



**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.

[Taper ici la dédicace. Cette
page est facultative ; l'éliminer s'il
n'y a pas de dédicace]

TABLE DES MATIÈRES

TABLE DES MATIÈRES	ix
LISTE DES SYMBOLES	xiii
Section non num	3
0.1 Phase quantitative / modéliser	3
0.1.1 L'importance des îles en biogéographie	3
0.2 Data availability	4
0.3 Information dans les distributions	4
0.3.1 Proxy	4
0.3.2 Potential interactions	4
0.3.3 Enjeux essentiels de la biogéographie	5
0.3.4 Une question d'échelle	5
0.4 Cadre théorique de la thèse	5
0.4.1 Développements théoriques en Biogéographie	5
0.4.2 Chapitre 2 TIB	7
0.4.3 Interactions écologique et TIB	7
0.4.4 La TIB : un modèle simple donnant une vision puissante	7
0.4.5 Environnement abiotique et distribution des espèces	9
0.4.6 Réseaux d'interactions : interdépendance des espèces	12
0.4.7 Plasticité phénotypique et processus évolutifs	14
0.4.8 Traits fonctionnels	15
0.4.9 Inférence en biogéographie	17
0.4.10 figures envisagées	19
0.4.11 L'espace en lui-même...	19
0.4.12 Remarques	19
0.4.13 La distribution des espèces des faits et des causes	20
0.4.14 Une question d'échelle	21

0.4.15	P2	22
0.4.16	Les relations de la biogéographie	23
0.5	Conclusion	24
0.6	Predire la répartition future des espèces	24
0.6.1	Modéliser	25
0.6.2	Correlative / Mechanism	26
0.6.3	Difficulté des modélisation	29
0.6.4	Les piliers théoriques de la biogéographie actuelle	29
0.6.5	DEB	29
0.7	Mes objectifs.	32
0.8	ce qu'il y a dedans la thèse.	32
0.9	Les interactions entre espèces vers la co-occurrence.	32
0.9.1	Et ben allons y	32

ARTICLE 1

A PROPOS DES INTERACTIONS BIOTIQUES ET DES CONTRAINTES ENVIRONNEMENTALES A L'ECHELLE BIOGEOGRAPHIQUE		35
1.1	RESUMÉ	35
1.2	TITLE	37
1.3	AUTHORS	37
1.4	ABSTRACT	37
1.5	Introduction	38
1.6	The model	41
1.6.1	A simple probabilistic biogeographical model	41
1.6.2	Integrating biotic interactions	44
1.6.3	Integrating environmental gradients	46
1.7	Exploring the model	48
1.7.1	Biodiversity distribution over environmental gradients	50
1.8	Discussion	52
1.9	Acknowledgment	57

1.10 Tables	64
1.11 Figures	64
ARTICLE 2	
THEORIE DE CO-OCCURRENCE DES ESPECES DANS LES RÉSEAUX D’INTERACTION	69
2.1 Résumé en français du deuxième article	69
2.2 TITLE	72
2.3 AUTHORS	72
2.4 ABSTRACT	72
2.5 KEYWORDS	73
2.6 Introduction	73
2.7 Definitions	77
2.8 ACKNOWLEDGMENTS	90
2.9 FIGURES	91
ARTICLE 3	
HOW ECOLOGICAL INTERACTIONS CAN AFFECT BIODIVERSITY FORECAST	95
ARTICLE 4	
ISLAND THEORY OF BIOGEOGRAPHY : FROM SPECIES TO FONCTIONNAL TRAITS	97

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.

CHAPITRE 1

INTRODUCTION GÉNÉRALE

Dans l'introduction de son livre "Island Life" paru en 1980 (soit 22 ans après la parution de *On the Tendency of Varieties to Depart Indefinitely From the Original Type*), le célèbre naturaliste Alfred Russel Wallace saisit le drôle de paradoxe suivant : bien que séparé traversé une bonne partie du globe terre les écosystème du Japon et du Royaume Uni sont très similaire notamment par leur composition en arbustes et en oiseaux alors que dans le même temps des êtres très rapprochées comme les îles indonesiennes Bali et Lombok séparées de quelques dizaines de kilomètres et peuvent être très différentes. Il évoque aussi la faible predictibility du climat pour comprendre les espèces en question, il prend pour exemple différence faune africaine et brésilienne malgré la similarité du climat. Face à ces paradoxes son ouvrage se veut une tentative pour comprendre ces raisons mais il reconnaît dès l'introduction que :

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.

Dans une vision simplifiée l'écologie détermine les liens entre les caractéristiques des espèces de la mort et des naissances et l'évolution regardera les conséquences sur ces caractéristiques de ces mêmes morts et naissances. En prenant ces définitions on comprend l'intrication de ces deux qui explique la demande pour réunir les deux de même que diverses disciplines ont été réunies avec succès lors de la théorie synthétique de l'évolution (Schoener, 2011).

problème d'intrication des processus

Spécial attention sur les îles

Premier chapitre plus d'espèce mais plus de paramètre plus de problème moins de prédiction

Je vais illustrer mon propos avec 2 (3 ?) récurrent exemple (mais d'autres aussi) le cas du Frelon asiatique (anglais : Yellow-legged hornet, *Vespa velutina*). Importance pour impact dans sur les abeilles domestiques mais très peu sur la faune locale et les oiseaux migrateurs dans le nord

L'introduction aux chapitres de ma thèse sera articulée autour de la question fondamentale suivant

Dans cette introduction à ma thèse j'ai choisi de prendre comme fil conducteur la question suivante : quelles informations refferment la distribution géographique des espèces ? Pour y apporter un maximum d'élément de réponse je regarderais les mécanismes sous-jacents à l'échelle d'une échelle avant d'aborder les espoir que soulève l'analyse de la variation de ces ranges dans le temps et l'espèce sur les différentes échelles de temps avant d'aborder les apporter un maximum d'é avant de montrer que l'analyse jointe semble révéler et les défis au regard.

: question d'échelle / de variation de co-variation / difficultés d'apprécier la proportions relative des différents mécanismes / mécanismes de coexistence coexistence vs co-occurrence

variabilité quelle espoir de généralisation

Crombie repris dans Macarthur => coexistence

Problème de coexistence

=> non reproductibilité des ranges / stochasticité des ranges Frelon asiatiques

=> degat sur la nouvelle faune local msiaune augmentation ++ du nombre de liens... reconfigurations des réseaux locaux.

=> ou est le curseur dans l'histoire (evolution) ou la géographie (l'écologie)

A quel point est-il pertinent d'évaluer le range d'une espèce sur juste une île.

Un problème d'identification.

classique experience de perte de la biodiv => et hope une histoire différenteds

raseemblé ecologie et Schoener (2011)

Section non num

1.1 Phase quantitative / modéliser

1.1.1 L'importance des îles en biogéographie

Décrire l'organisation spatiale des êtres vivants et en comprendre les mécanismes sous-jacents, tels sont les objectifs ambitieux de la biogéographie ?. Cette discipline a récemment percolée au sein de la société civile via le concept de biodiversité. Le regard des citoyens se posent attentivement sur le devenir de la biodiversité dans le contexte actuel des changements globaux. La biogéographie, par son essence, peut apporter des réponses à ce questionnement ambiant ?. Cependant, pour y parvenir, des défis techniques et théoriques majeurs restent à surmonter ?.

L'effort théorique nécessaire en biogéographie porte sur l'intégration ordonnée de concepts clés issus de différents champs de l'écologie ?. Ainsi, alors que les conditions climatiques et plus généralement la géographie physique sont classiquement évoquées pour expliquer la répartition des espèces ?, les interactions entre espèces sont quant à elles souvent occultées. De même, bien que les processus évolutifs soient souvent évoqués comme déterminants majeurs de la diversité des espèces ?, leurs effets à court terme sont souvent ignorés ? dans les scénarios décrivant la biodiversité de demain ?. La difficulté principale est alors de produire des modèles (théoriques en première instance) qui intègrent l'ensemble des processus et les

relations qu'ils entretiennent ? tout en gardant une relative simplicité. Une théorie intégrative en biogéographie pourrait être le meilleur point d'ancrage pour construire de nouvelles approches appliquées. Avec une telle théorie en main, nous pourrions aller vers l'enjeu majeur de ces dernières années en biogéographie : relâcher les hypothèses que les modèles classiques de répartition des espèces d'aujourd'hui utilisent (notamment en occultant les interactions) pour prédire la biodiversité de demain ?.

Dans le projet ici présenté, nous proposons de construire des modèles théoriques plus intégratifs en repartant d'un modèle théorique classique, celui de la théorie de la biogéographie des îles proposée par MacArthur et Wilson ?. Dans un premier temps, nous y ajoutons les interactions entre espèces et une relation explicite avec l'environnement abiotique au travers d'une approche communauté centrée qui étend le modèle classique. Dans un second temps, nous combinons une approche population centrée et les processus évolutifs pour une biogéographie insulaire plus mécaniste. Enfin, au regard des enjeux que soulève le rôle des interactions entre espèces dans la construction de la biodiversité, nous réfléchissons sur l'inférence d'espèces interdépendantes.

1.2 Data availability

1.3 Information dans les distributions

1.3.1 Proxy

1.3.2 Potential interactions

Vespa aussi au Amérqieu la densit. des traffic. . .

Multi couche de distrobution dans le cas du frelon asiatique Villemant et al. (???) ont montrés que superposition du genre *Vespa* et notamment au niveau asiatique énormément

ainsi l'inférence se fait sur des données qui comporte une empreinte de condition et localement éteinte alors que possiblement comtraite qui ne seront pas en France. . .

1.3.3 Enjeux essentiels de la biogéographie

1.3.4 Une question d'échelle

Question d'échelle La biogéographie avec au moins 3 problèmes d'échelles

=> spatiale peut-on avoir d

=> temporelle plus on augmente plus l'empreinte historiques est forte => grands événements géologique (lacitation mouvement des plaques) biogéographies historiques mais aussi forme un pool d'espèces

=> Mais aussi l'échelle taxonomique : la relation aire 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) mécanisme => diversité de milieu

contre exemple des chauves souris

1.4 Cadre théorique de la thèse

1.4.1 Développements théoriques en Biogéographie

équilibre => équation 3-3 repartir de 3-3

1.4.1.1 L’empreinte historique de la La Théorie de la Biogéographie des Iles de MacArthur et Wilson

=> impact enorme sur la conservation et encore aujourd’hui bien que simplifié les calculs permettent de comprendre simplement dans quelles directions nous allons [article New York Times] Malgré la 50 ans de depuis la publication du Livre et premier articles a la suzerie de l’auteur eux meme => publications récentes qui repartent de la théorie des îles ; l’école Warren et Gravel and all

Dans la réédition de 2001 [] Wilson rappelle que le problème :

“The flaws of the book lie in its oversimplification and incompleteness, which are endemic to most efforts at theory and synthesis.”

Preface de 67 :

Now we both call ourselves Biogeographers and are unable to see any real distinction between biogeography and ecology

Diminuer la composante historique à la recherche de loi et j’ajouterais aussi simple soit elle raffiner par la suite

1.4.1.2 La théorie des métapopulations

=> chapitre de Hanski

1.4.1.3 La théorie neutre de l’écologie et le débat qu’elle soulève

Ecological equivalence des individus OK mais peut-être que l’abondance des interactions explique aussi

=> chapitre dans revisited

Problème si explication alternatives possibles alors on n'est pas obligé de mettre pour expliquer quoi que ce soit. De plus savons nous si c'est discernable ??? Si les deux relation aire espèce sont différentes d'un groupe à l'autre alors oui... Mais sinon... Non.

1.4.2 Chapitre 2 TIB

area and number $S = CA^z$ ($z \in [0.2, 0.35]$) mais des exceptions C taxon dependance similarité avec les relation allométriques sample nom isolé même relation mais z différent

Preston 1962 a lié species abundance et

1.4.3 Interactions écologique et TIB

Wilson grand entomologist spécialistes des fourmis et MacArthur mathématicien + biologiste très oiseaux sont pleinement conscience et même comportement au que peut avoir la biogéographie c'est même souvent évoquer dans la théorie mais jamais inclus la théorie des

1.4.4 La TIB : un modèle simple donnant une vision puissante

Le travail remarquable de MacArthur et Wilson ? est l'un des cadres les plus robustes de la biogéographie actuelle. Plus de 40 ans après la parution de leur livre, la Théorie de la Biogéographie des Îles (abrégée dans la suite TBI) est encore une entrée bien adaptée en biogéographie et le point de départ de nombreux travaux ????. L'idée majeure de la TBI 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- la colonisation depuis le continent et 2- les extinctions locales. La TBI est une métaphore, le cas simple d'un territoire

isolé (l'île) où les flux d'individus depuis le pool d'espèce régional (le continent) sont facilement représentables. Le modèle peut 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 ?. Plus généralement, on peut adapter un tel modèle à un territoire quelconque avec l'hypothèse que le pool régional d'espèces est indépendant des conditions locales (aucune rétroaction de la communauté locale sur le pool régional). Ainsi, ce modèle a déjà été utilisé avec succès par Gravel *et al.* 2011 pour l'élaboration de leur théorie trophique de la biogéographie des îles ?.

La force de ce modèle théorique réside dans son élégance : avec très peu de processus invoqués, la TBI donne un cadre cohérent, biologiquement fondé pour comprendre la répartition locale de la biodiversité à la lumière de la richesse spécifique régionale. Au travers d'une équation simple (1.1), la TBI mêle ainsi subtilement les processus régionaux et locaux. Ainsi, la diversité locale S , s'enrichit par colonisation, c , depuis un pool continental d'espèce P et s'appauvrit par extinctions locale e .

$$\frac{dS}{dt} = c(P - S) - eS \quad (1.1)$$

Un telle vision imbriquant deux échelles de processus est aujourd'hui bien partagée. Il est en effet reconnu que la composition d'une communauté à l'échelle locale (S) est influencée par des facteurs biotiques et abiotiques (dont les conséquences sont capturées par e), mais également par les processus régionaux tels que l'histoire évolutive des espèces (qui façonne P) et la dispersion des individus (c) ??.

La TIB tient également sa notoriété des nombreuses prédictions supportées par les faits ?. En reliant la géographie physique des îles aux processus de colonisation et d'extinction, les auteurs démontrent la puissance de leur vision. Pour cela, ils admettent que le taux de colonisation des espèces dépend de la distance entre l'île et le continent. De plus, en considérant

que la taille de l'île conditionne les ressources et donc l'extinction. Ils parviennent alors à prédire, pour un groupe d'espèces donné, une relation pertinente entre taille de l'île, distance de l'île et richesse spécifique ?. Pour une île dont la superficie et la distance au continent sont connues, au cours du temps, le nombre d'espèces sur l'île accroît, de fait le nombre de nouvelles espèces potentielles diminuent (P étant constant), la colonisation diminue donc. De même, la richesse de l'île étant accrue, le risque d'extinction est plus élevé. Les forces d'extinction et de colonisation s'annulent alors pour un nombre d'espèce précis : la richesse spécifique à l'équilibre (figure 2). L'idée que la biodiversité atteint un équilibre à relier à la taille du territoire considéré a également été massivement utilisée en biologie de la conservation. En augmentant progressivement la taille de l'île, on obtient effectivement une relation entre aire et diversité ?. Cette relation a été appliquée pour estimer la richesse spécifique de divers territoires ?, déterminer ainsi des aires de protection ?? et estimer des taux d'extinction ?.

De part son pouvoir explicatif et son élégance, le modèle de MacArthur et Wilson est un point de départ approprié pour construire des modèles plus intégratifs en intégrant explicitement des processus écologiques et évolutifs. Cette idée n'est pas nouvelle et les auteurs de la TIB ont étudié un certain nombre de processus écologiques. Notamment, ils ont intégré les phénomènes de spéciation ? et réfléchis sur l'importance des interactions quant à la répartition des espèces ?. Néanmoins, dans le modèle classique, l'ensemble de ces aspects sont absents, l'idée que les processus écologiques importent peu aux larges échelles domine. Nous allons, dans ce projet, à l'encontre de cette idée et proposons de construire des modèles intégratifs qui étendent la TIB.

1.4.5 Environnement abiotique et distribution des espèces

Les atouts actuels de la biogéographie sont 1- une quantité importante d'information relative aux présences d'espèces et au climat et 2- des modèles corrélatifs puissants qui décrivent précisément le lien entre l'espèce et son environnement abiotique. Le terme abio-

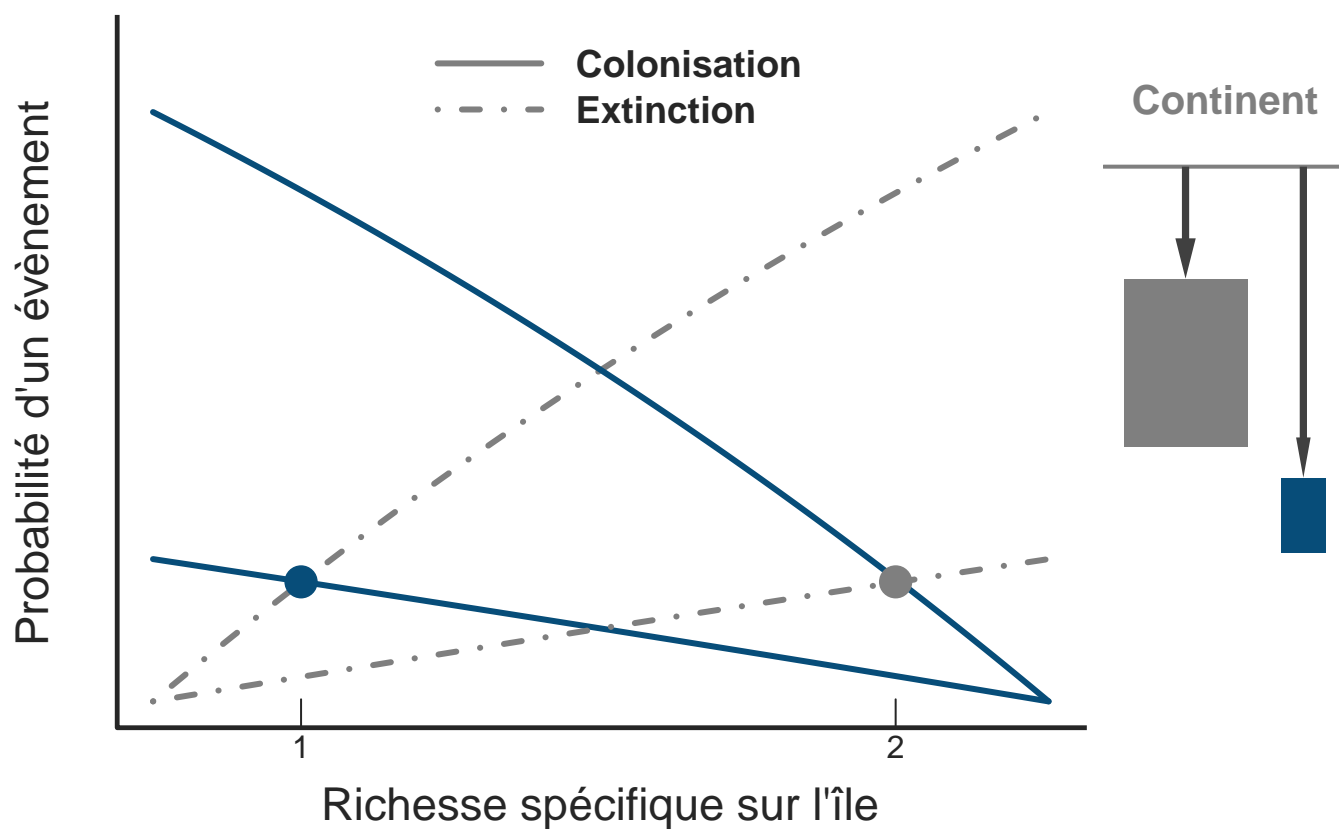


Figure 1: **La Théorie de la biogéographie des Iles.** L'évolution des taux de colonisation et d'extinction est présentée pour deux îles aux caractéristiques différentes. Les tailles relatives des îles et les distances qui les séparent du continent sont schématisées à droite du graphique, les couleurs associent les îles à leurs courbes respectives. Le pool d'espèce régional (P) est constitué de 100 espèces, les taux de colonisation et d'extinction sont exprimés en terme de probabilité d'évènement. Les points où colonisation et extinction s'équilibrent sont marqué par les symboles en gris.

tique peut prêter à confusion dans la mesure où les espèces elles-mêmes peuvent modifier des variables dites abiotiques. Par exemple, les végétaux peuvent avoir un grand impact sur les variables abiotiques locales comme la température et l'humidité du sol ?. Certains auteurs font une distinction précise en utilisant les termes de *scenopoetiques* pour les variables environnementales sur lesquels les espèces ne peuvent influencer et de *dynamiquement liées* pour les autres ?. Nous occulterons volontairement ces-dernières, l'environnement abiotique dont il est ici question n'est donc pas dynamiquement lié aux espèces.

Le premier pas pour expliquer la répartition des espèces est alors la recherche des variables environnementales les plus discriminantes pour comprendre la présence des espèces en un lieu donné ?. Au coeur de cette démarche existe un enracinement biologique profond. En effet, pour pouvoir s'installer sur un territoire donné, une espèce présente un certain nombre d'exigences physiologiques. De manière générale, l'espèce doit pouvoir répondre à l'ensemble de ses dépenses énergétiques pour survivre et éventuellement se reproduire ?. La dernière condition n'est pas indispensable : la présence d'une espèce peut résulter d'une permanente colonisation ?. Cet espace des variables environnementales dans lequel une survie d'une population est possible, nous l'appellerons niche écologique. Ce terme est l'objet de vif débat ? que nous éviterons en rappelant la définition employé. Nous parlerons ici de niche fondamentale pour désigner l'ensemble des variables *scenopoetiques* et niche réalisée lorsque la composante biotique intervient, même indirectement.

De nombreux travaux démontrent que les variables environnementales ont un grand pouvoir pour expliquer la présence des espèces ?. A partir de cette connaissance, il suffit de projeter l'espace environnemental sur l'espace géographique. Pour prédire la répartition de la biodiversité de demain, on couple des modèles d'évolution de l'environnement abiotique avec cette projection. Cette démarche rencontre actuellement un grand succès, les changements globaux induisant un effort de recherche important dans le domaine ??. Il est crucial que les modèles théoriques tel que le modèle de la TIB s'approprient le concept de niche fondamentale sous une forme simple mais cohérente. C'est en tout premier lieu par l'utilisation des

variables environnementales abiotiques que les modèles théoriques en biogéographie peuvent démontrer leur pertinence et attester de leur proximité avec les modèles plus corrélatifs et plus appliqués.

L'emploi des variables abiotiques seules pour comprendre la répartition des espèces demeurent problématique. Alors qu'il semble raisonnable de considérer des facteurs tels que la présence d'eau, de lumière et la température pour expliquer la distribution des végétaux, lorsqu'il s'agit d'espèces de niveaux trophiques plus élevés, les seules données de l'environnement abiotique ne suffisent pas ?? . Nous considérons, pour alimenter la réflexion, un exemple simple : un prédateur spécialiste et sa proie. De par l'étroite relation que les deux espèces entretiennent, il est peu efficace de regarder les seuls facteurs abiotiques pour comprendre la répartition future du prédateur. Il est alors plus pertinent d'examiner la répartition future de la proie et de s'interroger sur les possibilités de dispersion du prédateur.

1.4.6 Réseaux d'interactions : interdépendance des espèces

Il est difficile de concevoir les espèces comme indépendantes, elles partagent des espaces communs et des sources d'énergie, elles échangent de la matière, elles sont en permanentes interactions. Ces relations intra et inter spécifiques sont au coeur de l'écologie. En dynamique des populations, sont arrivés très vite des modèles classiques attestant les relations proies-prédateurs témoignant de l'importance de traiter la démographie de différentes espèces simultanément . L'écologie des réseaux pose des questions fondamentales comme celle de la stabilité des écosystèmes au regard de la structure des réseaux ?. Au delà des relations trophiques, les interactions peuvent se manifester sous de nombreuses formes ?. Le mutualisme, le commensalisme et la compétition sont des relations qui affectent la démographie des espèces sans que l'une d'entre elles se nourrisse d'une autre. La représentation en réseau des interactions est un outil puissant pour synthétiser la complexité des écosystèmes ??. Ils sont représentés par la matrice de communauté qui résume l'effet démographique des espèces par pair. Cette matrice renferme des informations précieuses telles que la connectance (me-

sure du nombre de liens constatés rapporté au nombre de liens possibles), la topologie des interactions entre espèces ? et les effets indirects ??.

Les interactions intra et inter spécifiques constituent un facteur rapidement pressenti comme responsable de la distribution spatiale des espèces ?. L'interdépendance des espèces conditionne, en effet, l'aspect favorable de l'environnement au sens large (biotique et abiotique). Ainsi Godsoe *et al.* 2012, mettent en équations le caractère favorable de l'environnement pour une espèce donnée en terme de probabilité de présence d'une autre espèce et de la nature de leur interaction ?. De même, Holt et Barfield 2009 montrent l'impact de la prédation sur la répartition d'espèces en compétition ? insistant ainsi sur le rôle majeur des interactions. Davis *et al.* 1998 ont montrés que, pour trois drosophiles en compétition, l'effet d'un parasitoïde n'est pas le même le long d'un gradient selon que les espèces sont seules ou ensemble ?. Récemment, des efforts ont été réalisés pour mettre en évidence l'importance de l'interdépendance des espèces dans les données aux larges échelles spatiales ?. On trouve actuellement dans la littérature une grande motivation pour les intégrer dans les modèles de distribution d'espèces ??. Des efforts théoriques sont encore nécessaires pour arriver à de telles approches. Néanmoins, rapprocher différents champs de l'écologie peut s'avérer d'une utilité majeure. Jabot et Bascompte ? 2012, ont d'ailleurs montré l'importance des interactions pour comprendre la distribution des espèces en rapprochant écologie des réseaux et un modèle de metacommunauté. De même Gravel *et al.* 2011 ? introduise l'interdépendance proie-prédateur dans le modèle classique de MacArthur et Wilson menant aux prémices d'une théorie trophique de la biogéographie des îles.

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 performances 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 ?. Defosse *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 ?.

1.4.7 Plasticité phénotypique et processus évolutifs

La vie telle que nous la connaissons pérennise l'information accumulée au cours du temps via à un support moléculaire, l'ADN. Cette molécule peut 1- renfermer une plasticité phénotypique offrant aux espèces des possibilités pour faire face aux stress environnementaux et 2- subir des altérations, des mutations, dont le relative avantage apporté peut assurer une survie accrue. Les espèces sont donc elles-mêmes porteuses potentielles de réponses face aux changements actuels ??. La plasticité phénotypique permet une réaction rapide des espèces à des changements environnementaux soudains. Tingley *et al.* 2009 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 ?. Les mutations sont quant à elles des événements relativement rares qui interviennent potentiellement à chaque génération, leur fréquence est donc dépendante, en premier lieu du temps de génération mais aussi de la tolérance des systèmes de réplication du matériel génétique. Pour des espèces aux temps de génération court, les processus micro-évolutifs peuvent donc être déterminants. Ainsi, Balanyá *et al.* 2009 ont montré des changements notables dans le génotype de *Drosophila subobscura* en 24 années avec des génotypes de basses latitudes plus répandus en réponses aux changements climatiques.

Il est capital de ne pas oublier les processus évolutifs dans un modèle de biogéographie afin d'envisager correctement la biodiversité de demain ?? La nature des processus à prendre en compte est dépendante de l'échelle de temps considérée. Ainsi, si l'on souhaite retracer l'histoire évolutive d'une région, les aspects adaptatifs relevant de la micro-évolution sont moins pertinents que les processus évolutifs de longue portée modifiant profondément les espèces. Il faut, à ce propos, rappeler que l'évolution peut conduire à un enrichissement du pool d'espèce d'une région donnée ?. Les mutations accumulées dans une population isolée géographiquement peuvent conduire à une incompatibilité reproductive avec les populations du pool dont elle est issue. Il y a alors spéciation, la biodiversité est augmentée. A court terme, les processus longs de spéciation peuvent être occultés mais prendre en compte les phénomènes d'adaptation et les processus d'évolution des espèces au temps de générations court est important. Il est aussi important de distinguer les réponses phénotypiques des réponses évolutives, les premières pouvant être plus rapide mais à porter moindre que les secondes plus lentes ?.

Les processus évolutifs peuvent être favorisés par les changements environnementaux mais également par les interactions entre espèces ?. Les étroites relations entre espèces peuvent favoriser ou contraindre les réponses évolutifs, qui elles-mêmes peuvent altérées ces interactions, il existe de fait des rétroactions permanentes entre évolution et écologie ?. Yoshida *et al.* 2003 montrent 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 ?. L'ensemble des trois éléments jusqu'ici évoqués (environnement abiotique, interaction, évolution) peuvent également être étroitement associé. Grant et Grant 2006 rapportent le cas de la compétition entre trois espèces de pinsons (dits de Darwin) sur l'île de Daphne (Galapagos) qui engendre une modification de la taille de leurs becs. Cette évolution liée à la compétition est elle même reliée à l'environnement abiotique car, par l'abondance ou l'absence de précipitations, il détermine la disponibilité des ressources et donc l'intensité de la compétition ?. A travers cet exemple, nous comprenons l'importance d'inclure l'ensemble des différents processus pour construire un modèle

intégratif en biogéographie. Un tel modèle serait capable, par exemple, de renseigner les risques d'exclusion compétitive dans l'exemple décrit par Grant et Grant.

1.4.8 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 compose l'organismes) ????. Un ensemble approprié de ces propriétés peut être un outil puissant pour décrire un ensemble d'espèce 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 ??.

1.4.9 Inférence en biogéographie

Rappelons les objectifs de la biogéographie : décrire et comprendre le lien entre le vivant et l'espace sur la Terre. Le coeur de l'inférence en biogéographie est donc de trouver les variables les plus pertinentes pour la répartition des espèces. Pour cela, les données spatialisées de présence ou d'abondance des organismes étudiés sont mises en relation avec des variables prédictives également spatialisées ??? . Idéalement, les échelles spatiales coïncident, sinon des transformations des données sont nécessaires. Si la variabilité capturée est satisfaisante, la combinaison retenue de variables explicatives éclairent alors les motifs de la présence des espèces en un lieu donné. Nous retiendrons le nom de modèle de distribution des espèces (MDE) pour référer à cette démarche de modélisation générale. Il y a cependant de nombreux aspects à discuter relatifs aux variables explicatives employées. Les MDE ont fourni des exemples attestant de leur pouvoir à décrire la niche fondamentale pour expliquer les présences des espèces ?. Si l'on considère des espèces mobiles, il est problématique de négliger leur mouvement, la dispersion et ses limites doivent alors être incorporés dans les modèles de distribution ?. De même, les espèces interagissant entre elles, elles influencent leurs distributions. Utiliser une espèce en tant que variable explicative pour la présence d'une autre peut s'avérer pertinent ?? mais soulève la question suivante : que faire lorsque nous essayons de prédire simultanément la présence de deux espèces dont les observations résultent elles-mêmes de leurs échanges ?

Dans le contexte actuel des changements globaux, il y a une concentration des efforts pour mieux cerner l'ensemble des réponses possibles des espèces face aux changements globaux ?. En guise de réponse, les MDE deviennent plus intégrateurs et de nouvelles approches émergent ?. Ainsi, Guisan et Rahbek 2011 proposent une démarche alliant les prédictions faites par les MDE sur un ensemble d'espèces et celles données par une approche de modélisation macroécologiques s'appuyant sur des règles de coexistence dans une unité géographique donnée ?. 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 com-

binaison 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.

Malgré leurs performances, les modèles de distribution actuels utilisés pour construire les scénarios de biodiversité de demain souffrent vraisemblablement d'un manque de théorie sous-jacent ??. La nécessité d'une approche théorique pour aller vers des approches plus appliquées est fondamentale, en témoignent, par exemple, l'histoire de la théorie de la biogéographie ? et de la théorie métabolique ?. Dans notre cas, partir d'une construction progressive assemblant les différents processus décrits ci-dessus nourrit, dans un premier temps, la réflexion sur l'ensemble des retroactions que peuvent exercer les différents processus les uns sur les autres?. Dans un second temps, le questionnement sur les échelles des phénomènes peut amener à isoler les processus que les futurs MDE ne doivent pas occulter au regard des échelles spatio-temporelles qu'ils considèrent. Troisièmement, les modèles théoriques fournissent des hypothèses à confronter aux faits, ce qui permet de conforter ou d'infirmer la théorie. Enfin, si l'agencement des processus entre eux est bien expliqué, de la théorie peut émerger de nouvelles méthodes pour traiter les données.

1.4.10 figures envisagées

- les problèmes d'échelles (figure qui cmontrent des paramètres qui captures si ou ça. . .)
- les reltions "solides" de la biogéographie

1.4.11 L'espace en lui-même...

1.4.12 Remarques

“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.” MCann Preface

L'objet de ma thèse est sur la distribution des espèces et les interactions et ce que la conjonction de tout ça. Elle est le plus souvent des articles qui sont de mon point de vue plus une réflexion des idées et pas nécessairement des démonstrations formelles et fermées mais la tentative de trouver des ouvertures d'appliquer des outils de manière un petit peu différente pour donner, ce que cherche tout le monde à donner de l'originalité à son travail. Chemin faisant j'ai passé bien du temps derrière l'ordi pour faire analyser faire des modèles mathématiques ensuite implémenté in silico. Dans cette introduction je ne peux donc pas faire l'impasse sur une mise en contexte générale de la biogéographie avec ces apports historiques ces contraintes mais aussi l'âge dans lequel nous sommes et les défis mais aussi toutes les aspects d'ordres computationnelle parler de modélisations de ces enjeux et valoriser les modèles théoriques fondamentaux qui s'éloignent parfois de la réalité mais sans jamais la déconsidérer.

Les paragraphes sont pour l'instant mis à titre indicatif avec aucune contrainte en terme de taille c'est juste pour y mettre les idées qui me viennent.

Dans la première partie de cette introduction je fais un tour très large de notion d'horizon biogéographie / pilier théorique / besoin d'hypothèse en biogéographie et finir sur la modélisation. Pour dans un deuxième temps les articuler autour de questions précises

1.4.13 La distribution des espèces des faits et des causes

Le concept récent de biodiversité.

However ecological equivalence in

“the niche is a mapping of population dynamics onto this space” (???)

1.4.13.1 De l'émerveillement

=> Partir de Wallace. L'inventaire de Wallace est impressionnant cet

1.4.13.2 Des causes / des mécanismes

vers le fonctionnement des écosystèmes levier d'action vers une approche plus utilitariste mais qui donne une certaine proximité avec les écosystèmes Loreau et al. (2001)

1.4.13.3 Challenge vers un espoir de généralisation

1.4.14 Une question d'échelle

Se problème est d'une grande importance. L'écologie porte sur l'ensemble du monde vivant quelques soient 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 historique 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 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 Hokaido (Jesse) 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, changement en composition Taranu et al. (2015) De Roos et al. (2008)

1.4.15 P2

La niche c'est quoi on en a deux définitions ultra classiques mais elles sont très problématiques. Il y a des tentatives de synthèse mais le problème est toujours là.

Partir du développement de la niche et des hypothèses clés comme l'hétérogénéité spatiale qui peut accroître la biodiversité un exemple c'est les écoulements à petites faibles échelles de l'hydrologie niche hydrologique à faibles échelles Letten et al. (2015) répartition hydrologique les hypothèses sont que qui explique selon les différents besoins des espèces (principes de la niche) que besoins différents me répartition des espèces. Cette idée est

A large échelle 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 y a des exemples convaincants comme celui de Gitelli.

Les interactions c'est quoi ce qu'on en fait.

Les interactions quelles pourraient être leurs conséquences à large échelle ?

Mais au-delà de cela il y a un besoin de règles. L'écologie cherche ces règles et essaye de faire le maximum sans trop de succès. Les traits sont un grand espoir. On a besoin d'une règle on reste descriptive il y a des relations EH-Biodiversité, SAR, Diversité-équilibre diversité fonctionnelle-

ne sont pas partiellement reliées et des théories débattues neutres théoriques de la niche Stein et al. (2014). Dans cette review Stein et al. (2014) montre que végétation est importante ce qui implique des interactions. Théorie allométrique prometteuse en ce sens qu'elle loi physiques. Différents concepts autour d'une même notion sur plusieurs paradigmes pour une même notion sur les métacommunités Leibold et al. (2004) il peuvent co-exister mais faudrait les savoir ce qui fait qu'on a l'un ou l'autre.

Exemple historique les Pinsons de Darwin (cf mon devis.) Les processus évolutifs peuvent être favorisés par les changements environnementaux mais également par les interactions entre espèces ?. Les étroites relations entre espèces peuvent favoriser ou contraindre les réponses évolutives, qui elles-mêmes peuvent altérer ces interactions, il existe de fait des rétroactions permanentes entre évolution et écologie ?. Yoshida *et al.* 2003 montrent 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 dynamique proie prédateur ?. L'ensemble des trois éléments jusqu'ici évoqués (environnement abiotique, interaction, évolution) peuvent également être étroitement associés. Grant et Grant 2006 rapportent le cas de la compétition entre trois espèces de pinsons (dits de Darwin) sur l'île de Daphne (Galapagos) qui engendre une modification de la taille de leurs becs. Cette évolution liée à la compétition est elle-même reliée à l'environnement abiotique car, par l'abondance ou l'absence de précipitations, il détermine la disponibilité des ressources et donc l'intensité de la compétition ?. A travers cet exemple, nous comprenons l'importance d'inclure l'ensemble des différents processus pour construire un modèle intégratif en biogéographie. Un tel modèle serait capable, par exemple, de renseigner les risques d'exclusion compétitive dans l'exemple décrit par Grant et Grant.

La puissance de la Biogéographie est aussi ses implications dans des cas très concrets Cirtwill and Stouffer (2015) mais aussi sa puissance exploratoire théorique Gravel et al. (2011) Cazelles et al. (2015) des îles

L'idée des interactions a déjà montré sa pertinence sur plusieurs exemples. Cirtwill and

Stouffer (2015)

1.4.16 Les relations de la biogéographie

Relation diversité des habitats Relation aire quantité d'espèces SAR / relation de la biogéographie des îles Compromis entre aire et hétérogénéité. . .

1.5 Conclusion

Generalist consumers should typically be weakly coupled to any one of their prey populations because, when feeding on many different species, they cannot be strongly coupled to any one of them Murdoch et al. (2002)

1.6 Predire la répartition future des espèces

On fait des atlas des 2 des entités avec un échantillon historique mais est-ce là important pour comprendre le réseau ? Hortal et al. (2011)

Bien sûr un objectif central est d'être en mesure de prédire la répartition des espèces L'écologie avance avec des rêves mais des pièces manquantes le lien entre les espèces et le foodweb le lien entre la présence des espèces et le fonctionnement. Il y a des problèmes fondamentaux et qui sont complexes. Par exemple les espèces peuvent avoir été invasives et causé des dégâts sur les écosystèmes, il y a les exemples de l'Australie véritable laboratoire avec les boeufs les crottes les importations mais d'un autre côté ça fonctionne encore.

Pour la prédiction des distributions d'espèces le processus est simple et bien expliqué dans le livre de Peterson. Finalement l'exercice consiste à trouver les contraintes environnementales qui sont les bonnes et ce placé ensuite un certain nombre de contraintes peuvent être les

même il y a de nombreux avantages à une telle métaphore compréhensible empiriquement fondée, intuitivement quand on va dans des endroits humides ou secs on connaît bien la différence de type d'écosystèmes mais il faut rajouter bien des couches. Ainsi la possibilité de colonisation est pour ça que des écosystèmes sont aussi éloignés mais pas les mêmes espèces dessous. Rajouter l'histoire évolutive pour bien comprendre la co-évolution des hummingbirds (cf mail de Bo)

La bonne unité d'analyse ? D'où partir ?

=> Faire un exemple tout automatiser et stocker sous Github.

Les systèmes écologiques sont des systèmes transitoires par excellence et on veut qu'il soit stable mais il y a une contingence encore assez dure à admettre les systèmes que nous observons sont ceux que nous cherchons à modéliser mais il faut pas oublier qu'ils ne resteront pas et qu'ils sont peut-être entraînés de disparaître que les moteurs de l'évolution rebat lentement les cartes

1.6.1 Modéliser

'The types of questions we pose and the types of observations we make bear witness to our preconceptions. There is no way to get rid of them. There is nothing wrong with this, but we should be aware of it. When we look around us we actually see mirrors of our ideas. We can try to change ourselves on the basis of what we see, but we cannot do without the projections we impose on reality. Observations and statements span the full range from facts via interpretation to abstract ideas. The more abstract the idea, the more important the mirror effect.' Kojima

Tentative de modéliser toutes les espèces à l'échelle de la terre entière alors qu'on est capable de générer des dynamiques chaotiques à partir d'une seule espèce. Attention je ne veux pas dire que les premières tentatives sont vaines et je ne m'appuie pas sur la dynamique

chaotique, j'indique simplement que s'il y a des cas de population isolés où la dynamique ne peut être connue à sans une précision initiale sur les conditions initiales on peut se demander comment cela peut être extrapoler. Mais la encore il y a quelques choses d'intéressant concernant l'abondance computationnelle exacte peuvent être connue à une échelle de temps courte = météorologie alors peut-être qu'à l'échelle plus large des entités plus grande = climatologie de la biodiversité !

approche modulaire => rupture de symétrie

J'explore quelques dualités propres à l'acte de modélisation que j'applique ultérieurement au champ de la biogéographie

1.6.2 Correlative / Mechanism

1.6.2.1 Règle et contingence

Finalement ce problème est aussi lié au problème d'échelle de travail ! il y a un problème d'échelle

Quand on se tourne vers les sciences de l'économie il y a un bon jeu de mots que j'ai entendu sous deux formes : - Les physiciens ont 5 règles pour expliquer 95% de l'univers et les 95 règles pour 5% - Les économistes ont prédit 12 des trois dernières crises économiques

Il y a un sentiment partagé avec l'écologie (beaucoup en commun dans la racine) administration/gestion de la maison que l'on rencontre plus dans le terme familial que / c'est la science de la maison, de l'habitat. et une complexité une légère jalousie des physiciens se serait qui ont des théories qui ont prédit des objets à une époque où pas les moyens de faire les installations Boson Onde gravitationnelle ou encore expliquer la loi de Fourier à partir de la mécanique statistiques

Rassurons nous les physiciens ont encore bien des parties sombres à explorer : matière

noire et energie noire et du boulot en masse pour ecologues / economistes peut-être que les foralimes que nous empreintons à ces disciplines ne sont pas les bons... Comme dit le phylodophe Sachs dans sa biodiversité c'est trop historiques qu'un concepte comme le fitness n'est peut-être pas bien mis en equation dans une forme physique

1.6.2.2 Un acte d'abstarction pour des défis très concret

ce n'est pas objectif, c'est se placer dans un cadre et c'est une façon de contruire le raisonemet. Exemple on peut vouloir modéliser la robabiliter d'interaction et alors uon peut commencer par une probabilité de rencontre qui est simplemnt la probabiliter de se detecter mais qui pourrait être calcluer de manière complexe ou alors juste un paramètre.

Il y a différent niveau la modelisation a pour but de donner une idée mais n'oublions as qu'il existe une progression, un raffinement et qu'ultimement, le réalisme de la simulation permet d'obetrnir précisément le phénomène données et on peut aller très loin à partor d'ima-gination d'un aller retour entre le réel et l'espace dans lequel on se place pour modéliser.

Même sur des dynamiqueues d'espèce l'échelle

Prendre quelques

Devons-nous prendre un modèle taille ou agr structuré quand on cherche la distribution des espàces à large échelle

Des courbes à fitetr qui sont parfois les m^mes et donc peut informatives...

prendre deux oiseaux il y a une relation phylo mais ecologiquement qui est le plus proche? Qui a quelle rôle? Les traist permettent cette abastraction qui ne doit pas non pplus laisser de cote la phylo, il faut bien sur regarder tout pour avoir l'image complete

Un modèle existe en lui soi on pourrait se demander Evaluation au regard de sa per-fomnce

1.6.2.3 Déterministes ou stochastiques

Il n'y a pas de meilleure manière de les traiter c'est sûr du stochastique et on peut mettre du stochastique avec une variance c'est le résultat d'un théorème mais il en est après moins que variance est quelques choses d'être utile et au bout aller des fois On peut avoir un modèle déterministe qui explique tout mais

Le hasard est quelques choses qui est dans le dé ce que Monod dans le hasard et la nécessité appelle un aléa opérationnel certain pense alors que le dé est un faux hasard que la détermination des lois de Newton et des forces initiales est totalement déterminée et on peut donc prédire que le résultat du dé. Cependant c'est l'ensemble croisé de série de contingence qui amènent au hasard et dans une partie de Yatz il sera bien difficile d'obtenir un score souhaité... Mais le hasard pur existe-t-il ? Dans l'œuvre de Monod le hasard existe et c'est l'ADN polymérase à l'origine des mutations qui en est l'expression. mais l'ensemble des séries de combinaison

Lien entre les deux Une manière de voir les choses est de donner par Clark (et repris par Dom) pour on essaye

Clark a suggéré que le stochastique n'est pas savoir et que le but c'est de passer du pas au maximum vers le déterministe. Dans leur défense de la théorie neutre de Rosindell et al rétorquent que c'est plus subtil il y a des modèles déterministes chaotiques (et dont les développements avancés font au contraire propriété statistique des attracteurs) que cela de même Gravel et collègues.. recherche que. Modèles en plus validés par des stats qui reposent sur des stochastiques. Argument d'Einstein Dieu ne joue pas au dé face à la quantique qui au final alors même que la gravité générale sans hasard pure sera fautive. Le chat de Schrödinger est peut-être vivant ici mais il est mort dans un autre univers.

Produire variance et moments d'ordre supérieurs...

Dans mon premier chapitre je fais appel à une chaîne de Markov qui est elle, un pro-

cessus aléatoires qui peut

1.6.3 Difficulté des modélisations

1.6.4 Les piliers théoriques de la biogéographie actuelle

Les principes de la biogéographie sont assez simples à comprendre,

Une des théories les plus fortes est la biogéographie des îles de MacArthur and Wilson dans les années 1960. MacArthur est mort en 1972, Wilson est célèbre pour les fourmis et une série de descendants comme Simberloff tout aussi connu.

Sont des entités avec histoire et beaucoup de publications viennent souligner le poids de l'histoire sur le pool d'espèces.

Si la compréhension est relativement simple, leur intrications est bien plus challenging.

Si les humains et leur commensaux sont reliés au point que la prise alimentaire est liée à la faune intestinale... Alors doit-on modéliser toutes les espèces de bactéries commensales pour déterminer le

1.6.5 DEB

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

1.6.5.1 Interaction et biogéographie

Accent sur les cascading effect est surtout un problème de l'instabilité (???) Il ya aussi l'article perturbant de Säterberg et al. (2013) qui montre que le fait qu'une espèce soit (ex. pêche) peut conduire à des extinctions d'autres espèces lié dans le réseau... Ces deux exemple montrent que les interactions peuvent mener à des problèmes de prédictions et donc problèmes sur prévoir les services écosystémiques et c'est appuyer par Cahill et al. (2013) qui nous indique en somme que le changement des interactions bioiques est la voie privilégiée d'extinction dans un contexte de changement climatique

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 qui est invoqué et ultimement il est vraisemblable que la destruction de la nature tel 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 même.

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).

Les rangs comme un fait (Wallace chap 2) des espèces avec des largeurs avec des grandes largeurs

Lodigèsie admirable (*Lodigesia mirabilis*) seul colibri de son genre vs Lièvre variable (*Lepus timidus*)

nombre d'espèces dans un genre varie beaucoup => un autre indice de solution pas

fructifiées. . .

Pithecia Monachus vs Pithecia pythecia séparé par une rivière

Geographical Ecology => patterns in the distribution of species

L'ouvrage de MacArthur de

2 espèces proches des ranges très séparées => species Bonobo et cChimpanzés

L'évolution = le hasard et la nécessité est un moteur de la répartition mais aussi la composante historique de la biologie. Cette dimension fascinante implique aussi nous focalisé sur des explications singulière souvent pas évident qui permettent de conformer le type de facteurs impliqué dans la variation des ranges mais nous amène pas encore à trouver des règles précises.

Wallace conclut :28 qu'une théorie générale doit tenir compte des variation range et proximité des espèces proches et des overlapp.

Both competition and predation appear now to be much more important in biogeography than people had formerly guesses

chap 2 geographical ecology

il prend comme exemple la compétition entre oiseau et un manque de ressource pour une année particulièrement sévère et que 19 ans pas assez pour voir et il conclut que

This is the main reason most evidence for competition is from biogeographers.

Distribution des fauvettes *Crateroscelis robusta* et *C.runa*

Mais le problème étant que le signal n'est visible que si on a des données sur 20ans.

Le problème

Parallèle entre information des traits sur le régime alimentaire et l'information dans les rangs est-ce cela qui conduit les écologistes à être des statisticiens. et l'info dans l'ADN

la question a été pourquoi il y a autant d'espèces mais je pense qu'une question légèrement différente n'a pas été assez investie : pourquoi peuvent-elles être si nombreuses. . . . La limite est toujours OK si assez pour 2 ou plus.

1.7 Mes objectifs.

1.8 ce qu'il y a dedans la thèse.

Alors on est loin du compte mais pourquoi. Sans vouloir détruire complètement le concept de niche il est demeuré lacunaire à bien des égards. Il y a quelque chose dans le concept de niche qu'on peut sentir..

1.9 Les interactions entre espèces vers la co-occurrence.

La distribution des plantes à fleurs et des insectes et une incroyable radiation évolutive mais aussi conséquence des dispersions.

1.9.1 Et ben allons y

Faire une figure avec les grands paramètres et discuter quelles hypothèses dessus ou pas. . .

Cahill, A.E., Aiello-Lammens, M.E., Fisher-Reid, M.C., Hua, X., Karanewsky, C.J., Ryu, H.Y., Sbeglia, G.C., Spagnolo, F., Waldron, J.B., Warsi, O., Wiens, J.J., 2013. How

does climate change cause extinction ? Proceedings. Biological sciences / The Royal Society 280, 20121890. doi : [10.1098/rspb.2012.1890](https://doi.org/10.1098/rspb.2012.1890)

Cazelles, K., Mouquet, N., Mouillot, D., Gravel, D., 2015. On the integration of biotic interaction and environmental constraints at the biogeographical scale. *Ecography* n/a–n/a. doi : [10.1111/ecog.01714](https://doi.org/10.1111/ecog.01714)

Cirtwill, A.R., Stouffer, D.B., 2015. Knowledge of predator-prey interactions improves predictions of immigration and extinction in island biogeography. *Global Ecology and Biogeography* n/a–n/a. doi : [10.1111/geb.12332](https://doi.org/10.1111/geb.12332)

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)

Gravel, D., Bell, T., Barbera, C., Bouvier, T., Pommier, T., Venail, P., Mouquet, N., 2011. Experimental niche evolution alters the strength of the diversity–productivity relationship. *Nature* 469, 89–92. doi : [10.1038/nature09592](https://doi.org/10.1038/nature09592)

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)

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)

Letten, A.D., Keith, D.a., Tozer, M.G., Hui, F.K., 2015. Fine-scale hydrological niche differentiation through the lens of multi-species co-occurrence models. *Journal of Ecology* 103, 1264–1275. doi :[10.1111/1365-2745.12428](https://doi.org/10.1111/1365-2745.12428)

Loreau, M., Naeem, S., Inchausti, P., Bengtsson, J., Grime, J.P., Hector, a, Hooper, D.U., Huston, M. a, Raffaelli, D., Schmid, B., Tilman, D., Wardle, D. a, 2001. Biodiversity and ecosystem functioning : current knowledge and future challenges. *Science* (New York, N.Y.) 294, 804–8. doi :[10.1126/science.1064088](https://doi.org/10.1126/science.1064088)

Murdoch, W.W., Kendall, B.E., Nisbet, R.M., Briggs, C.J., McCauley, E., Bolser, R., 2002. Single-species models for many-species food webs. *Nature* 417, 541–543. doi :[10.1038/417541a](https://doi.org/10.1038/417541a)

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)

Säterberg, T., Sellman, S., Ebenman, B., 2013. High frequency of functional extinctions in ecological networks. *Nature* 499, 468–70. doi :[10.1038/nature12277](https://doi.org/10.1038/nature12277)

Schoener, T.W., 2011. The newest synthesis : understanding the interplay of evolutionary and ecological dynamics. *Science* (New York, N.Y.) 331, 426–9. doi :[10.1126/science.1193954](https://doi.org/10.1126/science.1193954)

Stein, A., Gerstner, K., Kreft, H., 2014. Environmental heterogeneity as a universal driver of species richness across taxa, biomes and spatial scales. *Ecology Letters* n/a–n/a. doi :[10.1111/ele.12277](https://doi.org/10.1111/ele.12277)

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)

ARTICLE 2

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

2.1 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. Dependemment 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.

2.2 TITLE

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

2.3 AUTHORS

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

2.4 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.

2.5 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.

2.6 The model

2.6.1 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:

$$\frac{dp_i}{dt} = c_i(1 - p_i) - e_i p_i \quad (2.1)$$

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:

$$\begin{aligned} \mathbb{P}(X_{i,t+dt} = 1) &= \mathbb{P}(X_{i,t+dt} = 1|X_{i,t} = 1)\mathbb{P}(X_{i,t} = 1) \\ &\quad + \mathbb{P}(X_{i,t+dt} = 1|X_{i,t} = 0)\mathbb{P}(X_{i,t} = 0) \end{aligned} \quad (2.2)$$

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

$$\mathbb{P}(X_{i,t+dt} = 1) = c_i dt \mathbb{P}(X_{i,t} = 0) + (1 - e_i dt) \mathbb{P}(X_{i,t} = 1) \quad (2.3)$$

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:

$$\begin{aligned} \mathbb{P}(\mathbf{Y}_{t+dt} = \text{"state k"} | \mathbf{Y}_t = \text{"state l"}) &= \prod_{i_1 \in I_1} c_{i_1} dt \prod_{i_2 \in I_2} (1 - c_{i_2} dt) \\ &\quad \prod_{i_3 \in I_3} e_{i_3} dt \prod_{i_4 \in I_4} (1 - e_{i_4} dt) \end{aligned} \quad (2.4)$$

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.2, 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.2B). Finally, we can calculate

the expected number of species in a locality (Fig. 2C), 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.

2.6.2 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).

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

$$= e_{i,min} + \frac{1}{\frac{1}{e_{i,max}-e_{i,min}} + \left(\frac{1}{e_i-e_{i,min}} - \frac{1}{e_{i,max}-e_{i,min}}\right) \exp(d_i * v_i)} \quad (2.5)$$

$$g_i(\mathbf{v}) = c_i \quad (2.6)$$

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 3. 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.3B and Fig.3C). 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.3D).

2.6.3 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)):

$$\begin{aligned} \mu_{k,l}(\mathbf{w}, \mathbf{v}) = & \prod_{i_1 \in I_1} g_{i_1}(\mathbf{v}, \mathbf{w}) dt \prod_{i_2 \in I_2} (1 - g_{i_2}(\mathbf{v}, \mathbf{w}) dt) \\ & \prod_{i_3 \in I_3} f_{i_3}(\mathbf{v}, \mathbf{w}) dt \prod_{i_4 \in I_4} (1 - f_{i_4}(\mathbf{v}, \mathbf{w}) dt) \end{aligned} \quad (2.7)$$

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.

$$\lim_{\substack{l \rightarrow +\infty \\ l \in \mathbb{N}}} \mathbf{C}_0 \mathbf{M}^l = \mathbf{C}_{eq} \quad (2.8)$$

\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.

$$g_i(w_1) = g(w_1, (c_i, h_i, r_i)) = c_i * \exp\left(-\left(\frac{w_1 - h_i}{r_i}\right)^2\right) \quad (2.9)$$

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 4 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 3) 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.

2.7 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 5 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.5). 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. 5), 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. 5), 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. 5 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.

2.7.1 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 6.

For all scenarios, the richness is maximal at the center of the environmental gradient (Fig. 6). 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. 6: 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. 6). 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. 6), 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. 6). 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.

2.8 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.

2.9 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 $c_i = 10^{-5}$.

2.10 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_1dtc_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_1dtc_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

2.11 Figures

Figure 1

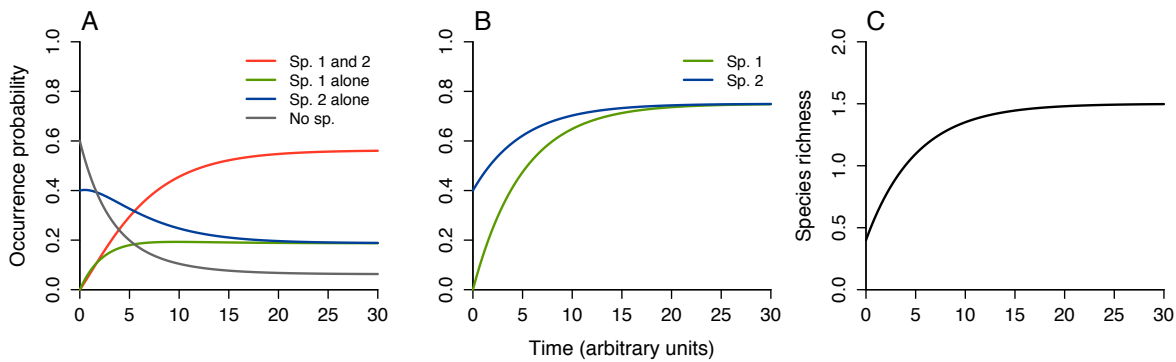


Figure 2

Figure 2

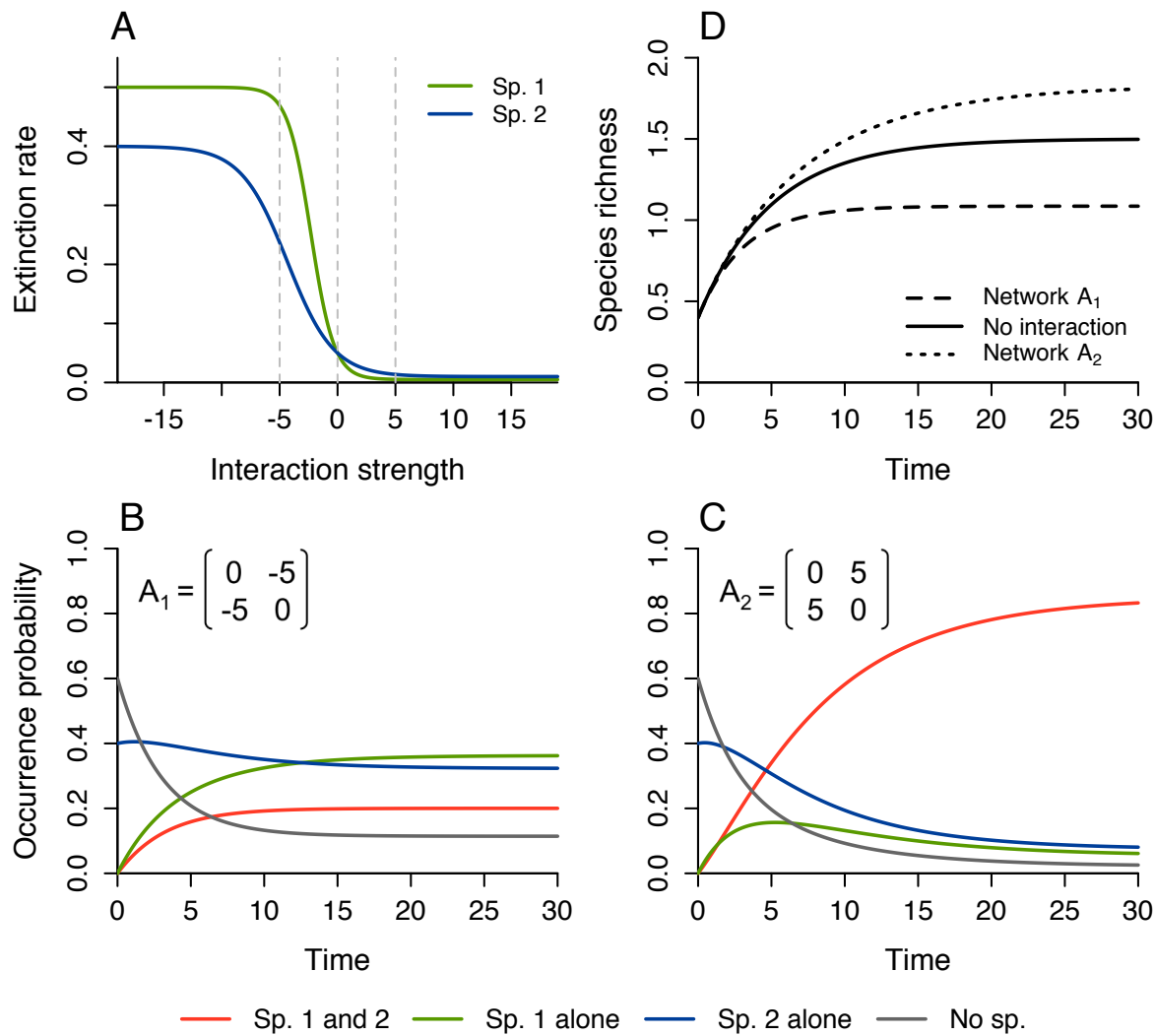


Figure 3

Figure 3

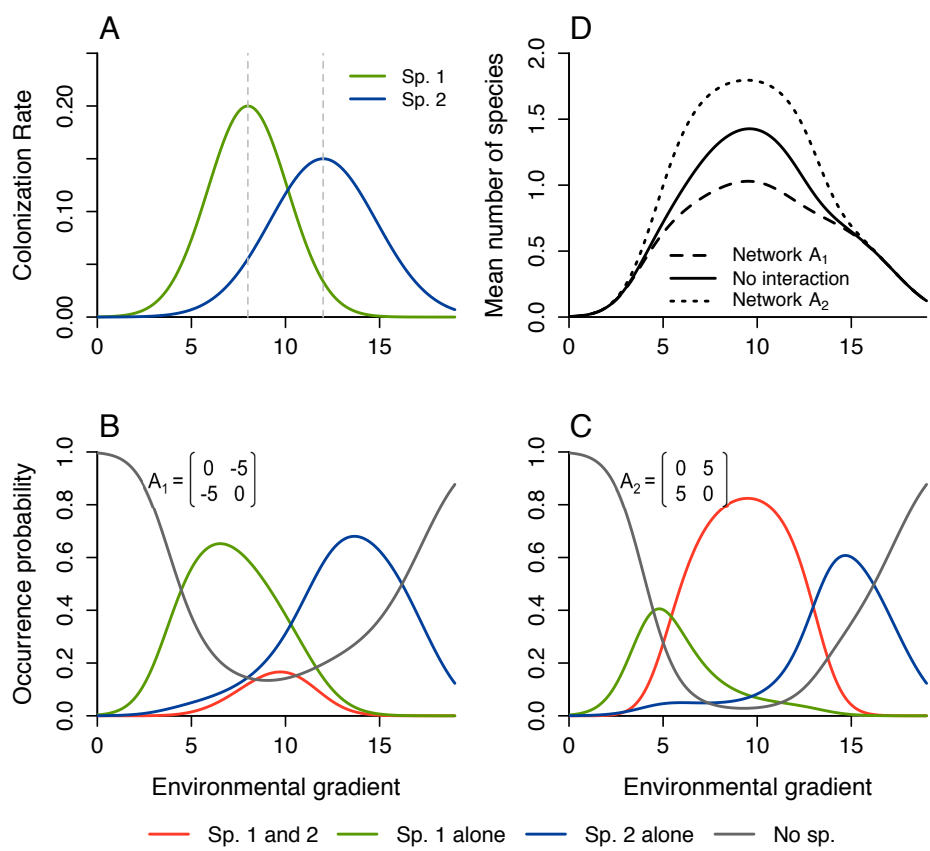


Figure 4

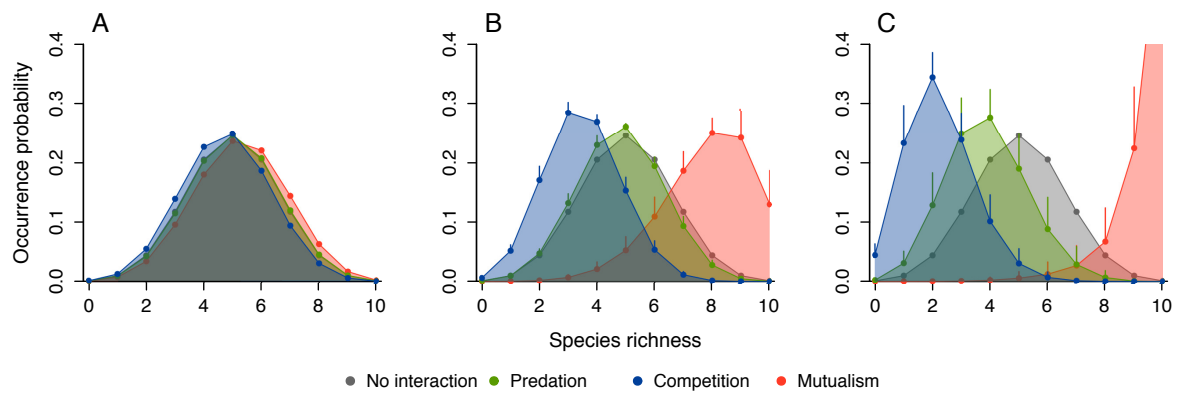
Figure 4**Figure 5**

Figure 5

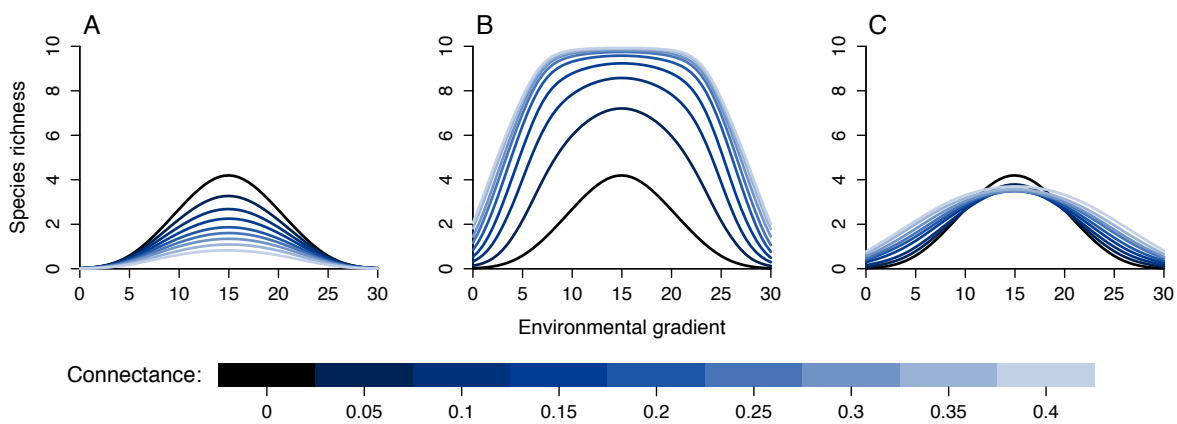


Figure 6

ARTICLE 3

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

3.1 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.

3.2 TITLE

A Theory for species co-occurrence in interaction networks

3.3 AUTHORS

Kévin Cazelles, Miguel Araújo, Nicolas Mouquet et Dominique Gravel

3.4 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.

3.5 KEYWORDS

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

3.6 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 χ^2 . A second index is the count of checkerboards χ^2 : “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 (χ^2). 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 χ^2 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., χ^2). Thanks to the theoretical work of χ^2 , 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 (χ^2 ; see also Table 2 in χ^2 for a description of 15 indices for co-occurrence analysis). A promising avenue is the one proposed by χ^2 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.

3.7 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:

$$\begin{aligned}
 p_1 &= \mathbb{P}(X_1 = 1 \cap X_2 = 1) + \mathbb{P}(X_1 = 1 \cap X_2 = 0) \\
 &= \mathbb{P}(X_1 = 1 | X_2 = 1) \mathbb{P}(X_2 = 1) \\
 &\quad + \mathbb{P}(X_1 = 1 | X_2 = 0) \mathbb{P}(X_2 = 0)
 \end{aligned} \tag{3.1}$$

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

$$\begin{cases} p_1 = p_{1|2} p_2 + p_{1|\bar{2}} (1 - p_2) \\ p_2 = p_{2|1} p_1 + p_{2|\bar{1}} (1 - p_1) \end{cases} \tag{3.2}$$

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

$$\begin{cases} p_1 = \frac{p_{1|\bar{2}} + p_{2|\bar{1}}(p_{1|2} - p_{1|\bar{2}})}{1 - (p_{2|1} - p_{2|\bar{1}})(p_{1|2} - p_{1|\bar{2}})} \\ p_2 = \frac{p_{2|\bar{1}} + p_{1|\bar{2}}(p_{2|2} - p_{2|\bar{1}})}{1 - (p_{2|1} - p_{2|\bar{1}})(p_{1|2} - p_{1|\bar{2}})} \end{cases} \tag{3.3}$$

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:

$$\begin{aligned}
 p_1 = & \mathbb{P}(X_1 = 1 \cap X_2 = 1 \cap X_3 = 1) + \mathbb{P}(X_1 = 1 \cap X_2 = 0 \cap X_3 = 1) \\
 & + \mathbb{P}(X_1 = 1 \cap X_2 = 1 \cap X_3 = 0) + \mathbb{P}(X_1 = 1 \cap X_2 = 0 \cap X_3 = 0)
 \end{aligned} \tag{3.4}$$

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

$$p_1 = \mathbb{P}(X_1 = 1|X_3 = 1)p_3 + \mathbb{P}(X_1 = 1|X_3 = 0)(1 - p_3) \quad (3.5)$$

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:

$$p_{1|\bar{3}} = \mathbb{P}(X_1 = 1|X_3 = 0) = \frac{p_{1|\bar{2}\bar{3}} + p_{2|\bar{1}\bar{3}}(p_{1|\bar{2}\bar{3}} - p_{1|\bar{2}\bar{3}})}{1 - (p_{2|\bar{1}\bar{3}} - p_{2|\bar{1}\bar{3}})(p_{1|\bar{2}\bar{3}} - p_{1|\bar{2}\bar{3}})} \quad (3.6)$$

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:

$$p_{1|3} = \mathbb{P}(X_1 = 1|X_3 = 1) = \frac{p_{1|\bar{2}3} + p_{2|\bar{1}3}(p_{1|\bar{2}3} - p_{1|\bar{2}3})}{1 - (p_{2|\bar{1}3} - p_{2|\bar{1}3})(p_{1|\bar{2}3} - p_{1|\bar{2}3})} \quad (3.7)$$

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

$$\begin{cases} p_1 = p_{1|3}p_3 + p_{1|\bar{3}}(1 - p_3) \\ p_2 = p_{2|3}p_3 + p_{2|\bar{3}}(1 - p_3) \\ p_3 = p_{3|2}p_2 + p_{3|\bar{2}}(1 - p_2) \end{cases} \quad (3.8)$$

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:

$$\begin{cases} p_1 = \frac{p_{1|\bar{3}} + p_{3|\bar{2}}(p_{1|3} - p_{1|\bar{3}}) + (p_{3|2} - p_{3|\bar{2}})(p_{1|3}p_{2|\bar{3}} - p_{1|\bar{3}}p_{2|3})}{1 - (p_{2|3} - p_{2|\bar{3}})(p_{3|2} - p_{3|\bar{2}})} \\ p_2 = \frac{p_{2|\bar{3}} + p_{3|\bar{2}}(p_{2|3} - p_{2|\bar{3}})}{1 - (p_{2|3} - p_{2|\bar{3}})(p_{3|2} - p_{3|\bar{2}})} \\ p_3 = \frac{p_{3|\bar{2}} + p_{2|\bar{3}}(p_{3|2} - p_{3|\bar{2}})}{1 - (p_{2|3} - p_{2|\bar{3}})(p_{3|2} - p_{3|\bar{2}})} \end{cases} \quad (3.9)$$

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.

3.7.0.0.1 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 :

$$\begin{aligned}
p_{i,k} &= p_i - p_{i,\bar{k}}(1 - p_k) \\
p_{i,k} &= p_i - \frac{p_{i|\bar{j}\bar{k}} + p_{j|\bar{i}\bar{k}}(p_{i|\bar{j}\bar{k}} - p_{i|\bar{j}\bar{k}})}{1 - (p_{j|\bar{i}\bar{k}} - p_{j|\bar{i}\bar{k}})(p_{i|\bar{j}\bar{k}} - p_{i|\bar{j}\bar{k}})}(1 - p_k)
\end{aligned} \tag{3.10}$$

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:

$$p_{1,3} = p_1 - \frac{p_{1|\bar{2}\bar{3}}}{1 - p_{2|\bar{1}\bar{3}}(p_{1|\bar{2}\bar{3}} - p_{1|\bar{2}\bar{3}})}(1 - p_3) \tag{3.11}$$

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:

$$p_{2,3} = p_2 - \frac{p_{1|\bar{2}\bar{3}}p_{2|\bar{1}\bar{3}}}{1 - (p_{1|\bar{2}\bar{3}} - p_{1|\bar{2}\bar{3}})p_{2|\bar{1}\bar{3}}}(1 - p_3) \tag{3.12}$$

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

$$p_{1,2} = p_1 - \frac{p_{1|\bar{2}\bar{3}}}{1 - p_{3|\bar{1}\bar{2}}(p_{1|\bar{2}\bar{3}} - p_{1|\bar{2}\bar{3}})}(1 - p_2) \tag{3.13}$$

3.7.0.0.2 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.

3.7.0.0.3 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 (?).

3.7.0.0.4 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 (?).

3.8 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.

3.9 FIGURES

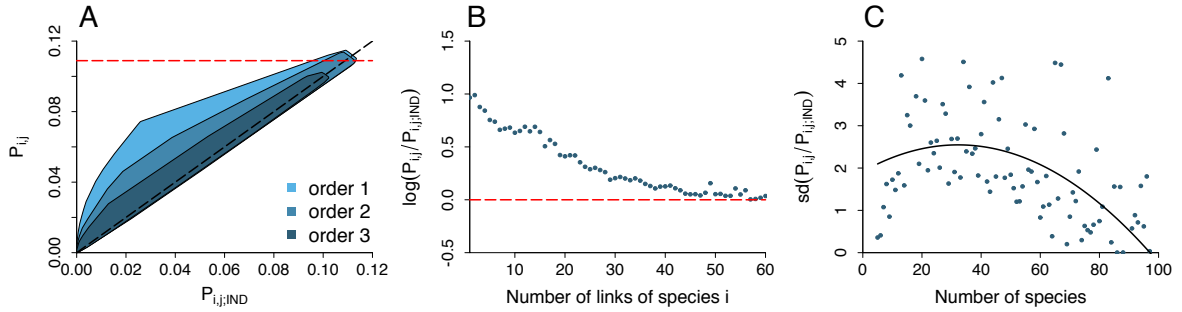


Figure 7: Co-occurrence in multi-species networks. (A) The disparity between observed co-occurrence ($P_{i,j}$) and independent co-occurrence ($P_{i,j,IND}$) decreases with the path length between nodes (species). The envelopes are drawn around the 5% and 95% quantiles of all of the data, from 100 replicated 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.

3.10 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
$\begin{pmatrix} p_{1 \bar{2}\bar{3}} & p_{1 2\bar{3}} & p_{1 \bar{2}3} \\ p_{2 \bar{1}\bar{3}} & p_{2 1\bar{3}} & p_{2 \bar{1}3} \\ p_{3 \bar{1}\bar{2}} & p_{3 1\bar{2}} & p_{3 \bar{1}2} \end{pmatrix}$	$\begin{pmatrix} p_{1 \bar{2}\bar{3}} & p_{1 2\bar{3}} & p_{1 \bar{2}3} \\ p_{2 \bar{1}\bar{3}} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$
Exploitative competition	Apparent competition
$\begin{pmatrix} p_{1 \bar{2}\bar{3}} & p_{1 2\bar{3}} & p_{1 \bar{2}3} \\ p_{2 \bar{1}\bar{3}} & 0 & 0 \\ p_{3 \bar{1}\bar{2}} & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} p_{1 \bar{2}\bar{3}} & p_{1 2\bar{3}} & p_{1 \bar{2}3} \\ p_{2 \bar{1}\bar{3}} & p_{2 1\bar{3}} & p_{2 \bar{1}3} \\ p_{3 \bar{1}\bar{2}} & p_{3 1\bar{2}} & 0 \end{pmatrix}$

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 $p_{i|j\bar{k}} = 0$, it means that species i cannot be found without k . When two species i and j do not interact directly, if the absence of species k do not impact species i survival then : $p_{i|j\bar{k}} = p_{i|\bar{j}\bar{k}}$. For apparent competition, if species 1 and 2 are interchangeable for species 3 then : $p_{3|1\bar{2}} = p_{3|\bar{1}2}$.

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 (???).

3.11 Material and Methods

3.11.1 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).

3.11.2 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.

3.12 Results

For two out of four datasets for which interactions were known, we obtained a difference between interacting and not-interacting species (?? 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 leafs 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 (?? 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 (?? 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 (?? 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 (??). 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 (?? A-D). (?? 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 cumulated range of its preys (?? 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 (??) 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.

3.12.1 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.

3.12.2 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 (?? 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 (?? 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.

3.12.3 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)

3.13 Figures

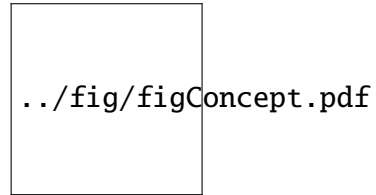
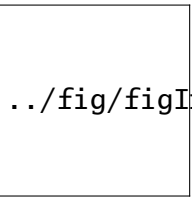


Figure 8: **Probabilistic description of fundamental and realized niches** For a three species network all the occurrence probabilities are derived along an environmental gradient assuming that A interactions are not limiting the distribution and B that species 3 needs at least of one of its preys, *i.e.* species 1 or 2. Horizontal dotted lines stand for the occurrence probabilities reached at an environmental optimum.



../fig/figIntVsNoInt.pdf

Figure 9: 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.

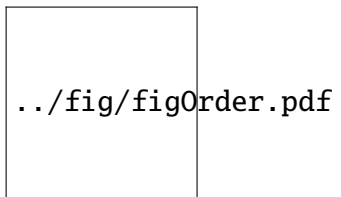


Figure 10: **Co-occurrence signal decays when the shortest path between a pair of species decay** The Z-score distribution are plotted against the shortest path for A willows-herbivores interactions, B herbivores-parasitoids interactions, C birds-plants interactions and D the pitcher plants network. 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.

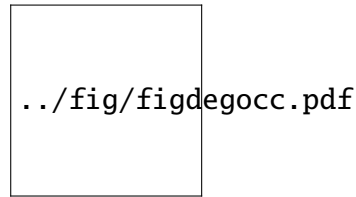


Figure 11: Co-occurrence significance decreases as the cumulated occupancy increases

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 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. 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)

3.14 Material and methods

All analysis have been performed using, table SX includes funciton and package that were used.

3.14.1 Datasets

3.14.1.1 Willows leafs network

3.14.1.2 Pitcher plants network

3.14.1.3 Caribbean Hummingbirds-Plant network

3.14.1.4 North American Trees datasets

3.14.1.5 French Breeding Birds Survey datasets

3.14.1.6 Keeping 1%

3.14.2 Traits-based distance

Kind of data / method of clustering (see table E)

For trees, we used a trait-based distance built of nine functional traits whose values retrieved from (??), see **Supplementary Table 3**.

3.14.3 Building metaweb

For datasets 1,2,3 we built network based on observed interactions and derived associated quantities, *i.e.* degrees of species and shortest-path, using the R package “igraph” (see Supplementary table 2).

3.14.4 Co-occurrence measurement

Following (???), 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 $E_{i,j}$ in light of their relative position of species within the network. The expected co-occurrence is derived based on the assumption that species are independent.

3.14.5 Hypergeometric distribution

This distribution has been mentioned in a different context (see ???) but have been fully exploited only 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 n sites in which two species 1 and 2 may occur. Species 1 is present in n_1 while 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, thi 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 calculated the expected co-occurrence $E_{i,j}$ under the hypothesis that all site were identical for all species.

3.14.6 GLM and RF

3.14.6.1 Climatic data

We used the global climate layers provided data WolrdClim, 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 axis in GLM and RF.

3.14.6.2 Generalized Linear Model

For all datasets, we performed a Generalized Linear Model (???) using all the axis provided by the PCA as polynomials of degree 2. To constraints the number of parameters, we did not evaluate the interactions among axis. 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.

3.14.6.3 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.

3.14.7 Mean Z-score log occupancy

Distribution of Z-score => normal $\frac{1}{\sqrt{n}}$

For all species we add occupancy meaning that degree plus total spaced occupied are integrated,

3.14.8 Model Evaluation

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 curve (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 Figure= (???)).

3.14.9 Trees

Based on table (???) we scale all the variable and used then to generate a synthetic distance variables are unweighted leading to a dendrograms present in figure ??.

3.15 Supplementary results

3.15.1 Commenting S6-S9

3.16 Supporting Tables

3.16.0.1 R packages used

Analysis	Function	Package name(s)	Version	Citation
PCA	dudi.pca	ade4	1.7.4	Dray and Dufour (2007)
GLM	glm step	stats	3.2.3	R Core Team (2015)
Random Forests	randomForest	randomForest	4.6.12	(???)
networks structure	-	igraph	1.0.1	(???)
AUROC	somers2	Hmisc	3.17.2	(???)
	tazxise	Hmisc	3.17.2	(???)

Supplementary Table 2: R and packages used for the analyses.

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

Number of species 2: For each datasets the total number of species (column *Total*),

the number of species present in more than 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’.

	Species	TSN	maxH	GR	WD	TolS	TolD	AM	EM	LMA	Nmass		
	Abies balsamea		18032		25	1	0.34	5.0	1.0	0	1	151.00	1.66
	Acer negundo		28749		20	3	0.44	3.5	3.0	1	0	37.04	2.50
	Acer rubrum		28728		25	3	0.49	3.4	1.8	1	0	71.09	1.91
	Acer saccharum		28731		35	1	0.56	4.8	2.3	1	0	70.63	1.83
	Betula alleghaniensis		19481		25	3	0.55	3.2	3.0	0	1	46.08	2.20
	Betula papyrifera		19489		25	3	0.48	1.5	2.0	0	1	77.88	2.31
	Carpinus caroliniana		19504		8	1	0.58	4.6	2.0	0	1	49.05	2.15
	Carya cordiformis		19227		25	1	0.60	2.1	4.0	0	1	44.05	2.60
	Fagus grandifolia		19462		25	1	0.56	4.8	1.5	0	1	61.22	2.04
	Fraxinus americana		32931		30	2	0.55	2.5	2.4	1	0	76.75	2.12
	Fraxinus nigra		32945		20	2	0.45	3.0	2.0	1	0	71.94	2.10
	Fraxinus pennsylvanica		32929		25	3	0.53	3.1	3.9	1	0	87.72	1.80
	Larix laricina		183412		25	3	0.48	1.0	2.0	0	1	120.00	1.36
	Ostrya virginiana		19511		12	1	0.63	4.6	3.3	1	0	37.04	2.20
	Picea glauca		183295		25	1	0.35	4.2	2.9	0	1	302.86	1.28
	Picea mariana		183302		20	1	0.41	4.1	2.0	0	1	294.12	1.12
	Picea rubens		18034		25	2	0.38	4.4	2.5	0	1	304.67	1.15
	Pinus banksiana		183319		20	3	0.42	1.4	4.0	0	1	243.90	1.24
	Pinus resinosa		183375		25	3	0.39	1.9	3.0	0	1	294.12	1.17
	Pinus strobus		183385		30	3	0.36	3.2	2.3	0	1	121.92	1.42
	Populus balsamifera		22453		25	3	0.37	1.3	1.8	1	1	83.46	1.95
	Populus grandidentata		22463		20	3	0.39	1.2	2.5	1	1	70.45	2.50
	Populus tremuloides		195773		25	3	0.37	1.2	1.8	1	1	82.02	2.16
	Prunus pensylvanica		24799		12	3	0.36	1.0	2.0	1	1	50.00	2.40

Species	TSN	maxH	GR	WD	TolS	TolD	AM	EM	LMA	Nmass		
Quercus alba		19290		35	1	0.60	2.9	3.6	0	1	81.21	2.39
Quercus macrocarpa		19287		15	1	0.58	2.7	3.9	0	1	92.74	2.27
Quercus rubra		19408		25	2	0.56	2.8	2.9	0	1	84.20	2.06
Thuja occidentalis		505490		15	1	0.30	3.5	2.7	1	0	223.00	1.02
Tsuga canadensis		183397		30	1	0.40	4.8	1.0	0	1	122.55	0.99
Ulmus americana		19049		35	3	0.46	3.1	2.9	1	0	79.47	2.07
Ulmus rubra		19050		25	3	0.48	3.3	3.0	1	0	59.88	2.50

Supplementary Table 3: Tree species and traits used TSN maxH GR WD Wood Density, TolS, TolD, AM, EM, LMA, Nmass(???).{#tbl:trees}

3.17 Supporting Figures

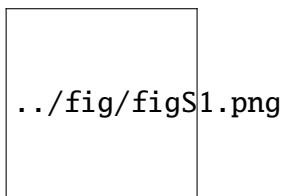
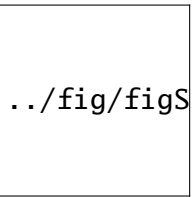


Figure 12: **Sites of the study**



../fig/figS2.pdf

Figure 13: **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.

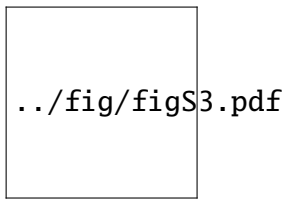
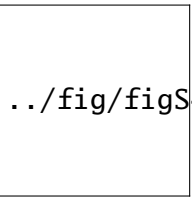


Figure 14: **Evaluation of the SDM approaches** For each dataset, the distributions of performance of generalized linear models (light grey symbols) and random Forest (dark grey symbols) for all species are presented.



../fig/figS4.pdf

Figure 15: **** 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).

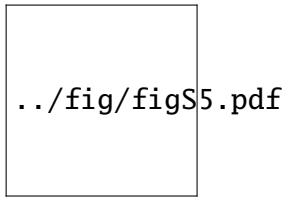
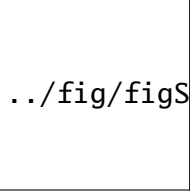


Figure 16: **Co-occurrence signal decays when the shortest path between a pair of species decay** Distribution of Z-scores for all interactions are grouped by shortest-path indicated by the first numbers below boxplots. The other figures below stand for the number of pairs of species included within the distributions.



../fig/figS6.pdf

Figure 17: *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.

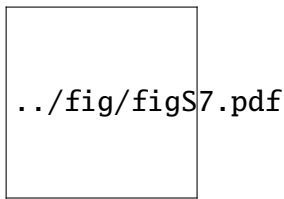
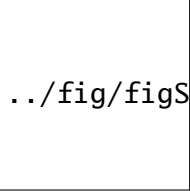


Figure 18: **The degree of species partially explains the decrease of the co-occurrence strength** For the herbivores (A-C) and the parasitoids in the willow leafs network datasets (D-F), the hummingbirds in the Caribbean hummingbirds datasets (G-I) and all species in the pitcher plants network that 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.



../fig/figS8.pdf

Figure 19: *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.

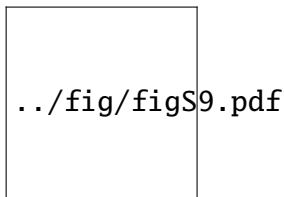


Figure 20: **Predicting herbivore distribution based on the distribution of willows** For the herbivores in the willow leafs network dataset, we compared the AUC obtained when using willow it does not interact with (circles) a willow in interacts with (squares) and the set of willow it interacts with (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 4

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.

