

UNIVERSITÉ DU QUÉBEC À RIMOUSKI

INFLUENCE DES INTERACTIONS BIOTIQUES SUR LA RÉPARTITION
GÉOGRAPHIQUE DES ESPÈCES

THÈSE
PRÉSENTÉE
COMME EXIGENCE PARTIELLE
DU DOCTORAT EN BIOLOGIE
EXTENSIONNÉ DE
L'UNIVERSITÉ DU QUÉBEC À MONTRÉAL

PAR
KÉVIN CAZELLES

MARS 2017

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion 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 travail de recherche de cycles supérieurs* (SDU-522 - Rév.01-2006). Cette autorisation stipule que « conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal 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 Montréal 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. »

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.

REMERCIEMENTS

Au royaume des idées, nous transformons les fragments de notre existence en songes. Chaque conversation est une caisse de résonance offerte à nos propres songes. Chaque lecture est une conversation, un échange avec d'autres qui, bien qu'étant absents, nous livrent leurs pensées. Comment alors ne pas croire que tout le travail ici présenté n'est autre que le récit de ma rencontre et des échanges avec tous ces autres ? Un récit qui à son tour deviendra une nouvelle source de songes pour des lecteurs que je ne connaîtrai peut-être jamais.

L'aventure de mon doctorat a été bien plus collective que ce j'imaginai. Longue est la liste de ceux qui apportèrent une pierre à l'édifice ici présenté. Des pierres les plus solides pour bâtir les fondations aux plus petits cailloux qui donnent un rien de charme et d'originalité à l'ensemble, nombreux sont les contributeurs au présent travail. À tous ces autres qui parfois ignorent ce qu'ils m'ont offert, j'adresse un chaleureux merci. Dans cette multitude de rencontres, certaines laissent des empreintes plus profondes et méritent que je leur destine des remerciements plus personnels.

Pour m'avoir donné ma chance et offert une confiance sans faille, je veux remercier vivement Dom, Nico et David, sans qui l'aventure n'aurait pas été la même. Pour faire ces premiers pas sur un chemin incertain, mieux vaut partir avec ceux qui pensent qu'on en trouvera la fin.

Un chemin qui m'amena sur deux continents, dans deux pays, à travailler dans deux universités au sein de deux laboratoires. Le laboratoire, ce n'est pas seulement un lieu, ce sont aussi des personnes, autant de sources de réflexion et d'inspiration qui jamais ne tarissent. Pour de précieux échanges à Rimouski, je remercie Hedou, Idal-

flex, Jacquette, Jo Brasco, Pippin le Solver, Tim. Solarik, les *behind the scenes* que nous avons abondamment échangés m'ont été d'une précieuse aide dans les derniers moments. Pour de passionnantes conversations à Montpellier, un grand merci Alain, Andreï, Claire, Clara, Florian, Isa, Marie, Pierre et Simon.

Dans les couloirs des laboratoires, il y a aussi des liens qui se nouent autour d'intérêts convergents et qui se transforment en une amitié très appréciée. Pour tout ce qu'ils m'ont apporté, je remercie Sonia (et Thomas), Albouy (et Séverine et Léo et Louis) et le joyeux Legagneux (et Aurélie et Margot et Juliette et Zélie et Romane). Dans ces mêmes couloirs, j'en ai croisé se frottant des yeux fatigués par la lumière des écrans. Pour mes geeks préférés, ceux qui connaissent comme moi l'appel du clavier, quand l'envie de coder devient trop forte, je désire taper un `\huge{merci}`. Merci Casajus, Team Zissou, Dave et Flaul pour les morceaux de code et bien plus : votre enthousiasme et votre insatiable curiosité.

D'un pays à l'autre, d'un laboratoire à l'autre, d'un couloir à l'autre, d'un projet à l'autre, l'aventure n'est, bien heureusement, nullement monochromatique et le rose de certains instants cède régulièrement la place à un gris parfois bien sombre. Partagés ses questionnements, ses colères, ou encore sa tristesse avec d'autres voyageurs est un soutien plus que précieux. Marion, Matoche, Clem, pour ce soutien qui a été si important, je vous remercie mille fois.

Pour ne pas faiblir face aux péripéties, le lieu de repos doit être choisi avec soin. Le temps du doctorat, il est important de trouver son havre de paix. Pour avoir été les colocataires de cet endroit merveilleux qu'a été la Maison des Courges, un immense merci à Camille, Élo, Gigi, Jean, Jerem, Jerem, Lau, Léo, Marie, Palardy, TriTri. Palette, bien sur, un remerciement tout particulier pour toi et pour tout ce que nous partageons.

Certains alliés étaient là avant même que l'aventure ne commence et constituèrent des repères essentiels dans un voyage quelques fois extrêmement déboussolant. À ces étoiles du sol qui nous guident quand la nuit vient de tomber, j'adresse un grand merci.

Ainsi, à mes amis du TerTer de Nanterre, qui me rappellent où j'ai grandi, les joies de mon parcours en banlieue parisienne, à ceux que j'ai connus il y a parfois plus de vingt ans, à vous Ariane, Bibi, Cendrars, Gomar, Gronico, K-wik, Matos, Misté, Rufo, Tinico, je vous dis un immense cimer ! Un spécial cimer pour toi Rhum, mon frerot, qui n'a jamais cessé de me surprendre.

À ma famille, ceux dont je partage parfois une partie de mon patrimoine génétique, mais souvent tellement plus. Un profond et chaleureux merci à Jean-Louis, Josette, Monique, Jean-Claude, Yvette, Tatoche, Tonton François, c'est toujours un immense bonheur de vous voir.

Papi, Mami, éternels soutiens, guides indispensables de mes premiers pas à aujourd'hui, merci. Des grands-parents comme vous, tous les petits enfants en rêve... Je vous le promets, regardez, je fais de la science, je vais être docteur, c'est forcément vrai. Sachez que l'aventure aurait été bien plus pénible sans vous.

Papa, père et pair, c'est quand même pas mal ! Ton appui au cours de ce long voyage m'a été d'une aide très précieuse. Tu as été un éclaireur génial vers qui j'ai pu me tourner quand l'horizon semblait encore bien loin, merci papa.

ClaCla et Nico le haricot, merci pour votre générosité, votre folie et pour le bonheur que vous respirez, je vous dois pas mal de sourire dans les moments où j'étais au plus bas. Merci Pépette pour ta spontanéité et ta fraîcheur, je sais que j'étais absent ces derniers temps, il faudra qu'on se rattrape !

Maman, la plus belle des mamans. Ton fils n'est pas souvent passé par la maison ces dernières années, eh oui, l'aventure m'a éloigné de mes terres d'origine. Rassures-toi, ton fils ne t'oublie pas, il sait très bien tout ce que tu as fait pour lui jusqu'à aujourd'hui et il t'en est éternellement reconnaissant.

à JASCERB,

AVANT-PROPOS

C'est animé par le désir de comprendre le fonctionnement des écosystèmes que j'envoyai, à l'automne 2011, un courriel à Nicolas Mouquet. Un courriel bien heureux car il me propulsa au Canada à la rencontre de Dominique Gravel, fraîchement devenu professeur à l'UQAR, et de Nicolas Mouquet lui-même alors en visite. C'est à Rimouski que j'ai découvert la biogéographie et que j'ai commencé mon stage de master (maîtrise dirait-on à Rimouski) qui déboucha sur la thèse de doctorat présentée dans les pages qui suivent. C'est avec Dominique Gravel, Nicolas Mouquet et David Mouillot que j'ai découvert le problème de l'intégration des interactions écologiques en biogéographie qui m'a profondément intéressé. Ces trois chercheurs m'ont mis le pied à l'étrier et je souhaite les remercier vivement pour cela.

Le doctorat que j'ai entrepris a été rendu possible grâce au soutien financier de plusieurs institutions que je souhaite remercier ici. Les trois premières années de mon doctorat ont été financées par une bourse du ministère français de l'éducation nationale, de l'enseignement supérieur et de la recherche, qu'il en soit remercié. J'ai également reçu un soutien financier du Conseil de Recherches en Sciences Naturelles et en Génie du Canada (CRSNG) que je remercie. Ce soutien m'a notamment permis de faire une quatrième année de thèse qui m'a été très profitable. De même, je remercie le programme des chaires de recherche du Canada qui, en plus de m'avoir offert un complément de bourse, a soutenu la mise en place du laboratoire de Dominique Gravel au sein duquel j'ai préparé la thèse ici présentée. Pour mes déplacements entre la France et le Canada, j'ai reçu un soutien du programme de mobilité FRONTENAC destiné aux doctorants inscrits en cotutelle de thèse franco-québécoise, je remercie tous ceux qui participent au bon fonctionnement de ce programme. J'adresse un grand merci au Centre de la

Science de la Biodiversité du Québec (CSBQ) et particulièrement à l'équipe qui participe à son dynamisme avec la mise en place d'initiatives qui participent fortement à rassembler les acteurs de la biodiversité au Québec. Je remercie aussi le CSBQ pour les bourses destinées aux étudiants qu'il met en place et dont j'ai personnellement bénéficié pour me rendre au laboratoire de Miguel Araújo à Madrid. Je profite de mentionner le nom de ce chercheur pour le remercier de son accueil et pour les précieux conseils qu'il m'a prodigués.

J'ai passé quatre ans à faire un doctorat en écologie et parmi les défis les plus délicats auxquels je me suis confronté, j'en ai relevé un bien singulier : celui d'expliquer, à des noms spécialistes, mon travail de thèse. Comment, en effet, comprendre que je m'intéresse à des questions importantes relatives à la biodiversité alors que je passe mon temps à coder derrière un ordinateur. Il faut une certaine habileté pour expliquer les tenants et les aboutissants de mon travail à celui qui me répond qu'il trie ses poubelles quand je déclare faire de l'écologie. L'effort de communication sur mes activités de recherche me semble, dans le contexte actuel, crucial et je souhaite prendre le temps à l'avenir pour faire cet effort.

J'ai passé quatre ans à étudier certains aspects de la biodiversité et je n'ai pourtant pas donné beaucoup de place à l'expression de mes convictions citoyennes sur le sujet. Je cherche encore à concilier le chercheur que je souhaite devenir et le citoyen que je suis déjà. Dans l'attente d'une telle conciliation, c'est par un message du moi citoyen que je souhaite ouvrir ma thèse en empruntant les mots du *Petit Prince* d'Antoine de Saint Exupéry :

— Les hommes de chez toi, dit le petit prince, cultivent cinq mille roses dans un même jardin. . . . et ils n'y trouvent pas ce qu'ils cherchent. . . .

— Ils ne trouvent pas, répondis-je. . .

— Et cependant ce qu'ils cherchent pourrait être trouvé dans une seule rose
ou un peu d'eau. . .

— Bien sûr, répondis-je.

Et le petit prince ajouta :

— Mais les yeux sont aveugles. Il faut chercher avec le coeur.

TABLE DES MATIÈRES

AVANT-PROPOS	ix
LISTE DES FIGURES	xvii
LISTE DES TABLEAUX	xxii
LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES	xxiv
RÉSUMÉ	xxv
ABSTRACT	xxvii
INTRODUCTION GÉNÉRALE	1
Des îles et des espèces	2
En suivant Wallace	2
En suivant MacArthur et Wilson	4
Quelles informations renferment les distributions d'espèces ?	7
Enjeux de la connaissance de la répartition géographique des espèces	9
Travail théorique et modélisation	11
Les processus qui façonnent les aires de répartition	15
Biogéographie historique	15
Capacités de dispersion	16
Contraintes abiotiques et niche écologique	18
Réseaux d'interactions : interdépendance des espèces	21
Échanges d'informations génétiques et processus micro-évolutifs	23
L'invasion européenne du frelon asiatique	25
Cadre théorique de la thèse	26
Une vision puissante de la dynamique des distributions d'espèces	27
Le modèle mathématique et les prédictions de la TIB	29
L'importance de la TIB dans des développements théoriques plus récents	31
Le rôle des interactions dans la distribution des espèces	35

Importance des interactions dans la distribution	36
Un problème d'échelle?	39
Vers une biogéographie énergétique	42
ARTICLE 1	
À PROPOS DES INTERACTIONS BIOTIQUES ET DES CONTRAINTES ENVIRONNEMENTALES À L'ÉCHELLE BIOGÉOGRAPHIQUE	44
1.1 Résumé en français	44
1.1.1 Contexte scientifique	44
1.1.2 Publication associée	45
1.1.3 Traduction du résumé de l'article publié	46
1.2 Title	47
1.3 Authors	47
1.4 Abstract	47
1.5 Introduction	48
1.6 The model	52
1.6.1 A simple probabilistic biogeographical model	52
1.6.2 Integrating biotic interactions	55
1.6.3 Integrating environmental gradients	58
1.7 Exploring the model	61
1.7.1 Biodiversity distribution over environmental gradients	64
1.8 Discussion	67
1.9 Acknowledgment	74
ARTICLE 2	
UNE THÉORIE DE LA COOCCURRENCE DES ESPÈCES DANS LES RÉSEAUX D'INTERACTIONS	75
2.1 Résumé en français du deuxième article	75
2.1.1 Contexte scientifique	75
2.1.2 Publication associée	77
2.1.3 Traduction du résumé de l'article publié	77

2.2	Titre	78
2.3	Auteurs	79
2.4	Résumé en anglais	79
2.5	Mot-clef	80
2.6	Introduction	80
2.7	Definitions	84
2.8	Acknowledgments	101

ARTICLE 3

QUELLES TRACES LES INTERACTIONS BIOTIQUES LAISSENT-ELLES DANS LES DONNÉES DE COOCCURRENCE DES ESPÈCES? 103

3.1	Résumé en français du troisième article	103
3.1.1	Publication envisagée	104
3.1.2	Traduction du résumé en anglais	105
3.2	Title	107
3.3	Abstract	107
3.4	Introduction	108
3.5	Material and Methods	110
3.5.1	Datasets	110
3.5.2	Quantifying co-occurrence	111
3.6	Results	112
3.6.1	Discussion	114
3.6.2	Box 1	118
3.6.3	Tables	121
3.7	Figures	121
3.8	Supporting Information	123
3.8.1	Material and methods	123
3.8.2	Co-occurrence measurement	127
3.8.3	Hypergeometric distribution	127

3.8.4	GLM and RF	129
3.8.5	Supporting Tables	130
3.9	Supporting Figures	139
ARTICLE 4		
VERS UNE THÉORIE MÉTABOLIQUE DE LA BIOGÉOGRAPHIE DES ÎLES		149
4.1	Résumé en français du troisième article	149
4.1.1	Contexte scientifique	149
4.1.2	Publication envisagée	150
4.1.3	Traduction du résumé en anglais	151
4.2	Title	152
4.3	Abstract	152
4.4	Introduction	153
4.5	Model description	157
4.5.1	Energetic constraints on local food webs	159
4.5.2	Energy fluxes and transfer efficiency	160
4.5.3	Extinctions	163
4.5.4	Simulations	164
4.6	Results	166
4.7	Discussion	168
CONCLUSION GÉNÉRALE		174
Interactions écologiques et répartition géographique des espèces		174
Des modèles théoriques à développer		174
Des théories pour mieux appréhender les données de cooccurrence		176
Vers une biogéographie prédictive ?		179
Les défis à relever dans un monde en changement		179
Des règles en écologie ?		182
Une théorie intégrative de la biogéographie		184
ANNEXE A		

COMMENT LA BIODIVERSITÉ S'INSTALLE EN TERRITOIRES ISOLÉS .	186
5.8 Préambule	186
5.9 Le modèle de MacArthur et Wilson	187
5.10 Comment un modèle déterministe émerge-t-il de phénomènes aléatoires ?	188
5.10.1 Modèle stochastique ?	188
5.10.2 Les objets mathématiques requis	189
5.10.3 Les briques élémentaires du modèle stochastique	189
5.10.4 Assemblons les briques !	191
5.10.5 Retrouvons le modèle classique	192
5.11 Applications et perspectives	193
ANNEXE B	
UN MODÈLE STOCHASTIQUE DE LA BIOGÉOGRAPHIE INSULAIRE POUR LES RÉSEAUX ÉCOLOGIQUES DANS UN ENVIRONNEMENT ABIOTIQUE VARIABLE	202
6.12 Stochastic rules in MacArthur & Wilson's model	202
6.13 A Markov chains model to describe island communities	205
6.13.1 Model for P non-interacting species	205
6.13.2 P Interacting species	206
6.14 Environmental gradient and island biogeography	208
6.15 Using Markov chains	208
6.16 Time continuous Markov chain	209
6.16.1 Solution for two species	210
6.16.2 Solution for P species	217
ANNEXE C	
PROBABILITÉS DE TRANSITION ENTRE RÉSEAUX TROPHIQUES . . .	220
7.17 The model	221
7.18 Reasonable approximations	224
7.19 Deriving species richness	225

LISTE DES FIGURES

0.1	La Théorie de la Biogéographie des Iles (TIB)	32
0.2	Intégration des interactions et des contraintes abiotiques dans la TIB. . .	39
1.1	Dynamics of the community assembly	58
1.2	Effects of biotic interactions on colonization-extinction dynamics	62
1.3	Equilibrium for interacting species along an environmental gradient . .	65
1.4	Probabilities of species richness for different types of interaction	68
1.5	Biodiversity distribution along environmental and connectance gradients	69
2.1	Co-occurrence in multi-species networks	102
3.1	Probabilistic description of fundamental and realized niches. For a four species network, all the occurrence probabilities are derived along an environmental gradient assuming that (A) interactions are not limiting the distribution and (B) that predators 3 and 4 needs at least of one of its preys, <i>i.e.</i> species 1 or 2. Horizontal dotted lines in (A) and (B) stand for the occurrence probabilities reached at an environmental optimum. The co-occurrence signal is calculated for the following pairs : a) predators 3 and 4; b) predator 3 and prey 1, predator 3 and prey 1; c) predator 3 and prey 1 or 2, predator 4 and prey 1 or 2. If no difference are found, 0 is expected.	122

- 3.2 **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 (light grey symbols). 124
- 3.3 **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. Note that we used the same y-axis for panels A and B as they regard two different kind of interaction of the same dataset. 125
- 3.4 **Co-occurrence significance decreases as the cumulated occupancy of preys increases.** For a given species, Z-scores are averaged over the all set species it interacts with and plotted against the joint distribution of the same set of species. We do so for the herbivores in the willows leafs network (panels A to C), the parasitoids in the willow leafs network (panels D to F), the hummingbirds in the Caribbean hummingbirds datasets (panels G to I) and all species in the pitcher plants network that consume other species (panels J to L). The x-axis is expressed as a log proportion of the total number of sites. Black symbols are mean Z-scores significantly different from 0 (see SI 3.8). 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 3.8). 126

3.5	Sites of the study	140
3.6	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.	141
3.7	Evaluation of the SDMs. For each dataset, the distributions of AUC for GLMs (light grey symbols) and RFs (dark grey symbols) for all species are presented.	142
3.8	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 different distances 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). They were also computed to show whether whether Z-score were greater for hypergeometric versus GLM and GLM versus RF (grey symbols).	143
3.9	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.	144
3.10	Changes in co-occurrence signal with increased 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.	145

- 3.11 **The degree of species partially explains the decrease of the co-occurrence strength.** For the herbivores (A-C) and the parasitoids in the willow leaf 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. 146
- 3.12 **Co-occurrence and cumulated occupancy of predators.** This figure corresponds to figure 3.4 in the main text, but the Z-score were calculated for preys (host plants) rather than for predators (pollinators). Mean Z-score are computed for willows (A-C) and herbivores (based on the herbivores-parasitoids only, D-F) of the willows leaf 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. 147
- 3.13 **Predicting herbivore distribution based on the distribution of willows.** For the herbivores in the willow leaf 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. 148
- 4.1 Species-Energy Relationship 165
- 4.2 Species occurrence probability along an energy gradient, grouped by trophic level 167
- 4.3 Species occurrence probability by energy gradient grouped by body mass class 169

4.4	Average body mass, number of preys, predators of predator species and the average number of interactions per species on the island	171
5.5	Représentation graphique de la solution de l'équation différentielle . . .	188
5.6	Dynamique de la biodiversité de trois zones protégées	194
5.7	Représentation de la formule des probabilités totales.	201
7.8	Simple food webs illustrating the TTTIB	224

LISTE DES TABLEAUX

1.1	Conditional probabilities between potential assemblages	56
2.1	Presence/absence dataset for three species and 10 sites.	91
2.2	Direct associations between pairs of species for different modules . . .	94
3.1	Data sets analyzed in this article.	121
3.2	R and packages used for the analyses. GLM: Generalized Linear Model, PCA: Principal Component Analysis. AUROC: Area Under the Re- ceiver Operating Characteristic.	131
3.3	For each datasets we provide the total number of species (column <i>Total</i>), the number of species present in more that 1% of the total number of sites (column <i>Selected</i>), and the number of species for which traits information are available (column <i>Traits</i>). The symbol ‘-’ means ‘not relevant’.	132
3.4	Tree species and traits used. Abbreviations are as follows: TSN - Taxonomic Serial Number defined by Integrated Taxonomic Information System (ITIS), maxH - Average maximum height, GR - Growth rate, WD - Wood Density, TolS - Shade tolerance, TolD - Drought tolerance, AM - Arbuscular mycorrhiza (Endomycorrhiza), EM - Ectomycorrhiza, LMA - Leaf mass per area, Nmass - Nitrogen content per leaf mass unit (?).	133
3.5	List of the Boolean traits used to compute Euclidean distances between all pairs of species in the French Breeding Birds Survey.	136
4.1	Hypotheses associated to the four scenarios.	164

7.2	Community states associated to simple networks	222
7.3	Transition matrix of the continuous-time Markov chain (part 1/2)	227
7.4	Transition matrix of the continuous-time Markov chain (part 2/2)	228

LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

DOI	<i>Digital Object Identifier</i> ; identifiant numérique d'objet.
GIEC	Groupe d'experts Intergouvernemental sur l'Évolution du Climat.
IPBES	<i>Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services</i> ; Plateforme intergouvernementale sur la biodiversité et les services écosystémiques.
JSDM	<i>Joint Species Distribution Model</i> ; modèle de distribution jointe d'espèce.
MVP	<i>Minimum Viable Population</i> ; population minimale viable.
SAR	<i>Species Area Relationship</i> ; relation aire-espèces.
SDM	<i>Species Distribution Model</i> ; modèle de distribution d'espèce.
SER	<i>Species Energy Relationship</i> ; relation énergie-espèces.
TIB	<i>Theory of Island Biogeography</i> ; théorie de la biogéographie des îles.
TTIB	<i>Trophic Theory of Island Biogeography</i> ; théorie trophique de la biogéographie des îles.

RÉSUMÉ

La biogéographie est l'étude des mécanismes et des processus influençant la répartition géographique des êtres vivants. Bien que ces mécanismes soient relativement bien identifiés, les biogéographes recherchent toujours une théorie cohérente pour les articuler. Parmi les problèmes les plus fréquemment soulevés en biogéographie, figure celui de l'intégration des interactions écologiques dans les modèles de distribution d'espèces. Bien que la littérature scientifique apporte un ensemble de preuves soulignant le rôle prépondérant des interactions dans la structuration des communautés locales, on trouve relativement peu d'études révélant les empreintes laissées par les interactions dans les données de distribution d'espèces. Proposer une explication simple et claire à ce problème demeure un défi important que la biogéographie doit mener. Le problème majeur que pose l'absence de réponse claire sur le rôle des interactions aux larges échelles spatiales est que la plupart des scénarios de changements de biodiversité partent de l'hypothèse que les interactions sont négligeables. Si cette hypothèse est remise en question, alors il devient nécessaire de réviser ces scénarios et soutenir le développement de méthodologies incluant les relations entre les espèces.

Dans le travail de thèse que je présente, je me confronte au problème de l'intégration des interactions dans la répartition des espèces et de leurs effets sur ces distributions. Je commence cette thèse par un travail théorique sur le sujet car les théories classiques en biogéographie relèguent souvent au second plan les interactions écologiques. Au premier chapitre, je traite de l'intégration des interactions écologiques dans un modèle théorique de distribution d'espèces issu d'une des théories les plus importantes en biogéographie : la théorie de la biogéographie des îles. Ce travail montre comment les effets conjoints des facteurs biotiques et abiotiques changent les attendus de la théorie classique. En m'appuyant sur ce premier chapitre, je montre au second chapitre comment les interactions peuvent se répercuter dans les données de cooccurrence d'espèces. Ces données indiquent la présence ou l'absence de plusieurs espèces sur un même ensemble de sites dispersés sur de larges étendues spatiales. À l'aide d'un modèle probabiliste, j'obtiens des résultats théoriques liant les données de cooccurrence et l'information contenue dans les réseaux écologiques. Je démontre clairement que les interactions affectent les données de cooccurrence. Je montre également que plus le nombre d'interactions séparant deux espèces est grand, plus il est délicat de déceler des traces de cette relation indirecte dans les données. De même, si une espèce entretient de nombreuses interactions, il sera difficile de trouver une quelconque trace des interactions dans les données de cooccurrence pour cette espèce.

Au troisième chapitre, je présente l'analyse de cinq jeux de données de cooccurrence pour lesquels la description des interactions était disponible. Avec ces données, j'ai été capable de confirmer mes hypothèses formulées au second chapitre en montrant que les espèces qui interagissent coexistent différemment de celles n'interagissant pas. Mes résultats indiquent aussi que l'abondance d'interactions est un frein à leur détection dans les données de cooccurrence. Cependant, en intégrant la similarité des facteurs abiotiques pour les différents sites, je montre que les signaux de cooccurrence s'affaiblissent pour parfois disparaître. Mes résultats suggèrent donc qu'en utilisant des facteurs abiotiques pour inférer les probabilités de cooccurrence, une partie du lien entre les espèces est capturée, mais cette part est entachée d'une grande incertitude. Ceci vient questionner la qualité des prédictions données par les modèles classiques de distribution d'espèces actuellement utilisés.

Les résultats de ma recherche apportent des éléments théoriques nouveaux sur le rôle des interactions écologiques dans le tracé des aires de répartition des espèces en plus de proposer une méthode originale pour étudier les données de cooccurrence d'espèces : les regarder à la lumière des réseaux écologiques. Avant de conclure ma thèse, je propose au chapitre 4 une démarche prometteuse pour aller encore plus loin dans l'intégration des interactions en biogéographie : les introduire par le biais des contraintes énergétiques, ce qui offre une base solide pour une théorie métabolique de la biogéographie.

MOTS-CLÉS : Biogéographie, interactions biotiques, réseaux écologiques, contraintes abiotiques, cooccurrence, théorie de la biogéographie des îles, théorie métabolique de l'écologie.

ABSTRACT

Biogeography is the study of the mechanisms and processes that control the geographical distribution of plants and animals. Although the list of mechanisms is clearly identified, biogeographers are still struggling to find a consistent theory to integrate their interaction. One of the most pressing challenges currently in the field of biogeography is the successful integration of ecological interactions in species distribution models. Although the scientific literature points out the evidence of the controlling role interactions play on local community structure, relatively few studies have demonstrated its importance over large geographical gradients. Developing a concise, clear explanation for this issue remains a significant challenge that biogeographers need to answer. The main issue associated to the lack of a clear answer concerning the role of interactions at broad spatial scales is that most of scenarios of biodiversity changes assume that interactions can be ignored. When tested, if this hypothesis is proven false, then a reconsideration of species distribution models and their development must be undertaken to include relationships among species.

In this thesis I tackle the issue of integrating the species interactions, and how this affects species distributions. I begin this thesis with a theoretical investigation on this topic, where classical theories have typically ignored ecological interactions. In the first chapter of the thesis I present the integration of interaction networks into a theoretical model of species distribution coming from one of the most important theory in biogeography : the theory of island biogeography. This work shows how together the biotic and abiotic factors can affect the expectations derived from the classical theory. Building upon the findings in the first chapter, in the second chapter, I show how interactions can affect co-occurrence (between species) data. Such data contains the presence or absence of several species for a similar set of sites dispersed along large latitudinal gradients. Using a probabilistic model, I obtain theoretical results linking co-occurrence data and the information included in ecological networks. I clearly demonstrate that interactions shape co-occurrence data. Furthermore, I show that the higher the number of links between two species, the more difficult it is to detect their indirect interaction. Similarly, if a species experiences many interactions, it is then challenging to detect any sign of interactions in static co-occurrence data for this species.

In the third chapter of the thesis, I assess five sets of co-occurrence data, which had descriptions of their interactions available. Using this data, I was able to confirm my hypotheses put forth in my second chapter, by showing that species co-occur differently from non-interacting one. These results also point out that the abundance of interaction

must preclude their detection in co-occurrence data. However, when accounting for abiotic similarities among sites, signals of interactions are weakened. Therefore, my results suggest that using abiotic factors to infer co-occurrence probabilities capture a part of the link between species and further pinpoint the uncertainty associated to this part. As a result of these findings, the predictive power of classical species distribution models used to date is brought into question.

My research findings bring new theoretical elements to the forefront when considering the influence of ecological interactions and how they shape species geographical distributions, while also introducing an original methodology for studying species co-occurrence : examining them in the light of ecological networks. Before concluding, my fourth and final chapter, I propose a promising new avenue to further investigate integrating species interactions in biogeography. Here, I introduce interactions in terms of energetic constraints, which will provides a sound basis for a metabolic theory of biogeography.

KEYWORDS : Biogeography, biotic interactions, ecological networks, abiotic constraints, co-occurrence, theory of island biogeography, metabolic theory of ecology.

INTRODUCTION GÉNÉRALE

La biogéographie est l'étude de la répartition géographique des espèces. Aujourd'hui, dans un monde où les écosystèmes sont très perturbés, les biogéographes ont pour mission, entre autres, l'établissement des scénarios de changement de la biodiversité. Cependant, les efforts de recherche accumulés sur plus d'un siècle ont révélé des obstacles majeurs à la compréhension de l'action simultanée des mécanismes dessinant les aires de distribution des espèces, faisant de leur prédiction un véritable défi. En réponse à ces difficultés, certains auteurs ont proposé d'aller vers un renouvellement de la théorie de la discipline. Parmi les étapes majeures de ce renouvellement, l'intégration des interactions écologiques au sein d'une théorie plus intégrative de la biogéographie tient une place particulièrement importante. C'est sur ce point qu'ont porté mes travaux de thèse. J'ai particulièrement réfléchi sur comment intégrer les interactions dans les modèles de distribution d'espèces, c'est-à-dire envisager le rôle joué par les liens écologiques dans la géométrie de leur répartition géographique.

Je propose, dans la présente introduction, de cerner un peu mieux les enjeux de ma thèse. Pour cela, je commence par montrer que la compréhension des mécanismes déterminant la distribution des espèces était très avancée dès la fin du XIX^{ème} siècle durant lequel, après plus de 150 ans de voyages d'exploration scientifique, la connaissance plus exhaustive de la richesse biologique terrestre a mené à la théorie de l'évolution. Je souligne que dès les années 1960, la biogéographie s'est dotée d'une théorie ambitieuse qui a grandement participé à la compréhension de l'ensemble des processus mis en jeu. Je présente ensuite différents cadres conceptuels en biogéographie qui, bien que présentant de nombreuses qualités sont aujourd'hui appelés à intégrer davantage de processus écologiques, notamment les interactions écologiques. Je reviens dans un dernier

temps plus en détails sur ce point pour contextualiser plus finement l'objet central de ma thèse qui propose des éléments de réponse à deux grandes questions : comment intégrer les interactions écologiques dans des modèles en biogéographie ? Quelles traces ces interactions laissent-elles sur la géométrie des aires de répartition ?

Des îles et des espèces

En suivant Wallace

Dans l'introduction de son livre *Island Life* paru en 1881, le célèbre naturaliste Alfred Russel Wallace nous rapporte deux faits étonnants qui justifient pleinement l'examen attentif de la répartition géographique des espèces (?). Premièrement, le biogéographe démontre, à l'aide de nombreux exemples, que l'éloignement entre deux régions du monde est un argument insuffisant pour conclure que leur composition faunistique et floristique sont dissimilaires. Ainsi, la comparaison des avifaunes de l'île japonaise d'Hokkaido et de l'Angleterre, séparées par des milliers de kilomètres, révèle une proximité des paysages ornithologiques très supérieure à celle constatée dans l'analyse comparée des oiseaux des îles indonésiennes de Bali et de Lombok pourtant distantes de quelques kilomètres seulement. Deuxièmement, en s'appuyant sur les différences entre la faune brésilienne et la faune africaine, Wallace souligne la faiblesse du pouvoir prédictif des variables climatiques pour décrire les compositions faunistiques présentes sous des latitudes similaires. Ces constatations soulignent l'utilité de croiser les informations relatives à la répartition géographique des espèces avec une analyse taxonomique pour aboutir sur une analyse avoir un regard plus profond sur leur histoire. Dans

le cadre de la théorie de l'évolution¹, en effet, cette analyse taxonomique est une analyse historique : Wallace montre que la compréhension d'un problème spatial, celui des aires de répartition de groupes d'espèces, n'est possible que par une compréhension temporelle, celle de l'histoire des espèces. Cette idée est clairement énoncée dans la suite de son introduction :

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

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

1. En 1881, la théorie de l'évolution est toute récente, Wallace est l'un des premiers penseurs à en décrire les grands principes, avec notamment son article de 1858 *On the Tendency of Varieties to Depart Indefinitely From the Original Type* qui témoigne clairement que ses idées sur les variations temporelles des espèces étaient très proches de celles de Charles Darwin à qui il avait d'ailleurs envoyé son manuscrit peu avant que *De l'origine des espèces* ne soit publié (?).

très larges², ce qui apporte un éclairage substantiel qui se double cependant d'un obstacle épistémologique majeur : si l'explication ultime de la présence d'une espèce en un point donné est le produit d'une série de contingences historiques, quelles peuvent être les fondations d'une théorie de la biogéographie ? Ce n'est qu'au XX^{ème} siècle que des propositions de cadres théoriques convaincants émergeront.

En suivant MacArthur et Wilson

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

Biogeography has long remained in a natural history phase, accumula-

2. L'âge de la terre est très débattu à l'époque. Bien que l'ensemble des savants s'accordent pour aller bien au-delà des 6000 ans bibliques, il n'y a alors pas de consensus. Wallace affirme à la page 212 du chapitre 10 de *Island Life* que la vie se développait il y a au moins 500 millions d'années (?), ce qui est audacieux pour l'époque mais bien en dessous de l'âge des plus anciens fossiles découverts à ce jour qui a estimé en 2011 à 3.4 milliards d'années (?).

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

ting information about the distribution of species and higher taxa and the taxonomic composition of biotas. Interpretative reasoning has been largely directed to the solution of special problems connected with the histories of individuals taxa and biotas. Without doubt this descriptive activity will continue to be of fundamental importance to the science, one of the most physically adventurous of all scientific enterprises and, in the richness of the detail it unfolds, esthetically pleasing. But biogeography is also in a position to enter an equally interesting experimental and theoretical phase.

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

Le paradigme de la TIB a eu un impact considérable sur les développements théoriques en écologie (?). Au centre du projet de la TIB, se loge la volonté de mettre l'espèce au coeur de la biogéographie afin de permettre à la discipline de s'enrichir des mécanismes biologiques qui sont un moteur essentiel de la variation dans la distribution des espèces. L'intérêt de leur « biogéographie de l'espèce » (le terme est mentionné à l'avant-dernière phrase de l'ouvrage ?, p.183) est dans l'affirmation qu'il faut regarder les contraintes conjointes de l'évolution (qui met un certain nombre d'espèces avec des caractéristiques propres en présence) et du contexte écologique qui détermine les conditions de colonisation et d'extinction. Cette intrication de l'écologie et de l'évolution est bien inscrite dans la pensée de MacArthur et Wilson, même si la puissance de leur vision réside dans le fait de les occulter en partie.

Près de 50 ans après la parution de leur livre, une des clefs en biologie semble être la compréhension des rétroactions qui existe entre écologie et évolution dans les variations spatiales et temporelles de la biodiversité. Je reprends ici trois aphorismes cités par ? concernant les liens entre biologie, écologie et évolution : « *Nothing in biology makes sense except in the light of evolution.* » (?); « *Nothing in evolutionary biology makes sense except in the light of ecology.* » (?); « *Nothing in evolution or ecology makes sense except in the light of the other.* » (?); La chronologie de ces citations est un indice de la reconnaissance actuelle du besoin (de la nécessité ?) de croiser écologie et évolution. Un parallèle avec les sciences humaines me semble pertinent dans lequel l'écologie serait à la biologie ce que la géographie est aux sciences humaines et l'évolution serait à la biologie ce que l'histoire est aux sciences humaines. Nous pouvons certes étudier l'une sans l'autre, mais le dialogue entre les deux disciplines est indispensable afin d'éviter que chacune avance en faisant des hypothèses fortes sur l'autre qui finiront éventuellement par nuire à une compréhension plus profonde de la biologie. Par exemple, supposer que les variations démographiques ont des origines pu-

rement écologiques devient problématique si les variations génétiques sont suffisantes pour expliquer qu'une partie importante de cette variation comme cela l'a été montré sur une population de moutons de Soay (?). Je ne cherche pas à nier l'utilité des savoirs acquis de manière autonome par un champ disciplinaire, j'insiste simplement sur l'importance de mettre ces connaissances en perspective les unes avec les autres en vue d'une synthèse indispensable pour décrypter l'information contenue dans les distributions d'espèces.

Quelles informations renferment les distributions d'espèces ?

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

Dans son ouvrage de 1972, MacArthur discute de l'ensemble de ces mécanismes, il considère aussi bien le rôle que peuvent jouer les variables climatiques que celui des interactions écologiques. En plus des exemples concrets amenés pour illustrer ses pro-

pos, MacArthur développe des modèles mathématiques pour prolonger la discussion. Dans son second chapitre, il formalise l'impact de la compétition sur la coexistence des espèces aboutissant ainsi sur un principe de ségrégation spatiale des espèces liées par ce type de relation : deux compétiteurs ne peuvent pas cooccurrencer (résider durablement au même endroit) sauf éventuellement sur une zone très restreinte de leur distribution (?). Toujours dans cet ouvrage, MacArthur évoque la distribution en damier que peuvent générer des espèces en compétition. La discussion de ce type de distribution sera approfondie par Jared Diamond (?) dont les travaux déclencheront un débat important sur la détermination de modèles nuls de cooccurrence (?) et sur laquelle mes travaux de thèse apportent de nouveaux éléments.

L'étude des limites géographiques d'un grand nombre d'espèces permet d'y déceler des généralités quant aux mécanismes qui les déterminent (?). L'examen des variations spatio-temporelles apporte une information très utile sur l'importance relative des divers mécanismes. Le contexte des changements climatiques est une bonne illustration de ce principe car les bouleversements actuels des répartitions géographiques des espèces permettent en effet de pointer le rôle majeur des mécanismes micro-évolutifs auparavant sous-estimés (?). Enfin, l'examen des distributions doit aussi être un examen des co-distributions : il faut s'intéresser à l'information de sous-ensembles d'espèces et notamment aux espèces en interaction afin de tester si la biologie laisse des empreintes dans la géométrie des aires de répartition. Par exemple, dans mes travaux de thèse, je propose de regarder l'intersection de l'aire associée à un ensemble de proies pour savoir ce qu'elle nous apprend sur la distribution de leur prédateur (présenté au chapitre 3).

Enjeux de la connaissance de la répartition géographique des espèces

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

En ciblant l'étude de la distribution de certaines espèces, la biogéographie rencontre des enjeux socio-économiques majeurs. Ainsi, pour un pays comme la France, la restriction des zones favorables à la croissance de la vigne envisagée à l'aide des scénarios de changements climatiques (?) pourrait conduire à des pertes économiques importantes et un bouleversement identitaire des grandes régions viticoles. De plus, détecter aujourd'hui un potentiel viticole futur dans des régions où cette production n'existe pas encore peut conduire à des augmentations drastiques du prix des terres agricoles. En guise de second exemple, je pose la question suivante : où seront les érablières de demain ? La réponse est donnée par la détermination de la répartition future des aires favorables à la croissance de l'érable à sucre (*Acer saccharum*) et à sa capacité à les atteindre afin de s'y établir (?).

Je termine avec un troisième exemple : la perte des pollinisateurs et notamment des abeilles. Pas moins de quatre grandes classes de facteurs d'origine anthropique les mettent en danger : les changements climatiques, le changement dans l'utilisation des terres⁴, l'apparition de nouveaux pathogènes (dont l'acarien parasite *Varroa destructa* vecteur de nombreux virus) et l'arrivée d'espèces invasives (comme le frelon asiatique ?). Le défi actuel est de prédire la distribution future des pollinisateurs en intégrant ces multiples aspects et leurs interactions. De plus, dans le cas des espèces invasives, il faut comprendre comment une espèce peut sortir de son aire de répartition naturelle et s'établir dans une nouvelle région dont elle était jusqu'alors absente.

Actuellement, les outils de prédiction des aires de répartition futures reposent essentiellement sur les scénarios de changements climatiques (?). Ces modèles dits de distribution d'espèces (que je noterai SDMs dans la suite en référence au terme anglais *Species Distribution Models*) proposent une démarche cohérente : la connaissance basée sur les corrélations de variables climatiques permet d'établir une relation climat-répartition d'espèce. En utilisant les résultats des climatologues qui dérivent les variations climatiques associées à des scénarios d'émission de gaz à effet de serre par les activités humaines, les biogéographes établissent les probabilités de présence des espèces dans les conditions climatiques futures. Cependant, la relation climat-présence n'est qu'une facette du lien qui unit les espèces à l'espace et chaque nouvelle invasion nous montre à quel point il est difficile de prédire les aires de répartition. Ces problèmes de qualité des prédictions sont le reflet de lacunes théoriques qui amènent plusieurs chercheurs à se positionner en faveur d'un renouvellement des fondations théoriques pour édifier une biogéographie plus intégrative (???). Bien sur, ces appels soulèvent des défis im-

4. Parmi les changements les plus problématiques relatifs aux terres agricoles figurent la monoculture et l'utilisation massive de pesticides. Les pesticides de la famille des néonicotinoïdes semblent affaiblir les colonies d'abeilles.

portants dont on ne peut qu'espérer qu'ils soient relevés au plus vite pour faire face à la crise environnementale actuelle.

Travail théorique et modélisation

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

Rassembler et intégrer des faits

Développer une théorie, c'est avant tout la mise en cohésion d'un certain nombre de faits, d'observations. Dans la TIB, par exemple, MacArthur et Wilson proposent une explication cohérente de l'augmentation de la richesse spécifique dans les îles de plus grandes tailles. De manière générale et poppérienne, trois principes encadrent la construction d'une théorie scientifique :

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

Une boutade, dont je suis malheureusement incapable de me souvenir de son auteur,

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

Des modèles pour explorer et tester la théorie

Utiliser un modèle c'est substituer volontairement un objet d'étude parfois complexe par un objet dont l'étude des propriétés sera plus accessible, tout en nous renseignant sur l'objet d'étude initial. Un organisme-modèle, par exemple, est un organisme souvent facile à élever et manipuler, et pour lequel beaucoup de connaissances ont été acquises et dont on admet que les résultats sont similaires pour l'ensemble des organismes proches; il sert souvent d'unité empirique pour un ou plusieurs groupes de recherche. Les modèles statistiques permettent, entre autres, de quantifier la plausibilité d'une hypothèse donnée au regard de données empiriques. De même, pour un travail de modélisation mathématique, la description du modèle peut être donnée par une série

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

d'équations dérivées d'une théorie validée par un certain nombre de savoirs empiriques. Les modèles, quelle qu'en soit leur nature, sont des objets avec lesquels il est possible d'explorer et de tester une théorie.

Les modèles sont souvent décrits comme une simplification de la réalité : comment, en effet, prétendre que les mécanismes biologiques décelés chez *Arabidopsis thaliana*⁶ sont les mêmes que ceux à l'œuvre pour l'ensemble des plantes à fleurs ? Pour combien de systèmes proie-prédateur le modèle de Lotka-Volterra est-il pertinent ? Les limites des modèles doivent être reconnues, mais il ne faut pas nier ce qu'ils apportent. Les modèles sont autant d'occasions pour explorer une ou plusieurs prédictions d'une théorie. Le choix du modèle employé est lié à l'histoire du chercheur qui l'utilise, à ses propensions à utiliser avec succès telle ou telle démarche scientifique, c'est ce que rappelle Kevin McCann dans la préface de son livre *Food Webs* (?) : « *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 of rabbits and lynx.* » En d'autres termes, certaines personnes ont plus de facilités pour penser en termes d'abstractions mathématiques alors que d'autres font meilleur usage de représentations plus concrètes. Je me réclame volontiers de la première catégorie de personnes, je pense que les mathématiques fournissent un cadre de pensée très puissant, comme l'indique le célèbre écologue Robert May (?, p. 791) :

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

6. Il s'agit de la plante modèle par excellence, son génome fut le premier à être séquencé chez les plantes (?).

Dans mes travaux, j'ai essayé d'utiliser les mathématiques pour développer des modèles dont le point de départ a été une réflexion collective autour du rôle que pouvaient jouer les interactions dans la répartition géographique des espèces. J'ai alors établi un cadre théorique avec lequel j'ai dérivé des prédictions dont certaines semblent être vérifiées dans les données empiriques.

Nouvelles prédictions

Après l'établissement d'une théorie expliquant un certain nombre de faits et pour laquelle un certain nombre de tests ont été réalisés, le raisonnement fondé sur celle-ci peut conduire à la production de nouvelles prédictions qui, si elles sont vérifiées, la renforceront. En revanche, l'apparition de faits expérimentaux en désaccord avec cette théorie peut conduire à des précisions relatives au domaine d'application de la théorie ou alors à l'émergence d'une théorie nouvelle pour les expliquer (?). Ces dernières années, la physique nous a donné deux exemples très probants de l'efficacité du pouvoir de l'imagination associé à un cadre mathématique rigoureux avec la vérification expérimentale de théories énoncées bien avant que les outils permettant de la vérifier existent. En 2012, c'est la détection du Boson de Higgs dont l'existence fut prédite en 1964⁷. Cette année, c'est la détection des ondes gravitationnelles, soit 100 ans après qu'Albert Einstein en ait prédit l'existence (?). En biogéographie, une théorie devrait être capable, par exemple, de dresser des cartes précises de risque d'invasion à l'échelle mondiale pour l'ensemble des espèces. Je pense que nous en sommes encore loin, néanmoins, le chemin pour y parvenir passe par une connaissance approfondie de l'ensemble des mécanismes qui interviennent dans le tracé des aires de répartition, c'est-à-dire connaître

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

leur nature, leur portée mais aussi leurs interactions et leurs importances relatives.

Les processus qui façonnent les aires de répartition

Biogéographie historique

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

Les processus de grandes amplitudes temporelles sont cependant dominés par leur com-

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

posante historique (et donc non reproductible) et prédire des phénomènes tels que l'extinction des dinosaures est, dans le meilleur des cas, extrêmement compliqué. Néanmoins, dans les mouvements de grandes amplitudes se manifestent des processus qui agissent en permanence. Ainsi, l'étude de la diversification des bousiers entreprise par Joachim Hortal et ses collègues (?) montre que la dernière glaciation a laissé des empreintes encore visibles dans la carte de répartition de la diversité de ce groupe : la limite de la thermocline 0°C durant le dernier maximum glaciaire (il y a 21000 ans environs) sépare les zones de forte diversité en bousiers des autres. De plus, ils montrent que la diversité phylogénique des espèces nordiques, c'est-à-dire plus tolérantes au froid, est un sous-ensemble phylogénétique bien identifié, il semble alors probable qu'une petite partie de l'embranchement des bousiers soit à l'origine des colonisations nordiques. Ainsi, après une contraction de la zone favorable au développement des bousiers, les mouvements de colonisation ont marqué à la fois la carte de répartition de la richesse spécifique de ce groupe, mais également la carte de répartition des différentes branches de l'arbre phylogénétique des bousiers européens (?).

Capacités de dispersion

La remontée nordique des bousiers depuis le dernier maximum glaciaire est le résultat d'événements de dispersions individuelles. Au cours de leur vie, les bousiers parcourent de grandes distances à la recherche de nourriture, on peut imaginer qu'au fil des générations, si les conditions environnementales le permettent, certains individus établissent des populations de plus en plus nordiques. Ce qui est vrai pour ce groupe d'espèces mobiles l'est aussi pour des espèces sessiles comme les plantes qui possèdent également des capacités de dispersion du fait de la dissémination de leurs graines par des mécanismes très diversifiés. Ce rapport à l'espace des différents organismes est une forme

de diffusion : des mouvements stochastiques conduisent à une augmentation de la répartition (c'est une question de probabilité), mais cette diffusion n'est pas totalement libre.

Plusieurs types de contraintes limitent l'élargissement de l'aire de répartition d'une espèce. Pour les espèces terrestres, les mers et les océans sont des obstacles majeurs à la colonisation de nouveaux territoires. À l'échelle régionale, les rivières, les hauts reliefs peuvent fortement limiter la dispersion d'une espèce. De même, pour les plantes dont la stratégie de dissémination est l'anémochorie, la vitesse et la direction des vents sont des facteurs primordiaux pour comprendre la propagation de l'espèce. Enfin, à l'échelle du paysage, il existe très souvent une mosaïque d'habitats plus ou moins favorables à la dispersion d'une espèce. Toutes ces possibilités sont complexes à intégrer et c'est en partie pour cela que la théorie de la biogéographie a été fondée sur les îles, les flux de colonisation y étant relativement faciles à identifier : de la côte la plus proche vers l'île.

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

Contraintes abiotiques et niche écologique

Au sixième chapitre de son livre de 1972, *Geographical Ecology*, MacArthur illustre l'importance des contraintes climatiques avec l'exemple de l'aire de répartition du cactus Saguaro (*Cereus giganteus* en 1972 mais aujourd'hui *Carnegiea gigantea*). Ce résident des hauteurs du désert de Sonora⁹ est sensible au gel et ne peut guère résister à une exposition de quelques dizaines d'heures au gel. Cette contrainte physiologique explique bien les limites nord et est de sa répartition. Pour la limite sud, il semblerait que l'abondance des pluies hivernales qu'il y trouve lui soit défavorable. En s'appuyant sur les conditions climatiques actuelles dans lesquelles le cactus se développe, des résultats récents prédisent que dans le cadre des changements climatiques, *Carnegiea gigantea* trouvera refuge à des altitudes supérieures. Cependant, ce mouvement pourrait être entravé par l'augmentation de la fréquence des feux (?).

Cette démarche de croisement de la limite des aires de répartition avec des variables climatiques est une forme répandue de la détermination de la niche écologique d'une espèce. Ce concept de niche est très débattu en écologie et son caractère éluif s'accompagne d'un certain nombre de problèmes de définition¹⁰. Afin d'éviter ces problèmes, je parlerai, dans la suite, de la niche au sens de Joseph Grinnell qui en tentant d'expliquer la restriction de la répartition du moqueur de Californie écrit :

9. Ce désert couvre le sud de l'Arizona et la Californie et une large partie des états mexicains de Basse-Californie et du Sonora pour une superficie de plus de 300 000 km².

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

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

Dans ses travaux, Grinnell montre que la présence du Moqueur de Californie est fortement corrélée à des conditions de température et d'humidité assez élevées (?). Ainsi, la niche écologique d'une espèce donnée, au sens de Grinnell, est un ensemble de conditions environnementales dans laquelle il est probable de trouver cette espèce. En allant au-delà des observations *in situ* et en caractérisant l'ensemble des conditions d'existence possibles, alors on détermine une niche écologique théorique appelée niche fondamentale. Cette caractérisation expérimentale a été poussée à son paroxysme dans l'article de Michael Kearney et Warren Porter sur le gecko nocturne australien *Heteronotia binoei* (?). Ils ont montré qu'en combinant des mesures physiologiques (dont le taux métabolique au repos, et des températures caractéristiques comme la température cumulée nécessaire au bon développement des œufs) avec des données climatiques, ils obtenaient une bonne concordance des probabilités d'occurrence et des observations, justifiant ainsi la démarche prédictive en s'appuyant sur des scénarios de changements climatiques pour essayer de comprendre les répartitions futures. De manière générale, cette méthode est la recherche de facteurs abiotiques limitant le développement d'une espèce à tous les niveaux de son cycle de développement et donc sa répartition géographique. Au niveau du Panama, par exemple, l'étude de ? a montré que les distributions locales et régionales de 48 espèces d'arbres s'expliquent par la sensibilité à la sécheresse, donc à une variation dans la disponibilité locale d'une ressource. Ces corrélations convaincantes fondent les SDMs qui sont des solutions techniques (statistiques) pour l'application de la méthode générale que je viens de décrire (??).

L'engouement actuel autour de ces modèles est lié à l'espoir de pouvoir faire des

prédictions fiables sur la géographie de la biodiversité mondiale de demain dans un contexte de changements climatiques. Cette démarche s'est appliquée avec succès à différents cas, par exemple en 2009, Tingley et ses collègues ont montré que sur 53 espèces d'oiseaux étudiés dans la Sierra Nevada, 48 ont colonisé de nouveaux sites où les conditions de températures et de précipitations leur étaient plus favorables (?). Une autre justification de l'utilisation abondante des SDMs est la relative facilité de leur mise en application grâce à l'abondance des données climatiques et d'occurrence à laquelle s'ajoute le partage des implémentations numériques de ces méthodes statistiques. Pour le premier type de données, WorldClim illustre cette facilité d'accès en proposant des données à l'échelle mondiale gratuitement téléchargeables (voir <http://worldclim.org>; ?). Pour les données d'occurrence, plusieurs initiatives offrent des données gratuites dont les plus exhaustives sont celles disponibles sur le portail de données sur la biodiversité à l'échelle mondiale GBIF (*Global Biodiversity Information Facility*, voir <http://www.gbif.org>) qui présentent cependant des biais liés à l'inégalité d'échantillonnage des régions du globe (?). Enfin pour ce qui est du partage des implémentations des SDMs, on peut évoquer le logiciel libre R (?) dont certaines extensions sont dédiées à l'utilisation des SDMs et sont largement utilisées dans la communauté scientifique des biogéographes.

Un des principaux problèmes posés par l'utilisation massive de ces approches est la faible remise en question des hypothèses sur lesquelles elles reposent. Le message délivré par les SDMs doit être pris comme une potentialité : étant données les conditions actuelles dans lesquelles une espèce est trouvée et sachant les variations climatiques données par les modèles climatologiques, s'il n'existe pas d'obstacles majeurs au mouvement de l'espèce en question, alors il est probable que celle-ci se déplace en suivant les conditions climatiques qui sont similaires à celles dans laquelle elle est actuellement trouvée, ce qui nous permet de savoir où sera l'espèce demain. Ce message est délivré

en supposant 1- une forme d'équilibre entre la distribution des espèces et les conditions climatiques actuelles et 2- que les espèces sont indépendantes (?). Ces deux hypothèses sont très fortes et demandent un examen approfondi. Etant donné que mon travail de thèse porte sur la seconde, je propose de la discuter dans le paragraphe suivant.

Réseaux d'interactions : interdépendance des espèces

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

Au-delà de la compétition, l'écologie des réseaux nous montre aujourd'hui la difficulté de concevoir les espèces comme étant des entités indépendantes ; elles sont en effet reliées par des relations de natures très diverses. Les relations trophiques sont les plus évidentes, il existe cependant une myriade d'interactions non trophiques qui affectent aussi la démographie des espèces (voir ?, pour une réflexion sur le sujet et une classification de ces interactions). De plus, aucun argument théorique ne justifie actuellement la

primauté d'un type d'interaction sur les autres. Récemment, les interactions trophiques et non trophiques ont été exhaustivement analysées pour 104 espèces des écosystèmes intertidaux rocheux de la partie centrale de la côte chilienne, révélant ainsi que les interactions non trophiques y étaient globalement plus abondantes et concentrées sur les bas niveaux trophiques (?).

L'écologie des réseaux est traversée de débats dont le plus important est sans doute celui de la relation qu'il existe entre la diversité spécifique d'un écosystème et sa stabilité (??). Autour de cette question, l'écologie s'est considérablement enrichi en outils mathématiques. Une preuve récente réside dans la mise en évidence par Stefano Allesina et Si Tang du caractère déstabilisant des interactions de compétition et de mutualisme et du rôle stabilisant des relations trophiques (?). Ce résultat est en effet la mise en application directe d'un résultat mathématique récent établi par Terence Tao et Vam Vu démontrant une loi générale sur la distribution des valeurs propres des matrices aléatoires (?). Les réseaux contiennent de nombreuses informations sur l'écologie des populations et doivent être placés au centre d'une théorie intégrative de la biogéographie. Cette idée était déjà suggérée par MacArthur et Wilson au dernier paragraphe de leur théorie de la biogéographie (?) :

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

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

Échanges d'informations génétiques et processus micro-évolutifs

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

Les causes de la divergence du code génétique entre deux populations d'une même espèce sont multiples. Ce qui m'intéresse ici, c'est le différentiel démographique positif que peut engendrer la variation du code dans un milieu nouvellement exploré par l'une des deux populations alors que cette même variation dans un autre milieu ne l'était

pas. La vitesse des mécanismes mis en jeu semble bien plus rapide au point que ceux-ci puissent jouer des rôles prépondérants dans la réponse des espèces aux changements climatiques (?). En 2009, Joan Balanyá et ses collègues publiaient un article dans lequel ils comparent la composition génétique de la mouche *Drosophila subobscura* entre des échantillons contemporains et des échantillons prélevés 24 années auparavant en Europe et en Amérique (où elle a été introduite accidentellement). Leurs résultats montrent que dans les zones de réchauffement climatique avéré, il y a aussi un changement de la composition génotypique avec une plus grande importance des génomes adaptés aux températures chaudes (?).

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

L'invasion européenne du frelon asiatique

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

Après son arrivée en 2004, le frelon s'étendait déjà en 2006 largement sur l'Aquitaine avec une aire de répartition française constituée d'une bande de 300 km du nord au sud et de 150 km d'est en ouest (?) et cela malgré l'éradication systématique des nids détectés. Alors que deux nids étaient observés en 2004, 1636 nids ont été observés en 2009 et en 2013 près des trois quarts des départements français étaient affectés (?). Des travaux récents tentent de caractériser les conditions climatiques favorables au dévelop-

pement de cette espèce (?) et révèlent qu'une large partie de l'Europe occidentale est une zone de développement probable. Un autre phénomène intéressant lié à cette invasion est l'arrivée concomitante en Corée du Sud du frelon asiatique où sa propagation a cependant été bien moins rapide (?). L'explication plausible de la différence de succès de la même espèce est une différence dans la composition en espèces proches des deux régions : en Europe, il n'y a qu'une espèce de frelon *Vespa crabro*, alors qu'il y en a de six en Corée du Sud dont *Vespa mandarinia* qui est une meilleure compétitrice (?). Cette nécessité de faire appel à la composition biologique pour comprendre les raisons d'un changement d'aire de répartition est de ce fait tout l'intérêt des travaux théoriques menés durant mon doctorat.

Cadre théorique de la thèse

Les développements entrepris dans mes travaux visent à incorporer les interactions écologiques dans la TIB. Je vais maintenant revenir sur cette théorie plus en détails afin de mieux comprendre pourquoi elle a une telle empreinte sur la biogéographie (et l'écologie de manière plus générale). Je signale d'ailleurs que ces idées étaient partagées par d'autres écologues et qu'il y a, à ma connaissance, deux autres découvertes indépendantes des idées qui ont conduit à la théorie. La première découverte est attribuée au spécialiste des lépidoptères, Eugene Gordon Munroe, qui a formulé dès 1948, des idées similaires dans 5 des 555 pages de sa dissertation de graduation (??). La seconde est celle de Richard Levins et Harold Heatwole qui publie en 1963, soit la même année que l'article fondateur de la TIB, l'idée d'un équilibre de la richesse spécifique régi par les mêmes processus que ceux décrits par MacArthur et Wilson (?). Néanmoins, ce sont sans aucun doute MacArthur et Wilson qui ont marqué les écologues par l'ensemble des développements présentés dans leur livre *The Theory of Island Biogeography* (?).

Une vision puissante de la dynamique des distributions d'espèces

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

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

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

Il y a une forme de hasard et de nécessité qui fait écho à l'œuvre de Jaques Monod (?). Le biologiste et prix Nobel de médecine y présente les mutations au niveau de l'ADN comme une source de hasard dont la persistance n'est rendue possible que dans un cadre physico-chimico-évolutifs précis, la nécessité. Dans les travaux de MacArthur et Wilson, l'événement de colonisation peut être interprété comme un pourvoyeur de stochasticité alors que les contraintes écologiques régissent l'organisation des communautés. Outre le fait que la prédiction de la colonisation ne peut se faire qu'en termes de fréquence, le caractère stochastique de cette dernière donne une dimension historique aux assemblages insulaires. L'arrivée d'une espèce est en fait un tirage aléatoire (éventuellement pondéré par les capacités respectives de dispersion) dans un réservoir régional d'une singularité historique car l'espèce en question possède une histoire évolutive propre et des singularités physiologiques qui en découlent. À son arrivée sur l'île, son éventuelle insertion est déterminée par la rencontre des singularités de l'espèce et du contexte biotique et abiotique de l'île. Les espèces installées sur une île ont ainsi été passées au crible des contraintes écologiques, de cette forme de nécessité qui est renouvelée à chaque nouvelle insertion. C'est ainsi que l'on peut décrire le moteur de la reconfiguration perpétuelle des réseaux écologiques locaux. Une telle dynamique peut être également analysée comme une imbrication de deux échelles de processus : régionalement, le réservoir d'espèce est façonné par une histoire évolutive de grande amplitude liée à des processus climatiques eux aussi de grande échelle, alors que les événements insulaires relèvent de processus de plus courte portée (?).

Enfin, la TIB, bien que cela soit rarement souligné, fait l'hypothèse de l'équivalence écologique des espèces considérées : il n'y a ni plantes ni animaux, ni proies ou prédateurs, elles sont toutes des unités de la richesse spécifique de la région étudiée. En analysant les exemples donnés par les auteurs en 1967, on est amené à penser que la théorie a été développée pour des groupes d'espèces aux rôles écologiques similaires

et phylogénétiquement proches. Ainsi, le premier exemple donné porte sur l’herpéto-faune (amphibiens et reptiles) antillaise et non sur un inventaire exhaustif de toutes les espèces des Antilles (?). Il est d’ailleurs possible que la validation de leur théorie ne soit possible qu’en fixant le niveau écologique à étudier. Paradoxalement, c’est en s’affranchissant de la diversité biologique qu’ils ont fait un grand bond vers la « biogéographie de l’espèce » qu’ils souhaitaient construire (?, p.183).

Le modèle mathématique et les prédictions de la TIB

Je ne rentre pas ici dans les détails mathématiques du modèle, ils sont néanmoins abordés au chapitre 1 ainsi que dans les deux annexes de la présente thèse ¹¹. J’écris ci-dessous l’équation qui résume à elle seule le paradigme livré par la TIB : les P espèces d’un continent colonisent l’île avec un taux individuel c , ce qui en augmente la richesse spécifique local S mais aussi les risques d’extinction dont le taux par espèce est noté e :

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

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

11. L’annexe 1 est un article de vulgarisation mathématique qui aborde de manière didactique la formulation la plus simple du modèle. L’annexe 2 détaille la formulation mathématique du modèle présenté au chapitre 1. Cela m’a conduit à contribuer à un article scientifique actuellement en révision dont je suis coauteur et pour lequel je présente ma contribution à l’annexe 3.

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

Cet équilibre est une prédiction très importante de la théorie, c'est même le point de départ des développements mathématiques dans le livre de 1967 (?). L'existence d'un tel équilibre a été validée par l'expérience de défaunation de Simberloff et Wilson mentionnée plus haut (?). Une seconde prédiction de la TIB est la variation de cet équilibre avec les caractéristiques de l'île. Dès leur article de 1963, MacArthur et Wilson présentent la taille de l'île comme un facteur affectant le taux d'extinction : plus l'île est grande, moins le risque d'extinction est grand (?). De même, ils supposent que l'isolement de l'île affecte le flux de migrants : plus l'île est isolée moins les événements de colonisation sont fréquents. J'ai résumé la vision classique de la TIB sur la figure 0.1 en y ajoutant les graphiques de l'article de 1963. Cette prédiction de la théorie est aussi l'origine : MacArthur et Wilson expliquent avec ces mécanismes que les îles de plus grandes tailles supportent plus d'espèces mais aussi que des exceptions liées à l'isolement peuvent exister. Cette relation est d'ailleurs présentée très tôt dans le livre de 1967 avec l'augmentation linéaire du nombre d'espèces de l'herpétofaune avec le logarithme de la surface des îles de l'ouest des Caraïbes (? , chapitre 2).

De manière plus générale, la TIB fournit une explication à la relation aire-espèce très discutée en écologie (?). Il s'agit de la courbe d'augmentation de la richesse spécifique (S) avec la surface d'échantillonnage (A). La question soulevée par l'étude de ces courbes porte sur la nature des mécanismes qui régissent les variations régionales. En liant la taille de l'île au taux d'extinction, La TIB propose une courbe de la forme $S = CA^z$ pertinente avec les observations présentées (?). La relation aire-espèce est

surtout connue pour ses applications dans le domaine de la conservation ¹². Elle permet, par exemple, d'estimer la taille qu'une zone de protection doit avoir pour atteindre un objectif de sauvegarde chiffré en nombre d'espèces (??). La relation peut être aussi utilisée dans le sens inverse pour apprécier les taux d'extinction liés à une dégradation d'habitat (?).

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

La théorie des métapopulations

Bien que ne représentant que cinq pourcents des terres émergées, ce sont bien les observations de la faune des îles qui ont mené à une vision paradigmatique de la biogéographie (?). L'importance des îles s'explique par leur relative abondance, leur disparité, leur diversité, la relative simplicité des assemblages biologiques qu'on y trouve et, comme je l'ai évoqué précédemment, la clarté des flux de migration (?). Cette dernière propriété est souvent absente pour des populations continentales ¹³. La théorie des métapopulations s'intéresse justement aux populations reliées entre elles par des flux de migrations (?). Le premier modèle de métapopulations a été proposé par Levins ¹⁴ lors d'une réflexion sur le contrôle démographique des ravageurs de cultures (?). Pour

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

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

14. Richard Levins qui avec Heatwole est un des codécouvreurs des idées de la TIB.

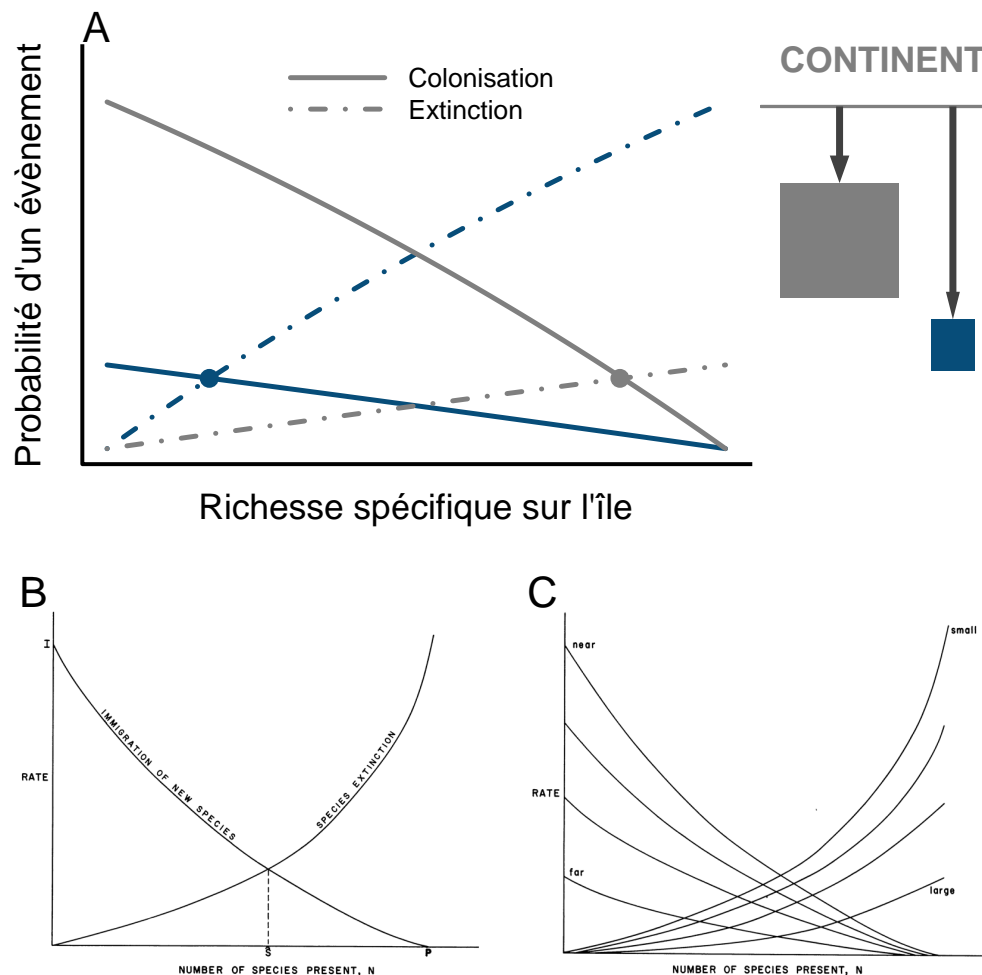


Figure 0.1: **La Théorie de la Biogéographie des Îles (TIB)** Le panel (A) illustre l'évolution des taux de colonisation et d'extinction pour deux îles aux caractéristiques différentes. Les tailles relatives des îles et les distances qui les séparent du continent sont schématisées sur la droite, les couleurs associent les îles à leurs courbes respectives. Le réservoir d'espèce régionale P est constitué de 100 espèces, les taux de colonisation et d'extinction sont exprimés en terme de probabilité d'événement (de colonisation ou d'extinction). Les points marquent les intersections entre les courbes d'extinction et de colonisation c'est-à-dire lorsque ces processus s'équilibrent. L'abscisse de ces points indique les richesses spécifiques de l'île à l'équilibre S_{eq} . (B) et (C) sont respectivement les Figures 4 et 5 extraites de l'article de 1963 de MacArthur et Wilson qui livre essentiellement le même message que celui illustré en (A) (?). La forme convexe des courbes de 1963 est justifiée par des facteurs biologiques qui ne sont pas intégrés dans l'équation qui confère une forme concave aux courbes, comme montré en (A).

un ravageur donné, les îlots de culture sont autant de patchs où une population peut se maintenir et disperser dans les autres patchs à proximité. Levins montre alors que les mesures de la lutte biologique doivent être conduites à large échelle pour en augmenter les probabilités de succès, c'est-à-dire d'extinction régionale du ravageur (?). Le modèle est simple et très proche de celui de la TIB : l'évolution de la proportion p est aussi gouvernée par des événements de colonisation c et d'extinction e :

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

La différence fondamentale avec la TIB est que la migration dépend de la proportion de patchs occupés : plus elle est importante plus la migration est importante. Parmi les démonstrations existantes, figurent les travaux menés par Ikkha Hanski sur les populations du mélitée du plantain (*Melitaea cinxia*) au sud-ouest de la Finlande (?). En plus de donner un cadre de pensée plus réaliste en termes de configuration spatiale, les dynamiques populationnelles associées sont bien comprises et mènent à des risques d'extinction mieux évalués (?). C'est aussi un cadre approprié pour l'étude des flux génétiques liés à l'arrangement spatial des différentes populations. Ainsi, toujours sur ces mêmes populations de papillons, Ilik Saccheri et ses collègues montrent qu'en ajoutant le degré d'hétérozygotie, ils obtiennent des prédictions précises quant à l'extinction locale des populations (?). Les travaux théoriques autour du concept de métapopulations proposent un certain nombre de paradigmes permettant d'évaluer le rôle que jouent les processus de colonisation et d'extinction dans les variations spatio-temporelles de la démographie d'une espèce et a été étendu à l'échelle de la communauté, on parle alors de métacommunauté (??). La prépondérance de ces mécanismes qui font la force de la TIB et de la théorie des métapopulations a été poussée à son paroxysme dans la théorie

neutre de la biogéographie.

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

La théorie neutre postule l'équivalence écologique entre les différents individus d'espèces éventuellement différentes et décrit les dynamiques populationnelles reposant sur les différences d'abondances relatives à l'échelle régionale et locale. Ainsi, en 1997, dans l'article fondateur de la théorie neutre, Stephen Hubbell décrit un modèle dans lequel le remplacement d'un individu mort dans une communauté locale est le résultat d'un tirage aléatoire : le nouvel individu peut être soit recruté localement et la probabilité que l'individu soit d'une espèce donnée dépend de l'abondance relative de cette dernière dans la communauté locale, soit le nouvel individu peut être un immigrant dont l'identité de l'espèce à laquelle il appartient est liée à l'abondance à l'échelle régionale de celle-ci (?). En plus des exemples donnés dans l'article de 1997, Hubbell montre de manière convaincante que dans la forêt tropicale du Panama, à la suite d'un chablis, le recrutement de l'arbre n'est pas prévisible par ses propriétés fonctionnelles et que le recrutement est similaire à la composition des alentours (?). La dynamique engendrée est appelée la dérive écologique, et est dominée par la stochasticité qui conduit presque certainement à l'extinction de toutes les espèces sauf une, phénomène contrebalancé par l'apparition d'espèces nouvelles (??).

La théorie neutre partage beaucoup de caractéristiques avec la TIB : les mécanismes fondamentaux sont l'extinction et la colonisation, l'hypothèse d'équivalence écologique et l'imbrication des échelles régionales et locales. Comme le fait remarquer Hubbell en 2010 dans le chapitre qu'il écrit dans *The Theory of Island Biogeography Revisited*, la théorie neutre place l'équivalence écologique au niveau des individus et non

plus au niveau des espèces (?). Une conséquence directe revendiquée par Hubbell est que cette hypothèse est cohérente avec la forme convexe des courbes de colonisation et d'extinction décrites par MacArthur et Wilson (voir la figure 0.1) qui n'est en fait pas expliquée par leur modèle (?). Le principe d'équivalence et la place importante que prend le hasard dans cette théorie a soulevé de très vifs débats et donné lieu à des argumentations à charge contre la véracité de cette théorie (voir par exemple ??). L'équivalence écologique doit, à mon sens, être comprise comme une abstraction de la singularité des espèces, une simplification de la diversité des systèmes biologiques, nécessaire à l'isolation d'une portion restreinte des phénomènes en cause dans la répartition géographique des espèces pour en évaluer le pouvoir explicatif. Bien qu'un certain nombre de cas d'études permettent de rejeter cette théorie (??), les défenseurs de la théorie neutre affirment qu'elle reste tout aussi utile quand une étude en démontre la fausseté (?). La théorie neutre peut en effet être présentée comme une jauge qui mesure l'importance des processus de différenciation de niches (?). Ainsi, pour certaines communautés, la dérive écologique est plus importante que pour d'autres, ce qui a été formalisé au travers de travaux théoriques proposant un continuum de la théorie neutre vers la théorie de la niche écologique (?). Malgré les possibilités offertes par ces deux théories, elles occultent largement les interactions écologiques qui sont factuelles ; si les observations donnent du crédit à ces théories, une théorie intégrative de la biogéographie doit expliquer pourquoi.

Le rôle des interactions dans la distribution des espèces

Mes travaux de thèse ont pour objectif de trouver des pistes pour intégrer les interactions écologiques dans la TIB. Plus précisément, je souhaite contribuer à i) la compréhension du rôle des interactions écologiques dans la géométrie de la répartition géo-

graphique des espèces, et ii) la recherche des traces qu'elles pourraient éventuellement laisser dans les données d'occurrence des espèces. Comme je l'ai mentionné plus haut cette idée est très ancienne, Wallace le remarque, dans son livre publié en 1881, avec la formule suivante : « *Both competition and predation appear now to be much more important in biogeography than people had formerly guessed* » (? , p.28).

Le problème de ces relations écologiques est leur spécificité, l'unicité de chacune d'entre elles, dont découlent un certain nombre de difficultés pour généraliser. Néanmoins, des travaux récents explorent des pistes prometteuses pour prédire les interactions, notamment sur la base de relations allométriques entre proie et prédateur (?). Du point de vue théorique et à l'examen des chapitres du dernier livre de MacArthur (?), il apparaît que l'intégration des interactions est une étape clef pour aller vers une biogéographie intégrative et c'est dans cette direction que j'ai mené ma thèse, essayant d'apporter des prémices de réponses pour arriver à une telle synthèse.

Importance des interactions dans la distribution

Dans la TIB, les interactions sont, en fait, omniprésentes en tant qu'une des composantes principales du processus d'extinction. Cependant, dans la formulation du modèle, elles ne sont jamais mentionnées explicitement, cachées dans le taux d'extinction e . Comme je le montre sur la figure 0.1, la différence dans l'allure des courbes dessinées par MacArthur et Wilson et celles obtenues en supposant un taux d'immigration et de colonisation sont différentes. D'après les auteurs, l'immigration devient plus difficile lorsque les espèces s'accumulent sur l'île et les extinctions sont de plus en plus fréquentes dues à l'intensification des interactions. Pour parler en termes de réseaux d'interactions, l'accumulation d'espèces sur l'île sature le réseau local et rend difficile

l'intégration d'une nouvelle espèce qui le rend par ailleurs de plus en plus instable. Une interprétation communauté-centrée de la TIB est tout à fait possible mais les liens entre les espèces ne sont pas formulés mathématiquement en 1967.

Depuis les années 60, la littérature théorique n'a cessé de discuter le rôle joué par les interactions intra- et inter-spécifiques dans la distribution spatiale des espèces. Il est reconnu que l'interdépendance des espèces détermine le caractère favorable de l'environnement au sens large (biotique et abiotique). En 2009, Robert Holt et Michael Barfield discutent de l'impact de la prédation sur la répartition d'espèces en compétition insistant alors sur le rôle majeur des interactions dans le dessin des aires de répartition (?). En 2012, William Godsoe et Luke H. Harmon introduisent les interactions dans un modèle simple de distribution d'espèce et montrent comment la probabilité de présence d'une espèce peut être affectée par la distribution d'une seconde et concluent alors que cela doit affecter vraisemblablement la qualité de prédictions des SDMs (?). La remise en cause des SDMs se concentrant sur les variables abiotiques est une étape fondamentale dans l'établissement des scénarios de biodiversité. Leur succès qui remonte à la fin du siècle dernier, a en effet relégué les interactions écologiques au second plan en démontrant que la corrélation avec les variables climatiques était peut-être suffisante, au moins en première approximation pour expliquer les aires de répartition (?). Pourtant, dès 1998, le travail précurseur d'Andrew Davis et ses collègues (?) avait fortement remis en question l'hypothèse d'indépendance des espèces (?). L'expérience, dont les résultats ont été publiés en 1998, est une analyse d'abondance de trois espèces de drosophile le long d'un gradient de température. Les comparaisons d'abondance sont menées pour toutes les combinaisons possibles de ces trois mouches (assemblages à 1, 2 ou 3 espèces) mais aussi en présence ou en absence d'un parasitoïde. La démonstration est sans appel : la compétition et le parasitisme affectent drastiquement la survie le long du gradient de température, les interactions écologiques affectant donc très probablement

les réponses aux changements climatiques.

Plus récemment, on constate une large engouement pour intégrer les relations écologiques dans les modèles de distribution d'espèces (??). Ainsi, des modèles de distributions jointes d'espèces (JSDMs, *Joint Species Distribution Models*) ont vu le jour dans les dernières années avec comme atout principal la prise en compte des corrélations entre les occurrences de différentes espèces (??). Néanmoins, ces efforts se heurtent à un manque de maturité des modèles et théories qui cherchent à rassembler distribution et interactions. Parmi les travaux récents, Franck Jabot et Jordi Bascompte ont rassemblé métacommunauté et écologie des réseaux pour souligner l'importance des relations écologiques dans la répartition géographique des espèces (?). De même, Dominique Gravel et ses collègues ont introduit en 2011 l'interdépendance proie-prédateur dans le modèle de la TIB menant aux prémices d'une théorie trophique de la biogéographie des îles (?) préfigurée par Holt (?). Ces travaux tentent de dépasser l'hypothèse d'équivalence écologique en vue de faire des prédictions plus précises concernant les compositions spécifiques attendues localement.

C'est dans la lignée de ces développements théoriques récents que s'inscrit mon premier chapitre de thèse. J'y montre comment l'intégration du concept de réseau écologique dans la TIB était possible tout en ajoutant la reconnaissance de performances plus ou moins importantes des espèces dans un contexte abiotique donné (niche écologique). Pour y arriver, je souligne l'intérêt de considérer les espèces sous la forme d'assemblage plutôt qu'une à une. Grâce à l'utilisation de probabilités conditionnelles d'assemblage dans un environnement abiotique donné, j'explore les conséquences simultanées des contraintes biotiques et abiotiques sur la distribution spatiale des espèces (voir la figure 0.2 pour une représentation schématique du modèle). Du point de vue technique, mon travail montre aussi qu'un retour aux processus stochastiques, tels que ceux présentés

en 1967, est une démarche puissante pour ajouter de nouveaux mécanismes dans la TIB.

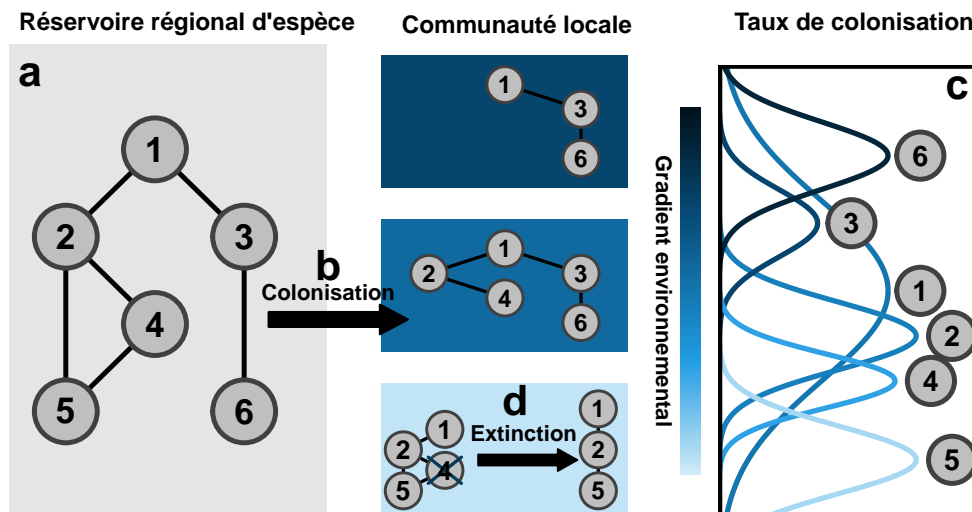


Figure 0.2: **Intégration des interactions et des contraintes abiotiques dans la TIB.** Pour intégrer les interactions, j'ai considéré non pas un ensemble d'espèces indépendantes, mais des espèces au sein d'un réseau qui est décrit à l'échelle régionale (a). Comme dans la TIB, ces espèces peuvent coloniser l'île (b), mais les taux de colonisation varient le long d'un gradient environnemental (c). Enfin, les interactions influencent les taux d'extinction locale (d). Voir le chapitre 1 pour une description complète du modèle.

Un problème d'échelle ?

En repartant de l'exemple classique de la ségrégation spatiale des tamias *Eutamias dorsalis* et *E. umbrinus* (?), j'ai précédemment mis en évidence qu'une information sur les interactions est contenue dans les aires de répartition de ces espèces. Il y a cependant deux caractéristiques qui peuvent conduire à la rareté de ce type de lecture : la singularité de l'interaction et son caractère local. Je reviens un peu plus bas sur la première propriété et m'arrête ici sur la seconde. Une idée dominante en biogéographie est que

les interactions ont des rôles majeurs à l'échelle locale mais que leurs conséquences sont de moins en moins perceptibles à mesure que l'on considère des échelles spatiales de plus en plus grandes (voir l'unique figure de ?). Du point de vue théorique, c'est tout à fait ce qui est décrit dans la TIB, car c'est à l'échelle locale que les interactions influencent l'extinction. Néanmoins, ces conséquences locales sont présentes sur l'ensemble de la distribution de l'espèce, il est alors pertinent de se demander pourquoi nous ne sommes pas capables de détecter les interactions en examinant les distributions d'espèces. En réalité, bien que cela soit rare, nous avons des preuves que cela soit possible dans certains cas. En 2010, Nicholas Gotelli et ses collègues divisent l'avifaune danoise en différentes catégories fondées sur la similarité écologique et démontrent que les répartitions des espèces d'une même catégorie se chevauchent moins qu'attendues aléatoirement (?). De même, en 2007, Risto Heikkinen et ses collègues ont obtenu des performances accrues de leurs modèles statistiques par l'utilisation de la répartition de six espèces de pics pour expliquer la présence de quatre espèces de hiboux (?). Dans cette même étude, le signal est plus fort lorsque les données sont dérivées de grilles spatiales à plus petites mailles (10x10 km contre 40x40 km), ce qui constitue un argument en faveur d'une dépendance à l'échelle, récemment supportée par d'autres travaux (?). Ce qui est remarquable dans les travaux de Gotelli et de Heikkinen est que l'utilisation de connaissances biologiques et écologiques a révélé une trace des interactions dans la distribution des espèces.

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

rayon d'action de ces interactions. Pour dépasser la question spatiale, il faut aussi envisager l'impact de la nature des interactions sur la répartition géographique. Ainsi, en 2014, Miguel Araújo et Alejandro Rozenfeld ont prouvé théoriquement que les interactions positives (mutualisme) se propageaient davantage que les interactions négatives (?). Par conséquent, la nature de la relation qui unit des espèces peut influencer la perte d'information contenue dans les aires de répartition. Suite à mes travaux sur l'intégration des interactions, je me suis penché sur un autre aspect qui peut influencer la perte d'information dans les données de présence : l'abondance des interactions. Au chapitre 2, je montre que les interactions directes et indirectes affectent les données de distributions mais aussi que l'abondance des interactions rend difficile de distinguer un quelconque signal : nous ne sommes plus en mesure de dire s'il y a des différences entre les paires d'espèces qui interagissent et celles qui n'interagissent pas. Ce qui est encore plus intéressant, c'est que j'ai accumulé un certain nombre d'indices dans des données de présence et d'absence réelle qui semblent confirmer nos prédictions. Je discute de ces résultats dans le chapitre 3 de cette thèse.

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

A few decades ago it was fashionable for ecologists to study communities in the arctic on the grounds that these would be very simple communities and hence easy to understand. Many excellent ecologists still follow this belief, but there are others who feel that it may be easier to understand the extremely complex communities of the tropics. This sounds paradoxical : How can a more complex communities by easier to understand ? A possible answer might be that the complex community has strong interactions among

species so that the lives of the separate species are less independent than in a simple community. Where there is greater interdependence, patterns may be more conspicuous.

Dans cet extrait, MacArthur suggère que la connectance du réseau (le nombre de liens entre espèces rapporté au nombre de liens possibles) est vraisemblablement une propriété importante pour comprendre la répartition des espèces. Peut-être qu'une biogéographie des réseaux serait une alternative porteuse de généralisations plus accessibles. Un nouveau problème d'échelle est soulevé, l'échelle biologique appropriée pour investiguer la répartition géographique des espèces : individus, population, communauté ou même réseau énergétique ?

Vers une biogéographie énergétique

Le problème d'échelle biologique est aussi un problème de catégorisation des espèces. J'ai suggéré que les prédictions étaient plus faciles pour des espèces généralistes que pour des espèces spécialistes. Malheureusement, le spectre est très large et plutôt balancé avec un continuum entre des espèces hyper-spécialistes de d'autres très généralistes (?). On peut néanmoins espérer que la réduction des espèces à un certain nombre de traits (??) doublée d'une réduction des réseaux à un nombre raisonnable de propriétés puissent permettre des généralisations utiles dans notre compréhension de la distribution des communautés. Il m'apparaît aujourd'hui important que le niveau de détails adéquate soit employé dans nos descriptions des systèmes écologiques quand il est question de prédire les futures aires de répartition des espèces.

Une piste prometteuse pour prolonger la recherche des propriétés est, me semble-t-il,

de s'appuyer sur la nature profonde des espèces : des systèmes énergétiques qui se perpétuent. La lecture de la théorie de la dynamique du budget énergétique de Sebastian A. L. M. Kooijman (?) m'a été très profitable pour cerner les possibilités offertes par une telle approche. S'il est possible, comme le suggèrent les travaux de Kooijman, de dériver de manière précise un grand nombre de propriétés énergétiques des espèces sur leur masse et leur forme, alors il semble raisonnable de nourrir l'espoir de pouvoir trouver des règles d'assemblages fiables des communautés et donc de comprendre d'un point de vue mécanistique les extinctions locales. Ces espoirs sont également attisés par la théorie métabolique de l'écologie qui rassemble des relations entre la taille des espèces, différentes de leurs propriétés (?) et qui offrent, en définitive, des pistes de recherche prometteuses pour aller au-delà de l'espèce (?). Mes réflexions sur l'intersection entre la TIB et une vision énergétique de l'écologie sont présentées au chapitre 4 de la présente thèse. Dans ce chapitre, j'explique en quoi l'approche énergétique est pertinente pour intégrer des interactions et les contraintes données par un flux énergétique limité. Je propose des pistes pour lever des difficultés posées par le calcul précis de la consommation de chaque niveau trophique et montre ce qu'apporte l'idée que les écosystèmes sont énergétiquement saturés. Ce chapitre est également une ouverture vers les projets de recherche que je souhaiterais mener dans un futur proche.

ARTICLE 1

À PROPOS DES INTERACTIONS BIOTIQUES ET DES CONTRAINTES ENVIRONNEMENTALES À L'ÉCHELLE BIOGÉOGRAPHIQUE

1.1 Résumé en français

1.1.1 Contexte scientifique

En 1967, Robert MacArthur et Edward O. Wilson publient leur théorie de la biogéographie des îles. Comme mentionner dans l'introduction, la TIB demeure le support de nombreux travaux de recherche, ce dont témoigne le livre publié en 2010 et édité par Jonathan B. Losos et Robert E. Ricklefs : *The Theory of Island Biogeography Revisited* (?). Ce livre souligne l'importance des travaux de MacArthur et Wilson et fait l'inventaire des questions qui restent à explorer. Parmi ces interrogations, figure celle relative au rôle des relations trophiques, développées au sixième chapitre par Robert D. Holt.

C'est précisément sur ce sujet que portent les travaux de Dominique Gravel et ses collègues présentés dans l'article *Trophic Theory of Island Biogeography* publié en 2011 dans *Ecology Letters* (?). Dans cet article, les auteurs montrent comment les résultats de la TIB 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 ses

collègues et 2- introduire les contraintes environnementales afin de comprendre dans quelle mesure les prédictions de la théorie classique en étaient affectées.

Pour y parvenir, la clef de mon travail a été de considérer les espèces non pas de manière indépendante, mais comme des assemblages, des communautés possibles. J'ai alors été capable de bâtir des probabilités de survie dépendantes de la configuration 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 comment le modèle a été construit et donné des prédictions simples, je me suis intéressé à des réseaux de 10 espèces présentant des types d'interactions différents : mutualisme, prédation et compétition, le long de gradients environnementaux. Ce qui ressort des simulations est un portrait des impacts potentiels des interactions sur la distribution des espèces et la nécessité de les intégrer dans les SDMs. Dépendamment de leur nature et de leur nombre, les interactions peuvent changer drastiquement la biodiversité attendue dans le cadre de la TIB. Cela pourrait avoir des conséquences majeures sur nos prévisions de richesse spécifique dans le contexte actuel des changements globaux. Le modèle développé dans ce chapitre est basé sur les chaînes de Markov, l'ensemble des aspects mathématiques de ce modèle sont présentés en détail à l'annexe 2. Certains des développements sont repris dans un article dont je suis coauteur et dont je présente ma contribution à cet article à l'annexe 3.

1.1.2 Publication associée

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 à l'automne 2015 dans le journal *Ecography* et a été publié en ligne en

décembre 2015. Le DOI associé est : **10.1111/ecog.01714**. La conception de l'article est le résultat de nombreux échanges entre l'ensemble des 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. Afin de promouvoir notre publication, j'ai également rédigé un article court sur le blog de la revue *Ecography* accessible en ligne ¹.

1.1.3 Traduction du résumé de l'article publié

La biogéographie est concernée en premier lieu par la répartition spatiale de la biodiversité dont elle produit les scénarios de variation dans un contexte de changement environnemental. Les efforts déployés pour développer les modèles de distribution d'espèces ont donné des outils qui, bien que prédictifs, sont restés corrélatifs et ont largement ignoré les interactions biotiques. Dans cet article, nous utilisons la théorie de la biogéographie des îles en tant que première approximation d'une dynamique d'assemblage des communautés locales dans un contexte de métacommunautés. Nous y superposons l'ensemble des types d'interactions et nous introduisons les contraintes environnementales sur la dynamique de colonisation et d'extinction. Nous développons une approche probabiliste reposant sur les chaînes de Markov et nous calculons les probabilités de réalisation des assemblages spécifiques plutôt que les probabilités de présence individuelle pour l'ensemble des espèces. Nous considérons la distribution de richesse spécifique pour les différents types d'interactions écologiques. De plus, nous illustrons également le potentiel de notre approche en étudiant l'articulation

1. <http://www.ecography.org/blog/towards-integrated-theory-biogeography>

entre les différences de besoins écologiques, les interactions et la distribution de la biodiversité le long d'un gradient environnemental. Notre approche supporte l'idée que les recherches futures en biogéographie requièrent une intégration cohérente de plusieurs concepts écologiques dans une théorie unique afin de mener vers des innovations méthodologiques et conceptuelles comme celui du changement des modèles de distributions espèce-centrés vers des modèles communauté-centrés.

Les sections suivantes sont celles de l'article publié.

1.2 Title

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

1.3 Authors

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

1.4 Abstract

Biogeography is primarily concerned with the spatial distribution of biodiversity, including performing scenarios in a changing environment. The efforts deployed to develop species distribution models have resulted in predictive tools, but have mostly

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

1.5 Introduction

Biogeography is concerned with the description of the distribution of biodiversity and understanding its underlying processes. The discipline is central to the simulation of future scenarios of biodiversity under climate change (?). The extensive development of statistical models of species distributions based on actual ranges and environmental data have provided valuable knowledge and predictions (?), but often remain purely correlative. There is now consensus that future developments in biogeography will require solving critical limitations of species distribution models (?) and incorporating explicitly biotic interactions and dispersal (?). This effort must be supported by theory in order to guide model development, maintain tractability and manage complexity.

Developing a mechanistic theory of species distribution will require an integration of three fundamental principles and their interplay (?): 1) how local and regional dynamics are linked, 2) how species interact with the abiotic environment and 3) how they are embedded in a network of biotic interactions. Each of these principles are discussed in detail below.

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

Species distribution is also constrained by physiological requirements, which is at the core of the niche concept (?). The niche is usually defined as a N-dimensional environmental and resource hyper-volume within which a species is able to maintain a viable population over the long term (?). Recent developments refined this definition based on demography and metapopulation dynamics (?). The abiotic niche, often referred as the Grinnelian niche, has been central to the development of species distribution models (SDMs, ?). Despite all of its criticisms, SDMs remain remarkably popular and oper-

ational 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 assess-

ment 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 assem-

blages and co-distribution at both the local and the regional scale. We illustrate how the interplay between biotic interactions and environmental requirements can affect the distribution of biodiversity over environmental gradients. Our results support the idea that the future research in biogeography require a consistent integration of several ecological concepts into a single framework to build promising approaches such as the switch from single-species distributions to community distributions.

1.6 The model

1.6.1 A simple probabilistic biogeographical model

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

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

Here, c_i and e_i are constant and a property of species i . In this widespread version of the TIB, also called the linear version of the TIB (?), the equilibrium occurrence probab-

ity of a species i is given by $p_{i,eq} = \frac{c_i}{e_i + c_i}$. Also, species are assumed to be independent, therefore, the richness S_{eq} is given by the sum of the P different $p_{i,eq}$. The linear TIB can be modified to include trophic interactions (after ?) and we propose to extend it to all types of interactions. To reach that goal, the first step is to find a way to derive the expected species composition at any time. This composition can actually be depicted at any time by a vector of P zeroes and ones indicating, respectively, presences and absences of each species considered, these combinations will be referred as assemblages. Following Mac-Arthur and Wilson, we use a stochastic modelling approach to describe the dynamics of assemblages. The simplest scenario is the one species case. Here there are only two assemblages for the locality: one with species i present and the other without. Let X_i be a random variable describing the occurrence of species i . When species i is present in the locality, X_i is 1, when it is absent X_i is 0; X_i is then a Bernoulli variable. We define this random variable at any time t which describes a stochastic process we denote $\mathbf{X}_{i,t>0}$. The occurrence probability of species i at time $t + dt$ (dt being a very small time step) is then given as follows:

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

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

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

This equation leads to (1.1) when dt tends to zero. This formulation keeps the simplicity of the original MacArthur and Wilson model, but can also more generally be used to consider the probability of any given assemblage. $\mathbb{P}(X_{i,t+dt}|X_{i,t})$ defines the rules to switch from one assemblage to one another during the interval dt . There are P occurrence probabilities we gather within $\mathbf{Y}_{t>0} = (\mathbf{X}_{1,t>0}, \mathbf{X}_{2,t>0}, \dots, \mathbf{X}_{P,t>0})$ which leads to the description of 2^P assemblages depicted by a given collection of zeros and ones. The conditional probabilities provide the transition from one local assemblage k to any other l during dt . For any species i there are only four possible cases: at time t either species i is locally absent and colonizes the locality (I_1) or not (I_2) during dt , either species i is present and goes extinct (I_3) or survives (I_4) during dt . The conditional probabilities between two communities states (l and k) can then be simply derived from these four probabilities:

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

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 1.1 presents these conditional probabilities (application of (1.4)). This is actually the transition matrix of a Markov chain we solve (by calculating one eigen value, see below). To illustrate the dynamics expected in TIB from our assemblage point of view, we simulate the model as follows: $c_1 = c_2 = 0.15$, $e_1 = e_2 = 0.05$, $\mathbb{P}(X_{1,0} = 0, X_{2,0} = 0) = 0.6$ and $\mathbb{P}(X_{1,0} = 1, X_{2,0} = 0) = 0.4$, so species 2 is absent at time $t = 0$. Just as for the single species situation, the probabilities of observing each community tend to an equilibrium (Fig.1.1, panel A). By summing the previous probabilities where a given species (1 or 2) is present (the conditional probabilities) we get its overall occurrence probability (marginal probability, Fig.1.1-B). Finally, we can calculate the expected number of species in a locality (Fig.1.1-C), in agreement with the TIB. Interestingly, this calculation is often achieved in the other way. Firstly, the presence probability of all species are computed: $\mathbb{P}(X_i) = \frac{c_i}{c_i + e_i}$. Then the richness is obtained under the assumption that species are independent and so $P(X_i, X_j) = \mathbb{P}(X_i)\mathbb{P}(X_j)$. We show below that occurrence probabilities of each assemblage is a key to introduce interactions among species.

1.6.2 Integrating biotic interactions

We start by representing the interaction network by a community matrix \mathbf{A} of P species that we incorporate into the Markovian TIB chain model. The elements $\alpha_{i,j}$ of \mathbf{A} quantify the effect of species j on the dynamics of species i . We first consider that interactions could alter both the colonization and the extinction probabilities (?). When $\alpha_{i,j}$ is negative, the colonization probability of species i decreases and/or its extinction probability increases when j is found locally. Inversely, when $\alpha_{i,j}$ is positive, the colonization probability increases and/or the extinction probability decreases. Note

that diagonal elements provide the extinction probability per time unit when no other species is present.

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

Table 1.1: 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.

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

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

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

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

To illustrate how interactions modify occurrence probabilities, we simulate the model for two networks: A_1 where all interactions are negative and A_2 where they are all positive. We consider null diagonal elements for both networks. Consequently, there is no difference with the model without interaction when one species is alone in the locality. Simulation results are presented at Figure 1.2. Panel A presents the functions f_1 and f_2 we chose for our two species example. For networks A_1 and A_2 , we show

how interactions alter the probabilities of observing different assemblages (respectively, Fig.1.2-B and Fig.1.2-C). The assemblage with both species present (solid red lines) is the most affected by interactions, switching from an occurrence probability of 0.2 (for negative interactions) to 0.8 (for positive interactions). Positive interactions enhance, as expected, co-occurrence while negative interactions prevent species from being found on the same island. Consequently, occurrence probabilities of single species states are lower in A_2 than in A_1 . According to a defined network, occurrence probabilities of the different assemblages are then modified, which affect the expected species richness (Fig.1.2-D).

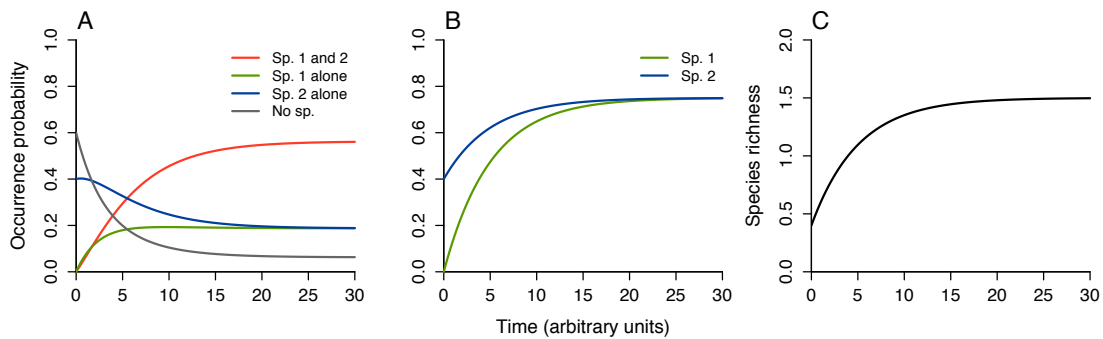


Figure 1.1: Dynamics of the community assembly. As a direct development of MacArthur and Wilson model, we simulate for two species, the dynamics of the four communities states with different initial conditions associated (A). By summing every states where one given species is present we get the occurrence 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.

1.6.3 Integrating environmental gradients

We now introduce the effect of abiotic conditions, such as climatic variables, on transition probabilities. We denote the vector of n environmental conditions by \mathbf{w} : $\mathbf{w} =$

(w_1, w_2, \dots, w_n) . We first assume that physiological constraints can affect both colonization and extinction probabilities through the functions f_i and g_i (affecting, respectively, extinction and colonization rates). Again the model in its general formulation does not presume any shape for these functions. We now have all the ingredients of an integrated model of biogeography as the transition probabilities at a location depend on 1) species-specific colonization and existence probabilities, 2) the network of interactions, 3) local community composition, and 4) local environmental conditions. In the general formulation of the model, functions f_i and g_i are functions of multiple variables (\mathbf{v} and \mathbf{w}).

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

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

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

the normalized left eigenvector associated to the first left eigenvalue.

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

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

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

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

This enables us to define an environmental optimum (h_i), a colonization probability per time unit (c_i) and also suitable range (r_i) for each species. Figure 1.3 presents the interplay between the three components of the integrated biogeographical model. The chosen functions for the environment-colonization relationship are illustrated in Panel A. For the two previous networks (A_1 and A_2 ; illustrated in Fig 1.2) we now compute the probabilities of observing the different assemblages at equilibrium, along the environmental gradient (Panel B and C). When interactions are negative (network A_1), species repulse each other and rarely co-occur, whatever the environment is. Most

of their occurrence follow their abiotic niche (blue and green lines) as they are barely found together. Inversely, when interactions are positive (for A_2 network) they often co-occur where their abiotic niches overlap, thereby decreasing the probability of an empty community (Panel D, solid grey line). Finally, we present how interactions modify the resulting community composition along the environmental gradient (Panel D). Species richness is constrained by the distribution of abiotic niches and the sign of the interactions. As expected, the role of interactions is strongest when abiotic niches largely overlap.

1.7 Exploring the model

In our exploration, we choose a regional pool P of 10 species to keep the number of assemblages reasonable ($2^{10} = 1024$) and to numerically compute the exact solution of the equilibrium distribution C_{eq} . We consider four types of interaction matrices A . The first situation corresponds to the classical MacArthur and Wilson model, where the 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 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 A positive to generate a mutualism network, or negative

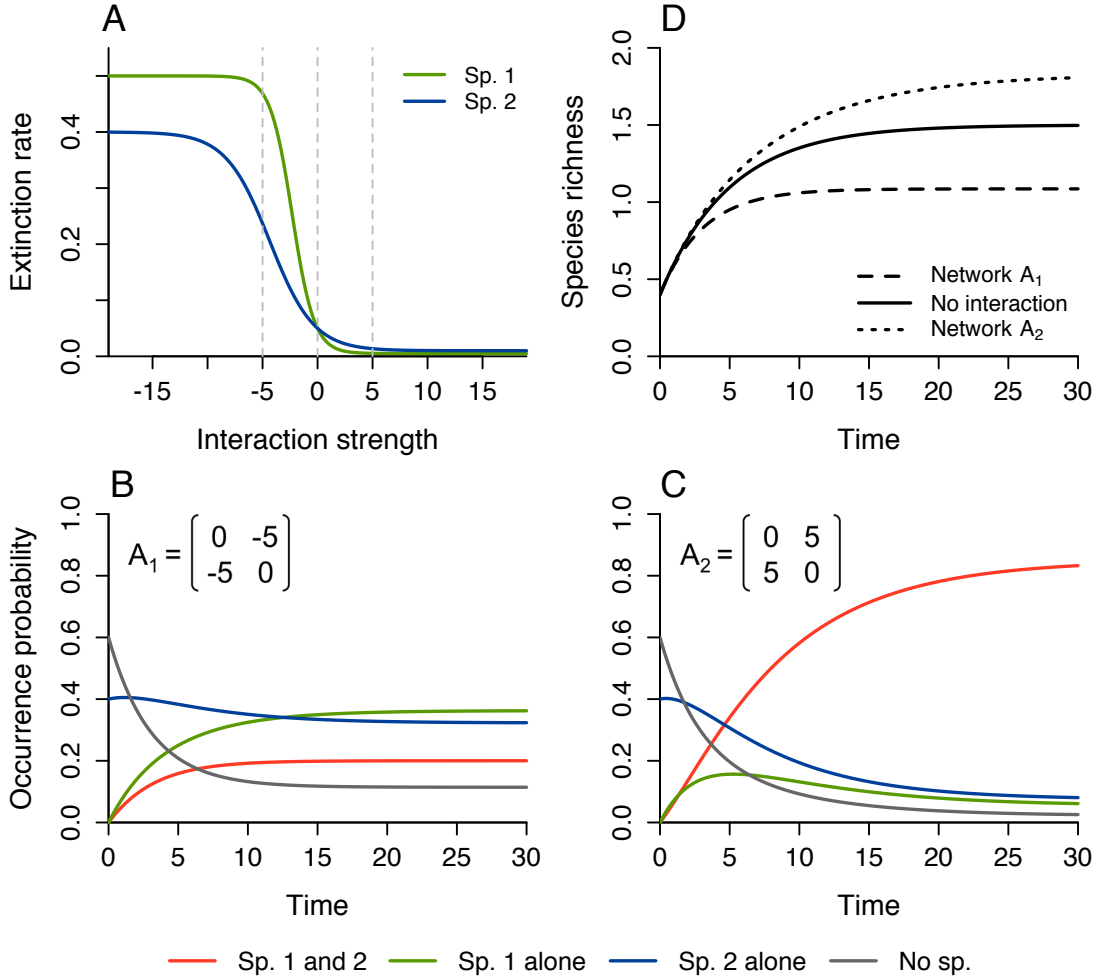


Figure 1.2: Effects of biotic interactions on colonization-extinction dynamics. For any species i , the extinction probability e_i is related to the strength of the interaction as shown in (A). The intersections of extinction curves with the grey dotted lines indicate the potential values of e_i according 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.

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 1.4 presents the results of such investigation, with Panels A to C corresponding to the results for the three different values of the d parameter.

As expected, positive interactions increase local species richness by diminishing extinction probabilities, while negative interactions weaken large communities (see the contrast between blue and red symbols on Fig.1.4). This is stressed as interaction strengths increase, that is for increasing values of d . Indeed, when d is low, there is almost no difference among scenarios because interactions do not impact strongly colonization and extinction dynamics; occurring species can be regarded as mostly independent. All scenarios converge to the classical TIB scenario (no-interaction, grey symbols), the

resulting species richness distribution is binomial (here for all species $p_{i,eq} = 0.5$ as $c_i = e_i = 10^{-5}$). Differences between interaction types increase with d . Species rich mutualistic communities are more likely to occur since positive interactions tends to promote co-occurrence. Therefore species occurrence can be dramatically affected by the strength of interactions: for $d = 10$ (Panel C in Fig. 1.4), the species richness is 9.46 for positive interactions (red symbols), 2.24 for the negative ones (blue symbols) and 5 without interactions. When positive and negative interactions are mixed (our predation scenario, green symbols on Fig.1.4), it seems that the negative effect of predators on their prey prevails and so predation reduces species richness, but less than for competitive networks.

As we introduce variability through the use of randomly-generated matrices, we also compute the standard deviation associated with occurrence probabilities. The variability is provided as the coloured vertical bars found in Fig. 1.4 which stand for 50% of the total standard deviation. Clearly, variability increases with (i) the strength of interaction and (ii) the occurrence probability. Although this can simply reflect the variability of values found in \mathbf{A} matrices, this could potentially be caused by the variability of the location of non-zero values in \mathbf{A} matrices; that is, the structure of the networks we use.

1.7.1 Biodiversity distribution over environmental gradients

In this section, we introduce an environmental gradient to emphasize the interplay between interactions species-specific requirements along an environmental gradient. Our environmental gradient takes values from 0 to 30, for each of them we calculate the expected species richness associated to all scenario. To do so, we start by computing the colonization functions (g_i functions): species optima h_i are drawn from a uniform dis-

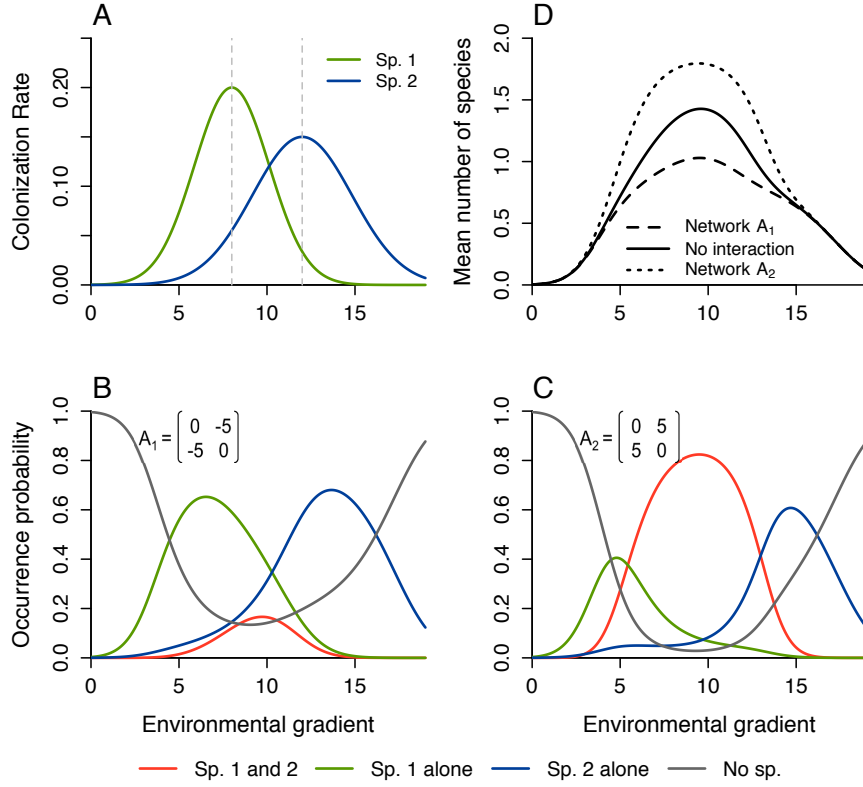


Figure 1.3: **Equilibrium for interacting species along an environmental gradient.** The colonization probability of species i , c_i , is related to the environment variable w according to species-specific requirements (A). The intersection of the colonization curve of species i with the grey dotted lines represents the value of c_i associated with its environmental optimum h_i . We compute equilibrium occurrence probabilities for the different assemblages along the environmental gradient, for the networks A_1 with negative interactions (B) and A_2 with positive one (C). We calculate the expected species richness on the locality for the two networks and without interaction (D).

tribution 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 \mathbf{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 1.5.

For all scenarios, the richness is maximal at the center of the environmental gradient (Fig. 1.5). This is due to the distribution of species optima in the range $[10, 20]$. Also this is the range of environmental values for which the effect of interaction are the most important. Indeed, the higher the colonization probabilities, the higher interactions occur, therefore, interactions strongly impact species richness for favourable abiotic conditions. We also find that changes in species richness increase with connectance, as depicted by the colour of the solid lines for the three panels of Fig. 1.5: from black (without interaction) to the lightest blue (connectance set to 0.4).

Species richness is inversely related to connectance when interactions are negative (Panel A in Fig. 1.5). Moreover, when abiotic conditions are favourable, the number of species expected tends to 1. At the centre of the gradient, even though colonization probabilities are maximal, many species colonize but likely go extinct because of competition. We expect the locality to be most often occupied by species that are not affected by competition. Alternatively, in the case of positive interactions (Panel B in Fig. 1.5), the expected species richness is strongly enhanced by interactions even for low connectance. The expected species richness tends to reach the total number of species from the most favourable to semi-harsh abiotic conditions. As the connectance increases the Gaussian shape of the richness profile turns into a hat shape, which has one major consequence: from favourable to semi-harsh conditions, the species richness is maintained thanks to positive interaction, but it also quickly collapses as the environment becomes slightly harsher.

Finally, when positive and negative interactions are mixed, the higher the connectance, the flatter the richness profile (Panel B in Fig. 1.5). The expected species richness declines as connectance increases but far less than it does for negative interactions only. We think this is caused by the colonization of numerous prey that promote the survival of predators which in turn prevent assemblages to be as large as they can be without interaction (as predators reduce the persistence of prey). Conversely, from harsh to intermediate environmental conditions, mixed sign interactions positively affect the species richness. We explain this as the consequence of the benefit predators take from the preys presence. Assemblages with few predators, promoted by positive effect of the prey on their predators, may be relatively stable. Since colonization is low, this assemblage may enhance species richness over time but they may also collapse as soon as an extra predator colonizes the island.

1.8 Discussion

Understanding how colonization-extinction dynamics influence species distribution and community structure remains a major challenge in biogeography (???). Here, we build upon the simplicity of the Theory of Island Biogeography (TIB) to integrate crucial ecological processes, namely biotic and abiotic dimensions of the niche. Using the formalism of Markov chains, we derive an exact general solution for the occurrence probabilities of all possible assemblages that we calculate numerically (up to 10 species). Our approach is in stark contrast to the classic TIB (?) where environmental gradients were not introduced and the co-occurrence among species was not modelled, despite empirical evidence of their impact (?). By taking these constraints together we reveal how they interplay and affect species richness. We believe our approach offers new perspectives on the theory of biogeography and will support the development of species

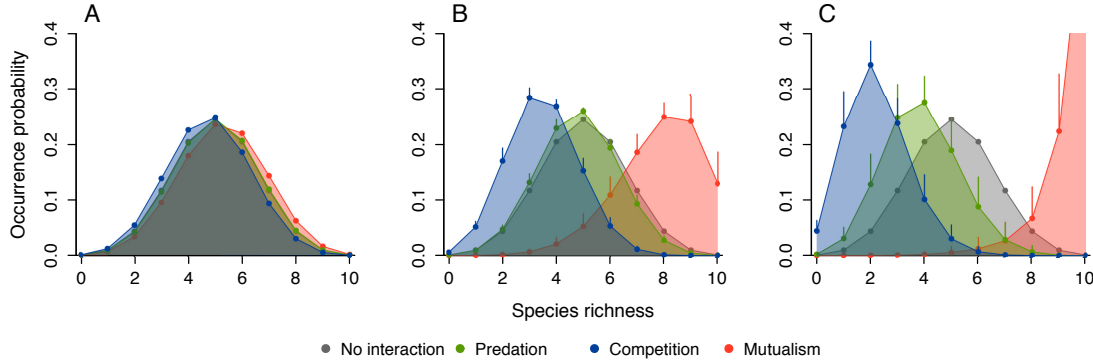


Figure 1.4: **Probabilities of species richness for different types of interaction.** We compute expected species richness at the equilibrium with the following set of parameters: $e_i = 10^{-5}$, $e_{i,min} = 10^{-3}e_i$, $e_{i,max} = 10^3e_i$ and $c_i = 10^{-5}$. We do so for three different interaction 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).

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.

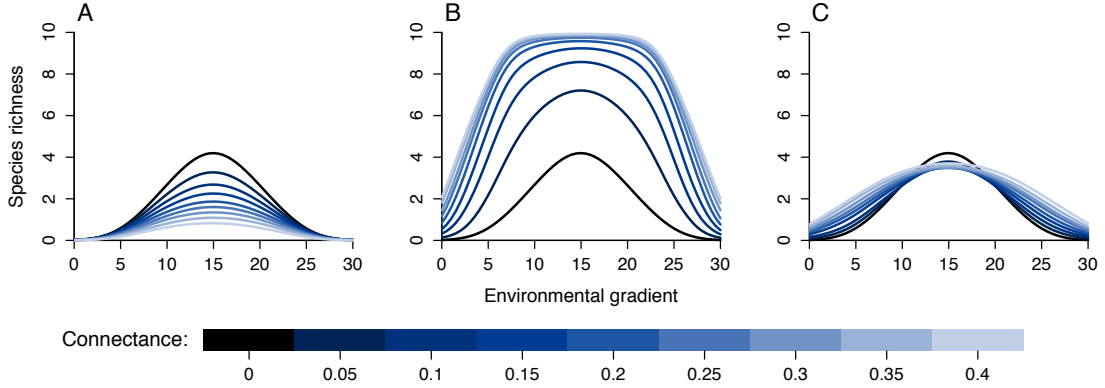


Figure 1.5: Biodiversity distribution along environmental and connectance gradients. We compute the expected species richness along an environmental gradient for competition (A), mutualism (B) and predation (C). We do so for different values of connectance depicted by the shades of blue. Species richness 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 $c_i = 10^{-5}$.

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 rationale that abiotic filters primarily determine the composition of assemblages (?). This argument persists even though theoretical (?) and empirical (?) evidence suggest the opposite. The key issue to solve this debate is thus to know how interactions can influence species assemblages with increasing spatial scale. Although the TIB

still provide insights into the assembly of natural communities, the success of recent approaches integrating interactions strongly support their relevance at large scales. Indeed the addition of network structure (?) or correlation between species (?) as proxies for interactions have adequately improved forecast accuracy. Here we do not solve this fundamental issue, however our model illustrates how species distribution at large scale will be impacted by the kind of interaction, their numbers and their distribution.

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

simple for applied situations. One solution may be to identify source and sink localities, (?). This requires us to consider i) species abundances and ii) spatial structures which would strongly increase the complexity of the model. One first step forward could be to apply the Levins model rather than the island-mainland model as Levins and Culver did to study the impact of competition (?).

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

Since the seminal work of Davis et al. (?), there is growing evidence that the response of species to climate change must be studied at the community scale (?). Even though species respond individually to climate change, they are constrained by complex direct

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

1.9 Acknowledgment

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

ARTICLE 2

UNE THÉORIE DE LA COOCCURRENCE DES ESPÈCES DANS LES RÉSEAUX D'INTERACTIONS

2.1 Résumé en français du deuxième article

2.1.1 Contexte scientifique

Au chapitre précédent, j'ai mis en évidence les impacts potentiels des interactions sur la présence locale des espèces dans le cadre de la TIB. Afin de faire un pas en avant vers le traitement des données typiques en biogéographie, les données de présence et d'absence des espèces, je me suis confronté aux problèmes de cooccurrence des espèces en interactions.

La cooccurrence observée dans les données est simplement le nombre total de sites où les espèces sont présentes ensemble rapporté au nombre total de sites étudiés. Pour pouvoir aller plus loin, avec mes collaborateurs, nous avons défini une mesure de cooccurrence sous l'hypothèse d'indépendance. Cela signifie que nous prenons l'occurrence respective des deux espèces et que nous les multiplions pour obtenir la mesure de cooccurrence sous cette hypothèse. Grâce à la comparaison entre ces deux valeurs nous avons été capables d'illustrer, dans l'article présenté dans ce chapitre, cinq grands principes relatifs à la cooccurrence des espèces en interaction :

1. Les interactions directes entre deux espèces affectent la cooccurrence. C'est une transposition directe des résultats du chapitre précédent sur la mesure de cooccurrence : s'il existe un lien entre deux espèces, leur probabilité d'être présentes simultanément dans une localité diffère de la probabilité attendue si elle se rencontraient aléatoirement.

2. Les interactions indirectes modifient la probabilité de cooccurrence. Malgré l'absence d'interaction directe entre deux espèces, ces dernières peuvent néanmoins être liées par une ou plusieurs autres espèces, l'interaction est alors qualifiée d'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 cooccurrence n'est pas symétrique. Il n'existe a priori aucune raison pour que ces effets soient symétriques. Néanmoins, en utilisant la mesure de cooccurrence 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 au sein du réseau, moins les conséquences des interactions indirectes seront perceptibles, nous montrons alors que les effets des interactions directes diminuent lors de leur propagation dans le réseau.

5. La force d'une association avec une autre espèce diminue avec le nombre d'interactions 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 cooccurrence sous hypothèse d'indépendance.

2.1.2 Publication associée

Pour ce second article, le contexte est particulier : Dominique Gravel a été invité à un numéro spécial de la revue scientifique *Theoretical Ecology* sur les réseaux écologiques. Dominique Gravel m'a alors proposé de travailler sur le prolongement de la réflexion menée au chapitre 1 et de l'appliquer sur les données de cooccurrence. J'ai alors conceptualisé un modèle probabiliste pour tenter de comprendre comment les interactions peuvent affecter la cooccurrence. Je me suis occupé de toute la partie modèle et des figures. Avec Dominique Gravel, nous avons rédigé l'essentiel du manuscrit avec des apports significatifs de Nicolas Mouquet et de Miguel B. Araújo qui sont co-auteurs. L'article a été publié en ligne en octobre 2015 avant d'être publié dans le numéro spécial (*species issue*) du volume 9 de l'année 2016 de *Theoretical Ecology*, le DOI associé est : **10.1007/s12080-015-0281-9 (?)**.

2.1.3 Traduction du résumé de l'article publié

L'étude de la cooccurrence des espèces a été essentielle en écologie des communautés depuis la fondation de la discipline. Cependant, les données de cooccurrence sont une source d'information négligée par les modèles de distribution d'espèces et les biogéographes se demandent encore quels sont les effets des interactions aux différentes échelles spatiales. Nous soutenons l'idée qu'une théorie de la cooccurrence des es-

pèces au sein d'un réseau écologique est nécessaire pour mieux interpréter les données de cooccurrence, pour formuler des hypothèses sur les mécanismes d'assemblage des communautés et pour étendre l'analyse des distributions d'espèces qui se concentre actuellement sur la relation entre les données d'occurrences et les facteurs abiotiques. L'objectif principal de cet article est de fournir les premières briques d'une théorie générale de la cooccurrence des espèces. Nous formalisons le problème avec différentes probabilités étudiées dans un contexte d'analyse de cooccurrence. Nous analysons les modules de trois espèces en interactions et nous conduisons des simulations pour documenter cinq principes influençant les associations au sein d'un réseau écologique : i) les interactions directes affectent la cooccurrence ii); les interactions indirectes affectent la cooccurrence; iii) la cooccurrence est rarement symétrique; iv) la force d'une association diminue avec la longueur du chemin le plus court entre deux espèces; v) la force d'une association diminue avec le nombre d'interactions qu'une espèce entretient. Nos analyses révèlent la difficulté d'interprétation des interactions entre espèces avec des données de cooccurrence. Nous discutons la possibilité d'inférence de la structure des réseaux d'interactions depuis ces données. Nous argumentons également que les modèles de distribution peuvent bénéficier de l'incorporation des probabilités conditionnelles en vue d'intégrer la contribution des interactions biotiques dans la distribution d'une seule espèce.

Les sections suivantes sont celles de l'article publié.

2.2 Titre

A Theory for species co-occurrence in interaction networks.

2.3 Auteurs

Kévin Cazelles, Miguel Araújo, Nicolas Mouquet et Dominique Gravel

2.4 Résumé en anglais

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

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

2.5 Mot-clef

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

2.6 Introduction

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

data.

Despite its historical importance for community ecology, co-occurrence data remain a neglected source of information in models of species distributions. Biogeographers are still debating the impact of biotic interactions on species distributions (????). The distribution of a species is thought to be first influenced by its physiological tolerance to environmental conditions, but also by interactions with other species (????). The question of whether such interactions leave imprints in the distributions of individual species at biogeographical scales is still open to debate (*e.g.* ?), but recent empirical ?, modeling (*e.g.* ?), and theoretical (?) evidence invites the interpretation that this might indeed be the case.

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

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

The analysis of species co-occurrences starts with a matrix representing the presence and absence of each species over a set of sites. There are two aspects to the quantitative study of co-occurrence. The first is the choice of the metric used to quantify the strength of associations (relationships between species occurrences) between pairs of species. The simplest measure of species co-occurrence is the number of species combinations, as defined by ?. A second index is the count of checkerboards ?: “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 (?). 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 (?) 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.*, ?). Thanks to the theoretical

work of ?, 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 (??); see also Table 2 in (?) for a description of 15 indices for co-occurrence analysis. A promising avenue is the one proposed by ? for the study of the matrix of species co-occurrence with tools borrowed from network theory.

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

The main objective of this paper is to provide the first building blocks of a general theory of species co-occurrences. We formalize the proposed theory with definitions of different quantities that are studied in the context of co-occurrence analyses. Herewith, we analyse three species interactions modules in order to document five principles influencing the association between pairs of species from an ecological network: i) direct interactions impact pairwise co-occurrence; ii) indirect interactions impact pairwise co-occurrence; iii) pairwise co-occurrence does not have to be symmetric; iv) the strength of an association decreases with the length of the shortest path between two species; v) strength of an association decreases with the number of interactions a species is experiencing. We base our mathematical argument on a general model of species distributions that is free of any assumption about how the ecological interactions operate. Finally we extend our analysis with simulations of multi-species networks in order to analyse how these mechanisms scale up in species rich communities.

2.7 Definitions

We start with definitions to formalize the quantities that can be computed from species distribution data and be used in the context of co-occurrence analyses. Let X_i be the random variable representing the presence of species i . $X_i = 1$ when species i is present, $X_i = 0$ otherwise. Then $X_{i,t>0}$ is the random process associated, giving the value that $X_{i,t}$ takes at any time t . Let $p_{i,t}$ standing for the probability $\mathbb{P}(X_{i,t} = 1)$. Also, to illustrate the definitions, we derive the quantities for a simple presence/absence dataset (see Table 2.1).

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 2.1 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 2.1, 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 $p_{i|\overline{K}} = \mathbb{P}(X_i = 1 | X_K = 0)$. The fundamental occurrence is conceptually equivalent to the fundamental niche of Hutchinson (1957) and represents the probability of observing a species in the absence of biotic interactions, i.e., when all other species are absent. By analogy, the marginal occurrence should be interpreted as the realized distribution. For species 1 in Table 2.1 we calculate $p_{1|\overline{23}} = \frac{p_{1,\overline{2},\overline{3}}}{p_{\overline{2},\overline{3}}} = \frac{0.2}{0.3} = 0.67$.

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

ergy fluxes, non-consumptive interactions, parasitism, direct interference, territoriality, space pre-emption, niche construction, etc. The impact of predator-prey interactions in a metapopulation setting with colonization and extinction dynamics will be considered for the multi-species simulations.

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

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

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

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

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

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

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

- i if species 1 cannot persist in absence of 2 (e.g., a parasite requiring its host), then $p_{1|\bar{2}} \rightarrow 0$, therefore $p_1 \rightarrow p_{1|2}p_2$
- ii if species 1 depends strongly on 2 thereby perfectly tracking its distribution 2, the $p_{1|\bar{2}} \rightarrow 0$ and $p_{1|2} \rightarrow 1$, and therefore $p_1 \rightarrow p_2$
- iii if species 2 excludes 1, then $p_{1|2} \rightarrow 0$ and $p_{2|1} \rightarrow 0$, 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 $p_2 = \frac{p_{2|\bar{1}} - p_{2|\bar{1}}p_{1|\bar{2}}}{1 - p_{2|\bar{1}}p_{1|\bar{2}}}$. 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:

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

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

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

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

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

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

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

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

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

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

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

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

Community modules

We now interpret these equations with examples of well-studied food web modules in community ecology: 1) linear food chain, 2) exploitative competition and 3) apparent competition. To do so, we consider matrices of direct associations representing the conditional co-occurrence probabilities among all pairs of species (see Table 2.2).

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

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

Table 2.1: Presence/absence dataset for three species and 10 sites.

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

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

Therefore the observed co-occurrence between species i and k depends on their respective interaction with species j ($p_{j|\bar{i}\bar{k}}, p_{j|\bar{i}\bar{k}}$ and $p_{j|\bar{i}\bar{k}}$). The conditional co-occurrence between two species could be null, but their observed co-occurrence be non-independent because of a shared interaction. This principle is best illustrated by the co-occurrence between a carnivore and a plant (species 3 and 1, respectively) in a linear food chain. In this situation, according to Table 2.2, we find that the observed co-occurrence between the plant and the carnivore is:

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

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

Similar observations could be made by studying the observed co-occurrence between

consumers (species 2 and 3) in the exploitative competition module:

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

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

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

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.

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

Table 2.2: 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 $p_{i|\bar{j}\bar{k}} = 0$, 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 : $p_{i|\bar{j}\bar{k}} = p_{i|\bar{j}k}$. For apparent competition, if species 1 and 2 are interchangeable for species 3 then : $p_{3|\bar{1}\bar{2}} = p_{3|\bar{1}2}$.

Multi-species simulations

Now we move to multi-species simulations of more complex networks to reveal the last two principles of our theory. To do so, we run simulations of the model of trophic island biogeography 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.

Distance decay of observed co-occurrence. The distribution of observed co-occurrence is illustrated for pairs of species separated by different path lengths at Fig. 2.1 A. 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 (?).

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. 2.1 B). We illustrated the relationship between the degree of a species and the observed co-occurrence for pairs of species with a direct association (Fig. 2.1 C). 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 $p_{i|j,\overline{K}}$). This is a $S \times S$ problem to solve and thus requires a significant amount of data. It might, however, be solved with large datasets where the number of sites N is much larger than S . There might also

be methods to reduce the dimension of the problem because usually only a small fraction of potential interactions are met in a network (corresponding to the connectance C). While a net interaction network is likely to be fully connected ($S \times S$ links), the direct interaction network has still only a fraction C of these links realized. Bayesian approach with latent variables could even further help reducing the dimension of the problem (*e.g.* ??). In such methods, latent variables are evaluated for each species to represent the underlying structure of the ecological network. It was found that between two and four parameters per species would be required to successfully represent more than 80% of interactions in a predator-prey network (?). This approach could, therefore, be used to represent the underlying structure of direct interactions and to evaluate numerically the non-null conditional probabilities. Note that these pairwise direct interactions should be interpreted specifically with reference to spatial dynamics because they would still represent phenomenologically the consequences of interactions, not the mechanisms of interactions. The next step in the development of a theory of species co-occurrence (and of species distribution) is the addition of environmental constraints. Our approach assumed a homogeneous environment, mainly for tractability of equations. We acknowledge that non-independent co-occurrence could also arise because of shared environmental requirements. The addition of environmental constraints would be easy to implement in our framework by simply making the conditional probability in absence in absence of interactions a function of the environment. Every quantity we derive after would be conditional on the environment. What would be more challenging but, nonetheless, feasible numerically, would be to make the direct interaction itself a function of the environment. There is now growing evidence that ecological interactions are context dependent (??). We view this integration as the next step to the derivation of a theory-driven species distribution model taking into account biotic interactions (?).

2.8 Acknowledgments

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

./chapitre2/fig1-eps-converted-to.pdf

Figure 2.1: Co-occurrence in multi-species networks. (A) The disparity between observed co-occurrence ($P_{i,j}$) and independent co-occurrence ($P_{i,j;\text{IND}}$) 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 ($\log(P_{i,j}/P_{i,j;\text{IND}})$) 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 ($sd(P_{i,j}/P_{i,j;\text{IND}})$) and thus the variance decreases with species richness. Taken together, (B) and (C) imply that species distributions converge to independence with increasing species richness.

ARTICLE 3

QUELLES TRACES LES INTERACTIONS BIOTIQUES LAISSENT-ELLES DANS LES DONNÉES DE COOCCURRENCE DES ESPÈCES?

3.1 Résumé en français du troisième article

Dans ce troisième article, je me suis intéressé aux données de cooccurrence. Ces données, très importantes en biogéographie, sont des enregistrements de présences et d'absences pour un certain nombre d'espèces le long d'un gradient géographique (par exemple, un gradient latitudinal). On dispose alors d'un grand nombre de sites sur lesquels les conditions abiotiques varient et pour lesquels nous disposons également de la connaissance d'une partie du contexte biotique (les présences de différentes espèces). En plus de ces informations, les jeux de données présentés sont munis d'observations relatives aux informations réelles avec lesquelles il est possible de construire des réseaux d'interaction. J'ai donc analysé la cooccurrence à la lumière des propriétés des réseaux trophiques, mutualistes ou de compétition avec des espèces plus ou moins généralistes ce qui m'a permis de tester les hypothèses faites au chapitre 2.

Nos résultats montrent que les interactions ont un effet sur la cooccurrence mais que la détection de cet effet est délicate quand 1) les espèces sont éloignées dans le réseaux, les interactions indirectes sont donc difficiles à mettre en évidence et 2) quand les interactions sont nombreuses, il est plus difficile de trouver des signaux de cooccurrence ¹

1. Il s'agit de déviations par rapport à des attendus sous l'hypothèse que les interactions n'ont pas

pour les espèces généralistes. De plus, en intégrant la similarité des facteurs abiotiques pour les différents sites, je montre que les signaux de cooccurrence s'affaiblissent et parfois disparaissent. Mes résultats suggèrent donc qu'en utilisant des facteurs abiotiques pour inférer les probabilités de cooccurrence, une partie du lien entre les espèces est capturée, mais cette part est entachée d'une grande incertitude. Ceci vient questionner la qualité des prédictions données par les modèles classiques de distribution d'espèces actuellement utilisés. Nos résultats propose un nouvel éclairage sur un débat classique de l'écologie concernant la détection des interactions à partir des aires de distribution. Nous montrons que la configuration du réseau est aussi importante et qu'il ne faut pas trop rapidement conclure que les espèces sont indépendantes, ce qui est souvent une hypothèse de départ dans les projections que nous faisons pour anticiper les conséquences des activités humaines sur les changements de biodiversité.

3.1.1 Publication envisagée

Le travail ici présenté a été rendu possible grâce à une collaboration entre différents chercheurs menée avec Dominique Gravel. Les jeux de données que j'ai analysé sont, en effet, difficiles à trouver et particulièrement intéressantes. Je remercie tous les chercheurs qui m'ont accordé leur confiance pour analyser les données et toutes les personnes qui ont contribué à la collecte des données. Pour les données sur les colibris et les plantes des Caraïbes, je remercie Bo Dalsgaard, Louise J. Lehmann, Ana M. Martín González, Andrea Baquero, Allan Timmermann. Pour le jeu de données relatif aux communautés associées aux feuilles de saules, je remercie Tomas Roslin. Pour les données sur les communautés microbiennes des Sarracénies (*Sarracenia purpurea*), je

remercie Benjamin Baiser. Pour les données du Suivi Temporel des Oiseaux Communs (STOC), je remercie Wilfried Thuiller et pour les données des arbres, je remercie Dominique Gravel et Steve Vissault.

L'article qui suit est le fruit de discussions avec Dominique Gravel sur le type d'analyse à mener pour donner suite au chapitre 2. Je me suis occupé de l'ensemble des analyses, des figures, de la rédaction de la première version relue par Dominique Gravel. J'ai également bénéficié d'une précieuse et attentive relecture de David Beauchesne que je remerci chaleureusement. Due à la qualité des jeux de données, à l'originalité de nos analyses et nos résultats, je nourris l'espoir que mes résultats intéresseront une revue généraliste. C'est pour cela que le chapitre est dans un format court avec une longue section "Supplementary information" (SI, 3.8). Se confronter à la rédaction en format court fut aussi un exercice très enrichissant bien que délicat.

3.1.2 Traduction du résumé en anglais

Un problème majeur en biogéographie et en modélisation de la biodiversité est de savoir si la cooccurrence des espèces contient une trace des interactions. L'hypothèse conventionnelle affirme que les interactions négatives et positives conduiraient respectivement à des répulsions et des attractions. La rareté des jeux de données combinant des distributions à larges échelles avec des données d'interactions a longtemps empêché les biogéographes d'obtenir des idées claires sur ce problème. Pour répondre à cette question, nous avons utilisé cinq jeux de données sur un large gradient de conditions climatiques incluant des informations sur les interactions écologiques, représentant un total de 793 espèces pour 354015 observations de cooccurrence. Nous avons comparé la cooccurrence des paires d'espèces interagissant avec celle de paire d'espèces

n'interagissant pas en intégrant ou non les variables abiotiques. Nous avons trouvé un effet des interactions sur la distribution des espèces de cooccurrence brute mais aucune preuve d'un tel effet lorsque les variables abiotiques étaient intégrées. De plus, nous avons trouvé que certaines propriétés du réseau d'interactions pouvaient masquer l'impact des interactions sur la cooccurrence. Nous avons ainsi montré que plus le nombre d'interactions qu'une espèce entretenait était grand, plus notre capacité à détecter le signal dans les données d'interactions statiques était faible. De plus, nous démontrons clairement que le signal de cooccurrence entre un prédateur et ses proies disparaît quand la proportion de sites couverts par l'ensemble de proies augmente. Dans un contexte où les écosystèmes sont fortement perturbés par l'activité humaine, nos résultats insistent sur le besoin d'intégrer les processus écologiques dans les modes de distribution d'espèces pour mieux prédire la biodiversité de demain.

Les sections qui suivent sont celles de l'article envisagé.

3.2 Title

Do interacting species co-occur differently from not-interacting species?

3.3 Abstract

A major problem in biogeography and biodiversity modeling is whether species co-occurrence holds an imprint of ecological interactions. The conventional hypothesis poses that negative and positive interactions should yield repulsion and attraction, respectively. The scarcity of datasets combining broad scale distributions with observed interactions has long prevented biogeographers from obtaining clear insights into this problem. To address this question, we compile five datasets of species distribution over a large range of climatic conditions, along with their ecological interactions, representing of a total 793 species and 354015 occurrence observations. We compared co-occurrence of interacting versus non-interacting pairs of species accounting or not for abiotic variables. We found an effect of interactions on species distribution in raw co-occurrence data but no clear evidence of such effect once when accounted for abiotic factors. Moreover we found that some properties of the network of interactions could hide the impact of interactions on co-occurrence. We showed that the larger the number of interactions a species is experiencing, the weaker our ability to detect signals in static co-occurrence data. Also, we clearly demonstrate that the signal of co-occurrence between a predator and its preys vanishes when the proportion of the sites covered by the set of preys increases. In a context where ecosystems are dramatically disturbed by human activities, our results emphasize the need for integrating ecological processes into distribution models to better predict tomorrow's biodiversity.

3.4 Introduction

Biogeographers are still debating about whether biotic interactions at the local scale impact large scale species distribution. If this is indeed the case, then we should expect pairs of species to exhibit non-random distribution at the large spatial scale. The observation of independent species occurrences would therefore support the use of classical species distribution models (hereinafter SDMs, ?) and confirm scenarios that have been proposed over the last decade (???). On the other hand, the observation of occurrence clearly related to interactions would give credit to methods interpreting co-occurrence as a proxy for ecological interactions (?) and encourage the use of joint species distribution models (hereafter JSDM, ??). In order to test whether interactions influence species distributions, the simplest avenue is to investigate species co-distribution in light of their ecological relationships. Such investigations started with Diamond's original study stating that species interacting by competition should avoid each other in space, leading to a 'checkerboard' distribution (?). This idea was rapidly criticized for the lack of an adequate null hypothesis (??) and has led to a major controversy that has lasted more than 30 years (?) and has underlined the need for null models in ecology (??). Furthermore, the absence of adequate datasets has long prevented this hypothesis to be test appropriately.

Species ranges are very often inferred from the realized distribution, which results from the joint impacts of abiotic and biotic environmental factors along with dispersal limitations and historical contingencies (????). Finding evidence of the effect of species interactions on distribution may prove technically difficult as it needs to discriminate the fundamental niche from co-occurrence data (Box 3.6.2), which could explain the scarcity of studies reporting such effect (but see ?). Fortunately, recent developments in co-occurrence theory promises novel avenues in which highly dimensional species

distribution data could be analyzed, accounting for the structure of the network of ecological interactions (?). It has notably been proposed that the higher the degree of a species, *i.e* the number of species with which it interacts, the weaker will be the detection of a co-occurrence signal. The geographical range of a species is indeed partially correlated with all the ranges of species it is linked to, which should blur the signal the interactions leave on pairwise co-occurrence. However, a signal may persist if the co-occurrence of the focal species with the joint distribution of the species it interacts with is considered.

Also, two co-occurring species may be part of the same ecological network but separated by several interactions. For such case, it has been suggested that the longer the shortest-path, *i.e* the number of links between two species, the smaller the signal of co-occurrence. Hence, based on co-occurrence data and information pertaining to ecological network, the hypotheses recently formulated on the signal of co-occurrence can be tested (Box 3.6.2).

Despite the fantastic number of studies investigating co-occurrence in order to understand community assembly, the hypothesis that species interactions should result in non-random associations in space has never been tested formally. It is rather implicitly assumed in every investigation. Our objective in this study is therefore to test this at the scale of species range. We analyzed species association within five different datasets ranging from plant-seed dispersers, host-parasitoids, microbes, birds and forest trees. We quantify co-occurrence and conditional co-occurrence accounting for abiotic constraints for species that are known to interact and known to not interact directly. This is a unique opportunity to test the hypotheses recently developed. We report some signal of positive co-occurrence among the datasets, suggesting a difference between pairs of interacting and pairs of not-interacting species. We also point out that the degree of a species in a network influences our ability to detect significant associations. Moreover,

we discover a clear relationship between the strength of species associations and the cumulated occupancy of the entire set of species with which a species interact. However, all the signal we detected are weakened or even disappear once abiotic similarities among site are integrated. We further discuss the integration of biotic interactions in distribution models in light of our results.

3.5 Material and Methods

3.5.1 Datasets

We analyzed five datasets spawning a large diversity of organisms and types of interactions (see Table 3.1), across an extensive gradient of environmental conditions (see Fig. 3.5 and SI 3.8). Our criteria for selecting datasets were as follow: i) the occurrence data must have been collected at the local scale, it could not be derived from range maps; ii) data included observed absences; iii) occurrence was observed for all species: interactions were known a priori; v) since we are computing co-occurrences, which is typically a fairly rare event, we needed an elevated number of sampling units. Direct interactions were documented for four datasets and represented by networks of ecological interactions: the Willow Leaf Network (WLN), the Pitcher Plants Network (PPN), the Caribbean Hummingbirds Network (CHN), the French Breeding Birds Survey (FBBS). We derived metawebs for each network, *i.e.*, the matrix recording all interactions, and computed three metrics among species from a regional pool: the connectance of metawebs (proportion of realized interactions), the degree of species (number of links for one species) and the shortest-path between all pairs of species (the minimum number of link from one species to the other, see SI 3.8). For the North American Trees (NAT) dataset, we approximated competitive interactions by comput-

ing functional distance between every pair of species (see Table 3.2 and Fig. 3.6). A similar proxy for ecological interactions was also computed for the FBBS (see SI 3.8). Only species that were present at least on 1% of the total number of observations were considered for the analysis (see SI 3.8).

3.5.2 Quantifying co-occurrence

For each pair of species i and j , we computed the number of observed co-occurrence $O_{i,j}$ and the expected number of co-occurrence under independent distribution $E_{i,j}$ and the standard deviation associated $SD_{i,j}$. We report the Z-score $\frac{O_{i,j} - E_{i,j}}{SD_{i,j}}$ (?), a standardized metric of co-occurrence, with positive (negative) values indicating more (less) co-occurrence relative to the random expectation. Expectations were derived using three different methods. First, we assumed that all sites were equivalent, meaning we ignored the potential influence of abiotic conditions across the entire range. The observed co-occurrence is an estimate computed from a limited number of observations (??) and therefore, we also considered an hypergeometric distribution, which corrects for limited sample size (see SI 3.8 for further details). We used two different SDMs in order to account for the abiotic environment and correct the random expectations, namely, Generalized Linear Model (hereafter GLM) and Random Forest (hereafter RF - see SI 3.8 for more details and Fig. 3.7 for the assessment of models' performances). The correction using SDMs accounts for the fact that species may co-occur simply because they have similar abiotic requirements, irrespective of if they interact or not.

3.6 Results

Without accounting the abiotic context, we observe a difference in the strength of spatial associations between interacting and not-interacting species (Fig. 3.2 panels A to D, white boxplots) for two out of four datasets for which direct interactions were known. In all cases, spatial associations were positive, indicating that interacting species are co-occurring more often than randomly expected, irrespective of the type of interaction. For the WLN, distinguishing herbivore-willow interactions from herbivore-parasitoids revealed that the strength of co-occurrence was greater for the former than the latter (Fig. 3.3 A-B). Interestingly, we noticed that detection of a signal of interactions in co-occurrence became more difficult as the mean degree of species in the dataset increased (Fig. 3.3 A-D). Therefore, based on co-occurrence data only, we conclude that interactions affect co-occurrence, but that detection trend becomes trickier in highly connected network.

We observed that the strength of spatial association was higher for pairs of similar species (Fig. 3.2 panels E and F) for the NAT and FBBS datasets, for which we inferred a distance based on functional traits. Functional similarity is often interpreted as a proxy for competition strength (?), which according to Diamond's hypothesis, should provide a negative association. Interestingly, we observe the opposite trend. This suggests that local competition is poorly detectable at large scale, which is theoretically supported (?), and the positive associations are likely the results of similar physiological requirements that cannot be taken into account given the sets of abiotic variables we used. This is further supported by the regressions of the Z-score against the distance (see Fig. 3.10 A-D). We also note that results for the FBBS dataset were identical irrespective the type of traits examined (Fig. 3.8). Hence, the detection of interactions from co-occurrence data varies according to the nature of the ecological relationships.

When we account for suitability discrepancies among sites using SDMs, the co-occurrence signals we previously observed are weakened or disappear. Co-occurrence distributions are shifted toward 0 and dramatically shrunk (Fig. 3.2 panels A to D, grey and dark grey boxplots). The tendencies are stronger for the RF approach than for the GLM one. The same trends are observed when we used functional distance rather than interactions (Fig. 3.2 panels E-F).

We investigated if the strength of associations between species interacting indirectly decreases with the shortest path (Fig. 3.3). We examined the co-occurrence distribution against the shortest-path for pairs of: herbivores and willows in WLN (A), herbivores and parasitoids in WLN (B), hummingbirds and hosts plants (C) and all pairs in PPN (D). Without accounting for the climatic factors, we observe a decrease of the standardized co-occurrence with an increased shortest-path for three out of four occurrence datasets (Fig. 3.3 A-D, white boxplots). This was predicted by the theory, but the observed decay is steeper than anticipated (?). We observed a similar but weaker decay when all pairs of species were taken into account (see Fig. 3.9). Moreover, when we calculated the mean degree of predators (pollinators), we also found the signal of co-occurrence stronger for specialists (Fig. 3.3 A) than for generalists (Fig. 3.3 C), suggesting that the abundance of interactions may prevent us from detecting them in static occurrence data. Again, the tendencies observed are strongly weakened when occurrence probability are derived from SDMs (Fig. 3.3 A-D grey and darker grey boxplots).

Finally, for all predators and pollinators in WLN, CHN and PPN, we average the standardized co-occurrence over the set of their respective preys (host plants) and plotted the obtained values against the total number of site covered by the same set of preys (hosts) they refer to (Fig. 3.4). When abiotic context is not included, we observe a

clear negative relationship between the two quantities for three out of four datasets. The associated linear regressions explain up to 69% of the variance (Fig. 3.4 panels A, D, G and J). However, the linear regression poorly performed for the PPN, in which the connectance is the highest. Additionally, we show that the linear regressions using total number of site covered by the set of preys (host plants) outperformed the ones using the degree of the species (Fig. 3.11 panels A, D, G and J) that was envisioned by the theory (?). Moreover, the relation is asymmetric: the decay is less convincing when the mean Z-scores of the preys are plotted against the cumulated range of their predators (Fig. 3.12 panels A, D, G and J). These results suggest that for a predator (pollinator) feeding upon a set of preys (hosts), the detection of interactions is possible if the part of the geographic range studied covered by the preys is low, when the part increases our ability to detect the signal fades away. Once again, the results are dramatically different once abiotic constraints were taken into account: the clear relationship obtained is either weakened or even reversed (Fig. 3.4 B, C, E, F, H, I, K, L). Even the strongest associations seem to be captured by the SDMs, meaning they are captured by the link with this environment but this questions how the inference is made when the presence of one species is best explained by the presence of its prey. Indeed, we show the presence of the whole set of prey as predictor to assign the presence of specialist predator outperform GLMs but not RFs (see Fig. 3.13).

3.6.1 Discussion

The historical debate on co-occurrence triggered by Diamond's original work was based on the idea that competitors must repulse each other (?). Interestingly, we were not able to detect any negative signal for species we assumed to compete. This support the idea that local repulsion that must occur does not leave any imprint on co-

occurrence data over a broad geographical gradient (?). However, we demonstrate that other signals, not mentioned in the controversy, can actually be revealed. Mutualism and predation were indeed detected as positive signals in the standardized co-occurrence. Therefore, the interdependency between a predator (pollinator) and its preys (host plant) increases co-occurrence over random expectations and this effect is appreciable at large scale (?). Moreover, we find that the signal of co-occurrence is blurred by the number of interactions a generalist species experience, in agreement with recent co-occurrence theory (?). As a consequence, the distribution of a given species is more strongly related to the distribution of the entire set of species it interacts with, than each species taken individually. This indicates that the role of ecological interactions may not only be a matter of spatial scale (?), but also a matter of picking the most consistent biological unit to investigate distribution over large spatial scales.

The signals of co-occurrence were dramatically reduced once we accounted for the abiotic context. This could lead us to conclude that co-occurrences are likely driven by the abiotic differences among sites. Climatic data unsurprisingly explains a large part of the co-occurrence observed. However, our findings suggest a potential methodological issue. Current SDMs always use abiotic constraints as the main drivers of geographic distributions. In doing so, they capture, at least partially, the impact of biotic interactions on distribution (as it infers the distribution from the realized distribution). A predator is therefore linked to its preys through the abiotic similarities of the sites where they are found. However, the strength of an association may not be perfectly captured by SDMs (*i.e.* abiotic factors only). This would explain why the co-occurrence signal vanishes in Fig. 3.4 when SDMs were used even though the standardized co-occurrences remain positive.

SDMs are based on the assumption that species are distributed independently from

each other and that scenarios of tomorrow's biodiversity can be anticipated using abiotic variables only (?). Some recent developments in statistics relaxed this hypothesis by representing the distribution of assemblages, based on correlations between species (??). The development of these powerful statistical tools is essential, however it can only provide a partial solution to some limitations of SDMs.

A substantial part of the solution lies in the understanding of the role of ecological and evolutionary mechanisms shaping distributions (?). Here, we underline the need for not hypothesizing *a priori* the independence among species. Rather, the assumption must be proved to apply, otherwise a relevant assemblage must be modeled. We suggest that for generalist species, the assumption of independence is reasonable, while for specialists the relative position with the other species within the network should help in deciding which set of species are to be modeled. We therefore suggest that ecological interactions should be integrated as prior information in SDMs to strengthen our predictions. For instance, the range of a predator is bounded by the range of the set of its preys (??), this reality should be taken into account when we predict the distribution of food webs, and predators' ranges must be computed consistently with the predicted ranges of their prey.

Our findings lead us to conclude increased interactions diversity causes species to co-occur independently. Only the specialized species are strongly associated with their partners, as generalists experience a much more diffuse set of constraints from interacting species. Similarly, it has been recently highlighted that net species interaction strength between pairs of species is better predicted when the species richness increases (?). Further, press perturbations applied to specialist species do have much stronger indirect impacts on the other species than perturbations applied to generalists (?). All of these results are congruent with MacArthur's vision, who more than 40 years ago noticed the following paradox "*How can a more complex community be easier to un-*

derstand? A possible answer might be that the complex community have strong interactions among species so that the lives of the separate species are less independent than in a simple community. Where there is greater interdependence, patterns may be more conspicuous.” (?, p.199). In the present study, the more the interactions among species the easier the forecasting of species distribution. However, under the ongoing mass extinction, a myriad of links vanishes while new ones emerge with the changes in the composition of local communities. Therefore, even if under current conditions the assumption of independence may be valid, under dramatic modifications of ecosystem as currently observed, it may often prove false. As a step forward, new biodiversity scenarios must not solely map the future ranges of individual species but the entire community including the consequences of potential extinction on community structure.

3.6.2 Box 1

The fundamental niche is here defined as the occurrence probability under the assumptions that (1) biotic factors are not limiting the occupancy and (2) the occupancy is estimated in the absence of immigration (?). In this case, only abiotic factors (such as water availability, temperature variability or edaphic variables) limit survival and/or reproduction success, and then the occurrence probability. For a four species network made of three trophic levels: two primary producers, a predator feeding upon them and a top predator, we first consider the fundamental niches f_i (Fig. 3.1 A). For the top predator, we derive the fundamental niche under the assumption of the presence of its prey:

$$f_4(w) = P(X_4 = 1 | X_3 = 1, G = w)$$

where G denotes the environmental gradient and X_i represents the random variable of the presence of species i . For the lower predator, we consider its preys substitutable and we assume the presence of at least one prey to allow the predator to be present:

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

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 dynamics of studied species. The realized niche (Fig. 3.1 B) includes these factors.

In our simplified example, fundamental and realized niches of preys are identical. The realized niche of the top predator is constrained by the one of predator 3, r_3 , which it-self is constrained by the joint realized niches of its preys:

$$r_3(w) = P(X_3 = 1 | X_2 + X_1 > 0, G = w) P(X_2 + X_1 > 0, G = w)$$

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

Then, assuming that species 4 do not alter r_3 , we have:

$$r_4(w) = f_4(w) r_3(w)$$

$$r_4(w) = f_4(w) f_3(w) (1 - (1 - r_1(w))(1 - r_2(w)))$$

The above expressions may often be more complicated due to immigration fluxes as

well as the size and the structure of the interaction network. For instance, we do not consider the apparent competition between species 1 and 2, although it must affect their co-distribution. Integrating the impact of many interactions may be possible using occurrence probabilities of species assemblages rather than single species (?). Here, the realized niches of predators 3 and 4 depends on the fundamental niches of preys 1 and 2.

The link between the fundamental niche and the distribution goes through different ecological processes such as immigration and ecological interaction (??). The distribution we infer is based on realized niche that includes the effect of abiotic variables and biotic interactions. If we now examine the co-occurrence signals, *i.e.* the difference between the expected co-occurrence and the observed co-occurrence (?), in empirical data, in order to ascribe a particular signal to an ecological interaction, five difficulties emerge:

1. the signal of co-occurrence is not constant along the environmental gradient (Fig. 3.1 C-a),
2. for species experiencing several interactions, the association with a particular species may be detectable only for a portion of the environmental gradient (3.1 C-b)
3. for species experiencing several interactions, the signal of co-occurrence with the set of species it interacts with might be more informative (Fig. 3.1 C-c),
4. for species experiencing several interactions, the increase of interaction weakens the signal (this explain the drop for medium environmental values in Fig. 3.1 C-c),
5. when the shortest-path between two species increase, the signal decreases (the red lines in Fig. 3.1 C-c) is below the grey one).

Using datasets including occurrence data for a large number of species together with information pertaining to the ecological relationships, we propose to test these theoretical expectations.

3.6.3 Tables

Table 3.1: Data sets analyzed in this article.

Type	No. of sites	No. of species	Interaction type	Connectance	References
Willow Leaf Network	374	156	Trophic / Parasitism	0.042	unpublished
Pitcher Plants Network	39x20	53	Trophic	0.44	?
Caribbean Hummingbirds Network	32	62	Mutualism	0.011	?, ?, ?
North American Trees	128891	31	Competition	-	unpublished
French Breeding Birds Survey	2354	179	Trophic	0.018	?

3.7 Figures

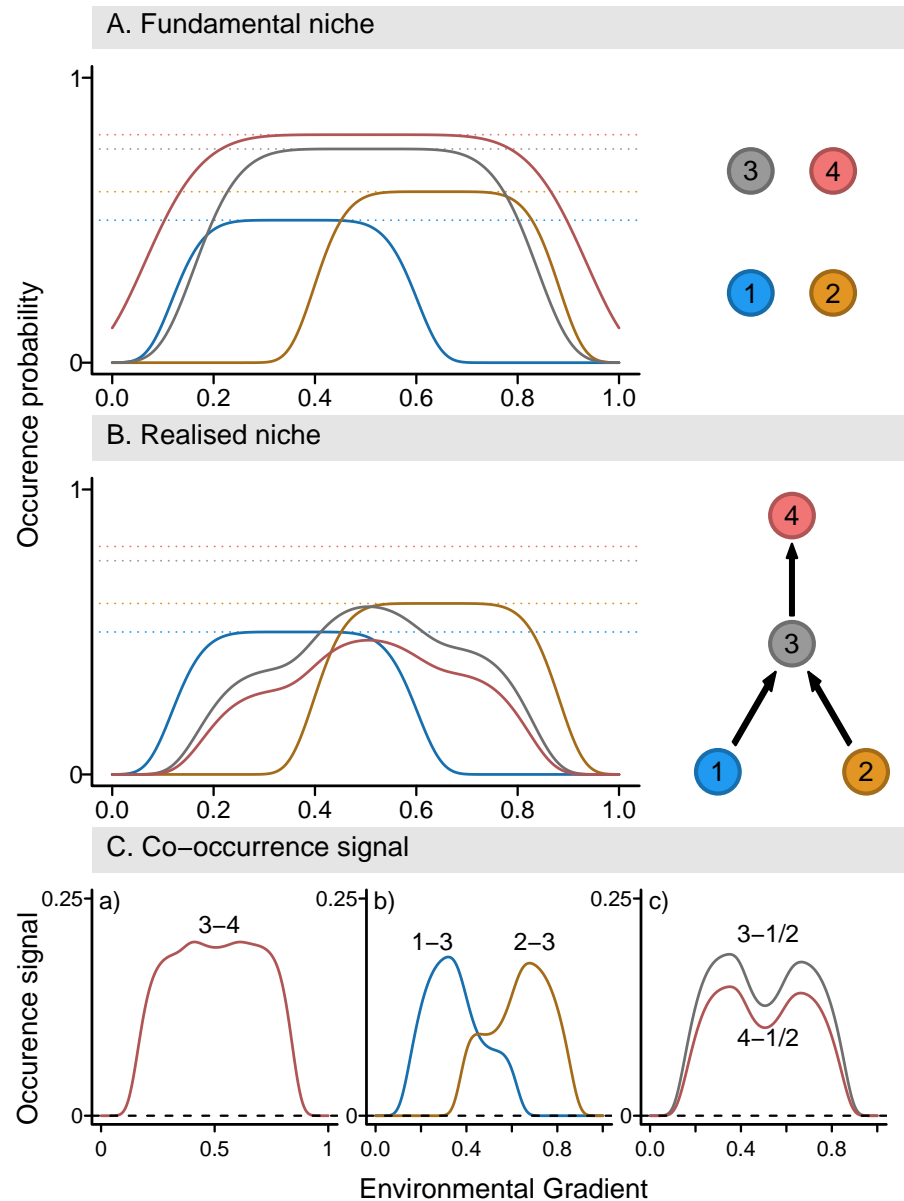


Figure 3.1: **Probabilistic description of fundamental and realized niches.** For a four species network, all the occurrence probabilities are derived along an environmental gradient assuming that (A) interactions are not limiting the distribution and (B) that predators 3 and 4 needs at least of one of its preys, *i.e.* species 1 or 2. Horizontal dotted lines in (A) and (B) stand for the occurrence probabilities reached at an environmental optimum. The co-occurrence signal is calculated for the following pairs : a) predators 3 and 4; b) predator 3 and prey 1, predator 3 and prey 2; c) predator 3 and prey 1 or 2, predator 4 and prey 1 or 2. If no difference are found, 0 is expected.

3.8 Supporting Information

3.8.1 Material and methods

In this section, we present in more details, the datasets and the methodology we used. All analyses have been performed using the R environment software and Table 3.2 present the functions and packages we used.

3.8.1.1 Datasets

Sites for the five datasets are reported on five maps gathered in Fig. 3.5. The total number of species, the number of species present in at least 1% of the total number of sites and the number of species for which traits information were available are reported in Table 3.3.

3.8.1.1.1 Pitcher Plants Network *Sarracenia purpurea* is a carnivorous plant that occurs along the east coast of North America from the panhandle of Florida to Canada and across southern Canada to British Columbia (?). *S. purpurea* has tube-shaped leaves that fill with rainwater and house a food web consisting of bacteria, protozoa, rotifers, and dipteran larvae among other taxa (?). This inquiline food web decomposes prey items releasing nutrients to the plant (?). We used pitcher plant data from 39 sites across North America (Fig. 1B) collected by ?. This dataset contains abundance data and feeding interactions for 20 pitcher plant food webs at each site for a total of 769 food webs (11 pitchers were dropped due to missing data). In total, there are 90 species and morpho species. Interaction structure (*i.e.*, who eats whom) is from

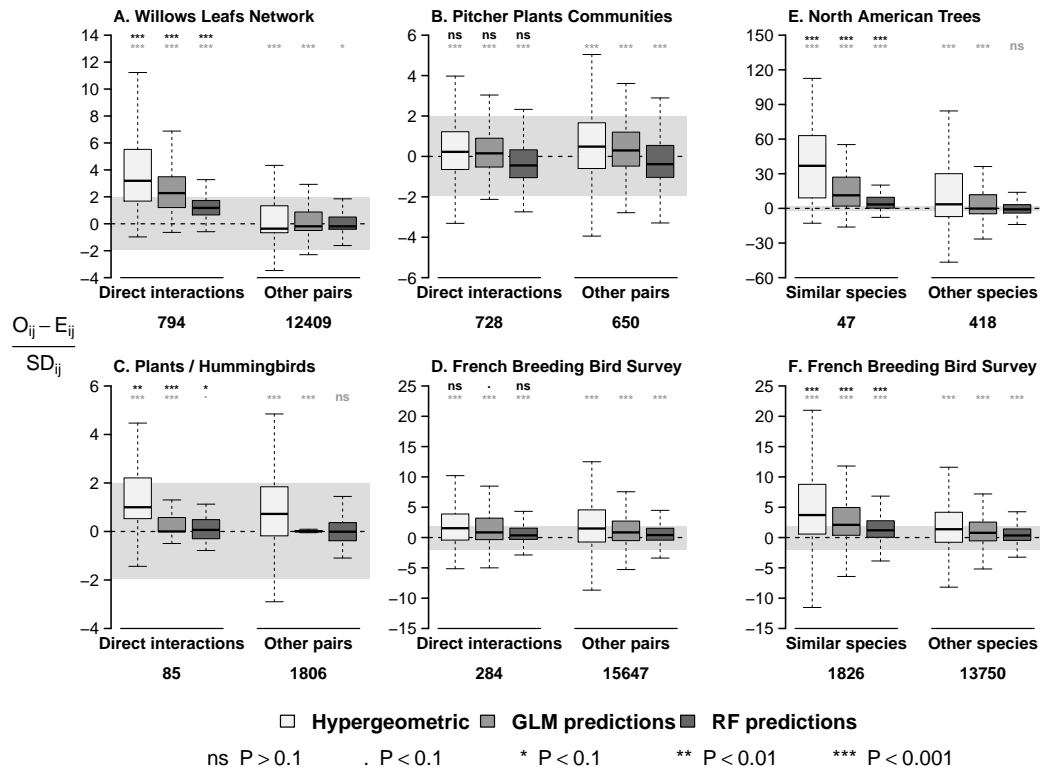


Figure 3.2: Co-occurrence of interacting versus not-interacting pairs of species. Figures under each groups of boxplots indicate the number of pairs to which the Z-score distributions refer. The light grey rectangle corresponds to the 95% confidence interval for the standard normal distribution which 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 (light grey symbols).

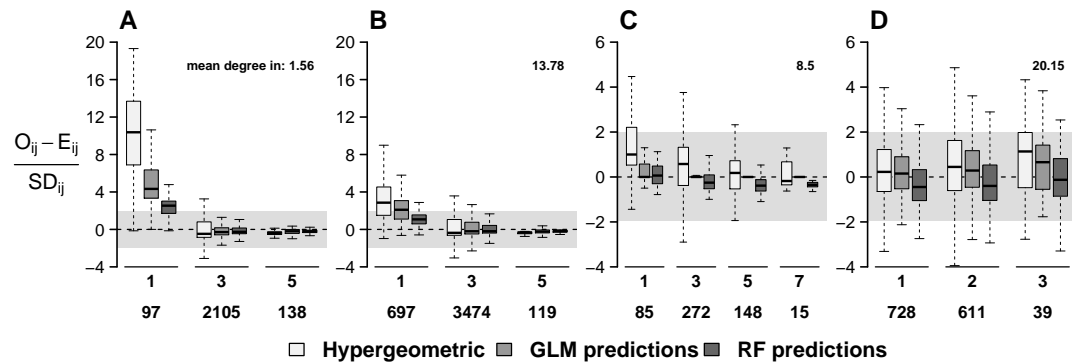


Figure 3.3: **Co-occurrence signal decays when the shortest path between a pair of species decay.** The Z-score distribution are plotted against the shortest path for A willows-herbivores interactions, B herbivores-parasitoids interactions, C birds-plants interactions and D the pitcher plants network. 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. Note that we used the same y-axis for panels A and B as they regard two different kind of interaction of the same dataset.

? which is based on previous studies (*e.g.*, ??, ?) and direct observation of feeding interactions.

3.8.1.1.2 North American Trees We used a distance built upon nine functional traits whose values were retrieved from (?), see Supplementary Table 3.4 available at <http://onlinelibrary.wiley.com/doi/10.1111/j.1466-8238.2010.00592.x/supinfo>. Each of the nine selected variables were centered and scaled (R functions used reported in Table 3.2) then used as is to derive Euclidean distances for all pairs of species. Then, we used agglomeration clustering with the Ward's method (implemented in the *hclust()* function, see Table 3.2) to obtain the dendrogram presented in 3.6.



Figure 3.4: **Co-occurrence significance decreases as the cumulated occupancy of preys increases.** For a given species, Z-scores are averaged over the all set species it interacts with and plotted against the joint distribution of the same set of species. We do so for the herbivores in the willows leafs network (panels A to C), the parasitoids in the willow leafs network (panels D to F), the hummingbirds in the Caribbean hummingbirds datasets (panels G to I) and all species in the pitcher plants network that consume other species (panels J to L). The x-axis is expressed as a log proportion of the total number of sites. Black symbols are mean Z-scores significantly different from 0 (see SI 3.8). 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 3.8).

3.8.1.1.3 French Breeding Birds Survey datasets We used 73 traits that are boolean variables reported in Table 3.5 based on which we derive Euclidean distances between all pairs of species.

3.8.1.2 Building metawebs

For four datasets, we built network based on all observed interactions and derived associated quantities, *i.e.* the connectance of the metawebs, the degrees of species and the shortest-path, using the R package *igraph* (Table 3.2).

3.8.2 Co-occurrence measurement

As mentioned in the main text, for three different scenarios, we derive a standardized co-occurrence: for a given pair of species i and j , we examined $\frac{O_{i,j} - E_{i,j}}{SD_{i,j}}$. Here, we provide more information about the three methods we used to analyze co-occurrence.

3.8.3 Hypergeometric distribution

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

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

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

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

by rearrangement:

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

once sorted out, this yields:

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

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

3.8.4 GLM and RF

For GLM and RF, $E_{i,j}$ corresponds to probabilities of occurrence computed based on climatic data. R functions are reported in Table 3.2.

3.8.4.1 Climatic data

We used the global climate layers provided by the data WolrdClim, version 1.4, available online at <http://www.worldclim.org> (?). For each dataset, we performed a principal component analysis and kept as many axes as needed to explain 90% of the total inertia. We used these axes for GLMs and RFs.

3.8.4.2 Generalized Linear Model

For all datasets, we performed a Generalized Linear Model (?) using all the axes provided by the PCA as polynomials of degree 2. To constrain the number of parameters, we did not evaluate the interactions among axis. We also performed model selection based on the Akaike's information criterion (AIC) in a Stepwise Algorithm (?). R functions used to carry out the analyses are indexed in Table 3.2.

3.8.4.3 Random Forests

Random Forests (?) were performed using the same formula as for GLMs. For all species, 10000 trees were computed and the probability of a species being in a given

site were calculated based on the number of votes the sites were granted.

3.8.4.4 Evaluating the models

For all species, we assessed the performance of the Species Distribution Models, *i.e.* Generalized Linear Model and Random Forest, using the Area Under the Receiver Operating Characteristic (AUROC, ?). We present the results as a cumulative sum of frequencies corresponding to the score for all species for each of the four ecological systems we studied (see Fig. 3.13).

3.8.5 Supporting Tables

3.8.5.1 R packages used

Table 3.2: R and packages used for the analyses. GLM: Generalized Linear Model, PCA: Principal Component Analysis. AUROC: Area Under the Receiver Operating Characteristic.

Analysis	Function	Package name	Version	Citation
Scaling and Centering	scale	base	3.3.1	?
Euclidean distance	dist	stats	3.3.1	?
Clustering	hclust	stats	3.3.1	?
PCA	dudi.pca	ade4	1.7.4	?
GLM	glm	stats	3.3.1	?
GLM Selection	step	stats	3.3.1	?
Random Forests	randomForest	randomForest	4.6.12	?
Degree of species	degree	igraph	1.0.1	?
Shortest Paths	shortest.paths	igraph	1.0.1	?
AUROC	somers2	Hmisc	3.17.2	?
TSN retrieving	get_tsn	taxize	0.7.4	?
Wilcoxon tests	wilcox.test	stats	3.3.1	?

Table 3.3: For each datasets we provide the total number of species (column *Total*), the number of species present in more that 1% of the total number of sites (column *Selected*), and the number of species for which traits information are available (column *Traits*). The symbol ‘-’ means ‘not relevant’.

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

Table 3.4: Tree species and traits used. Abbreviations are as follows: TSN - Taxonomic Serial Number defined by Integrated Taxonomic Information System (ITIS), maxH - Average maximum height, GR - Growth rate, WD - Wood Density, TolS - Shade tolerance, TolD - Drought tolerance, AM - Arbuscular mycorrhiza (Endomycorrhiza), EM - Ectomycorrhiza, LMA - Leaf mass per area, Nmass - Nitrogen content per leaf mass unit (?).

Species	TSN	maxH	GR	WD	TolS	TolD	AM	EM	LMA	Nmass
<i>Abies balsamea</i>	18032	25	1	0.34	5.0	1.0	0	1	151.00	1.66
<i>Acer negundo</i>	28749	20	3	0.44	3.5	3.0	1	0	37.04	2.50
<i>Acer rubrum</i>	28728	25	3	0.49	3.4	1.8	1	0	71.09	1.91
<i>Acer saccharum</i>	28731	35	1	0.56	4.8	2.3	1	0	70.63	1.83
<i>Betula alleghaniensis</i>	19481	25	3	0.55	3.2	3.0	0	1	46.08	2.20
<i>Betula papyrifera</i>	19489	25	3	0.48	1.5	2.0	0	1	77.88	2.31
<i>Carpinus caroliniana</i>	19504	8	1	0.58	4.6	2.0	0	1	49.05	2.15
<i>Carya cordiformis</i>	19227	25	1	0.60	2.1	4.0	0	1	44.05	2.60
<i>Fagus grandifolia</i>	19462	25	1	0.56	4.8	1.5	0	1	61.22	2.04
<i>Fraxinus americana</i>	32931	30	2	0.55	2.5	2.4	1	0	76.75	2.12
<i>Fraxinus nigra</i>	32945	20	2	0.45	3.0	2.0	1	0	71.94	2.10
<i>Fraxinus pennsylvanica</i>	32929	25	3	0.53	3.1	3.9	1	0	87.72	1.80

Species	TSN	maxH	GR	WD	TolS	TolD	AM	EM	LMA	Nmass
<i>Larix laricina</i>	183412	25	3	0.48	1.0	2.0	0	1	120.00	1.36
<i>Ostrya virginiana</i>	19511	12	1	0.63	4.6	3.3	1	0	37.04	2.20
<i>Picea glauca</i>	183295	25	1	0.35	4.2	2.9	0	1	302.86	1.28
<i>Picea mariana</i>	183302	20	1	0.41	4.1	2.0	0	1	294.12	1.12
<i>Picea rubens</i>	18034	25	2	0.38	4.4	2.5	0	1	304.67	1.15
<i>Pinus banksiana</i>	183319	20	3	0.42	1.4	4.0	0	1	243.90	1.24
<i>Pinus resinosa</i>	183375	25	3	0.39	1.9	3.0	0	1	294.12	1.17
<i>Pinus strobus</i>	183385	30	3	0.36	3.2	2.3	0	1	121.92	1.42
<i>Populus balsamifera</i>	22453	25	3	0.37	1.3	1.8	1	1	83.46	1.95
<i>Populus grandidentata</i>	22463	20	3	0.39	1.2	2.5	1	1	70.45	2.50
<i>Populus tremuloides</i>	195773	25	3	0.37	1.2	1.8	1	1	82.02	2.16
<i>Prunus pensylvanica</i>	24799	12	3	0.36	1.0	2.0	1	1	50.00	2.40
<i>Quercus alba</i>	19290	35	1	0.60	2.9	3.6	0	1	81.21	2.39
<i>Quercus macrocarpa</i>	19287	15	1	0.58	2.7	3.9	0	1	92.74	2.27
<i>Quercus rubra</i>	19408	25	2	0.56	2.8	2.9	0	1	84.20	2.06
<i>Thuja occidentalis</i>	505490	15	1	0.30	3.5	2.7	1	0	223.00	1.02
<i>Tsuga canadensis</i>	183397	30	1	0.40	4.8	1.0	0	1	122.55	0.99

Species	TSN	maxH	GR	WD	TolS	TolD	AM	EM	LMA	Nmass
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

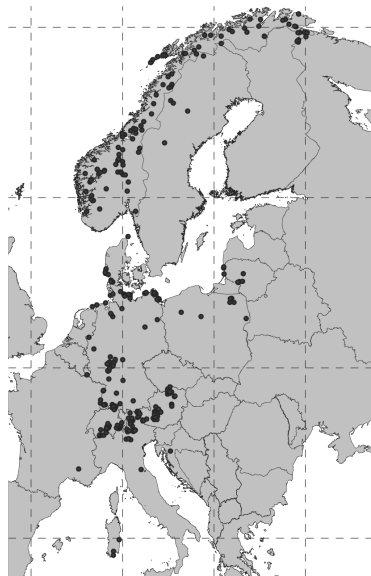
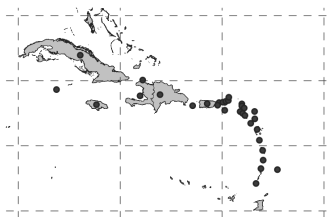
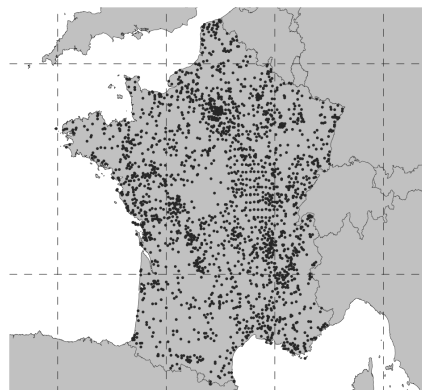
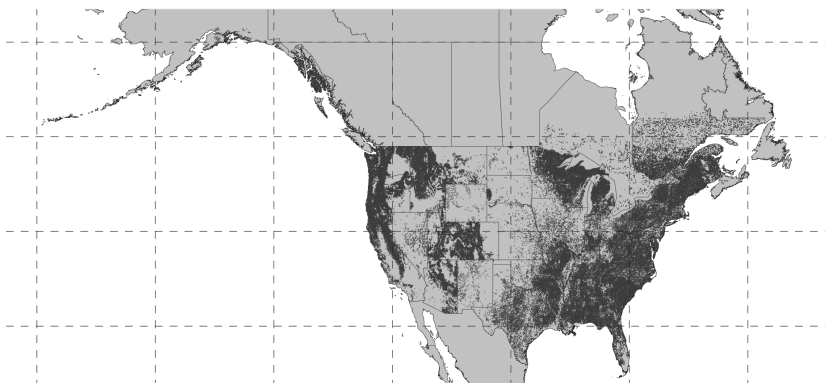
Table 3.5: List of the Boolean traits used to compute Euclidean distances between all pairs of species in the French Breeding Birds Survey.

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

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

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

3.9 Supporting Figures

A. Willow Leaf Communities**B. Pitcher Plants Communities****C. Hummingbirds / Plants sites****D. French Breeding Bird Survey****E. North American Trees****Figure 3.5: Sites of the study**

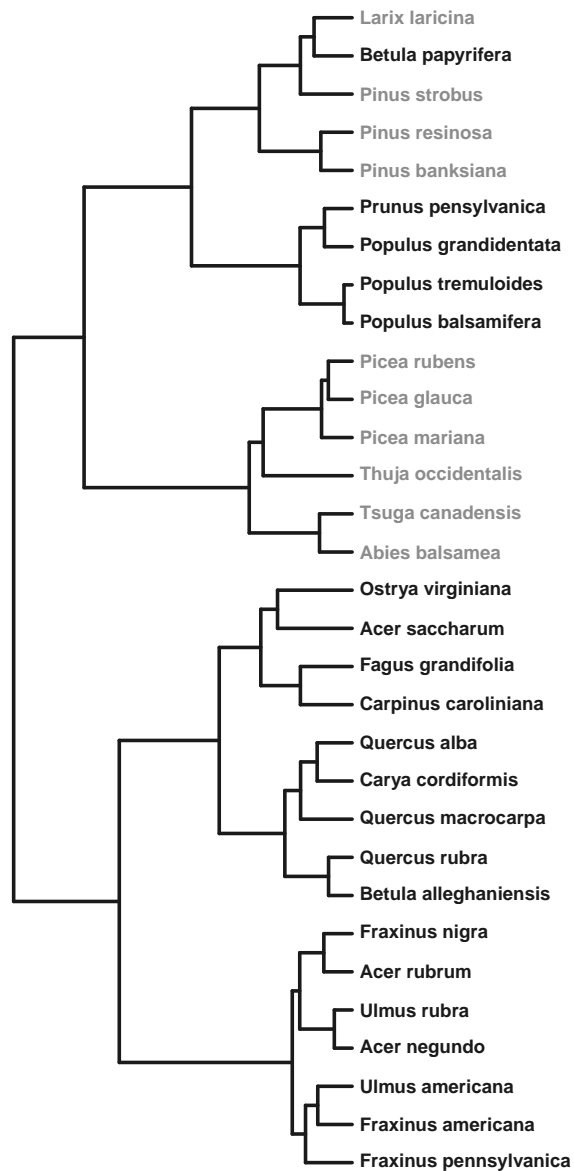


Figure 3.6: **Dendrogram representing the trait-based distances between the 31 species studied in the North American tree datasets.** Names of angiosperm species are written in dark grey while names of Gymnosperm species are in a lighter grey.

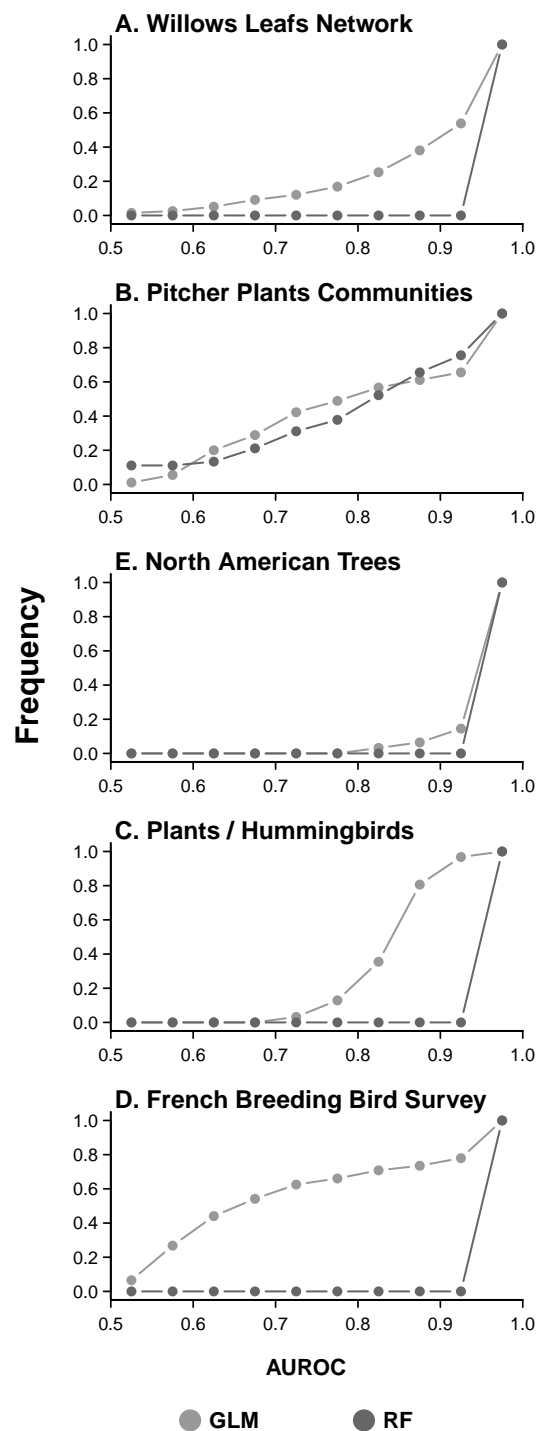


Figure 3.7: **Evaluation of the SDMs.** For each dataset, the distributions of AUC for GLMs (light grey symbols) and RFs (dark grey symbols) for all species are presented.

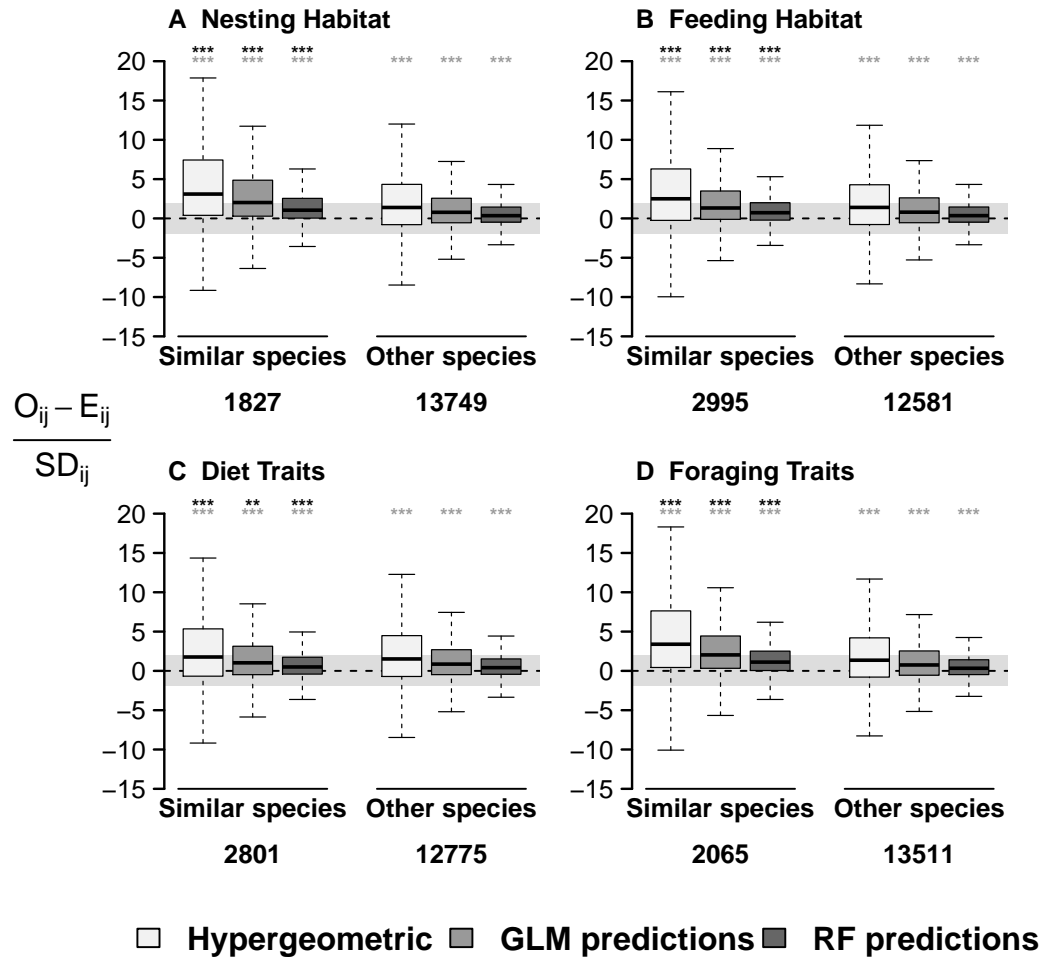


Figure 3.8: **Co-occurrence and the nature of the trait-based distance in the FBBS dataset.** The different panels correspond to four different set of trait upon which different distances are built. Similar species are defined as the species for which the trait-based distance is less than or equal to 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). They were also computed to show whether whether Z-score were greater for hypergeometric versus GLM and GLM versus RF (grey symbols).

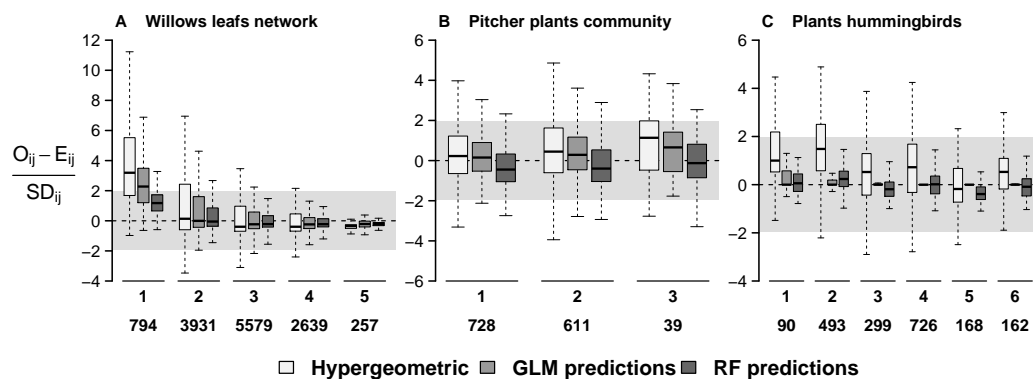


Figure 3.9: **Co-occurrence signal decays when the shortest path between a pair of species decay** Distribution of Z-scores for all interactions are grouped by shortest-path indicated by the first numbers below boxplots. The other figures below stand for the number of pairs of species included within the distributions.

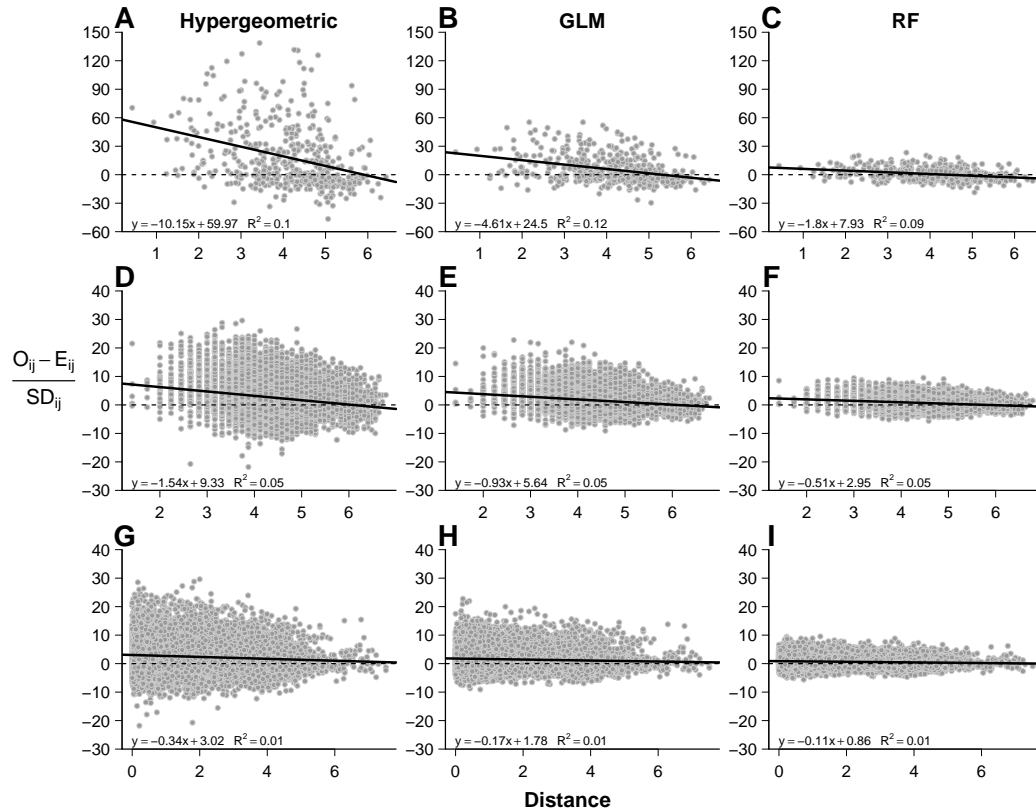


Figure 3.10: **Changes in co-occurrence signal with increased distance between two species** Points represent the result for all pairs of interaction for two datasets: the North American Tree dataset (A=C) and the FBBS (D-I). For the latter, we used the trait-based distance computed with all available 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.

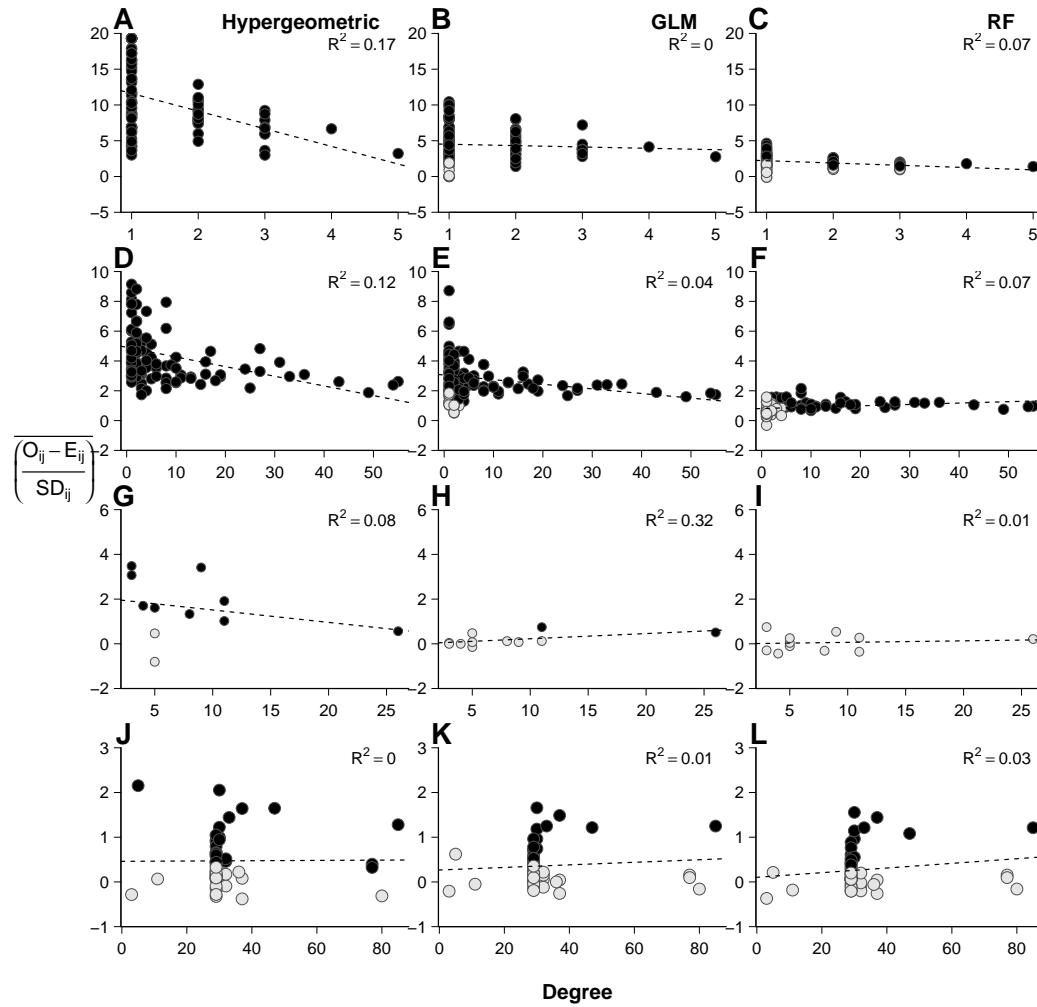


Figure 3.11: The degree of species partially explains the decrease of the co-occurrence strength. For the herbivores (A-C) and the parasitoids in the willow leaves network datasets (D-F), the hummingbirds in the Caribbean hummingbirds datasets (G-I) and all species in the pitcher plants network that consume other species (J-L) the mean Z-score is plotted against the degree of the species. Black symbols are mean Z-scores significantly different from 0 (see SI Text). In each panel, the dotted line represents the linear regression $y = ax + b$ for which the R^2 is provided.

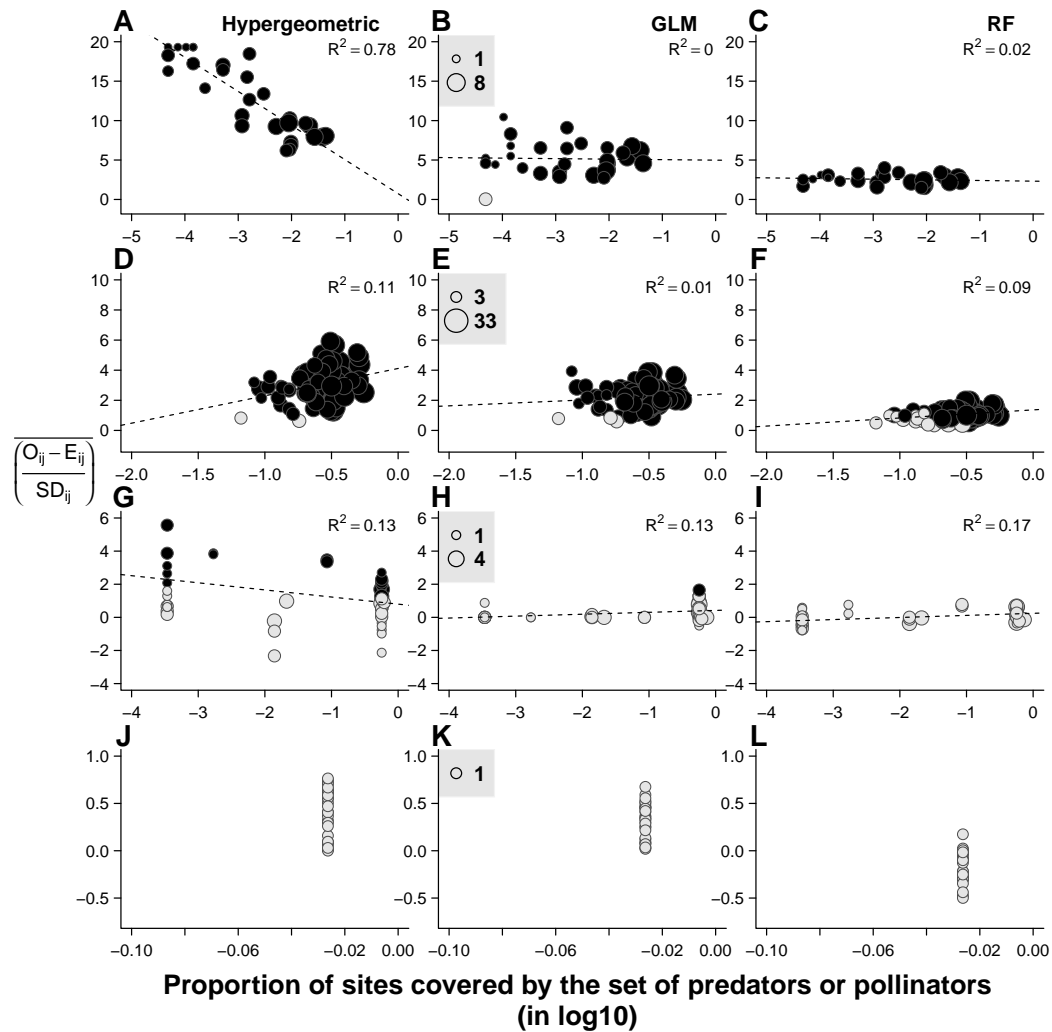


Figure 3.12: **Co-occurrence and cumulated occupancy of predators.** This figure corresponds to figure 3.4 in the main text, but the Z-score were calculated for preys (host plants) rather than for predators (pollinators). Mean Z-score are computed for willows (A-C) and herbivores (based on the herbivores-parasitoids only, D-F) of the willows leaves 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.

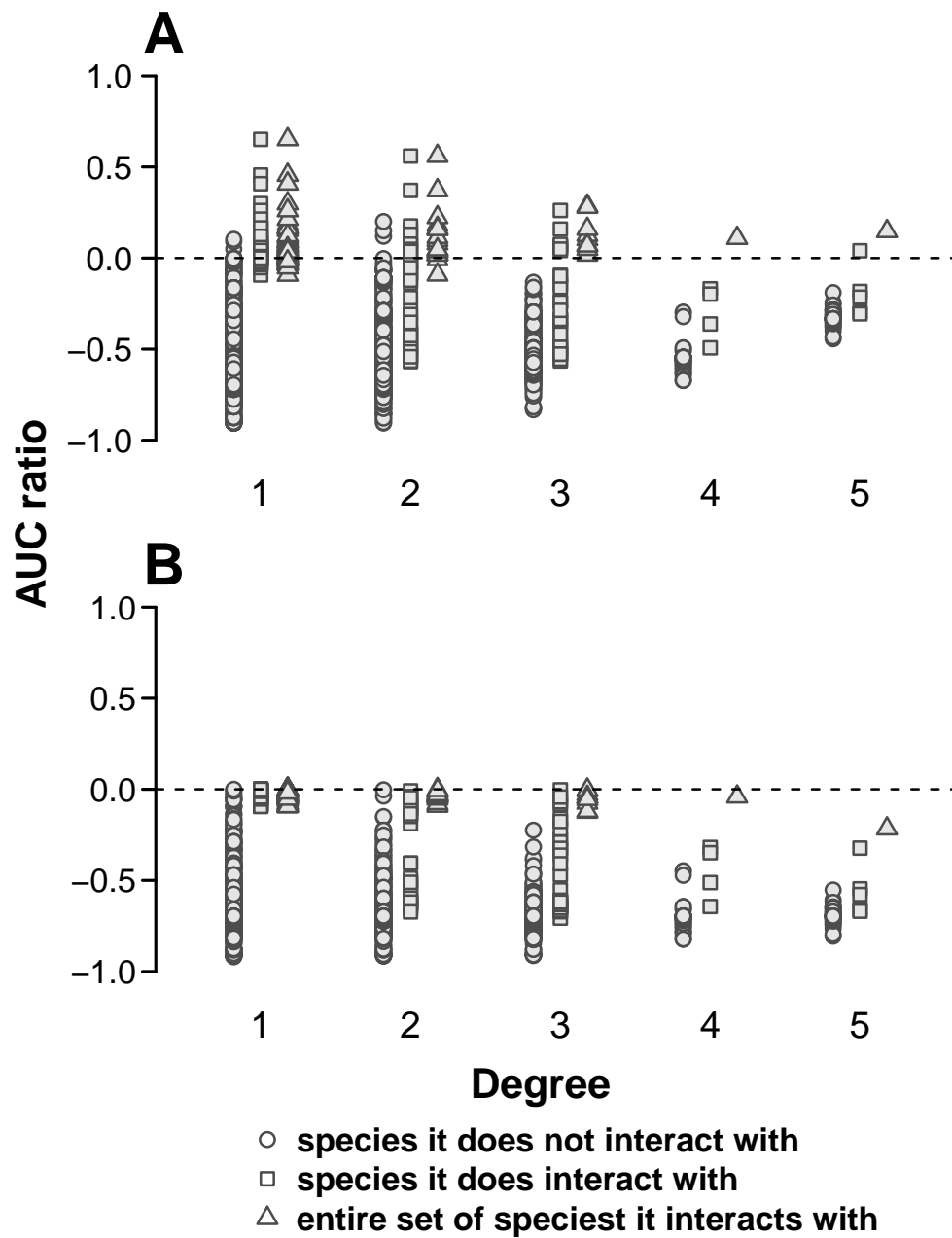


Figure 3.13: **Predicting herbivore distribution based on the distribution of willows.** For the herbivores in the willow leaves network dataset, we compared the AUC obtained when using willow it does not interact with (circles) a willow in interacts with (squares) and the set of willow it interacts with (triangles) to AUC obtained for GLM (A) and RF (B). Positive values indicated that species based model outperformed the SDM model.

ARTICLE 4

VERS UNE THÉORIE MÉTABOLIQUE DE LA BIOGÉOGRAPHIE DES ÎLES

4.1 Résumé en français du troisième article

4.1.1 Contexte scientifique

Dans ce chapitre, je présente une formulation énergétique du modèle de la TTIB (?). Pour cela, je transforme l'île de la TIB en une quantité de producteurs primaires. À partir de l'énergie générée par ces derniers, je m'intéresse à la dynamique de colonisation de l'ensemble des niveaux trophiques supérieurs du réseau écologique décrit à l'échelle régionale (le metaweb tel que décrit aux chapitres précédents). Dans ce modèle, une espèce du metaweb colonise une île sans difficulté si 1- elle y trouve une ressource qu'elle est en mesure d'exploiter et si 2- il y a assez d'énergie disponible pour que les espèces présentes puissent toutes maintenir une population locale d'une taille minimale, si ces conditions ne sont pas remplies alors la colonisation est impossible ou un processus d'extinction est enclenché. Je propose dans ce chapitre un calcul simple mais astucieux pour intégrer les différences de consommations entre les différents niveaux trophiques. Le calcul est basé sur la biomasse des espèces mais aussi sur le transfert énergétique, qui représente la quantité d'énergie qui passe d'un réseau à l'autre. À partir de cette nouvelle formulation de la TTIB, j'obtiens une relation espèce-énergie (SER) informative sur la structure du réseau.

L'approche que je propose est, à mon avis, un pas significatif vers une intégration plus aboutie des mécanismes écologiques dans la TIB, notamment pour les interactions biotiques. Je prolonge le travail de la TTIB et je fais un premier rapprochement avec la théorie métabolique de l'écologie. Bien que le modèle proposé soit relativement simple, il est une pierre solide pour aller plus loin dans l'exploration d'une théorie métabolique de la biogéographie. De plus, au regard de l'abondante littérature relative aux relations allométriques, il me semble essentiel de comprendre comment les contraintes énergétiques, exercées à l'échelle de l'individu, se propagent à l'échelle des populations puis aux échelles biogéographiques.

4.1.2 Publication envisagée

Présentement, l'article offre des perspectives intéressantes sur une théorie métabolique en biogéographie mais aussi des pistes de réflexion pour s'affranchir simplement de certaines contraintes de modélisation qui pourraient hautement complexifier le modèle (par exemple, je ne considère pas explicitement la dynamique des populations sur l'île). Je vois deux opportunités associées au travail ici présenté. Premièrement, orienter davantage mon travail sur les aspects techniques et mathématiques du modèle et essayer, par exemple, de dériver une solution approchée du modèle. Cela m'orienterait vers une publication plus technique pour un journal spécialisé sur les aspects théoriques en écologie. Deuxièmement, aller plus loin dans l'exploration des possibilités offertes par le modèle, insister sur les perspectives, notamment sur l'impact de la structure des réseaux sur la relation énergie-espèces. Cela m'orienterait plutôt vers un article de perspectives, pour un journal d'écologie généraliste.

Pour cet article ma réflexion a été alimentée par discussion avec Dominique Gravel,

Miguel Araújo et Loïc Pelissier que je remercie. Je me suis occupé de concevoir le modèle, d'implémenter le modèle et de sortir les résultats et de l'analyser. J'ai écrit la première version du modèle et j'ai bénéficié d'apports significatifs de Dominique Gravel qui a fait une relecture complète du manuscrit. Je remercie également chaleureusement Kevin Solarik qui m'a apporté de nombreuses relectures mais aussi des suggestions pertinentes. Cet article est également la base de réflexion du travail que j'aimerais mener dans un projet de recherche post-doctoral.

4.1.3 Traduction du résumé en anglais

L'énergie façonne la biodiversité; l'abondance de l'énergie solaire et la disponibilité en eau décrivent bien les écosystèmes terrestres de même que la température de surface de la mer et la disponibilité en nutriments écrivent les écosystèmes marins. À l'échelle de la communauté, les réseaux trophiques et leurs structures déterminent les échanges d'énergie des producteurs primaires jusqu'aux prédateurs du haut de la chaîne alimentaire. En dépit de sa grande importance, l'introduction de l'énergie dans les théories classiques en biogéographie a été négligée, ces théories considèrent les espèces équivalentes (c'est le cas de la Théorie de la Biogéographie des Îles) ou se concentrent sur le concept de niche écologique. Pour rectifier cela, nous développons une théorie pour inclure les contraintes énergétiques pour décrire la répartition de la biodiversité. Comme illustration des perspectives offertes par cette approche, nous étudions la richesse spécifique et la structure des réseaux le long d'un gradient d'énergie. Avec l'augmentation de l'énergie disponible, nous obtenons une succession claire des espèces selon leur statut trophique : des plus bas niveaux trophiques sur les îles pauvres en énergie jusqu'aux plus hauts niveaux sur les îles où l'énergie est abondante. C'est une piste prometteuse pour aller au-delà de la prédiction quantitative des pertes d'espèces lorsque les habitats

sont fragmentés et prédire le rôle qu'elles ont au sein du réseau écologique. Nous pensons que l'utilisation de ce modèle sera très utile pour une intégration de l'écologie des communautés en biogéographie, tout en permettant aussi une meilleure paramétrisation des modèles de distribution d'espèces.

Les sections qui suivent sont celles de l'article publié.

4.2 Title

Towards a metabolic theory of island biogeography.

4.3 Abstract

Energy availability shapes biodiversity; solar radiation and water availability describes the terrestrial ecosystems well, just as sea surface temperature and nutrients availability describe marine ecosystems. At the community scale, food webs and their structures control energy exchanges from primary producers to top predators. Despite its crucial importance, the implementation of energy has been overlooked in classical theories in biogeography, which often assume species are equivalent (*e.g.* the Theory of Island Biogeography) or either remain too focused on the concept of the ecological niche. In order to rectify this, we develop a theory to include energy constraints to describe the distribution of biodiversity. As an illustration of the perspectives offered by this framework, we derive expected species richness and network structure along an energy gradient. When increasing the energy availability, we obtain a clear succession of species according to their trophic status, from the lowest trophic level on poor energy

islands to the highest species when energy is abundant. This a promising avenue to predict more than a expected number of species we loss when habitat is fragmented but also their role in the ecological network. We are confident the use of this model will be proven beneficial for the integration of community ecology with biogeography, while also better parameterize species distribution models.

4.4 Introduction

Disentangling the relative contribution of the different processes shaping biodiversity distribution is the central objective of biogeography. While biogeographers clearly envision the list of ingredients needed to understand species distribution (?), they miss the framework to mix them in the right amount and to investigate the biogeography of ecological communities. We will likely fail to predict accurately biodiversity responses to global changes as we keep our focus strictly on abiotic factors *i.e* temperature and precipitation, while continuing to overlook biotic interactions (?) and short-term evolutionary responses (?). At the core of this issue is the need for an integrated approach at the crossroad between species distribution modelling and community ecology.

Biotic interactions should be integrated as a constraint for species co-distribution in meta-communities (??). Ecological interactions may explain, at least partially, the dynamics of local extinctions, which in turn would further explain some properties of the geometrical shape of species ranges (*e.g.* nested distributions of parasitoid and its host, ?). We have however limited evidence of the effect of biotic interactions on large scale species distribution (see Chapter 3 - but see ?). Incidentally, two of the most influential models in biogeography assume ecological equivalence of species; the Theory of Island Biogeography of MacArthur and Wilson (hereafter TIB, ?) overlook the variation

among species characteristics and ignore their arrangement in organized ecological networks. In his neutral theory, Hubbell assumes that individuals of different species are ecologically equivalent and predicts the distribution of abundance without considering interactions (??). These two theoretical models have been proven relevant for certain groups of species and inadequate for others, and none of them were initially intended to describe exhaustively the structure of ecological communities. The integration of ecological interactions into biogeography however offers promising perspectives and a much extended range of predictions (??).

The TIB is well-suited to explore the consequences of ecological interactions on community structure at broad spatial scales, as it includes two fundamental processes of biogeography, immigration and extinction, in an elegant fashion that eases its extension (??). Building upon the classical model of MacArthur and Wilson, recent studies have included various forms of interactions in the TIB (??). The regional species pool becomes a *metaweb*, not only listing potential colonists but also their interactions. With this approach, species are inter-dependent entities with specificities (*e.g.* a given trophic level) rather than indistinguishable units. As a consequence, colonization and extinction rates vary with respect to species identity *and* the composition of the local community. ? proposed the Trophic TIB (hereinafter TTIB) to represent food web assembly on isolated communities, where predators can persist locally as long as they find at least one preys. More generally, ? presented a Lotka-Volterra like model in which the composition of the local community determines both colonization and extinction rates. Adding new ecological processes in the TIB increases realism, at the cost however of reducing its simplicity. It provides an extended list of predictions, which have yet to be proven significantly better than the original theory (see ?). Extending the TIB while preserving its elegance is a challenging and technical issue. Here we propose to reformulate the TTIB in terms of energetic constraints in attempt to solve this

challenge.

The TIB was originally meant to predict and explain the common observation of increasing species richness with island area. Another ubiquitous and fundamental observation in biogeography is the latitudinal diversity gradient (LDG, ???). The productivity of primary producers is generally found the best predictor of species richness at large spatial scales (??). Others found a strong correlation between richness, water availability and temperature (??). Several mechanisms have been proposed to explain this relationship (see ??, for a review), including higher speciation rates in tropical areas, greater biomass and finer niche partitioning. Recently, some authors investigated how networks could change with latitude. Based on the meta-analysis of 196 empirical food webs, ? shown that ecological networks of interactions also vary along this gradient but that the density of links remains constant throughout. In the majority of ecosystems theses authors studied, an increased energy results in a wider niche space, rather than an increase of niche breadth poleward. Similarly, ? reconstructed marine networks across the globe and found significant latitudinal-network gradients (LNG). These studies are however descriptive only, as there is currently no theory to explain how and why energy constraints should impact food web structure. In 1983, ? developed the species-energy theory and replaced the “area” with “available energy” to derive a meaningful Species Energy Relationship (SER). Although area and energy may bring more information taken together (?), the rationale behind it allows the derivation of species abundance and occurrence probability based on energetic constraint (?). Wright’s approach however do not specifically account for network structure and thereby remains limited to species richness.

Species do not escape from the laws of thermodynamics and as a consequence, despite the complexity of evolutionary trajectories, many key ecological properties and pro-

cesses scale with body mass (*i.e.* metabolic theory of ecology ??). The scaling of the metabolic rates with body mass is the fundamental observation underlying this theory (?), which typically follows a power function with coefficients often between $2/3$ and $3/4$ (?). Even if all the relationships are not well-understood (see the case of abundances reviewed in ?; and the recent relationship between prey and predator biomasses ?), the ubiquity of allometric relationships makes body mass distribution a key variable to reduce the complexity of ecosystems. Allometric relationships and energy flows also provide the mean to parameterize models of population dynamics (?). Recent developments of food web theory have convincingly shown that allometric relationships are the cog to analyze the properties of ecological networks (??), their dynamics (?) and to characterize the role of species within them (?).

Our objective in this study is to develop a theory of biodiversity distribution based on first principles of energetic constraints in order to investigate the impact of trophic interactions on the latitudinal-diversity gradient. We propose to extend the model of the TTIB following the vision of ?. As a consequence, our theory encompasses both the species-area and the species-energy relationships, making it a very general framework to understand and predict biodiversity distribution. In addition, the explicit integration of trophic interactions allows the development of a vast array of testable and insightful predictions, including the expected latitudinal-network gradient. To do so, we propose a theoretical model of food web assembly driven by colonization from the regional metaweb and energetic constraints. We derive expected species richness and network structure along gradients of primary productivity. We use “Lindeman” (?) transfer efficiency (*i.e.* energy transferred to another trophic level) to integrate species-specific differences in their source of energy. We derive a SER and the relationship between body-mass distribution and available energy. Our model formalizes the initial hypothesis of Lindeman and offers immense possibilities for the integration of community

ecology with biogeography.

4.5 Model description

The model developed by MacArthur and Wilson predicts species richness on an island according to the size of the island and its distance from the mainland (?). One promising direction to extend this theory is to include biotic interactions into the classical model (???). Following this avenue, we consider explicit interactions among the regional pool of species based on allometric relationships. Body mass is used to determine the quantity of energy a given species requires to maintain its local population on an island. Extinction follows if the island cannot sustain a minimal population of any given species. Therefore, in the model described below, we simulate the TIB with purely stochastic colonization events together with deterministic extinction based on an energy rational.

We consider an island as an isolated piece of land covered by a maximal quantity of primary producers, which determine the amount of available energy upon which a food web can develop. We denote E_0 the maximal amount of energy available for herbivores and consider it to vary proportionally to island area. Here we do not enter the specifics of how productivity scales with area, as in some situation it could scale linearly or not *e.g.* for lakes when the volume determines productivity (?). As a consequence, the model explains both the scaling of richness with island area and productivity, with the same underlying principle that a minimal population size is required for persistence. We make two additional simplifications: the diversity of primary producers is not taken into account, and the production is constant over the time.

The regional pool of species in the TIB is the number of species P and the trophic interactions among them. Following ?, potential interactions among them are determined using the niche model (?). We furthermore assume the niche axis to be the body size of the species (?) and that species which are without any links in the metaweb are determined to be herbivores. Note that primary producers are not included in the niche model but the source of energy of herbivores. They are rather included in the total energy E_0 (see below) consumed by herbivores. As a consequence, trophic level is strongly correlated to body mass, with herbivores often the smallest and top predators the largest. Although this assumption is an oversimplification of the determinants of ecological interactions, it remains reasonable for marine ecosystems (?), where allometric relationships have been used to infer food web structure (?).

Following ?, we assume that the colonization of herbivores/predators is successful only if they find at least one of their habitat/prey on the given island. Additionally, we propose that colonization events are possible only if the energetic requirements to maintain local populations exceeds the primary production. In this case, energetic constraints and network topology determine the identity of species to go extinct. Colonization events are assumed to be purely stochastic, while extinctions are deterministic, controlled by the energetic constraints (this difference in stochastic nature between these fundamental processes of biogeography has been recently supported in ?, where the authors show that extinction probability decreases with the presence of resource/prey on the island).

4.5.1 Energetic constraints on local food webs

The rationale for energetic constraints is simple: local populations need a certain amount of energy to maintain a minimum level of activity otherwise they die from starvation. Under this constraint, species richness locally increases until the energy available on the island (E_0) is no longer sufficient for all populations because of inefficient energy transfer at each trophic interactions. Therefore all population on the island are maintained as long as the inequality below holds true:

$$\sum_i e_i \leq E_0 \quad (4.1)$$

where e_i is the energy uptake of the insular population of species i . We now need to define a rule to set the the minimal energy a population needs to survive on the island. For a given species i , energy requirements of any individual is derived from allometric relationships proposed by the metabolic theory of ecology (?), of the form $c_i m_i^b$ where c_i is a constant, specific constant, m_i the body mass of species i and b the coefficient of power (between 2/3 and 3/4 ?). We do not integrate intraspecific variability in body mass and thus m_i is a constant for a species. The energy uptake associated to a minimal viable population (hereafter MVP) of n_i individuals becomes $n_i m_i^b$. ? defined the MVP as “[...] the smallest isolated population having a 99% chance of remaining extant for 1000 years despite the foreseeable effects of demographic, environmental and genetic stochasticity, and natural catastrophe”, highlighting that the smallest portion of a population has the highest possible extinction risk (extrapolated as time to extinction). The key problem is whether or not the MVP is size dependent. The most conservative assumption is to consider it to be a constant. ? however shown that the time to extinction is affected by the mean population growth rate, which underlies that species character-

istics may lead to a heterogeneity in MVP. Moreover, ? had developed a framework within which they proved the growth rate to be proportional to m_i^{-b} . Based on these results, we explore two different rules for extinctions: (1) MVP is equal for all species: $n_i = n_0$, and (2) MVP scales with the growth rate $n_i = n_0 m_i^{-b}$. For both scenarios, species i can survive only if the energy expenditures can be covered, *i.e* if the energy available is greater than: $n_i c_i m_i^b$.

4.5.2 Energy fluxes and transfer efficiency

Although the expression of energy consumption is similar among species, they obtain it from different sources: herbivores feed on primary producers, whereas predators feed on a set number of secondary consumers. The primary production must therefore be tracked as it moves up through the food chain. To deal with this, we propose to convert the energy costs for maintaining predator populations into additional populations of herbivores to be maintained. To exemplify this idea, we start with the simplest trophic network where a predator j feeds upon a herbivore i . The cost to maintain the MVP of i is $n_i c_i m_i^b$ and $n_j c_j m_j^b$ for j . For the latter, we convert n_j into an extra population of i , $n_{i,j}$ herbivore individuals dedicated to j consumption. Furthermore, we account for the energy loss during consumption by including a transfer efficiency (*sensu* ?) τ . Hence, the conversion from n_j to $n_{i,j}$ is given by the following equation:

$$\tau n_{i,j} c_i m_i^b = n_j c_j m_j^b \quad (4.2)$$

the energy requirements of n_j are converted into a energy expressed in term of a population of i based on which we obtain the population size of i needed to maintain a MVP

of j :

$$n_{i,j} = \frac{n_j c_j}{\tau c_i} \left(\frac{m_j}{m_i} \right)^b \quad (4.3)$$

We postulate that the transfer efficiency is constant across trophic levels, which is likely an oversimplification of the reality as suggested by the sparse empirical data available (??). We now add a new predator k feeding on j (the predator of i) to the insular food web. According to our reasoning, we must covert n_k into $n_{i,k}$. To do so, we start by converting n_k into a population of j :

$$n_{j,k} = \frac{n_k c_k}{\tau c_j} \left(\frac{m_k}{m_j} \right)^b \quad (4.4)$$

We now turn $n_{j,k}$ into a herbivore population:

$$n_{i,k} = \frac{\frac{n_k c_k}{\tau c_j} \left(\frac{m_k}{m_j} \right)^b c_j}{\tau c_i} \left(\frac{m_j}{m_i} \right)^b \quad (4.5)$$

which gives:

$$n_{i,k} = \frac{n_k c_k}{\tau^2 c_i} \left(\frac{m_k}{m_i} \right)^b \quad (4.6)$$

In a similar fashion, for a linear trophic chain, we can demonstrate that the additional population of herbivore i necessary to maintain predator j of level l is:

$$n_{i,k} = \frac{n_j c_j}{\tau^{l_j} c_i} \left(\frac{m_j}{m_i} \right)^b \quad (4.7)$$

In many cases, a predator feeds on an array of prey species rather than a single one. As such, it uptakes its energy from the different sources and we must account for all of them. We assume that energy costs associated with maintaining the local predator are minimal. Therefore, n_j is converted into i , the herbivore linked to j for which (4.7) is minimal. Basically, i should be a large and separated from j by a low number of species. Hence, on the island, the species richness increases as long as the inequality below holds true:

$$\sum_i c_i n_i m_i^b + \sum_j \frac{n_j c_j}{\tau^{l_j} c_i} \left(\frac{m_j}{m_{i_j}} \right)^b \leq E_0 \quad (4.8)$$

The first term is the energy cost for maintaining populations of herbivores, the second term is associated to higher trophic levels: predator j is converted into herbivore i_j from which it is separated by l_j links. For the sake of simplicity, we make an extra assumptions: c_i values are constant among species and set to 1. Therefore, if we assume that MVP is constant (scenario 1), inequality 4.8 becomes:

$$\sum_i m_i^b + \sum_j \frac{1}{\tau^{l_j}} \left(\frac{m_j}{m_{i_j}} \right)^b \leq \frac{E_0}{n_0} \quad (4.9)$$

Similarly, if we assume $n_i = n_0 m_i^{-b}$ and $n_j = n_0 m_j^{-b}$ (scenario 2), then:

$$\sum_i 1 + \sum_j \frac{1}{\tau^{l_j}} \left(\frac{1}{m_{i_j}} \right)^b \leq \frac{E_0}{n_0} \quad (4.10)$$

The left side of this equation provides the minimal energy needed to sustain all populations while the right side is the total amount of energy available. Therefore, the maximal population of species i without any extinction is given by converting the extra amount of energy available into an additional population of species i , the population of i we get is denote $n_{i,max}$. The range $[n_i, n_{i,max}]$ is then the range of possible fluctuations of species i without any new extinction event.

4.5.3 Extinctions

When the local primary production cannot sustain the establishment of a new immigrant, *i.e.* when inequality 4.8 is no longer verified, its arrival is either impossible or lead to extinction of other species. In the latter situation, we must determine the identity of species that will go extinct, which remains a significant challenge that will not be undertaken here given its complexity highlighted by recent theoretical studies (??). To overcome this concern, we use two scenarios: (1) *random extinction*: any species can go extinct, and (2) *costs-based extinctions*: the probability of extinction is proportional to the energetic costs of the species. Once an extinction occurs, we ensure that all species remain linked to at least one herbivore, otherwise unlinked species will too go extinct. Finally, extinctions are set to occur until 4.8 is satisfied. Note that the deterministic nature of extinctions lays in in the fact that they occur when energy is lacking, the identity of the species that actually goes extinct remains stochastic.

Although the assumptions we have made here are likely unrealistic, our intentions are

Table 4.1: Hypotheses associated to the four scenarios.

Scenario	MVP	extinction
A	n_0	random
B	n_0	costs-based
C	$n_0 m^{-.75}$	random
D	$n_0 m^{-.75}$	costs-based

merely to examine the model under the special case of an optimal energy allocation on an island. These assumptions remain acceptable as long as we focus on the qualitative consequences in term of community dynamics. As an important remark, the energy cost associated to one predator is not based on inherent properties only, as it is determined also by the identity of its preys present on the island (in other words, via apparent competition). Hypotheses made for the four scenario are recapitulated in Table 4.1.

4.5.4 Simulations

Every simulation starts with the generation of the regional metaweb of 100 species. To do so, we first draw a number p_i for all species from a uniform distribution from the range $[0,3]$. Numbers p_i are then sorted, the lowest values are assigned to the herbivores and we use the vector obtained as the niche axis in the niche model, with connectance set to 0.05 (?). Species that are not connected are assumed to be herbivores and to allow comparisons among simulation we standardize the number of herbivore to 10. To do so we keep generating new metawebs until one we obtain one metweb with 10 species that were not connected. We also use the p_i number to derive the biomass of all species, we do so by using 10^{p_i} . The smallest species has the smallest niche axis.

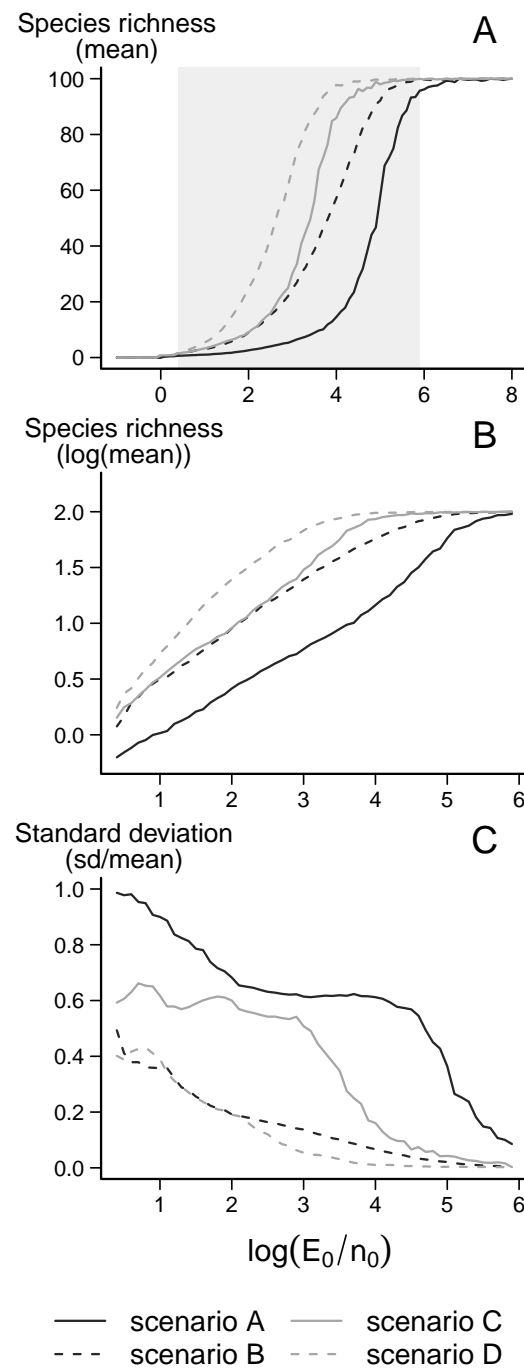


Figure 4.1: **Species-Energy Relationship** using 100 replicates for each of the four scenarios for (A) mean species richness, (B) logarithm of the mean, and (C) coefficient of variation.

We simulated the model along a gradient of energy ranging in $\frac{E_0}{n_0}$ from .1 to 10^8 . We start with an empty island and at any time step, any species of the metaweb has a probability $c = 0.001$ of colonizing the island. If a successful colonization occurs if available energy on the island allow to sustain a MVP of the new species. For all time step we record the species on the island. We perform 125,000 iterations and discard 25,000 burn-in iterations and the 100,000 remaining to do our analyses. For all the 91 values of the energy gradient we used 100 replicates, meaning 100 different regional metawebs.

4.6 Results

We found a sigmoidal SER for all scenarios (Fig. 4.1 A). Under scenario A, for instance, species richness increased exponentially with the increase in energy, where more species were allowed to colonize the island (Fig. 4.1). SER reached saturation much faster under scenarios C and D than under the scenarios A and B, due to the difference in energy uptake. Under scenarios C and D the MVP decreased with body mass, which allows higher species richness or a comparable amount of energy. Remarkably, species richness was smaller for scenarios where the extinction rate was cost-based than randomly based (Fig. 4.1 A-B), which also results in less variability (Fig. 4.1 C).

Occurrence probability increases with available energy and decreases with trophic level for all scenarios (4.2). Herbivores were found to colonize the island first, followed by predators of increasing trophic level. The trend was similar for all scenarios, and more pronounced for those where extinction is cost-based (scenarios B and D). When species occurrence probabilities were grouped by body mass, only significant differences were

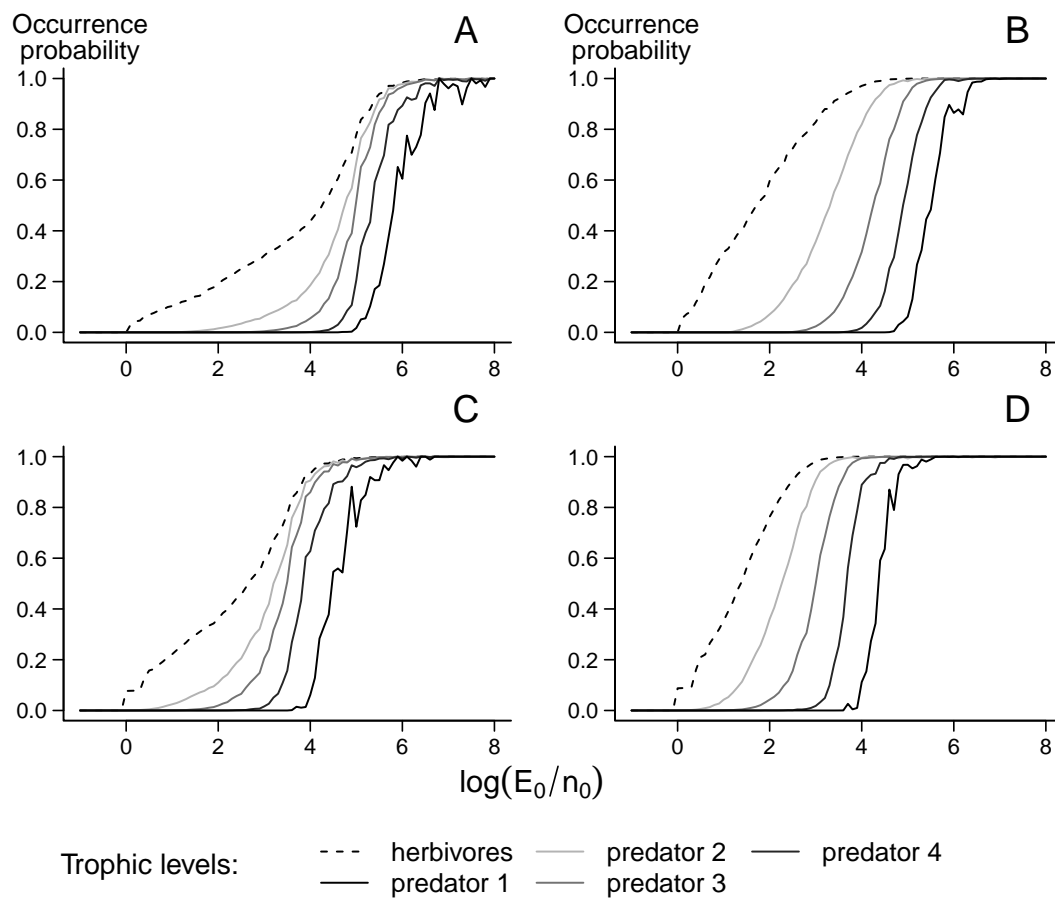


Figure 4.2: **Species occurrence probability along an energy gradient, grouped by trophic level; herbivores, or by their shortest path linking to a herbivore species.** The top right letters indicates the scenarios.

found within scenario B (4.3). Taken together, these results highlight the importance of network structure. Here, we outline the importance of colonization, where the shortest path between herbivore and predator is extremely important especially given it strongly impacts the cost of a species; suggesting that this is the cause of the main differences among random and cost-based extinction rates.

The average species body mass increased with energy under scenarios A and B, but decreased under scenarios C and D. The reduction of MVP with body mass would therefore allow for larger species to maintain on small island (4.4 A-D). We see a similar trend for the number of prey species, which increases with energy in scenarios A and B, while it decreases in scenarios C and D (4.4 E-H). Ultimately, generalist species have a much higher rate of successful colonization under low energy in scenarios C and D. When the MVP is smaller, larger species can colonize and their probability of success is higher as they have more preys in average (due to the use of the niche model and body mass as the niche axis ??). Finally, under scenarios A and B small species with several predators colonize low energy islands while under scenarios C and D the same species colonize the island with higher energy availability as they are preceded by larger and more generalist species (4.4 I-L). Consequently, the number of interactions per species increases faster under scenarios C and D than under scenarios A and B (4.4 M-P).

4.7 Discussion

Here, we formulated the classical TIB in terms of first energetic principles, which turns the well known colonization/extinction dynamics into a successional assembly process, beginning with herbivores and then builds upon the food web (until top predators).

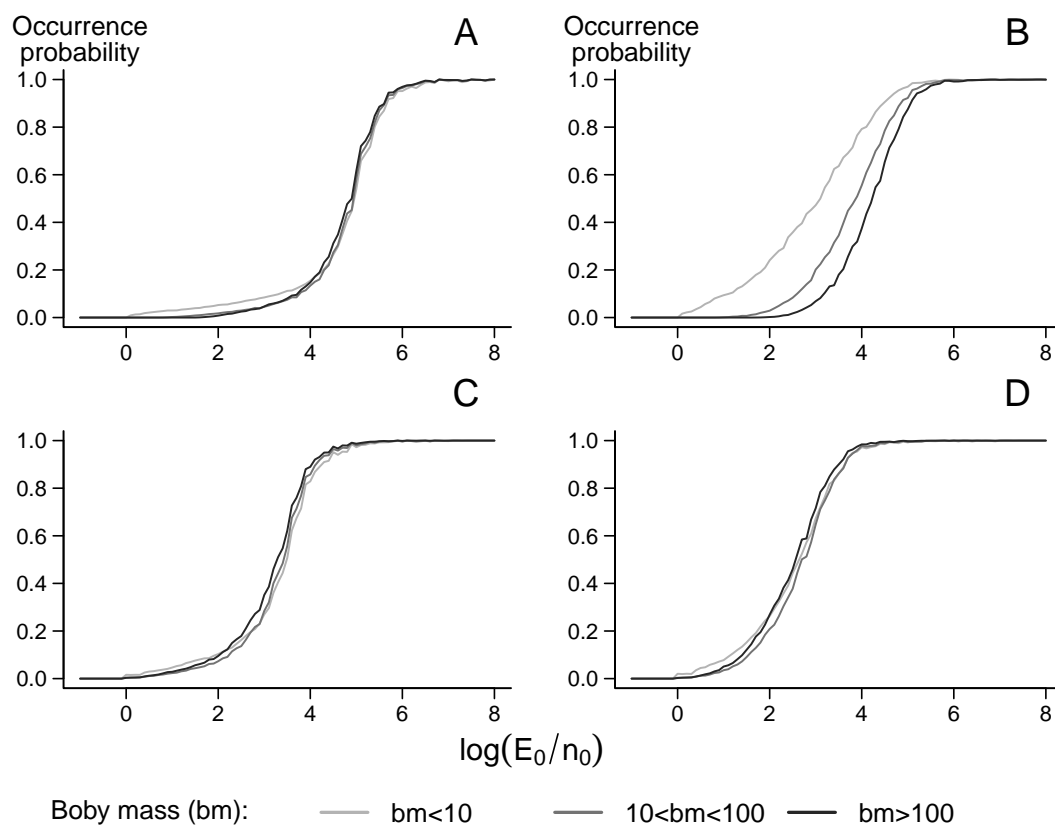


Figure 4.3: **Species occurrence probability by energy gradient grouped by body mass class.** The top right letters indicates the scenarios.

Here, we link species extinctions to their energy requirements, thus making extinctions more mechanistic than in the classical TIB. Moreover, species are no longer ecologically equivalent as in TIB (?), where they now have a body mass, trophic level, and specific interactions. Based on these characteristics, we compute a cost association to the MVP on an island. Therefore, an insular community becomes an assemblage of a population that should not exceed a given amount of energy. Although the energy allocation we use here is simple, we believe it provides a significant step forward towards implementing the interactions of community dynamics into biogeography theory. The next step would be to integrate more realistic energy partitioning (?), while also taking into account variations of population caused by bottom up and top-down effects we introduced here (??). Ultimately, the downscaling from species to populations we begin here, will allow for the integration of the principles of population dynamics as well as concepts of community ecology (*i.e.*, stability of a community ?). Furthermore, this would lead us down the path paved by MacArthur and Wilson 50 years ago: “[...] biogeography appears to us to have developed to the extent that it can be reformulated in the terms of the first principles of population ecology and genetics?” (? , p.183).

The formulation of the TIB in terms of energy offers a great opportunity to integrate to a mechanistic biogeography, one which strongly eases the integration of various ecological processes. For example, here we show how three ecological necessities are continually overlooked in classical biogeographical theories: (1)- the necessity for a minimum population in order to sustain presence on an island, (2)- the necessity of finding resource/prey to maintain this population, and (3)- the difference in energy uptake due to body mass and trophic level variations among species. Based on this integration, we showed how the structure of the community is constrained by the energy availability to the species in question on islands. We also point out that the structure of the food webs, which determines the energy requirements of the whole community,

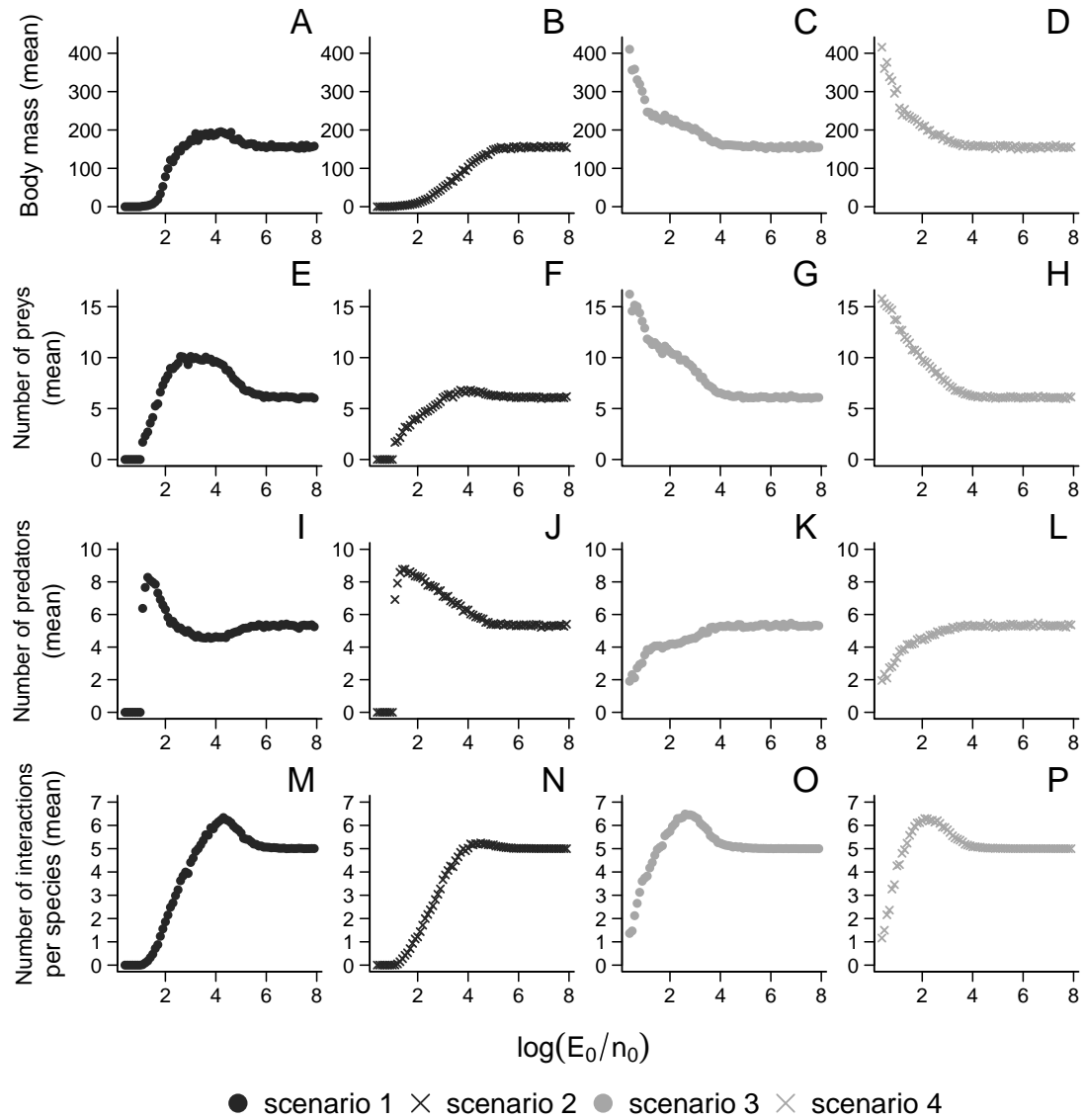


Figure 4.4: **Average body mass, number of preys, predators of predator species and the average number of interactions per species on the island** in regards to the energy gradient for four scenarios: (A) A,E,I, (B) B, F, J, M (C) C, G, K, O and (D) D, H, L, P.

controls the species richness and the characteristics of species presence of the island.

In our study, we study the SER under two different scenarios considering the MVP either decreased with the body mass (C and D) or did not (A and B). The results were qualitatively similar, however, were also quantitatively different. In our first case, we find that larger species can colonize earlier – leading to increases in species richness for the island. Where having a lower MVP for larger species can allow for energy compensation costs associated with the maintenance of a larger body. Furthermore, this could explain the “energetic equivalence rules” (?) - species have an equal share of energy allocated irrespective to their body mass (??). In our case, this holds true as long as the island is “saturated” with species. Remarkably, under this condition, the relationship between the logarithm of species richness and the logarithm of available energy (4.1 B) is linear, which is consistent with empirical observations (see Fig 3. and Fig 4. in ?). Over evolutionary scales, these observation would also be consistent with the Ecological Limits hypothesis to diversification dynamics (?).

Our findings suggest that there is a clear relationship between the trophic status of species and its occurrence probability: species of higher trophic levels, irrespective of their body mass, colonize islands with higher energy. This indicates that the food-chain length is limited by the energy availability. Remarkably, a comparable relationship has already been observed by ?, who has found a correlation among maximal food-chain length and lake size, which is consistent with the idea that higher trophic level predators need larger areas that are accompanied with sufficient amount of prey species in order to survive in these lakes (larger area = more potential energy to consume).

Using our framework, we showed that under scenarios C and D (for which MVP decreases with the body mass), generalist species are the first to colonize the island. Re-

markably, the early colonization of generalist species has been highlighted by ? when they revisited the classical study by ?. Also, reading Fig 4.4 G-H in the direction of a loss of energy, we observe an increase in generalist species similarly found by ?. The classical TIB has provided a theoretical basis for the SAR (?), that has been largely used in conservation practices to help determine the size of protected areas (?). However, as in the TIB, the SAR ignores the identity of species, there is currently no theory integrating the position of a species within the network it is embedded to predict its distribution. The role of a species in the network must be of first importance to help better understand the consequences of its loss (?). The model we described here offers a testable hypothesis regarding the role of the species that are expected to go extinct first, beginning with the highest trophic level cascading to the lowest.

Finally, our framework is an opportunity to review the way we build SDMs. Here, our model describes the food web dynamic associated to a patch of primary producers. Fortunately, that primary production have been proven highly correlated to climate variables (???) and are readily quantifiable using indices such as the normalized difference vegetation index (NDVI, ?). Here, such a quantity is not only a proxy to assess the diversity of primary producers and their productivity, but also a potential metric of the species richness (?), and a predictor for network structure (*e.g.* a predictor of the food chain length for instance). Current species distribution models are based on the correlations between abiotic factors and their occurrence probabilities. It is described as a species abiotic niche, while its presence could be best explained by the presence of its preys (see chapter 4). Therefore, focusing on the prediction of the distribution of primary producers while using information gathered concerning the associated energy that they could provide to higher trophic levels provides a promising avenue to enhance our understandings of tomorrow's biodiversity.

CONCLUSION GÉNÉRALE

Interactions écologiques et répartition géographique des espèces

Des modèles théoriques à développer

Dans son appel pour un renouvellement de la théorie de la biogéographie des îles, Mark Lomolino soulignait le besoin d'intégrer davantage de mécanismes écologiques et évolutifs autour des trois processus fondamentaux de la biogéographie : colonisation, extinction et évolution (?). Au chapitre 1, je me suis confronté à ce problème en proposant une démarche visant à incorporer les interactions et les contraintes abiotiques dans la TIB. J'ai proposé d'utiliser un cadre mathématique général (détaillé à l'annexe 2) reliant ces facteurs aux probabilités d'extinction et de colonisation. Cette approche est simple et prometteuse, je la considère aujourd'hui comme une extension de la TIB en ce sens que la théorie classique devient un cas particulier pour lequel les interactions et les contraintes abiotiques n'ont pas d'influence sur les taux de colonisation et d'extinction.

Le modèle proposé au chapitre 1 souligne le potentiel d'une approche communauté-centrée dans laquelle le raisonnement repose sur des probabilités de transition entre différents assemblages (?) plutôt que sur le traitement indépendant de la présence de plusieurs espèces. Je pense que cette idée était aussi en germe dès 1972 dans le livre de MacArthur, mais sans avoir été mathématiquement complètement formalisée (?). Partir des assemblages pour comprendre la distribution des espèces pose néanmoins un problème majeur : le nombre de communautés à envisager devient rapidement très élevé (pour n espèces, ce sont 2^n assemblages possibles), ce qui limite la mise en ap-

plication du modèle développé au chapitre 1 dans sa formulation actuelle. Cela étant dit, il est possible que des moyens émergent pour en réduire la complexité et qu'il soit progressivement transformé en une méthode d'inférence statistique efficace.

En me confrontant à l'incorporation des interactions écologiques dans la TIB (chapitre 1), je me suis aperçu à quel point il était délicat de construire des modèles simples et élégants expliquant un grand nombre de faits. En conséquence, je ne suis pas étonné que la TIB soit toujours abondamment utilisée comme point de départ de nombreuses études (?) et cela en dépit de ces défauts reconnus dans la réédition de 2001 de *The Theory of Island Biogeography* par Wilson lui-même : « *The flaws of the book lie in its oversimplification and incompleteness, which are endemic to most efforts at theory and synthesis.* »

L'objet *interactions* n'est pas simple à manipuler ; à l'échelle de la communauté, les interactions ne peuvent pas être traitées isolément : elles forment des réseaux. Un champ entier de la mathématique est dédié à l'étude de tels objets qui sont appelés graphes. La théorie qui traite de ces objets est utilisée dans différents domaines pour appréhender les réseaux, qu'ils soient sociaux ou neuronaux. Les neurosciences d'ailleurs, en cherchant à expliquer la cognition à partir des réseaux neuronaux, pointent, comme l'écologie, les difficultés à comprendre les systèmes caractérisés par l'interdépendance de leurs unités (?). Avec de tels objets à modéliser, il n'est pas étonnant de voir des questions importantes de l'écologie traitées par des mathématiques assez pointues (??).

En parallèle des questionnements très précis, il me semble également important que des réflexions soient menées pour aller vers des approches simplifiées et davantage intégratives. En forçant un peu le trait, en écologie nous avons d'un côté des modèles qui, avec très peu de populations considérées, engendrent des dynamiques complexes

voir chaotiques (dont l'existence est validée expérimentalement ??) et de l'autre des modèles comme celui de la TIB qui, avec une équation différentielle simple, propose une vision profonde de la biogéographie (?). J'ajoute que je suis convaincu qu'il est tout aussi pertinent d'essayer de partir de l'échelle la plus large pour aller vers des échelles plus petites que de mener la démarche inverse. Il est, par ailleurs, possible que les objets finaux à prédire, c'est-à-dire d'un côté l'abondance relative de populations en interaction et de l'autre la composition spécifique sur des larges échelles spatiotemporelles, ne puissent être prédits de la même façon, comme le suggère ?, ce qui serait une forme de *rupture de symétrie* (au sens de ?). Quoi qu'il en soit, c'est bien en essayant d'utiliser une approche simplifiée mais plus intégrative que j'ai réussi à mieux cerner le fait qu'elles pouvaient être les traces laissées par les interactions écologiques sur les distributions d'espèces.

Des théories pour mieux appréhender les données de cooccurrence

Le chapitre 2 bien que théorique, est une avancée importante en direction de l'application de mes réflexions à des données empiriques. Il y est question, en effet, d'étudier des données de cooccurrence pour lesquels les réseaux écologiques sont connus. Les données d'occurrence constituent une source de réflexion centrale pour les biogéographes car elles concentrent des efforts de développements méthodologiques importants dans le domaine (???). Les données de cooccurrence sont issues de la considération simultanée de données d'occurrence de plusieurs espèces sur un ensemble de sites dispersés le long d'un gradient géographique. Exploiter ces données permet, par exemple, d'envisager la structure des assemblages de demain (?). En proposant une réflexion relative à l'impact des interactions écologiques sur les données de cooccurrence, j'ai essayé d'améliorer la compréhension de la nature de l'information que pouvaient contenir les

données de cooccurrence. Ce travail est crucial pour orienter le développement des outils avec lesquels les scénarios de changement de la biodiversité sont établis (?). De manière générale, il s'agit de comprendre le lien qu'il existe entre la distribution d'une espèce et sa niche hutchinsonienne (??). En utilisant un modèle basée sur des probabilités simples et la version trophique de la TIB (?), j'ai découvert des attentes théoriques précises sur les données de cooccurrence et j'ai montré que l'empreinte des interactions écologiques sur les données de cooccurrence n'est pas appréciable, notamment lorsque les interactions sont nombreuses.

L'article présenté au chapitre 3 propose de tester la théorie du chapitre 2. En analysant des données de cooccurrence pour des systèmes dont les interactions étaient documentées, j'ai montré que celles-ci laissent des traces visibles dans les données statiques de cooccurrence. La détection de signaux de cooccurrence imputables aux liens écologiques liant les espèces n'est cependant possible que sous certaines conditions, à savoir lorsque les espèces interagissent directement, soit quand le nombre d'interactions est limité. De manière cohérente, la distribution d'un prédateur spécialiste est intimement liée à celle de sa proie alors que dans le cas d'un prédateur généraliste, celui-ci voit sa distribution partiellement corrélée avec un grand nombre de distributions, celles de ses proies. Dans ce dernier cas, il devient difficile de déceler un signal clair dans la cooccurrence d'un généraliste avec l'une de ces proies. Un signal peut néanmoins exister lorsque l'on examine la corrélation de la distribution de ce prédateur et la répartition géographique jointe de l'ensemble de ses proies.

En travaillant sur les cooccurrences avec des données de distribution d'espèces en interaction, j'ai aussi pointé du doigt un problème important concernant l'application des SDMs pour des réseaux écologiques. La cooccurrence forte de deux espèces est souvent interprétée comme le témoin de la similarité de leurs besoins physiologiques,

ce qui justifie d'utiliser des projections à l'échelle de l'espèce pour prédire des communautés (??). Cela dit, en partant de ce principe et en optant pour l'utilisation d'un espace explicatif fait exclusivement de variables abiotiques, le risque est très fort d'attribuer la cooccurrence aux variables abiotiques alors que la cause profonde de cette cooccurrence pourrait être l'interaction entre les espèces. Il en est de même en inférant la distribution d'une espèce à partir de la distribution des autres espèces du réseau : en nous concentrons sur les variables biotiques, nous occulterions le signal abiotique. Ce problème du choix des variables est illustré au chapitre 3 dans lequel j'ai montré que l'utilisation de SDMs affaiblissait considérablement le signal observé sur les données de cooccurrence brutes (pour lequel les facteurs abiotiques ne sont pas intégrés). Au lieu de conclure que la cooccurrence est contrainte par les variables pédoclimatiques, j'ai cherché la cohérence de cet affaiblissement et constaté que même les associations les plus fortes (pour les prédateurs spécialistes et leurs proies) disparaissaient et qu'un modèle simple basé sur la présence de proie était plus performant que certains SDMs. Ainsi, l'association entre deux espèces interagissant n'est pas nécessairement capturée de manière adéquate par les SDMs, ce qui jette le doute sur la qualité des prédictions basées sur les SDMs. Pour remédier à cela, une fusion méthodique des deux types d'information est nécessaire (??). Dans le cas précis d'un prédateur et de ses proies¹, il y a un lien évident entre les distributions : le prédateur est nécessairement limité par la distribution conjointe de ces proies (??). La reconnaissance des réalités des interactions biotiques auxquelles les espèces sont soumises doit être au coeur d'un renouvellement des approches pour prédire des espèces en réseaux (?).

Tout au long des chapitres 2 et 3, j'ai souligné l'intérêt des développements théoriques pour mieux comprendre les données empiriques. En partant initialement de la ques-

1. C'est aussi valable pour un pollinisateur et les plantes que pollinise ce dernier, ou encore pour un parasite et ses hôtes.

tion *est-ce que les espèces qui interagissent cooccurrent différemment que celles qui n'interagissent pas ?*, j'ai compris qu'il n'y avait pas de réponse tranchée, mais plutôt une réponse qui dépendait de la nature du réseau. Ce résultat sera, je pense, très utile pour amener une lumière nouvelle sur le débat qui anime la communauté des biogéographes autour de la question du rôle des interactions dans la distribution des espèces aux larges échelles spatiales (??). Bien que dans les dernières années des avancements méthodologiques significatifs ont été fait avec l'essor des JSDMs (???), il faut prolonger l'effort pour intégrer davantage d'informations biologiques dans les SDMs car l'intégration systématique des cooccurrences n'est peut-être pas suffisante. Je n'ai été capable d'interpréter de manière cohérente les données de cooccurrence qu'avec l'apport d'une source d'information extérieure : celles des réseaux, grâce à laquelle j'ai pu montré que l'importance des interactions n'était pas seulement une question d'échelle spatiale (??), mais aussi un problème de topologie du réseau d'interaction du système étudié.

Vers une biogéographie prédictive ?

Les défis à relever dans un monde en changement

Érosion de la biodiversité, extinctions de masse, perte de services écosystémiques, la liste est longue des bouleversements biologiques de notre époque. Le cinquième rapport d'évaluation du Groupe d'experts Intergouvernemental sur l'Evolution du Climat (le GIEC²) souligne qu'il y a très peu de doutes sur le lien entre les activités humaines

2. Tous les documents émis par le GIEC sont disponibles en ligne sur le site du GIEC <https://ipcc.ch> et certaines initiatives rassemblent en des documents synthétiques leurs conclusions, voir par exemple <http://leclimatchange.fr.>

et ces changements. Pour y faire face, les gouvernements d'un grand nombre de pays doivent agir de concert et, ces dernières années nous donnent des raisons d'espérer. En janvier 2013 se tenait la première réunion de la Plate-forme Intergouvernementale scientifique et politique sur la Biodiversité et les Services Écosystémiques (l'IPBES³) qui se veut être un acteur privilégié de la sauvegarde de la biodiversité à l'échelle mondiale (?). En décembre 2015, l'accord de Paris sur le climat signé formellement par 175 pays lors de la 21^{ème} conférence des parties (COP21), et ratifié par la Chine et les États-Unis au début du mois de septembre 2016, vise des réductions considérables de gaz à effet de serre à l'échelle mondiale. Ajoutées à ces initiatives internationales, des initiatives nationales attestent de la prise de conscience des enjeux de la biodiversité. Ainsi, en France, le projet de loi pour la reconquête de la biodiversité, de la nature et des paysages a été adopté le 20 juillet 2016 et prévoit la mise en place de l'agence française pour la biodiversité. En dépit des défis écologiques à relever, il faut également reconnaître les problèmes liés à l'état d'avancements des connaissances en écologie. La jeunesse de l'écologie, le caractère partiel des théories actuelles est, en effet, un obstacle considérable : comment la communauté scientifique peut-elle envoyer un message fort et cohérent aux décideurs politiques quand elle est traversée par des débats loin d'être résolus ? Pour rebondir sur un exemple donné en introduction : si l'érable à sucre ne peut migrer assez vite pour suivre les conditions favorables, nous pouvons artificiellement déplacer des arbres pour pallier ce problème, c'est ce qu'on appelle la migration assistée. Mais comment anticiper les conséquences de l'utilisation massive d'une telle pratique ? Un immense effort de synthèse sur les différentes positions du débat scientifique devrait alors être mené afin d'intégrer les connaissances dans le domaine pour mettre en place un cadre législatif cohérent entourant cette pratique,

3. L'IPBES a été construit sur le même modèle que le GIEC en ayant pourtant un mandat plus large : en plus de faire une synthèse des connaissances en vue de guider les décideurs politiques, elle compte parmi ses prérogatives la production de nouvelles connaissances (?).

un défi considérable (?).

La crise d'extinction d'espèces que nous traversons actuellement (?) est un moment de restructuration intense des communautés. Ce phénomène est particulièrement visible dans les aires urbaines où les communautés natives sont remplacées par des communautés adaptées aux milieux anthropiques. Depuis la seconde partie du XIX^{ème} siècle, la ville de New York a ainsi perdu 578 espèces natives de plantes vasculaires ; en revanche, la ville a gagné 411 espèces non-natives (?). Ces communautés urbaines sont, par ailleurs, plus homogènes à travers les États-Unis que les communautés endémiques (?). De manière générale, les espèces spécialistes subissent un recul plus marqué que celui des espèces généralistes, ce qui conduit à une homogénéisation fonctionnelle qui s'ajoute à une homogénéisation taxonomique (?). Face à ces restructurations, deux attitudes opposées existent : d'un côté la restauration des communautés qui postule l'existence d'une intégrité écologique des systèmes naturels (?) et de l'autre la primauté du bon fonctionnement des systèmes écologiques, quelle qu'en soit la composition, ce que résume la formule *Don't judge a species on their origin*, utilisé par Mark Davis pour intituler un article développant cette idée (?).

Alors, doit-on lutter contre le frelon asiatique ? Au-delà d'un positionnement éthique, il y a une forme de spéculation dans la réponse, car il faut répondre à une autre interrogation : quelles seraient les conséquences de l'inaction ? Si nous laissons la population de frelon grandir, il est probable qu'une partie de l'entomofaune européenne disparaîtra. Il est cependant également probable que les abeilles domestiques puissent adopter des stratégies de défense plus efficaces, ce qui stabiliserait la population du frelon. Les théories actuelles ne nous permettent pas de savoir avec exactitude lequel des scénarios est le plus probable, et nous constatons que dans les systèmes vivants, la surprise est généralement la règle. Ainsi, dans une étude très récente sur le Diamant mandarin (*Tae-*

niopygia guttata), un oiseau commun du centre de l'Australie, Mylene M. Mariette et Katherine L. Buchanan, ont montré qu'au-dessus de 26°C le mâle, seul adulte dans le nid, produit un chant particulier destiné à ses oeufs induisant une réduction significative de la taille des futurs oiseaux adultes dont la fertilité est augmentée (?). Face à cette étude, on mesure les difficultés pour intégrer l'ensemble des réponses possibles des espèces face à une perturbation. Leur connaissance exhaustive semble pourtant nécessaire pour prédire de manière adéquate la réponse des écosystèmes aux changements actuels. Les systèmes biologiques sont d'une haute complexité et nos connaissances sont très incomplètes. Nous devons reconnaître nos carences théoriques mais rester ambitieux pour construire une écologie prédictive (?).

Des règles en écologie ?

Dans un article publié en 1999, John H. Lawton pose la question suivante : y a-t-il des lois générales en écologie ? Lawton y relève que les systèmes écologiques sont contraints par des lois fondamentales, incluant notamment les principes de la thermodynamique et l'évolution. Cet auteur place au niveau de ces lois fondamentales les interactions écologiques qu'il présente comme une *observation* (selon ces mots, ?). En dehors de ces lois, Lawton argumente qu'il n'y a pas de règle universelle, mais plutôt des règles applicables à un nombre restreint systèmes écologiques. Cette idée est résumée par la formule suivante (p.178) : « *The universal laws do not allow us to predict the existence of kangaroos.* » Dans son article, Lawton dresse un portrait d'une écologie dominée par la contingence et suggère d'aller chercher des règles dans la macro-écologie. En dépit de la pertinence des arguments développés, Lawton oublie, à mon avis, de répondre à une question centrale : quels sont les objets sur lesquels les prédictions doivent porter ? Cherchons-nous vraiment à prédire l'existence des kangourous ? Peut-être que

l'objet de la prédiction est plutôt la probabilité de l'émergence, en Australie, dans un contexte biotique et abiotique donné, d'un marsupial herbivore d'une taille donnée. Il me semble que présenté ainsi, l'objectif de prédictions fiables semble bien plus réalisable que celui présenté par Lawton. La contingence des systèmes biologiques nous indique qu'il faut un formalisme qui envisage des possibilités et non pas la supposition d'existence de règles dans ces systèmes. Ainsi, en suivant le physicien James C. Maxwell, le second principe de la thermodynamique nous indique que lorsque l'on met en présence de l'eau chaude et de l'eau froide initialement séparées, obtenir de l'eau tiède après un certain temps est simplement un évènement très probable.

Les communautés (au sens d'ensemble structuré d'espèce) ont vraisemblablement des fonctionnements et des propriétés qui leur sont propres, néanmoins, il y a aussi des rôles écologiques similaires entre différents réseaux. Les prédateurs en haut de la chaîne trophique par exemple : ces espèces des niveaux trophiques les plus élevés structurent profondément les communautés, leur extinction ont de fortes influences sur l'ensemble de la communautés (????). En plus de relever les similarités, il faut essayer de comprendre les différences de fonctionnement au regard des différences de composition spécifique. De même que la loi universelle de la gravitation dépend de la masse des deux corps en interaction et de la distance qui les séparent, les lois de l'écologie pourraient se décliner aisément en fonction de la composition spécifique ou même de la combinaison des caractéristiques de ces espèces (de traits fonctionnels). C'est peut-être, à la suite de MacArthur et Wilson (?), dans la réduction des espèces à quelques propriétés que des règles émergeront (??). Une des premières propriétés est tout simplement la masse des espèces dont dépendent un grand nombre de propriétés écologiques (?). Le lien avec les lois de la thermodynamique est évident : les êtres vivants, pour se maintenir (et donc respecter les principes thermodynamiques) ont des besoins énergétiques qui augmentent avec la taille. La taille impose donc des contraintes essentielles pour

construire des approches cohérentes (ce que j'ai essayé de faire au chapitre 4) sans lesquelles la recherche de règle éventuelle serait délicate. Il est difficile d'argumenter que l'écologie a des règles universelles, mais il est encore plus délicat de démontrer qu'il n'y en a pas (et la démonstration formelle de ce principe serait une révolution). Aussi, constatant l'absence actuelle de lois universelles en écologie, je dirais que nous avons besoin de trouver les chemins par lesquelles elles pourraient éventuellement être révélées, de même que pour trouver des traces d'interactions écologiques dans des données statiques de cooccurrence, c'est la méthode qu'il m'a d'abord trouver en premier.

Une théorie intégrative de la biogéographie

Malgré les avancements théoriques des dernières années en biogéographie, l'appel lancé par Lomolino en 2000 reste d'actualité (?). L'effort d'intégration ordonné de concepts clés issus de différents champs de l'écologie dans une théorie intégrative de la biogéographie doit être poursuivi (?). Avec l'éclairage que je propose sur la relation entre la distribution des espèces et les interactions écologiques, j'espère que mes travaux participeront à avancer vers une telle théorie. Comme je l'ai indiqué tout au long de ma thèse, ce n'est pas tant la liste des mécanismes qui pose problème, mais bien leur articulation cohérente sans laquelle on ne peut connaître leur importance relative, ce qui freine l'émergence des méthodes plus cohérentes pour prédire la répartition géographique des espèces. Néanmoins, comme le relevaient Miguel B. Araújo et Carsten Rahbek, pendant que la recherche progresse, il faut essayer de tirer le meilleur parti de l'état actuel des connaissances (?) :

We will never be able to predict the future with accuracy, but we need a strategy for using existing knowledge and bioclimatic modeling to improve

understanding of the likely effects of future climate on biodiversity.

Les écosystèmes sont des objets complexes, des réseaux dynamiques d'objets tout aussi complexes. Prédire la composition précise d'un écosystème donné semble un objectif encore très éloigné, mais estimer le temps de survie d'une espèce dans un contexte biotique et abiotique est à notre portée. Pour y arriver, pour avoir une vision cohérente de la dynamique des écosystèmes sur des échelles de temps plus longues, on ne peut négliger ni les processus écologiques ni les processus évolutifs : nous devons construire des approches intégratives. Pour cela, nous devons chercher et trouver le liant entre tous ces processus et examiner soigneusement les contraintes auxquelles les espèces n'échappent pas. Les contraintes énergétiques (voir chapitre 4) en sont un bon exemple. Même l'évolution, toute pourvoyeuse de contingence qu'elle soit, apporte son lot de contraintes : elle met en jeu une variation qui engendre un différentiel démographique dans un contexte biotique et abiotique donné. Ainsi, comme le notait Wallace au troisième chapitre de son livre *Island Life* : « *In the first place we must remember that new species can only be formed when and where there is room for them* ».

ANNEXE A

COMMENT LA BIODIVERSITÉ S'INSTALLE EN TERRITOIRES ISOLÉS

L'article qui suit est le fruit de ma rencontre avec Philippe Etchécopar, membre du comité éditorial du journal *Accroamth*. Ce journal de vulgarisation mathématique est destiné aux étudiants et enseignants en mathématiques du Cégep⁴. Dans cet article, je reprends les bases du modèle de la TIB et je montre comment en partant d'une équation simple sur les probabilités, il est possible d'obtenir une équation différentielle déterministe. L'article est accessible en ligne⁵

5.8 Préambule

La biogéographie est la science qui s'interroge sur les causes de la répartition de la biodiversité dans les différentes parties du globe (voir encadré 1). Le modèle déterministe de MacArthur et Wilson décrit l'évolution de la richesse spécifique sur les îles vers

4. Le cégep est la dernière étape avant l'entrée à l'université pour ceux désirant poursuivre leurs études.

5. <http://accromath.uqam.ca/2014/02/la-biodiversite-en-territoires-isoles>

un équilibre, mais la migration des espèces vers des îles et leur extinction potentielle sont des phénomènes aléatoires. Comment un modèle déterministe peut-il renfermer ce hasard ? C'est ce que nous allons voir.

5.9 Le modèle de MacArthur et Wilson

Un des modèles les plus puissants en biogéographie est celui proposé par MacArthur et Wilson dans leur passionnante théorie de la biogéographie des îles. Ces deux illustres prédécesseurs nous ont livré un paradigme simple et puissant pour envisager la construction de la biodiversité sur une île. Notons dès maintenant que l'île n'est pas nécessairement un monticule de sable fin au milieu de l'océan mais plutôt -et plus généralement- un territoire isolé. Le modèle permet de comprendre l'impact des capacités de dispersion et de survie des espèces sur la biodiversité S de l'île considérée. Considérons que cette île est accessible depuis un continent qui présente un ensemble constant d'espèces P . Les P espèces sont potentiellement celles que nous retrouverons sur l'île, nous avons donc $S \leq P$. À un temps t donné, l'ensemble des S espèces de l'île peuvent s'éteindre avec un taux e . De même, les $P - S$ espèces du continent (absentes de l'île) peuvent coloniser l'île avec un taux c . En langage mathématique, nous obtenons l'équation différentielle déterministe suivante pour décrire l'évolution temporelle de S :

$$\begin{aligned} \frac{dS}{dt} &= c(P - S) - eS \\ &= cP - (c + e)S \end{aligned} \tag{5.1}$$

Cette équation est linéaire, nous pouvons la résoudre facilement par la méthode de la variation de la constante (essayez puis regardez l'encadré 2 !). Pour les valeurs arbi-

traies $c = 0.2$, $e = 0.1$ et la condition initiale suivante : à $t = 0$, l'île est inhabité ($S(0) = 0$), nous représentons graphiquement la solution de l'équation à la figure 5.5. Pour un temps infini, la biodiversité tend vers la valeur : $S_{eq} = P\left(\frac{c}{c+e}\right)$, interprétée comme le nombre d'espèces provenant du continent et présentes sur l'île à l'équilibre.

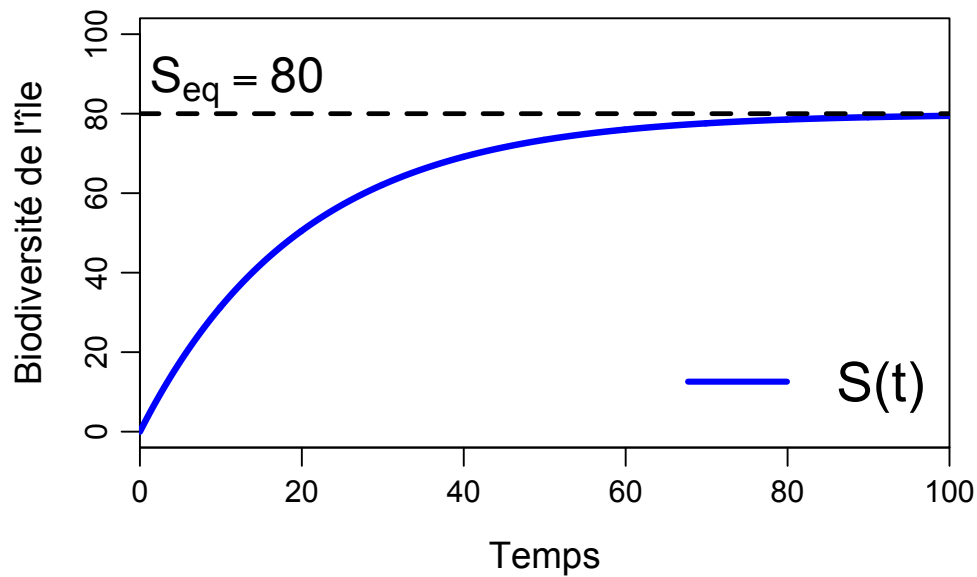


Figure 5.5: **Représentation graphique de la solution de l'équation différentielle (5.1)** pour les conditions suivantes : $c = 0.04$, $e = 0.01$ et $S(0) = 0$. La ligne en pointillés représente la valeur de la biodiversité S_{eq} que nous obtenons à un temps infini.

5.10 Comment un modèle déterministe émerge-t-il de phénomènes aléatoires ?

5.10.1 Modèle stochastique ?

Il est très intéressant de réaliser que l'équation (5.1), d'apparence déterministe, renferme un modèle stochastique. "Stochastique" est un terme signifiant aléatoire, avec

une part de hasard. Un modèle stochastique est donc un modèle dont l'issue contient une part de hasard ; deux réalisations du modèle ne donneront donc pas nécessairement le même résultat. Pour comprendre où se cache le hasard dans le modèle que nous étudions, nous avons besoin d'objets mathématiques particuliers appartenant au domaine des probabilités : variables aléatoires et processus stochastiques (reportez-vous à l'encadré 3).

5.10.2 Les objets mathématiques requis

Comment savoir combien d'espèces sont présentes sur l'île ? Il suffit de compter 1 pour chaque espèce présente sur l'île. Nous introduisons donc X_i , la variable aléatoire de présence sur l'île de l'espèce i choisie parmi les P espèces du continent. X_i est égale à 0 si l'espèce n'est pas sur l'île et 1 si elle est sur l'île. C'est une variable aléatoire de Bernoulli (encadré 4). Ainsi, le nombre d'espèces présentes sur l'île, S , est égale à la somme des X_i . Nous allons encore plus loin en enregistrant ces valeurs au cours du temps. C'est ainsi que nous définissons le processus stochastique $X_{i,t>0}$, qui n'est autre qu'une succession de 1 et de 0 indiquant à chaque instant si oui ou non l'espèce i est sur l'île. A priori, les suites $X_{i,t}$ ne sont pas prévisibles, ce qui n'exclut pas que la fonction S présente, en moyenne, un comportement tout à fait régulier.

5.10.3 Les briques élémentaires du modèle stochastique

Au temps $t = 0$, l'espèce i n'est pas sur l'île. Pour construire la suite de son histoire, il nous faut une règle simple pour décrire l'évolution de sa présence entre deux pas de temps très proches. Le modèle de MacArthur et Wilson postule que si l'espèce consi-

dérée est sur l'île au temps t , elle s'éteint avec un taux e ; si elle est sur le continent, elle le colonise avec un taux c . Pour simplifier, nous supposons que les taux d'extinction et de colonisation sont les mêmes pour toutes les espèces. Nous donnons maintenant une signification en terme de probabilité à ces taux : c est la probabilité de colonisation par unité de temps (on parle de densité du processus), e est la probabilité d'extinction par unité de temps. Ainsi, edt désigne la probabilité d'extinction pendant l'intervalle de temps dt . De même, cdt est la probabilité de colonisation pendant dt . En faisant appel aux probabilités conditionnelles (encadré 5), en supposant dt assez petit, nous différencions quatre cas :

$$\forall t \in \mathbb{R}^+, P(X_{i,t+dt} = 1 | X_{i,t} = 0) = cdt \quad (5.2)$$

$$P(X_{i,t+dt} = 0 | X_{i,t} = 1) = edt \quad (5.3)$$

$$P(X_{i,t+dt} = 0 | X_{i,t} = 0) = (1 - cdt) \quad (5.4)$$

$$P(X_{i,t+dt} = 1 | X_{i,t} = 1) = (1 - edt) \quad (5.5)$$

Il faut comprendre la première équation ainsi : sachant que l'espèce i était absente au temps t , la probabilité qu'elle soit sur l'île au temps $t + dt$ est égale à la probabilité qu'elle colonise l'île durant dt , c'est-à-dire cdt . Vous pourriez avancer que durant dt , une espèce peut coloniser, s'éteindre et re-coloniser et que nous n'en parlons pas ! Tous ces évènements sont en fait beaucoup moins probables et nous pouvons les ignorer complètement à la limite quand dt tend vers 0. Les trois autres équations s'interprètent avec un raisonnement similaire. Ces quatre équations sont les briques du modèle stochastique que nous construisons. Nous devons maintenant les assembler correctement.

5.10.4 Assemblons les briques !

L'assemblage consiste en l'utilisation de la formule des probabilités totales expliquée dans l'encadré 5. Nous pouvons en déduire la probabilité de présence de l'espèce i à l'instant $t + dt$:

$$\begin{aligned} P(X_{i,t+dt} = 1) &= P(X_{i,t+dt} = 1 | X_{i,t} = 0)P(X_{i,t} = 0) \\ &\quad + P(X_{i,t+dt} = 1 | X_{i,t} = 1)P(X_{i,t} = 1) \end{aligned} \quad (5.6)$$

Il s'agit d'une somme de deux termes couvrant toutes les possibilités puisque, à l'instant t , l'espèce i était soit absente de l'île soit présente, c'est un système complet d'événements (encadré 5). Si l'espèce i était absente (premier terme) à t , elle sera présente à $t + dt$ si elle colonise pendant dt (5.2). Si elle était déjà sur l'île à t , elle s'y maintient à condition de ne pas s'éteindre (5.5). En remarquant que $P(X_{i,t} = 0) = 1 - P(X_{i,t} = 1)$, nous pouvons écrire :

$$P(X_{i,t+dt} = 1) = cdt(1 - P(X_{i,t} = 1)) + (1 - edt)P(X_{i,t} = 1) \quad (5.7)$$

Pour simplifier l'écriture, $s_{i,t}$ représente $P(X_{i,t} = 1)$:

$$s_{i,t+dt} = cdt(1 - s_{i,t}) + (1 - edt)s_{i,t} \quad (5.8)$$

Divisons par dt :

$$\frac{s_{i,t+dt} - s_{i,t}}{dt} = c(1 - s_{i,t}) - es_{i,t} \quad (5.9)$$

Nous faisons alors tendre dt vers 0. Nous obtenons alors :

$$\frac{ds_i}{dt} = c(1 - s_i) - es_i \quad (5.10)$$

C'est l'équation que nous avons résolu dans l'encadré 1, avec $P = 1$:

$$s_i(t) = \frac{c}{c+e} (1 - \exp(-(c+e)t)) \quad (5.11)$$

5.10.5 Retrouvons le modèle classique

Nous avons presque retrouvé la solution de l'équation déterministe ; vous pourriez penser qu'il faut simplement multiplier par le nombre d'espèces P . L'idée est bonne, mais demande justification ! Nous allons effectivement considérer non pas une mais bien les P espèces du continent pour lesquelles l'équation (5.11) est valable. Nous allons alors définir un nouveau processus stochastique qui est simplement la somme des P processus de présence que nous supposons indépendants $X_{i,t>0}$, $Y_t > 0 = X_{1,t>0} + X_{2,t>0} + \dots + X_{P,t}$. Le problème est alors de connaître à tout instant t la probabilité d'avoir un nombre donné k d'espèces présentes sur l'île. À un instant donné t nous faisons une somme de variable aléatoires de Bernoulli. La somme de P variables de Bernoulli indépendantes est une variable aléatoire suivant une loi binomiale (encadré 5). Nous avons donc :

$$\forall t \in \mathbb{R}^+, P(Y_t = k) = \binom{P}{k} s(t)^k (1 - s(t))^{P-k} \quad (5.12)$$

De plus, on peut démontrer les résultats suivants :

$$\forall t \in \mathbb{R}^+, E(Y_t) = Ps(t) \quad (5.13)$$

$$\forall t \in \mathbb{R}^+, V(Y_t) = Ps(t)(1 - s(t)) \quad (5.14)$$

Grâce à ces résultats de probabilité, nous avons donc :

$$\forall t \in \mathbb{R}^+, E(Y_t) = P \frac{c}{c+e} (1 - \exp(-(c+e)t)) \quad (5.15)$$

Nous retrouvons la solution déterministe ! Ainsi, le modèle déterministe peut être considéré comme reflétant l'évolution de l'espérance des variables Y_t pour $t > 0$, c'est donc l'espérance du processus stochastique. Nous pouvons alors simuler le modèle stochastique grâce au modèle déterministe : à chaque instant t , il suffit de simuler une loi binomiale de paramètres P et $E(Y_t)$. Regardez la figure 5.6, il y a des différences fondamentales entre les deux approches. Le modèle déterministe donnera des valeurs continues et, pour des paramètres donnés, toujours le même résultat. De son côté, le modèle stochastique livrera des valeurs discrètes. De plus, deux simulations du modèle aléatoire ne donneront pas nécessairement les mêmes courbes.

5.11 Applications et perspectives

Bien que datant du milieu des années 1960, le modèle de MacArthur et Wilson demeure très précieux, notamment pour étudier les habitats fragmentés (des îles !) et prendre des décisions relatives à la conservation des espèces. Prenons un exemple : vous êtes le gestionnaire d'une nouvelle zone protégée que nous assimilerons à une île. Cette zone a longtemps été exploitée par l'homme de sorte qu'au temps $t = 0$ le nombre d'espèces

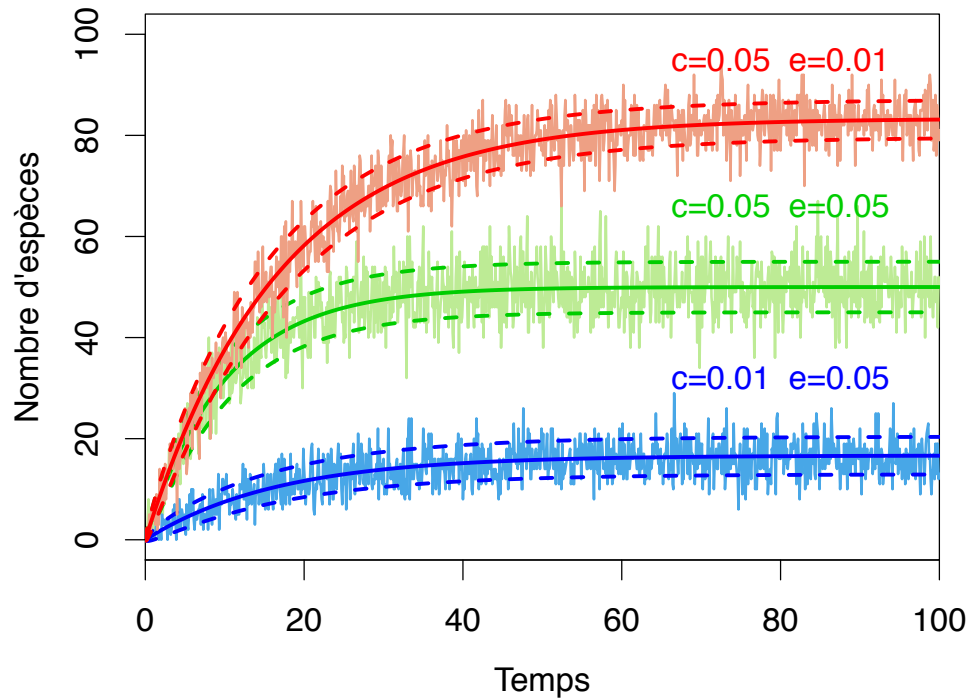


Figure 5.6: **Dynamique de la biodiversité de trois zones protégées.** Les caractéristiques de la zone protégée sont données par les valeurs de c et e . Les symboles rouges sont relatifs à une zone protégée de grande taille et facilement accessible, en vert, ils font référence à une petite zone facilement accessible ; enfin, en bleu, sont présentées les résultats pour une petite zone difficilement accessible. Les courbes striées en couleurs pastels sont les réalisations du modèle stochastique. Les courbes en traits pleins, font référence au modèle déterministe qui est également la moyenne du modèle stochastique. Enfin, les courbes en pointillés représentent la moyenne du modèle stochastique plus ou moins l'écart type.

est $S = 0$. Dans la région, la richesse en espèces est de $P = 100$, c'est-à-dire que vous pouvez rencontrer jusqu'à 100 espèces dans votre aire de gestion. À la suite de MacArthur et Wilson, nous allons supposer que la valeur du taux de colonisation c dépend de la difficulté d'accès. De plus, la survie d'une espèce sur cette zone dépend de sa superficie : ainsi plus l'aire de conservation est grande, moins les espèces s'éteignent ; e est donc plus faible pour les grandes aires de protection. Prenons trois situations :

- Grande zone et accès aisé : $e=0.01$ et $c=0.05$
- Petite zone et accès aisé : $e=0.05$ et $c=0.05$
- Petite zone et accès délicat : $e=0.05$ et $c=0.01$

Pour mettre en place un suivi et savoir comment se comporte votre réserve, il faut avoir une idée de la dynamique attendue ! Ce que nous permet le modèle étudié, que nous simulons pour les trois situations, avec les deux approches du modèle. Les résultats sont donnés par la figure 1. Cette simple démarche vous permet d'estimer l'évolution de la biodiversité au sein de votre réserve. La paramétrisation est plus compliquée dans les faits, mais l'idée est très utilisée pour apprécier la santé des réserves naturelles.

Dans cet exemple, le modèle stochastique n'a finalement que peu d'intérêt. Il est cependant très utile si vous êtes intéressé par la variance du modèle (que nous obtenons avec l'équation (5.14)). C'est aussi un bon point de départ pour les démarches statistiques, indispensables pour le suivi de votre réserve. De plus, l'approche stochastique révèle toute sa force pour enrichir le modèle. En partant de l'équation (5.1), si vous souhaitiez introduire un effet des espèces les unes sur les autres (les interactions entre espèces), vous auriez quelques difficultés. Avec le travail réalisé ici vous pouvez trouver et surtout justifier des approches fertiles en réfléchissant en terme de probabilité. Ces approches pourraient vous amener à résoudre des systèmes d'équations différentielles déterministes un peu plus compliqués !

Encadré 1 : la Biogéographie

La biogéographie est la science qui s'interroge sur les causes de l'agencement spatial des espèces à la surface de notre planète. Comment, par exemple, pouvons-nous expliquer la répartition et les différences entre les grands biomes terrestres ? Le premier élément de réponse est l'implication des facteurs climatiques. La température, l'humidité, les précipitations sont autant de variables qui contraignent les conditions d'existence des espèces qui composent et modèlent les paysages terrestres. C'est avec ces contraintes que les écologues élaborent des modèles dits "bioclimatiques" afin de décrire l'évolution de la biodiversité avec les contraintes climatiques de demain. Cependant, les facteurs qui expliquent la distribution des espèces ne sont pas toujours climatiques, les causes sont multiples. Les mouvements des espèces et leurs histoires évolutives sont ainsi des moteurs fondamentaux en biogéographie. Les capacités de dispersion des espèces leur donnent accès à de nouveaux territoires qui, avec le temps peut-être, les verront disparaître. C'est ainsi qu'un territoire voit sa composition en espèces évoluer, avec l'arrivée de nouvelles espèces et l'extinction d'autres. C'est autour de cette idée majeure que s'articule le modèle présenté.

Encadré 2

Il est facile de vérifier que la solution générale de l'équation homogène :

$$\frac{dS}{dt} = -(c + e)S$$

est donnée par :

$$\forall t \in \mathbb{R}^+, S(t) = C \exp(-(c + e)t)$$

où C est une constante, la solution de l'équation non homogène est alors de la forme :

$$S(t) = f(t) \exp(-(c + e)t)$$

En remplaçant dans l'équation dif-

férentielle on obtient :

$$f'(t) \exp(-(c + e)t) = cP$$

ce qui nous donne :

$$f'(t) = cP \exp((c + e)t)$$

En intégrant on obtient :

$$f(t) = P \frac{c}{c + e} \exp((c + e)t) + K$$

où K est une constante arbitraire.

$$S(t) = P \frac{c}{c + e} + K \exp(-(c + e)t)$$

Avec $S(0) = 0$, $K = -P \frac{c}{c + e}$, nous obtenons donc :

$$S(t) = P \frac{c}{c + e} (1 - \exp(-(c + e)t))$$

Encadré 3

Les **variables aléatoires** sont des fonctions définies sur l'ensemble des résultats possibles d'un événement aléatoire. Comme le résultat de l'événement est aléatoire, on associe à chaque valeur de la variable sa probabilité, de sorte que la somme des probabilités sur l'ensemble des valeurs possibles soit égale à 1. Dans notre exemple, X_i est une variable aléatoire, 0 et 1 sont ses valeurs, et nous nous intéressons à leur probabilité $P(X_i = 1)$ et $P(X_i = 0)$ avec $P(X_i = 1) + P(X_i = 0) = 1$.

Les **processus aléatoires ou stochastiques** sont des collections de variables aléatoires indexées par le temps t , c'est-à-dire ordonnées. Dans notre exemple, l'indexation est faite selon le temps t . Nous suivons ainsi l'évolution de X_i au cours du temps d'où la notation $X_{i,t>0}$. Attention ! $X_{i,t>0}$ est un processus aléatoire mais $X_{i,t}$ est l'une des variables aléatoires du processus.

Encadré 4

Une variable aléatoire X suit un **schéma de Bernoulli** lorsqu'elle prend uniquement deux valeurs 0 ou 1. Si nous notons $p = P(X = 1)$, nous avons :

- $P(X = 0) = 1 - p$
- $E(X) = p$
- $V(X) = p(1 - p)$

X est alors une variable aléatoire de Bernoulli de paramètre p .

Une variable aléatoire X suit une **loi Binomiale** de paramètres $n \in \mathbb{N}$ et $p \in [0, 1]$ lorsqu'elle est la somme de n variables de Bernoulli indépendantes de paramètre p . On peut alors montrer les résultats suivants :

- $\forall k \leq n, P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$
- $E(X) = np$
- $V(X) = np(1 - p)$

Encadré 5

Une **probabilité conditionnelle**, est une probabilité qu'un évènement se réalise à condition qu'un autre soit déjà réalisé. Soit deux évènements A et B , la probabilité de A sachant B , notée ici $P(A|B)$ est la probabilité que A se réalise lorsque B est réalisé.

Un **système complet d'évènements** est un ensemble d'évènements exclusifs dont la somme des probabilités est 1. Par exemple, l'évènement B et son complémentaire \bar{B} forme un tel système (voir figure 5.7).

La **formule des probabilités totales** nous donne la probabilité de réalisation d'un évènement à partir de la connaissance de probabilités conditionnelles sous réserve que l'ensemble des évènements du conditionnement forment un système complet d'évènements. Ainsi pour connaître la probabilité de A , nous pouvons écrire :

$$\begin{aligned} P(A) &= P(A \cap B) + P(A \cap \bar{B}) \\ &= P(A|B)P(B) + P(A|\bar{B})P(\bar{B}). \end{aligned}$$

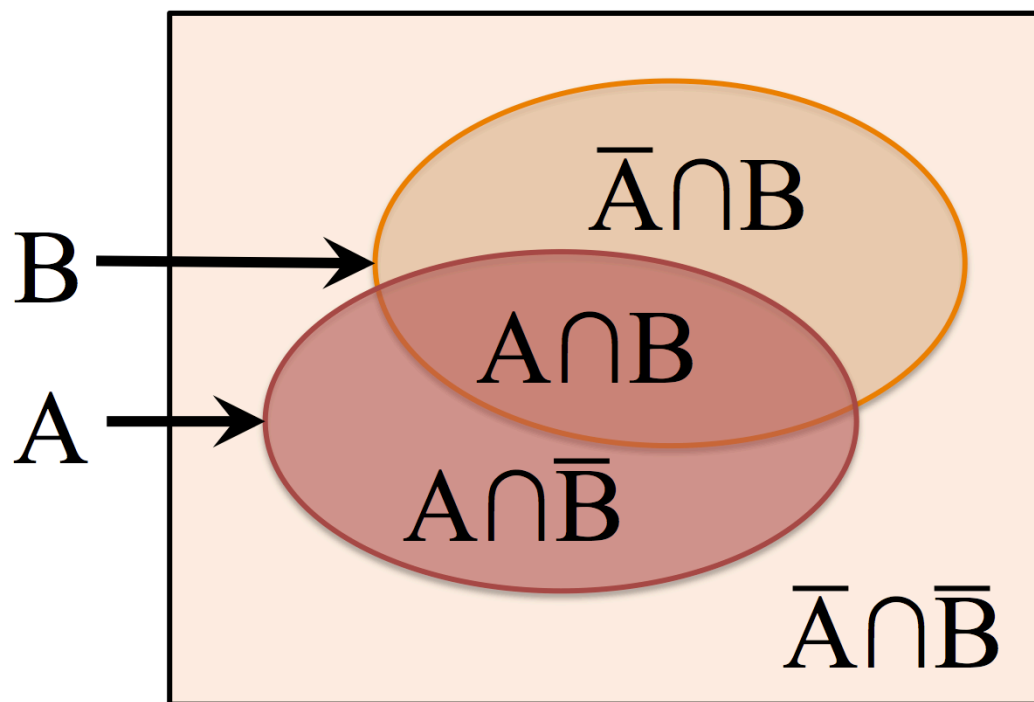


Figure 5.7: **Représentation de la formule des probabilités totales.** A et B sont deux événements. La probabilité de A peut être décrite comme la somme de l'intersection avec B et \bar{B} qui constituent un système complet d'événements.

ANNEXE B

UN MODÈLE STOCHASTIQUE DE LA BIOGÉOGRAPHIE INSULAIRE POUR LES RÉSEAUX ÉCOLOGIQUES DANS UN ENVIRONNEMENT ABIOTIQUE VARIABLE

Dans cette annexe, je montre comment l'intégration des interactions écologiques et des contraintes environnementales abiotiques est possible mathématiquement. Je démontre que le problème fait finalement appel à des objets mathématiques connus et qu'il existe un principe simple pour obtenir la solution analytique générale au modèle présenté au chapitre 1. La notation $\mathbb{P}(X)$ représente la probabilité de l'événement X .

6.12 Stochastic rules in MacArthur & Wilson's model

Following the TIB ?, we based our work on stochastic models. Let X_i be the random variable of presence on islands of a species i ; $X_i = 1$ means "species i is present on the island" while $X_i = 0$ means "species i is not found on the island"; X_i is a Bernoulli random variable. We define this probability at any time $t > 0$ and $X_{i,t}$ is the associated stochastic process. Moreover, let c_i (e_i) be the probability of colonization (extinction) of species i per time unit. To compute $X_t + dt$ based on X_t , we have to

derive $\mathbb{P}(X_{i,t+dt}|X_{i,t})$. As $X_{i,t}$ has two possible values leading to four possibilities:

$$\forall(t, c_i, e_i, dt) \in (\mathbb{R}^+)^4 :$$

$$\mathbb{P}(X_{i,t+dt} = 1|X_{i,t} = 0) = c_i dt + o(dt) \quad (6.1)$$

$$\mathbb{P}(X_{i,t+dt} = 0|X_{i,t} = 1) = e_i dt + o(dt) \quad (6.2)$$

$$\mathbb{P}(X_{i,t+dt} = 0|X_{i,t} = 0) = (1 - c_i dt) + o(dt) \quad (6.3)$$

$$\mathbb{P}(X_{i,t+dt} = 1|X_{i,t} = 1) = (1 - e_i dt) + o(dt) \quad (6.4)$$

Where dt is defined such as $e_i dt < 1$ and $c_i dt < 1$. For the remaining analyses, we use the symbol $o(dt)$ defines as follows:

$$\lim_{\substack{dt \rightarrow 0 \\ dt > 0}} \frac{o(dt)}{dt} = 0$$

According to equation (6.1), during dt , species i has a probability of $c_i dt$ of colonizing the island by a single colonization event and $o(dt)$ of colonizing it by multiple colonization/extinction events (*e.g* colonization-extinction-colonization). These multiple events are less likely and neglected when dt tends towards 0. Similarly, (6.2) explicits the probability of species i becoming extinct during dt , (6.3) gives us the probability of species i maintaining it-self on island and (6.4) provides probability of species i staying out of the island. The distribution $\mathcal{L}(X_{i,t+dt}|X_{i,t})$ solely depends on the duration dt not on t , $X_{i,t}$ is a no memory process, also called a first order discrete Markov chain. As $\{X_{i,t} = 0, X_{i,t} = 1\}$ is a partition:

$$\begin{aligned} \mathbb{P}(X_{i,t+dt} = 1) &= \mathbb{P}(X_{i,t+dt} = 1|X_{i,t} = 0)\mathbb{P}(X_{i,t} = 0) + \\ &\quad \mathbb{P}(X_{i,t+dt} = 1|X_{i,t} = 1)\mathbb{P}(X_{i,t} = 1) \end{aligned} \quad (6.5)$$

At time $t + dt$, species i will be on the island either because species i has colonized

during dt or because it has not died out from there. By using $\mathbb{P}(X_{i,t} = 0) = 1 - \mathbb{P}(X_{i,t} = 1)$:

$$\mathbb{P}(X_{i,t+dt} = 1) = c_i dt(1 - \mathbb{P}(X_{i,t} = 1)) + (1 - e_i dt)\mathbb{P}(X_{i,t} = 1) + o(dt) \quad (6.6)$$

Let $p_{i,t}$ stand for $\mathbb{P}(X_{i,t} = 1)$:

$$p_{i,t+dt} = c_i dt(1 - p_{i,t}) + (1 - e_i dt)p_{i,t} + o(dt) \quad (6.7)$$

$dt > 0$, then:

$$\frac{p_{i,t+dt} - p_{i,t}}{dt} = c_i(1 - p_{i,t}) - e_i p_{i,t} + \frac{o(dt)}{dt} \quad (6.8)$$

By passing to the limit, we finally find MacArthur and Wilson's model for one species:

$$\frac{dp_i}{dt} = c_i(1 - p_i) - e_i p_i \quad (6.9)$$

Similarly:

$$\frac{d(1 - p_i)}{dt} = e_i(1 - p_i) - c_i p_i \quad (6.10)$$

Equation (6.9) describes distribution of $X_{i,t>0}$: for any t , X_t follows a Bernoulli distribution with parameter $p_i(t)$. Equations (6.9) and (6.10) jointly describe a continuous time Markov Chain. We now consider the vector $\mathbf{C}(t)$ defined for any positive real number t as:

$$\mathbf{C}(t) = \begin{pmatrix} p_{i,t} & 1 - p_{i,t} \end{pmatrix} \quad (6.11)$$

the derivative is them:

$$\mathbf{C}'(t) = \mathbf{C}(t) \begin{pmatrix} -e_i & e_i \\ c_i & -c_i \end{pmatrix} = \mathbf{C}(t)\mathbf{G} \quad (6.12)$$

\mathbf{G} is the generator matrix of a continuous-time Markov chain associated to the classical model of MacArthur and Wilson. This provides the system of differential equations depicting the dynamics of the two possible states (with or without species i) the island can be found.

6.13 A Markov chains model to describe island communities

6.13.1 Model for P non-interacting species

We now consider a pool of P species (P is a natural number). When species are independent, the species richness on the island can be described as a sum of the random processes associated to the P species: $\mathbf{S}_{t>0} = \mathbf{X}_{1,t>0} + \mathbf{X}_{2,t>0} + \dots + \mathbf{X}_{P,t>0}$. As species are supposed to be independent, at any time t :

$$\mathbb{E}(S_t) = \sum_{i=0}^P p_i(t) \quad (6.13)$$

For homogenous colonization and extinction rates among species, we directly obtain a solution: for any time t , $S_t \sim \mathcal{B}(P, Pp_i(t))$. $\mathbb{E}(S_t)$ stands then for the solution of the classical differential equation with P species.

6.13.2 P Interacting species

When species are assumed to interact, the composition of insular communities must be integrated and must influence the colonization/extinction dynamics. Consequently, we gather species processes within $\mathbf{Y}_{t>0} = (\mathbf{X}_{1,t>0}, \mathbf{X}_{2,t>0}, \dots, \mathbf{X}_{P,t>0})$. For any t value, the line vector $\mathbf{Y}_t = (X_{1,t}, X_{2,t}, \dots, X_{P,t})$ contains presence and absence on the island for all the species of the network. Each of \mathbf{Y}_t elements takes a values of 0 or 1, then $\mathbf{Y}_t \in \{0, 1\}^P$. Elements of matrix \mathbf{A} , $\alpha_{i,j}$, describe the demographical impact of species j on species i . At time t , the total influence of insular species on a given species i , v_i is:

$$v_{i,t} = (\mathbf{A}\mathbf{Y}_t^T)_i = \sum_{j=1}^P \alpha_{ij} * x_{j,t} \quad (6.14)$$

Where T denotes the transposition operator, $(\)_i$ denotes the i^{th} column and $x_{j,t}$ the values of $X_{j,t}$ (0 or 1). We then use a function to change extinction and colonization rates according to v_i . Extinction rate of species i is therefore denoted f_i highlighting that it is a function of v_i . Similarly, g_i stands for the colonization rate, this is a function of v_i .

The conditional probability $\mathbb{P}(\mathbf{Y}_{t+\text{dt}}|\mathbf{Y}_t)$ is now examined. For P species, there is 2^P possible values for \mathbf{Y}_t . Let T_k ($k \in \{1, 2, \dots, 2^P\}$) represent one of these values (a given species assemblage). We have to split species into four different categories: I_1 , I_2 , I_3 et I_4 relatively to their presence on the island. This refers to the four potential situations

we have noticed earlier (see (6.1) to (6.4)).

$\forall t > 0, \forall (k, j) \in \{1, 2, \dots, 2^P\}^2 :$

$$\begin{aligned} \mathbb{P}(\mathbf{Y}_{t+dt} = \mathbf{T}_1 | \mathbf{Y}_t = \mathbf{T}_k) = & \mathbb{P} \left(\bigcap_{i_1 \in I_1} (X_{i_1, t+dt} = 1 | X_{i_1, t} = 0) , \right. \\ & \bigcap_{i_2 \in I_2} (X_{i_2, t+dt} = 0 | X_{i_2, t} = 1), \\ & \bigcap_{i_3 \in I_3} (X_{i_3, t+dt} = 1 | X_{i_3, t} = 1), \\ & \left. \bigcap_{i_4 \in I_4} (X_{i_4, t+dt} = 0 | X_{i_4, t} = 0) \right) \quad (6.15) \end{aligned}$$

Species are interdependent which apparently prevents from getting simple results. Nevertheless, with dt enough small, the island composition could be regarded as constant during dt . Extinction probability is thus calculated at time t and fixed until $t + dt$:

$$\begin{aligned} \mathbb{P}(\mathbf{Y}_{t+dt} = \mathbf{T}_1 | \mathbf{Y}_t = \mathbf{T}_k) = & \prod_{i_1 \in I_1} \mathbb{P}(X_{i_1, t+dt} = 1 | X_{i_1, t} = 0) \\ & \prod_{i_2 \in I_2} \mathbb{P}(X_{i_2, t+dt} = 0 | X_{i_2, t} = 1) \\ & \prod_{i_3 \in I_3} \mathbb{P}(X_{i_3, t+dt} = 1 | X_{i_3, t} = 1) \\ & \prod_{i_4 \in I_4} \mathbb{P}(X_{i_4, t+dt} = 0 | X_{i_4, t} = 0) \quad (6.16) \end{aligned}$$

The previous assumption leads us to consider multiple events as null-probability events, so we assume dt enough small to get $o(dt) = 0$.

$$\begin{aligned} \mathbb{P}(\mathbf{Y}_{t+dt} = \mathbf{T}_1 | \mathbf{Y}_t = \mathbf{T}_k) = & \prod_{i_1 \in I_1} g_{i_1}(v_{i_1, t}) dt \prod_{i_2 \in I_2} f_{i_2}(v_{i_2, t}) dt \\ & \prod_{i_3 \in I_3} (1 - f_{i_3}(v_{i_3, t}) dt) \prod_{i_4 \in I_4} (1 - g_{i_4}(v_{i_4, t}) dt) \quad (6.17) \end{aligned}$$

6.14 Environmental gradient and island biogeography

Let $\mathbf{W} = (W_1, W_2, \dots, W_n)$ be the vector gathering the n components of the environmental gradient considered; \mathbf{w} will be a vector giving a specific set of values for the environmental gradient. Colonization and extinction rates are influenced by environmental gradients. Consequently, for species i , functions g_i and f_i become multiple inputs functions. Equation (6.17) then becomes:

$$\begin{aligned} \mathbb{P}(\mathbf{Y}_{t+dt} = \mathbf{T}_k | \mathbf{Y}_t = \mathbf{T}_1, \mathbf{W} = \mathbf{w}) = & \prod_{i_1 \in I_1} g_{i_1}(v_{i_1,t}, \mathbf{w}) dt \prod_{i_2 \in I_2} f_{i_2}(v_{i_2,t}, \mathbf{w}) dt \\ & \prod_{i_3 \in I_3} (1 - f_{i_3}(v_{i_3,t}, \mathbf{w}) dt) \\ & \prod_{i_4 \in I_4} (1 - g_{i_4}(\mathbf{w}, v_{i_4,t}) dt) \end{aligned} \quad (6.18)$$

6.15 Using Markov chains

Based on equation (6.18), we spawn the transition matrix of a discrete Markov chain $\mathbf{M}_{\mathbf{w}}^{dt}$. For a given environment \mathbf{w} , this matrix describes the probabilities to switch from one state to another during dt . Its coefficients are obtained as follows:

$$\forall (k, l) \in \{1, 2, \dots, 2^P\}^2, \mu_{k,l} = \mathbb{P}(\mathbf{Y}_{t+dt} = \mathbf{T}_k | \mathbf{Y}_t = \mathbf{T}_l, \mathbf{W} = \mathbf{w}) \quad (6.19)$$

Now, let $\mathbf{C}_{\mathbf{w}}(t)$ be the line vector defines at each time t by:

$$\begin{aligned} \mathbf{C}_{\mathbf{w}}(t) = & (\mathbb{P}(\mathbf{Y}_t = \mathbf{T}_1 | \mathbf{W} = \mathbf{w}), \mathbb{P}(\mathbf{Y}_t = \mathbf{T}_2 | \mathbf{W} = \mathbf{w}), \dots, \\ & \mathbb{P}(\mathbf{Y}_t = \mathbf{T}_{2^n} | \mathbf{W} = \mathbf{w})) \end{aligned} \quad (6.20)$$

This vector describes probabilities of each possible island composition at any time t , we then derive $C_w(t + dt)$ from $C_w(t)$ as follows:

$$C_w(t + dt) = C_w(t)M_w^{dt} \quad (6.21)$$

We assume that none of the M_w^{dt} is null, which yields a regular Markov chain. In such case, $C(t)$ converges to an equilibrium value C_{eq} when t increases:

$$\lim_{l \rightarrow +\infty} C_0(M_w^{dt})^l = C_{eq} \quad (6.22)$$

This C_{eq} satisfies:

$$C_{eq}(M_w^{dt}) = C_{eq} \quad (6.23)$$

$$||C_{eq}|| = 1 \quad (6.24)$$

Therefore, C_{eq} is given by the normalized left Eigen vector associated to left Eigen value 1.

6.16 Time continuous Markov chain

We show here how we can get the generator matrix of the time-continuous Markov chain associated to the transition matrix M_w^{dt} . We then provide an explicit solution of the system of differential equations we derived.

6.16.1 Solution for two species

We start with $P = 2$, we denote: $\mathbf{T}_1 = (1, 1)$, $\mathbf{T}_2 = (1, 0)$, $\mathbf{T}_3 = (0, 1)$ and $\mathbf{T}_4 = (0, 0)$. We consider here that \mathbf{W} is set to \mathbf{w} and so, for instance, $\mathbb{P}(\mathbf{Y}_t = \mathbf{T}_1)$ means $\mathbb{P}(\mathbf{Y}_t = \mathbf{T}_1 | \mathbf{W} = \mathbf{w})$.

$$\begin{aligned}
 \mathbb{P}(\mathbf{Y}_{t+dt} = \mathbf{T}_1) &= \mathbb{P}(\mathbf{Y}_{t+dt} = \mathbf{T}_1 | \mathbf{Y}_t = \mathbf{T}_1) \mathbb{P}(\mathbf{Y}_t = \mathbf{T}_1) \\
 &\quad + \mathbb{P}(\mathbf{Y}_{t+dt} = \mathbf{T}_1 | \mathbf{Y}_t = \mathbf{T}_2) \mathbb{P}(\mathbf{Y}_t = \mathbf{T}_2) \\
 &\quad + \mathbb{P}(\mathbf{Y}_{t+dt} = \mathbf{T}_1 | \mathbf{Y}_t = \mathbf{T}_3) \mathbb{P}(\mathbf{Y}_t = \mathbf{T}_3) \\
 &\quad + \mathbb{P}(\mathbf{Y}_{t+dt} = \mathbf{T}_1 | \mathbf{Y}_t = \mathbf{T}_4) \mathbb{P}(\mathbf{Y}_t = \mathbf{T}_4) \quad (6.25)
 \end{aligned}$$

As in this stage, as \mathbf{Y}_t do not refer to the same values for the whole equation, we slightly change our denotation: v_{i,\mathbf{T}_j} represents $v_{i,t}$ when $\mathbf{Y}_t = \mathbf{T}_j$. According to (6.18), we get:

$$\begin{aligned}
 \mathbb{P}(\mathbf{Y}_{t+dt} = \mathbf{T}_1) &= (1 - f_1(v_{1,\mathbf{T}_1}, \mathbf{w})dt)(1 - f_2(v_{2,\mathbf{T}_1}, \mathbf{w})dt) \mathbb{P}(\mathbf{Y}_t = \mathbf{T}_1) \\
 &\quad + (1 - f_1(v_{1,\mathbf{T}_2}, \mathbf{w})dt)g_2(v_{2,\mathbf{T}_2}, \mathbf{w})dt \mathbb{P}(\mathbf{Y}_t = \mathbf{T}_2) \\
 &\quad + g_1(v_{1,\mathbf{T}_3}, \mathbf{w})dt(1 - f_2(v_{2,\mathbf{T}_3}, \mathbf{w})dt) \mathbb{P}(\mathbf{Y}_t = \mathbf{T}_3) \\
 &\quad + g_1(v_{1,\mathbf{T}_4}, \mathbf{w})g_2(v_{2,\mathbf{T}_4}, \mathbf{w})dt^2 \mathbb{P}(\mathbf{Y}_t = \mathbf{T}_4) + o(dt) \quad (6.26)
 \end{aligned}$$

This yields:

$$\begin{aligned}
\mathbb{P}(\mathbf{Y}_{t+dt} = \mathbf{T}_1) &= (1 - f_1(v_{1,\mathbf{T}_1}, \mathbf{w})dt - f_2(v_{2,\mathbf{T}_1}, \mathbf{w})dt \\
&\quad + f_1(v_{1,\mathbf{T}_1}, \mathbf{w})f_2(v_{2,\mathbf{T}_1}, \mathbf{w})dt^2)\mathbb{P}(\mathbf{Y}_t = \mathbf{T}_1) \\
&\quad + ((g_2(v_{2,\mathbf{T}_2}, \mathbf{w})dt - g_2(v_{2,\mathbf{T}_2}, \mathbf{w})f_1(v_{1,\mathbf{T}_2}, \mathbf{w})dt^2))\mathbb{P}(\mathbf{Y}_t = \mathbf{T}_2) \\
&\quad + (g_1(v_{1,\mathbf{T}_3}, \mathbf{w})dt - g_1(v_{1,\mathbf{T}_3}, \mathbf{w})f_2(v_{2,\mathbf{T}_3}, \mathbf{w})dt^2)\mathbb{P}(\mathbf{Y}_t = \mathbf{T}_3) \\
&\quad + g_1(v_{1,\mathbf{T}_4}, \mathbf{w})g_2(v_{2,\mathbf{T}_4}, \mathbf{w})dt^2\mathbb{P}(\mathbf{Y}_t = \mathbf{T}_4) + o(dt) \quad (6.27)
\end{aligned}$$

Assuming $dt > 0$:

$$\begin{aligned}
\frac{d\mathbb{P}(\mathbf{Y}_t = \mathbf{T}_1)}{dt} &= \frac{\mathbb{P}(\mathbf{Y}_{t+dt} = \mathbf{T}_1) - \mathbb{P}(\mathbf{Y}_t = \mathbf{T}_1)}{dt} \\
&= -(f_1(v_{1,\mathbf{T}_1}, \mathbf{w}) + f_2(v_{2,\mathbf{T}_1}, \mathbf{w})) \\
&\quad f_1(v_{1,\mathbf{T}_1}, \mathbf{w})f_2(v_{2,\mathbf{T}_1}, \mathbf{w})dt)\mathbb{P}(\mathbf{Y}_t = \mathbf{T}_1) \\
&\quad + ((g_2(v_{2,\mathbf{T}_2}, \mathbf{w}) - g_2(v_{2,\mathbf{T}_2}, \mathbf{w})f_1(v_{1,\mathbf{T}_2}, \mathbf{w})dt))\mathbb{P}(\mathbf{Y}_t = \mathbf{T}_2) \\
&\quad + (g_1(v_{1,\mathbf{T}_3}, \mathbf{w}) - g_1(v_{1,\mathbf{T}_3}, \mathbf{w})f_2(v_{2,\mathbf{T}_3}, \mathbf{w})dt)\mathbb{P}(\mathbf{Y}_t = \mathbf{T}_3) \\
&\quad + g_1(v_{1,\mathbf{T}_4}, \mathbf{w})g_2(v_{2,\mathbf{T}_4}, \mathbf{w})dt\mathbb{P}(\mathbf{Y}_t = \mathbf{T}_4) + \frac{o(dt)}{dt} \quad (6.28)
\end{aligned}$$

When passing to the limit, we derive the following master equation:

$$\begin{aligned}
\frac{d\mathbb{P}(\mathbf{Y}_t = \mathbf{T}_1)}{dt} &= -(f_1(v_{1,\mathbf{T}_1}, \mathbf{w}) + f_2(v_{2,\mathbf{T}_1}, \mathbf{w}))\mathbb{P}(\mathbf{Y}_t = \mathbf{T}_1) + g_2(v_{2,\mathbf{T}_2}, \mathbf{w})\mathbb{P}(\mathbf{Y}_t = \mathbf{T}_2) \\
&\quad + g_1(v_{1,\mathbf{T}_3}, \mathbf{w})\mathbb{P}(\mathbf{Y}_t = \mathbf{T}_3) \quad (6.29)
\end{aligned}$$

We do so for the \mathbf{T}_2 , \mathbf{T}_3 and \mathbf{T}_4 . Let $\mathbf{C}(t)$ be the column vector define for all real $t > 0$ such as $\mathbf{C}(t) = (\mathbb{P}(\mathbf{Y}_t = \mathbf{T}_1), \mathbb{P}(\mathbf{Y}_t = \mathbf{T}_2), \mathbb{P}(\mathbf{Y}_t = \mathbf{T}_3), \mathbb{P}(\mathbf{Y}_t = \mathbf{T}_4))$. We thus have the following relationship:

$$\mathbf{C}'(t) = \mathbf{C}(t)\mathbf{G}_{\mathbf{w}} \quad (6.30)$$

where $\mathbf{G}_{\mathbf{w}}$ is:

$$\begin{pmatrix} -(f_1(v_{1,\mathbf{T}_1}, \mathbf{w}) + f_2(v_{2,\mathbf{T}_1}, \mathbf{w})) & f_1(v_{1,\mathbf{T}_1}, \mathbf{w}) & f_2(v_{2,\mathbf{T}_1}, \mathbf{w}) & 0 \\ g_2(v_{2,\mathbf{T}_2}, \mathbf{w}) & -(f_1(v_{1,\mathbf{T}_2}, \mathbf{w}) + g_2(v_{2,\mathbf{T}_2}, \mathbf{w})) & 0 & f_1(v_{1,\mathbf{T}_2}, \mathbf{w}) \\ g_1(v_{1,\mathbf{T}_3}, \mathbf{w}) & 0 & -(g_1(v_{1,\mathbf{T}_3}, \mathbf{w}) + f_2(v_{2,\mathbf{T}_3}, \mathbf{w})) & f_2(v_{2,\mathbf{T}_3}, \mathbf{w}) \\ 0 & g_1(v_{1,\mathbf{T}_4}, \mathbf{w}) & g_2(v_{2,\mathbf{T}_4}, \mathbf{w}) & -(g_1(v_{1,\mathbf{T}_4}, \mathbf{w}) + g_2(v_{2,\mathbf{T}_4}, \mathbf{w})) \end{pmatrix}$$

At the equilibrium, the solution is given by \mathbf{C}_{eq} which verifies:

$$\mathbf{C}_{\text{eq}} \mathbf{G}_{\mathbf{w}} = 0 \quad (6.31)$$

$$\|\mathbf{C}_{\text{eq}}\| = 1 \quad (6.32)$$

This is the normalized left Eigen vector associated to the left Eigen values 0. We now solve the linear system of differential (6.30). First, as $\{\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3, \mathbf{T}_4\}$ is a partition (which also justifies that 0 is a left Eigen values):

$$\sum_{i=1}^4 \mathbb{P}(\mathbf{Y}_{\mathbf{t}} = \mathbf{T}_i) = 1 \quad (6.33)$$

so, we express $\mathbb{P}(\mathbf{Y}_{\mathbf{t}} = \mathbf{T}_3)$ using the probabilities:

$$\begin{aligned} \frac{d\mathbb{P}(\mathbf{Y}_{\mathbf{t}} = \mathbf{T}_1)}{dt} &= -(f_1(v_{1,\mathbf{T}_1}, \mathbf{w}) + f_2(v_{2,\mathbf{T}_1}, \mathbf{w}))\mathbb{P}(\mathbf{Y}_{\mathbf{t}} = \mathbf{T}_1) \\ &\quad + g_2(v_{2,\mathbf{T}_2}, \mathbf{w})\mathbb{P}(\mathbf{Y}_{\mathbf{t}} = \mathbf{T}_2) + g_1(v_{1,\mathbf{T}_3})\mathbb{P}(\mathbf{Y}_{\mathbf{t}} = \mathbf{T}_3) \end{aligned} \quad (6.34)$$

$$\begin{aligned} \frac{d\mathbb{P}(\mathbf{Y}_{\mathbf{t}} = \mathbf{T}_2)}{dt} &= g_1(v_{1,\mathbf{T}_4}, \mathbf{w}) + (f_2(v_{2,\mathbf{T}_1}, \mathbf{w}) - g_1(v_{1,\mathbf{T}_4}, \mathbf{w}))\mathbb{P}(\mathbf{Y}_{\mathbf{t}} = \mathbf{T}_1) \\ &\quad - (f_1(v_{1,\mathbf{T}_2}, \mathbf{w}) + g_2(v_{2,\mathbf{T}_2}, \mathbf{w})g_1(v_{1,\mathbf{T}_4}, \mathbf{w}))\mathbb{P}(\mathbf{Y}_{\mathbf{t}} = \mathbf{T}_2) \\ &\quad - g_1(v_{1,\mathbf{T}_4}, \mathbf{w})\mathbb{P}(\mathbf{Y}_{\mathbf{t}} = \mathbf{T}_3) \end{aligned} \quad (6.35)$$

$$\begin{aligned} \frac{d\mathbb{P}(\mathbf{Y}_{\mathbf{t}} = \mathbf{T}_3)}{dt} &= g_2(v_{2,\mathbf{T}_4}, \mathbf{w}) + (f_1(v_{1,\mathbf{T}_1}, \mathbf{w}) - g_2(v_{2,\mathbf{T}_4}, \mathbf{w}))\mathbb{P}(\mathbf{Y}_{\mathbf{t}} = \mathbf{T}_1) \\ &\quad - g_2(v_{2,\mathbf{T}_4}, \mathbf{w})\mathbb{P}(\mathbf{Y}_{\mathbf{t}} = \mathbf{T}_2) - (g_1(v_{1,\mathbf{T}_3}, \mathbf{w}) + f_2(v_{2,\mathbf{T}_3}, \mathbf{w}) \\ &\quad + g_2(v_{2,\mathbf{T}_4}, \mathbf{w}))\mathbb{P}(\mathbf{Y}_{\mathbf{t}} = \mathbf{T}_3) \end{aligned} \quad (6.36)$$

we denote:

$$\mathbb{P}^*(\mathbf{Y}_{\mathbf{t}} = \mathbf{T}_2) = \mathbb{P}(\mathbf{Y}_{\mathbf{t}} = \mathbf{T}_2) - \frac{g_1(v_{1,\mathbf{T}_4}, \mathbf{w})}{f_1(v_{1,\mathbf{T}_2}, \mathbf{w}) + g_2(v_{2,\mathbf{T}_2}, \mathbf{w}) + g_1(v_{1,\mathbf{T}_4}, \mathbf{w})} \quad (6.37)$$

$$\mathbb{P}^*(\mathbf{Y}_{\mathbf{t}} = \mathbf{T}_3) = \mathbb{P}(\mathbf{Y}_{\mathbf{t}} = \mathbf{T}_3) - \frac{g_2(v_{2,\mathbf{T}_4}, \mathbf{w})}{g_1(v_{1,\mathbf{T}_3}, \mathbf{w}) + f_2(v_{2,\mathbf{T}_3}, \mathbf{w}) + g_2(v_{2,\mathbf{T}_4}, \mathbf{w})} \quad (6.38)$$

as:

$$\frac{dP^*(\mathbf{Y}_t = \mathbf{T}_2)}{dt} = \frac{dP(\mathbf{Y}_t = \mathbf{T}_2)}{dt} \quad (6.39)$$

$$\frac{dP^*(\mathbf{Y}_t = \mathbf{T}_3)}{dt} = \frac{dP(\mathbf{Y}_t = \mathbf{T}_3)}{dt} \quad (6.40)$$

yielding:

$$\mathbf{C}^{*'}(t) = \mathbf{C}^*(t) \mathbf{G}_w^* \quad (6.41)$$

where:

$$\begin{aligned} \mathbf{C}^{*'} &= \begin{pmatrix} \frac{dP(\mathbf{Y}_t = \mathbf{T}_1)}{dt} & \frac{dP^*(\mathbf{Y}_t = \mathbf{T}_2)}{dt} & \frac{dP^*(\mathbf{Y}_t = \mathbf{T}_3)}{dt} \end{pmatrix} \\ \mathbf{C}^* &= \begin{pmatrix} P(\mathbf{Y}_t = \mathbf{T}_1) & P^*(\mathbf{Y}_t = \mathbf{T}_2) & P^*(\mathbf{Y}_t = \mathbf{T}_3) \end{pmatrix} \end{aligned}$$

$$\mathbf{G}_{\mathbf{w}}^* = \begin{pmatrix} -(f_1(v_{1,\mathbf{T}_1}, \mathbf{w}) + f_2(v_{2,\mathbf{T}_1}, \mathbf{w})) & f_2(v_{2,\mathbf{T}_1}, \mathbf{w}) - g_1(v_{1,\mathbf{T}_4}, \mathbf{w}) & (f_1(v_{1,\mathbf{T}_1}, \mathbf{w}) - g_2(v_{2,\mathbf{T}_4}, \mathbf{w})) \\ (g_2(v_{2,\mathbf{T}_2}, \mathbf{w}) & -(f_1(v_{1,\mathbf{T}_2}, \mathbf{w}) + g_2(v_{2,\mathbf{T}_2}, \mathbf{w}) + g_1(v_{1,\mathbf{T}_4}, \mathbf{w})) & -g_2(v_{2,\mathbf{T}_4}, \mathbf{w}) \\ g_1(v_{1,\mathbf{T}_3}, \mathbf{w}) & -g_1(v_{1,\mathbf{T}_4}, \mathbf{w}) & -(g_1(v_{1,\mathbf{T}_3}, \mathbf{w}) + f_2(v_{2,\mathbf{T}_3}, \mathbf{w}) + g_2(v_{2,\mathbf{T}_4}, \mathbf{w})) \end{pmatrix}$$

then, if \mathbf{G}_w^* is diagonalizable, we readily solve (6.41):

$$\mathbf{C}^{*'} = \mathbf{CZDZ}^{-1} \quad (6.42)$$

where \mathbf{D} is the diagonal matrix containing the Eigen values of \mathbf{G}_w^* and \mathbf{Z} , the matrix permitting the change of basis:

$$\mathbf{C}^{*'}\mathbf{Z} = \mathbf{CZD} \quad (6.43)$$

Thus we have to solve a homogenous system of differential equations:

$$\mathbf{C}^*(t)\mathbf{Z} = \mathbf{K} \exp(\Lambda t) \quad (6.44)$$

Where:

$$\exp(\Lambda t) = \begin{pmatrix} \exp(\lambda_1 t) & 0 & 0 \\ 0 & \exp(\lambda_2 t) & 0 \\ 0 & 0 & \exp(\lambda_3 t) \end{pmatrix}$$

where λ_i stands for the i^{th} Eigen values of \mathbf{G}_w^* , therefore:

$$\mathbf{C}^*(t) = \mathbf{K} \exp(\Lambda t) \mathbf{Z}^{-1} \quad (6.45)$$

Furthermore, given:

$$\mathbf{C}^*(0) = \mathbf{KZ}^{-1} \quad (6.46)$$

we obtain:

$$\mathbf{C}^*(t) = \mathbf{C}^*(0) \mathbf{Z} \exp(\Lambda t) \mathbf{Z}^{-1} \quad (6.47)$$

$\mathbf{C}(t)$ is finally obtained by adding the two constants we have subtracted in (6.37) and (6.38). This allows us to express the expected values of local richness $S(t)$ as the following matrix product:

$$\mathbb{E}(S(t)) = \mathbf{C}(t)\mathbf{N}^T \quad (6.48)$$

where \mathbf{N} is the vector defined as $\mathbf{N} = (||\mathbf{T}_1||^2, ||\mathbf{T}_2||^2, ||\mathbf{T}_3||^2, ||\mathbf{T}_4||^2)$, where $|| \cdot ||$ denotes the euclidian norm. In our example, we have $\mathbf{N} = (2, 1, 1, 0)$.

6.16.2 Solution for P species

We first build \mathbf{M}_w^{dt} for P species, we then recall equation (6.18) and consider the expression of the conditional probabilities between any pair of states ($\mathbf{Y}_{t+\text{dt}} = \mathbf{T}_i, \mathbf{Y}_t = \mathbf{T}_j$). We assume simultaneous colonization and/or extinction of different species to be neglectable when dt tends towards 0 (this is shown using (6.18), this kind of events are indeed multiplied by dt^m with $m \geq 2$). Consequently, we distinguish the different cases associated to the number of changes in presence status of species. To do so, we consider all values $||\mathbf{T}_i - \mathbf{T}_j||$ takes. Note that when $i = j$ (diagonal terms), the island composition is unchanged during dt .

$$\mathbb{P}(\mathbf{Y}_{t+\text{dt}} = \mathbf{T}_j | \mathbf{Y}_t = \mathbf{T}_j) = \prod_{i_3 \in I_3} (1 - f_{i_3}(v_{i_3, \mathbf{T}_j}, \mathbf{w})dt) \prod_{i_4 \in I_4} (1 - g_{i_4}(v_{i_4, \mathbf{T}_j}, \mathbf{w})dt) \quad (6.49)$$

yielding:

$$\mu_{j,j} = 1 - \sum_{i_3 \in I_3} f_{i_3}(v_{i_3, \mathbf{T}_j}, \mathbf{w})dt - \sum_{i_4 \in I_4} g_{i_4}(v_{i_4, \mathbf{T}_j}, \mathbf{w})dt + o(dt) \quad (6.50)$$

when $i \neq j$, we consider three cases:

$$\|\mathbf{T}_i - \mathbf{T}_j\| = 1, (\mathbf{T}_i - \mathbf{T}_j)_l = 1 \Rightarrow \mu_{j,i} = g_l(v_{l,\mathbf{T}_j}, \mathbf{w})dt + o(dt) \quad (6.51)$$

$$\|\mathbf{T}_i - \mathbf{T}_j\| = 1, (\mathbf{T}_i - \mathbf{T}_j)_l = -1 \Rightarrow \mu_{j,i} = f_l(v_{l,\mathbf{T}_j}, \mathbf{w})dt + o(dt) \quad (6.52)$$

$$\|\mathbf{T}_i - \mathbf{T}_j\| > 1 \Rightarrow \mu_{j,i} = o(dt) \quad (6.53)$$

Here l denotes the identity of the species that go extinct or colonize the island during dt . Then, we derive \mathbf{G}_w from the previous results. Let K_1 and K_2 be the groups of states which respectively correspond to (6.51) and (6.52). For any state \mathbf{T}_i :

$$\begin{aligned} \mathbb{P}(\mathbf{Y}_{t+dt} = \mathbf{T}_i) &= \sum_{k_1 \in K_1} g_{l(k_1)}(v_{l(k_1)}, \mathbf{T}_{k_1}, \mathbf{w})dt \mathbb{P}(\mathbf{Y}_t = \mathbf{T}_{k_1}) \\ &+ \sum_{k_2 \in K_2} f_{l(k_2)}(v_{l(k_2)}, \mathbf{T}_{k_2}, \mathbf{w})dt \mathbb{P}(\mathbf{Y}_t = \mathbf{T}_{k_2}) \\ &+ \left(1 - \sum_{i_3 \in I_3} f_{i_3}(v_{i_3, \mathbf{T}_j}, \mathbf{w})dt - \sum_{i_4 \in I_4} g_{i_4}(v_{i_4, \mathbf{T}_j}, \mathbf{w})dt \right) \mathbb{P}(\mathbf{Y}_t = \mathbf{T}_i) \\ &+ o(dt) \end{aligned} \quad (6.54)$$

Therefore:

$$\begin{aligned} \frac{d\mathbb{P}(\mathbf{Y}_t = \mathbf{T}_i)}{dt} &= \sum_{k_1 \in K_1} g_{l(k_1)}(v_{l(k_1)}, \mathbf{T}_{k_1}, \mathbf{w})\mathbb{P}(\mathbf{Y}_t = \mathbf{T}_{k_1}) \\ &+ \sum_{k_2 \in K_2} f_{l(k_2)}(v_{l(k_2)}, \mathbf{T}_{k_2}, \mathbf{w})\mathbb{P}(\mathbf{Y}_t = \mathbf{T}_{k_2}) + \\ &\left(- \sum_{i_3 \in I_3} f_{i_3}(v_{i_3, \mathbf{T}_j}, \mathbf{w}) - \sum_{i_4 \in I_4} g_{i_4}(v_{i_4, \mathbf{T}_j}, \mathbf{w}) \right) \mathbb{P}(\mathbf{Y}_t = \mathbf{T}_i) \end{aligned} \quad (6.55)$$

Hence, $\mathbf{G}_{\mathbf{w}}$ is as follows:

$$\|\mathbf{T}_{\mathbf{j}} - \mathbf{T}_{\mathbf{i}}\| = 1, (\mathbf{T}_{\mathbf{j}} - \mathbf{T}_{\mathbf{i}})_l = 1 \Rightarrow \gamma_{i,j} = f(v_{l,\mathbf{T}_{\mathbf{i}}}, \mathbf{w}) \quad (6.56)$$

$$\|\mathbf{T}_{\mathbf{j}} - \mathbf{T}_{\mathbf{i}}\| = 1, (\mathbf{T}_{\mathbf{j}} - \mathbf{T}_{\mathbf{i}})_l = -1 \Rightarrow \gamma_{i,j} = g(v_{l,\mathbf{T}_{\mathbf{i}}}, \mathbf{w}) \quad (6.57)$$

$$\|\mathbf{T}_{\mathbf{j}} - \mathbf{T}_{\mathbf{i}}\| > 1 \Rightarrow \gamma_{i,j} = 0 \quad (6.58)$$

for the diagonal elements, we get:

$$\gamma_{j,j} = \left(- \sum_{i_3 \in I_3} f_{i_3}(v_{i_3,\mathbf{T}_{\mathbf{j}}}, \mathbf{w}) - \sum_{i_4 \in I_4} g_{i_4}(v_{i_4,\mathbf{T}_{\mathbf{j}}}, \mathbf{w}) \right) \quad (6.59)$$

Finally, the solution of the expected richness is given by following the different steps from (6.33) to (6.48).

ANNEXE C

PROBABILITÉS DE TRANSITION ENTRE RÉSEAUX TROPHIQUES

La TTIB proposée par ? est issue d'une collaboration entre cinq auteurs : Dominique Gravel, François Massol, Elsa Canard, David Mouillot et Nicolas Mouquet. Parmi ces auteurs, François Massol a récemment été invité à soumettre une contribution au volume 56 du journal scientifique *Advances in Ecological Research* traitant du lien entre les réseaux écologiques et l'invasion des espèces, volume dont il est coéditeur. Ce fut une opportunité pour montrer les possibilités offertes par la TTIB et aussi une occasion de pouvoir revenir en détail sur les mathématiques en arrière du modèle. L'article final, est une collaboration entre François Massol (premier auteur), Maxime Dubart, Vincent Calcagno, Claire Jacquet, Sonia Kéfi, Dominique Gravel (dernier auteur) et moi-même.

C'est pour mon travail sur l'extension de la TIB présenté au chapitre 1 que j'ai été invité à contribuer à cet article. Ma participation détaille quelques aspects techniques du modèle mathématique ainsi que les nouvelles possibilités qu'il ajoute à la TIB. Dans cette annexe, je présente mon travail tel qu'il était avant d'être intégré dans l'article qui est actuellement en cours de révision. La figure qui y est présentée est de François Massol qui m'a gentiment permis de la réutiliser. La notation P_X représente la probabilité de l'événement X , c'est donc l'équivalent de la notation $\mathbb{P}(X)$ à l'annexe 2.

7.17 The model

An alternative to study analytically the model of the TTIB proposed by ? is to derive master equations associated to community states (rather than for individual species). The mathematical object is then the random process $(C_{t>0} = \cup_{i=1}^T X_i, i.e. a vector of 0 and 1 describing the presence and absence of all species at any time t . For T species, there are 2^T community states, Table 7.2 provides an example for species B, C, D and E presented in Figure 7.8 .$

Deriving the master equation associated to a given community state S_k necessitates the consideration of transition probabilities among community states between t and $t + dt$ (where dt is assumed to be small enough to permit only one transition). As all community states are included in the set $S_k, k \in \{1, 2, \dots, 2^T\}$, following the law of total probability, for any community states:

$$P_{C_{t+dt}=S_k} = \sum_{l=0}^{2^T-1} P_{C_{t+dt}=S_k|C_t=S_l} P_{C_t=S_l} \quad (7.1)$$

$P_{C_{t+dt}=S_k|C_t=S_l}$ is now assumed to be a linear function of dt , this assumption is needed valid only for a very small time step. For $k \neq l$:

$$P_{C_{t+dt}=S_k|C_t=S_l} = f(S_k, S_l)dt \quad (7.2)$$

and, as $\sum_l P_{C_{t+dt}=S_l|C_t=S_k} = 1$:

Table 7.2: Community states associated to the B-C-D-E network of figure 7.8. They represent all combinations of presence and absence of species. At any time t , $C_{t>0}$ represents the species composition of the island and thus is in one of the sixteen possible S_k states listed. Here, all communities are considered regardless whether or not they are TTIB-compatible as highlighted in the right column.

Community states	B	C	D	E	TTIB-compatible
S_0	0	0	0	0	yes
S_1	0	0	0	1	no
S_2	0	0	1	0	yes
S_3	0	0	1	1	yes
S_4	0	1	0	0	yes
S_5	0	1	0	1	yes
S_6	0	1	1	0	yes
S_7	0	1	1	1	no
S_8	1	0	0	0	no
S_9	1	0	0	1	no
S_{10}	1	0	1	0	no
S_{11}	1	0	1	1	yes
S_{12}	1	1	0	0	yes
S_{13}	1	1	0	1	yes
S_{14}	1	1	1	0	yes
S_{15}	1	1	1	1	yes

$$P_{C_{t+dt}=S_k|C_t=S_k} = 1 - \sum_{l \neq k} f(S_l, S_k) dt \quad (7.3)$$

$f(S_k, S_l)$ reflects how easy the transition between S_k and S_l is, it may be greater than one as long as $f(S_k, S_l) dt \leq 1$. The TTIB actually acknowledges the existence of trophic interactions since it assumes:

1. $f(S_k, S_l) = 0$ when the transition from S_l to S_k involves the colonization of a predator without any prey,
2. $f(S_k, S_l) dt = 1$ when S_l is a community states where a predator is present without any prey and S_k the same community without the predator.

Plugging the two latter equations in equation (1):

$$\frac{P_{C_{t+dt}=S_k} - P_{C_t=S_k}}{dt} = - \left(\sum_{l \neq k} f(S_l, S_k) \right) P_{C_t=S_k} + \sum_{l \neq k} f(S_k, S_l) P_{C_t=S_l} \quad (7.4)$$

When $dt \rightarrow 0$, it provides the master equation than can be written in a vector format to integrate the dynamics of all community states $\mathbf{P} = (P_{S_1}, P_{S_2}, \dots, P_{S_{2T}})$:

$$\frac{d\mathbf{P}}{dt} = \mathbf{M}\mathbf{P} \quad (7.5)$$

Here \mathbf{P} includes all the community states and the coefficient of the matrix \mathbf{M} are a characteristic of the community. Nevertheless, the solution has a similar expression:

$$\mathbf{P}(t) = e^{t\mathbf{M}}\mathbf{P}_0 \quad (7.6)$$

This actually describes a continuous-time Markov chain. When all the community states are linked (*i.e.* the Markov chain is irreducible), an equilibrium \mathbf{P} is reached and given by the vector in the kernel of \mathbf{M} whose elements sum to one.

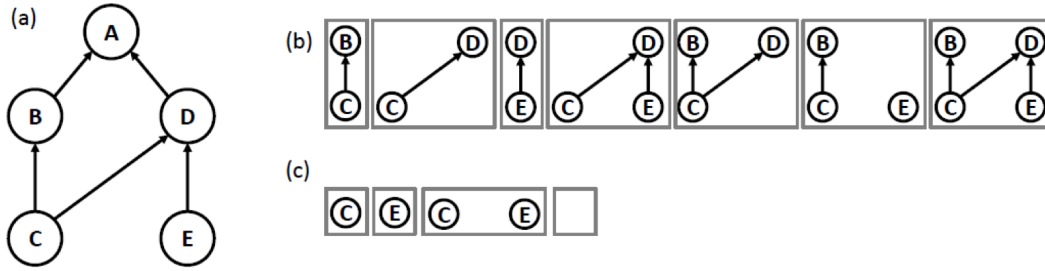


Figure 7.8: Simple food webs used to illustrate the TTTIB ?. (a) The complete food web (on the mainland) consists in five different species (A, B, C, D, E); (b) sub-food webs that species A can colonize; (c) sub-food webs that species A cannot colonize.

7.18 Reasonable approximations

The formulation above allows the study of assemblages of non-independent species but suffers from its generality: for T species, the matrix \mathbf{M} must be filled out with $(2^T - 1) \times (2^T - 1)$ coefficients (‘-1’ acknowledges that for any time t the elements of $\mathbf{P}(t)$ sum to one). Even if the knowledge of a particular network may help finding these coefficients, reasonable assumptions can be made to decrease the complexity of \mathbf{M} . First, community compositions between t and $t + dt$ cannot differ in more than one species, *i.e.* $f(S_k, S_l) = 0$ if $|S_k - S_l| > 1$. This turns \mathbf{M} into a sparse matrix: only $T \times (2^T - 1)$ are not zero. This assumption is not possible in the TTIB as the extinction

of a prey leads to more than one extinction. However, this issue is easy to circumvent by allowing the predator to survive a (very) short period alone on the island which would be assumed using large value for $f(S_k, S_l)$ that measures the transition of a community with a predator and the same community without it.

A second hypothesis is that colonization processes may be independent of interactions. That is, a predator may actually colonize an island without any prey. This is reasonable if the extinction probability of this predator on such island is high. Therefore, $f(S_k, S_l) = c$ when $\text{sum}\{S_k - S_l\} = -1$. The latter assumption is also useful to integrate variability among species' dispersal capacities (?). Tables 7.3 and 7.4 present the matrix \mathbf{M} for species C, D and E, *i.e.* community states from S_0 to S_8 . Once the colonization probability is determined, the remaining $T \times (2^T - 2)$ coefficients needed can be found based on the biological knowledge of species studies, *e.g.* the nature and the strength of interactions.

7.19 Deriving species richness

The solution \mathbf{P} includes the probabilities of all community states at equilibrium. This information is more than the knowledge of individual presences and species richness. In order to obtain the probability of species i being present on the island at equilibrium, the following sum of probabilities is computed:

$$P_{X_i=1} = \sum_{k|S_{k,i}=1} \mathbf{P}_k \quad (7.7)$$

where $S_{i,k}$ is the i^{th} component of S_i (the one pertaining to species i) and \mathbf{P}_i the k^{th} component of \mathbf{P}_i . Similarly, the species richness is given by the sum of \mathbf{P} weighted by the cardinal of the community states it refers to:

$$S = \sum_{k=0}^{2^T-1} |S_k| \mathbf{P}_k \quad (7.8)$$

In a similar fashion, many probabilities can be derived regarding either a particular set of species either a particular property of the community. For instance, this framework allows to derive the probability of finding a given set of predators but also the mean trophic level expected and even the probability of having a trophic chains of at least p levels. In all these situations, calculus necessitate the identification of the community states to be summed.

Table 7.3: The transition matrix of the continuous-time Markov chain associated to all combinations of C, D and E species (part 1/2). An empty cell means 0.

	S_0	S_1	S_2	S_3	S_4
S_0	$1 - c - c - c$	$f(S_1, S_0)$	$f(S_2, S_0)$		$f(S_4, S_0)$
S_1	c	$1 - f(S_1, S_0) - c - c$		$f(S_3, S_1)$	
S_2	c		$1 - f(S_2, S_0) - c - c$	$f(S_3, S_2)$	
S_3		c	c	$1 - f(S_3, S_1) - f(S_3, S_2) - c$	
S_4	c				$1 - f(S_4, S_0) - c - c$
S_5		c			c
S_6			c		c
S_7				c	

Table 7.4: The transition matrix of the continuous-time Markov chain associated to all combinations of C,D and E species (part 2/2). An empty cell means 0.

	S_5	S_6	S_7
S_0			
S_1	$f(S_5, S_1)$		
S_2		$f(S_6, S_2)$	
S_3			$f(S_7, S_3)$
S_4	$f(S_5, S_4)$	$f(S_6, S_4)$	
S_5	$1 - f(S_5, S_1) - f(S_5, S_4) - c$		$f(S_7, S_5)$
S_6		$1 - f(S_6, S_2) - f(S_6, S_4) - c$	$f(S_7, S_6)$
S_7	c	c	$1 - f(S_7, S_3) - f(S_7, S_5) - f(S_7, S_6)$