-prey system. *Nature* 424, 303–6. doi :[10.1038/nature01767](https://doi.org/10.1038/nature01767)

ARTICLE 1

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

1.1 RESUMÉ

En 1967, Robert MacArthur et Edward Osborne Wilson publient leur théorie de la biogéographie des îles. Leur connaissance de l'organisation spatiale du vivant, acquise lors de nombreuses expériences de terrain, les conduisent à une vision très puissante de la biogéographie qui reste aujourd'hui un des piliers de l'écologie. Le tour de force de ces auteurs a été de confiner dans un modèle très simple décrivant la relation qu'il existe entre la diversité à l'échelle locale (une île) et celle à l'échelle régionale (un continent). Une île est en fait un espace géographique limité que les espèces du continent peuvent venir coloniser avec un certain succès qui dépend de la facilité d'accès de l'île en question. De plus, les auteurs ajoutent que la probabilité de survie des espèces est liée à la quantité de ressource présente sur l'île, ce qu'ils relient à sa taille. Avec ces hypothèses peu contraignantes, ils parviennent à expliquer de manière cohérente la répartition de la biodiversité dans différents archipels et plusieurs travaux dans les années suivantes étayeront leur propos.

La théorie de la biogéographie des îles 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 précisément 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 résultats 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 article 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 prédictions 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 probabilités de colonisation des espèces du continent ont été reliées aux conditions environnementales de l'île. Après avoir montré et développé comment le modèle a été construit et donné des prédictions simples, nous nous sommes intéressés à des scénarios portant sur 10 espèces et pour des types d'interactions différents : mutualistes, prédation et compétitions le long de gradients environnementaux. Ce qui apparaît ressort de mes simulations est un portrait des impacts potentielles des interactions sur la distributions des espèces. Dependamment de leur nature et de leur nombre, les interactions peuvent changer drastiquement la biodiversité attendue dans le cadre de la 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 développé 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.

1.2 TITLE

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

1.3 AUTHORS

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

1.4 ABSTRACT

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 (?). The extensive development of statistical models of species distributions based on actual ranges and environmental data have provided valuable knowledge and predictions (?), but often remain purely correlative. There is now consensus that future developments in biogeography will require solving critical limitations of species distribution models (?) and incorporating explicitly biotic interactions and dispersal (?). 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 (?): 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 (?). The metacommunity concept has been proposed as a simple framework to link different spatial scales in ecology (?). 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 (?), and spatial food webs (??). 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 (?). 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 (?). Recent developments refined this definition based on demography and metapopulation dynamics (?). The abiotic niche, often referred as the Grinnelian niche, has been central to the development of species distribution models (SDMs, ?). Despite all of its criticisms, SDMs remain remarkably popular and operational for conservation ecology (?). 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 (?). 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 (?). Theory predicts that interactions in small community modules (2-4 species) should influence range limits (?), 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 (?). However an increasing number of studies emphasizes the role of local interactions as a major factor influencing geographical ranges (??). The representation of interactions in a network is a convenient method to summarize the type and strength of interactions among species, their organization (?) and their consequences on dynamics (?). Food webs were first considered in the development of a trophic theory of biogeography (?), 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 (?). Mutualism, competition and indirect effects (?), for instance, also impact local environmental suitability (?). Tools and knowledge acquired through the study of local ecological networks, such as the community matrix and metrics of structure (?), 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) (??) 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 (??). The TIB emphasizes the role of regional processes to local community assembly. Indeed it can be regarded as the simplest representation of metacommunity dynamics (?). Furthermore, the model is easily expandable. Following (?), ? 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 generalize 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 (??) to derive expected assemblages and co-distribution at both the local and the re-

gional scale. We illustrate how the interplay between biotic interactions and environmental requirements can affect the distribution of biodiversity 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 (?)) 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:

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

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 (?), 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 (?)) 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:

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

$\mathbb{P}(X_{i,t+dt}|X_{i,t})$ 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:

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

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 $\mathbf{Y}_{t>0} = (\mathbf{X}_{1,t>0}, \mathbf{X}_{2,t>0}, \dots, \mathbf{X}_{P,t>0})$ which leads to the description of 2^P assemblages 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:

$$\begin{aligned} \mathbb{P}(\mathbf{Y}_{t+dt} = \text{"state k"} | \mathbf{Y}_t = \text{"state l"}) &= \prod_{i_1 \in I_1} c_{i_1} dt \prod_{i_2 \in I_2} (1 - c_{i_2} dt) \\ &\quad \prod_{i_3 \in I_3} e_{i_3} dt \prod_{i_4 \in I_4} (1 - e_{i_4} dt) \end{aligned} \quad (1.4)$$

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.3, 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.3B). Finally, we can calculate

the expected number of species in a locality (Fig.3C), 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 (?). 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.

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 $\mathbf{v} = \mathbf{A}\mathbf{Y}_t^T$ (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 (? considered that at least one interaction was required to persist, which is very similar).

$$f_i(\mathbf{v}) = f(\mathbf{v}, (e_i, e_{i,min}, e_{i,max}, d_i))$$

$$= e_{i,min} + \frac{1}{\frac{1}{e_{i,max}-e_{i,min}} + \left(\frac{1}{e_i-e_{i,min}} - \frac{1}{e_{i,max}-e_{i,min}} \right) \exp(d_i * v_i)} \quad (1.5)$$

$$g_i(\mathbf{v}) = c_i \quad (1.6)$$

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 4. 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.4B and Fig.4C). 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.4D).

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} : $\mathbf{w} = (w_1, w_2, \dots, w_n)$. 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 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)):

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

Note that the dimension of $\mathbf{M}(\mathbf{w})$ 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: $\mathbf{C}_t = (\mathbb{P}(\mathbf{Y}_t = \text{"state 1"}), \mathbb{P}(\mathbf{Y}_t = \text{"state 2"}), \dots, \mathbb{P}(\mathbf{Y}_t = \text{"state } 2^P\text{"}))$. The Markov Chain formalism defines the probability of the future community composition at time $t + dt$ as $\mathbf{C}_{t+dt} = \mathbf{C}_t \mathbf{M}$. \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.

$$\lim_{\substack{l \rightarrow +\infty \\ l \in \mathbb{N}}} \mathbf{C}_0 \mathbf{M}^l = \mathbf{C}_{eq} \quad (1.8)$$

\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.

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

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 5 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 4) 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 (?). 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 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 \mathbf{A} : (i) no interaction \mathbf{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 \mathbf{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 \mathbf{A} matrices generated (connectance set to 0.2) and we calculate the vector \mathbf{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 6 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.6). 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 strength of interactions: for $d = 10$ (Panel C in Fig. 6), 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. 6), 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. 6 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 \mathbf{A} matrices, this could potentially be caused by the variability of the location of non-zero values in \mathbf{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 \mathbf{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 parameters are set as follows: $e_i = 10^{-5}$, $e_{i,min} = 10^{-3}e_i$, $e_{i,max} = 10^3e_i$ and $c_i = 10^{-5}$. Like so we obtain the profile of species richness we report on Figure 7.

For all scenarios, the richness is maximal at the center of the environmental gradient (Fig. 7). 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. 7: 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. 7). 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. 7), 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. 7). 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.

1.8 Discussion

Understanding how colonization-extinction dynamics influence species distribution and community structure remains a major challenge in biogeography (??). 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 (?) where environmental gradients were not introduced and the co-occurrence among species was not modelled, despite empirical evidence of their impact (?). By taking these constraints together we reveal how they interplay 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 (?). 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 (?). 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 (?). 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 (?). A strength of our model is that it not only takes all direct interactions into account, but also indirect ones (?). 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 (?). 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 be used to study the emergence of indirect interactions in networks at the large spatial scale.

The challenge of developing joint species distribution models (??) have recently motivated researchers to investigate co-occurrence (??). 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 demography by definition. This reality enters the model by either enhancing or 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 (?) 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 (?). 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 (?). 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 (?).

The importance of interactions across different scales is still debated (??). 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 (?). This argument persists even though theoretical (?) and empirical (?) 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 (?) or correlation between species (?) 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 (?). 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 (?). 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 (?), 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 (?). The island-mainland approximation remains elegant but might be too simple for applied situations. One solution may be to identify source and sink localities, (?). 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 (?).

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 (?). Beyond body size, other functional traits (?) could help us to escape from species singularity toward more general rules. Hence, approximating food web structure could be done using traits (?) and energetic requirements can be easily quantified through body size and local temperature constraints (?). 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. (?), there is growing evidence that the response of species to climate change must be studied at the community scale (?). Even though species respond individually to climate change, they are constrained by complex direct and indirect biotic interactions emerging from large scale organization (?). The study of Cahill and colleagues (?) 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 (?)). For instance, species contributing to the persistence of plant-pollinator networks are paradoxically the most vulnerable to extinction (?), 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 (?). 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 (?). We think that the assemblage-based approach we propose here is a promising perspective to introduce interactions in biogeographical models.

1.9 Acknowledgment

We thank Chantal L. Hutchison for insightful comments on the manuscript. KC was supported by a grant from the Ministry of Higher Education and Research of France. DG was supported by a NSERC Discovery grant and the Canadian Research Chair program. NM was supported by the CNRS.

Captions

Tables

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 matrix of our markov chain where species are assumed to be independent. We release this hypothesis by linking extinction coefficients and species assemblages.

Figures

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.

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.

Figure 3

Equilibrium for interacting species along an environmental gradient. The colonization probability of species i , c_i , is related to the environment variable \mathbf{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).

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: $e_i = 10^{-5}$, $e_{i,min} = 10^{-3}e_i$, $e_{i,max} = 10^3e_i$ 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 (?) 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).

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: $e_i = 10^{-5}$, $e_{i,min} = 10^{-3}e_i$, $e_{i,max} = 10^3e_i$ and $c_i = 10^{-5}$.

1.10 Tables

Table 1

$(\mathbf{X}_{1,t}, \mathbf{X}_{2,t})$	$(\mathbf{X}_{1,t+dt}, \mathbf{X}_{2,t+dt})$			
	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$
$(0,0)$	$(1 - c_1dt)(1 - c_2dt)$	$(1 - c_1dt)c_2dt$	$c_1dt(1 - c_2dt)$	c_1dte_2dt
$(0,1)$	$(1 - c_1dt)e_2dt$	$(1 - c_1dt)(1 - e_2dt)$	c_1dte_2dt	$c_1dt(1 - e_2dt)$
$(1,0)$	$e_1dt(1 - c_2dt)$	e_1dte_2dt	$(1 - e_1dt)(1 - c_2dt)$	$(1 - e_1dt)c_2dt$
$(1,1)$	e_1dte_2dt	$e_1dt(1 - e_2dt)$	$(1 - e_1dt)e_2dt$	$(1 - e_1dt)(1 - e_2dt)$

Table 1

1.11 Figures

Figure 1

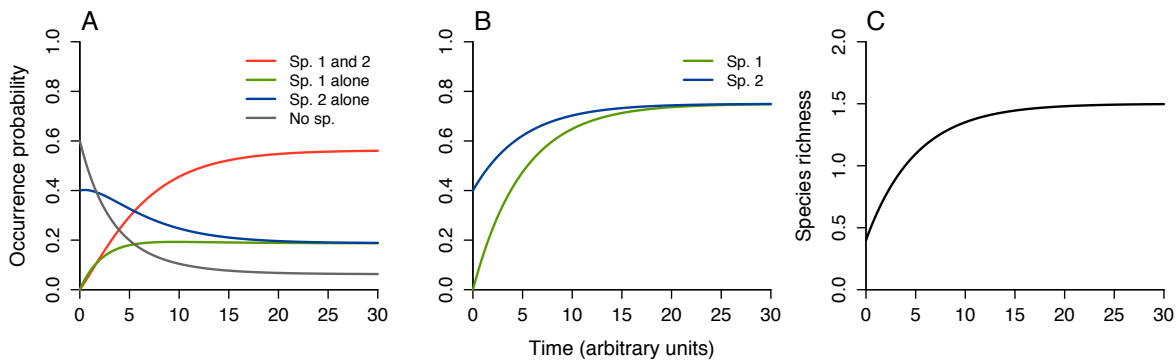


Figure 3

Figure 2

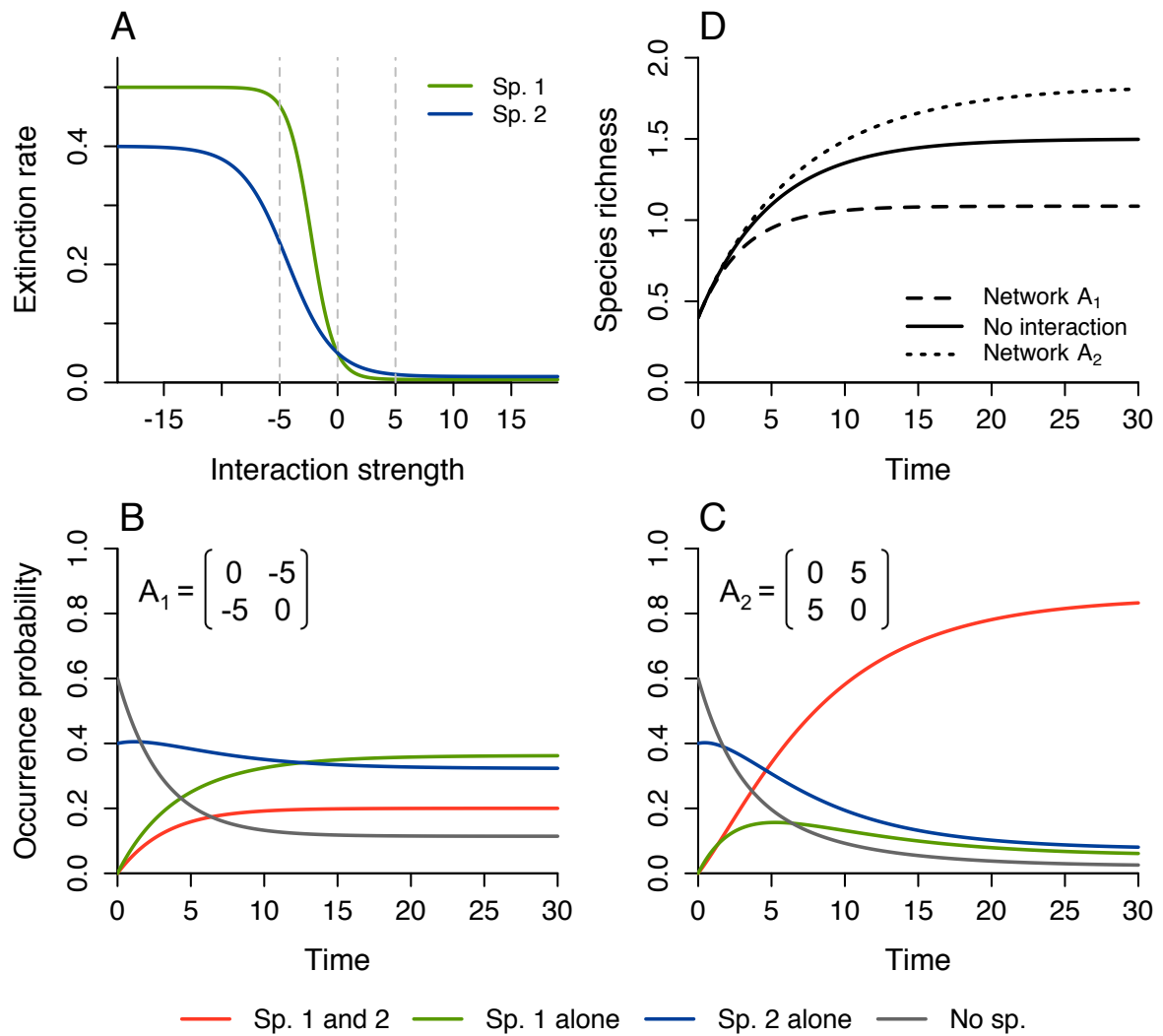


Figure 4

Figure 3

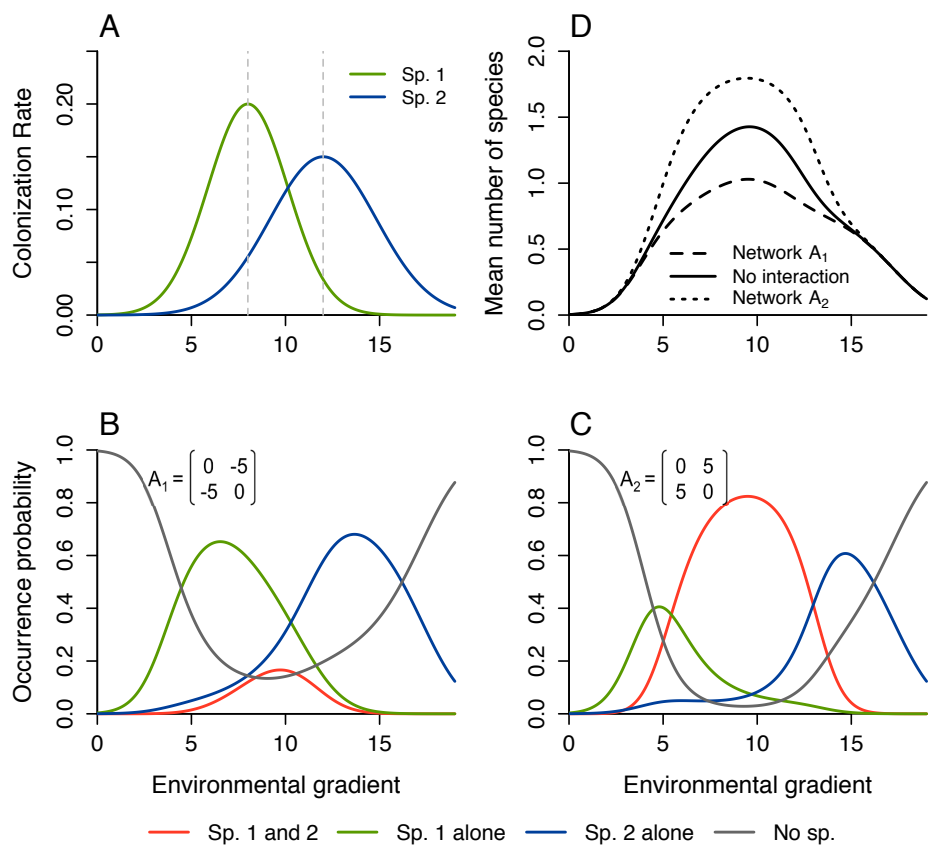


Figure 5

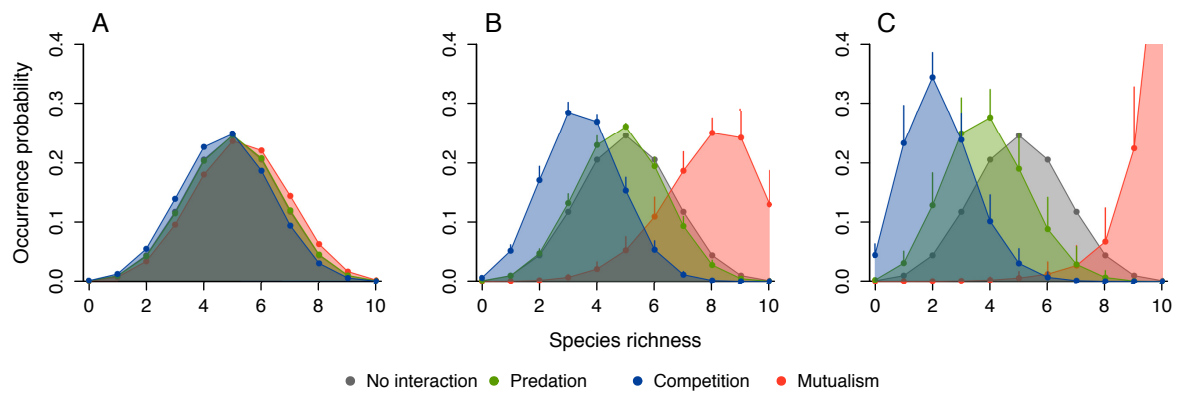
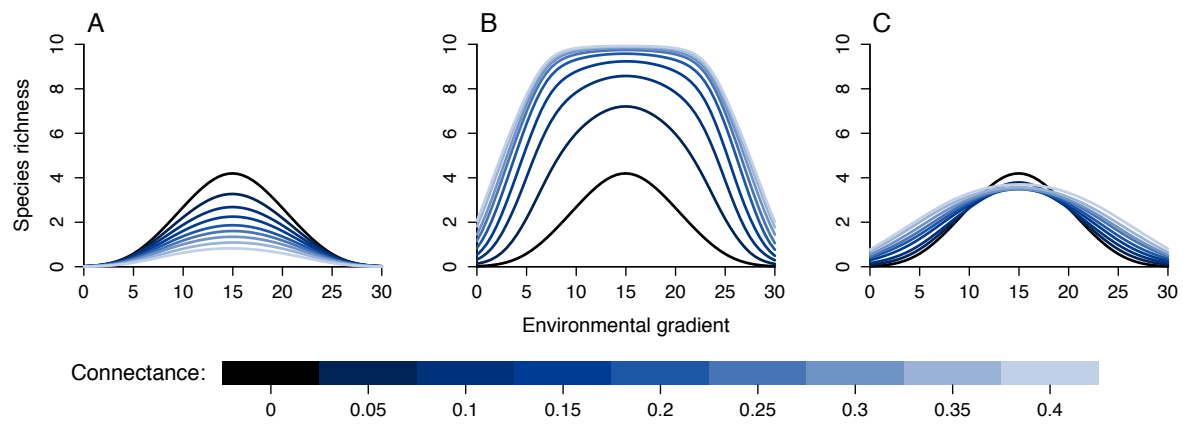
Figure 4**Figure 6**

Figure 5**Figure 7**

ARTICLE 2

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

2.1 Résumé en français du deuxième article

Dans le chapitre précédent, nous avons 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 littérature 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 interactions dans les modèles en biogéographie, nous nous sommes intéressés à leur impact sur une mesure majeure en biogéographie : la co-occurrence.

La co-occurrence 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 aller plus loin, nous avons défini une mesure de co-occurrence sous l'hypothèse d'indépendance de deux espèces. Cela signifie simplement que nous prenons l'occurrence respective des deux espèces et que nous les multiplions pour obtenir 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-occurrence des espèces en interaction :

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

occurrence. 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 simultanément dans une localité diffère de la probabilité attendue si elle se recontrait aléatoirement.

2. Les interactions indirectes modifient leur probabilité de co-occurrence. Malgré 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 conséquences des interactions directes se propagent à travers le réseau via ces relations indirectes, il est alors possible que la 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 décrite 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'intractions 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 entretient avec les espèces se rapprochera de la co-occurrence sous hypothèse d'indépendance.

Pour ce second papier, le contexte est particulier : Dominique Gravel a été invité à un numéro spécial de *Theoretical Ecology*. Dominique Gravel m'a alors proposé de travailler sur le prolongement de la réflexion menée 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 réflexion menée ayant été entamée 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 substantiellement à la rédaction du manuscrit.

2.2 TITLE

A Theory for species co-occurrence in interaction networks

2.3 AUTHORS

Kévin Cazelles, Miguel Araújo, Nicolas Mouquet et Dominique Gravel

2.4 ABSTRACT

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 KEYWORDS

Co-occurrence, 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 ? pioneered the analysis of species co-occurrence in geographical space and, together with the controversy triggered by ?, it stimulated the development of a new field of research in numerical ecology (???). 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., ?), 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 (????). 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 (????). The question of whether such interactions leave imprints in the distributions of individual species at biogeographical scales is still open to debate (*e.g.* ?), but recent empirical ?, modeling (*e.g.* ?), and theoretical (?) 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 - (?)) 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 (???). 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 (?). 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 (?). The problem with such approaches is that data to parametrize interactions mechanistically are generally lacking (?); 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 (1975). A second index is the count of checkerboards (Pielou, 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 (Pielou, 1975). 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 Diamond (1975) 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., Pielou, 1975). Thanks to the theoretical work of Pielou (1975), 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 (Pielou, 1975; see also Table 2 in Pielou, 1975 for a description of 15 indices for co-occurrence analysis). A promising avenue is the one proposed by Pielou (1975) 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 (?). For instance, plant and carnivore abundances are expected to correlate across a productivity gradient (??) because of top-down control on the herbivore population. Similarly, the propagation of indirect interactions has been studied in more complex interaction networks (?). Indirect interactions could reverse the net interaction in a surprising way, such that predator-prey abundances could be positively related (?). 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; ??). 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 definitions, 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 ?. 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: $p_{i|j,\bar{K}} = \mathbb{P}(X_i = 1 | X_j = 1, X_K = 0)$, 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 $p_{1|2,\bar{3}} = \frac{p_{1,2,\bar{3}}}{p_{2,\bar{3}}} = \frac{0.2}{0.3} = 0.67$.

Following the same logic, we define the **fundamental occurrence** as $p_{i|\bar{K}} = \mathbb{P}(X_i = 1 | X_K = 0)$. 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 $p_{1|\bar{23}} = \frac{p_{1,\bar{2},\bar{3}}}{p_{\bar{2},\bar{3}}} = \frac{0.2}{0.3} = 0.67$.

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 (?). 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 (?). 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:

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

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 $p_{1|\bar{2}} \rightarrow 0$, therefore $p_1 \rightarrow p_{1|2}p_2$
- ii if species 1 depends strongly on 2 thereby perfectly tracking its distribution 2, the $p_{1|\bar{2}} \rightarrow 0$ and $p_{1|2} \rightarrow 1$, and therefore $p_1 \rightarrow p_2$
- iii if species 2 excludes 1, then $p_{1|2} \rightarrow 0$ and $p_{2|\bar{1}} \rightarrow 0$, 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 $p_2 = \frac{p_{2|\bar{1}} - p_{2|1}p_{1|\bar{2}}}{1 - p_{2|\bar{1}}p_{1|\bar{2}}}$.
Therefore, if $p_{1|\bar{2}} \rightarrow 1$, then $p_1 \rightarrow 1$ and $p_2 \rightarrow 0$.

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 (???). 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:

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

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:

$$p_{1|\bar{3}} = \mathbb{P}(X_1 = 1|X_3 = 0) = \frac{p_{1|\bar{2}\bar{3}} + p_{2|\bar{1}\bar{3}}(p_{1|\bar{2}\bar{3}} - p_{1|\bar{2}\bar{3}})}{1 - (p_{2|\bar{1}\bar{3}} - p_{2|\bar{1}\bar{3}})(p_{1|\bar{2}\bar{3}} - p_{1|\bar{2}\bar{3}})} \quad (2.6)$$

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:

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

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

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

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:

$$\begin{cases} p_1 = \frac{p_{1|\bar{3}} + p_{3|\bar{2}}(p_{1|3} - p_{1|\bar{3}}) + (p_{3|2} - p_{3|\bar{2}})(p_{1|3}p_{2|\bar{3}} - p_{1|\bar{3}}p_{2|3})}{1 - (p_{2|3} - p_{2|\bar{3}})(p_{3|2} - p_{3|\bar{2}})} \\ p_2 = \frac{p_{2|\bar{3}} + p_{3|\bar{2}}(p_{2|3} - p_{2|\bar{3}})}{1 - (p_{2|3} - p_{2|\bar{3}})(p_{3|2} - p_{3|\bar{2}})} \\ p_3 = \frac{p_{3|\bar{2}} + p_{2|\bar{3}}(p_{3|2} - p_{3|\bar{2}})}{1 - (p_{2|3} - p_{2|\bar{3}})(p_{3|2} - p_{3|\bar{2}})} \end{cases} \quad (2.9)$$

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.

2.7.0.0.1 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 :

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

Therefore the observed co-occurrence between species i and k depends on their respective interaction with species j ($p_{j|\bar{i}\bar{k}}, p_{j|\bar{i}\bar{k}}$ and $p_{j|\bar{i}\bar{k}}$). 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:

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

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:

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

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

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

2.7.0.0.2 Associations do not have to be symmetrical. Many studies of co-occurrence assume pairwise associations to be symmetrical (but see ??). 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 (?), 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.

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 developed by ?. 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 ?. 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 ?).

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 (?). 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.

2.7.0.0.3 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 (?).

2.7.0.0.4 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 (?), 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 (?).

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 (?). We also illustrate principles that have been overlooked in most studies of co-occurrence (?); **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 ?). Previous studies already made the argument that positive and negative interactions could balance each other (?) and consequently associations should be studied on a pairwise basis (?). At least, some measure of the variability of the associations is required, and at best metrics such as network analyses (?) 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 (??). 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 $p_{i|j,\bar{K}}$). 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. ??). 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 (?). 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, feasible 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 (??). We view this integration as the next step to the derivation of a theory-driven species distribution model taking into account biotic interactions (?).

2.8 ACKNOWLEDGMENTS

This work was inspired by discussions with T. Poisot, D. Stouffer, A. Cyrtwill and A. Rozenfield. Thanks to Matt Talluto and Isabelle Boulangeat for helpful comments on a previous version of the manuscript. Financial support was provided by the Canada Research Chair program and a NSERC-Discovery grant to D. Gravel. M. Araújo acknowledges support from Imperial College's Grand Challenges in Ecosystems and Environment Initiative.

2.9 FIGURES

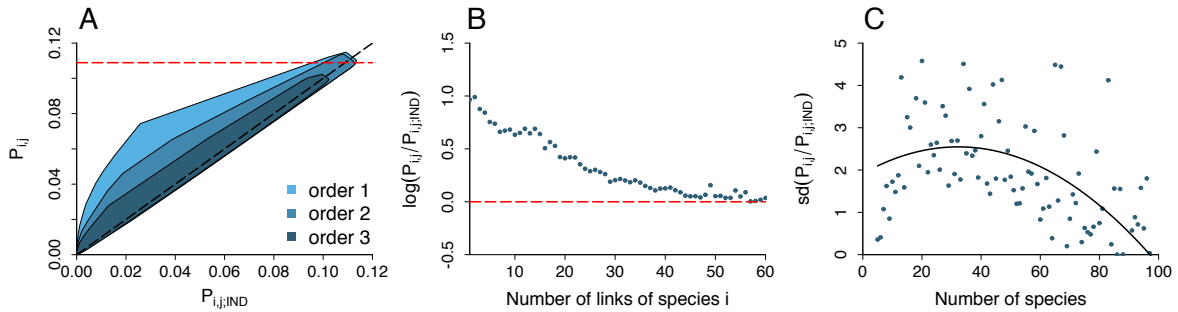


Figure 8: Co-occurrence in multi-species networks. (A) The disparity between observed co-occurrence ($P_{i,j}$) and independent co-occurrence ($P_{i,j,IND}$) decreases with the path length between nodes (species). The envelopes are drawn around the 5% and 95% quantiles of all of the data, from 100 replicated simulations for every species richness value (5 to 100 species). (B) The strength of co-occurrence ($\log(P_{i,j}/P_{i,j,IND})$) 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 reflects the variance associated. At least 3000 values were used for each point. (C) The standard deviation of the strength of association ($sd(P_{i,j}/P_{i,j,IND})$) and thus the variance decreases with species richness. Taken together, (B) and (C) imply that species distributions converge to independence with increasing species richness.

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.

2.10 Context

Biogeographers have long been fascinated by the picture of species distributions and questioned how it could have been made, i.e. searching for the processes shaping biodiversity on Earth (???, (???)). Starting from the clear relationship between abiotic variables and the physiological constraints of organisms, large-scale studies have been conducted in a pattern-driven perspective making Biogeography a realm of correlations (???). Such an approach has provided many valuable knowledge along with the development of efficient statistical tools. However, in the context of global changes, many researchers claim for strengthening the theoretical foundations of the field towards a Biogeography mechanism-driven providing reliable biodiversity forecasts (???, (???), (???)).

The importance of biotic constraints on species distribution are one of the many concerns regarding this request (???, Araújo and Rozenfeld (2014)). In order to test whether

General case	Linear chain
$\begin{pmatrix} p_{1 \overline{23}} & p_{1 2\overline{3}} & p_{1 \overline{2}3} \\ p_{2 \overline{13}} & p_{2 \overline{1}\overline{3}} & p_{2 \overline{1}3} \\ p_{3 \overline{12}} & p_{3 \overline{1}\overline{2}} & p_{3 \overline{1}2} \end{pmatrix}$	$\begin{pmatrix} p_{1 \overline{23}} & p_{1 2\overline{3}} & p_{1 \overline{2}3} \\ p_{2 \overline{13}} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$
Exploitative competition	Apparent competition
$\begin{pmatrix} p_{1 \overline{23}} & p_{1 2\overline{3}} & p_{1 \overline{2}3} \\ p_{2 \overline{13}} & 0 & 0 \\ p_{3 \overline{12}} & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} p_{1 \overline{23}} & p_{1 2\overline{3}} & p_{1 \overline{2}3} \\ p_{2 \overline{13}} & p_{2 \overline{1}\overline{3}} & p_{2 \overline{1}3} \\ p_{3 \overline{12}} & p_{3 \overline{1}\overline{2}} & 0 \end{pmatrix}$

Table 3: Direct associations between pairs of species for different modules. Entries indicate the fundamental conditional probabilities of occurrence 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 $p_{i|\overline{j}\overline{k}} = 0$, 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 : $p_{i|\overline{j}\overline{k}} = p_{i|\overline{j}k}$. For apparent competition, if species 1 and 2 are interchangeable for species 3 then : $p_{3|\overline{1}\overline{2}} = p_{3|\overline{1}2}$.

interactions influence species distributions, the simplest avenue is to investigate the species co-distribution in light of their ecological relationships. Such investigation started with Diamond's original study stating that species interacting by competition should avoid each other in space, leading to a 'checkerboard' distribution (Diamond, 1975). This idea was rapidly criticized for the lack of an adequate null hypothesis (???, (???)). Nevertheless, the resulting debate captured the attention of biogeographers as it must unravel whether co-occurrence data are more than the sum of occurrence information (???). The answer to this question as direct and major consequences: a negative one would support the use of classical species distribution models (hereafter SDMs, (???)) whereas a positive one would give credit to methods taking co-occurrence information as a proxy for ecological interactions (???) and would support the development of methods including network into species distribution models (???, (???)).

Recent theoretical developments have proposed mechanisms explaining how ecological interactions must affect the fundamental niche (see Box 1) and how they could impact occurrence data (???, Araújo and Rozenfeld (2014)). However, ranges of species are very often inferred from the realized niche which includes the impact of abiotic and biotic factors. Therefore, finding evidences of interaction signals may prove difficult which could explain the scarcity of studies reporting such effect (but see (???)). Fortunately, the co-occurrence theory in interaction networks has been formalized and suggests that the repercussion of interactions in co-occurrence data depends on the structure of the network (???). Notably, the higher the degree of a species, *i.e* the number of species which with it interacts, the harder it becomes to link a co-occurrence to an ecological relationship rather than to a random co-occurrence. Finding such relationship in empirical data would support the idea that interactions shape geographic ranges even if for many pairs of interacting species, no significant co-occurrences are found.

Here, we examined occurrence data in the light of these recent theoretical developments for five different datasets for which interactions are observed or assessed to determine whether ecological interactions impact the distribution of species. We report that the analysis of co-occurrence failed to clearly reveal a difference between pairs of interacting species and pairs of not-interacting species. However our results suggest that the degree of species influences our ability to detect significant association making co-occurrence information more than a collection of co-occurrence only for species with a limited number of link. Moreover we discover a clear relationship between the co-occurrence strength of a species and the cumulated occupancy of the entire set of species with which a species interact. Interestingly, we point out that the relation vanishes when we used classical SDMs. This results questions the capacity reliability of SDM for forecasting relevant assemblage of species and support the need for integrating ecological information into SDMs (???).

2.11 Material and Methods

2.11.1 Datasets

We analyzed five datasets spawning a large range of environmental conditions (see Fig S1 and SI Text), a large diversity of organisms and covering all fundamental type of interactions (see (???)). Four of them came with observed interactions based on which we derived metawebs and computed the connectance associated, the degree of species and the shortest-path between all pairs of species (see SI Text). For the North American Trees datasets, we derive a distance based on functional traits (see table S1 and Fig S2). For the French Breeding Birds Survey, we also derived different trait-based distance (see table S2). For all datasets, we kept only species that were present at least on 1% of the total number of sites (see SI Text).

2.11.2 Measures of co-occurrence

For each pair of species, we determined the number of observed co-occurrence $O_{i,j}$ and we calculated the expected co-occurrence values $E_{i,j}$ and its standard deviation $SD_{i,j}$ to compute a Z-score $O_{i,j} - E_{i,j}/SD_{i,j}$ (???) whose positive (negative) values indicates more (less) co-occurrence than randomly expected. Expectations were derived using three different methods. First, we assumed that all sites were equivalent, meaning we occulted the potential influence of abiotic conditions. The distribution of co-occurrence for a limited number of sites have been already studied elsewhere (???, (???)), therefore, we used an hypergeometric distribution (see SI Text for further details). For the two other expected values, we used two different classical SDMs, namely, Generalized Linear Model (hereafter GLM) and Random Forest (hereafter RF) in order to assign a probability of being presence in a given site for all species (see SI Text for more details and Fig S3 for the assessment of performances of the models). Hence, we integrate the possibility that species may often co-occur simply because they have similar abiotic requirements.

2.12 Results

For two out of four datasets for which interactions were known, we obtained a difference between interacting and not-interacting species (10 panels A to D). Therefore, when integrating all pairs of species we did not obtain a clear evidence that interacting species co-occur differently from not-interacting one. For the willows leaves network, distinguishing herbivore-willow interactions from herbivore-parasitoids revealed that the strength of co-occurrence was stronger for the former interactions than the latter ones (11 A-B). Interestingly, we noticed that the higher the mean degree of species in the dataset, the more difficult the detection of a signal of interactions in co-occurrence was (11 A-D).

For the two datasets for which we inferred a distance based on functional traits, we found that co-occurrence was higher for pairs of similar species (10 panels E and F). As similarity could be taken as a proxy for competition strength (???), this result suggests that competition is poorly detectable at large scale which is theoretically supported (Araújo and Rozenfeld, 2014). Therefore, co-occurrences of similar species are likely driven by the similarity of their abiotic requirements. The results for the FBBS dataset were identical irrespective the type of traits examined (Fig S4). This hypothesis was further supported by the decrease of the Z-score with the distance for both datasets (fig S6 A and D).

For all datasets, we report that taking environmental context into account shrinks the distribution and shift it toward 0 (10). Hence, assuming that sites are not identical for species due to the abiotic context makes the signal of co-occurrence decrease and sometimes vanished. In the pitcher dataset, we found that the signal is even reversed but the quality of the SDM approaches were low (fig S3 B).

Z-scores quickly tend to 0 when the shortest between the two species in the pair examined increases irrespective the methods employed to calculate the expected co-occurrence (11 A-D). (11 A-D). Although this was predicted by the theory (???), the decay observed is steeper. Therefore the imprint of indirect interaction in static co-occurrence data sounds un-

appreciable. From a prediction perspective, this results suggests that if species are separated by more than two links, they can be considered statistically independent. The decay was valid when we all pairs pf species were examined (see Fig S5).

When abiotic context is not taken into account, we showed that the mean Z-score of a predator (pollinator), *i.e* Z-scores averaged over all the set of its preys (host plants), decreases with the total number of site covered by its preys (12 panels A, D, G and J). The associated linear regression outperformed the one using the degree of the species that has been envisioned by the theory (Fig S7). Therefore, when a predator feeds on a set of preys that jointly cover a large part of the geographic range studied, the impact of species interactions is undetectable, but when the joint repartition of the prey is restricted, the imprint of interactions remains appreciable. Additionally, we show this relation asymmetric: the decay is less convincing when the the mean Z-scores of the preys are plotted against the cumulated range of their predators (fig S8). Hence the imprints of interactions in static occurrence are appreciated once relevant pair of specials species are student. When species are highly linked with other species and when these species have ranges that do not completely overlap, we cannot make clearly co-occurrence to interactions. This suggest that the range of the set of species should be examined rather that individual range of prey. Interestingly, we found that using the presence of the whole set of prey as predictor to assign the presence of species outperformed GLMs (see Fig S9). When abiotic constrains are taken into account, the relationship is weakened or even reversed (12) meaning the signals of co-occurrence for specialists are no longer different from the one of generalists. This illustrates that inferring species distribution from abiotic requirements cannot reflect meaningful biological properties of the ecological system studied.

2.12.1 Discussion (~ 4000 char)

to be written:

- Our results imply :
- the absence of signal at large scale often observed does not mean that interaction are unimportant rather absence of evidence for significant co-occurrence may be due to the abundance of interaction.
- co-occurrence studies must be conducted in the light of network properties. At least spatial knowledge about the system may help searching for pattern of occurrence.
- co-occurrence data have an imprint for specialist and must include it.
- Abundance of interaction occult a signal of co-occurrence.
- For specialists, the relative position of two species within an ecological network is a valuable source of information that species distribution models must integrate to better deal with the assumption that species are independent.
- Biological consistency of SDMs must be questioned. What part of interaction are actually hidden by SDM approach? JSMD approaches do a better job ?
- Using the whole set of species as one to improve predictions?
- Co-occurrence can be used as a proxy interaction ? In very special case (in microbiology it is relevant) or with other source of informations *e.g* time series get the covariation of ranges that must be a richer information.
- Mechanism-based approaches are needed.
- The ongoing mass extinction is a decline of the total number of species on Earth but also a strong drop in the number of links. Our results highlight that predictions when interactions are abundant may be easier than when they are scarcer and dramatically changed. Hence many of the current forecast may prove wrong.

2.12.2 Box 1

The fundamental niche is here described as the occurrence probability under the assumptions that (1) biotic factors are not limiting occupancy and (2) that dispersion is unconstrained. In this case, only abiotic factors (such as water availability, temperature variability and edaphic variables) limits survival and/or reproduction success, and then the occurrence probability. Consequently, predators occupancy is computed assuming that preys are abundant enough all along the environmental gradient. Similarly, the fundamental niche of any prey is not influenced neither by predators nor by competitors.

For a three species network made of one predator and its two preys, we derive the three fundamental niches f_i (9 A). Regarding the predator (species 3), we assume its prey are equivalent and that the presence of at least one prey is sufficient to release all the biological constraints:

$$f_3(w) = P(X_3 = 1 | X_2 + X_1 > 0, G = w)$$

where G denotes the environmental gradient and X_i is the random variable associated to the presence of species i . Similarly, f_1 and f_2 are obtained assuming that 3 is absent :

$$f_2(w) = P(X_2 = 1 | X_3 = 0, G = w)$$

and:

$$f_1(w) = P(X_1 = 1 | X_3 = 0, G = w)$$

Once projected on a map, the fundamental niche unravels the potential distribution of a species (???). The expected distribution can be compared to real observations and could

reveal whether dispersal limits and ecological interactions are prevalent in the occupancy dynamic of studied species. The realized niche (9 B) includes these factors.

In our simplified example, fundamental and realized niches of preys are identical. The realized niche of the predator, r_3 , is controlled by the joint realized niches of its preys:

$$r_3(w) = f_3(w) (1 - (1 - r_1(w))(1 - r_2(w)))$$

The above expression may often be more complicated due to the size and the structure of the network. For instance, we do not consider the apparent competition between 1 and 2 although it must affect the distribution of all species. Integrating the impact of many interactions may be possible using occurrence probabilities of species assemblages rather than single species (???). Integrating network information to shed light upon species distribution is also crucial to understand what kind of co-occurrence is biologically relevant. Consider as an example the co-occurrence between species 1 and 3: the co-occurrence may be strong if we restrict the analysis to the suitable conditions for species 1 but it must be weak if the entire environmental gradient is sampled. However, if we examine the co-occurrence between 3 and the assemblage made of species 1 plus 2, the co-occurrence may always be strong. Although this is meaningful in a biological point of view, co-occurrence studies often remain focus on pairs of species.

2.12.3 Tables

Table 4: Data sets analyzed in this article.

Type	No. of sites	No. of species	Interaction type	Observed	Expected	Connectance	References
Willow Leaf Network	374	156	Trophic / Parasitism	yes	no	0.042	unpublished
Pitcher Plants Network	39x20	53	Trophic	yes	no	0.44	(???)
Caribbean Humming- birds Network	32	62	Mutualist	yes	no	0.011	(???), (???), (???)
North American Trees	128891	31	Competition	yes	no	-	unpublished
French Breeding Birds Survey	2354	179	Trophic	yes	no	0.018	Gaüzère et al. (2015)

2.13 Figures

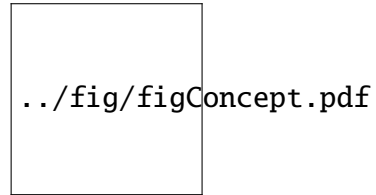
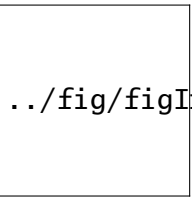


Figure 9: **Probabilistic description of fundamental and realized niches** For a three species network all the occurrence probabilities are derived along an environmental gradient assuming that A interactions are not limiting the distribution and B that species 3 needs at least of one of its preys, *i.e.* species 1 or 2. Horizontal dotted lines stand for the occurrence probabilities reached at an environmental optimum.



../fig/figIntVsNoInt.pdf

Figure 10: Co-occurrence of interacting versus not-interacting pairs of species Figures under each groups of boxplots indicate the number of pairs to which the Z-score distributions refer. The light grey rectangle corresponds to the 95% confidence interval for the standard normal distribution which gives insight into the proportion of pairs of species significantly different from 0. The comparison made in panels A to D is based on direct interactions observed. For panels E and F, similar species are defined as the species for which the trait-based distance is less than or equal to the lower decile of this distance distribution. Note that outliers are not displayed. P values were computed using the Wilcoxon rank sum test, to compare interacting versus not-interacting Z-score distribution calculated for the three different methods (black symbols) and to show whether the distribution is symmetric about 0.

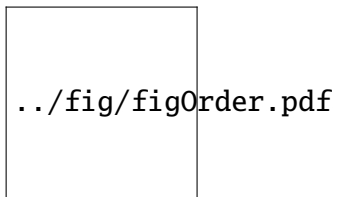


Figure 11: **Co-occurrence signal decays when the shortest path between a pair of species decay** The Z-score distribution are plotted against the shortest path for A willows-herbivores interactions, B herbivores-parasitoids interactions, C birds-plants interactions and D the pitcher plants network. First figures under each grouped boxplots indicate the shortest path associated while the figures below provide the number of pair to which the distribution refers.

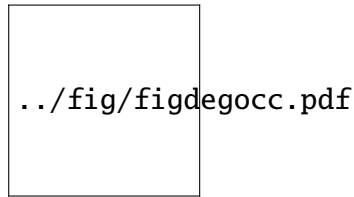


Figure 12: Co-occurrence significance decreases as the cumulated occupancy increases

For a given species, Z-scores are averaged over the all set species it interacts with and plotted against the joint distribution of the same set of species. We do so for the herbivores in the willows leafs network (panels A to C), the parasitoids in the willow leafs network (panels D to F), the hummingbirds in the Caribbean hummingbirds datasets (panels G to I) and all species in the pitcher plants network that consume other species (panels J to L). The x-axis is expressed as a log proportion of the total number of sites. Black symbols are mean Z-scores significantly different from 0 (see SI Text). In each panel, the dotted line represents the linear regression $y = ax + b$ for which the R^2 is provided. The size of circles reflects the degree of species for which the Z-score was calculated, the relation size-degree for each row is given in the middle panel. For the hummingbirds dataset (panels G to I), the triangle represent the values obtained for the former distribution of a species already analyzed (see SI text).

Reference

Araújo, M.B., Rozenfeld, A., 2014. The geographic scaling of biotic interactions. *Ecography* 37, 406–415. doi:[10.1111/j.1600-0587.2013.00643.x](https://doi.org/10.1111/j.1600-0587.2013.00643.x)

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.

Gaüzère, P., Jiguet, F., Devictor, V., 2015. Rapid adjustment of bird community compositions to local climatic variations and its functional consequences. *Global Change Biology* n/a–n/a. doi:[10.1111/gcb.12917](https://doi.org/10.1111/gcb.12917)

2.14 Material and methods

In this section, we present in more details, the datasets and the methodology we used. All analyses have been performed using R environment software (table S1 includes functions and packages we used).

2.14.1 Datasets

Sites for the five datasets are reported on five maps gathered in Fig S1. Below, we describe in more details the five datasets. The total number of species, the number of species present in at least 1% of the total number of site and the number of species for which traits information were available are reported in table S2.

2.14.1.1 Willows leafs network

2.14.1.2 Pitcher plants network

2.14.1.3 Caribbean Hummingbirds-Plant network

2.14.1.4 North American Trees datasets

2.14.1.4.1 Traits-based distance We used a distance built upon nine functional traits whose values were retrieved from (???), see **Supplementary Table 3** available at <http://onlinelibrary.wiley.com/doi/10.1111/j.1466-8238.2010.00592.x/supinfo>. Each of the nine selected variables were centered and scaled (R functions used reported in table S1) then used as is to derive Euclidean distances for all pairs of species. Then, we use agglomeration clustering with Ward's method (implemented in the *hclust()* function we used, see Table S1) to obtain the dendrogram presented in 14.

2.14.1.5 French Breeding Birds Survey datasets

2.14.1.5.1 Traits-based distance We used 73 traits that are boolean variable (see Table S4) we kept as is to derive Euclidean distances for all pairs of species.

2.14.2 Building metawebs

For four datasets, we built network based on all observed interactions and derived associated quantities, *i.e.* the connectance of the metawebs, the degrees of species and the shortest-path, using the R package “igraph” (table S1).

2.14.3 Co-occurrence measurement

For a given pair of species i and j , we examined the relationship between the observed co-occurrence $O_{i,j}$ and the expected co-occurrence $E_{i,j}$. Here, we provide more information about the three methods we used to analyse co-occurrence.

2.14.4 Hypergeometric distribution

This distribution has been mentioned in a different context (see ???) and have been fully exploited in (???) despite the author never mentioned it is a classical distribution. To clarify this, we start from the distribution written in equation (1) in Veech (2013). We consider the co-occurrence of two species on n sites. Species 1 is present in n_1 while species 2 is present in n_2 sites. The probability of having j co-occurrence, p_j is:

$$p_j = \frac{\binom{n}{j} \binom{n-j}{n_2-j} \binom{n-n_2}{n_1-j}}{\binom{n}{n_2} \binom{n}{n_1}}$$

if $\max 0, n_1 + n_2 - n \leq j \leq \min n_1, n_2$ and 0 otherwise. The expression above yields:

$$p_j = \frac{n!}{(n-j)!j!} \frac{(n-j)!}{(n-j-n_2+j)!(n_2-j)!} \frac{(n-n_2)!}{(n-n_2-n_1+j)!(n_1-j)!} \frac{(n-n_1)!n_1!}{n!} \frac{1}{\binom{n}{n_2}}$$

by rearrangement:

$$p_j = \frac{1}{j!} \frac{1}{(n_2-j)!} \frac{1}{(n-n_2-n_1+j)!(n_1-j)!} \frac{(n-n_1)!n_1!}{1} \frac{1}{\binom{n}{n_2}}$$

once sorted out, this results in:

$$p_j = \frac{\binom{n_1}{j} \binom{n-n_1}{n_2-j}}{\binom{n}{n_2}}$$

Thus, the number of co-occurrence follows a hypergeometric distribution of parameters (n, n_1, n_2) we used to calculate the expected co-occurrence $E_{i,j}$ under the hypothesis that all sites were identical for all species.

2.14.5 GLM and RF

For GLM and RF, $E_{i,j}$ correspond to probabilities of occurrence computed based on climatic data. R functions are reported in Table S1.

2.14.5.1 Climatic data

We used the global climate layers provided data WorldClim, version 1.4, available at <http://www.worldclim.org> (???). For each dataset, we performed a principal component analysis and keep as many axes as needed to explain 90% of the total inertia. We used these axes in GLM and RF.

2.14.5.2 Generalized Linear Model

For all datasets, we performed a Generalized Linear Model (???) using all the axes provided by the PCA as polynomials of degree 2. To constraints the number of parameters, we did not evaluate the interactions among axes. We also performed a selection model based on the Akaike's information criterion (AIC) C in a Stepwise Algorithm. R functions used to carry out the analyses are indexed in table SX.

2.14.5.3 Random Forests

Random Forests (???) were performed using the same formula as for GLMs. For all species, 10000 trees were computed and the probability for a species being in a given site were calculated based on the number of votes the sites were granted.

2.14.5.4 Evaluating the models

For all species, we assess the performance of the Species Distribution Models we used, *i.e.* Generalized Linear Model and Random Forest, using the Area Under the Receiver Operating Characteristic (AUROC) (???). We present the results as a cumulative sum of frequencies corresponding to the score for all species for each of the four ecological systems we studied (see Fig (???)).

2.15 Supporting Tables

2.15.0.1 R packages used

Analysis	Function	Package name	Version	Citation
Scaling and Centering	scale	base	3.3.1	R Core Team (2015)
Euclidean distance	dist	stats	3.3.1	R Core Team (2015)
Clustering	hclust	stats	3.3.1	R Core Team (2015)
PCA	dudi.pca	ade4	1.7.4	Dray and Dufour (2007)
GLM	glm	stats	3.3.1	R Core Team (2015)
GLM Selection	step	stats	3.3.1	R Core Team (2015)
Degree of species	degree	igraph	1.0.1	(???)
Shortest Paths	shortest.paths	igraph	1.0.1	(???)
AUROC	somers2	Hmisc	3.17.2	(???)

Analysis	Function	Package name	Version	Citation
TSN retrieving	get_tsn	taxize	0.7.4	(???), (???)
Wilcoxon tests	wilcox.test	stats	3.3.1	R Core Team (2015)

Supplementary Table 1: R and packages used for the analyses. GLM: Generalized Linear Model, PCA: Principal Component Analysis. AUROC: Area Under the Receiver Operating Characteristic. {#tbl:code}

Type	Total	Selected	Traits
Willow Leaf Network	274	156	-
Pitcher Plants Network	91	53	-
Caribbean Hummingbirds Network	62	62	-
North American Trees	31	31	31
French Breeding Birds Survey	340	179	321

Supplementary Table 2: For each datasets the total number of species (column *Total*), the number of species present in more that 1% of the total number of sites (column *Selected*), and the number of species for which traits information are available (column *Traits*). The symbol ‘-’ means ‘not relevant’. {#tbl:numsp}

Species	TSN	maxH	GR	WD	TolS	TolD	AM	EM	LMA	Nmass
<i>Abies balsamea</i>	18032	25	1	0.34	5.0	1.0	0	1	151.00	1.66
<i>Acer negundo</i>	28749	20	3	0.44	3.5	3.0	1	0	37.04	2.50
<i>Acer rubrum</i>	28728	25	3	0.49	3.4	1.8	1	0	71.09	1.91
<i>Acer saccharum</i>	28731	35	1	0.56	4.8	2.3	1	0	70.63	1.83
<i>Betula alleghaniensis</i>	19481	25	3	0.55	3.2	3.0	0	1	46.08	2.20
<i>Betula papyrifera</i>	19489	25	3	0.48	1.5	2.0	0	1	77.88	2.31

Species	TSN	maxH	GR	WD	TolS	TolD	AM	EM	LMA	Nmass
<i>Carpinus caroliniana</i>	19504	8	1	0.58	4.6	2.0	0	1	49.05	2.15
<i>Carya cordiformis</i>	19227	25	1	0.60	2.1	4.0	0	1	44.05	2.60
<i>Fagus grandifolia</i>	19462	25	1	0.56	4.8	1.5	0	1	61.22	2.04
<i>Fraxinus americana</i>	32931	30	2	0.55	2.5	2.4	1	0	76.75	2.12
<i>Fraxinus nigra</i>	32945	20	2	0.45	3.0	2.0	1	0	71.94	2.10
<i>Fraxinus pennsylvanica</i>	32929	25	3	0.53	3.1	3.9	1	0	87.72	1.80
<i>Larix laricina</i>	183412	25	3	0.48	1.0	2.0	0	1	120.00	1.36
<i>Ostrya virginiana</i>	19511	12	1	0.63	4.6	3.3	1	0	37.04	2.20
<i>Picea glauca</i>	183295	25	1	0.35	4.2	2.9	0	1	302.86	1.28
<i>Picea mariana</i>	183302	20	1	0.41	4.1	2.0	0	1	294.12	1.12
<i>Picea rubens</i>	18034	25	2	0.38	4.4	2.5	0	1	304.67	1.15
<i>Pinus banksiana</i>	183319	20	3	0.42	1.4	4.0	0	1	243.90	1.24
<i>Pinus resinosa</i>	183375	25	3	0.39	1.9	3.0	0	1	294.12	1.17
<i>Pinus strobus</i>	183385	30	3	0.36	3.2	2.3	0	1	121.92	1.42
<i>Populus balsamifera</i>	22453	25	3	0.37	1.3	1.8	1	1	83.46	1.95
<i>Populus grandidentata</i>	22463	20	3	0.39	1.2	2.5	1	1	70.45	2.50
<i>Populus tremuloides</i>	195773	25	3	0.37	1.2	1.8	1	1	82.02	2.16
<i>Prunus pensylvanica</i>	24799	12	3	0.36	1.0	2.0	1	1	50.00	2.40
<i>Quercus alba</i>	19290	35	1	0.60	2.9	3.6	0	1	81.21	2.39
<i>Quercus macrocarpa</i>	19287	15	1	0.58	2.7	3.9	0	1	92.74	2.27
<i>Quercus rubra</i>	19408	25	2	0.56	2.8	2.9	0	1	84.20	2.06
<i>Thuja occidentalis</i>	505490	15	1	0.30	3.5	2.7	1	0	223.00	1.02
<i>Tsuga canadensis</i>	183397	30	1	0.40	4.8	1.0	0	1	122.55	0.99
<i>Ulmus americana</i>	19049	35	3	0.46	3.1	2.9	1	0	79.47	2.07
<i>Ulmus rubra</i>	19050	25	3	0.48	3.3	3.0	1	0	59.88	2.50

Supplementary Table 3: Tree species and traits used. Abbreviations are as follows:

TSN - Taxonomic Serial Number defined by Integrated Taxonomic Information System (ITIS), maxH - Average maximum height, GR - Growth rate, WD - Wood Density, TolS - Shade tolerance, TolD - Drought tolerance, AM - Arbuscular mycorrhiza (Endomycorrhiza), EM - Ectomycorrhiza, LMA - Leaf mass per area, Nmass - Nitrogen content per leaf mass unit(???)
 available at <http://onlinelibrary.wiley.com/doi/10.1111/j.1466-8238.2010.00592.x/supinfo.{#tbl:trees}>

Category	Trait name
Activity	Nocturnal
Activity	Crepuscular
Activity	Diurnal
Diet	Seeds, nuts or grain
Diet	Fruits / frugivory
Diet	Vegetative
Diet	invert
Diet	fish
Diet	Very small mammals
Diet	Large mammals
Diet	Herptile
Diet	Small birds
Diet	Long birds
Diet	Vertebrate
Diet	Bones
Diet	Carrion
Feeding behavior	Pursuit (air and/or aquatic)
Feeding behavior	Sally
Feeding behavior	Foliage gleaning
Feeding behavior	Pouncing

Category	Trait name
Feeding behavior	Grazing
Feeding behavior	Picking, pecking or stabbing
Feeding behavior	Digging
Feeding behavior	Overturning
Feeding behavior	Probing
Feeding behavior	Filtering
Feeding habitat	Water-surface
Feeding habitat	Underwater
Feeding habitat	Water
Feeding habitat	Mud
Feeding habitat	Ground
Feeding habitat	Canopy
Feeding habitat	Shrub (low and high)
Feeding habitat	Vegetation
Feeding habitat	Air
Foraging habitat	Wet grassland, meadows, fens, sedges or tundra
Foraging habitat	Dry grassland
Foraging habitat	Rocky slope
Foraging habitat	Fast river/stream
Foraging habitat	Slow river/stream
Foraging habitat	Shore (marine)
Foraging habitat	Salt marsh
Foraging habitat	Mud or silt
Foraging habitat	Sandy gravel/beach
Foraging habitat	Reed marshes
Foraging habitat	Conifer
Foraging habitat	Mixed forest

Category	Trait name
Foraging habitat	Deciduous
Foraging habitat	Mediterranean oak or other
Foraging habitat	Open/low forest
Foraging habitat	Forest or habitat edge
Foraging habitat	Shrub/bush
Foraging habitat	Urban
Foraging habitat	Garden
Foraging habitat	High air
Nesting habitat	Wet grassland, meadows, fens, sedges or tundra
Nesting habitat	Dry grassland
Nesting habitat	Banks/sand/mud
Nesting habitat	Rock surface/outcrops
Nesting habitat	Near water/shore/island
Nesting habitat	Sand gravel/beach
Nesting habitat	Reed marshes
Nesting habitat	Conifer
Nesting habitat	Mixed forest
Nesting habitat	Deciduous
Nesting habitat	Mediterranean oak and other
Nesting habitat	Open/low forest
Nesting habitat	Shrub/bush
Nesting habitat	Urban
Nesting habitat	Garden
Nesting location	Elevated
Nesting location	Tree hole
Nesting location	Ground

Supplementary Table 4: List of the Boolean traits used to derive Euclidean distances between all pairs of species in the French Breeding Birds Survey.

2.16 Supporting Figures

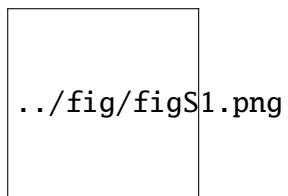


Figure 13: **Sites of the study**

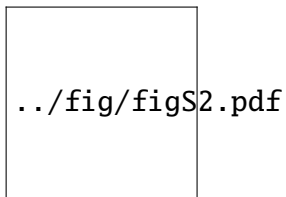


Figure 14: **Dendrogram representing the trait-based distances between the 31 species studied in the North American tree datasets.** Names of angiosperm species are written in dark grey while names of Gymnosperm species are in a lighter grey.

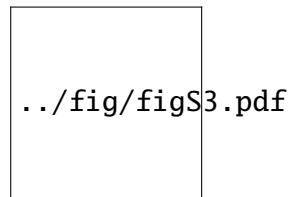


Figure 15: **Evaluation of the SDM approaches** For each dataset, the distributions of performance of generalized linear models (light grey symbols) and random Forest (dark grey symbols) for all species are presented.

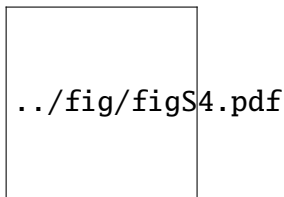
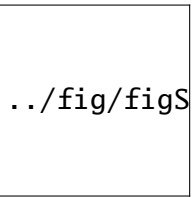


Figure 16: ** Co-occurrence and the nature of the trait-based distance in the FBBS dataset**
 The different panels correspond to four different set of trait upon which for different distance are built. Similar species are defined as the species for which the trait-based distance is less than or equal to the lower decile of this distance distribution. Note that outliers are not displayed. The light grey rectangle corresponds to the 95% confidence interval for the standard normal distribution which gives insight into the proportion of pairs of species significantly different from 0. P values were computed using the Wilcoxon rank sum test, to compare interacting versus not-interacting Z-score distribution calculated for the three different methods (black symbols) and to show whether whether Z-score were greater for hypergeometric versus GLM and whether GLM versus RF (grey symbols).



../fig/figS5.pdf

Figure 17: **Co-occurrence signal decays when the shortest path between a pair of species decay** Distribution of Z-scores for all interactions are grouped by shortest-path indicated by the first numbers below boxplots. The other figures below stand for the number of pairs of species included within the distributions.

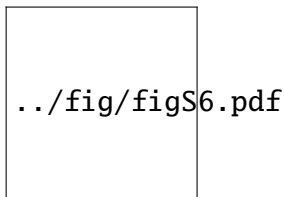
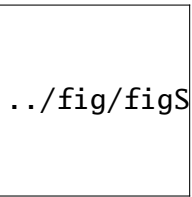


Figure 18: *Changes co-occurrence signal when increasing the distance between two species*
 Points represent the result for all pairs of interaction for two datasets: the North American Tree dataset (A=C) and the FBBS (D-I). For the latter, we used the trait-based distance computed with all available traits (D-F) and the body-size ratios (the lighter species over the heavier, panels G-I). In each panel, the equation on the bottom-left corner indicated the results of the linear regression depicted by the dotted line.



../fig/figS7.pdf

Figure 19: **The degree of species partially explains the decrease of the co-occurrence strength** For the herbivores (A-C) and the parasitoids in the willow leaves network datasets (D-F), the hummingbirds in the Caribbean hummingbirds datasets (G-I) and all species in the pitcher plants network that consume other species (J-L) the mean Z-score is plotted against the degree of the species. Black symbols are mean Z-scores significantly different from 0 (see SI Text). In each panel, the dotted line represents the linear regression $y = ax + b$ for which the R^2 is provided.

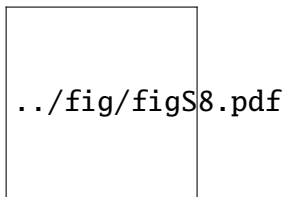
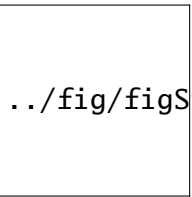


Figure 20: *Reversed figure 4** This figures correspond to the figure 4 in the main text but the Z-score are calculated for preys (host plants) rather than for predators 9pollinators). Mean Z-score are computed for willows (A-C) and herbivores (based on the herbivores-parasitoids only, D-F) of the willows leafs network, the hosts plants in the Caribbean hummingbirds datasets (G-I) and species that feed on the detritus in the pitcher plants network (panels J-L). The x-axis is expressed as a log proportion of the total number of sites included in the considered dataset. Black symbols are mean Z-scores significantly different from 0 (see SI Text). In each panel, the dotted line represents the linear regression $y = ax + b$ for which the R^2 is provided. The size of circles reflects the degree of species for which the Z-score was calculated, the relation size-degree for each row is given in the middle panel.



../fig/figS9.pdf

Figure 21: **Predicting herbivore distribution based on the distribution of willows** For the herbivores in the willow leafs network dataset, we compared the AUC obtained when using willow it does not interact with (circles) a willow in interacts with (squares) and the set of willow it interacts with (triangles) to AUC obtained for GLM (A) and RF (B). Positive values indicated that species based model outperformed the SDM model.

Supporting References

Dray, S., Dufour, A.B., 2007. The ade4 Package: Implementing the Duality Diagram for Ecologists. *Journal of Statistical Software* 22, 1–20. doi:[10.1.1.177.8850](https://doi.org/10.1.1.177.8850)

R Core Team, 2015. R: A Language and Environment for Statistical Computing.

Veech, J.A., 2013. A probabilistic model for analysing species co-occurrence. *Global Ecology and Biogeography* 22, 252–260. doi:[10.1111/j.1466-8238.2012.00789.x](https://doi.org/10.1111/j.1466-8238.2012.00789.x)

ARTICLE 3

ISLAND THEORY OF BIOGEOGRAPHY : FROM SPECIES TO FONCTIONNAL TRAITS

Le chapitre est en cours de dévelpppement. L'avancement est indiqué dans le rapport du troisième compte-rendu de comité de thèse.

CONCLUSION GÉNÉRALE

Que peut-on prédire ?

Quantifié les flux les fréquences aléatoire et contraintes les extinctions

3.1 Catégorie

Il existe en fait une catégorisation non déclarée par un de fait. La TIB bne s'intéresse pas au île auand MacArthur s'intéresse au arbres, il en fait des îlesm un cadre de font. Pourrions nous avoir des groupes neutres (coraux arbres) et d,utre pas (OK pour théories neutres), si findé sur un poinds relatifs des processus c'est super good ! Théorie sur la composition ou un trait ! mais il fait savoir ce qui chnage la compositions des traits !! On ignore le mycorrhix il y aura très probablement un résaeux de cette tforme la avec telle conséquence économique

Et si on faisait rien pour le frelo et si le comportement des abeilles européennes changeait comment prédire cela...

3.2 Quelles type de prédictions pouvons nous faire ?

Bon objet et ce qui la concerne ou pas

problème de réflexion sur l'unité pertinente

des convergences contraintes physiques / intelligence pour se soustraire à la niche.. / la chitine

3.2.1 Une question d'échelle

L'écologie porte sur l'ensemble du monde vivant quel que soit leur taille mais les différents champs ne sont pas toutes relatives à la même échelle alors il y a bien des échelles de temps, les échelles spatiales mais il y a le niveau d'organisation. Il est bien important de comprendre cela !

Un schéma avec des variables qui émergent avec différents paramètres et quelques exemples de théorie ! (DEB Evolution foodweb...) et l'action de

Repartition des espèces des passages historiques dans l'origine des espèces et dans Wallace. Le principe même de l'écologie (la définition de l'écologie). On arrive à l'idée de la niche. Exemple historique. Dans son ouvrage, le grand biogéographe Wallace reconnaît en introduction le caractère fascinant de la répartition de la biodiversité des îles avec des faits intrigants quant à la faune et la flore. Ainsi il constate qu'il peut y avoir plus de différence entre une île très éloignée et deux îles très proches. Il écrit que la faune et la flore sont plus dissimilaires entre les deux îles des Galapagos Bali et Lombok qu'entre Hokkaido (Japon) et la Grande Bretagne ou encore la Nouvelle Zélande et l'Australie,

Exemple classique de Grinnell et des Trasher + évolution avec les caractéristiques de déplacement.

Nous accumulons des évidences quand aux impacts du changement anthropique. A différentes échelles la diminution de la biodiversité, changements en composition Taranu et al. (2015) De Roos et al. (2008)

La biogéographie avec au moins 3 problèmes d'échelles => spatiale => temporelle plus on augmente plus l'empreinte historique est forte => grands événements géologiques (l'extinction des dinosaures) biogéographies historiques mais aussi forme un pool d'espèces => Mais aussi l'échelle taxonomique : la relation entre espèce est décrite à l'intérieur des taxons les relations allométriques à l'intérieur des taxons E O Wilson a commencé à rapporter des relations sur les formes les exemples du livre sont herpète faune (reptile plus amphibien) méca-

nisme => diversité de milieu

3.2.2 Des classes d'espèces ?

Wallace n'aurait-il pas eu plus de mal à comprendre les zones aujourd'hui. Si naïvement on réduit aux villes, l'homogénéité ++ mais avec les espèces invasives le signal est fortement brouillé aussi !

Je pense qu'on est à un tournant de la biologie vers un changement de paradigme communautaire centré qui ne nait pas des travaux précédents mais les suit.

La défense des modèles climatiques bioclimatiques enveloppe de Pearson comme une première approximation utilise se fait sur 3 exemples de plantes Pearson and Dawson (2003)

3.2.3 Prédire des communautés

=> des interactions changer de paradigme

On nous fait miroiter que finalement que l'érosion de la biodiversité est dramatique et le ressort actuel pour faire un levier face à cela c'est les services écosystémiques qui sont actuellement l'argument choc pour renforcer la production de la nature. Il y a un côté pervers qui est la financiarisation et la substituabilité l'argent peut alors être utilisée pour intervenir ou alors remplacer un type d'écosystème par un autre ailleurs... En fait on a l'impression que c'est plus un principe de précaution qu'on invoque et ultimement il est vraisemblable que la destruction de la nature telle que nous la connaissons soit dans le futur un générateur de conflit... et ultimement on a à craindre de faire une planète invivable pour nous-mêmes. Mais les changements sont des remplacements et pour la conservation on peut se demander la stratégie. Dans son article 'Don't judge a species on their origin' Mark Davis prend à revers un certain nombre d'idées reçues et souligne que les effets des envahisseurs peuvent être positifs Davis et al. (2011).

Vers une biogéographie intégrative

3.2.4 Les données

Comme souvent en écologie / science nous avons besoin de données, mais ce n'est pas une question vaine, L'Accumulation des données doit se faire avec une certaine normalisation pour utiliser les. Il est souvent difficile et la conséquence c'est de trouver des difficultés pour réintégrer des anciennes données Tingley and Beissinger (2009) celle des muséum Shaffer et al. (1998) Malgré les espoirs des remplacer les ordinateurs pour formuler les hypothèses, toujours besoin d'un développement théorique plus de que de corrélations essayer d'estimer aujourd'hui en utilisant le plus près possible la méthode d'hier pour savoir quel biais probable il y avait. Ici si on détecte beaucoup plus bas qu'avant avec la même méthode, alors on peut se dire que le fait que ce soit des fausses absences est faible. Par contre si on essaye d'avoir des comparaisons et que les résultats sont dus à la période de l'année... C'est plus compliqué !
Aller vers des occupancy models

3.2.5 L'abstraction des espèces

Traits fonctionnels

Les traits fonctionnels sont des propriétés mesurables sur les organismes en relation avec leurs performances et leur rôle dans l'écosystème ?. Les traits étudiés peuvent être de différentes natures, 1-morphologiques : taille de différentes parties du corps, position des yeux, taille des oeufs chez les organismes ovipares, taille des graines pour les végétaux, 2-physiologiques : taux métaboliques de bases, stœchiométrie (rapport de la concentration entre divers éléments qui composent l'organismes) ????. Un ensemble approprié de ces propriétés peut être un outil puissant pour décrire un ensemble d'espèces dans un même espace. Leur proximité dans l'espace des traits est alors un indice précieux d'une proximité fonctionnelle.

Ainsi, à l'aide de 13 traits ecomorphologiques, Albouy *et al.* 2011 parviennent à prédire les guildes trophiques de 35 espèces de poissons de la Méditerranée ?. Edwards *et al.* 2013 montrent que l'effet saisonnier sur une communauté de phytoplancton dans la Manche peut être capturé à l'aide de traits décrivant : le taux maximal de croissance, la compétitivité pour la lumière et l'azote ?. La distribution des traits fonctionnels au sein de la biodiversité est aussi une entrée de choix pour réfléchir quand à la fragilité potentielle des fonctions remplies par les écosystèmes ?. %DG : je comprends cette citation de Mouillot, mais juste une mise en garde contre ce type de référence. Mouillot se base sur l'hypothèse que les traits nous informent du fonctionnement, sans jamais documenter cette relation. Ce qui est souvent le cas, et par conséquent contribue à bâtir des mythes dans la littérature qui à l'occasion ne sont pas toujours bien appuyés. L'approche par traits est un bel exemple, on a édifié rapidement une structure conceptuelle sur les traits, mais on n'a pas solidement appuyé le concept sur de bonnes bases empiriques.

L'approche de la biodiversité par les traits fonctionnels est plus quantitative que l'approche taxonomique et permet de déduire un grand nombre de propriétés en se passant de la connaissance de leur identité. Ainsi McGill, dans son article d'opinion de 2006, propose une approche nouvelle de l'écologie des communautés qui transforme les questions centrées autour des espèces par des questions qui interrogent la répartition et la variabilité des traits ?. L'emploi des traits fonctionnels est en fait un appel à une écologie plus mécaniste, qui se penche sur la physiologie des organismes, en prend les faits les plus importants (relativement au problème traité) pour les placer dans un espace de traits commun. Cette approche est aussi en lien avec la controversée théorie métabolique en écologie ??. Dans cette théorie un certain nombre de grandeurs (comme le taux métabolique) sont reliées à la biomasse corporelles de l'adulte, fournissant ainsi en un seul trait de nombreuses relations pour des groupes d'organismes très différents. Par ces nouvelles approches, l'espérance de s'extraire de la seule identité des espèces est accrue, l'idée d'avoir des règles générales se concrétise.

Dans une théorie intégrative de la biogéographie, les traits fonctionnels peuvent être

un pivot très intéressant pour rassembler les différents concepts que nous avons développés dans les paragraphes précédents. Les traits peuvent tout d'abord être mis en relation avec le milieu abiotique. Le taux métabolique ou encore la sensibilité à la sécheresse sont des indices performant pour décrire la survie dans un milieu donné ?? que l'on peut capturer sous forme de traits. Kearney *et al.* 2010 propose une approche prometteuse dans laquelle, l'environnement physique, la disponibilité des ressources et la dynamique énergétique sont reliées par les traits fonctionnelles le tout aboutissant à un modèle de distribution très mécanistes. La structure d'un réseaux peut également être dérivée à partir de l'espace des traits. Dans leur méthode proposée cette année, Gravel *et al.* infèrent les paramètres du modèle de niche de Williams et Martinez ? à partir des relations de masse du corps entre proie et prédateurs ?. Ils sont alors en mesure de dériver un réseau global pour un ensemble d'espèce donné. Enfin, en tant qu'expression phénotypique, les traits fonctionnels sont soumis aux processus évolutifs. Sur les temps longs, l'expression de l'évolution résulte en la modification progressive des traits qui se répercute sur l'ensemble des propriétés qui en découle. Ainsi la considération d'une modification des traits est une approche simple et réaliste pour introduire les processus évolutifs et leurs conséquences ??.

L'abstraction de l'espèce Poisot et al. (2015) pour des questions centrales : - quelles esp-
ceace av interagir avce qui ?% Une chance pour voir des communautés chnager et des com-
munatés compltement affecté et en tirer des conclusion ou alors le contraire des inférences
des règles valableque dans les milieux perturbés... qui ont leur règles...

des prédictions fiables ?

Les dangers d'aller trop vite

There is also a danger that predictions grow faster than our understanding of eco-
logical systems, resulting in a gap between the scientists generating the predictions
and stakeholders using them

“Predictive ecology in a changing world” (???)

Un monde en changement : entre espoir et illusion

Une érosion de la biodiversité affolantes

L'érosion de la biodiversité exerce une certaine nostalgie qui parfois conduit une forme de fatalisme chez certains experts. Relevons la tête il va falloir trouver les solutions dans le mimétisme ?

Allant jusqu'à des problèmes de santé La tique la souris le réservoir et des hommes des problèmes de productions

Un monde biaisé ?

Sommes nous en train de biaiser le signal phylogénétique ? (cf article Thuillier)

3.2.6 Avons-nous des espoirs vains ?

Le royaume de la contingence du à l'impact historique de l'histoire évolutive. Alors comment trouver des espoirs de généralité quand le moteur repose sur de la stochasticité Mais cette loi mène à des prédictions exoécologie Les bactéries mais comment généraliser alors que l'évolution a fait émerger bon nombre d'organismes qui en soi loin quoique complètement imbriqués on a plus de micro-organismes que de cellules...

inertie historique comment imaginer des plantes sans mycorrhize mais d'autres systèmes auraient pu marcher En fait quand on pense à la plante on pense à l'immunité de la plante + mycorrhize et quand on pense à un vertébrés on inclut tout ces bactéries on ne peut certes pas

comprendre comment l'un marche sans l'autre mais pour on a pas besoin de tout connaître c'est un problème de rupture de symétrie.

Les conséquences sont compliqués des changements climatiques sont nombreuses et certaines espèce voir le range grandir d'autre diminuer pour cds espèce de co existent et donc à un changemnet prononc. de al morphologie des communautés alors que le nombre d'espèce peut être peu affecté Moritz et al. (2008)

3.2.7 DEB

Le travail de Gotelli *et al.* est également un exemple de démarche intégrative où un nombre important de processus peuvent être inclus via un système de combinaison de scénarios et tester par simulations stochastiques ?. Enfin, en construisant des réseaux basés sur la cooccurrence des espèces, Araújo *et al.* revisitent le problème de l'interdépendance des espèces ? : ils s'interrogent sur la résistance des réseaux de cooccurrence obtenus face aux futurs changement climatiques, ils mettent ainsi en évidence des risques accrus de perte des espèces les moins connectés (celles qui cooccurrent moins). Ces travaux témoignent de la volonté d'une biogéographie intégrative.

C'est impressionnant de voir comment un auteur en repartant de simple considération telle que la taille le volume peut arriver à construire une théorie à la fois simple, fondée et predictive. mettant de la cohérence dans une accumulation de fait.

=> problème SDMS quand inférence fait sur les données d'espèces la force c'est d'avoir des mesures ++ et indépendante quelque part c'est vrai mais la source d'information est très brouillé et on peut se demander se que l'on peut obtenir comme information. . . .

Nous contraignons énormément les ranges d'espèces alors nous sort de tout ça. . .

L'ajout des interactions dans un modèle incluant l'environnement abiotique interroge la relation que les deux processus entretiennent. Si les espèces n'ont pas les mêmes perfor-

mances dans différents milieux du fait de leur physiologie, pour les mêmes espèces considérées, les réseaux n'ont pas de raison d'être identiques d'un milieu à un autre. C'est sur ce fait que Poisot *et al.* 2012 ont proposé une mesure de dissimilarité des réseaux ?. Defossez *et al.* montrent que les interactions négatives entre l'hêtre commun (*Fagus Sylvatica*) et les micro-organismes du sol diminuent avec l'altitude ?. Ainsi, les contraintes biotiques sont à relier à l'environnement ?? et un modèle intégratif doit donner un cadre cohérent à ces rétroactions entre processus. Enfin, l'importance des interactions est à mettre en relation avec l'échelle considérée ?. Pour deux espèces en interaction, plus l'échelle d'étude est large, moins les effets des interactions locales sont susceptibles d'être capturés, le pouvoir explicatif de la présence d'une espèce sur l'autre peut être alors discutable ?. Comprendre quels sont les processus à prendre en compte aux différentes échelles spatio-temporelles et comprendre comment le changement d'échelle affecte nos prédictions est aussi un véritable challenge en biogéographie ?.

A large espes répartition de la biodiversité on quantifie la différence depuis les mesures classiques. Simpson, alpha gamma beta qui sont étendues au réseau Poisot et al. (2012). Mais quand on change d'échelle on arrive rarement à quelques choses de concluant pour l'intégration des interactions. Pourtant il ya des exemples convaincant comme celui de Gitelli.

Le travail le plus dur devient d'utiliser un ensemble de connaissance pur déterminer des cartes de zone à risque mais la qualité des cartes est théoriquement dépendant mais comprendre avec plus de fiabilité les prochaines zones où *Vespa* sera...

Information dans les distributions gecko australien généraliste *Heteronotia binoei* => alors peut être que ça marche bien mais sur une espèce spécialiste ??

Davis, M. a, Chew, M.K., Hobbs, R.J., Lugo, A.E., Ewel, J.J., Vermeij, G.J., Brown, J.H., Rosenzweig, M.L., Gardener, M.R., Carroll, S.P., Thompson, K., Pickett, S.T. a, Stromberg, J.C., Del Tredici, P., Suding, K.N., Ehrenfeld, J.G., Grime, J.P., Mascaro, J., Briggs,

J.C., 2011. Don't judge species on their origins. *Nature* 474, 153–4. doi :[10.1038/474153a](https://doi.org/10.1038/474153a)

De Roos, A.M., Schellekens, T., Van Kooten, T., Persson, L., 2008. Stage-specific predator species help each other to persist while competing for a single prey. *Proceedings of the National Academy of Sciences of the United States of America* 105, 13930–5. doi :[10.1073/pnas.0803834105](https://doi.org/10.1073/pnas.0803834105)

Moritz, C., Patton, J., Conroy, C., Parra, J., 2008. Impact of a century of climate change on small-mammal communities in Yosemite National Park, USA. *Science* 322, 261–4. doi :[10.1126/science.1163428](https://doi.org/10.1126/science.1163428)

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

Poisot, T., Canard, E., Mouillot, D., Mouquet, N., Gravel, D., Jordan, F., 2012. The dissimilarity of species interaction networks. *Ecology letters* 15, 1353–61. doi :[10.1111/ele.12002](https://doi.org/10.1111/ele.12002)

Poisot, T., Stouffer, D.B., Gravel, D., 2015. Beyond species : why ecological interactions vary through space and time. *Oikos* 124, 243–251. doi :[10.1101/001677](https://doi.org/10.1101/001677)

Shaffer, H., Fisher, R.N., Davidson, C., 1998. The role of natural history collections in documenting species declines. *Trends in Ecology & Evolution* 13, 27–30. doi :[10.1016/S0169-5347\(97\)01177-4](https://doi.org/10.1016/S0169-5347(97)01177-4)

Taranu, Z.E., Gregory-Eaves, I., Leavitt, P.R., Bunting, L., Buchaca, T., Catalan, J., Domaizon, I., Guilizzoni, P., Lami, A., McGowan, S., Moorhouse, H., Morabito, G., Pick, F.R., Stevenson, M.A., Thompson, P.L., Vinebrooke, R.D., 2015. Acceleration of cyanobacterial dominance in north temperate-subarctic lakes during the Anthropocene. *Ecology Letters* 18, 375–384. doi :[10.1111/ele.12420](https://doi.org/10.1111/ele.12420)

Tingley, M.W., Beissinger, S.R., 2009. Detecting range shifts from historical species occurrences : new perspectives on old data. *Trends in Ecology and Evolution* 24, 625–633. doi :[10.1016/j.tree.2009.05.009](https://doi.org/10.1016/j.tree.2009.05.009)