Large Lotka-Volterra model : when random matrix theory meets theoretical ecology

Maxime Clenet

Université Gustave Eiffel

Étude des grands systèmes de Lotka-Volterra : l'écologie théorique à travers les matrices aléatoires

Thèse de doctorat de l'Université Gustave Eiffel

École doctorale n°532, Mathématiques et STIC (MSTIC)
Spécialité : Mathématiques appliquées
Unité de recherche : Laboratoire d'Informatique Gaspard Monge (LIGM)

Thèse présentée et soutenue à l'Université Gustave Eiffel, le 9 Décembre 2022, par :

Maxime CLENET

Composition du Jury

Vincent BANSAYE Professeur, École Polytechnique	Examinateur
Vincent CALCAGNO Directeur de recherche, INRAE, CNRS, Université Côte d'Azur	Rapporteur
Walid HACHEM Directeur de recherche, CNRS, Université Gustave Eiffel	Invité
Christian MAZZA Professeur, Université de Fribourg	Rapporteur
Florence MERLEVÈDE Professeure, Université Gustave Eiffel	Examinatrice
Hélène LEMAN Chargée de recherche, INRIA, ENS Lyon	Examinatrice

Encadrement de la thèse

François MASSOL Directeur de recherche CNRS, Université de Lille, INSERM, CHU, Institut Pasteur de Lille	Directeur de thèse
Jamal NAJIM Directeur de recherche CNRS, Université Gustave Eiffel	Directeur de thèse

Remerciements

Mes premiers remerciements s'adressent naturellement à mes directeurs qui m'ont suivi et encadré d'une manière exceptionnelle pendant cette thèse et sans qui je n'aurais jamais pu accomplir ce travail. Ils m'ont ouvert la voie et aiguillé dans leur domaine de recherche respectif: les matrices aléatoires et l'écologie théorique. À Jamal Najim avec qui j'ai toujours eu des échanges fructueux lors de nos nombreuses réunions autant à l'université qu'à distance. Merci de ton pragmatisme et ta rigueur face à tous les problèmes de mathématiques sur lesquels on a planché et ton accompagnement minutieux pendant ces trois années. Je te remercie de m'avoir toujours conseillé vers le bon chemin et de m'avoir toujours poussé à présenter mes travaux de recherche lors de nombreuses conférences. À François Massol qui m'a fait découvrir le monde de l'écologie théorique et m'a donné envie de continuer à faire de la recherche dans ce domaine. Merci pour tous les échanges que l'on a eu pendant la thèse, la distance n'a jamais été un frein à nos discussions. Je suis et resterai toujours très impressionné de toutes les nombreuses références (et connaissances) que tu m'as transmises dans toutes les branches de l'écologie mais également en mathématiques. Merci également de m'avoir accueilli de nombreuses fois à Pasteur pour trouver des solutions à nos problèmes pas toujours facile...

J'aimerai ensuite remercier tous les membres du jury qui me font l'honneur de leur présence lors de ma soutenance de thèse et qui me donnent le privilège d'avoir un jury remarquable et ambivalent. Merci à Vincent Calcagno et Christian Mazza d'avoir rapporté ma thèse et toutes vos remarques pertinentes sur le manuscrit. Vos remarques sont d'autant plus importantes que vos travaux forment une pierre angulaire de cette thèse. Je remercie Florence Merlevède, Hélène Leman, Vincent Bansaye et Walid Hachem d'avoir accepté de faire partie de mon jury.

Cette thèse n'aurait probablement jamais eu lieu si je n'avais pas eu la chance d'avoir en professeur et directeur de mémoire Viet Chi Tran lors d'une partie de mes études à l'Université de Lille. Sa pédagogie, son parcours et ses innombrables connaissances sont un exemple pour moi. Merci de m'avoir accompagné lorsque je cherchais une thèse et de m'avoir conseillé de travailler avec Jamal et François.

Pendant ces trois années, j'ai pu collaborer avec de nombreux chercheurs dont les discussions fructueuses ont permis ou permettront l'élaboration de divers articles scientifiques. Merci à Hafedh El Ferchichi pour nos nombreuses discussions "elliptiques" qui ont permis une collaboration scientifique prolifique. Merci à mes nouveaux collaborateurs américains Zach, Stefano et Katja dont la collaboration inattendue donne lieu à une contribution scientifique ambitieuse. Merci à toutes les personnes qui participent au projet KARATE. Merci à Imane Akjouj, Mylène Maïda et Walid Hachem pour nos nombreuses réunions et échanges. Merci à Matthieu Barbier pour tous les conseils scientifiques que tu as partagés avec moi pendant ces trois ans. J'espère enfin pouvoir comprendre tous les outils des physiciens !

Je souhaite remercier tous les membres du laboratoire LIGM qui m'ont accueilli (quand c'était possible) pendant ces trois années. Merci aux membres de l'équipe MMSID : Théo, Philippe, Alexis et François-Xavier avec qui j'ai pu avoir des discussions scientifiques et non scientifiques lors de mon séjour au laboratoire. Je remercie en particulier Claire, Nadime, Nicolas, Philippe, Stéphane et toutes les personnes qui m'ont accompagné et aidé lors de mes missions d'enseignement. Enfin, je ne remercierai jamais assez Nathalie, Severine et Corinne de m'avoir aidé dans mes démarches, votre bonne humeur face à mes

milliards de questions et toutes nos super discussions. Grâce à vous, je n'ai jamais eu de problèmes administratifs et tout a toujours été très simple pour voyager à des événements scientifiques.

Merci à tous les doctorants du LIGM et également du LAMA avec qui j'ai échangé pendant de nombreux repas au RU. Merci à Doriann et Hélène pour nos super "séminaires du soir", à Thomas pour les inarrêtables discussions que l'on a pu avoir. Je remercie également Kayané, Florent et Zéphyr qui sont super géniaux et que j'ai plaisir à voir en dehors des murs de la fac ! Merci à Quentin et Elias, pour toutes nos discussions, nos courses à pied le samedi matin et les après-midi BU à la magnifique médiathèque de Vincennes qui porte bien son nom : coeur de ville. Merci à Adrien, Aaron, Julien, Josué, Niloufar, Corentin, Théo, Charlie et tous ceux que j'ai oubliés et croisés dans les couloirs du bâtiment Copernic.

Cette thèse a été rythmée par notre chère et tendre Covid qui nous a confiné chez nous pendant une partie des 3 ans... Je souhaite remercier Fanfan et Toch qui m'ont accueilli les bras ouverts pendant deux confinements, pour leur joie de vivre au quotidien et les nombreuses parties de palets et de fléchettes qui m'ont permis de garder une cadence de travail rythmée pendant ces quelques mois. J'aimerai remercier nos deux autres colocataires Ben et Romane avec qui j'ai bien rigolé !

Un gros big-up à mes copains Coco, Félix, Gabin, Greg, Jonathan, Jordane, Lieb et Louis-Quentin que je n'ai pas pu voir autant que je souhaitais pendant cette thèse mais avec qui j'ai toujours autant de plaisir à faire des randonnées, des soirées et pleins d'autres aventures qui m'ont permis de décompresser.

Bien évidemment je remercie ma famille qui m'a toujours accompagné et motivé pendant ces trois années. Je remercie Annick et Jean pour leur accueil toujours très chaleureux. Je remercie Isa, Flo et Louane pour toutes nos activités endiablées. Je remercie Marie alias Dauphinou avec qui j'ai pu partager de nombreux séjours à Challans city pour bosser (ou glander). Je remercie Maryvonne et Jacky de m'avoir accueilli les bras ouverts pendant toutes mes retraites studieuses qui ont été un vrai vecteur de ma réussite pendant mes études. Votre maison a toujours été un havre de paix pour moi et m'a permis de prendre du recul sur mes travaux de recherche et la vie ! Je remercierai jamais assez mon père de m'avoir donné goût à la recherche, ma mère pour son optimisme en toute circonstance et sa présence dans les moments difficiles et mes soeurs Claire et Nono pour leur bonne humeur, joie de vivre, conseils et d'être vous même parce que je vous kiffe !

Enfin, j'aimerai remercier la personne avec qui je vis maintenant à plein temps Coline. Merci de m'avoir supporté et soutenu pendant ces trois années, de remettre un peu de stabilité et de bons sens dans mes projets et d'être aussi enthousiaste que moi de partir pour de nouvelles aventures ensemble au Canada !

"Not only in research, but also in the everyday world of politics and economics, we would all be better off if more people realised that simple nonlinear systems do not necessarily possess simple dynamical properties." Robert M. May [May76].

Contents

List of Figures

Publications

In Prep:

1. M. Clenet, F. Massol, J. Najim, *Impact of a block structure in large systems of Lotka-Volterra* (Expected: early 2023)
2. S. Allesina, M. Clenet, K. Della Libera, F. Massol, Z. Miller, *A probabilistic perspective of the hierarchical competition-colonization trade-off model* (Expected: early 2023)
3. S. Allesina, M. Clenet, K. Della Libera, F. Massol, Z. Miller, *Sequences of invasions and community assembly in the hierarchical competition-colonization trade-off model* (Expected: early 2023)
4. I. Akjouj, M. Barbier, M. Clenet, W. Hachem, M. Maïda, F. Massol, J. Najim, V. Chi Tran, *Complex systems in ecology: a guided tour with the large Lotka-Volterra model and random matrices* (Expected: 2022)

Preprints:

1. M. Clenet, F. Massol, J. Najim, *Equilibrium and surviving species in a large Lotka-Volterra system of differential equations*, arXiv:2205.00735 (2022)

Published articles:

1. M. Clenet, H. El Ferchichi, J. Najim, *Equilibrium in a large Lotka-Volterra system with pairwise correlated interactions*, Stochastic Processes and their Applications (2022)

Conference papers:

1. M. Clenet, F. Massol, J. Najim, *Surviving species in a large Lotka-Volterra system of differential equations*, GRETSI (2022)
2. P. Bizeul, M. Clenet, J. Najim, *Positive Solutions for Large Random Linear Systems*, IEEE International Conference on Acoustics, Speech and Signal Processing (2020)

Code:

1. M. Clenet, *Equilibrium and surviving species in a large Lotka-Volterra system of differential equations* (Chapter 1): https://github.com/maxime-clenet/Equilibrium-and-surviving-species-in-a-large-Lotka-Volterra-system
2. M. Clenet, *Equilibrium in a large Lotka-Volterra system with pairwise correlated interactions* (Chapter 2): https://github.com/maxime-clenet/Feasibility-in-a-large-Lotka-Volterra-system-with-pairwise-correlated-interactions

Introduction en français

Enjeux écologiques

Une meilleure compréhension des écosystèmes

L'écologie est étymologiquement la science de la maison (du grec ancien oîkos). Par définition, c'est la science des êtres vivants (animaux, micro-organismes, etc.) dans un environnement spécifique à une échelle particulière (populations, espèces, communautés) et de leurs relations avec les autres êtres vivants. En écologie, une espèce est constituée d'individus qui peuvent se reproduire entre eux et produire une descendance fertile, formant ainsi des populations. Les individus sont généralement mutuellement dépendants les uns des autres pour leur survie.

Les individus de différentes espèces vivant dans une même région forment une communauté. L'ensemble des êtres vivants dans leur environnement forme un écosystème (savane, forêt, intestin, etc.). Historiquement, le concept d'écosystème est récent [Tan39]. La compréhension de ces écosystèmes (écologie des communautés, lorsque l'accent est mis sur l'interaction entre les espèces) et de leurs mécanismes de fonctionnement sous-jacents constitue un défi majeur en écologie [MW67].

Le nombre d'espèces dans un écosystème est souvent lié à une mesure d'échelle. Sur Terre, certains rares écosystèmes sont petits (2-3 espèces), cependant la plupart ont un très grand nombre d'espèces. Cette grande diversité d'espèces est nécessaire à la survie des êtres vivants. A ce jour, les scientifiques ont répertorié plus de 2 millions d'espèces sur Terre. Par exemple, la forêt amazonienne abrite 10^6 espèces. A notre échelle, notre microbiome abrite un ordre de grandeur de 10^3 espèces et 10^{18} cellules [CSF15]. De nombreuses questions sur ces grands écosystèmes restent sans réponse car les systèmes plus complexes demandent beaucoup plus de données empiriques pour être "compris" et les études expérimentales ne sont pas adaptées à leur étude. Une meilleure compréhension permettrait de gérer durablement les populations animales pour protéger les populations menacées, ou d'avoir une meilleure gestion des antibiotiques sur notre flore intestinale.

Les écologues mènent de nombreuses études expérimentales sur des systèmes à peu d'espèces, alors que dans les grands systèmes, il devient rapidement impossible de récolter des données à grande échelle. Cependant, ces dernières années, de nombreux outils technologiques ont été développés en laboratoire pour étudier des systèmes microbiologiques et permettraient de faire des comparaisons avec les études théoriques [HAB+21]. Par exemple, de nombreux processus sont automatisés avec l'émergence du deep learning pour reconnaître les espèces et les compter, notamment sur des images prises par avion ou drone. Ce manque de données peut être compensé par l'utilisation et l'étude de modèles. Ces modèles n'ont pas forcément pour seul but de prédire la dynamique de l'écosystème mais de comprendre les mécanismes qui permettent une grande diversité.

L'un des principaux débats en écologie concerne la relation entre la diversité et la stabilité d'un écosystème. Depuis longtemps, de nombreux écologues ont suggéré que la diversité des communautés renforçait la stabilité des écosystèmes [Mac55, May73]. Cependant, à partir d'un modèle théorique qu'il a introduit dans les années 70 [May72], May a remis en question la relation diversité-stabilité en utilisant une analyse de stabilité linéaire sur un modèle de communauté construit de manière aléatoire et a découvert que la diversité tend à déstabiliser le système. Cela a conduit au débat sur la diversité-stabilité [May73, Yod81, McC00, IC07, JMM$^+$16, LMB$^+$18] où les enjeux théoriques consistent à trouver les arguments ou mécanismes manquants du modèle de May.

Dans cette thèse, j'étudie les grands écosystèmes afin de comprendre l'un des principaux facteurs écologiques affectant leur diversité et leur dynamique : les interactions biotiques entre espèces. Dans une communauté, l'ensemble des interactions entre ses composantes est représenté comme un réseau d'interactions. Il existe deux grandes classes d'interactions : les interactions intra-spécifiques et les interactions inter-spécifiques. La première classe correspond aux interactions au sein d'une même espèce qui peuvent être négatives (compétition) et positives (effet Allee). La seconde, plus diversifiée, correspond aux interactions entre deux espèces différentes, par exemple la compétition, le mutualisme, la prédation, etc. L'écologie des communautés est une sous-discipline de l'écologie qui s'attache à comprendre l'évolution des abondances (:=nombre d'individus) des espèces qui composent une communauté au cours du temps. En bref, j'utilise un modèle qui décrit la dynamique de la communauté pour comprendre l'impact du réseau d'interactions sur les propriétés d'un équilibre telles que l'existence, la diversité, la stabilité, etc.

Cours éclair sur les EDOs

En écologie, la dynamique des populations peut être modélisée en temps continu ou discret. En temps continu, des équations différentielles ordinaires (EDO) sont utilisées pour décrire l'évolution de l'abondance $\boldsymbol{x} = (x_1, ..., x_n)$ d'un système à n espèces. On commence par rappeler le problème de Cauchy.

Soit U un ouvert de $\mathbb{R}^{n+1}$, $f : U \to \mathbb{R}^{n+1}$ continue par rapport à $(t, \boldsymbol{x})$,

$$\begin{cases} \frac{d\boldsymbol{x}(t)}{dt} = f(t, \boldsymbol{x}(t)) \,, \\ \boldsymbol{x}(0) = x_0 \,. \end{cases} \tag{1}$$

Un système est dit autonome si f ne dépend pas de t.

Le théorème de Cauchy-Lipschitz établit que si f est $\mathcal{C}^1$ par rapport à $\boldsymbol{x}$, alors pour toute condition initiale, le problème (1) admet une unique solution maximale (I, γ), $\gamma : I \mapsto \mathbb{R}^n$. De plus, toute autre solution du problème (1) est une restriction de la solution maximale.

Si le système est autonome, le principe de majoration a priori indique que si $f : U \to \mathbb{R}^n$ est continue et localement lipstchitzienne, (I, u) une solution maximale du théorème de Cauchy et sup(I) $= \infty$, alors il existe une solution globale au problème (1).
Dans la suite, on s'intéresse au problème autonome :

$$\begin{cases} \frac{d\boldsymbol{x}(t)}{dt} = f(\boldsymbol{x}(t)) \,, \\ \boldsymbol{x}(0) = x_0 \,. \end{cases} \tag{2}$$

Afin d'étudier le problème (2), une question importante consiste à obtenir des informations sur l'existence et l'unicité des équilibres et leurs propriétés. Un équilibre $\boldsymbol{x}^*$ du système

(2) est une solution de l'équation :

$$\frac{d\boldsymbol{x}(t)}{dt} = 0 \quad \Leftrightarrow \quad f(\boldsymbol{x}(t)) = 0\,.$$

La fonction f peut être complexe, les solutions de ce système ne sont pas nécessairement triviales et il peut y avoir plusieurs équilibres avec des propriétés différentes. Un équilibre $\boldsymbol{x}^*$ est dit faisable si toutes ces composantes sont strictement positives c.a.d.

$$\boldsymbol{x}^* > 0 \quad \Leftrightarrow \quad x_k > 0,\, \forall k \in [n]\,.$$

Une propriété majeure d'un point d'équilibre est sa stabilité. Un équilibre est stable s'il revient à sa valeur d'équilibre après une petite perturbation du vecteur d'abondance $\boldsymbol{x}$. Dans le cas d'une équation différentielle linéaire, l'étude de la stabilité est triviale et dépend des valeurs propres de l'opérateur linéaire. Dans le cas non linéaire, elle est plus complexe. Cependant, on peut linéariser le système pour obtenir des informations sur la stabilité locale autour de l'équilibre.

Soit $\boldsymbol{x}^*$ un point d'équilibre de (2), on dit que $\boldsymbol{x}^*$ est asymptotiquement stable si il est stable et si $\exists\, \delta > 0$, $\forall (I, \boldsymbol{x})$ solution de (2)

$$\exists\, t_0 \in I, |\boldsymbol{x}(t_0) - \boldsymbol{x}^*| \leqslant \delta \quad \Rightarrow \quad \boldsymbol{x}(t) \xrightarrow[t\to\infty]{} \boldsymbol{x}^*\,.$$

Théorème 0.1 (Stabilité d'un équilibre, cas non-linéaire)**.** *Soit $\boldsymbol{x}^*$ un équilibre d'un système différentiel non linéaire autonome où f est différentiable en $\boldsymbol{x}^*$ et soit $Df(\boldsymbol{x}^*) = J|_{\boldsymbol{x}^*}$ sa Jacobienne au point d'équilibre. Soit $\Lambda = \mathrm{Sp}(J|_{\boldsymbol{x}^*})$ l'ensemble des valeurs propres de $J|_{\boldsymbol{x}^*}$ et $\mathcal{R}(\Lambda)$ l'ensemble de la partie réelle des valeurs propres de Λ.*

1. *Si $\forall \lambda \in \Lambda, \mathcal{R}(\Lambda) < 0$, alors $\boldsymbol{x}^*$ est asymptotiquement stable et on a*

$$\forall \mu \in]0, \min -\mathcal{R}(\Lambda)[, \forall \epsilon > 0, \exists \delta > 0,\ |\boldsymbol{x}(t_0) - \boldsymbol{x}^*| < \delta \\ \Rightarrow \forall t \geqslant t_0, \boldsymbol{x}(t) \text{ existe et } |\boldsymbol{x}(t) - \boldsymbol{x}^*| \leqslant \epsilon e^{-\mu(t-t_0)}\,.$$

2. *Si $\exists \lambda \in \Lambda, \mathcal{R}(\Lambda) > 0$, alors $\boldsymbol{x}^*$ est instable.*

3. *Si $\forall \lambda \in \Lambda, \mathcal{R}(\Lambda) \leqslant 0$ et il y a des valeurs propres purement imaginaires, alors on ne peut pas conclure.*

Soit $\boldsymbol{x}^*$ un équilibre pour un système différentiel non linéaire autonome (2), on dit que $\boldsymbol{x}^*$ est asymptotiquement globalement stable si pour tout $\boldsymbol{x}_0 > 0$, la solution de (2) qui commence à $\boldsymbol{x}(0) = \boldsymbol{x}_0$ satisfait

$$\boldsymbol{x}(t) \xrightarrow[t\to\infty]{} \boldsymbol{x}^*\,.$$

Remarque 0.1. Dans cette thèse, j'étudie exclusivement la stabilité asymptotique, c'est-à-dire la stabilité et la convergence vers le point d'équilibre. Par abus de notation, je réfère l'étude de la stabilité à la stabilité asymptotique. Je conseille le livre de Hirsch *et al.* [HSD74] pour un cours complet sur les EDOs.

Le modèle de Lotka-Volterra

Les équations différentielles sont fréquemment utilisées en biologie pour décrire un système d'espèces en interaction. Une forme particulièrement utilisée est le modèle densité-dépendant :

$$\begin{cases} \frac{d\boldsymbol{x}(t)}{dt} = \boldsymbol{x}(t) f(\boldsymbol{x}(t)) \,, \\ \boldsymbol{x}(0) = x_0 \,, \end{cases} \tag{3}$$

où f est communément appelé valeur sélective ou taux de croissance d'une espèce. Quand $f(\boldsymbol{x}(t)) < 0$ la dynamique sera décroissante et inversement lorsque $f(\boldsymbol{x}(t)) > 0$. Si le système est défini dans l'orthant positif $\mathbb{R}^n_+$, alors le système est dit invariant : si $\forall\, x_i(0) \geqslant 0$, alors $\forall t > 0 : x_i(t) \geqslant 0$ [HS98]. L'un des modèles densité-dépendant les plus utilisés en écologie est le modèle de Lotka-Volterra qui est une pierre angulaire de cette thèse.

Historiquement, Thomas Robert Malthus (1766-1834) s'est intéressé à la modélisation des fluctuations d'une population. Sa conclusion était que sans contraintes, l'abondance d'une population croît de façon exponentielle

$$\begin{cases} \frac{dx(t)}{dt} = rx(t) \,, \\ x(0) = x_0 \,, \end{cases}$$

où r = naissance − mort et la solution analytique est $x(t) = x_0 e^{rt}$. Une solution qui manque un peu de réalisme.

Plus tard, Pierre François Verhulst (1804-1849) s'est intéressé à un modèle plus réaliste en supposant que le modèle est limité par une taille maximale $K > 0$ (capacité de charge)

$$\begin{cases} \frac{dx(t)}{dt} = rx(t) \left(1 - \frac{x(t)}{K}\right) \,, \\ x(0) = x_0 \,. \end{cases}$$

Son modèle logistique représente, par exemple, la limite de croissance d'une population de zèbres dans la savane en raison de la pénurie de ressources.

Dans un second temps, les scientifiques se sont intéressés à la modélisation des interactions entre espèces. Lorsqu'on ajoute les interactions entre les populations, le modèle le plus simple porte le nom de deux scientifiques, Lotka et Volterra, qui l'ont formulé indépendamment à la fin des années 20 [Lot25, Vol26]. Classiquement étudié sous la forme d'un modèle proie-prédateur à 2 dimensions, il a été comparé à des données issues de populations naturelles [Huf58].

D'un point de vue général et dans des dimensions supérieures, les équations de Lotka-Volterra ou le modèle de Lotka-Volterra généralisé jouent un rôle clé dans l'étude de la dynamique des populations dans le temps. Ce modèle s'analyse mathématiquement, il est aussi très polyvalent et robuste, et constitue une première étape dans le développement des modèles écologiques. Ce modèle a été étudié à la fois en écologie [Wan78, Jan87, LB92] en en mathématiques [GJ77, Goh77, Tay88, HS98, Tak96].

D'un point de vue mathématique, ce modèle décrit la dynamique de la population d'un système à n espèces. Il est défini par un système à n équations différentielles :

$$\frac{dx_k(t)}{dt} = x_k(t) \left(r_k - \theta x_k(t) + \sum_{\ell \in [n]} B_{k\ell} x_\ell(t) \right) \,, \tag{4}$$

où $k \in [n] = \{1, \cdots, n\}$. L'abondance de l'espèce k au temps t est représentée par $x_k(t)$ et $\boldsymbol{x} = (x_1, \cdots, x_n)$ est le vecteur des abondances des différentes espèces. Le paramètre θ est le coefficient d'autorégulation ou l'interaction intraspécifique de chaque espèce. Le paramètre r_k correspond au taux de croissance intrinsèque de l'espèce k. Le coefficient $B_{k\ell}$ est l'effet per capita des espèces ℓ sur le taux de croissance de l'espèce k. La matrice B, représentant la structure du réseau d'interactions, peut souvent être décomposée sous différentes formes : blocs, cascade, réseaux multiplex, graphons, etc. [CN88, SCG+05, LIPJ+06, PEM12].

L'objectif de nombreux mathématiciens et écologues est de comprendre le comportement du système en fonction de ces différents paramètres d'entrée. Par exemple, le nombre d'équilibres, leur stabilité et leur faisabilité pour comprendre les implications écologiques qui en découlent.

Comme nous l'avons déjà mentionné, le problème majeur du travail avec de grands systèmes est la difficulté d'observer ou d'estimer les informations sur la matrice d'interaction. Un choix naturel consiste à remplacer les interactions par des coefficients aléatoires dont les propriétés statistiques (moyenne, variance, etc.) et la structure (blocs, cascade, etc.) codent certaines des véritables propriétés du réseau trophique. La matrice B devient un objet mathématique complexe : une matrice aléatoire. Cet objet mathématique représente la deuxième pierre angulaire de cette thèse.

L'émergence des matrices aléatoires en écologie

Dans les années 70, suite aux travaux de Gardner et Ashby [GA70], Robert May a réouvert le débat de longue date sur la relation entre diversité et stabilité en écologie [Mac55]. Son article fondateur [May72] a motivé l'émergence de matrices aléatoires en tant qu'outil mathématique clé pour caractériser les écosystèmes en grande dimension. Une meilleure compréhension de ces outils a amélioré notre compréhension de la nature des interactions et des réseaux trophiques pour parvenir à la stabilité [Tay88, AT12, TPA14]. Dans son étude, May s'est intéressé au modèle (2) [May73], en supposant que le système est à un équilibre faisable $\boldsymbol{x}^*$. D'après le théorème 0.1, l'étude de la stabilité locale correspond à l'étude des valeurs propres réelles de la matrice Jacobienne du système au point d'équilibre. La matrice Jacobienne du système (2) est :

$$J = (J_{k\ell})_{n\times n}\,, \quad J_{k\ell} = \frac{\partial f_k(\boldsymbol{x})}{\partial x_\ell}\,.$$

Il existe une matrice $M = J|_{x^*}$, appelée "matrice des communautés" (Jacobienne) qui décrit l'effet de l'espèce ℓ (colonne) sur l'espèce k (ligne) autour du point d'équilibre.

May a remis en question une croyance centrale en écologie en affirmant que des réseaux écologiques suffisamment grands ou complexes ont une probabilité d'être stables proches de zéro. Pour établir ce point, il a analysé la stabilité de grands réseaux dans lesquels les espèces interagissent au hasard. Dans ce cas, la matrice Jacobienne est une matrice aléatoire non-Hermitienne

$$M = -I + A\,,$$

où A est une matrice aléatoire centrée $n \times n$ composée de variables aléatoires $A_{k\ell} \sim \mathcal{N}(0, \sigma^2)$ qui apparaissent avec une probabilité C et valent 0 sinon (le paramètre C est appelé connectance). Lorsque n est grand, May a prouvé que la probabilité que le point

d'équilibre soit stable est proche de 0 chaque fois que la "complexité" satisfait :

$$\sigma\sqrt{nC} > 1\,.$$

Les valeurs propres de la matrice M sont distribuées selon la loi circulaire, dans un disque de centre $(-1, 0)$ et de rayon $\sigma\sqrt{nC}$ (voir Figure 1). La condition de stabilité est $\mathcal{R}(\mathrm{Sp}(M)) < 0$. Si le modèle a un grand nombre d'espèces connectées avec de fortes interactions, alors le modèle est susceptible d'être plus instable.

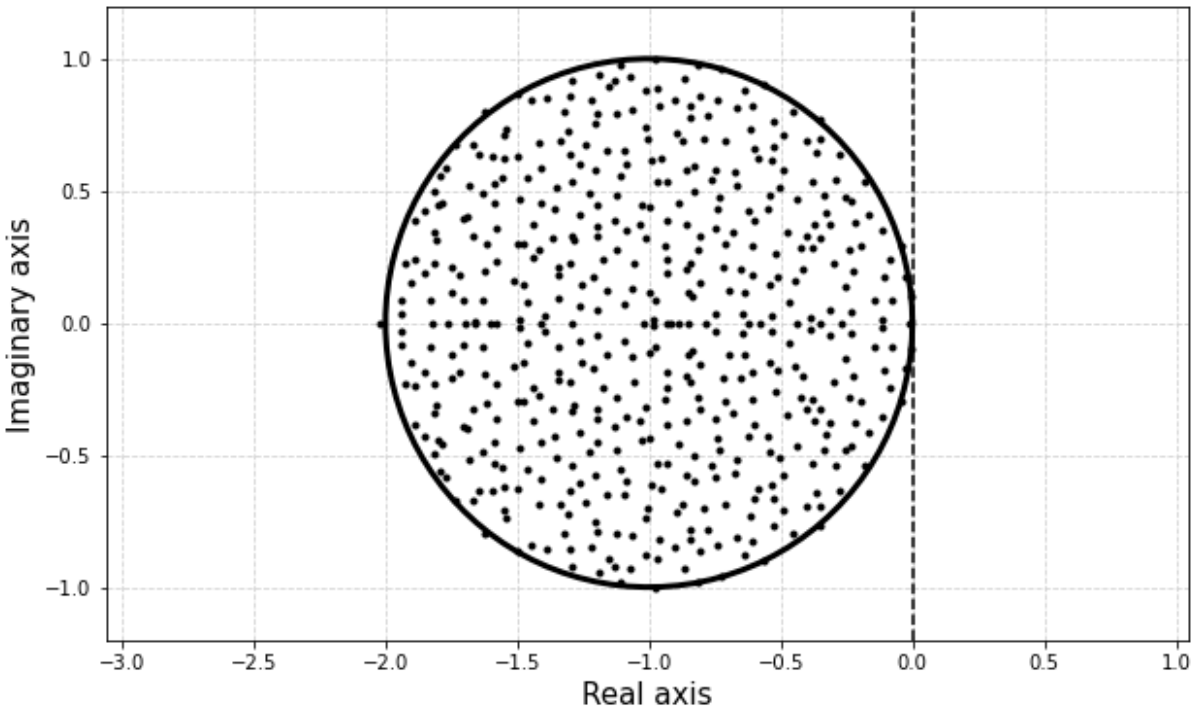


Figure 1: Spectre de la matrice aléatoire Jacobienne (matrice non-Hermitienne) $M = -I + A$ dans le plan complexe ($n = 500$, $\sigma = 1$, $C = 1$). Le cercle en trait plein représente la limite de la loi circulaire. La ligne pointillée représente le seuil à ne pas dépasser pour la partie réelle des valeurs propres pour que le système soit stable.

Dans un modèle densité-dépendant (3), la Jacobienne est évaluée par

$$J_{k\ell} = \delta_{k\ell} f_k(\boldsymbol{x}) + x_k \frac{\partial f_k(\boldsymbol{x})}{\partial x_\ell}\,.$$

Dans le système de Lotka-Volterra (4), à l'équilibre faisable x^*, la matrice Jacobienne dépend de l'abondance des espèces à l'équilibre

$$J|_{\boldsymbol{x}^*} = \mathrm{diag}(\boldsymbol{x}^*)(-I + B)\,,$$

où B est introduite à (4). L'étude de la Jacobienne est plus difficile car $\boldsymbol{x}^*$ et B ne sont plus indépendants. Cependant, Stone [Sto18] et Gibbs *et al.* [GGRA18] ont montré que les intuitions restent similaires, la stabilité des grands systèmes LV est uniquement déterminée par la matrice d'interaction.

Formulation finale du modèle

Dans la majeure partie de la thèse, nous nous concentrons sur le modèle (4) où $r_k = 1, \forall k \in [n]$, $\theta = 1$ et nous ajoutons un paramètre de normalisation générique dans la matrice d'interaction. Des taux de croissance égaux sont choisis permettant de réduire le nombre de paramètres et de simplifier grandement les calculs. La compréhension de l'impact de la matrice d'interaction B dans le système LV pose de nombreux problèmes ouverts. Cependant, Rohr *et al.* [RSB14] se sont intéréssés à l'étude des taux de croissance permettant la coexistence entre les espèces. Leurs travaux sur la stabilité structurelle ont été étendus par Cenci *et al.* [CS18], Saavedra *et al.* [SRB+17] et Grilli *et al.* [GAS+17].

Le choix de garder le paramètre d'autorégulation $\theta = 1$ apporte de l'intelligibilité. Dans le cas $\theta \neq 1$, on peut redimensionner le système (32) pour supprimer le paramètre θ en posant $\tilde{x}_k := \theta x_k$, $\tilde{B}_{k\ell} := B_{k\ell}/\theta$. Dans les faits, Barabás *et al.* [BMSA17] ont étudié l'importance d'avoir un terme d'autorégulation fort pour que le système ait un équilibre stable. A noter que les valeurs de la diagonale de la matrice B ne sont pas fixées à 0, mais leurs valeurs microscopiques ont un impact négligeable sur les résultats asymptotiques.

$$\frac{dx_k(t)}{dt} = x_k(t)\left(1 - x_k(t) + \sum_{\ell\in[n]} B_{k\ell}x_\ell(t)\right) . \tag{5}$$

Le dernier détail important qui diffère de la notation standard du modèle de Lotka-Volterra est que l'on ajoute un paramètre de normalisation $1/\sqrt{n}$ dans la matrice B. La raison théorique majeure est de limiter l'impact des paramètres d'interaction sur les autres termes tout en gardant une influence asymptotique c.a.d.

$$\mathbb{E}\left(\sum_{\ell\in[n]} B_{k\ell}x_\ell(t)\right) \sim O(1) \quad ; \quad \mathrm{Var}\left(\sum_{\ell\in[n]} B_{k\ell}x_\ell(t)\right) \sim O(1)\,.$$

D'un point de vue écologique, on peut imaginer que quand le nombre d'espèces augmente dans un écosystème, alors la force des interactions d'une espèce sur les autres a tendance à diminuer.

Généralement, le modèle peut finalement s'écrire sous la forme compacte :

$$\frac{dx_k}{dt} = x_k\,(1 - x_k + (B\boldsymbol{x})_k) \;, \quad k \in [n]\,, \tag{6}$$

où B reste à déterminer.

Le réseau d'interactions

L'enjeu majeur de cette thèse est de comprendre l'impact de la matrice d'interaction B sur la dynamique du modèle Lotka-Volterra. Dans la nature, B correspond au réseau d'interactions entre les espèces ou peut être considéré comme le réseau trophique de l'écosystème (dans le sens "qui mange qui ?"). Les réseaux écologiques vastes et fortement connectés sont fréquents dans la nature [DWM02, PLC91].

Dans le système (5), un modèle général pour la matrice d'interaction B est une matrice aléatoire non centrée avec des interactions corrélées deux à deux combiné avec une structure de graphe :

$$B = S \circ \left(\frac{A}{\alpha\sqrt{n}} + \frac{\mu}{n}\mathbf{1}_n\mathbf{1}_n^\top\right) , \tag{7}$$

où $\circ$ est le produit d'Hadamard c.a.d. $(X \circ Y)_{ij} := X_{ij}Y_{ij}$ et le vecteur $\mathbf{1}_n$ de taille $n \times 1$ est composé de uns. La matrice aléatoire $A = (A_{k\ell})_{k,\ell\in[n]}$ satisfait les conditions suivantes :

1. $(A_{k\ell}, k \leqslant \ell)$ sont des variables aléatoires indépendantes identiquement distribuées (i.i.d.) et $\mathbb{E}(A_{k\ell}) = 0$, $\mathbb{E}(|A_{k\ell}|^2) = 1$ et $\mathbb{E}(|A_{k\ell}|^4) < \infty \, \forall \, 1 \leqslant k \leqslant \ell$.

2. pour $k < \ell$ le vecteur $(A_{k\ell}, A_{\ell k})$ a une distribution standard bivariée, indépendante des autres variables aléatoires, avec une covariance $\mathrm{cov}(A_{k\ell}, A_{\ell k}) = \mathbb{E}(A_{k\ell}A_{\ell k}) = \rho$ avec $|\rho| \leqslant 1$.

$S := (S_{k\ell})_{k,\ell\in[n]}$ est la matrice d'adjacence d'un graphe c.a.d. si on représente chaque espèce par un nœud et une interaction entre deux espèces par un sommet dirigé, alors

$$S_{k\ell} = \begin{cases} 1 & \text{si il existe un impact de l'espèce } \ell \text{ sur } k, \\ 0 & \text{sinon.} \end{cases}$$

En l'absence de toute autre information, on peut supposer que S est la matrice d'adjacence d'un graphe d'Erdös Renyi (ER) [ER60]. C'est un graphe avec n sommets. On suppose qu'il existe une arête entre deux sommets avec une probabilité p indépendante de toute autre arête.

D'un point de vue écologique, deux types de structures peuvent affecter le type d'interactions et l'existence de communautés avec des interactions préférentielles.

D'une part, le type d'interaction est différent selon les espèces. Ces paramètres sont gérés par des choix sur les propriétés statistiques des variables aléatoires B_{ij}. L'ensemble des paramètres (α, μ, ρ) peut représenter une gamme de types d'interaction. Tout d'abord, la force de l'interaction est représentée par α, une grande valeur de α représente un système avec des interactions faibles, au contraire une petite valeur de α représente des interactions très fortes. Les paramètres μ et ρ décrivent la nature des interactions du système. Quand $\rho < 0$, les interactions entre partenaires ont un impact opposé l'une sur l'autre, comme dans les interactions antagonistes (le prédateur est influencé positivement par l'abondance de sa proie tandis que la proie est affectée négativement par celle du prédateur). Lorsque $\rho > 0$, les interactions entre les partenaires ont un impact similaire les uns sur les autres, c'est-à-dire qu'ils sont engagés dans des interactions mutualistes ou compétitives. Le paramètre d'interaction moyen μ augmente la proportion d'interactions compétitives ou mutualistes en fonction de son signe. Étant donné une interaction par paire $B_{k\ell}/B_{\ell k}$ dans le système, les trois motifs dominants sont :

- compétition (relation -/-), ce qui se produit plus souvent lorsque $\rho > 0$, $\mu < 0$ [Mac70, Zee95],
- mutualisme (relation +/+), ce qui se produit plus souvent lorsque $\rho > 0$, $\mu > 0$ [SGB+15, Sto20],
- prédation (relation +/-), ce qui se produit plus souvent lorsque $\rho < 0$, $\mu \approx 0$ [AT12].

Il existe d'autres types d'interactions comme le commensalisme ou l'amensalisme [BTH06].

D'autre part, la structure du réseau d'interactions diffère selon les écosystèmes. La structure de la matrice d'interaction peut également être affectée par l'existence de communautés, c'est-à-dire de groupes d'espèces qui interagissent préférentiellement entre

eux [TF10, AGB+15]. Dans l'équation (5), le réseau est représenté par S une matrice d'adjacence d'un graphe donné. Plusieurs types de structures sont largement étudiés en écologie et peuvent être modélisés par le graphe S.

Tout d'abord, les structures modulaires telles que la compartimentalisation des réseaux trophiques, également appelée modularité, qui sont la propension des nœuds à être connectés préférentiellement au sein des groupes plutôt qu'entre les groupes ([GSSP+10, GRA16]).

Deuxièmement, les structures imbriquées où chaque espèce joue un rôle différent dans l'écosystème. Ces modèles sont généralement appelés imbriqués en raison de leur structure dans laquelle certaines espèces ont plus d'interactions que d'autres [BJMO03, BFPG+09, SKA13, PBHM19].

Dans les grands écosystèmes, toutes les espèces n'interagissent pas les unes avec les autres, d'où l'intérêt d'étudier les écosystèmes épars [BSHM17]. Dans les faits, May [May72] considère la connectance comme un paramètre clé lié à la complexité d'un système.

Enfin, les modèles de traits avec des structures latentes [EJK+13] comme le modèle de niche de Williams et Martinez [WM00], le modèle en "cascade" [CBN90] qui établit une structure de prédation dans le réseau trophique. Chaque espèce peut se nourrir aux niveaux trophiques inférieurs mais pas aux niveaux supérieurs [HMS16, PBG22]. A noter que toutes ces structures peuvent être modélisées par un graphon (chaque nœud est associé à une variable aléatoire et les connexions dépendent d'une fonction des variables associées aux deux nœuds).

Pour un aperçu général des différents modèles des communautés écologiques complexes du point de vue du physicien, voir Barbier *et al.* [BABL18].

Métacommunauté et dynamique spatiale

À des échelles spatiales plus grandes, les écologistes s'intéressent également aux interactions entre les populations (plutôt qu'entre les individus) afin de comprendre les schémas de la diversité des espèces dans l'espace et le temps. Dans ce contexte, les interactions mutualistes, compétitives et prédatrices sont remplacées par les processus de colonisation, d'extinction et de remplacement de populations entières. La théorie de l'écologie spatiale trouve ses racines dans les travaux de MacArthur et Wilson [MW63] et Levins et Heatwole [LH63] et plus tard MacArthur sur la biologie et géographie des populations [Mac84]. En particulier, *The Theory of Island Biogeography* (TIB) est une pierre angulaire de la théorie de la dynamique spatiale [MW67]. Le TIB décrit comment la biodiversité des îles est maintenue par un équilibre entre l'immigration et l'extinction des espèces.

Depuis ce jour, la théorie de l'écologie spatiale a évolué pour comprendre les mécanismes de coexistence qui sous-tendent le modèle de métacommunauté introduit par Wilson. [Wil92]. L'étude des modèles de métacommunautés s'est particulièrement développée en raison de la prise de conscience de l'hétérogénéité spatiale des écosystèmes. Leibold *et al.* [LHM+04] ont décrit les différents mécanismes à l'échelle spatiale : la colonisation (les organismes se déplacent d'un site à l'autre entre les générations), l'habitat de niche (les espèces peuvent être plus ou moins bien adaptées à un environnement donné) et la stochasticité (si les espèces sont équivalentes en termes de traits, de compétitivité, etc., on ne s'attend pas nécessairement à ce qu'elles coexistent, mais on sait qu'il faudra un certain temps avant que l'une ou l'autre ne prenne complètement le contrôle du site, c'est-à-dire

un modèle neutre).

Considérant que la capacité des espèces à coloniser de nouveaux habitats est cruciale pour le maintien des populations, un type de modèle de métacommunauté mathématiquement analysable est connu sous le nom de modèle d'occupation avec un compromis compétition-colonisation. La légitimité de ce modèle repose sur l'existence d'un compromis entre la capacité d'une espèce à coloniser de nouveaux patchs et sa capacité concurrentielle. Cette capacité concurrentielle affecte sa résistance à la colonisation par une autre espèce et sa propre capacité à remplacer d'autres espèces. Formellement, il s'agit d'un modèle d'occupation de parcelles où chaque espèce a la capacité de coloniser de nouvelles parcelles en compétition avec d'autres espèces. La variable d'intérêt est la proportion de l'habitat occupée par chaque espèce.

Initiallement étudié par Levins [Lev69] et Levins et Culver [LC71] dans le cas de deux espèces, ce modèle a suscité un intérêt particulier dans sa version à n espèces où la compétition est hiérarchique [Has80, NM92, Til94]. Dans un cadre plus général, il a été étudié dans un cas où la compétition n'est pas hiérarchique [Ama03, YW01, CMJD06a] et aussi dans un contexte épidémiologique d'une dynamique d'interactions hôte-parasite [MN94, NM94]. Dans un cadre plus général, des approches de matrice aléatoire ont été utilisées pour étudier la stabilité dans un contexte de méta-écosystème. Chaque parcelle a sa propre dynamique et la dispersion de toutes les espèces relie les différentes parcelles, voir Gravel *et al.* [GML16].

La dynamique spatiale d'un système à n espèces avec un dilemme compétition colonisation [CMJD06a] est de la forme

$$\frac{dp_k}{dt} = c_k p_k \left(1 - \sum_{\ell=1}^{n} p_\ell\right) - m_k p_k + c_k p_k \sum_{\ell \neq k} p_\ell \eta_{k\ell} - p_k \sum_{\ell \neq k} c_\ell p_\ell \eta_{\ell k}\,, \tag{8}$$

où p_k représente le taux d'occupation de l'espèce k, m_k est le taux d'extinction de l'espèce k, c_k représente le taux de colonisation de l'espèce k, $\eta_{k\ell}$ correspond à la probabilité de remplacement de l'espèce ℓ par k.

Ces équations peuvent être représentées comme un modèle de compétition de Lotka-Volterra avec des interactions asymétriques :

$$\frac{dp_k}{dt} = p_k \left[c_k - m_k + \sum_{\ell=1}^{n} p_\ell \left(c_k \eta_{k\ell} - c_\ell \eta_{\ell k} - c_k\right)\right].$$

Dans le contexte du modèle de Lotka-Volterra avec de la dispersion, des travaux supplémentaires ont été effectués en introduisant un paramètre de migration [BG20, PNJ21, VPNJ22]. Dans le contexte du méta-réseau trophique avec des paramètres de diffusion, Brechtel *et al.* [BGR+18] ont étudié la formation de motifs par diffusion dans les réseaux.

Théorie des matrices aléatoires : une visite guidée

Historiquement, la théorie des matrices aléatoires trouve ses racines dans les travaux du statisticien John Wishart dont le but était d'étudier les matrices aléatoires de covariance empirique d'échantillons gaussiens multivariés [Wis28]. Par la suite, dans les années 50, une deuxième impulsion a été donnée par Eugène Wigner [Wig55] dont le but était d'expliquer la distribution des niveaux d'énergie dans les noyaux atomiques. L'approche

innovante utilisée par Wigner [Wig67] pour décrire le spectre d'une matrice aléatoire Hermitienne a été reprise par d'autres physiciens pour résoudre des problèmes de physique nucléaire [Dys62] et de sciences physiques. Par la suite, de nouvelles structures matricielles ont été étudiées, de nombreux travaux ont été réalisés par Marchenko et Pastur [MP67] sur les grandes matrices de covariance et Girko [Gir85], Bai [Bai97] et Silverstein [SC95, BS10] sur l'extension des résultats aux matrices non-Hermitiennes. Jusqu'à aujourd'hui, une multitude de travaux ont été publiés dans des domaines très divers des mathématiques tels que la combinatoire, les graphes aléatoires, la théorie des probabilités libres, la théorie du signal, la théorie des nombres, etc.

La force de la théorie des matrices aléatoires vient de la stabilisation du spectre des matrices (aléatoire et a priori compliqué) lorsque la dimension de la matrice tend vers l'infini. Dans ce cadre, la distribution des valeurs propres de la matrice devient complètement déterministe. De manière très simplifiée, il s'agit d'une équivalence de la loi des grands nombres pour le spectre d'une matrice. Les enjeux et motivations de la théorie des matrices aléatoires reposent sur la description des propriétés standard du spectre des matrices : valeurs propres, vecteurs propres, plus grande valeur propre, etc. D'un point de vue technique, il s'agit d'un mélange équilibré d'algèbre linéaire, de probabilités, d'analyse complexe et de combinatoire.

Quelques definitions

Soit $A \in \mathcal{M}(\mathbb{C})$, $A := (A_{k\ell})_{n\times n}$, une matrice carrée de taille n avec des coefficients appartenant à l'ensemble complexe $\mathbb{C}$. On note $A^* := \overline{A}^\top$. Soit un vecteur $\boldsymbol{x} \in \mathbb{R}^n$, on note $\|x\|_2$ sa norme Euclidienne :

$$\|\boldsymbol{x}\|_2 = \left(\sum_{k=1}^{n} |x_k|^2\right)^{1/2}.$$

Définition 0.1 (Valeurs propres)**.** On définit $\lambda_1(A)$, $\lambda_2(A)$, …, $\lambda_n(A)$ les valeurs propres de A c.a.d. les racines de son polynôme caractéristique, tel que

$$|\lambda_1(A)| \geqslant \cdots \geqslant |\lambda_n(A)|\,.$$

L'ensemble des valeurs propres de A est appelé le spectre de A et noté $\mathrm{Sp}(A)$.

Définition 0.2 (Rayon spectral)**.** La rayon spectral de la matrice A, que l'on note $\rho(A) = |\lambda_1(A)|$, est le module de la valeur propre ayant le plus grand module.

Définition 0.3 (Valeurs singulières)**.** Les valeurs singulières $\sigma_1(A)$, $\sigma_2(A)$, …, $\sigma_n(A)$ de la matrice A sont la racine carrée des valeurs propres de la matrice hermitienne A^*A c.a.d.

$$\sigma_i(A) := \sqrt{\lambda_i(A^*A)}\,, \quad \forall i \in [n]\,.$$

Définition 0.4 (La norme spectrale)**.** La norme spectrale de la matrice A notée par $\|A\|$ est définie par sa plus grande valeur singulière

$$\|A\| := \max\left(\sqrt{\lambda}, \lambda \text{ valeur propre de } A^*A\right) = \sigma_1(A)\,.$$

En probabilité, la mesure spectrale caractérise le spectre d'une matrice. Dans le domaine des RMT, elle est utilisée pour exprimer les résultats de convergence du spectre vers une mesure déterministe. Soit $I \subset \mathbb{C}$, on note δ_λ la mesure de Dirac au point λ définie par

$$\delta_\lambda(I) = \begin{cases} 1 \text{ si } \lambda \in I\,, \\ 0 \text{ sinon }. \end{cases}$$

Définition 0.5 (Mesure spectrale empirique)**.** Soit $A \in \mathcal{M}_n(\mathbb{C})$ et ses valeurs propres $\lambda_1(A), \ldots, \lambda_n(A)$, on définit la mesure spectrale empirique de la matrice A dans $(\mathbb{C}, \mathcal{B}(\mathbb{C}))$ par

$$\mu_A := \frac{1}{n}\sum_{k=1}^{n} \delta_{\lambda_k(A)}\,.$$

Pour tout sous-ensemble $E \subset \mathbb{C}$, la quantité :

$$\mu_A(E) = \frac{\text{card}\{1 \leqslant k \leqslant n : \lambda_k(A) \in E\}}{n}\,,$$

est la proportion de valeurs propres de A dans E.

La convergence faible d'une mesure spectrale empirique vers une mesure déterministe décrit de nombreux résultats de matrices aléatoires.

Définition 0.6 (Convergence faible)**.** On dit que μ_A convergence faiblement vers une mesure de probabilité μ c.a.d. $\mu_A \xrightarrow[n\to\infty]{\mathcal{D}} \mu$, si pour toute fonction f continue et bornée sur $\mathbb{R}$

$$\int f(u)\mu_A(du) = \frac{1}{n}\sum_{k=1}^{n} f(\lambda_k) \xrightarrow[n\to\infty]{} \int f(u)\mu(du)\,.$$

Remarque 0.2. Si A est aléatoire, alors μ_A est une distribution de probabilité aléatoire discrète, cela implique $\int f\mu_A(du)$ sont aussi des variables aléatoires. Nous dirons alors que presque sûrement (p.s.) μ_A convergence faiblement vers μ

$$(p.s.) \quad \mu_A \xrightarrow[n\to\infty]{\mathcal{D}} \mu\,.$$

Définition 0.7 (Résolvante)**.** Soit $A \in \mathcal{M}_n(\mathbb{C})$, on appelle résolvante de la matrice A $Q := (Q_{k\ell})_{n\times n}$ définie par

$$Q(z) = (A - zI)^{-1},\ z \notin \text{Sp}(A)\,.$$

On note

$$\mathbb{C}^+ := \{z \in \mathbb{C} \,:\, \text{Im}(z) > 0\}$$

la moitié supérieure du plan complexe.

Définition 0.8 (Transformée de Stieltjes)**.** Soit $\mu \in \mathcal{P}(\mathbb{R})$ une mesure de probabilité. La transformée de Stieltjes de μ notée $g_\mu : \mathbb{C}^+ \to \mathbb{C}$ est définie par

$$g_\mu(z) = \int \frac{1}{\lambda - z}\mu(d\lambda)\,, z \in \mathbb{C}^+\,.$$

Remarque 0.3. Soit μ_A la mesure empirique des valeurs propres $\lambda_1(A), \cdots, \lambda_n(A)$ de la matrice symétrique A, alors la transformée de Stieljes associée est donnée par

$$g_{\mu_A}(z) = \int \frac{1}{\lambda - z}\mu_A(d\lambda) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\lambda_i - z} = \frac{1}{n}\mathrm{Tr}\left((A - zI)^{-1}\right) ,$$

où $Q = (A - zI)^{-1}$ est la résolvante de la matrice A et $\mathrm{Tr}(Q)$ est la trace de la matrice Q.

Proposition 0.2 (Inversion de Stieltjes)**.** *Soit g_μ la transformée de Stieltjes de la mesure μ de masse finie $\mu(\mathbb{R})$. Si $a, b \in \mathbb{R}$ et $\mu(\{a\}) = \mu(\{b\}) = 0$, alors*

$$\mu(a, b) = \frac{1}{\pi}\lim_{y\to 0^+}\mathrm{Im}\int_a^b g_\mu(x + iy)dx ,$$

et

$$\forall x \in \mathbb{R},\ \mu(\{x\}) = \frac{1}{\pi}\lim_{y\to 0^+}\mathrm{Im}(g_\mu(x + iy)) .$$

Proposition 0.3 (Identité de Woodbury)**.** *Soit A une matrice de taille $n \times n$, U une matrice $n \times m$, B une matrice $m \times m$, V une matrice $m \times n$. On suppose que toutes les inverses des matrices considérées existent, alors :*

$$(A + UBV)^{-1} = A^{-1} - A^{-1}U(B^{-1} + VA^{-1}U)^{-1}VA^{-1} .$$

L'identité de Woodbury pour une perturbation de rang 1 est souvent utilisée et appelée identité de Sherman-Morrison.

Proposition 0.4 (Identité de Sherman-Morrison)**.** *Soit A une matrice $n \times n$ et u, v deux vecteurs de dimension n. On suppose que toutes les inverses des matrices considérées existent, alors :*

$$(A + uv^*)^{-1} = A^{-1} - \frac{A^{-1}uv^*A^{-1}}{1 + v^*A^{-1}u} .$$

Proposition 0.5 (Inégalité de Poincaré)**.** *Une mesure de probabilité $\mathbb{P}$ sur $\mathbb{R}^n$ satisfait l'inégalité de Poincaré avec une constante $c > 0$ si, pour toute fonction continue différentielle $f : \mathbb{R}^n \to \mathbb{C}$,*

$$\mathrm{Var}_{\mathbb{P}}(f) = \mathbb{E}_{\mathbb{P}}(|f(x) - \mathbb{E}_{\mathbb{P}}(f(x))|^2) \leqslant \frac{1}{c}\mathbb{E}_{\mathbb{P}}|\nabla f(x)|^2 .$$

Matrice de Wigner

Définition 0.9. Soit W_n une matrice Hermitienne $n \times n$, $W_n = W_n^*$ tel que $W_n := (W_{k\ell}, 1 \leqslant k \leqslant \ell \leqslant n)$ sont des variables aléatoires i.i.d. avec $\mathbb{E}(W_{k\ell}) = 0$, $\forall\, 1 \leqslant k \leqslant \ell$ et $\mathbb{E}(|W_{k\ell}|^2) < \infty$, $\forall\, 1 \leqslant k \leqslant \ell$. $W_n/\sqrt{n}$ est appelée matrice de Wigner.

Théorème 0.6 (Universalité du théorème de Wigner et de la loi semi-circulaire)**.** *Soit W_n une matrice de Wigner définie par $W_n := (W_{k\ell}, 1 \leqslant k \leqslant \ell)$ variables aléatoires i.i.d. telles que*

1. $\mathbb{E}(W_{k\ell}) = 0$, $\forall\, 1 \leqslant k \leqslant \ell$,
2. $\mathbb{E}(|W_{k\ell}|^2) = \sigma^2 < \infty$, $\forall\, 1 \leqslant k \leqslant \ell$ *et* $\sigma > 0$.

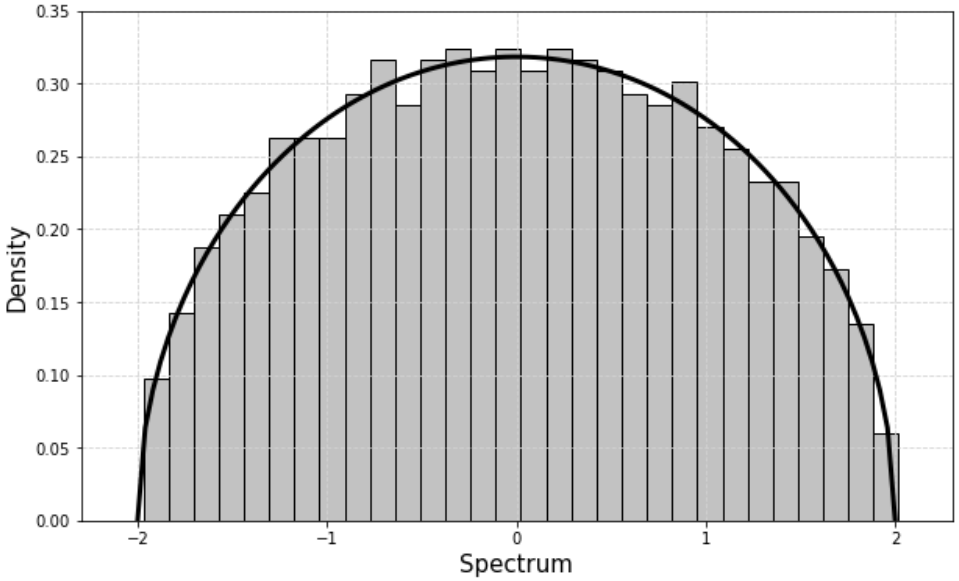


Figure 2: Spectre (histogramme) de la matrice aléatoire de Wigner $W_n/\sqrt{n}$ ($n = 1000$, $\sigma = 1$). La ligne continue représente la loi semi-circulaire.

Alors presque surement, la mesure spectrale empirique de $W_n/\sqrt{n}$ *converge faiblement vers la loi semi-circulaire :*

$$(p.s.) \quad \mu_{\frac{W_n}{\sqrt{n}}} := \frac{1}{n}\sum_{k=1}^{n} \delta_{\lambda_k\left(\frac{W_n}{\sqrt{n}}\right)} \xrightarrow[n\to\infty]{\mathcal{D}} \mu_{sc}\,,$$

où μ_{sc} *est définie par*

$$d\mu_{sc}(t) = \frac{1}{2\pi\sigma^2}\sqrt{(4\sigma^2 - t^2)}\mathbf{1}_{[-2\sigma,2\sigma]}(t)dt\,.$$

Les valeurs propres de la matrice $W_n/\sqrt{n}$ sont réelles. Dans la Figure 2, un histogramme des valeurs propres d'une matrice aléatoire de Wigner est illustré par rapport à la distribution théorique donnée par le théorème 0.6.

Comportement local du spectre pour les matrices de Wigner

On note

$$\lambda_{\max}(W_n) = \max_{k\in[n]} \lambda_k(W_n) \text{ et } \lambda_{\min}(W_n) = \min_{k\in[n]} \lambda_k(W_n)\,.$$

Pour traiter certaines questions, il est nécessaire d'avoir des informations exactes sur la position de la plus grande valeur propre de la matrice. Dans le cas des matrices de Wigner, de nombreux travaux ont été réalisés et affinés dans les années 80' [FK81, BY88].

Théorème 0.7 (Convergence des valeurs propres extrêmes)**.** *Si* $\mathbb{E}(|W_{k\ell}|^4) < \infty$, $\forall\, 1 \leqslant k \leqslant \ell$, *alors*

$$\lambda_{\max}(W_n/\sqrt{n}) \xrightarrow[n\to\infty]{p.s.} 2\sigma \;,\; \lambda_{\min}(W_n/\sqrt{n}) \xrightarrow[n\to\infty]{p.s.} -2\sigma\,.$$

En particulier,

$$\left\|W_n/\sqrt{n}\right\| = \max(|\lambda_{\max}(W_n/\sqrt{n})|, |\lambda_{\min}(W_n/\sqrt{n})|) \xrightarrow[n\to\infty]{p.s.} 2\sigma\,.$$

Si $\mathbb{E}(|W_{k\ell}|^4) = \infty$, $\forall\, 1 \leqslant k \leqslant \ell$, *alors*

$$\lambda_{\max}(W_n/\sqrt{n}) \xrightarrow[n\to\infty]{p.s.} +\infty\,.$$

Matrice de Wigner déformée

La dernière propriété spécifique sur les matrices de Wigner à traiter dans le cadre de cette thèse est la distribution des valeurs propres lorsque la matrice de Wigner est perturbée par une déformation de rang fini. Ce type de modèle est fréquemment appelé "spike". Selon le type de déformation, certaines valeurs propres aberrantes peuvent s'échapper du support de la distribution. Soit W_n une matrice de Wigner aléatoire et

1. $\mathbb{E}(W_{k\ell}) = 0$, $\forall\, 1 \leqslant k \leqslant \ell$,
2. $\mathbb{E}(|W_{k\ell}|^2) = \sigma^2 < \infty$, $\forall\, 1 \leqslant k \leqslant \ell$ et $\sigma > 0$,
3. $\sup_{k\neq\ell} \mathbb{E}[|W_{k\ell}|^4] < \infty$.

Soit P_n une matrice symétrique réelle déterministe de rang fixe r. Nous sommes intéressés par les propriétés du spectre de la matrice $\frac{1}{\sqrt{n}}W_n + P_n$.

Depuis l'article pionnier de Füredi et Komlòs [FK81], de nombreux scientifiques ont étudié les propriétés spectrales des matrices de Wigner déformées [Pé06, CDMF09, PRS13, RS13]. On note $\theta_1, ..., \theta_r$ les valeurs propres ordonnées de P_n où $\forall j \in [1, r]$, θ_j est de multiplicité k_j et indépendant de n. Soit r_0, l'index associé au seuil 0 c.a.d. $\theta_{r_0} = 0$ et P_n a $r_0 - 1$ valeurs propres distinctes. Soit $r_{+\sigma}$ (resp $r_{-\sigma}$) est le nombre de j tel que $\theta_j > \sigma$ (resp $\theta_j < -\sigma$).

Théorème 0.8 (Théorème de Wigner deformé - [CDMF09, PRS13])**.** *Soit W_n est une matrice de Wigner réelle aléatoire satisfaisant la condition (1)-(3) et P_n est une matrice Hermitienne réelle déterministe de rang r fini fixe comme ci-dessus. Soit*

$$\rho_{\theta_j} = \theta_j + \frac{\sigma^2}{\theta_j}\,.$$

Alors, les conditions suivantes sont vraies :

1. *Pour* $1 \leqslant j \leqslant r_{+\sigma}, 1 \leqslant i \leqslant k_j, \lambda_{k_1+...+k_{j-1}+i} \to \rho_{\theta_j}$,
2. $\lambda_{k_1+...+k_{r_{+\sigma}}+1} \to 2\sigma$,
3. $\lambda_{k_1+...+k_{r-r_{-\sigma}}} \to -2\sigma$,
4. *Pour* $j \geqslant r - r_{-\sigma} + 1, 1 \leqslant i \leqslant k_j, \lambda_{k_1+...+k_{j-1}+i} \to \rho_{\theta_j}$.

La convergence (1)-(4) est en probabilité.

Remarque 0.4. Dans [CDMF09], Capitaine, Donati-Martin et Féral montrent que si la distribution des entrées satisfait l'inégalité de Poincaré (proposition 0.5), la convergence du théorème 0.8 est vraie presque sûrement.

Dans la Figure 3, un histogramme des valeurs propres de la matrice de Wigner déformée est illustré par rapport à la distribution théorique donnée par le théorème 0.8.

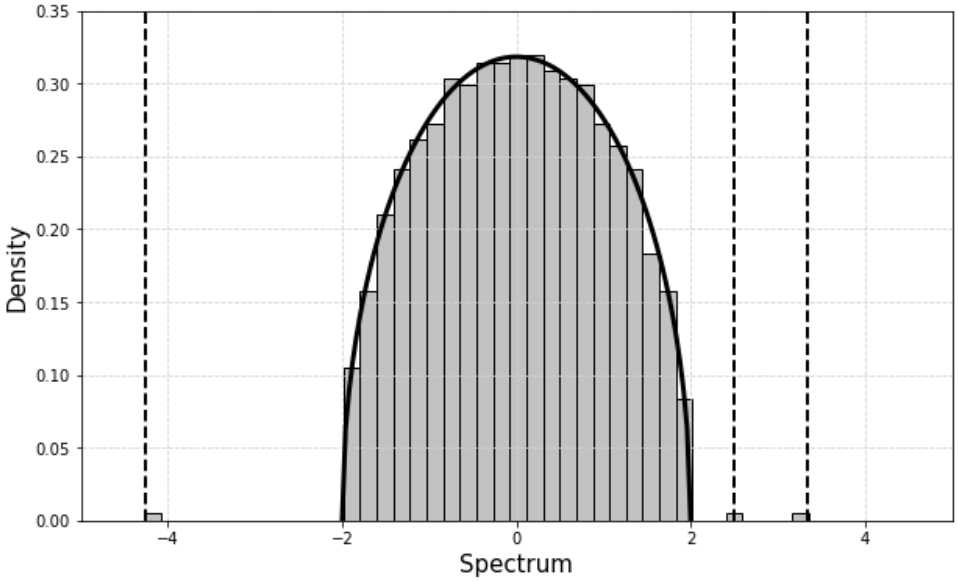


Figure 3: Spectre (histogramme) de la matrice aléatoire de Wigner deformée $W_n/\sqrt{n}+P_n$ ($n = 1000$, $\sigma = 1$). La matrice de perturbation est $P_n = \mathrm{diag}(-4, 2, 3, 0, ..., 0)$. La ligne continue représente la loi semi-circulaire. Les lignes pointillées indiquent la valeur théorique des valeurs aberrantes à $-4-1/4, 2+1/2, 3+1/3$ comme prédit par le théorème 0.8.

La loi circulaire

Dans la deuxième partie de cette introduction sur les matrices aléatoires, nous nous concentrons sur les matrices non-Hermitiennes. Soit $Y_n \in \mathcal{M}_n(\mathbb{C})$ une matrice aléatoire carrée de dimension $n \times n$ dont les entrées sont i.i.d. centrées de variance σ^2. Les valeurs propres de Y_n ne sont plus réelles mais complexes. Le résultat principal concerne la convergence de la mesure spectrale empirique de $Y_n/\sqrt{n}$ vers la loi circulaire dans le plan complexe. Initiallement prouvé par Mehta [Meh67] pour la distribution spectrale empirique moyenne dans le cas d'une gaussienne complexe suite aux travaux de Ginibre sur la formule explicite du spectre [Gin65], Edelman [Ede97] a établi la loi circulaire dans le cas de variables aléatoires gaussiennes réelles. Silverstein a donné un argument pour passer de la convergence moyenne à la convergence presque sûre. Girko a travaillé sur la version universelle (pour d'autres types de distribution) [Gir85] en fournissant quelques éléments de preuves tels que la technique d'hermétisation. Cependant, c'est finalement Tao et Vu [TVK10] qui ont prouvé le cas général. Je conseille au lecteur de consulter Bordenave et Chafaï [BC12].

Théorème 0.9. *Soit Y_n une matrice aléatoire $\mathcal{M}_n(\mathbb{C})$ telle que $Y_n := (Y_{k\ell}, 1 \leqslant k, \ell \leqslant n)$ sont des variables aléatoires i.i.d. telles que $\mathbb{E}(Y_{k\ell}) = 0$, $\forall\, 1 \leqslant k, \ell \leqslant n$ et $\mathbb{E}(|Y_{k\ell}|^2) = \sigma^2$, $\forall\, 1 \leqslant k, \ell \leqslant n$. Alors presque sûrement, la mesure spectrale empirique de $Y_n/\sqrt{n}$ converge faiblement vers la loi circulaire*

$$\mu_{\frac{Y_n}{\sqrt{n}}} \xrightarrow[n\to\infty]{\mathcal{D}} \mu_c\,,$$

où μ_c est la loi circulaire c.a.d. la loi uniforme sur le disque de rayon σ de $\mathbb{C}$ avec comme densité

$$d\mu_c(z) = \frac{1}{\pi\sigma^2}\mathbf{1}_{z\in\mathbb{C},|z|\leqslant\sigma}dz\,.$$

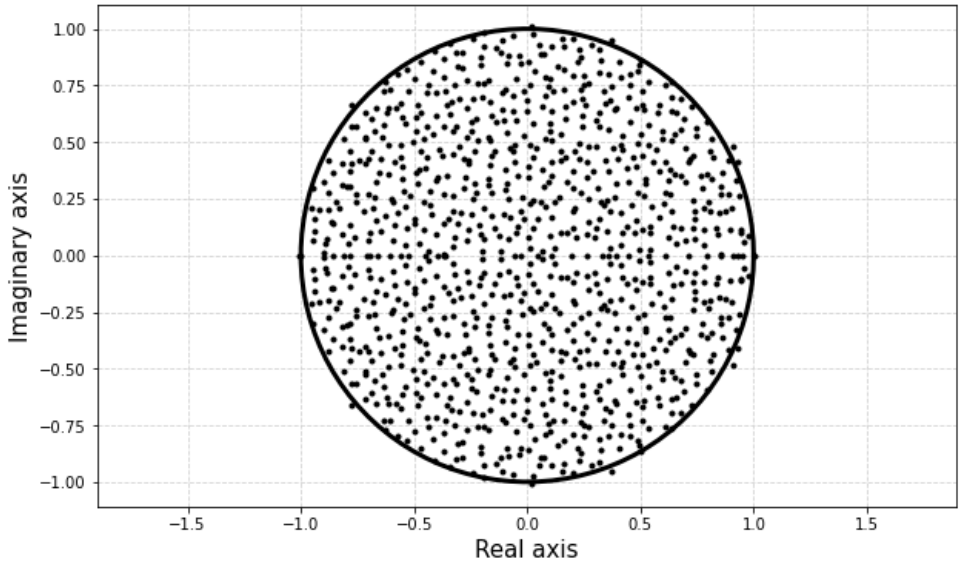


Figure 4: Spectre de la matrice aléatoire non-Hermitienne $Y_n/\sqrt{n}$ dans le plan complexe ($n = 1000$, $\sigma = 1$). Le cercle de ligne continue représente la frontière de la loi circulaire.

Dans la Figure 4, les valeurs propres d'une matrice aléatoire non-Hermitienne dans le plan complexe sont représentées par rapport à la distribution théorique donnée par le théorème 0.9.

Comportement local du spectre pour les matrices non-Hermitiennes

Dans le cas de la loi du cercle, il est important d'avoir des informations sur la position du rayon spectral. Nous avons vu précédemment dans les travaux de May [May72], la nécessité de décrire le seuil de perte de stabilité par la plus grande valeur propre de la partie réelle de la matrice Jacobienne. De nombreux travaux ont été réalisés sur ce sujet, en particulier par Bai [BSY88, BS10].

Théorème 0.10 (Convergence des valeurs propres extrêmes)**.** *Si* $\mathbb{E}(Y_{k\ell}) = 0$ *et* $\mathbb{E}(|Y_{k\ell}|^4) < \infty$, $\forall\, 1 \leqslant k, \ell \leqslant n$, *alors*

$$\left\| \frac{Y_n}{\sqrt{n}} \right\| \xrightarrow[n\to\infty]{p.s.} 2\sigma \quad \textit{et} \quad \rho\left(\frac{Y_n}{\sqrt{n}} \right) \xrightarrow[n\to\infty]{p.s.} \sigma\,.$$

Valeurs aberrantes dans le spectre des matrices non-Hermitiennes

Comme dans le cas de Wigner, nous pouvons considérer une perturbation de rang fini de la matrice non-Hermitienne. Ce problème a été considéré par Tao [Tao13, Théorème 1.7] et a été étendu par Benaych-Georges et Rochet [BGR16] pour l'étude des fluctuations des valeurs propres aberrantes.

Théorème 0.11 (Spectre déformé d'une matrice aléatoire non-Hermitienne)**.** *Soit* Y_n *une matrice aléatoire i.i.d. avec* $\mathbb{E}(Y_{k\ell}) = 0$, $\mathbb{E}(|Y_{k\ell}|^2) = 1$ *et* $\mathbb{E}(|Y_{k\ell}|^4) < \infty$, $\forall\, 1 \leqslant k, \ell \leqslant n$ *et pour chaque* n, *soit* P_n *une matrice déterministe de rang* $O(1)$ *et de norme opérateur* $O(1)$. *Soit* $\varepsilon > 0$, *et supposons que pour tout* n *suffisamment grand, il n'y a pas de valeurs*

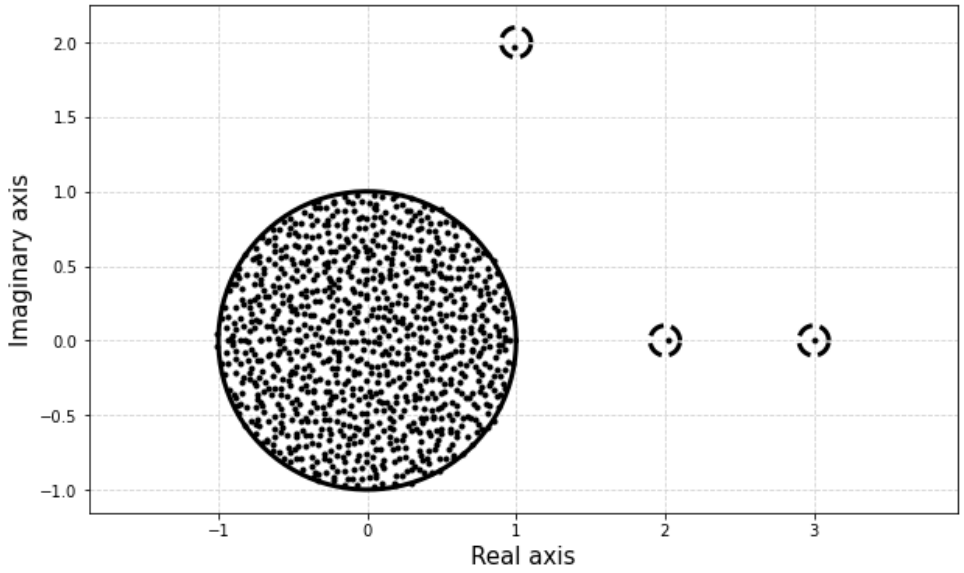


Figure 5: Spectre de la matrice non-Hermitienne $Y_n/\sqrt{n} + P_n$ dans le plan complexe ($n = 1000$, $\sigma = 1$). La matrice de perturbation est $P_n = \mathrm{diag}(1 + 2i, 2, 3, 0, ..., 0)$. Le cercle en trait plein représente la limite de la loi circulaire. Il y a trois valeurs propres dans les petits cercles en pointillés centrés sur $1 + 2i$, 2, 3 comme prédit par le théorème 0.11.

propres de P_n dans l'anneau $\{z \in \mathbb{C} : 1 + \varepsilon < |z| < 1 + 3\varepsilon\}$, et il y a j valeurs propres $\lambda_1(P_n), ..., \lambda_j(P_n)$ pour $j = O(1)$ dans la région $\{z \in \mathbb{C} : |z| \geqslant 1 + 3\varepsilon\}$.

Alors, p.s., pour n suffisamment grand, il y a précisément j valeurs propres

$$\lambda_1\left(\frac{Y_n}{\sqrt{n}} + P_n\right), ..., \lambda_j\left(\frac{Y_n}{\sqrt{n}} + P_n\right),$$

de $\frac{Y_n}{\sqrt{n}} + P_n$ dans la région $\{z \in \mathbb{C} : |z| \geqslant 1 + 2\varepsilon\}$, et après avoir étiqueté correctement ces valeurs propres, $\lambda_i(\frac{Y_n}{\sqrt{n}} + P_n) = \lambda_i(P_n) + o(1)$ lorsque $n \to \infty$ pour chaque $1 \leqslant i \leqslant j$.

Dans la Figure 5, les valeurs propres d'une matrice aléatoire non Hermitienne déformée avec des valeurs aberrantes sont représentées par rapport à la distribution théorique donnée par le théorème 0.11.

Le modèle elliptique

Dans la configuration de la matrice de Wigner, on considère que l'interaction d'une espèce sur l'autre est la même. Pour les matrices non-Hermitiennes, toutes les interactions sont indépendantes. Cependant, en écologie, les effets réciproques d'une espèce k sur une autre espèce ℓ ($X_{k\ell} \leftrightarrow X_{\ell k}$) sont liés. Mathématiquement, on considère une corrélation par paire entre les entrées de la matrice. Ceci peut être utilisé pour décrire des processus biologiques tels que la prédation lorsque le signe des interactions est inversé et que la corrélation est négative. Dans une matrice aléatoire, lorsque les interactions par paire sont tirées d'une distribution bi-variée, nous sommes dans le cadre du modèle elliptique. Introduit à l'origine par Girko [Gir86], ce modèle a depuis été largement étudié [Gir95, Nau12, NO15, OR14].

Définition 0.10 (Modèle elliptique aléatoire)**.** Soit X_n une matrice aléatoire réelle qui satisfait aux trois conditions suivantes :

1. Les paires $(X_{k\ell}, X_{\ell k})$, $k \neq \ell$ sont des variables aléatoires i.i.d. avec

$$\forall\, k \neq \ell,\ \mathbb{E}(X_{k\ell}) = 0,\ \mathbb{E}(|X_{k\ell}|^2) = 1 \text{ et } \mathbb{E}(|X_{k\ell}|^4) < \infty\,.$$

2. Pour $k < \ell$ le vecteur $(X_{k\ell}, X_{\ell k})$ est tiré d'une distribution bivariée, indépendemment des variables aléatoires restantes, avec une covariance $\mathbb{E}(X_{k\ell}X_{\ell k}) = \rho$ avec $|\rho| \leqslant 1$.

3. $(X_{kk}, 1 \leqslant k \leqslant n)$ sont des variables aléatoires i.i.d., indépendantes des entrées hors-diagonales avec $\mathbb{E}(X_{kk}) = 0$ and $\mathbb{E}(|X_{kk}|^2) = 1$.

Pour $\rho \in (-1, 1)$, on définit l'ellipsoïde

$$\mathcal{E}_\rho := \left\{ z = x + \boldsymbol{i}y \in \mathbb{C} : \frac{x^2}{(1+\rho)^2} + \frac{y^2}{(1-\rho)^2} \leqslant 1 \right\}\,.$$

Remarque 0.5.

1. Pour $\rho = 1$, $\mathcal{E}_1$ est l'intervalle $[-2, 2]$ sur l'axe réel et pour $\rho = -1$, $\mathcal{E}_{-1}$ est l'intervalle $[-2, 2]$ sur l'axe imaginaire.

2. Si $\rho = 1$, X_n est une matrice de Wigner.

3. Si $\rho = 0$, X_n est une matrice non-Hermitienne c.a.d. définie par le théorème 0.9.

Théorème 0.12 (Loi elliptique)**.** *Soit X_n une variable aléatoire elliptique satisfaisant les conditions de la définition 0.10. Alors presque sûrement, la mesure spectrale empirique de $X_n/\sqrt{n}$ converge faiblement vers la loi elliptique :*

$$(p.s.) \quad \mu_{\frac{X_n}{\sqrt{n}}} \xrightarrow[n\to\infty]{\mathcal{D}} \mu_\rho\,,$$

où μ_ρ est la mesure de probabilité uniforme sur l'ellipsoïde $\mathcal{E}_\rho$ de densité

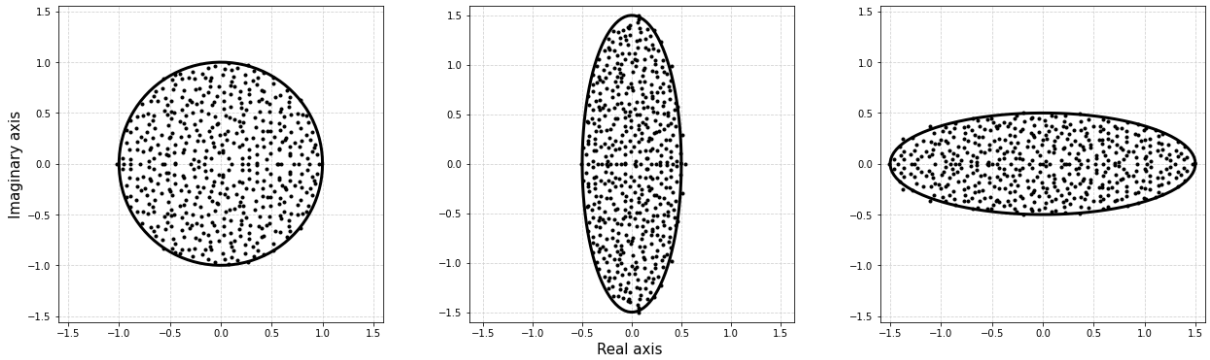
$$\mu_\rho(z) = \begin{cases} \frac{1}{\pi(1-\rho^2)} & \text{si } z \in \mathcal{E}_\rho\,, \\ 0 & \text{sinon}\ . \end{cases}$$

Dans la Figure 6, les valeurs propres d'une matrice aléatoire elliptique dans le plan complexe sont représentées par rapport à la distribution théorique donnée par le théorème 0.12.

Le corollaire 2.3 dans O'Rourke et Renfrew [OR14] fournit des informations sur le rayon spectral d'une matrice elliptique.

Proposition 0.13 (Rayon spectral d'une matrice aléatoire elliptique)**.** *Soit X_n une matrice aléatoire elliptique définie dans la définition 0.10, alors*

$$\rho\left(\frac{X_n}{\sqrt{n}}\right) \xrightarrow[n\to\infty]{p.s.} 1 + |\rho|\,.$$



(a) $\rho = 0$ (b) $\rho = -0.5$ (c) $\rho = 0.5$

Figure 6: Spectre de la matrice elliptique X_n ($n = 500$) avec des paramètres distincts $\rho \in \{-0.5,\ 0,\ 0.5\}$. La ligne continue représente une ellipse $\{z = x+\boldsymbol{i}y \in \mathbb{C},\ \frac{x^2}{(1+\rho)^2}+\frac{y^2}{(1-\rho)^2} = 1\}$ qui est la limite du support de la distribution spectrale limite pour un modèle elliptique.

Valeurs aberrantes dans le modèle elliptique

L'étude des valeurs aberrantes dans le cas d'une matrice elliptique déformée a été réalisée par O'Rourke et Renfrew [OR14].

On définit le voisinage pour tout $\delta > 0$

$$\mathcal{E}_{\rho,\delta} := \{z \in \mathbb{C} : \mathrm{dist}(z, \mathcal{E}_\rho) \leqslant \delta\}\ .$$

Théorème 0.14 (Matrice aléatoire elliptique déformée)**.** *Soit $k \geqslant 1$ et $\delta > 0$. Soit X_n une matrice aléatoire elliptique spécifiée par la définition 0.10. Soit P_n une matrice déterministe de taille $n \times n$ de rang fini k et $\sup_n \|P_n\| = O(1)$. On suppose que pour n suffisamment grand, il n'y a pas de valeurs propres non nulles de P_n qui satisfont*

$$\lambda_i(P_n) + \frac{\rho}{\lambda_i(P_n)} \in \mathcal{E}_{\rho,3\delta} \backslash \mathcal{E}_{\rho,\delta} \text{ avec } |\lambda_i(P_n)| > 1,$$

et il y a j valeurs propres $\lambda_1(P_n), ..., \lambda_j(P_n)$ pour $j \leqslant k$ qui satisfont

$$\lambda_i(P_n) + \frac{\rho}{\lambda_i(P_n)} \in \mathbb{C} \backslash \mathcal{E}_{\rho,3\delta} \text{ avec } |\lambda_i(P_n)| > 1\,.$$

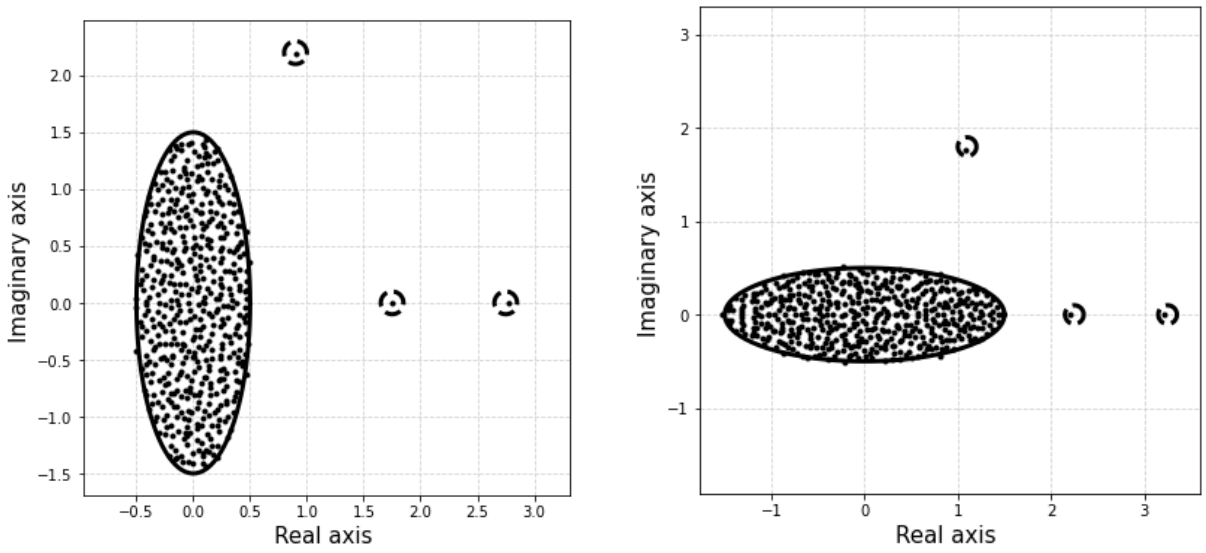
Alors, p.s., pour n suffisamment grand, il y a exactement j valeurs propres de $\frac{1}{\sqrt{n}}X_n + P_n$ dans la région $\mathbb{C}\backslash\mathcal{E}_{\rho,2\delta}$ et après avoir étiqueté correctement les valeurs propres,

$$\lambda_i\left(\frac{X_n}{\sqrt{n}} + P_n\right) = \lambda_i(P_n) + \frac{\rho}{\lambda_i(P_n)} + o(1), \forall\, 1 \leqslant i \leqslant j\,.$$

Dans la Figure 7, les valeurs propres d'une matrice aléatoire elliptique déformée dans le plan complexe sont représentées par rapport à la distribution théorique donnée par le théorème 0.14.

Cadre théorique de la thèse

La compréhension des points d'équilibre du système de Lotka-Volterra (6) et leur stabilité permet de mieux comprendre l'impact du réseau trophique, représenté par la matrice



(a) $\rho = -0.5$ (b) $\rho = 0.5$

Figure 7: Spectre de la matrice elliptique deformée $X_n + P_n$ ($n = 500$) avec des paramètres distincts $\rho \in \{-0.5,\ 0.5\}$. La matrice de perturbation est $P_n = \mathrm{diag}(1 + 2i, 2, 3, 0, ..., 0)$. La ligne continue représente une ellipse $\{z = x + \boldsymbol{i}y \in \mathbb{C},\ \frac{x^2}{(1+\rho)^2} + \frac{y^2}{(1-\rho)^2} = 1\}$ qui est la limite du support de la distribution spectrale limite pour un modèle elliptique. Il y a trois valeurs propres dans les petits cercles en pointillé centrés sur $1 + \frac{\rho}{5} + (2 - \frac{2\rho}{5})i$, $2 + \frac{\rho}{2}$, $3 + \frac{\rho}{3}$ comme prédit par le théorème 0.14

d'interaction B, sur l'abondance des espèces. En particulier, le réseau trophique a un impact sur la persistance des espèces qui le composent (:= nombre d'espèces persistantes), la faisabilité du système (c'est-à-dire s'il existe un équilibre avec toutes les espèces à des abondances non nulles) et la stabilité de l'équilibre. On rappelle le système d'équations (6),

$$\frac{dx_k}{dt} = x_k\ (1 - x_k + (B\boldsymbol{x})_k)\ \ , \quad k \in [n]\,.$$

Une caractéristique essentielle pour comprendre la dynamique du système LV (6) est l'existence d'un équilibre $\boldsymbol{x}^* = (x_k^*)_{k\in[n]}$ tel que :

$$\begin{cases} x_k^*\,(1 - x_k^* + (B\boldsymbol{x}^*)_k) = 0\,, \quad \forall k \in [n]\,, \\ x_k^* \geqslant 0. \end{cases} \tag{9}$$

Une question naturelle est de savoir si un équilibre existe et s'il est unique. Si c'est le cas, une autre considération est de savoir si le système converge vers cet équilibre, c'est-à-dire la convergence d'une solution $\boldsymbol{x}$ vers l'équilibre $\boldsymbol{x}^*$: $\boldsymbol{x}(t) \xrightarrow[t\to\infty]{} \boldsymbol{x}^*$ si $\boldsymbol{x}(0)$ est suffisamment proche de $\boldsymbol{x}^*$. La dernière étape consiste à décrire la stabilité : locale, globale, résilience (:= capacité d'un système à retrouver sa structure initiale suite à une perturbation), etc. Le système de Lotka-Volterra est invariant c'est-à-dire que $\boldsymbol{x}(0) > 0$ (composante par composante) implique $\boldsymbol{x}(t) > 0$ pour tout $t > 0$. Cependant, certaines de ces composantes $x_k(t)$ peuvent converger vers zéro si l'équilibre $\boldsymbol{x}^*$ a des composantes nulles.

Faisabilité

La question de la faisabilité d'un équilibre $\boldsymbol{x}^* > 0$ avait déjà été abordée par Goh [GJ77] pour le modèle Lotka-Volterra dans les années 70', puis Logofet s'est intéressé à ce problème dans le cas d'un système compétitif [Log93]. Rossberg a également étudié le nombre moyen d'espèces pouvant coexister dans des communautés compétitives [Ros13]. Récemment, Grilli *et al.* [GAS+17] ont étudié l'impact des propriétés du réseau trophique sur le taux de croissance pour maintenir un équilibre faisable en utilisant les méthodes de stabilité structurelle introduites par Rohr *et al.* [RSB14].

D'après (7), nous considérons une matrice d'interaction avec des entrées gaussiennes

$$B_n = \frac{1}{\alpha_n \sqrt{n}} A_n,$$

où $A_n := (A_{k\ell}, 1 \leqslant k, \ell \leqslant n)$ sont des variables aléatoires i.i.d. avec $A_{k\ell} \sim \mathcal{N}(0,1)$.

Basé sur les travaux de Geman et Hwang [GH82], Dougoud *et al.*[DVR+18] ont montré que dans le cadre d'entrées aléatoires de la matrice d'interaction, si $\alpha_n > 0$ est fixé et indépendant de n alors nécessairement certaines espèces s'éteignent. Le seuil d'existence d'un équilibre faisable du modèle (6) a été étudié par Bizeul et Najim [BN21]. Dans leur article, ils montrent également que la faisabilité implique la stabilité. Ce type de résultat avait déjà été observé par Stone [Sto16] qui a montré que le seuil de stabilité est franchi avant le seuil de faisabilité.

En partant de l'équation (9), si $\boldsymbol{x}^* > 0$, l'ensemble des équations d'équilibre devient une équation linéaire :

$$\boldsymbol{x}^* = \mathbf{1} + B_n \boldsymbol{x}^* \,. \tag{10}$$

On se limitera ici au cas non trivial dans lequel $\alpha_n \to \infty$ et on définit le seuil de faisabilité par $\alpha_n^* = \sqrt{2\log(n)}$.

Théorème 0.15 (Theorème 1.1 [BN21])**.** *Soit $\alpha_n \xrightarrow[n\to\infty]{} \infty$. Soit $\boldsymbol{x}^* = (x_k^*)_{k\in[n]}$ la solution de* (10) *et $A_n := (A_{k\ell}, 1 \leqslant k, \ell \leqslant n)$ sont des variables aléatoires i.i.d. avec $A_{k\ell} \sim \mathcal{N}(0,1)$.*

1. *Si il existe $\varepsilon > 0$ de telle sorte que $\alpha_n \leqslant (1-\varepsilon)\alpha_n^*$ alors*

$$\mathbb{P}\left\{\min_{k\in[n]} x_k^* > 0\right\} \xrightarrow[n\to\infty]{} 0\,.$$

2. *Si il existe $\varepsilon > 0$ de telle sorte que $\alpha_n \geqslant (1+\varepsilon)\alpha_n^*$ alors*

$$\mathbb{P}\left\{\min_{k\in[n]} x_k^* > 0\right\} \xrightarrow[n\to\infty]{} 1\,.$$

Remarque 0.6 (Esquisse de la preuve). Soit $Q_n = \left(I - \frac{A_n}{\alpha\sqrt{n}}\right)^{-1}$, la résolvante de la matrice A_n. Le problème est défini par :

$$\boldsymbol{x}^* = \mathbf{1} + \left(\frac{A_n \boldsymbol{x}^*}{\alpha_n \sqrt{n}}\right) \quad \Leftrightarrow \quad \boldsymbol{x}^* = \left(I - \frac{A_n}{\alpha_n \sqrt{n}}\right)^{-1} \mathbf{1}\,.$$

Pour chaque entrée du vecteur $\boldsymbol{x}^*$

$$\begin{aligned}\forall k \in [n],\ x_k^* &= \sum_{i=0}^{\infty} e_k^* \left(\frac{A_n}{\alpha_n\sqrt{n}}\right)^i \mathbf{1}\,,\\ &= 1 + \frac{1}{\alpha_n} e_k^* \frac{A_n}{\sqrt{n}}\mathbf{1} + \frac{1}{\alpha_n^2} e_k^* \left(\frac{A_n}{\alpha_n}\right)^2 Q_n \mathbf{1}\,,\\ &= 1 + \frac{Z_k}{\alpha_n} + \frac{R_k}{\alpha_n^2}\,,\\ &\approx 1 + \frac{Z_k}{\alpha_n},\ Z_k \sim \mathcal{N}(0,1)\,.\end{aligned}$$

On peut montrer que le terme R_k est négligeable en utilisant les propriétés de concentration gaussienne (pour plus de détails, voir [BN21]). À partir des propriétés des valeurs extrêmes d'une famille de variables aléatoires gaussiennes, on peut déduire

$$\min_{k\in[n]} x_k^* > 0 \quad \Leftrightarrow \quad \min_{k\in[n]} Z_k \sim -\sqrt{2\log(n)} > -\alpha_n\,.$$

Les résultats d'existence et d'unicité d'un équilibre faisable ont été étendus dans le cas d'un réseau trophique creux par Akjouj et Najim [AN21]. Ils supposent que chaque espèce est en relation avec d autres espèces. La magnitude de d par rapport à n reflète la densité de connexions du modèle. Deux cas distincts sont étudiés, d'une part d est proportionnel à n. D'autre part, $d \geqslant \log(n)$ dans le cas d'une structure en blocs particulière. De plus, ils démontrent également la stabilité globale de l'équilibre.

Condition de non-invasion

Le problème (9) devient beaucoup plus complexe lorsque l'on considère un équilibre dans lequel $\boldsymbol{x}^*$ a des composantes nulles. Le système d'équations n'est plus linéaire et l'équation devient un problème d'optimisation non linéaire. Une solution naïve et immédiate pour résoudre ce problème est de choisir un sous-ensemble $\mathcal{I} \subset [n]$, définir les composantes correspondantes $x_{\mathcal{I}} = (x_i^*)_{i\in\mathcal{I}}$ à zéro, et de résoudre le système linéaire restant :

$$\boldsymbol{x}_{\mathcal{I}^c} = \mathbf{1}_{|\mathcal{I}^c|} + B_{\mathcal{I}^c}\boldsymbol{x}_{\mathcal{I}^c}\ .$$

Si il existe $\boldsymbol{x}_{\mathcal{I}^c} \geqslant 0$ qui résout l'équation précédente, alors $\boldsymbol{x} = \begin{pmatrix}\boldsymbol{x}_{\mathcal{I}}\\ \boldsymbol{x}_{\mathcal{I}^c}\end{pmatrix}$ satisfait (9) et est un potentiel équilibre. Le nombre de sous-cas $\mathcal{I} \subset [n]$ est 2^n et, en particulier, croît de manière exponentielle lorsque $n \to \infty$.

Les équations d'équilibre deviennent mal posées car il peut y avoir plusieurs équilibres. Une condition connue en écologie pour les systèmes dynamiques est la condition de non-invasion [LM96, JS98] associé à l'équilibre saturé. Un équilibre est saturé s'il est résistant à l'invasion d'une espèce absente. L'étude des équilibres saturés et de la permanence est un sujet de recherche important dans le domaine des systèmes dynamiques (pour plus de détails, voir le livre fondateur de Hofbauer et Sigmund [HS98]).

Définition 0.11 (Equilibre saturé)**.** Soit $\mathcal{I}^c$ l'ensemble des espèces persistantes,

- $\boldsymbol{x}$ est saturé $\Leftrightarrow \forall k \in \mathcal{I} : 1 - x_k^* + (B\boldsymbol{x}^*)_k \leqslant 0$,

- $\boldsymbol{x}$ est strictement saturé $\Leftrightarrow \forall k \in \mathcal{I} : 1 - x_k^* + (B\boldsymbol{x}^*)_k < 0$.

Lemme 0.16.

1. *Si il y a une solution strictement positive* $\boldsymbol{x}(t) > 0$, *tel que* $\boldsymbol{x}(t) \xrightarrow[t\to\infty]{} \boldsymbol{x}^*$, *alors* $\boldsymbol{x}^*$ *est un équilibre saturé.*

2. *Si* $\boldsymbol{x}^*$ *est strictement saturé, alors il existe une solution strictement positive* $\boldsymbol{x}(t) > 0$, *tel que* $\boldsymbol{x}(t) \xrightarrow[t\to\infty]{} \boldsymbol{x}^*$.

On remarque qu'en s'appuyant sur les propriétés standard des systèmes dynamiques, voir par exemple [Tak96, Theorem 3.2.5], une condition nécessaire pour que l'équilibre $\boldsymbol{x}^*$ soit stable est que

$$1 - x_k^* + (B\boldsymbol{x}^*)_k \leqslant 0\,. \tag{11}$$

La condition (11) diminue le nombre de solutions potentielles au système (9). En référence à l'EDO (6), l'exigence pour une espèce donnée $k \in [n]$ pour être non invasive est équivalent à:

$$\left(\frac{1}{x_k}\frac{dx_k}{dt}\right)_{x_k\to 0^+} \leqslant 0\,. \tag{12}$$

Par conséquent, on se concentrera maintenant sur l'ensemble des conditions suivantes :

$$\begin{cases} x_k^*\,(1 - x_k^* + (B\boldsymbol{x}^*)_k) & = & 0 & \text{pour } k \in [n]\,, \\ 1 - x_k^* + (B\boldsymbol{x}^*)_k & \leqslant & 0 & \text{for } k \in [n]\,, \\ \boldsymbol{x}^* & \geqslant & 0 & \text{par composante}\,. \end{cases} \tag{13}$$

Le problème de la recherche d'un équilibre positif entre ainsi dans la classe des problèmes de complémentarité linéaire (LCP), que nous décrivons ci-après.

Problème de complémentarité linéaire

Le LCP est une classe de problèmes issue de l'optimisation mathématique qui englobe notamment les problèmes de programmation linéaire et quadratique ; les références standard sont les suivantes [Mur88, CPS09]. Soit une matrice M de taille $n \times n$ et un vecteur $\boldsymbol{q}$ de taille $n \times 1$, le LCP associé désigné par $LCP(M, \boldsymbol{q})$ consiste à trouver deux vecteurs $\boldsymbol{z}, \boldsymbol{w}$ de taille $n \times 1$ satisfaisant l'ensemble des contraintes suivantes :

$$\begin{cases} \boldsymbol{z} & \geqslant & 0\,, & & & \\ \boldsymbol{w} = M\boldsymbol{z} + \boldsymbol{q} & \geqslant & 0\,, & & & \\ \boldsymbol{w}^\top \boldsymbol{z} & = & 0 & \Leftrightarrow & w_k z_k = 0 & \text{pour tout } k \in [n]\,. \end{cases} \tag{14}$$

Comme $\boldsymbol{w}$ peut être inféré de $\boldsymbol{z}$, on note $\boldsymbol{z} \in LCP(M, \boldsymbol{q})$ si $(\boldsymbol{w}, \boldsymbol{z})$ est une solution de (14).

L'étude du LCP remonte aux travaux de Lemke [Lem65] et Cottle et Danzig [CD68]. Lemke et Howson [LH64] ont développé un algorithme basé sur des étapes du pivot pour résoudre le problème (14).

Introduit par Fielder et Pták [FP66], la classe des P-matrices est reliée au problème de complémentarité linéaire. Murty [Mur72] montre que si M est une P-matrice alors il existe une unique solution au LCP.

Définition 0.12 (P-matrice)**.** Une matrice carrée M est une P-matrice si tous ses principaux mineurs (sous-déterminants) sont strictement positifs

$$\det(M_{\mathcal{I}}) > 0 \ , \quad \forall \mathcal{I} \subset [n] \ , \quad M_{\mathcal{I}} = (M_{k\ell})_{k,\ell \in \mathcal{I}} \ .$$

De nombreuses propriétés sur les conditions nécessaires et suffisantes pour qu'une matrice réelle soit une P-matrice ont été étudiées par Rump [Rum03] et Rohn [Roh12]. Numériquement, vérifier qu'une matrice est une P-matrice est co-NP complet [Cox94].

Théorème 0.17 (Existence et unicité d'une solution au problème LCP [Mur72])**.** *Une matrice M est une P-matrice si et seulement si le $LCP(M, \boldsymbol{q})$ a une unique solution $(\boldsymbol{w}, \boldsymbol{z})$ pour tout $\boldsymbol{q} \in \mathbb{R}^n$.*

Dans le cas du modèle de Lotka-Volterra et en considérant (13), nous cherchons $\boldsymbol{x}^* \in LCP(I - B, -\mathbf{1})$.

Définition 0.13 (M-matrice)**.** Une matrice carrée A est une M-matrice si elle peut être exprimée sous la forme $A = sI - C$, où $C = (C_{k\ell})$ avec $C_{k\ell} \geqslant 0, 1 \leqslant k, \ell \leqslant n$, et $s > \rho(C)$, le rayon spectral de C.

Le nom M-matrice a été donné par Ostrowski [Ost56] en référence à Hermann Minkowski. De nombreuses propriétés sur les M-matrices ont été introduites par Fiedler et Pták [FP62] et étendues par Plemmons [Ple77].

Remarque 0.7.

- L'ensemble des M-matrices non singulières est un sous-ensemble de la classe des P-matrices.
- L'ensemble des M-matrices non singulières est un sous-ensemble de la classe des matrices avec une inverse positive, c'est-à-dire que

$$A^{-1} \text{ existe et } A^{-1} \geqslant 0 \, .$$

Stabilité globale

Le theorème 0.17 donne une condition suffisante et nécessaire pour l'existence d'un équilibre unique non envahissable à l'équation (9). Dans le cas d'un équilibre faisable $\boldsymbol{x}^* > 0$, Bizeul et Najim [BN21] ont montré qu'il existe un équilibre globalement stable. Dans le cas d'un équilibre avec des espèces qui s'éteignent, il est nécessaire de revenir aux propriétés des fonctions de Lyapunov.

Le théorème de Lyapunov dit qu'une matrice A est stable (ses valeurs propres ont une partie réelle strictement négative) si et seulement s'il existe une matrice définie positive H telle que $HA + A^\top H$ est définie négative. Cette condition remonte aux travaux de Lyapunov [Lia07] qui ont été améliorés et étudiés par Barker *et al.* [BBP78] et Logofet [Log05] qui fait un résumé de toutes les conditions sur les matrices représenté par une fleur.

Définition 0.14 (Stabilité diagonale de Lyapunov)**.** Une matrice M est Lyapunov diagonallement stable, noté $M \in S_\omega$, si et seulement s'il existe une matrice diagonale D à éléments positifs telle que $DM + M^\top D$ est définie négative, c'est-à-dire que toutes ses valeurs propres sont négatives.

Proposition 0.18 (Takeuchi *et al.* [TAT78]). *Si $M \in S_\omega$ alors $-M$ est une P-matrice.*

Soit le système (4) et on considère que la matrice B est arbitraire,

$$\frac{d\,y_k(t)}{dt} = y_k(r_k + (-\theta I + B\boldsymbol{y})_k)\,, \quad k \in [n]\ . \tag{15}$$

Takeuchi et Adachi (voir par exemple [Tak96, Th. 3.2.1]) fournissent un critère pour l'existence d'un équilibre unique $\boldsymbol{y}^*$ et la stabilité globale du système LV.

Théorème 0.19 (Takeuchi et Adachi [TA80]). *Si $-\theta I + B \in S_\omega$, alors $LCP(\theta I - B, \boldsymbol{r})$ admet une unique solution. En particulier, pour tout $\boldsymbol{r} \in \mathbb{R}^n$, il y a un unique équilibre $\boldsymbol{y}^*$ à (15), qui est globalement stable dans le sens où pour chaque $\boldsymbol{y}_0 > 0$, la solution de (15) qui démarre à $\boldsymbol{y}(0) = \boldsymbol{y}_0$ satisfait*

$$\boldsymbol{y}(t) \xrightarrow[t\to\infty]{} \boldsymbol{y}^*\,.$$

Le modèle Lotka-Volterra vu par un physicien

Dans la section précédente, des conditions mathématiques ont été données pour l'existence d'un équilibre faisable et l'unicité d'un équilibre globalement stable dans (4) où certaines espèces peuvent disparaître. Cependant, la richesse des équations de Lotka-Volterra provient de la diversité de ses comportements dynamiques. Le manque de connaissances mathématiques est complété par des méthodes issues de la physique pour améliorer la compréhension de ces divers comportements dynamiques (propriétés de l'équilibre, dynamique hors équilibre, sophistication du modèle).

Depuis longtemps, la théorie de la mécanique statistique des systèmes désordonnés a été développée pour étudier les verres de spin par le système des répliques (voir Mezard *et al.* [MPV86] pour une revue).

L'utilisation de ces méthodes pour étudier les systèmes biologiques a été introduite pour la première fois par Diederich et Opper [DO89, OD92] et utilisé pour étudier la dynamique des équations des répliques multi-espèces (modèle équivalent au système Lotka-Volterra) par Tokita [Tok04]. Plus récemment, ces méthodes de physique statistique des systèmes désordonnés ont été adaptées pour résoudre des problèmes d'écologie théorique. En particulier, la méthode de la cavité dynamique est utilisée pour analyser la dynamique des communautés où des interactions aléatoires entre les espèces sont considérées. Les physiciens ont divisé l'espace des paramètres (μ, α, ρ) en un diagramme de phase où la question majeure est d'identifier les limites entre les différentes phases : point fixe stable unique, chaos avec des attracteurs multiples, croissance infinie etc.

La méthode de la cavité permet de dériver une équation de champ moyen approchant un problème non linéaire de haute dimension. Le concept clé consiste à supposer qu'un point fixe unique existe et à introduire une nouvelle espèce avec de nouvelles interactions dans le système existant. Après l'établissement de la nouvelle espèces, une analogie entre les propriétés des solutions avec n et $n+1$ espèces est vérifiée. Cette méthode est utilisée pour étudier le système (4) qui peut admettre un point fixe stable unique mais aussi une dynamique hors équilibre : une phase chaotique avec des attracteurs multiples. Bunin [Bun16, Bun17] a utilisé la méthode de la cavité dynamique pour mener et étendre des résultats plus généraux (propriétés des espèces persistantes, phase des attracteurs multiples) pour le diagramme de phase du système de Lotka-Volterra (4). Ces méthodes ont

été utilisées pour résoudre de nombreux problèmes en écologie théorique. En particulier, Barbier *et al.* [BABL18] présente des comportements génériques dans les communautés complexes. Pour une revue de la méthode de la cavité appliquée aux problèmes de dynamique des communautés, voir Barbier et Arnoldi [BA17].

Le modèle étudié par les physiciens est plus générique, le système désordonné de Lotka-Volterra est donné par

$$\frac{dx_k(t)}{dt} = x_k(t)\left(r_k - x_k(t) + \sum_{\ell\in[n]} B_{k\ell}x_\ell(t)\right) + \lambda_k + \omega_k\sqrt{2x_k(t)}\eta_k(t)\ , \tag{16}$$

où les λ_k sont des constantes de migration et $\omega_k\sqrt{2x_k(t)}\eta_k(t)$ est un terme de bruit démographique où $\eta_k(t)$ est une fonction aléatoire variant dans le temps.

Le diagramme de phase étudié par Bunin [Bun17] avait déjà été étudié numériquement par Kessler et Shnerb [KS15] avec la présence du paramètre de migration λ_k. Le modèle de Lotka Volterra avec un terme de bruit démographique a été étudié récemment par Bunin [Bun21] et Altieri *et al.* [ARCB21]. La méthode de la cavité dynamique permet de dériver l'équation dynamique du champ moyen pour les applications hors équilibre [RBBC19, RBBB20, ABC20], en particulier pour l'étude de la dynamique des attracteurs multiples représentée dans la Figure 8. En outre, Biroli *et al.* [BBC18] montrent que le régime des attracteurs multiples est analogue à une phase critique de verre de spin. Pour une revue dans les systèmes écologiques, voir Altieri [Alt22].

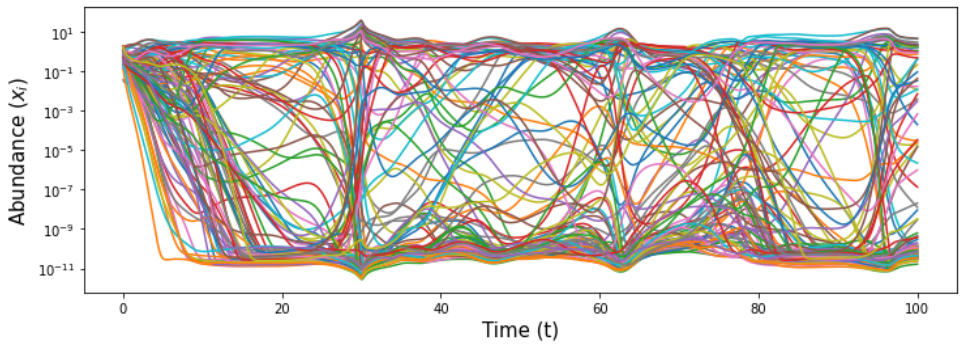


Figure 8: Dynamique d'un système LV de 100 espèces avec migration (16) dans la phase chaotique avec des attracteurs multiples et des paramètres $\mu = 4$, $\alpha = 0.5$, $\forall i \in [n]\,, \lambda_i = 10^{-10}$. L'axe des ordonnées est en échelle logarithmique.

Une autre méthode utilisant des techniques fonctionnelles génératrices pour établir des équations de champ moyen similaires afin d'étudier la phase d'équilibre dans le système LV a été utilisée par Galla [Gal18]. Des méthodes identiques ont été utilisées pour établir les valeurs propres des matrices aléatoires [BJRG22b, BJRG22a] et analyser les modèles de Lotka Volterra avec différentes structures d'interaction, comme le modèle en cascade [PBG22].

D'autres applications sont possibles, comme les travaux récents de Fraboul *et al.* [FBM22] sur les mutations dans le modèle LV ou l'étude de l'impact de l'effet Allee sur le diagramme de phase par Altieri *et al.* [AB22].

Contributions

Chapitre 1 - Equilibrium and persisting species in a large Lotka-Volterra system of differential equations

Ce chapitre est basé sur une prépublication de Clenet, Massol et Najim [CMN22].

Dans le Chapitre 1, nous nous concentrons sur le modèle (6) où la matrice d'interaction B est une version simplifiée de (7) qui admet la représentation suivante :

$$B = \frac{A}{\alpha\sqrt{n}} + \frac{\mu}{n}\mathbf{1}_n\mathbf{1}_n^\top ,$$

où $A = (A_{k\ell})$ est une matrice avec des variables aléatoires normalisées ($\mathbb{E}A_{k\ell} = 0$ et $\mathrm{Var}(A_{k\ell}) = 1$) indépendantes et identiquement distribuées (i.i.d.) avec un moment d'ordre quatre fini, $\alpha > 0$ est un paramètre supplémentaire reflétant la force d'interaction et $\mu \in \mathbb{R}$ représente une tendance arbitraire des interactions.

Dans le theorème 0.15, Bizeul et Najim prouvent l'existence d'un seuil $\alpha \sim \sqrt{2\log(n)}$ dans le cas $\mu = 0$, qui garantie la faisabilité de l'équilibre $\boldsymbol{x}^*$ de (6). Cependant, Dougoud *et al.* [DVR+18] ont montré que certaines espèces s'éteignent lorsque $\alpha > 0$ est indépendant de n. L'objectif de ce chapitre est de décrire l'impact de la force d'interaction α et de la tendance d'interaction μ sur les conditions de coexistence des espèces en interaction.

Premièrement, en combinant les résultats de Takeuchi et Adachi du théorème 0.19 avec les résultats standard de RMT du théorème (0.27), nous fournissons des conditions suffisantes sur les paramètres α et μ pour assurer l'existence d'un équilibre unique globalement stable $\boldsymbol{x}^*$ en grande dimension $n \to \infty$. L'équilibre est composé d'espèces persistantes et d'espèces disparues.

Par la suite, étant donné un équilibre unique $\boldsymbol{x}^*$, nous décrivons les propriétés des espèces persistantes. Dans cette perspective, nous fournissons une heuristique pour calculer asymptotiquement la proportion d'espèces persistantes et nous analysons via un système d'équations la dépendance entre les paramètres α, μ et la proportion d'espèces persistantes. De plus, nous montrons à l'aide d'une heuristique que la distribution de l'abondance des espèces persistantes est une gaussienne tronquée (voir Figure 9).

Dans la nature, les interactions entre espèces sont en constante évolution et sont affectées par l'environnement. Sous l'hypothèse que les conditions environnementales influencent la force d'interaction, nous étudions les conséquences d'un changement soudain des conditions environnementales, exprimé par un changement brutal du paramètre α. Lorsque α varie pour la même matrice A et le même paramètre μ, le système peut présenter différents états. Lorsque la valeur de α augmente au-delà d'une certaine valeur critique, toutes les espèces coexistent ; à l'inverse, pour des valeurs suffisamment faibles de α, les espèces peuvent disparaître tout en conservant un équilibre stable unique. Nous décrivons le changement entre ces deux états et comment la proportion d'espèces persistantes varie en résolvant numériquement le système de Lotka-Volterra. Nous observons qu'une diminution de α affecte négativement la richesse spécifique de l'équilibre (voir Figure 10).

Enfin, nous analysons un indice de diversité (nombres de Hill d'ordre 1) pour avoir une représentation plus précise de la dynamique de la biodiversité. La dynamique de cette

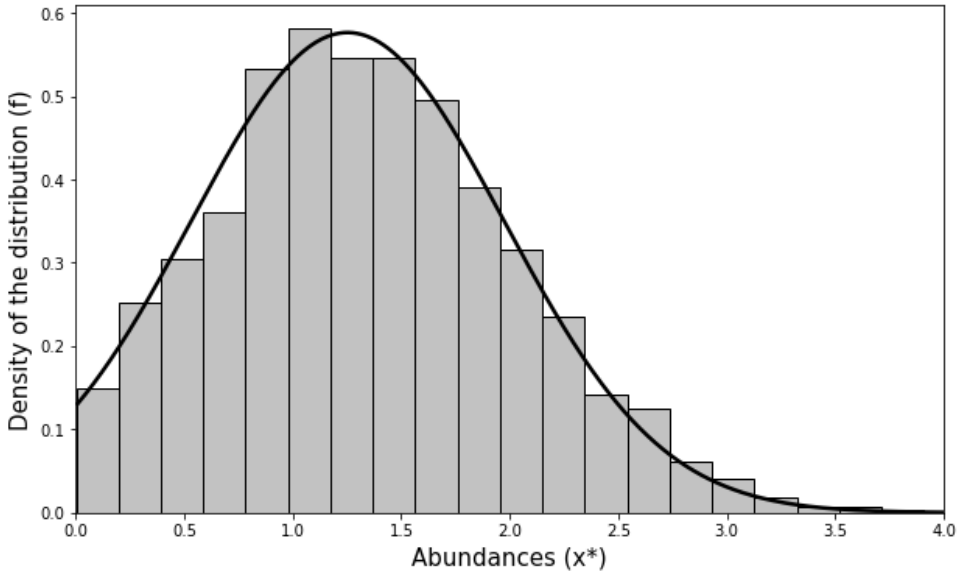


Figure 9: Distribution de l'abondance des espèces persistantes. L'axe des abscisses représente la valeur des abondances et l'histogramme est construit sur les composantes positives de l'équilibre $\boldsymbol{x}^*$. La ligne solide représente la distribution théorique des paramètres (α, μ) donnée par l'heuristique. Les entrées sont gaussiennes $\mathcal{N}(0, 1)$ et les paramètres sont fixés à $(n = 2000, \alpha = 2, \mu = 0.2)$.

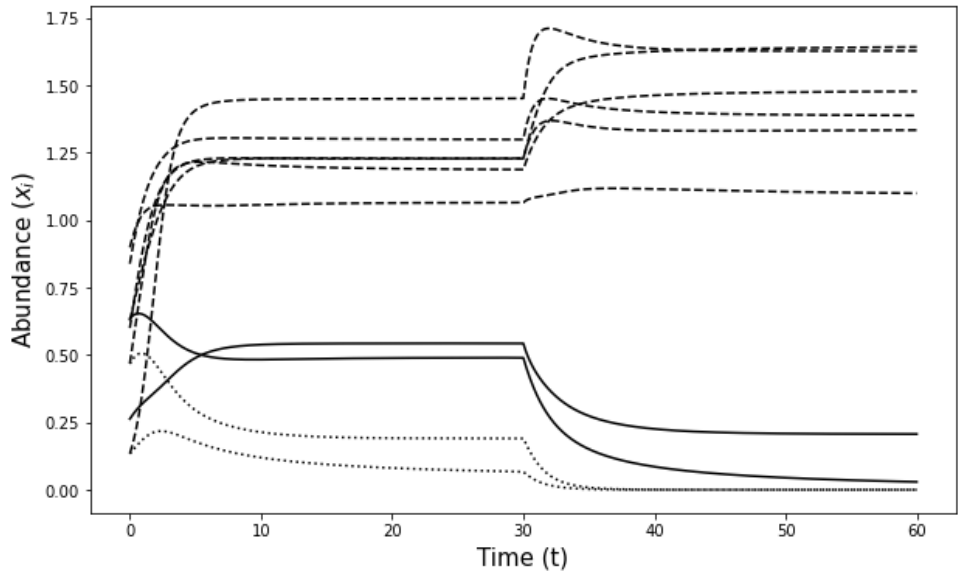


Figure 10: Dynamique des abondances dans le cas d'une communauté de dix espèces. La matrice des interactions A et les conditions initiales sont communes et nous appliquons une variation brutale de $\alpha(t)$ à $t = 30$. Les lignes hachées représentent les espèces qui bénéficient de la variation de l'habitat ; les lignes pleines représentent les espèces qui souffrent du changement. Les lignes en pointillés représentent les espèces en voie d'extinction.

mesure de diversité suggère que la moyenne des coefficients d'interaction, μ, affecte la durée de la dynamique transitoire, une dynamique transitoire plus courte étant associée à des interactions plus mutualistes (c'est-à-dire des valeurs positives plus élevées de μ).

Chapitre 2 - Equilibrium in a large Lotka-Volterra system with pairwise correlated interactions

Ce chapitre est basé sur l'article écrit par Clenet, El Ferchichi et Najim publié dans le journal Stochastic Processes and its Applications (Novembre 2022) [CEFN22].

Dans le Chapitre 2, nous nous concentrons sur le modèle (7) où nous étendons le résultat sur le seuil de faisabilité de Bizeul et Najim (0.15). La matrice d'interaction B_n est une matrice aléatoire non centrée avec des entrées corrélées par paire :

$$B_n = \frac{A_n}{\alpha_n \sqrt{n}} + \frac{\mu}{n} \mathbf{1}_n \mathbf{1}_n^\top ,$$

où $A_n = (A_{k\ell})_{k,\ell \in [n]}$ est une matrice aléatoire qui satisfait deux conditions (i) $(A_{k\ell}, k \leqslant \ell)$ sont des variables aléatoires gaussiennes standard $\mathcal{N}(0,1)$ indépendantes et identiquement distribuées (ii) pour $k < \ell$ le vecteur $(A_{k\ell}, A_{\ell k})$ est un vecteur gaussien standard bivarié, indépendant des autres variables aléatoires, avec une covariance $\mathrm{cov}(A_{k\ell}, A_{\ell k}) = \rho$ et $|\rho| \leqslant 1$. La suite de nombres positifs (α_n) est soit fixe, soit infinie. Le paramètre μ est un nombre réel fixé.

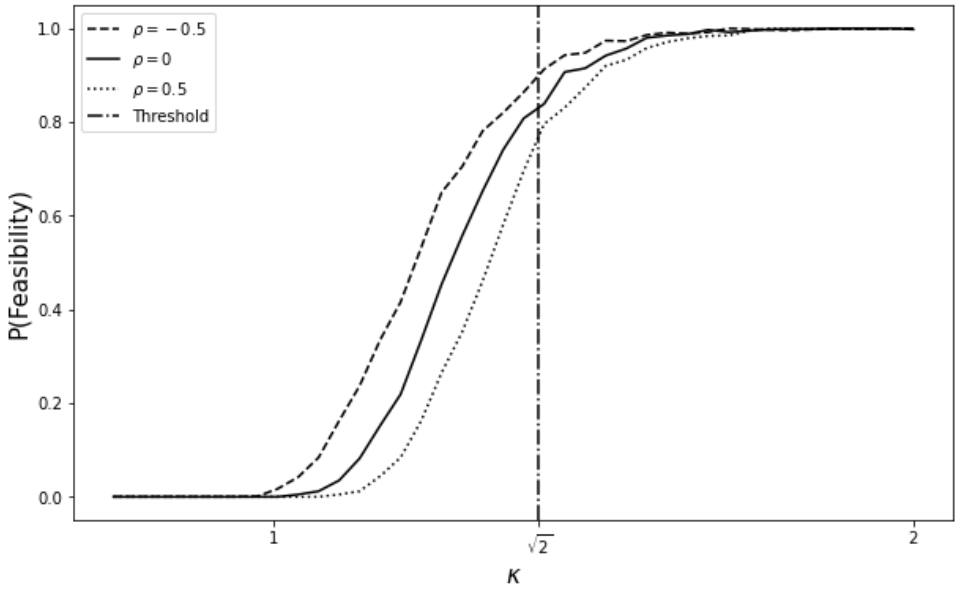


Figure 11: Transition vers la faisabilité pour le modèle elliptique. Pour chaque κ sur l'axe des abscisses, nous simulons 1000 matrices B_n de taille $n = 1000$, puis nous calculons la solution $\boldsymbol{x}_n$ du théorème de faisabilité à l'échelle $\alpha_n(\kappa) = \kappa\sqrt{\log(n)}$ puis de tracer la proportion de solutions faisables obtenues pour les 1000 simulations. Chaque courbe représente la proportion de solutions faisables $\boldsymbol{x}_n$ pour trois valeurs distinctes $\rho \in \{-0.5,\ 0,\ 0.5\}$. La ligne verticale en pointillés correspond à $\kappa = \sqrt{2}$ c.a.d. l'échelle critique $\alpha_n^* = \sqrt{2\log(n)}$.

On prouve que la faisabilité est atteinte lorsque $\alpha_n \gg \sqrt{2\log(n)}$ et $\mu < 1$, et qu'il n'y a pas de faisabilité autrement. De plus, le paramètre de corrélation ρ n'a aucune influence puisque le seuil de transition de phase est le même que dans le cas i.i.d. [BN21] : les corrélations induites entre les composantes x_k de la solution $\boldsymbol{x}_n$ sont trop faibles (voir Figure 11). De plus, nous prouvons que la même transition de phase se produit si

nous considérons un profil de covariance ($\rho_{k\ell}$, $k < \ell$) où $\rho_{k\ell} = \mathrm{cov}(A_{k\ell}, A_{\ell k})$ au lieu d'un paramètre de covariance fixe ρ.

En utilisant les résultats de Takeuchi et Adachi (0.19) sur la stabilité des systèmes LV avec les résultats RMT du théorème 0.14, nous établissons des conditions suffisantes pour l'existence d'un équilibre stable unique où certaines espèces peuvent disparaître, ce qui représente une extension du Chapitre 1.

Enfin, nous concluons avec un résultat important sur l'estimation de la proportion d'espèces persistantes. En utilisant les arguments des physiciens, nous énonçons le problème ouvert, nous rappelons les équations de Bunin et Galla et nous fournissons des simulations d'un système d'équations pour calculer la proportion d'espèces persistantes.

Chapitre 3 - Impact of a block structure in large systems of Lotka-Volterra

Ce chapitre est un projet en cours entre Clenet, Massol et Najim.

Dans le Chapitre 3, nous visons à développer les résultats des Chapitres 1 et 2 dans un écosystème comportant de nombreuses communautés. Dans la nature, les réseaux d'interactions sont plutôt structurés, ce qui contribue à la stabilité du système. Afin d'étendre les résultats du modèle de Lotka-Volterra pour décrire les propriétés d'une dynamique multi-communautés, nous définissons une matrice d'interaction par blocs dans laquelle nous pouvons adapter les interactions intra et inter-communautés. Pour des raisons d'interprétation dues à la complexité du modèle, nous considérons le cas de deux communautés en interaction (voir Figure 12). Dans le cadre de 2 communautés, la matrice $B = (B_{k\ell})_{n,n}$ est définie comme suit :

$$B = \frac{1}{\sqrt{n}} \begin{pmatrix} \frac{A_{11}}{\alpha_{11}} & \frac{A_{12}}{\alpha_{12}} \\ \frac{A_{21}}{\alpha_{21}} & \frac{A_{22}}{\alpha_{22}} \end{pmatrix} + \frac{1}{n} \begin{pmatrix} \mu_{11}\mathbf{1}_{\mathcal{I}_1}\mathbf{1}_{\mathcal{I}_1}^\top & \mu_{12}\mathbf{1}_{\mathcal{I}_1}\mathbf{1}_{\mathcal{I}_2}^\top \\ \mu_{21}\mathbf{1}_{\mathcal{I}_2}\mathbf{1}_{\mathcal{I}_1}^\top & \mu_{22}\mathbf{1}_{\mathcal{I}_2}\mathbf{1}_{\mathcal{I}_2}^\top \end{pmatrix}, \tag{17}$$

où :

$$\boldsymbol{s} = \begin{pmatrix} 1/\alpha_{11} & 1/\alpha_{12} \\ 1/\alpha_{21} & 1/\alpha_{22} \end{pmatrix}, \ \boldsymbol{\mu} = \begin{pmatrix} \mu_{11} & \mu_{12} \\ \mu_{21} & \mu_{22} \end{pmatrix},$$

Le paramètre $\boldsymbol{\beta} = (\beta_1, \beta_2)$, $\sum_{i=1}^2 \beta_i = 1$ est la taille en proportion de chacun des blocs, $\mathcal{I}_i$ est un sous-ensemble de $[n]$ de taille $|\mathcal{I}_i| := \beta_i n$ correspondant à l'indice des espèces appartenant à la communauté i, $\mathbf{1}_{\mathcal{I}_i}$ est un vecteur d'entrée de 1 de taille $\beta_i n$. La matrice A_{ij} est une matrice aléatoire non Hermitienne de taille $(\beta_i n, \beta_j n)$ avec des entrées gaussiennes centrées réduites, c'est-à-dire $\mathcal{N}(0,1)$.

La matrice $\boldsymbol{s}$ représente la force d'interaction dans chaque bloc. La matrice de tendance $\boldsymbol{\mu}$ permet d'ajuster en moyenne le type d'interaction (mutualisme, compétition) de chaque bloc.

Dans une première section, nous étendons le résultat de faisabilité de Bizeul et Najim (0.15) pour une matrice d'interaction par blocs où $\boldsymbol{\mu} = \mathbf{0}$. En utilisant ce résultat, nous étudions le maintien de la faisabilité de deux communautés lorsqu'on ajoute des interactions entre elles. Les interactions entre les communautés réduisent la faisabilité et si nous supposons que les communautés peuvent varier en taille et que les interactions intra-communauté sont différentes, la communauté avec la force d'interaction la plus faible est avantagée, c'est-à-dire que la taille de la communauté peut être plus grande. Nous

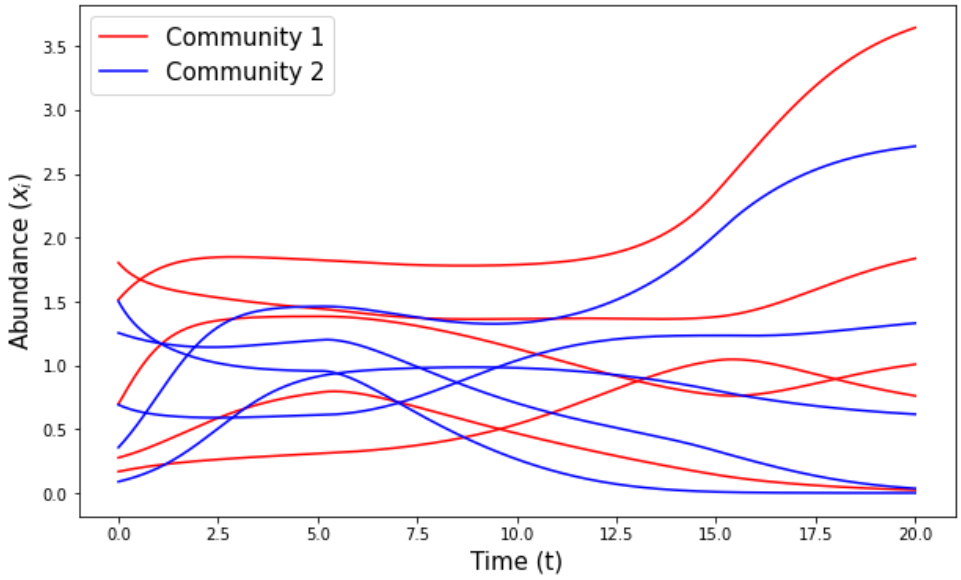


Figure 12: Dynamique du modèle (7) de 2 communautés distinctes composées de 5 espèces avec une matrice d'interaction (17). À $t = 0$, les deux communautés convergent vers leur point d'équilibre faisable et ne sont pas en interaction. À $t = 5$, les deux communautés commencent à interagir c.a.d. α_{12} et α_{21} augmentent d'une façon linéaire jusqu'à $t = 15$. Ensuite, les deux communautés convergent vers leur nouveau point d'équilibre avec des espèces persistantes et éteintes dans les deux communautés.

concluons cette première partie en étudiant le cas non centré $\boldsymbol{\mu} \neq \mathbf{0}$, où nous donnons des conditions de faisabilité en utilisant des propriétés sur les M-matrices.

Les deuxième et troisième parties de ce chapitre sont une extension des résultats du Chapitre 1 et une interprétation écologique des résultats. D'une part, dans la deuxième partie, nous étudions l'existence d'un équilibre unique globalement stable où les espèces peuvent s'éteindre en utilisant le théorème 0.19 de Takeuchi et Adachi et des résultats de RMT, en particulier la théorie sur l'équation vectorielle quadratique. Nous établissons un théorème pour le cas $\boldsymbol{\mu} = 0$, puis nous étudions le cas non centré $\boldsymbol{\mu} \neq 0$ lorsque la force d'interaction est similaire dans chaque bloc. Contrairement aux résultats obtenus dans le cas d'une seule communauté, l'augmentation de la compétition inter-communautés peut déstabiliser le système.

D'autre part, la troisième partie décrit les heuristiques sur les propriétés et la distribution des abondances des espèces persistantes dans chaque bloc (voir Figure 13). Une interprétation graphique de ces heuristiques met en évidence plusieurs résultats. Il existe une contagion de la diversité : plus la persistance d'une communauté est élevée, moins son impact sera néfaste sur les autres communautés. Le déclin de la persistance entre deux communautés en interaction n'est pas linéaire mais a un double effet négatif, d'où l'importance de maintenir des communautés persistantes et de ne pas négliger les phénomènes de rétroaction dans les interactions entre les communautés. Nous concluons par une étude de l'impact des interactions mutualistes et compétitives.

Dans une quatrième et dernière section, une étude de similarité numérique est réalisée entre un modèle où la force d'interaction varie et un modèle où la connectance dans chacune des communautés est variable, ce qui donne une matrice d'adjacence du graphe d'interaction connu sous le nom de Bernouilli Stochastic Block Model (voir Figure 14).

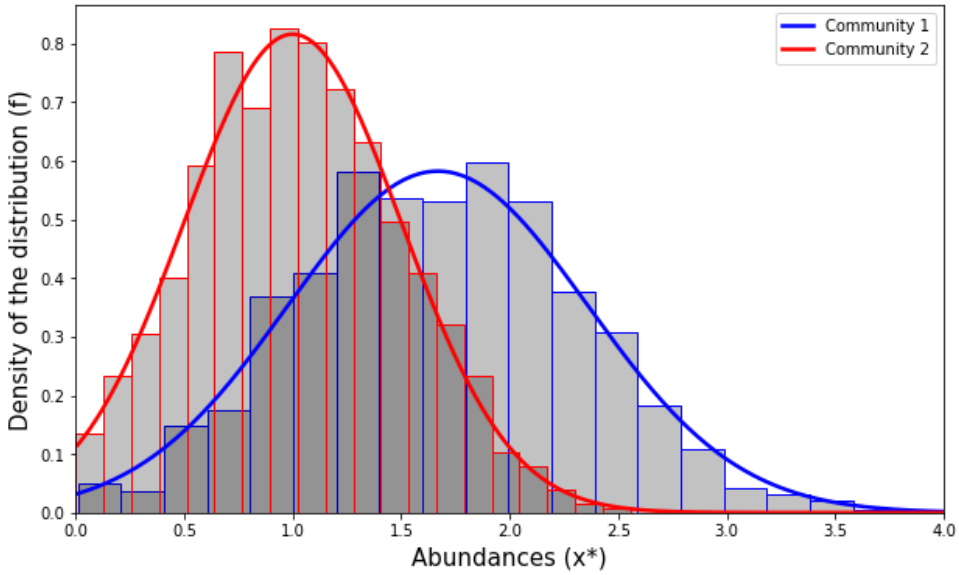


Figure 13: Distribution des abondances des espèces persistantes pour chaque communauté. L'axe des abscisses représente la valeur des abondances et l'histogramme est construit à partir des composantes positives de l'équilibre $\boldsymbol{x}^*$ associé à chaque communauté. La ligne continue bleue (resp. la ligne continue rouge) représente la distribution théorique de la communauté 1 (resp. communauté 2) en fonction des paramètres $(\boldsymbol{\alpha}, \boldsymbol{\mu})$ donnés par les heuristiques. Les entrées sont gaussiennes $\mathcal{N}(0,1)$ et les paramètres sont fixés à

$$n = 1000\ ,\ \boldsymbol{\mu} = \begin{pmatrix} 0.5 & 0.5 \\ 0 & 0 \end{pmatrix}\ ,\ \boldsymbol{\alpha} = \begin{pmatrix} 2 & 3 \\ 3 & 3 \end{pmatrix}\ ,\ \boldsymbol{\beta} = \left(\frac{1}{2}, \frac{1}{2}\right)\ .$$

Cette similitude est analysée à travers la condition de stabilité donnée historiquement par May [May72].

Chapitre 4 - A probabilistic perspective of the hierarchical competition-colonization trade-off model

Ce chapitre est un projet en cours entre Allesina, Clenet, Della Libera, Massol et Miller.

Dans le Chapitre 4, nous étudions le modèle compétition-colonisation à plusieurs espèces (8), dans le cas d'une compétition hiérarchique c.a.d.

$$\eta_{k\ell} = \begin{cases} 1 & \text{si } k < \ell\,, \\ 0 & \text{sinon}\ . \end{cases}$$

La dynamique des abondances de chaque espèce au sein de l'habitat dépend principalement de son taux de colonisation et de son taux d'extinction. Nous proposons une interprétation probabiliste du modèle en échantillonnant les taux de colonisation à partir d'une distribution de probabilité donnée avec le même taux d'extinction pour toutes les espèces. Dans ce cadre, nous étudions deux types différents de processus d'assemblage.

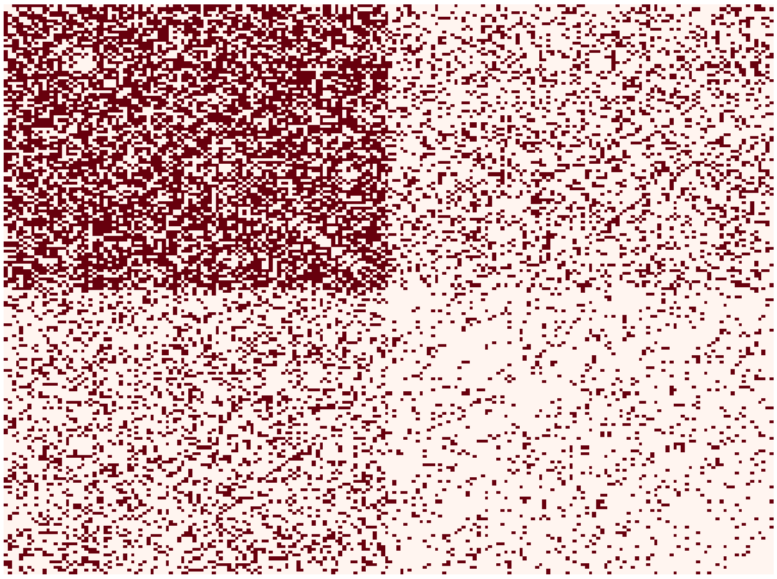


Figure 14: Représentation d'une matrice d'adjacence des interactions d'un écosystème de taille $n = 200$. Un graphe d'un Stochastic Block model symétrique de paramètre $P = \begin{pmatrix} 0.6 & 0.25 \\ 0.25 & 0.1 \end{pmatrix}$. Une cellule de couleur rouge indique $S_{k\ell} = 1$, au contraire, une cellule de couleur blanche indique qu'il n'y a pas d'interaction $S_{k\ell} = 0$.

D'une part, nous supposons que initialement la communauté contient un pool de n espèces (processus de métacommunauté "tout en une fois") et nous laissons les occupations de toutes les espèces évoluer dans le temps selon l'équation (8) où certaines espèces persistent alors que d'autres peuvent disparaître. De manière surprenante, nous obtenons un résultat d'universalité de la distribution du nombre d'espèces persistantes. Pour une large gamme de distribution, en moyenne la proportion d'espèces persistantes est de un demi. De plus, nous montrons (pour la distribution uniforme) que la distribution des espèces persistantes est une distribution binomiale $B(n, 1/2)$ (voir Figure 15).

Outre le résultat d'universalité, nous décrivons les propriétés des espèces persistantes en rappelant quelques résultats de Kinzig *et al.* [KLD+99] sur la fraction des parcelles vides et des occupations. Nous terminons cette section en clarifiant l'hypothèse de perturbation intermédiaire qui avait été observée par Hastings [Has80] où la coexistence optimale entre les espèces se produit lorsque le taux de mortalité est intermédiaire.

D'autre part, nous étudions un processus d'invasion séquentielle. Partant d'un habitat vide, celui-ci est rempli par l'introduction séquentielle d'espèces dont les taux de colonisation sont tirés selon une distribution spécifique. Nous observons que le nombre d'espèces persistantes sature avec une croissance logarithmique due aux contingences historiques et aux cascades d'extinction (voir Figure 16). Nous analysons les propriétés des contingences historiques dû au phénomène des cascades d'extinction qui est un élément clé du phénomène de saturation.

Nous donnons quelques éléments de réponse théoriques avant de procéder à une analyse numérique du modèle du processus d'assemblage. Le résultat d'universalité n'est plus vrai et une différence majeure est observée entre les distributions à queue régulière et à queue lourde. En général, plus la queue est lourde, plus la diversité est grande. Cette hypothèse de compromis entre compétition et colonisation montre l'importance de trouver

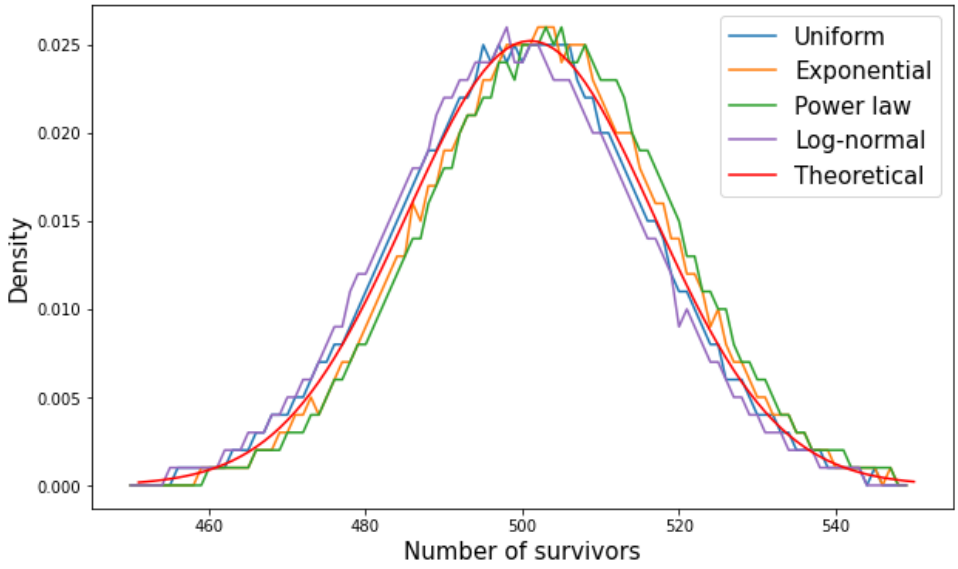


Figure 15: Représentation de la distribution du nombre d'espèces persistantes pour un pool initial de $n = 1000$ espèces et pour différentes distributions du taux de colonisation. Chaque courbe est obtenue par des expériences de Monte Carlo en calculant $P = 100000$ fois l'algorithme et en stockant les valeurs obtenues pour former le contour d'un histogramme. La courbe rouge correspond à la fonction de densité de la distribution binomiale $\mathcal{B}(n, \frac{1}{2})$.

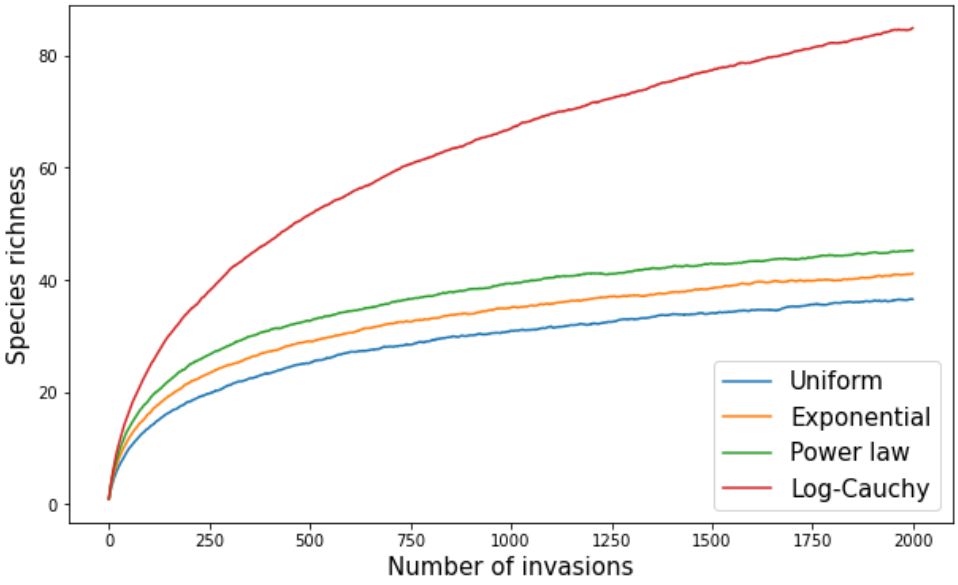


Figure 16: Représentation de la richesse spécifique du modèle d'invasion séquentielle en fonction du nombre d'invasions pour différentes distributions. La courbe est obtenue à l'aide de simulations de Monte Carlo en calculant $P = 2000$ fois et en calculant la moyenne du nombre d'espèces persistantes.

un équilibre entre les compétiteurs et les colonisateurs.

Pour conclure, cette perspective probabiliste du modèle hiérarchique de compétition-colonisation multi-espèces met en avant et compare deux types différents d'assemblages

distincts et donne les conditions pour que de nombreuses espèces coexistent sous le compromis compétition-colonisation.

Introduction

Ecological issues

A better understanding of ecosystems

Ecology is etymologically the science of the house (from ancient Greek oîkos). By definition, it is the science of living beings (animals, micro-organisms, etc.) in a specific environment at a particular scale (populations, species, communities) and their relations with other living beings. In ecology, a species gathers only individuals that can reproduce with one another and that can produce fertile offspring and thus form populations. Individuals are usually mutually dependent on each other for their survival.

Individuals of different species living in the same region form a community. All living beings in their environment form an ecosystem (savannah, forest, intestine, etc.). Historically, the concept of ecosystem is not old [Tan39]. Understanding these ecosystems (community ecology, when the focus is on the interaction between species) and their underlying functioning mechanisms is a major challenge in ecology [MW67].

The number of species in an ecosystem is often related to a measure of scale. On earth, some rare ecosystems are small (2-3 species), however, there are many ecosystems with a very large number of species. This great diversity of species is necessary for the survival of living beings. To date, scientists have listed more than 2 million species on earth. For example the Amazon forest is home to 10^6 species. On our own scale, our microbiome hosts an order of magnitude of 10^3 species and 10^{18} of cells [CSF15]. Many questions about these large ecosystems remain unanswered because more complex systems ask for much more empirical data to be "understood" and experimental studies are not suited to the study of large ecosystems. A better understanding would allow to manage animal populations sustainably to protect endangered populations, or to have a better management of antibiotics on our intestinal flora.

Ecologists carry out many experimental studies on small systems of species, whereas in large systems this quickly becomes intractable to collect data at a large scale. However, in recent years, many technological tools have been developed in laboratories for studies in microbiology systems and would allow comparisons with theoretical studies [HAB+21]. For example, many processes are automated with the emergence of deep learning to recognize species and count them, especially on images taken by plane or drone. This lack of data can be compensated by the use and study of models. Such models do not necessarily have the sole purpose of predicting the evolution of the ecosystem but to understand the mechanisms that allow a great diversity i.e. toy models.

One of the major debates in ecology is the relationship between diversity and stability in an ecosystem. In the 70', many ecologists suggest that diverse communities enhanced ecosystem stability [Mac55, May73]. However, initiated by a theoretical model

that he introduced in the 70' [May72], May challenged the diversity-stability relationship by using linear stability analysis on randomly constructed community model and found that diversity tends to destabilize the system. This lead to the diversity-stability debate [May73, Yod81, McC00, IC07, JMM+16, LMB+18] where the theoretical issues are to find the missing arguments or mechanisms of the model of May.

In this thesis, I study large ecosystems in order to understand one of the major ecological factors affecting their diversity and dynamics: biotic interactions between species. In a community, all the interactions among its component are represented as a network of interactions. There are two main classes of interactions: intra-specific and inter-specific interactions. The first class corresponds to the interactions within the same species that can be negative (competition) and positive (Allee effect). The second, more diversified, corresponds to interactions between two different species, for example competition, mutualism, predation, etc. Community ecology is a sub-discipline of ecology which focuses on understanding the evolution of abundances (:=number of individuals) of the species that compose a community over time. To summarize, I use a model that describes community dynamics to understand the impact of the interactions network the properties of an equilibrium such as existence, diversity, stability, etc.

Crash course on ODEs

In ecology, population dynamics can be modeled in continuous or discrete time. In continuous time, ordinary differential equations (ODE) are used to describe the evolution of the abundance $\boldsymbol{x} = (x_1, ..., x_n)$ of a n-species system. We start by recalling the Cauchy problem.

Let U be an open set of $\mathbb{R}^{n+1}$, $f : U \to \mathbb{R}^{n+1}$ continuous with respect to $(t, \boldsymbol{x})$,

$$\begin{cases} \frac{d\boldsymbol{x}(t)}{dt} = f(t, \boldsymbol{x}(t)) \,, \\ \boldsymbol{x}(0) = x_0 \,. \end{cases} \tag{18}$$

A system is said to be autonomous if f does not depend on t.

The Cauchy-Lipschitz theorem states that if f is $\mathcal{C}^1$ with respect to $\boldsymbol{x}$, then for any initial condition the problem (18) admits a unique maximal solution (I, γ), $\gamma : I \mapsto \mathbb{R}^n$. Moreover, any other solution of the problem (18) is a restriction of the maximal solution.

If the system is autonomous, the theorem of a priori majorations indicates that if $f : U \to \mathbb{R}^n$ is locally Lipschitz continuous, (I, u) a maximal solution of Cauchy's theorem and $\sup(\mathrm{I}) = \infty$, then there exists a global solution of the problem (18).
In the following, we focus on the autonomous problem:

$$\begin{cases} \frac{d\boldsymbol{x}(t)}{dt} = f(\boldsymbol{x}(t)) \,, \\ \boldsymbol{x}(0) = x_0 \,. \end{cases} \tag{19}$$

In order to study the problem (19), an important issue consists in obtaining information on the existence and uniqueness of the equilibrium and their properties. An equilibrium $\boldsymbol{x}^*$ of the system (19) is a solution of the equation:

$$\frac{d\boldsymbol{x}(t)}{dt} = 0 \quad \Leftrightarrow \quad f(\boldsymbol{x}(t)) = 0 \,.$$

The function f can be complex, the solutions of this system are not necessarily trivial and there can be several equilibria with different properties. An equilibrium $\boldsymbol{x}^*$ is feasible if all its components are positive i.e.

$$\boldsymbol{x}^* > 0 \quad \Leftrightarrow \quad x_i > 0, \forall i \in [n] \,.$$

A major property of an equilibrium point is its stability. An equilibrium is stable if it returns to its equilibrium value after a small perturbation of the abundance vector $\boldsymbol{x}$. In the case of a linear differential equation, the study of stability is trivial and depends on the eigenvalues of the linear operator. In the non-linear case, it is more complex. However, one can linearize the system to obtain information on the local stability around the equilibrium.

Given $\boldsymbol{x}^*$ an equilibrium point of (19), we say that $\boldsymbol{x}^*$ is asymptotically stable if it is stable and if $\exists\, \delta > 0$, $\forall (I, \boldsymbol{x})$ solution of (19)

$$\exists\, t_0 \in I, |\boldsymbol{x}(t_0) - \boldsymbol{x}^*| \leqslant \delta \quad \Rightarrow \quad \boldsymbol{x}(t) \xrightarrow[t\to\infty]{} \boldsymbol{x}^* \,.$$

Theorem 0.20 (Stability of an equilibrium, non-linear case)**.** *Let $\boldsymbol{x}^*$ be an equilibrium for an autonomous nonlinear differential system where f is differentiable in $\boldsymbol{x}^*$ and let $Df(\boldsymbol{x}^*) = J|_{\boldsymbol{x}^*}$ be its Jacobian. Let $\Lambda = \mathrm{Sp}(J|_{\boldsymbol{x}^*})$ be the set of eigenvalues of $J|_{\boldsymbol{x}^*}$ and $\mathcal{R}(\Lambda)$ the set of real parts of eigenvalues of Λ.*

1. *If $\forall \lambda \in \Lambda, \mathcal{R}(\Lambda) < 0$, then $\boldsymbol{x}^*$ is asymptotically stable and we have*

$$\forall \mu \in]0, \min -\mathcal{R}(\Lambda)[, \forall \epsilon > 0, \exists \delta > 0,\ |\boldsymbol{x}(t_0) - \boldsymbol{x}^*| < \delta \\ \Rightarrow \forall t \geqslant t_0, \boldsymbol{x}(t) \text{ exist and } |\boldsymbol{x}(t) - \boldsymbol{x}^*| \leqslant \epsilon e^{-\mu(t-t_0)} \,.$$

2. *If $\exists \lambda \in \Lambda, \mathcal{R}(\Lambda) > 0$, then $\boldsymbol{x}^*$ is unstable.*

3. *If $\forall \lambda \in \Lambda, \mathcal{R}(\Lambda) \leqslant 0$ and there are pure imaginary eigenvalues, then we cannot conclude.*

Let $\boldsymbol{x}^*$ be an equilibrium for an autonomous nonlinear differential system (19), we say that $\boldsymbol{x}^*$ is asymptotically globally stable if for every $\boldsymbol{x}_0 > 0$, the solution to (19) which starts at $\boldsymbol{x}(0) = \boldsymbol{x}_0$ satisfies

$$\boldsymbol{x}(t) \xrightarrow[t\to\infty]{} \boldsymbol{x}^* \,.$$

Remark 0.8. In this thesis, we investigate exclusively asymptotic stability i.e. stability and convergence to the equilibrium point. By abuse of notation, we refer the study of the stability towards the asymptotic stability. I advise the reader to look at the book of Hirsch *et al.* [HSD74] for a complete review on ODEs.

Lotka-Volterra model

Differential equations are frequently used in biology to describe a system of interacting species. A particularly used form is the density-dependent model:

$$\begin{cases} \frac{d\boldsymbol{x}(t)}{dt} = \boldsymbol{x}(t) f(\boldsymbol{x}(t)) \,, \\ \boldsymbol{x}(0) = x_0 \,, \end{cases} \tag{20}$$

where f is commonly called the fitness or growth rate of a species. When $f(\boldsymbol{x}(t)) < 0$ the dynamics will be decreasing and the opposite when $f(\boldsymbol{x}(t)) > 0$. If the system is defined in the nonnegative orthant $\mathbb{R}^n_+$, then the system is forward invariant: if $\forall\, x_i(0) \geqslant 0$, then $\forall t > 0 : x_i(t) \geqslant 0$ [HS98]. One of the most widely used density-dependent models in ecology is the Lotka-Volterra model which is the cornerstone of this thesis.

Historically, Thomas Robert Malthus (1766-1834) was interested in modeling the fluctuations of a population. His conclusion was that without constraints, abundance in his model grows exponentially.

$$\begin{cases} \frac{dx(t)}{dt} = rx(t)\,, \\ x(0) = x_0\,, \end{cases}$$

where $r = \text{birth} - \text{death}$ and the analytical solution is $x(t) = x_0 e^{rt}$. A solution that lacks a bit of realism.

Later, Pierre François Verhulst (1804-1849) was interested in a more realistic model by assuming that the model is limited by a maximum size $K > 0$ (carrying capacity)

$$\begin{cases} \frac{dx(t)}{dt} = rx(t)\left(1 - \frac{x(t)}{K}\right)\,, \\ x(0) = x_0\,. \end{cases}$$

His logistic model represents, for example, the growth limit of a zebra population in the savanna due to the shortage of resources.

In a second time, scientists were interested in modeling interactions between species. When adding interactions between populations, the simplest model is named after two scientists, Lotka and Volterra who formulated it independently at the end of the 20's [Lot25, Vol26]. Classically studied in the form of a 2-dimensional prey-predator model, it has been compared with data from natural populations [Huf58].

From a general point of view and in higher dimensions, the Lotka-Volterra equations or Generalized Lotka-Volterra model play a key role in the study of population dynamics over time. This model is mathematically tractable but also very versatile and robust, and forms a first step in the development of ecological models. This model has been studied both in ecology [Wan78, Jan87, LB92] and in mathematics [GJ77, Goh77, Tay88, HS98, Tak96].

From a mathematical point of view, this model describes the population dynamics of a n-species system. It is defined by a system of n differential equations:

$$\frac{dx_k(t)}{dt} = x_k(t)\left(r_k - \theta x_k(t) + \sum_{\ell\in[n]} B_{k\ell}x_\ell(t)\right)\,, \tag{21}$$

where $k \in [n] = \{1, \cdots, n\}$. The abundance of species k at time t is represented by $x_k(t)$ and $\boldsymbol{x} = (x_1, \cdots, x_n)$ is the vector of abundances of the various species. Parameter θ is the self regulation coefficient or intra-specific interaction of each species. Parameter r_k corresponds to the intrinsic growth rate of species k. The coefficient $B_{k\ell}$ is the per capita effect of species ℓ on the growth rate of species k. The matrix B, representing the interaction network structure, can often be decomposed in different forms i.e. blocks, cascade, multiplex networks, graphons, etc [CN88, SCG+05, LIPJ+06, PEM12].

The goal of many mathematicians and ecologists is to understand the behavior of the system as a function of these different input parameters. For example, the number of equilibria, their stability and feasibility to understand the resulting ecological implications.

As mentioned before, the major issue of working with large systems is the difficulty of observing or estimating information on the interaction matrix. A natural choice is to replace the interactions by random coefficients whose statistical properties (mean, variance, etc) and structure (block, cascade, etc) encode some of the true properties of the food web. The matrix B becomes a complex mathematical object: a random matrix. This mathematical object represents the second cornerstone of this thesis.

Emergence of random matrices in ecology

In the 70's, following the work of Gardner and Ashby [GA70], Robert May reopened the long-standing diversity vs stability debate in ecology [Mac55]. His seminal work [May72] motivated the emergence of random matrices as a key mathematical tool for characterizing high dimensional ecosystems. A better understanding of these tools has expanded our understanding of the nature of interactions and food webs to achieve stability [Tay88, AT12, TPA14]. In his study, May was interested in the model (19) [May73], assuming that the system is at a feasible equilibrium $\boldsymbol{x}^*$. According to Theorem 0.20, the study of local stability corresponds to studying the real eigenvalues of the Jacobian matrix of the system at the equilibrium point. The Jacobian matrix of the system (19) is:

$$J = (J_{k\ell})_{n\times n}\,,\ \ J_{k\ell} = \frac{\partial f_k(\boldsymbol{x})}{\partial x_\ell}\,.$$

There exists a matrix $M = J|_{x^*}$, the so-called "community matrix" (Jacobian), describing the effect a species ℓ (column) has on species k (row) around the equilibrium point.

May questioned a central belief in ecology by proving that sufficiently large or complex ecological networks have probability to be stable close to zero. To prove this point, he analyzed the stability of large networks in which species interact at random. In this case, the Jacobian matrix is a non-Hermitian random matrix

$$M = -I + A\,,$$

where A is a centered random matrix $n \times n$ with element $\mathcal{N}(0, \sigma^2)$ with probability C and 0 otherwise (C is called the connectance). For large n, May proved that the probability of stability is close to 0 whenever the "complexity" satisfies:

$$\sigma\sqrt{nC} > 1\,.$$

The eigenvalues of matrix M are distributed according to the circular law, in a disk of center $(-1, 0)$ and radius $\sigma\sqrt{nC}$ (see Figure 17). The stability condition is $\mathcal{R}(\mathrm{Sp}(M)) < 0$. If the model has a large number of highly connected species with strong interactions, then the model is likely to be more unstable.

In a density-dependent model (20), the Jacobian is evaluated as

$$J_{k\ell} = \delta_{k\ell} f_k(\boldsymbol{x}) + x_k \frac{\partial f_k(\boldsymbol{x})}{\partial x_\ell}\,.$$

In the Lotka-Volterra system (21), at a feasible equilibrium x^*, the Jacobian matrix depends on species abundances at equilibrium

$$J|_{\boldsymbol{x}^*} = \mathrm{diag}(\boldsymbol{x}^*)(-I + B)\,,$$

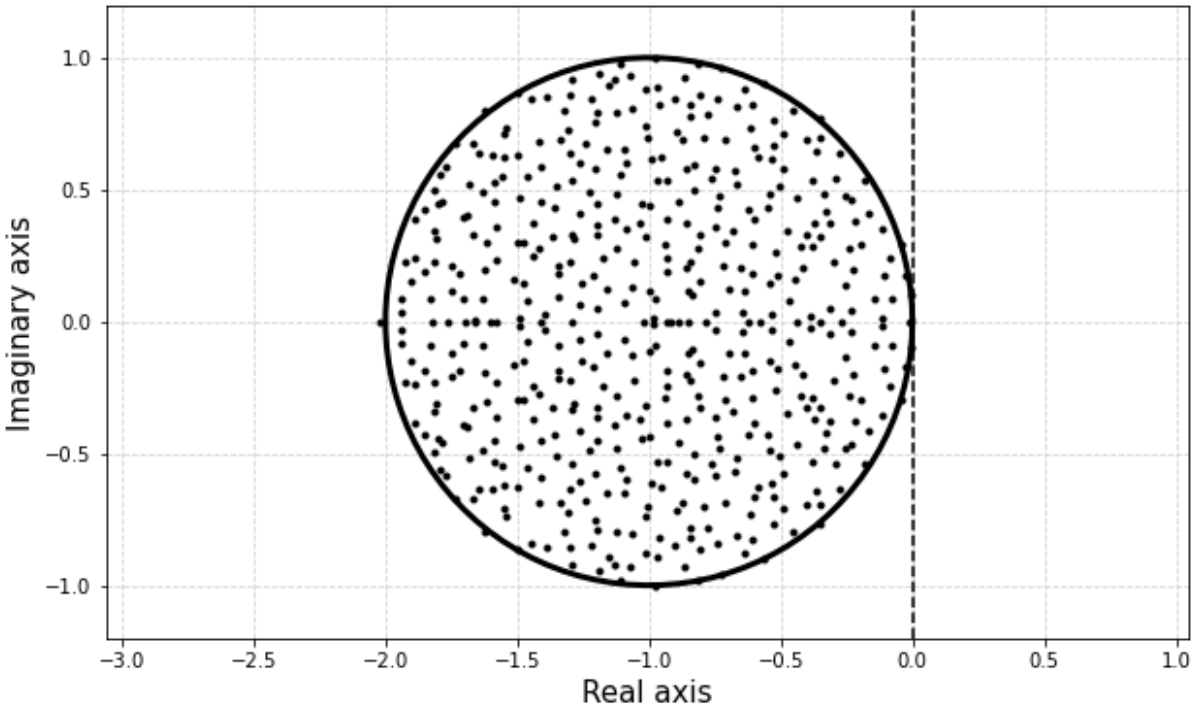


Figure 17: Spectrum of Jacobian random matrix (non-Hermitian matrix) $M = -I + A$ in the complex plan ($n = 500$, $\sigma = 1$, $C = 1$). The solid line circle represents the boundary of the circular law. The dashed line represents the threshold not to be exceeded for the real part of the eigenvalues for the system to be stable.

where B is introduced in (21). The study of this Jacobian is more difficult because there is no independence between $\boldsymbol{x}^*$ and B. However, Stone [Sto18] and Gibbs *et al.* [GGRA18] showed that the intuitions remain similar, the stability of large LV systems is uniquely determined by the interaction matrix.

Final model formulation

In most of the thesis, we focus on the model (21) where $r_k = 1, \forall k \in [n]$, $\theta = 1$ and we add a generic normalization parameter in the interaction matrix. Choosing equal growth rates releases the number of parameters, greatly simplifies the computations and discharge the complexity of the model. Understanding the impact of the interaction matrix B in the LV system comprises many open-problems. However, Rohr *et al.* [RSB14] were concerned on the possible growth rate leading to coexistence. Their work on structural stability has been extended by Cenci *et al.* [CS18], Saavedra *et al.* [SRB+17] and Grilli *et al.* [GAS+17].

The choice to keep the self-regulation term $\theta = 1$ brings more clarity. In a $\theta \neq 1$ case, we can redimension the system (32) to avoid the parameter θ by setting $\tilde{x}_k := \theta x_k$, $\tilde{B}_{k\ell} := B_{k\ell}/\theta$. In fact, Barabás *et al.* [BMSA17] studied the importance of strong self-regulation term for the food web to have a stable equilibrium. Note that the values of the diagonal of the matrix B are not set to 0, but their microscopic values have a negligible impact on

the asymptotic results.

$$\frac{dx_k(t)}{dt} = x_k(t)\left(1 - x_k(t) + \sum_{\ell\in[n]} B_{k\ell}x_\ell(t)\right) . \tag{22}$$

The last important detail that differs from the standard Lotka-Volterra notation is that we assume a normalization parameter $1/\sqrt{n}$ in the matrix B. The major theoretical reason is to limit the impact of the interaction parameters on the other terms and to make it macroscopic i.e.

$$\mathbb{E}\left(\sum_{\ell\in[n]} B_{k\ell}x_\ell(t)\right) \sim O(1) \quad ; \quad \mathrm{Var}\left(\sum_{\ell\in[n]} B_{k\ell}x_\ell(t)\right) \sim O(1)\,.$$

From an ecological point of view, one could imagine that when the number of species increases in an ecosystem, then the strength of the sum of the interactions of one species with all the others will not tend to increase.

Generally, the model can finally be written in the compact form:

$$\frac{dx_k}{dt} = x_k\,\left(1 - x_k + (B\boldsymbol{x})_k\right) \ , \quad k \in [n]\,, \tag{23}$$

where B is yet to be determined.

Network

The major challenge of this thesis is to understand the impact of the interaction matrix B on the dynamics of the Lotka-Volterra model. In nature, B corresponds to the network of interactions between species or considered as the food web of the ecosystem (in the sense: "who eats whom?"). Large and highly connected ecological networks are common in nature [DWM02, PLC91].

In the system (22), a general model for the interaction matrix B is a non-centered random matrix with pairwise correlated interactions combined with a graph structure:

$$B = S \circ \left(\frac{A}{\alpha\sqrt{n}} + \frac{\mu}{n}\mathbf{1}_n\mathbf{1}_n^\top\right) , \tag{24}$$

where $\circ$ is the Hadamard product i.e. $(X \circ Y)_{ij} := X_{ij}Y_{ij}$ and the $n \times 1$ vector $\mathbf{1}_n$ is a vector of ones. $A = (A_{k\ell})_{k,\ell\in[n]}$ is a random matrix satisfying the following conditions

1. $(A_{k\ell}, k \leqslant \ell)$ are independent and identically distributed (i.i.d.) random variables and $\mathbb{E}(A_{k\ell}) = 0$, $\mathbb{E}(|A_{k\ell}|^2) = 1$ and $\mathbb{E}(|A_{k\ell}|^4) < \infty\, \forall\, 1 \leqslant k \leqslant \ell$.

2. for $k < \ell$ the vector $(A_{k\ell}, A_{\ell k})$ has a standard bivariate distribution, independent from the remaining random variables, with covariance $\mathrm{cov}(A_{k\ell}, A_{\ell k}) = \mathbb{E}(A_{k\ell}A_{\ell k}) = \rho$ with $|\rho| \leqslant 1$.

$S := (S_{k\ell})_{k,\ell\in[n]}$ is an adjacency matrix from a graph i.e. if we represent each species by a node and an interaction between two species by a directed vertex then

$$S_{k\ell} = \begin{cases} 1 & \text{if there exists an impact of species } \ell \text{ on } k, \\ 0 & \text{otherwise.} \end{cases}$$

In the absence of any other information, we may assume that S is the adjacency matrix of an Erdös Renyi graph (ER) [ER60]. It is a graph with n vertices. It is assumed that there is an edge between two vertices with probability p independent from every other edge.

From an ecological standpoint, two types of structures can affect the type of interactions and the existence of communities with preferential interactions.

On the one hand, the type of interaction is different between the species. These settings are managed by choices about the statistical properties of the random variables B_{ij}. The parameter set (α, μ, ρ) can represent a range of interaction types. First of all, the interaction strength is represented by α, a large value of α represents a system with weak interactions, conversely a small value of α represents very strong interactions. The parameters μ and ρ describe the nature of the interaction in the system. When $\rho < 0$, interaction partners have opposite impact on one another, i.e. as in antagonistic interactions (the predator being positively influenced by the abundance of it prey while the prey is negatively affected by that of the predator). When $\rho > 0$, interaction partners have similar impact on one another, i.e. they are engaged in either mutualistic or competitive interactions. The mean interaction parameter μ increases the proportion of competitors or mutualists depending of its sign. Given a pairwise interaction $B_{k\ell}/B_{\ell k}$ in the system, the three predominant patterns are:

- competition (-/- relationship), which happens more often when $\rho > 0$, $\mu < 0$ [Mac70, Zee95],
- mutualism (+/+ relationship), which happens more often when $\rho > 0, \mu > 0$ [SGB+15, Sto20],
- predation (+/- relationship), which happens more often when $\rho < 0$, $\mu \approx 0$ [AT12].

There are other types of interactions such as commensalism, amensalism [BTH06].

On the other hand, the structure of the interaction network differs between ecosystems. The structure of the interaction matrix can also be affected by the existence of communities, i.e. groups of species that interact preferentially among them [TF10, AGB+15]. In (22), the network is represented by S an adjacency matrix of a given graph. Several types of structures are widely studied in ecology and be modeled by the graph S.

First, modular structures such as the compartmentalization of food webs, also called modularity, is the tendency of nodes to be connected preferentially within groups than between groups ([GSSP+10, GRA16]).

Second, nested structures where each species plays a different role in the ecosystem. These models are generally called nested because of their structure where some species have more interactions than others [BJMO03, BFPG+09, SKA13, PBHM19].

In large ecosystems, not all species interact with each other, thus the relevance of studying sparse ecosystems is of considerable interest [BSHM17]. In fact, May [May72] considers connectance as a key parameter linked to the complexity of a system.

Last, trait models with latent structures [EJK+13] such as the niche model of Williams and Martinez [WM00], the so called "cascade" model [CBN90] which establishes a predation structure in the food web. Each species can eat on the lower trophic level but not on the subsequent ones [HMS16, PBG22]. Note that all these structures can be modeled with

a graphon (each node is associated with a random variable and the connections depend on a function of the variables associated with the two nodes).

For a general review of the different patterns in complex ecological communities from a physicist's point of view see Barbier *et al.* [BABL18].

Metacommunity and spatial dynamics

At larger spatial scales, ecologists are also interested by interactions between populations (rather than between individuals) in order to understand the patterns of species diversity in space and time. In this context, mutualistic, competitive and predatory interactions are replaced by the processes of colonization, extinction and replacement of whole populations. The theory of spatial ecology finds its roots in the works of MacArthur and Wilson [MW63] and Levins and Heatwole [LH63] and later on by MacArthur on population biology and geography [Mac84]. In particular, *The Theory of Island Biogeography* (TIB) is a fundamental cornerstone of spatial dynamics theory [MW67]. The TIB describes how the biodiversity on islands is maintained by a balance between immigration and extinction of species.

Since then, spatial ecology theory has evolved to understand coexistence mechanisms behind the metacommunity model introduced by Wilson [Wil92]. The study of metacommunity models has been particularly developed due to an awareness of the spatial heterogeneity of ecosystems. Leibold *et al.* [LHM+04] have described the different mechanisms at the spatial scale: colonization (organisms move from one site to another between generations), niche habitat (species may be more or less well adapted to a given environment) and stochasticity (if species are equivalent in terms of traits, competitiveness, etc., we don't necessarily expect them to coexist, but we do know that it will take a certain amount of time before one or the other completely takes over the site, i.e. a neutral model).

Considering that the ability of species to colonize new habitats is crucial to the maintenance of populations, one type of mathematically tractable metacommunity model is known as an occupancy model with a competition-colonization trade-off. The belief associated to this model is the existence of a trade-off between the ability of a species to colonize new patches and its competitive ability, which affects its resistance to colonization by another competing species and its own ability to replace other species. Formally, it is a patch occupancy model where each species has the ability to colonize new patches in competition with other species. The variable of interest is the proportion of habitat occupied by each species.

Initially studied by Levins [Lev69] and Levins and Culver [LC71] in the case of a two species, this model has been of special interest in its n-species version where the competition is hierarchical [Has80, NM92, Til94]. In a more general case, it has been studied when the competition is not hierarchical [Ama03, YW01, CMJD06a] and also in an epidemiological context of a dynamic of host-parasite interactions [MN94, NM94]. In a more general framework, random matrix approaches have been used to study stability in a meta-ecosystem context. Each patch has its own dynamics and dispersal of all species connects the different patches, see Gravel *et al.* [GML16].

The spatial dynamics of an n-species system in a competition colonization trade-off

[CMJD06a] is of the form

$$\frac{dp_k}{dt} = c_k p_k \left(1 - \sum_{\ell=1}^{n} p_\ell\right) - m_k p_k + c_k p_k \sum_{\ell \neq k} p_\ell \eta_{k\ell} - p_k \sum_{\ell \neq k} c_\ell p_\ell \eta_{\ell k}\,, \tag{25}$$

where p_k represents the occupancy of species k, m_k is the extinction rate of species k, c_k represents the colonization rate of species k, $\eta_{k\ell}$ corresponds to the probability of replacement of species ℓ by k.

These equations can be represented as a Lotka Volterra competition model with asymmetric interactions

$$\frac{dp_k}{dt} = p_k \left[c_k - m_k + \sum_{\ell=1}^{n} p_\ell \left(c_k \eta_{k\ell} - c_\ell \eta_{\ell k} - c_k\right) \right].$$

In the context of the Lotka-Volterra model with dispersal, further work has been done by introducing a migration parameter [BG20, PNJ21, VPNJ22]. In the context of meta-food web with diffusion parameters, Brechtel *et al.* [BGR+18] studied diffusion-driven pattern formation in networks.

Guided tour on random matrices

Historically, the theory of random matrices has its roots in the work of the statistician John Wishart whose purpose was to study a random matrix of empirical covariance of multivariate Gaussian samples [Wis28]. Subsequently, in the 50's, a second impulse was given by Eugene Wigner [Wig55] whose aim was to explain the distribution of energy levels in atomic nuclei. The innovative approach used by Wigner [Wig67] to describe the spectrum of a Hermitian random matrix was taken up by other physicists to solve problems in nuclear physics [Dys62] and physical sciences. Later on, new matrix structures were studied, many works were done by Marchenko and Pastur [MP67] on large covariance matrices and Girko [Gir85], Bai [Bai97] and Silverstein [SC95, BS10] extended results to non-hermitian matrices. Until today, a multitude of works have been published in very diverse fields of mathematics such as combinatorics, random graphs, free probability theory, signal theory, number theory, etc.

The strength of random matrix theory comes from the stabilization of its spectrum (random and a priori complicated) when the dimension of the matrix tends towards infinity. Within this framework, the distribution of the eigenvalues of the matrix becomes completely deterministic. In a very simplified way, this is an equivalence of the law of large numbers for the spectrum of a matrix. The stakes and motivations of the theory of random matrices are based on the description of the standard properties of the spectrum of matrices: eigenvalues, eigenvectors, largest eigenvalue, etc. It is a balanced mixture of linear algebra, probability, complex analysis, combinatorics.

A few definitions

Let $A \in \mathcal{M}(\mathbb{C})$, $A := (A_{k\ell})_{n\times n}$, a square matrix of size n with coefficient in the set of complex numbers $\mathbb{C}$. We denote by $A^* := \overline{A}^\top$. Given a vector $\boldsymbol{x} \in \mathbb{R}^n$, we note $\|x\|_2$ its

Euclidean norm:

$$\|\boldsymbol{x}\|_2 = \left(\sum_{k=1}^{n} |x_k|^2\right)^{1/2} .$$

Definition 0.15 (Eigenvalues)**.** Denote by $\lambda_1(A)$, $\lambda_2(A)$, ..., $\lambda_n(A)$ the eigenvalues of A i.e. the roots of its characteristic polynomial, such that

$$|\lambda_1(A)| \geqslant \cdots \geqslant |\lambda_n(A)| .$$

The set of eigenvalues of A is called the spectrum of A and noted $\mathrm{Sp}(A)$.

Definition 0.16 (Spectral radius)**.** The spectral radius of the matrix A, that we note $\rho(A) = |\lambda_1(A)|$, is the modulus of the eigenvalue with the largest modulus.

Definition 0.17 (Singular values)**.** The singular values $\sigma_1(A)$, $\sigma_2(A)$, ..., $\sigma_n(A)$ of the matrix A are the square root of the eigenvalues of the Hermitian matrix A^*A i.e.

$$\sigma_i(A) := \sqrt{\lambda_i(A^*A)} , \quad \forall i \in [n] .$$

Definition 0.18 (Spectral norm)**.** The spectral norm of the matrix A denoted by $\|A\|$ is defined by its largest singular value

$$\|A\| := \max\left(\sqrt{\lambda}, \lambda \text{ eigenvalue of } A^*A\right) = \sigma_1(A) .$$

In probability, the spectral measure characterizes the spectrum of a matrix. In the RMT field, it is used to express results of convergence of the spectrum to a deterministic measure. Given $I \subset \mathbb{C}$, denote by δ_λ the Dirac measure at the point λ defined by

$$\delta_\lambda(I) = \begin{cases} 1 \text{ if } \lambda \in I , \\ 0 \text{ otherwise } . \end{cases}$$

Definition 0.19 (Empirical spectral measure)**.** If $A \in \mathcal{M}_n(\mathbb{C})$ with eigenvalues $\lambda_1(A)$, ..., $\lambda_n(A)$, we define the empirical measure of eigenvalues in $(\mathbb{C}, \mathcal{B}(\mathbb{C}))$ by

$$\mu_A := \frac{1}{n}\sum_{k=1}^{n} \delta_{\lambda_k(A)}.$$

For every subset $E \subset \mathbb{C}$, the quantity:

$$\mu_A(E) = \frac{\mathrm{card}\{1 \leqslant k \leqslant n : \lambda_k(A) \in E\}}{n} ,$$

is the proportion of eigenvalues of A in E.

The weak convergence of a empirical spectral measure to a deterministic measure describes many random matrix results.

Definition 0.20 (Weak convergence)**.** It is said that μ_A converges weakly to a probability measure μ i.e. $\mu_A \xrightarrow[n\to\infty]{\mathcal{D}} \mu$, if for any function f continuous and bounded on $\mathbb{R}$

$$\int f(u)\mu_A(du) = \frac{1}{n}\sum_{k=1}^{n} f(\lambda_k) \xrightarrow[n\to\infty]{} \int f(u)\mu(du) .$$

Remark 0.9. If A is random, then μ_A is a random discrete probability distribution, this implies $\int f\mu_A(du)$ are also random variables. We will then say that almost surely (a.s.) μ_A converges weakly to μ

$$(a.s.) \quad \mu_A \xrightarrow[n\to\infty]{\mathcal{D}} \mu\,.$$

Definition 0.21 (Resolvent)**.** Let $A \in \mathcal{M}_n(\mathbb{C})$, we call the resolvent of A the matrix $Q := (Q_{k\ell})_{n\times n}$ defined by

$$Q(z) = (A - zI)^{-1},\ z \notin \mathrm{Sp}(A)\,.$$

We denote by

$$\mathbb{C}^+ := \{z \in \mathbb{C} \,:\, \mathrm{Im}(z) > 0\}$$

the upper half of the complex plane.

Definition 0.22 (Stieltjes transform)**.** Let $\mu \in \mathcal{P}(\mathbb{R})$ a probability measure. The Stieltjes transform of μ denoted by $g_\mu : \mathbb{C}^+ \to \mathbb{C}$ is defined by

$$g_\mu(z) = \int \frac{1}{\lambda - z}\mu(d\lambda)\,, z \in \mathbb{C}^+\,.$$

Remark 0.10. Let μ_A the empirical measure of the eigenvalues $\lambda_1(A), \cdots, \lambda_n(A)$ of the symmetric matrix A, then the associated Stieltjes transform is given by

$$g_{\mu_A}(z) = \int \frac{1}{\lambda - z}\mu_A(d\lambda) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\lambda_i - z} = \frac{1}{n}\mathrm{Tr}\left((A - zI)^{-1}\right)\,,$$

where $Q = (A - zI)^{-1}$ is the resolvent of the matrix A and $\mathrm{Tr}(Q)$ is the trace of matrix Q.

Proposition 0.21 (Stieltjes inversion)**.** *Let g_μ the Stieltjes transform of the measure μ of finite mass $\mu(\mathbb{R})$. If $a, b \in \mathbb{R}$ and $\mu(\{a\}) = \mu(\{b\}) = 0$, then*

$$\mu(a, b) = \frac{1}{\pi}\lim_{y\to 0^+}\mathrm{Im}\int_a^b g_\mu(x + iy)dx\,,$$

and

$$\forall x \in \mathbb{R},\ \mu(\{x\}) = \frac{1}{\pi}\lim_{y\to 0^+}\mathrm{Im}(g_\mu(x + iy))\,.$$

Proposition 0.22 (Woodbury identity)**.** *Let A be a matrix $n \times n$, U a matrix $n \times m$, B a matrix $m \times m$, V a matrix $m \times n$. It is assumed that all the considered matrix inverses exist, then:*

$$(A + UBV)^{-1} = A^{-1} - A^{-1}U(B^{-1} + VA^{-1}U)^{-1}VA^{-1}\,.$$

The Woodbury identity for a rank 1 perturbation is often used and referred to as Sherman-Morrison identity.

Proposition 0.23 (Sherman-Morrison identity)**.** *Let A a matrix $n \times n$ and u, v two vectors of dimension n. It is assumed that all the considered matrix inverses exist, then:*

$$(A + uv^*)^{-1} = A^{-1} - \frac{A^{-1}uv^*A^{-1}}{1 + v^*A^{-1}u}\,.$$

Proposition 0.24 (Poincaré inequality)**.** *A probability measure $\mathbb{P}$ on $\mathbb{R}^n$ satisfies a Poincaré inequality with constant $c > 0$ if, for all continuously differentiable functions $f : \mathbb{R}^n \to \mathbb{C}$,*

$$\mathrm{Var}_{\mathbb{P}}(f) = \mathbb{E}_{\mathbb{P}}(|f(x) - \mathbb{E}_{\mathbb{P}}(f(x))|^2) \leqslant \frac{1}{c}\mathbb{E}_{\mathbb{P}}|\nabla f(x)|^2\,.$$

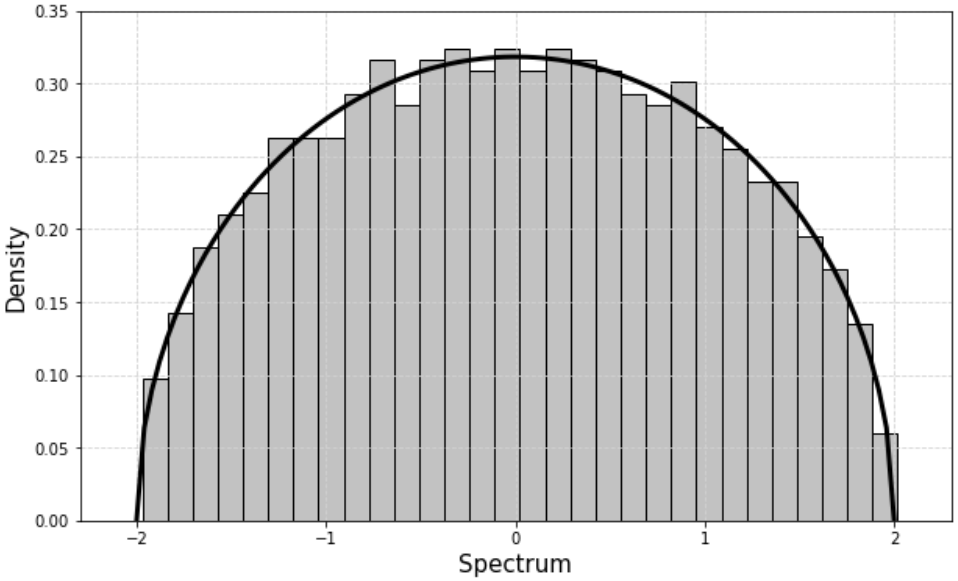


Figure 18: Spectrum (histogram) of the Wigner random matrix $W_n/\sqrt{n}$ ($n = 1000$, $\sigma = 1$). The solid line represents the semi-circular law.

Wigner matrices

Definition 0.23. Given W_n a hermitian matrix $n \times n$, $W_n = W_n^*$ such that $W_n := (W_{k\ell}, 1 \leqslant k \leqslant \ell \leqslant n)$ are i.i.d. random variables with $\mathbb{E}(W_{k\ell}) = 0$, $\forall\, 1 \leqslant k \leqslant \ell$ and $\mathbb{E}(|W_{k\ell}|^2) < \infty$, $\forall\, 1 \leqslant k \leqslant \ell$. $W_n/\sqrt{n}$ is called Wigner matrix.

Theorem 0.25 (Universality of the Wigner theorem and semi-circular law)**.** *Let W_n a Wigner matrix defined by $W_n := (W_{k\ell}, 1 \leqslant k \leqslant \ell)$ i.i.d. random variables such that*

1. $\mathbb{E}(W_{k\ell}) = 0$, $\forall\, 1 \leqslant k \leqslant \ell$,
2. $\mathbb{E}(|W_{k\ell}|^2) = \sigma^2 < \infty$, $\forall\, 1 \leqslant k \leqslant \ell$ *and* $\sigma > 0$.

Then almost surely, the empirical spectral measure of $W_n/\sqrt{n}$ converges weakly to the semi-circular law:

$$(a.s.) \quad \mu_{\frac{W_n}{\sqrt{n}}} := \frac{1}{n}\sum_{k=1}^{n} \delta_{\lambda_k\left(\frac{W_n}{\sqrt{n}}\right)} \xrightarrow[n\to\infty]{\mathcal{D}} \mu_{sc}\,,$$

where μ_{sc} is defined by

$$d\mu_{sc}(t) = \frac{1}{2\pi\sigma^2}\sqrt{(4\sigma^2 - t^2)}\mathbf{1}_{[-2\sigma,2\sigma]}(t)dt\,.$$

The eigenvalues of the matrix $W_n/\sqrt{n}$ are real. In Figure 18, a histogram of the eigenvalues of a Wigner random matrix is illustrated compared to the theoretical distribution given by Theorem 0.25.

Local spectrum behavior for Wigner matrices

Denote by

$$\lambda_{\max}(W_n) = \max_{k\in[n]} \lambda_k(W_n) \text{ and } \lambda_{\min}(W_n) = \min_{k\in[n]} \lambda_k(W_n)\,.$$

To address certain issues, it is necessary to have exact information on the position of the largest eigenvalue of the matrix. In the case of Wigner matrices, many works have been done and refined in the 80' [FK81, BY88].

Theorem 0.26 (Convergence of extremes eigenvalues)**.** *If* $\mathbb{E}(|W_{k\ell}|^4) < \infty$, $\forall\, 1 \leqslant k \leqslant \ell$, *then*

$$\lambda_{\max}(W_n/\sqrt{n}) \xrightarrow[n\to\infty]{a.s.} 2\sigma \ , \ \lambda_{\min}(W_n/\sqrt{n}) \xrightarrow[n\to\infty]{a.s.} -2\sigma \, .$$

In particular,

$$\left\|W_n/\sqrt{n}\right\| = \max(|\lambda_{\max}(W_n/\sqrt{n})|, |\lambda_{\min}(W_n/\sqrt{n})|) \xrightarrow[n\to\infty]{a.s.} 2\sigma \, .$$

If $\mathbb{E}(|W_{k\ell}|^4) = \infty$, $\forall\, 1 \leqslant k \leqslant \ell$, *then*

$$\lambda_{\max}(W_n/\sqrt{n}) \xrightarrow[n\to\infty]{a.s.} +\infty \, .$$

Deformed Wigner matrix

The last specific property to be addressed for this thesis is the distribution of eigenvalues when the Wigner matrix is perturbed by a finite rank deformation. This type of model is frequently called "spike". Depending on the type of deformation, some spike eigenvalues may escape from the bulk of the distribution.

Let W_n be a random Wigner matrix and

1. $\mathbb{E}(W_{k\ell}) = 0$, $\forall\, 1 \leqslant k \leqslant \ell$,
2. $\mathbb{E}(|W_{k\ell}|^2) = \sigma^2 < \infty$, $\forall\, 1 \leqslant k \leqslant \ell$ and $\sigma > 0$,
3. $\sup\limits_{k\neq\ell} \mathbb{E}[|W_{k\ell}|^4] < \infty$.

Let P_n a deterministic real symmetric matrix of fixed rank r. We are interested in the properties of the spectrum of the matrix $\frac{1}{\sqrt{n}}W_n + P_n$.

Since the pioneering paper of Füredi and Komlòs [FK81], many scientists have studied the spectral properties of deformed Wigner matrices [Pé06, CDMF09, PRS13, RS13]. Denote by $\theta_1, ..., \theta_r$ the ordered eigenvalues of P_n, θ_j has multiplicity of k_j and they are independent of n. Let r_0, the index associated to the threshold 0 i.e. $\theta_{r_0} = 0$ and P_n has $r_0 - 1$ distinct positive eigenvalues. Let $r_{+\sigma}$ (resp $r_{-\sigma}$) be the number of j such that $\theta_j > \sigma$ (resp $\theta_j < -\sigma$).

Theorem 0.27 (Deformed Wigner Theorem - [CDMF09, PRS13])**.** *Let* W_n *be a random real Wigner matrix satisfying condition (1)-(3) and* P_n *be a deterministic real hermitian matrix of fixed finite rank* r *as above. Let*

$$\rho_{\theta_j} = \theta_j + \frac{\sigma^2}{\theta_j} \, .$$

Then the following holds:

1. *For* $1 \leqslant j \leqslant r_{+\sigma}, 1 \leqslant i \leqslant k_j, \lambda_{k_1+...+k_{j-1}+i} \to \rho_{\theta_j}$,
2. $\lambda_{k_1+...+k_{r_{+\sigma}}+1} \to 2\sigma$,

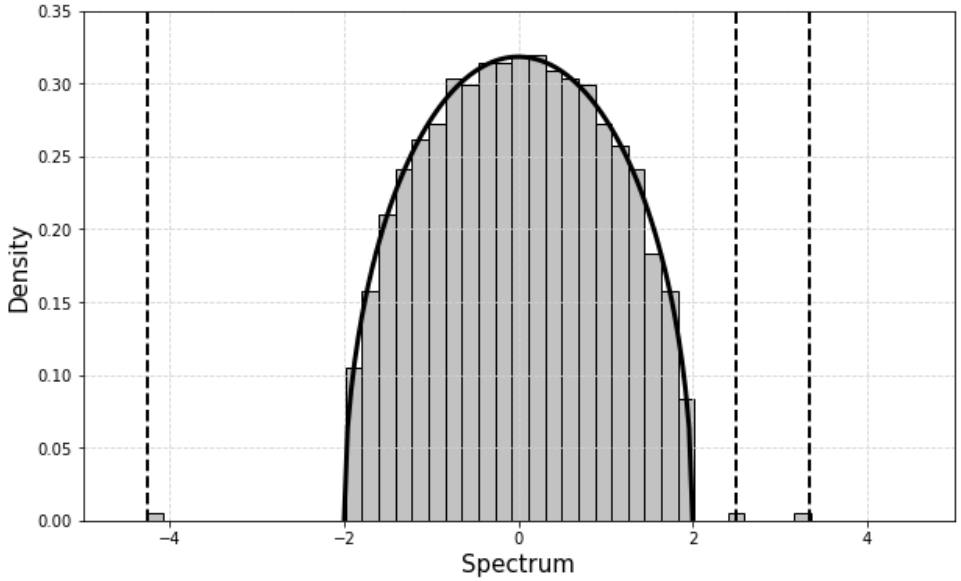


Figure 19: Spectrum (histogram) of the deformed Wigner random matrix $W_n/\sqrt{n} + P_n$ ($n = 1000$, $\sigma = 1$). The perturbed matrix is $P_n = \mathrm{diag}(-4, 2, 3, 0, ..., 0)$. The solid line represents the semi-circular law. The dashed lines indicate the theoretical value of the outliers at $-4 - 1/4, 2 + 1/2, 3 + 1/3$ as predicted by Theorem 0.27.

3. $\lambda_{k_1+...+k_{r-r_{-\sigma}}} \to -2\sigma$,

4. *For* $j \geqslant r - r_{-\sigma} + 1, 1 \leqslant i \leqslant k_j, \lambda_{k_1+...+k_{j-1}+i} \to \rho_{\theta_j}$.

The convergence in (1)-(4) is in probability.

Remark 0.11. In [CDMF09], Capitaine, Donati-Martin and Féral show that if the entries distribution satisfy a Poincaré inequality (proposition 0.24), the convergence in Theorem 0.27 holds almost surely.

In Figure 19, a histogram of the eigenvalues of the deformed Wigner random matrix is illustrated compared to the theoretical distribution given by theorem 0.27.

Circular law

In a second part of the tour, we focus on non-Hermitian matrices. Let $Y_n \in \mathcal{M}_n(\mathbb{C})$ be a square random matrix of dimension $n \times n$ whose entries are i.i.d. centered of variance σ^2. The eigenvalues of Y_n are no longer real but complex. The major outcome concerns the convergence of the empirical spectral measure of $Y_n/\sqrt{n}$ toward the circular law in the complex plane. Initially proved by Mehta [Meh67] for the expected empirical spectral distribution in the complex Gaussian case as a result of Ginibre's work [Gin65] of the explicit formula for the spectrum. Edelman [Ede97] established the circular law in the case of real Gaussian random variables. Silverstein gave an argument to pass from expected convergence to almost surely convergence. Girko worked in the universal version (for other types of distribution) [Gir85] by providing some insights of proof such as the Hermitization technique. However, it is finally Tao and Vu [TVK10] who proved the general case. I advice the reader to look at Bordenave and Chafaï [BC12].

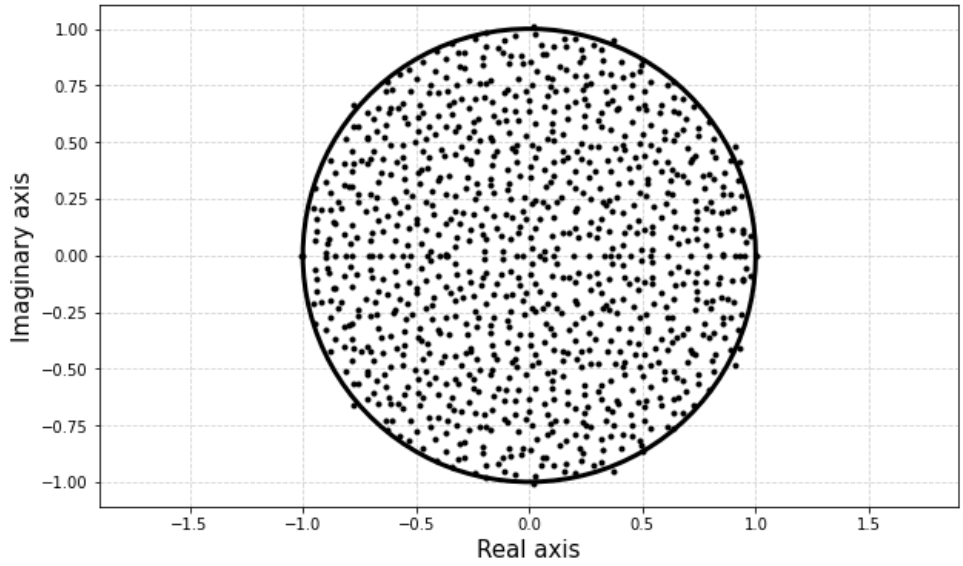


Figure 20: Spectrum of a non-Hermitian random matrix $Y_n/\sqrt{n}$ in the complex plan ($n = 1000$, $\sigma = 1$). The solid line circle represents the boundary of the circular law.

Theorem 0.28. *Let Y_n be a random matrix $\mathcal{M}_n(\mathbb{C})$ such that $Y_n := (Y_{k\ell}, 1 \leqslant k, \ell \leqslant n)$ are i.i.d. random variables such that $\mathbb{E}(Y_{k\ell}) = 0$, $\forall\, 1 \leqslant k, \ell \leqslant n$ and $\mathbb{E}(|Y_{k\ell}|^2) = \sigma^2$, $\forall\, 1 \leqslant k, \ell \leqslant n$. Then almost surely, the empirical spectral measure of $Y_n/\sqrt{n}$ converges weakly to the circular law.*

$$\mu_{\frac{Y_n}{\sqrt{n}}} \xrightarrow[n\to\infty]{\mathcal{D}} \mu_c\,,$$

where μ_c is the circular law which is the uniform law on the disc of radius σ of $\mathbb{C}$ with density

$$d\mu_c(z) = \frac{1}{\pi\sigma^2}\mathbf{1}_{z\in\mathbb{C},|z|\leqslant\sigma}dz\,.$$

In Figure 20, the eigenvalues of a non-Hermitian random matrix in the complex plan are illustrated compared to the theoretical distribution given by Theorem 0.28.

Local spectrum behavior for non-Hermitian matrix

In the case of the circle law, it is important to have information on the position of the spectral radius. We have seen previously in the work of May [May72], the necessity to describe the stability transition by the largest real part eigenvalue of the Jacobian matrix. Many works have been done on this subject, in particular by Bai [BSY88, BS10].

Theorem 0.29 (Convergence of extremes eigenvalues)**.** *If $\mathbb{E}(Y_{k\ell}) = 0$ and $\mathbb{E}(|Y_{k\ell}|^4) < \infty$, $\forall\, 1 \leqslant k, \ell \leqslant n$, then,*

$$\left\|\frac{Y_n}{\sqrt{n}}\right\| \xrightarrow[n\to\infty]{a.s.} 2\sigma \quad \textit{and} \quad \rho\left(\frac{Y_n}{\sqrt{n}}\right) \xrightarrow[n\to\infty]{a.s.} \sigma\,.$$

Outliers in the spectrum of non-Hermitian matrix

As in the Wigner case, we can consider a finite rank perturbation of the non-Hermitian matrix. This result has been proved by Tao [Tao13, Theorem 1.7] and has been extended

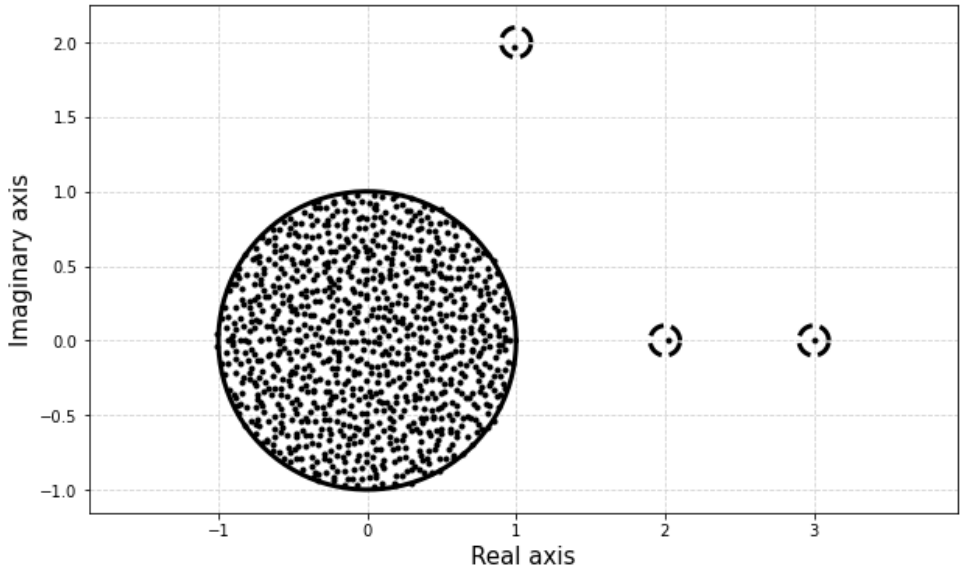


Figure 21: Spectrum of a non-Hermitian random matrix $Y_n/\sqrt{n} + P_n$ in the complex plan ($n = 1000$, $\sigma = 1$). The perturbed matrix is $P_n = \text{diag}(1 + 2i, 2, 3, 0, ..., 0)$. The solid line circle represents the boundary of the circular law. There are three eigenvalues in the small dashed circles centered at $1 + 2i$, 2, 3 as predicted by Theorem 0.30.

in Benaych-Georges and Rochet [BGR16] where they studied the fluctuations of the outlier eigenvalues.

Theorem 0.30 (Deformed spectrum of non-Hermitian matrix)**.** *Let Y_n a i.i.d. random matrix with $\mathbb{E}(Y_{k\ell}) = 0$, $\mathbb{E}(|Y_{k\ell}|^2) = 1$ and $\mathbb{E}(|Y_{k\ell}|^4) < \infty$, $\forall\, 1 \leqslant k, \ell \leqslant n$ and for each n, let P_n be a deterministic matrix with rank $O(1)$ and operator norm $O(1)$. Let $\varepsilon > 0$, and suppose that for all sufficiently large n, there are no eigenvalues of P_n in the annulus $\{z \in \mathbb{C} : 1 + \varepsilon < |z| < 1 + 3\varepsilon\}$, and there are j eigenvalues $\lambda_1(P_n), ..., \lambda_j(P_n)$ for some $j = O(1)$ in the region $\{z \in \mathbb{C} : |z| \geqslant 1 + 3\varepsilon\}$.*

Then, a.s., for sufficiently large n, there are precisely j eigenvalues

$$\lambda_1\left(\frac{Y_n}{\sqrt{n}} + P_n\right), ..., \lambda_j\left(\frac{Y_n}{\sqrt{n}} + P_n\right),$$

of $\frac{Y_n}{\sqrt{n}} + P_n$ in the region $\{z \in \mathbb{C} : |z| \geqslant 1 + 2\varepsilon\}$, and after labeling these eigenvalues properly, $\lambda_i(\frac{Y_n}{\sqrt{n}} + P_n) = \lambda_i(P_n) + o(1)$ as $n \to \infty$ for each $1 \leqslant i \leqslant j$.

In Figure 21, the eigenvalues of a deformed non-Hermitian random matrix with outliers are illustrated compared to the theoretical distribution given by theorem 0.30.

Elliptic model

In the Wigner matrix configuration, the interaction of one species on the other is considered to be the same. For non-Hermitian matrices, all interactions are independent. However, in ecology, the reciprocal effects of a species k on another species ℓ ($X_{k\ell} \leftrightarrow X_{\ell k}$) are linked. Mathematically, we consider a pairwise correlation between the entries of the matrix. This can be used to describe biological processes such as predation when the sign

of the interactions is reversed and the correlation is negative. In random matrix, when the pairwise interactions are drawn from a bi-variate distribution, we are in the framework of the elliptic model. Originally introduced by Girko [Gir86], this model has since been widely studied [Gir95, Nau12, NO15, OR14].

Definition 0.24 (Random elliptic model)**.** Let X_n be a real random matrix satisfying the following three conditions:

1. Pairs $(X_{k\ell}, X_{\ell k})$, $k \neq \ell$ are i.i.d. random variables with
$$\forall\, k \neq \ell,\ \mathbb{E}(X_{k\ell}) = 0,\ \mathbb{E}(|X_{k\ell}|^2) = 1 \text{ and } \mathbb{E}(|X_{k\ell}|^4) < \infty\,.$$
2. For $k < \ell$ the vector $(X_{k\ell}, X_{\ell k})$ is sample from a bivariate distribution, independent from the remaining random variables, with covariance $\mathbb{E}(X_{k\ell}X_{\ell k}) = \rho$ with $|\rho| \leqslant 1$.
3. $(X_{kk}, 1 \leqslant k \leqslant n)$ are i.i.d. random variables, independent of off-diagonal entries with $\mathbb{E}(X_{kk}) = 0$ and $\mathbb{E}(|X_{kk}|^2) = 1$.

For $\rho \in (-1, 1)$, define the ellipsoid
$$\mathcal{E}_\rho := \left\{ z = x + \boldsymbol{i}y \in \mathbb{C} : \frac{x^2}{(1+\rho)^2} + \frac{y^2}{(1-\rho)^2} \leqslant 1 \right\}\,.$$

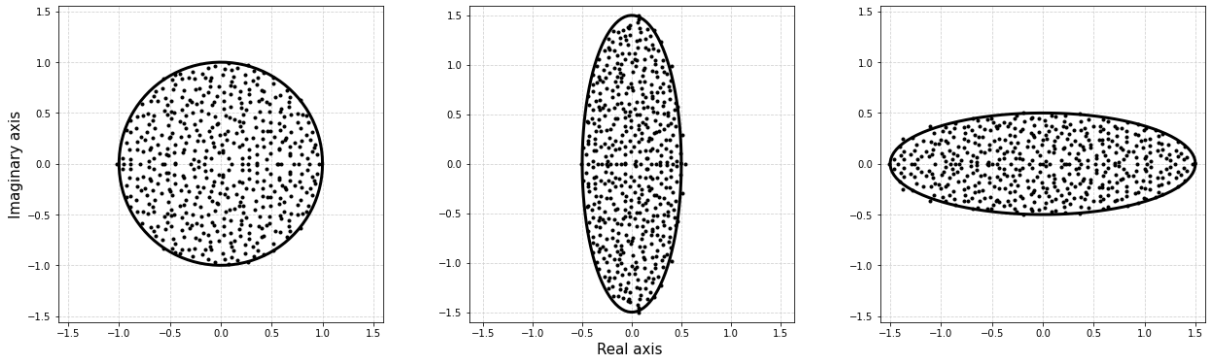
Remark 0.12.
1. For $\rho = 1$, $\mathcal{E}_1$ is the line segment $[-2, 2]$ on the real axis and for $\rho = -1$, $\mathcal{E}_{-1}$ is the line segment $[-2, 2]$ on the imaginary axis.
2. If $\rho = 1$, X_n is a Wigner matrix.
3. If $\rho = 0$, X_n is a non-Hermitian matrix i.e. defined by Theorem 0.28.

Theorem 0.31 (Elliptic law)**.** *Let X_n an elliptic random matrix satisfying conditions in definition 0.24. Then almost surely, the empirical spectral measure of $X_n/\sqrt{n}$ converges weakly to the elliptic law:*
$$(a.s.) \quad \mu_{\frac{X_n}{\sqrt{n}}} \xrightarrow[n\to\infty]{\mathcal{D}} \mu_\rho\,,$$
where μ_ρ is the uniform probability measure on the ellipsoid $\mathcal{E}_\rho$ with density
$$\mu_\rho(z) = \begin{cases} \frac{1}{\pi(1-\rho^2)} & \text{if } z \in \mathcal{E}_\rho\,, \\ 0 & \text{otherwise}\ . \end{cases}$$

In Figure 22, the eigenvalues of an elliptic random matrix in the complex plan are represented compared to the theoretical distribution given by theorem 0.31.

Corollary 2.3 in O'Rourke and Renfrew [OR14] provides information about the spectral radius of an elliptic matrix.

Proposition 0.32 (Spectral radius of elliptic random matrix)**.** *Let X_n a elliptic random matrix defined in definition 0.24, then*
$$\rho\left(\frac{X_n}{\sqrt{n}}\right) \xrightarrow[n\to\infty]{a.s.} 1 + |\rho|\,.$$



(a) $\rho = 0$ (b) $\rho = -0.5$ (c) $\rho = 0.5$

Figure 22: Spectrum of an elliptic matrix X_n ($n = 500$) with distinct parameters $\rho \in \{-0.5,\ 0,\ 0.5\}$. The solid line represents the ellipse $\{z = x + \boldsymbol{i}y \in \mathbb{C},\ \frac{x^2}{(1+\rho)^2} + \frac{y^2}{(1-\rho)^2} = 1\}$ which is the boundary of the support of the limiting spectral distribution for an elliptic model.

Outliers in the elliptic model

The study of outliers in the case of a deformed elliptic random matrix has been done by O'Rourke and Renfrew [OR14].

We define the neighborhoods for any $\delta > 0$

$$\mathcal{E}_{\rho,\delta} := \{z \in \mathbb{C} : \mathrm{dist}(z, \mathcal{E}_\rho) \leqslant \delta\}\ .$$

Theorem 0.33 (Deformed elliptic random matrix)**.** *Let $k \geqslant 1$ and $\delta > 0$. Let X_n be an elliptic random matrix defined in definition 0.24. Let P_n a deterministic matrix $n \times n$ of finite rank k and $\sup_n \|P_n\| = O(1)$. Suppose for n sufficiently large, there are no nonzero eigenvalues of P_n which satisfy*

$$\lambda_i(P_n) + \frac{\rho}{\lambda_i(P_n)} \in \mathcal{E}_{\rho,3\delta} \backslash \mathcal{E}_{\rho,\delta} \ \textit{with}\ |\lambda_i(P_n)| > 1,$$

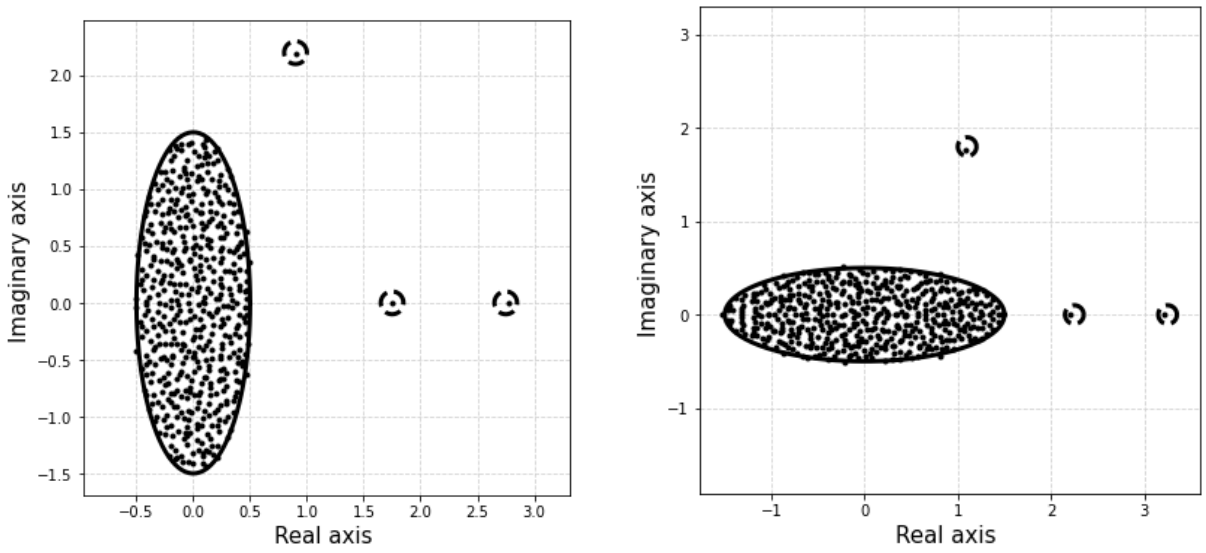
and there are j eigenvalues $\lambda_1(P_n), ..., \lambda_j(P_n)$ for some $j \leqslant k$ which satisfy

$$\lambda_i(P_n) + \frac{\rho}{\lambda_i(P_n)} \in \mathbb{C} \backslash \mathcal{E}_{\rho,3\delta} \ \textit{with}\ |\lambda_i(P_n)| > 1\,.$$

Then, a.s., for n sufficiently large, there are exactly j eigenvalues of $\frac{1}{\sqrt{n}} X_n + P_n$ in the region $\mathbb{C} \backslash \mathcal{E}_{\rho,2\delta}$ and after labeling the eigenvalues properly,

$$\lambda_i\left(\frac{X_n}{\sqrt{n}} + P_n\right) = \lambda_i(P_n) + \frac{\rho}{\lambda_i(P_n)} + o(1), \forall\, 1 \leqslant i \leqslant j\,.$$

In Figure 23, the eigenvalues of a deformed elliptic random matrix in the complex plan are illustrated compared to the theoretical distribution given by theorem 0.33.



(a) $\rho = -0.5$ (b) $\rho = 0.5$

Figure 23: Spectrum of a deformed elliptic matrix $X_n + P_n$ ($n = 500$) with distinct parameter $\rho \in \{-0.5,\ 0.5\}$. The perturbed matrix is $P_n = \mathrm{diag}(1 + 2i, 2, 3, 0, ..., 0)$. The solid line represents the ellipse $\{z = x + \boldsymbol{i}y \in \mathbb{C},\ \frac{x^2}{(1+\rho)^2} + \frac{y^2}{(1-\rho)^2} = 1\}$ which is the boundary of the support of the limiting spectral distribution for an elliptic model. There is three eigenvalues in the small dashed circles centered at $1 + \frac{\rho}{5} + (2 - \frac{2\rho}{5})i$, $2 + \frac{\rho}{2}$, $3 + \frac{\rho}{3}$ as predicted by Theorem 0.33

Theoretical background of the thesis

Understanding the equilibrium points of the Lotka-Volterra system (23) and their stability provides a better understanding of the impact of the food web, represented by the interaction matrix B, on the abundances of the species. In particular, the food web has an impact on the persistence of its component species (:= number of persisting species), the feasibility of the system (i.e. whether there exists an equilibrium with all species at non-zero abundances) and the stability of the equilibrium. Recall (23),

$$\frac{dx_k}{dt} = x_k\ (1 - x_k + (B\boldsymbol{x})_k)\ \ , \quad k \in [n]$$

an essential feature to understand the dynamics of the LV system (23) is the existence of an equilibrium $\boldsymbol{x}^* = (x_k^*)_{k\in[n]}$ such that

$$\begin{cases} x_k^*\,(1 - x_k^* + (B\boldsymbol{x}^*)_k) = 0\,, \quad \forall k \in [n]\,, \\ x_k^* \geqslant 0. \end{cases} \tag{26}$$

A natural question is whether an equilibrium exists and whether it is unique. If so, a further consideration is whether the system converges to this equilibrium i.e. the convergence of a solution $\boldsymbol{x}$ to the equilibrium $\boldsymbol{x}^*$: $\boldsymbol{x}(t) \xrightarrow[t\to\infty]{} \boldsymbol{x}^*$ if $\boldsymbol{x}(0)$ is sufficiently close to $\boldsymbol{x}^*$. The last step is to describe the stability: local, global, resilience (:= ability of an system to regain its initial structure following a perturbation), etc. The Lotka-Volterra system is forward invariant i.e. $\boldsymbol{x}(0) > 0$ (componentwise) implies $\boldsymbol{x}(t) > 0$ for every

$t > 0$. However, some of these components $x_k(t)$ may converge to zero if the equilibrium $\boldsymbol{x}^*$ has zero components.

Feasibility

The question of the feasibility of a equilibrium $\boldsymbol{x}^* > 0$ had already been addressed by Goh [GJ77] for the Lotka-Volterra model in the 70's', then Logofet was interested in this problem in the case of a competitive system [Log93]. Rossberg also investigated the mean number of species that can coexist in competitive communities [Ros13]. Recently, the work of Grilli *et al.* [GAS+17], using structural stability methods introduced by Rohr *et al.* [RSB14], studied the impact of food web properties on the growth rate to maintain a feasible equilibrium.

From (24), we consider a matrix of interactions with Gaussian entries

$$B_n = \frac{1}{\alpha_n \sqrt{n}} A_n,$$

where $A_n := (A_{k\ell}, 1 \leqslant k, \ell \leqslant n)$ are i.i.d. random variables with $A_{k\ell} \sim \mathcal{N}(0,1)$.

Based on the work by Geman and Hwang [GH82], Dougoud *et al.*[DVR+18] have shown that in the framework of random entries of the interaction matrix, if $\alpha_n > 0$ is fixed and independent of n then necessarily some species will become extinct. The threshold of existence of a feasible equilibrium of the model (23) has been studied by Bizeul and Najim [BN21]. In their paper, they also show that feasibility involves stability. This type of result had already been observed by Stone [Sto16] where the stability threshold is crossed before the feasibility threshold.

Starting from (26), if $\boldsymbol{x}^* > 0$, the equilibrium set of equations becomes a linear equation:

$$\boldsymbol{x}^* = \mathbf{1} + B_n \boldsymbol{x}^* \,. \tag{27}$$

We will restrict here to the non-trivial case in which $\alpha_n \to \infty$ and define the feasibility threshold by $\alpha_n^* = \sqrt{2\log(n)}$.

Theorem 0.34 (Theorem 1.1 [BN21]). *Let $\alpha_n \xrightarrow[n\to\infty]{} \infty$. Let $\boldsymbol{x}^* = (x_k^*)_{k\in[n]}$ be the solution of (27) and recall $A_n := (A_{k\ell}, 1 \leqslant k, \ell \leqslant n)$ are i.i.d. random variables with $A_{k\ell} \sim \mathcal{N}(0,1)$.*

1. *If there exists $\varepsilon > 0$ such that eventually $\alpha_n \leqslant (1-\varepsilon)\alpha_n^*$ then*

$$\mathbb{P}\left\{\min_{k\in[n]} x_k^* > 0\right\} \xrightarrow[n\to\infty]{} 0\,.$$

2. *If there exists $\varepsilon > 0$ such that eventually $\alpha_n \geqslant (1+\varepsilon)\alpha_n^*$ then*

$$\mathbb{P}\left\{\min_{k\in[n]} x_k^* > 0\right\} \xrightarrow[n\to\infty]{} 1\,.$$

Remark 0.13 (Sketch of proof). Let $Q_n = \left(I - \frac{A_n}{\alpha\sqrt{n}}\right)^{-1}$, the resolvent of the matrix A_n. The problem is defined by:

$$\boldsymbol{x}^* = \mathbf{1} + \left(\frac{A_n \boldsymbol{x}^*}{\alpha_n\sqrt{n}}\right) \quad \Leftrightarrow \quad \boldsymbol{x}^* = \left(I - \frac{A_n}{\alpha_n\sqrt{n}}\right)^{-1} \mathbf{1}\,.$$

For each entry of the vector $\boldsymbol{x}^*$

$$\begin{aligned}\forall k \in [n],\ x_k^* &= \sum_{i=0}^{\infty} e_k^* \left(\frac{A_n}{\alpha_n\sqrt{n}}\right)^i \mathbf{1}\,,\\ &= 1 + \frac{1}{\alpha_n} e_k^* \frac{A_n}{\sqrt{n}}\mathbf{1} + \frac{1}{\alpha_n^2} e_k^* \left(\frac{A_n}{\alpha_n}\right)^2 Q_n \mathbf{1}\,,\\ &= 1 + \frac{Z_k}{\alpha_n} + \frac{R_k}{\alpha_n^2}\,,\\ &\approx 1 + \frac{Z_k}{\alpha_n},\ Z_k \sim \mathcal{N}(0,1)\,.\end{aligned}$$

We can show that the R_k term is negligible by using Gaussian concentration properties (for more details, see [BN21]). From the properties of the extreme values of a family of Gaussian random variables, one can deduce

$$\min_{k\in[n]} x_k^* > 0 \quad \Leftrightarrow \quad \min_{k\in[n]} Z_k \sim -\sqrt{2\log(n)} > -\alpha_n\,.$$

The existence and uniqueness results of a feasible equilibrium have been extended in the case of a sparse food web by Akjouj and Najim [AN21]. They assume that each species is in relation with d other species. The magnitude of d with respect to n reflects the sparsity of the model. Two distinct cases are studied, on the one hand d is proportional to n. On the other hand $d \geqslant \log(n)$ in the case of a particular block structure. Moreover, they also demonstrate the global stability of the equilibrium.

Invadability condition

The problem (26) becomes much more complex when we consider an equilibrium where $\boldsymbol{x}^*$ has some vanishing components. The system of equations is no longer linear and the equation becomes a non-linear optimization problem. A naive and immediate solution to solve this problem is to choose a subset $\mathcal{I} \subset [n]$ and set the corresponding components $x_{\mathcal{I}} = (x_i^*)_{i\in\mathcal{I}}$ to zero, and to solve the remaining linear system:

$$\boldsymbol{x}_{\mathcal{I}^c} = \mathbf{1}_{|\mathcal{I}^c|} + B_{\mathcal{I}^c}\boldsymbol{x}_{\mathcal{I}^c}\ .$$

If there exists $\boldsymbol{x}_{\mathcal{I}^c} \geqslant 0$ that solves the previous equation, then $\boldsymbol{x} = \begin{pmatrix}\boldsymbol{x}_{\mathcal{I}}\\ \boldsymbol{x}_{\mathcal{I}^c}\end{pmatrix}$ satisfies (26) and is a potential equilibrium. The number of sub-cases $\mathcal{I} \subset [n]$ is 2^n and in particular grows exponentially as $n \to \infty$.

The equilibrium equations become ill-posed as there might be many equilibria. A known condition in ecology for dynamical system is the non-invadability condition [LM96, JS98] associated to saturated equilibrium. An equilibrium is saturated if it is resistant against the invasion by an absent species. The study of saturated equilibrium and permanence is an important research topic in the field of dynamical systems (for more details, see the seminal book of Hofbauer and Sigmund [HS98]).

Definition 0.25 (Saturated equilibrium)**.** Given $\mathcal{I}^c$ the set of persisting species,

- $\boldsymbol{x}$ is saturated $\Leftrightarrow \forall k \in \mathcal{I} : 1 - x_k^* + (B\boldsymbol{x}^*)_k \leqslant 0$,

- $\boldsymbol{x}$ is strictly saturated $\Leftrightarrow \forall k \in \mathcal{I} : 1 - x_k^* + (B\boldsymbol{x}^*)_k < 0$.

Lemme 0.35.

1. *If there is a positive solution $\boldsymbol{x}(t) > 0$, such that $\boldsymbol{x}(t) \xrightarrow[t\to\infty]{} \boldsymbol{x}^*$, then $\boldsymbol{x}^*$ is a saturated equilibrium.*
2. *If $\boldsymbol{x}^*$ is strictly saturated, then there exists a positive solution $\boldsymbol{x}(t) > 0$, such that $\boldsymbol{x}(t) \xrightarrow[t\to\infty]{} \boldsymbol{x}^*$.*

Notice that relying on standard properties of dynamical systems, see for instance [Tak96, Theorem 3.2.5], a necessary condition for the equilibrium $\boldsymbol{x}^*$ to be stable is that

$$1 - x_k^* + (B\boldsymbol{x}^*)_k \leqslant 0\,. \tag{28}$$

The condition (28) decreases the number of potential solutions to system (26). In reference to the ODE (23), the requirement for a given species $k \in [n]$ to be non-invadable is equivalent to:

$$\left(\frac{1}{x_k}\frac{dx_k}{dt}\right)_{x_k\to 0^+} \leqslant 0\,. \tag{29}$$

As a consequence, we will now focus on the following set of conditions:

$$\begin{cases} x_k^*\,(1 - x_k^* + (B\boldsymbol{x}^*)_k) &= 0 \quad \text{for } k \in [n]\,, \\ 1 - x_k^* + (B\boldsymbol{x}^*)_k &\leqslant 0 \quad \text{for } k \in [n]\,, \\ \boldsymbol{x}^* &\geqslant 0 \quad \text{componentwise}\,. \end{cases} \tag{30}$$

This casts the problem of finding a non negative equilibrium into the class of Linear Complementarity Problems (LCP), which we describe hereafter.

Linear complementary problem

LCP is a class of problems from mathematical optimization which in particular encompasses linear and quadratic programs; standard references are [Mur88, CPS09]. Given a $n \times n$ matrix M and a $n \times 1$ vector $\boldsymbol{q}$, the associated LCP denoted by $LCP(M, \boldsymbol{q})$ consists in finding two $n \times 1$ vectors $\boldsymbol{z}, \boldsymbol{w}$ satisfying the following set of constraints:

$$\begin{cases} \boldsymbol{z} &\geqslant 0\,, \\ \boldsymbol{w} = M\boldsymbol{z} + \boldsymbol{q} &\geqslant 0\,, \\ \boldsymbol{w}^\top \boldsymbol{z} &= 0 \quad \Leftrightarrow \quad w_k z_k = 0 \quad \text{for all } k \in [n]\,. \end{cases} \tag{31}$$

Since $\boldsymbol{w}$ can be inferred from $\boldsymbol{z}$, we denote $\boldsymbol{z} \in LCP(M, \boldsymbol{q})$ if $(\boldsymbol{w}, \boldsymbol{z})$ is a solution of (31).

The LCP problem study goes back to the work of Lemke [Lem65] and Cottle *et al.* Danzig [CD68]. Lemke and Howson [LH64] developed a algorithm based on pivot steps to solve the problem (31).

Introduced by Fielder and Pták [FP66], the class of P-matrices is related to the linear complementarity problem. Murty [Mur72] showed that a P-matrix give necessary and sufficient conditions to have a unique equilibrium to the LCP problem.

Definition 0.26 (P-matrix)**.** A square matrix M is called a P-matrix if all its principal minors (sub-determinants) are strictly positive

$$\det(M_{\mathcal{I}}) > 0\,, \quad \forall\, \mathcal{I} \subset [n]\,, \quad M_{\mathcal{I}} = (M_{k\ell})_{k,\ell\in\mathcal{I}}\,.$$

Many properties on the necessary and sufficient conditions for a real matrix to be a P-matrix has been studied by Rump [Rum03] and Rohn [Roh12]. Numerically, checking that a matrix is a P-matrix is co-NP complete [Cox94].

Theorem 0.36 (Existence and uniqueness of a solution to the LCP problem [Mur72])**.** *A matrix M is a P-matrix if and only if the $LCP(M, \boldsymbol{q})$ has a unique solution $(\boldsymbol{w}, \boldsymbol{z})$ for all $\boldsymbol{q} \in \mathbb{R}^n$.*

In the case of the Lotka-Volterra and in view of (30), we look for $\boldsymbol{x}^* \in LCP(I-B, -\mathbf{1})$.

Definition 0.27 (M-matrix)**.** A square matrix A is called M-matrix if it can be expressed in form $A = sI - C$, where $C = (C_{k\ell})$ with $C_{k\ell} \geqslant 0, 1 \leqslant k, \ell \leqslant n$, and $s > \rho(C)$, the spectral radius of C.

The name M-matrix was given by Ostrowski [Ost56] refering to Hermann Minkowski. Many properties of M-matrices has been introduced by Fiedler and Pták [FP62] and extended by Plemmons [Ple77].

Remark 0.14.

- The set of non-singular M-matrices are a subset of the class of P-matrices.
- The set of non-singular M-matrices are a subset of the class of inverse-positive matrices i.e.

$$A^{-1} \text{ exists and } A^{-1} \geqslant 0\,.$$

Global stability

Theorem 0.36 gives a sufficient and necessary condition for the existence of a unique non-invadable equilibrium to the equation (26). In the case of a feasible equilibrium $\boldsymbol{x}^* > 0$, Bizeul and Najim [BN21] have shown that there exists a globally stable equilibrium. In the case of an equilibrium with vanishing species, it is necessary to go back to the properties of Lyapunov functions.

Lyapunov's theorem says that a matrix A is stable (its eigenvalues have a strictly negative real part) if and only if there exists a positive definite matrix H such that $HA + A^\top H$ is negative definite. This condition goes back to the work of Lyapunov [Lia07] which has been improved and studied by Barker *et al.* [BBP78] and Logofet [Log05] which makes a summary of all the matrix conditions in the form of a flower.

Definition 0.28 (Lyapunov diagonal stability)**.** Matrix M is called Lyapunov diagonally stable, denoted by $M \in S_\omega$, if and only if there exists a diagonal matrix D with positive elements such that $DM + M^\top D$ is negative definite i.e. all eigenvalues are negative.

Proposition 0.37 (Takeuchi *et al.* [TAT78])**.** *If $M \in S_\omega$ then $-M$ is a P-matrix.*

Recall the system (21) and consider the matrix B is arbitrary,

$$\frac{d\,y_k(t)}{dt} = y_k(r_k + (-\theta I + B\boldsymbol{y})_k)\,, \quad k \in [n]\,. \tag{32}$$

Takeuchi and Adachi (see for instance [Tak96, Th. 3.2.1]) provide a criterion for the existence of a unique equilibrium $\boldsymbol{y}^*$ and the global stability of the LV system.

Theorem 0.38 (Takeuchi and Adachi [TA80]). *If $-\theta I + B \in S_\omega$, then $LCP(\theta I - B, \boldsymbol{r})$ admits a unique solution. In particular, for every $\boldsymbol{r} \in \mathbb{R}^n$, there is a unique equilibrium $\boldsymbol{y}^*$ to (32), which is globally stable in the sense that for every $\boldsymbol{y}_0 > 0$, the solution to (32) which starts at $\boldsymbol{y}(0) = \boldsymbol{y}_0$ satisfies*

$$\boldsymbol{y}(t) \xrightarrow[t\to\infty]{} \boldsymbol{y}^* .$$

The Lotka-Volterra model from a physicist standpoint

Previously, mathematical conditions were given for the existence of a feasible equilibrium and uniqueness of a globally stable equilibrium in (21) where some species may vanish. However, the richness of the Lotka-Volterra equations stems from the diversity of its dynamical behaviors. The lack of mathematical knowledge is supplemented by methods from physics to improve the understanding on these diverse dynamical behaviors (properties of the equilibrium, out-of equilibrium dynamics, model sophistication).

Since a long time, the theory of statistical mechanics of disorder systems have been developed to study spin glasses and replicator system (see Mezard *et al.* [MPV86] for a review).

The use of these methods to study biological systems were firstly introduced by Diederich and Opper [DO89, OD92] and used to study multi-species replicator dynamics (model equivalent to the Lotka-Volterra system) by Tokita [Tok04]. More recently, these methods of statistical physics of disorder system have resurfaced to solve problems in theoretical ecology. In particular, the dynamical cavity method is used to analyze community dynamics where random interactions between species are considered. Physicists have divided the space of parameters (μ, α, ρ) in a phase diagram where the major issue is to identify the boundaries between the different phases: unique stable fixed point, chaos with multiple attractors, unbounded growth, etc.

The cavity method enables to derive mean-field equation approximating a high-dimensional nonlinear problem. The key concept consist of assuming a unique fixed point exist and introduce a new species with new interactions in the existing system. After the establishment of the new species, an analogy between the properties of the solutions with n and $n+1$ species is verified. This method is used to study the system (21) that admits a unique stable fixed point but also out of equilibrium dynamics. The dynamics may fall into a chaotic phase with multiple attractors. Bunin [Bun16, Bun17] used the dynamical cavity method to conducts and extends more general results (properties of the persisting species, multiple attractors phase) for the phase diagram of the Lotka-Volterra system (4). These methods have been used to solve many problems in theoretical ecology. In particular, Barbier *et al.* [BABL18] exhibits generic behaviors in complex communities. For a review on the cavity method applied to community dynamics problems see Barbier and Arnoldi [BA17].

The model studied by physicists is more generic, disordered Lotka-Volterra system is given by

$$\frac{dx_k(t)}{dt} = x_k(t)\left(r_k - x_k(t) + \sum_{\ell\in[n]} B_{k\ell}x_\ell(t)\right) + \lambda_k + \omega_k\sqrt{2x_k(t)}\eta_k(t) \ , \tag{33}$$

where the λ_k are a migration constants and $\omega_k\sqrt{2x_k(t)}\eta_k(t)$ is a demographic noise term, $\eta_k(t)$ is random time-varying function.

The phase diagram studied in Bunin [Bun17] had already been investigated numerically by Kessler and Shnerb [KS15] with the presence of the migration parameter λ_k. The Lotka Volterra model with demographic noise term has been investigated recently by Bunin [Bun21] and Altieri *et al.* [ARCB21]. The dynamical cavity method enables to derive dynamical mean-field equation for out-of-equilibrium applications [RBBC19, RBBB20, ABC20], in particular for the study of the multiple attractors dynamics represented in Figure 24. Furthermore, Biroli *et al.* [BBC18] show that multiple equilibrium regime is analogous to a critical spin-glass phase. For a review of glassy phases in ecological systems see Altieri [Alt22].

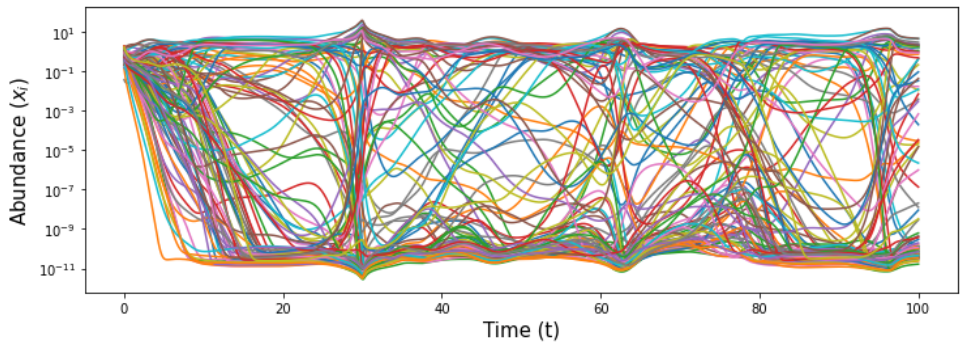


Figure 24: Dynamics of a 100 species LV system with migration (33) in the chaotic phase with multiple attractors and parameters $\mu = 4$, $\alpha = 0.5$, $\forall i \in [n]\,, \lambda_i = 10^{-10}$. The y-axis is in a log-scale.

An alternative method using generating functional techniques for deriving similar mean-field equations to study the equilibrium phase in the LV system was used by Galla [Gal18]. Identical methods was used to derive the eigenvalues of random matrices [BJRG22b, BJRG22a] and analyze the Lotka Volterra models with different interaction structure such as cascade model [PBG22].

Other applications are possible, such as the recent work of Fraboul *et al.* [FBM22] on mutations in the LV model or the study of the impact on the Allee effect on the phase diagram by Altieri *et al.* [AB22].

Contributions

Chapter 1 - Equilibrium and persisting species in a large Lotka-Volterra system of differential equations

This Chapter is based on a preprint by Clenet, Massol and Najim [CMN22].

In Chapter 1, we focus on the model (23) where matrix B is a simpler version of (24) admitting the following representation:

$$B = \frac{A}{\alpha\sqrt{n}} + \frac{\mu}{n}\mathbf{1}_n\mathbf{1}_n^\top\,,$$

where $A = (A_{k\ell})$ is a matrix with random standardized ($\mathbb{E}A_{k\ell} = 0$ and $\mathrm{Var}(A_{k\ell}) = 1$) independent and identically distributed (i.i.d.) entries with finite fourth moment, $\alpha > 0$

is an extra parameter reflecting the interaction strength and $\mu \in \mathbb{R}$ represents an arbitrary trend of the interactions.

In Theorem 0.34, Bizeul and Najim proved the existence of a threshold $\alpha \sim \sqrt{2\log(n)}$ in the case $\mu = 0$, which guarantees the feasibility of the equilibrium $\boldsymbol{x}^*$ of (23). However, Dougoud *et al.* [DVR+18] showed that some species will go to extinction if $\alpha > 0$ is independent of n. The aim of the chapter is to describe the impact of the interaction strength α and interaction trend μ on the conditions of coexistence of the interacting species.

First, combining results from Takeuchi and Adachi in Theorem 0.38 with standard RMT results in Theorem (0.27), we provide sufficient conditions on the parameters α and μ to ensure the existence of a unique globally stable equilibrium $\boldsymbol{x}^*$ in large dimension $n \to \infty$. The equilibrium is composed of persisting species and vanishing species.

Later on, given a unique equilibrium $\boldsymbol{x}^*$, we describe the properties of the persisting species. In this perspective, we provide a heuristics to compute asymptotically the proportion of persisting species and understand via a system of equations the dependence between parameters α and μ and the proportion of persisting species. Furthermore, we show that the distribution of abundance of persistent species is a truncated Gaussian (see Figure 25).

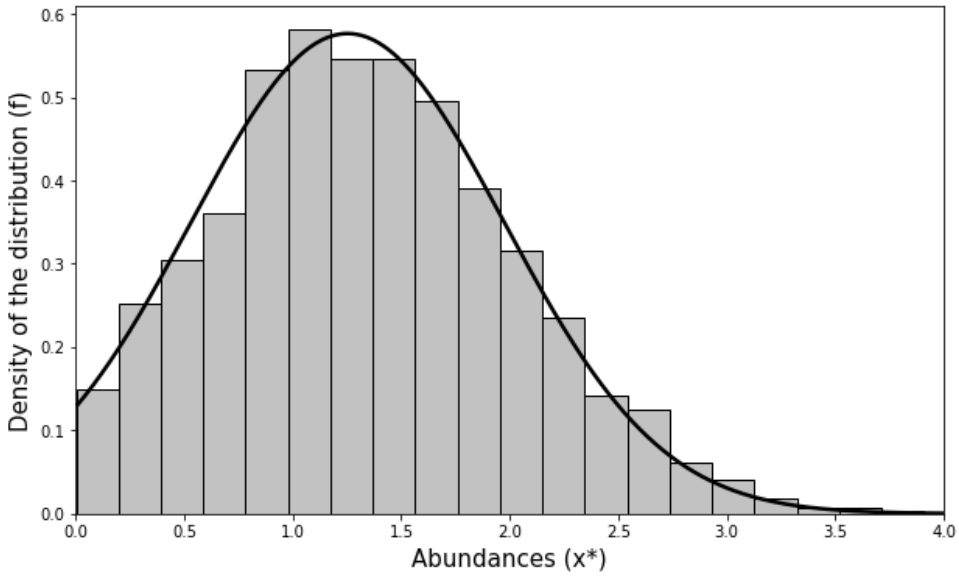


Figure 25: Distribution of the abundance of persisting species. The x-axis represents the value of the abundances and the histogram is built upon the positive components of equilibrium $\boldsymbol{x}^*$. The solid line represents the theoretical distribution for parameters (α, μ) as given by the heuristics. The entries are Gaussian $\mathcal{N}(0,1)$ and the parameters are set to $(n = 2000, \alpha = 2, \mu = 0.2)$.

In nature, interactions between species are constantly changing and affected by the environment. Under the assumptions that environmental conditions influence interaction strengths, we study the consequences of a sudden change of environmental conditions, expressed through a decrease in parameter α. When α varies for the same matrix A and the same parameter μ, the system can display different states. When the value of α increases above a certain critical value, all species will coexist; conversely, for sufficiently low values of α, species may vanish while keeping a unique stable equilibrium. We describe

the change between these two states and how the proportion of persisting species varies by solving numerically the Lotka-Volterra system. We observe that a decrease in α negatively affects equilibrium species richness (see Figure 26).

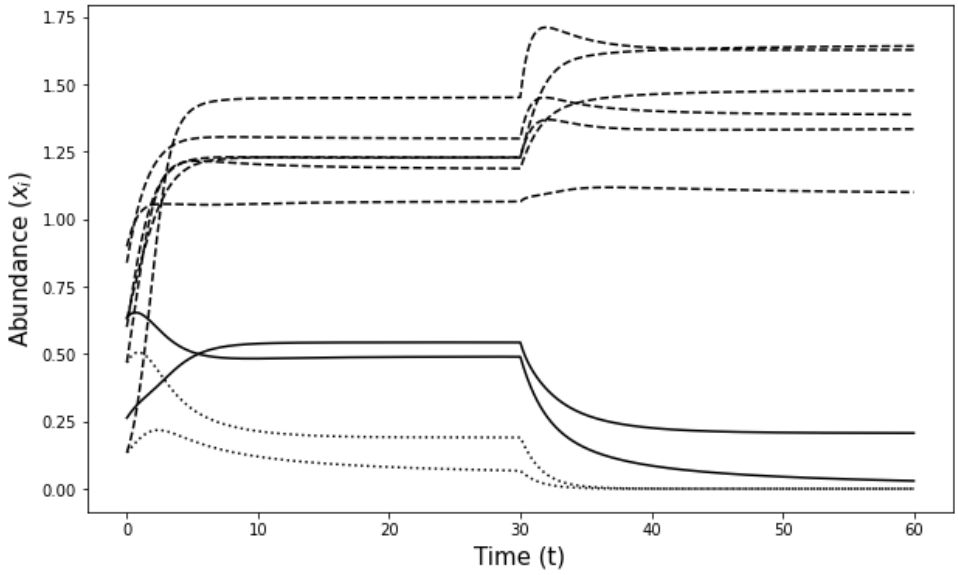


Figure 26: Abundance dynamics in the case of a community of ten species. The matrix of interactions A and the initial conditions are common and we apply a abrupt variation of $\alpha(t)$ at $t = 30$. The dashed lines represents species which benefit from habitat variation; solid lines represent species suffering from the change. Dotted lines represent species undergoing extinction.

Finally, we analyze a diversity index (Hill numbers of order 1) to have a more precise representation of biodiversity dynamics. The dynamics of this diversity measure suggests that the mean of interaction coefficients, μ, affects the duration of transient dynamics, with shorter transient dynamics being associated with more mutualistic interactions (i.e. higher positive values of μ).

Chapter 2 - Equilibrium in a large Lotka-Volterra system with pairwise correlated interactions

This Chapter is based on the article by Clenet, El Ferchichi and Najim and will be published in Stochastic Processes and its Applications in November 2022 [CEFN22].

In Chapter 2, we focus on the model (24) where we extend the result on the feasibility threshold of Bizeul and Najim (0.34). The interaction matrix B_n is a non-centered random matrix with pairwise correlated entries:

$$B_n = \frac{A_n}{\alpha_n \sqrt{n}} + \frac{\mu}{n} \mathbf{1}_n \mathbf{1}_n^\top ,$$

where $A_n = (A_{k\ell})_{k,\ell \in [n]}$ is a random matrix satisfying the two conditions (i) $(A_{k\ell}, k \leqslant \ell)$ are standard Gaussian $\mathcal{N}(0,1)$ independent and identically distributed (i.i.d.) random variables (ii) for $k < \ell$ the vector $(A_{k\ell}, A_{\ell k})$ is a standard bivariate Gaussian vector, independent from the remaining random variables, with covariance $\mathrm{cov}(A_{k\ell}, A_{\ell k}) = \rho$

with $|\rho| \leqslant 1$. The sequence of positive numbers (α_n) is either fixed or goes to infinity. Parameter μ is a fixed real number.

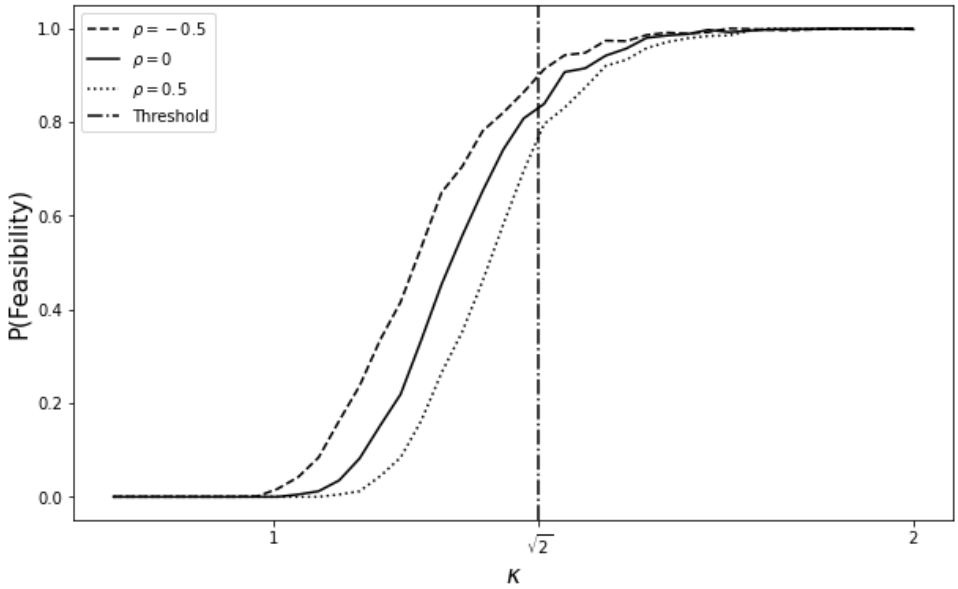


Figure 27: Transition towards feasibility for the elliptic model. For each κ on the x-axis, we simulate 1000 matrices B_n of size $n = 1000$, compute the solution $\boldsymbol{x}_n$ of the feasibility Theorem at the scaling $\alpha_n(\kappa) = \kappa\sqrt{\log(n)}$ and then plot the proportion of feasible solutions obtained for the 1000 simulations. Each curve represents the proportion of feasible solutions $\boldsymbol{x}_n$ for three distinct values $\rho \in \{-0.5,\ 0,\ 0.5\}$. The dot-dashed vertical line corresponds to $\kappa = \sqrt{2}$ i.e. the critical scaling $\alpha_n^* = \sqrt{2\log(n)}$.

We prove that feasibility is reached whenever $\alpha_n \gg \sqrt{2\log(n)}$ and $\mu < 1$, and that there is no feasibility otherwise. Furthermore, the correlation parameter ρ has no influence since the phase transition threshold is the same as in the i.i.d. case [BN21]: the induced correlations between components x_k's of solution $\boldsymbol{x}_n$ are too weak (see Figure 27). In addition, we prove that the same phase transition holds if we consider a covariance profile $(\rho_{k\ell},\ k < \ell)$ where $\rho_{k\ell} = \mathrm{cov}(A_{k\ell}, A_{\ell k})$ instead of a fixed covariance parameter ρ.

Using results by Takeuchi and Adachi (0.38) on stability of LV systems with RMT results in Theorem 0.33, we establish sufficient conditions for the existence of a unique stable equilibrium where some species may vanish which represents an extension of Chapter 1.

We finally conclude with an important outcome on estimating the proportion of persisting species. Using physicists' arguments, we state the open problem, recall Bunin's and Galla's equations and provide simulations of a closed-form system of equations to compute the proportion of persisting species.

Chapter 3 - Impact of a block structure in large systems of Lotka-Volterra

This Chapter is an ongoing project between Clenet, Massol and Najim.

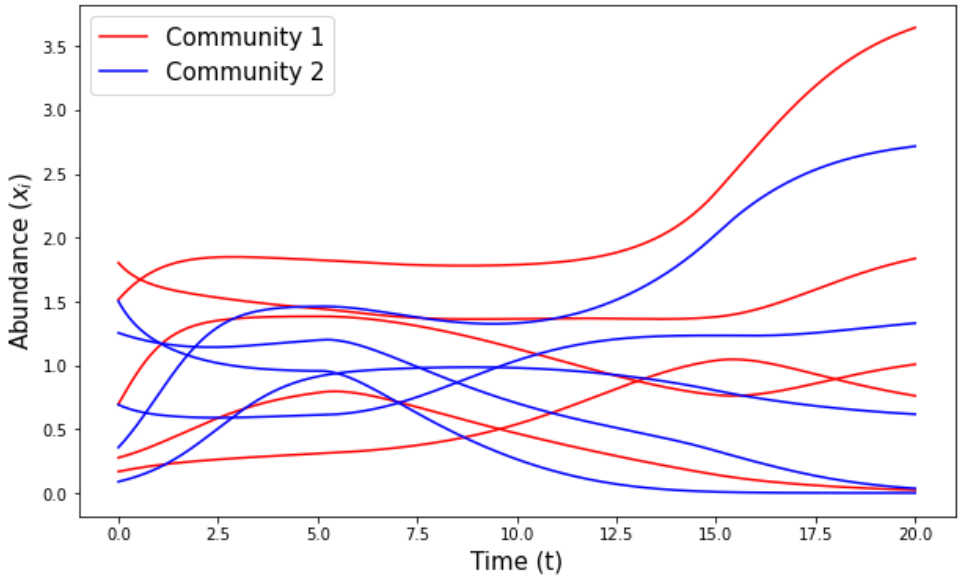


Figure 28: Dynamics of model (24) of 2 distinct communities of 5 species with interaction matrix (34). At $t = 0$, the two communities converges to their feasible equilibrium point and are not interacting. At $t = 5$, the two communities start to interact i.e. α_{12} and α_{21} increase with a linear growth until $t = 15$. Then, the two communities converges to their new equilibrium point with persisting and extinct species in both communities.

In Chapter 3, we aim at developing the results of chapter 1 and 2 in an ecosystem with many communities. In nature, interaction networks are rather structured, which contributes to the stability of the system. To extend the results of the Lotka-Volterra model to describe the properties of a multi-community dynamic, we define a block interaction matrix in which we can adapt the intra and inter-community interactions. For interpretation purposes due to the complexity of the model, we consider the case of two interacting communities (see Figure 28). Within the framework of 2 communities, the matrix $B = (B_{k\ell})_{n,n}$ is defined as

$$B = \frac{1}{\sqrt{n}} \begin{pmatrix} \frac{A_{11}}{\alpha_{11}} & \frac{A_{12}}{\alpha_{12}} \\ \frac{A_{21}}{\alpha_{21}} & \frac{A_{22}}{\alpha_{22}} \end{pmatrix} + \frac{1}{n} \begin{pmatrix} \mu_{11}\mathbf{1}_{\mathcal{I}_1}\mathbf{1}_{\mathcal{I}_1}^\top & \mu_{12}\mathbf{1}_{\mathcal{I}_1}\mathbf{1}_{\mathcal{I}_2}^\top \\ \mu_{21}\mathbf{1}_{\mathcal{I}_2}\mathbf{1}_{\mathcal{I}_1}^\top & \mu_{22}\mathbf{1}_{\mathcal{I}_2}\mathbf{1}_{\mathcal{I}_2}^\top \end{pmatrix}, \tag{34}$$

where:

$$\boldsymbol{s} = \begin{pmatrix} 1/\alpha_{11} & 1/\alpha_{12} \\ 1/\alpha_{21} & 1/\alpha_{22} \end{pmatrix}, \ \boldsymbol{\mu} = \begin{pmatrix} \mu_{11} & \mu_{12} \\ \mu_{21} & \mu_{22} \end{pmatrix},$$

$\boldsymbol{\beta} = (\beta_1, \beta_2)$, $\sum_{i=1}^{2} \beta_i = 1$ is the size by proportion of each of the blocks, $\mathcal{I}_i$ is a subset of $[n]$ of size $|\mathcal{I}_i| := \beta_i n$ matching the index of species belonging to community i, $\mathbf{1}_{\mathcal{I}_i}$ is a entry wise vector of 1 of size $\beta_i n$. A_{ij} is a non-Hermitian random matrix of size $(\beta_i n, \beta_j n)$ with reduced centered Gaussian entries i.e. $\mathcal{N}(0, 1)$. The matrix $\boldsymbol{s}$ represents the interaction strength in each block. The trend matrix $\boldsymbol{\mu}$ allows to adjust on average the type of interaction (mutualism, competition) of each block.

In a first section, we extend the feasibility result of Bizeul and Najim (0.34) for a block interaction matrix where $\boldsymbol{\mu} = \mathbf{0}$. Using this result, we study the maintenance of the feasibility of two communities when adding interactions between them. The interactions between communities reduce the feasibility and if we assume that the communities can

vary in size and the intra-community interactions are different, the community with the lowest interaction strength is advantaged i.e. the size of the community can be larger. We conclude this first part by studying the non-centered case $\boldsymbol{\mu} \neq \mathbf{0}$, where we give feasibility conditions using properties on M-matrices.

The second and third parts of this chapter are an extension of the results of Chapter 1 and an ecological interpretation of the results. On the one hand, in the second part, we study the existence of a unique globally stable equilibrium where species can become extinct by using a result of Takeuchi and Adachi in Theorem 0.38 and RMT, in particular the theory on quadratic vector equation. We establish a theorem for the case $\boldsymbol{\mu} = 0$, then we study the non-centered case $\boldsymbol{\mu} \neq 0$ when the strength of interactions is similar in each block. In contrast to the results in the single-community case, increasing inter-community competition can destabilize the system. On the other hand, the third part describes the

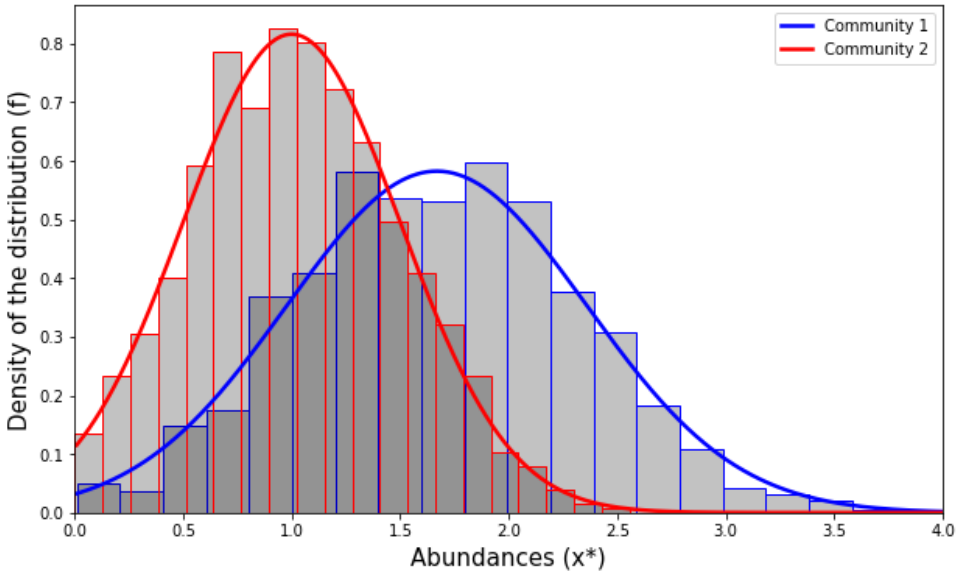


Figure 29: Distribution of abundance of the persisting species in each community. The x-axis represents the value of the abundances and the histogram is built upon the positive components of equilibrium $\boldsymbol{x}^*$ associated to each community. The blue-solid line (resp. red-solid line) represents the theoretical distribution of community 1 (resp. community 2) for parameters $(\boldsymbol{\alpha}, \boldsymbol{\mu})$ as given the heuristics. The entries are Gaussian $\mathcal{N}(0,1)$ and the parameters are set to

$$n = 1000\ ,\ \boldsymbol{\mu} = \begin{pmatrix} 0.5 & 0.5 \\ 0 & 0 \end{pmatrix}\ ,\ \boldsymbol{\alpha} = \begin{pmatrix} 2 & 3 \\ 3 & 3 \end{pmatrix}\ ,\ \boldsymbol{\beta} = \begin{pmatrix} \frac{1}{2}, \frac{1}{2} \end{pmatrix}\ .$$

heuristics on the properties and distribution of abundances of persistent species in each block (see Figure 29). A graphical interpretation of these heuristics highlights several elements, there is a contagion of diversity: the higher the persistence of a community, the less its impact will be harmful on the other communities. The decline in persistence between two interacting communities is not linear but has a double negative effect, hence the importance of maintaining persistent communities and not neglecting feedback phenomena in community interactions. We conclude with a study of the impact of mutual and competitive interactions.

In an fourth section, a numerical similarity study is performed between a model where the interaction strength varies and a model where the connectance in each of the communities is varying, which gives an adjacency matrix of the interaction graph known as the Bernouilli Stochastic Block Model (see Figure 30). This similarity is analyzed through the stability condition given historically by May [May72].

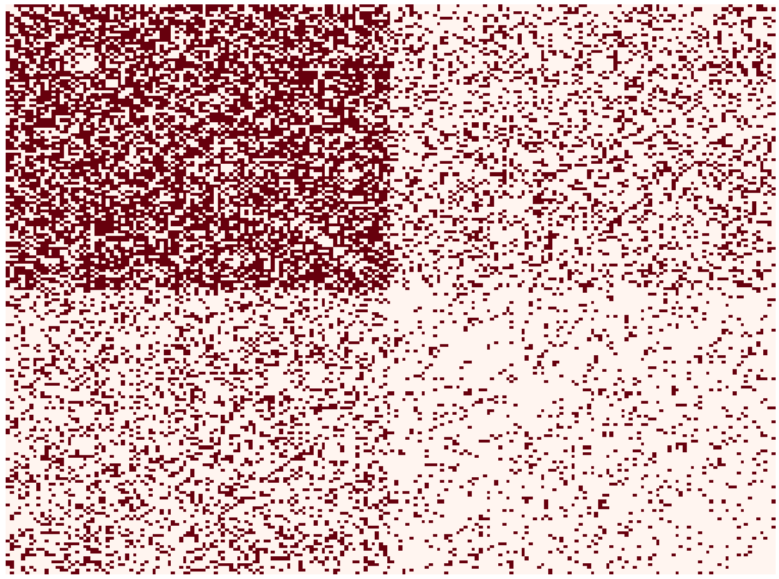


Figure 30: Representation of an adjacency matrix of the interactions of an ecosystem of size $n = 200$. A graph of a symmetric Stochastic Block model of parameter $P = \begin{pmatrix} 0.6 & 0.25 \\ 0.25 & 0.1 \end{pmatrix}$. A red colored cell indicates $S_{k\ell} = 1$, on the contrary, a white colored cell indicates that there is no interaction $S_{k\ell} = 0$.

Chapter 4 - A probabilistic perspective of the hierarchical competition-colonization trade-off model

This Chapter is an ongoing project between Allesina, Clenet, Della Libera, Massol and Miller.

In Chapter 4, we study the hierarchical competition-colonization trade-off model (25), in the case of a hierarchical competition i.e.

$$\eta_{k\ell} = \begin{cases} 1 & \text{if } k < \ell\,, \\ 0 & \text{otherwise.} \end{cases}$$

The dynamics of each species' abundances within the habitat depends mainly on its colonization rate and its extinction rate. We propose a probabilistic interpretation of the model by sampling the colonization rates from a given probability distribution with the same extinction rate for all species. In this framework, we investigate two different types of assembly processes.

On the one hand, it is assumed that initially the community contains a pool of n species (all-at-once metacommunity process) and we let the occupancies of all species

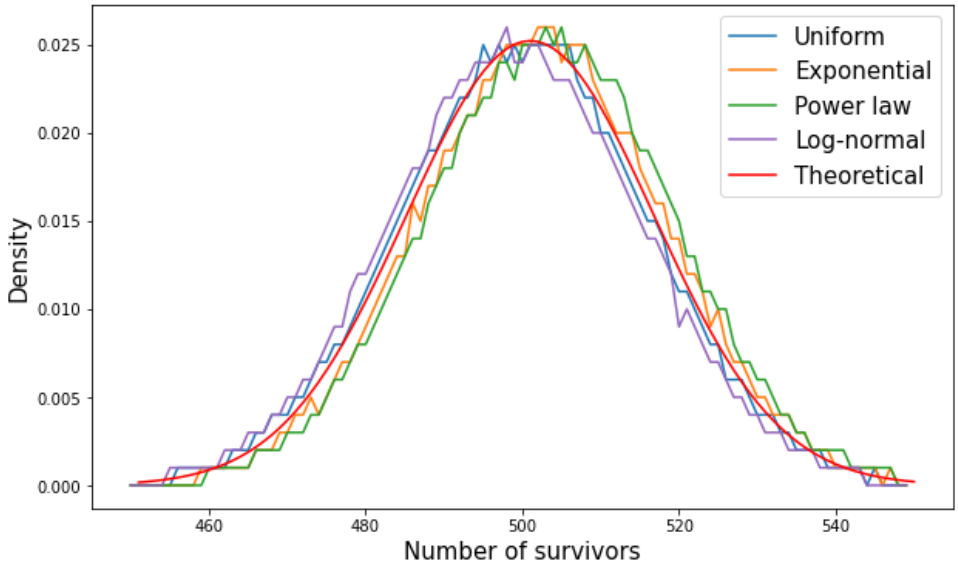


Figure 31: Representation of the distribution of the number of persistent species for a initial pool of $n = 1000$ species and for different distribution of colonization rate. Each curve is derived using Monte Carlo experiments by computing $P = 100000$ times the all-at-once metacommunity process and store the values obtained to form the outline of a histogram. The red curve corresponds to the density function of the binomial distribution $\mathcal{B}(n, \frac{1}{2})$.

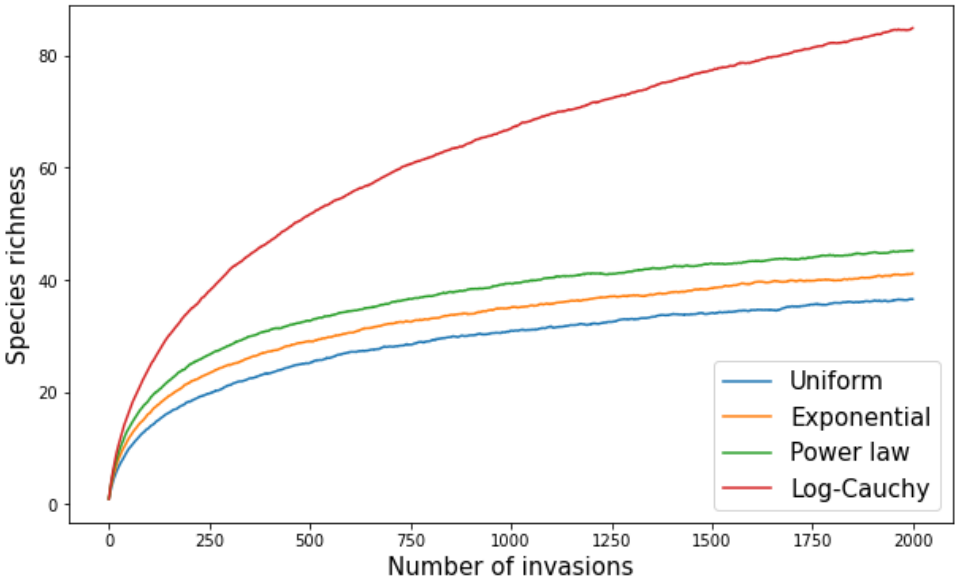


Figure 32: Representation of the species richness of the sequential invasion model as a function of the number of invasions for different distributions. The curve is derived using Monte Carlo simulations by computing $P = 2000$ times and averaging the number of persistent species.

change in time according to equation (25) where some species persist while others may vanish. Surprisingly, we obtain a universality result of the distribution of the number of persistent species. For a wide range of distribution, on average the proportion of

persistent species is one half. Furthermore, we show (for the uniform distribution) that the distribution of persistent species is a binomial distribution $B(n, 1/2)$ (see Figure 31).

In addition to the universality result, we describe the properties of persistent species by recalling some results of Kinzig *et al.* [KLD+99] on fraction of empty patches and occupancies. We end-up this section by clarifying the intermediate disturbance hypothesis that had been observed by Hastings [Has80] where optimal coexistence between species occurs when the mortality rate is intermediate.

On the other hand, we study a sequential invasion process. Beginning with an empty habitat, it is filled by sequentially introducing species whose colonization rates are drawn according to a specific distribution. We observe that the number of persisting species saturates with a logarithmic growth due to historical contingencies and extinction cascades (see Figure 32). We analyze the properties of historical contingencies due to the phenomenon of extinction cascades which is a key element of the saturation phenomenon.

We give some theoretical elements of answer before carrying out a numerical analysis of the assembly process model. The universality result is no longer true and a major difference is observed between regular and heavy-tailed distributions. In general, the heavier the tail, the greater the diversity. This assumption of competition colonization trade-off show the importance of finding a balance between competitors and colonizers.

To conclude, this probabilistic perspective of the hierarchical competition-colonization trade-off model put forward and compare two different types of distinct assemblages and gives conditions for many species to coexist under the competition-colonization trade-off.

Chapter 1

Equilibrium and persisting species in a large Lotka-Volterra system of differential equations

Abstract

Lotka-Volterra (LV) equations play a key role in the mathematical modeling of various ecological, biological and chemical systems. When the number of species (or, depending on the viewpoint, chemical components) becomes large, basic but fundamental questions such as computing the number of persisting species still lack theoretical answers. In this paper, we consider a large system of LV equations where the interactions between the various species are a realization of a random matrix. We provide conditions to have a unique equilibrium and present a heuristics to compute the number of surviving species. This heuristics combines arguments from Random Matrix Theory, mathematical optimization (LCP), and standard extreme value theory. Numerical simulations, together with an empirical study where the strength of interactions evolves with time, illustrate the accuracy and scope of the results.

1.1 Introduction

Since May's seminal work [May72] and for the past decades, many theoretical studies addressed the issue of the coexistence of species in ecosystems.

Introduced in the 1920s by Lotka [Lot25] and Volterra [Vol26], the Lotka-Volterra (LV) model is a well-known classic in theoretical ecology and mathematics. It represents a first step in our understanding of ecosystems through the variety of its dynamical behaviours (single or multiple equilibria, cycles, chaos), its flexibility (many models can be approximated in the form of a LV model) and its mathematical calculability.

In this article, we consider large LV models with random parameters. Leveraging on the asymptotic understanding of large random matrices which naturally appear enables us to provide insights on equilibria and species coexistence for such models.

Model and assumptions.

Large Lotka-Volterra systems of differential equations arise in various scientific fields such as biology, ecology, chemistry, etc. Although our results are generic in nature and not specific to a given field, we will rely on the ecological terminology in the sequel.

A large system of Lotka-Volterra equations is a system of coupled ordinary differential equations (ODE) that write:

$$\frac{dx_k(t)}{dt} = x_k(t)\left(r_k - \theta x_k(t) + \sum_{\ell\in[n]} B_{k\ell}x_\ell(t)\right) , \tag{1.1}$$

where $k \in [n] = \{1, \cdots, n\}$.

Here, n represents the number of species in a food web or community, the unknown vector $\boldsymbol{x} = (x_k)_{k\in[n]}$ is the vector of abundances of the various species and evolves with time $t > 0$ according to the dynamics (1.1). Parameter r_k represents the intrinsic growth rate of species k, θ is an intraspecific feedback coefficient (most often positive due to competition), and $B_{k\ell}$ is the per capita effect of species ℓ on species k.

Remark 1.1. Notice that without interactions, i.e. $B = (B_{k\ell})_{k,\ell\in[n]} = 0$, system (1.1) is simply a system of uncoupled logistic differential equations.

We shall focus on the model where $r_k = \theta = 1$:

$$\frac{dx_k}{dt} = x_k\left(1 - x_k + (B\boldsymbol{x})_k\right) , \quad k \in [n] \tag{1.2}$$

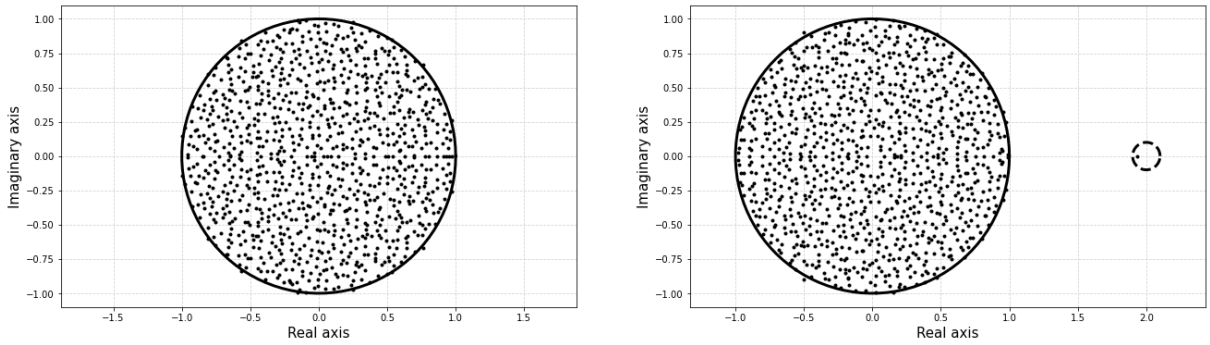
with matrix B admitting the following representation:

$$B = \frac{A}{\alpha\sqrt{n}} + \frac{\mu}{n}\mathbf{1}\mathbf{1}^T ,$$

where $A = (A_{ij})$ is a matrix with random standardized ($\mathbb{E}A_{ij} = 0$ and $\mathrm{var}(A_{ij}) = 1$) independent and identically distributed (i.i.d.) entries with finite fourth moment, $\alpha > 0$ is an extra parameter reflecting the interaction strength, and $\mu \in \mathbb{R}$ represents an arbitrary trend of the interactions. The $n \times 1$ vector $\mathbf{1}$ is a vector of ones.

Remark 1.2. Although matrix B is a complex random object, a result by Tao [Tao13, Theorem 1.7] fully describes its asymptotic spectrum: Assume that $|\mu| > 1/\alpha$, then for any fixed $\varepsilon > 0$, almost surely eventually all the eigenvalues of B but one are in the disk $\{z \in \mathbb{C} : |z| \leqslant 1/\alpha + \varepsilon\}$ while one extra eigenvalue takes the value $\mu + o(1)$.

Remark 1.2 is illustrated in Fig. 1.1.



(a) No outlier ($\mu = 0$) (b) Presence of an outlier ($\mu = 2$)

Figure 1.1: Spectrum of non-Hermitian matrix B in the complex plan ($n = 1000$, $\alpha = 1$). In Fig. 1.1a, $\mu = 0$ and the solid line circle represents the boundary of the circular law. In Fig. 1.1b, $\mu = 2$ and there is an eigenvalue in the small dashed circle centered at 2, as predicted by [Tao13, Th. 1.7] - see also Remark 1.2.

Presentation of the main results.

Unique equilibrium. In the study of the behaviour of $\boldsymbol{x}(t)$ as $t \to \infty$ the existence of an equilibrium $\boldsymbol{x}^*$ to Eq. (1.2) is an important prior to any stability property of $\boldsymbol{x}(t)$. By equilibrium, we mean the existence of a vector $\boldsymbol{x}^* = (x_k^*)$ satisfying

$$x_k^*(1 - x_k^* + (B\boldsymbol{x}^*)_k) = 0 \quad \text{for} \quad k \in [n]\,.$$

General results on LV systems state that $\boldsymbol{x}(t) > 0$ (componentwise) as long as $\boldsymbol{x}(0) > 0$ [HS98]. However, a possible equilibrium $\boldsymbol{x}^*$ will only verify $\boldsymbol{x}^* \geqslant 0$, i.e. some components x_k^* may take the value zero.

In Theorem 1.2, we provide sufficient conditions on the parameters α and μ to ensure the existence of a unique equilibrium. These conditions rely on the "typical" behaviour of the random matrix B in large dimension $n \to \infty$.

Evaluating the number of surviving species. Given a unique equilibrium $\boldsymbol{x}^*$, an important question is to describe the set of surviving/vanishing species. In this perspective, we introduce the set

$$\mathcal{S} = \{i \in [n],\ x_i^* > 0\} \tag{1.3}$$

of surviving species. In Section 1.3, we provide a heuristics to compute asymptotically the ratio $\frac{|\mathcal{S}|}{n}$ and understand via a system of equations the dependence between parameters α and μ and the number of surviving species. A complementary result addressing the elliptic random matrix model by means of theoretical physics methods can be found in [Bun17] (dynamical cavity method) and in [Gal18] (generating functional techniques).

Notice that in [BN21], Bizeul and Najim have studied a different normalization for α in the case $\mu = 0$, namely $\alpha \sim \sqrt{2\log(n)}$, to guarantee the survival of every species (feasibility of the equilibrium). Indeed, a consequence of Dougoud *et al.*'s results [DVR+18] is that some species will go to extinction if $\alpha > 0$ is fixed (i.e. does not increase sufficiently with n).

An empirical study of LV systems with changing interaction strengths. Equipped with results on the existence of a unique equilibrium, one pending question is

to understand what happens when the coefficient α varies for the same matrix A and the same parameter μ. In particular, when the value of α increases above a certain critical value, all species will coexist [BN21]; conversely, for sufficiently low values of α, the existence of a feasible equilibrium is not warranted anymore, however a unique and stable equilibrium may exist. How species equilibrium abundances change between these two states and how $|\mathcal{S}|$ varies will be the focus of Section 1.4.

Notations

Denote by $\rho(C)$ the spectral radius of matrix C, by $\|C\|$ the spectral norm of matrix C, and by $\|\boldsymbol{u}\|$ the euclidean norm of vector $\boldsymbol{u}$. We represent by δ_x the Dirac measure at x:

$$\delta_x(E) = \begin{cases} 1 & \text{if } x \in E \\ 0 & \text{else} \end{cases}.$$

We denote by $\xrightarrow{a.s.}$ the almost sure convergence of random quantities and by $\xrightarrow{weak}$ the weak convergence of measures. Given a set S, we denote by $|S|$ its cardinality.

1.2 Equilibrium and stability results

A primer on Random Matrix Theory. We first recall some results on Random Matrix Theory (RMT), which provides a number of valuable insights to understand the asymptotic behaviour of A. We begin by the almost sure (a.s.) convergence of the spectral radius and the spectral norm:

$$\rho(A/\sqrt{n}) \xrightarrow[n\to\infty]{a.s.} 1 \quad \text{and} \quad \|A/\sqrt{n}\| \xrightarrow[n\to\infty]{a.s.} 2\,.$$

We also have the a.s. weak convergence of the spectral measure of $A/\sqrt{n}$ to the circular law (see for instance [BC12]):

$$(a.s.) \quad \frac{1}{n}\sum_{k\in[n]} \delta_{\lambda_k(A/\sqrt{n})} \xrightarrow[n\to\infty]{weak} \frac{\mathbf{1}_{\{x^2+y^2\leqslant 1\}}}{\pi}\,dx\,dy\,,$$

where $(\lambda_k(A/\sqrt{n}); k \in [n])$ is the spectrum of $A/\sqrt{n}$. This convergence is illustrated in Fig. 1.1a.

The description of the spectral norm of the deterministic part of matrix B is more straightforward:

$$\left\| \mu \frac{\mathbf{1}\mathbf{1}^T}{n} \right\| = |\mu|\,.$$

Notice that both the random and deterministic parts of matrix B do not vanish asymptotically and thus have a macroscopic effect on the dynamics of system (1.1), as recalled in Remark 1.2 where the asymptotic spectrum of B is described.

The non-invadability condition. A key element to understand the dynamics of the LV system (1.1) is the existence of an equilibrium $\boldsymbol{x}^* = (x_k^*)_{k\in[n]}$ such that

$$\begin{cases} x_k^*\,(1 - x_k^* + (B\boldsymbol{x}^*)_k) = 0\,, \quad \forall k \in [n]\,, \\ x_k^* \geqslant 0. \end{cases} \tag{1.4}$$

and the study of its stability, that is the convergence of a solution $\boldsymbol{x}$ to the equilibrium $\boldsymbol{x}^*$: $\boldsymbol{x}(t) \xrightarrow[t\to\infty]{} \boldsymbol{x}^*$ if $\boldsymbol{x}(0)$ is sufficiently close to $\boldsymbol{x}^*$.

It is well known that for LV equations, the fact that $\boldsymbol{x}(0) > 0$ (componentwise) implies that $\boldsymbol{x}(t) > 0$ for every $t > 0$, but one can have some components $x_k(t)$ of $\boldsymbol{x}(t)$ vanishing to zero. As a consequence, we will only consider non negative equilibria $\boldsymbol{x}^* \geqslant 0$ with possibly vanishing components.

Notice that the situation substantially differs whether $\boldsymbol{x}^* > 0$ or $\boldsymbol{x}^*$ has vanishing component. In the former case, the equilibrium set of equations becomes a linear equation:

$$\boldsymbol{x}^* = \mathbf{1} + B\boldsymbol{x}^* \,.$$

In the latter case, the equilibrium equations are no longer linear.

In the centered case $\mu = 0$, the existence of a positive solution has been studied in [BN21] and requires $\alpha \gg \sqrt{2\log(n)}$ (while we consider α fixed here).

A naive and systematic way to solve (1.4) is to choose a priori a subset $\mathcal{I} \subset [n]$, to set the corresponding components $\boldsymbol{x}_{\mathcal{I}} = (x_i^*)_{i\in\mathcal{I}}$ to zero, and to solve the remaining linear system:

$$\boldsymbol{x}_{\mathcal{I}^c} = \mathbf{1}_{|\mathcal{I}^c|} + B_{\mathcal{I}^c}\boldsymbol{x}_{\mathcal{I}^c} \,.$$

If there exists $\boldsymbol{x}_{\mathcal{I}^c} \geqslant 0$ that solves the previous equation, then $\boldsymbol{x} = \begin{pmatrix} \boldsymbol{x}_{\mathcal{I}} \\ \boldsymbol{x}_{\mathcal{I}^c} \end{pmatrix}$ satisfies (1.4) and is a potential equilibrium. The number of subcases $\mathcal{I} \subset [n]$ is 2^n and in particular exponentially grows as $n \to \infty$.

In order to decrease the number of potential solutions to (1.4), we first notice that relying on standard properties of dynamical systems, see for instance [Tak96, Theorem 3.2.5], a necessary condition for the equilibrium $\boldsymbol{x}^*$ to be stable is that

$$1 - x_k^* + (B\boldsymbol{x}^*)_k \leqslant 0 \,. \tag{1.5}$$

The condition (1.5) is better known in ecology as the non-invadability condition [LM96]. In reference to the ODE (1.2), the requirement for a given species $k \in [n]$ to be non-invasive is equivalent to:

$$\left(\frac{1}{x_k}\frac{dx_k}{dt}\right)_{x_k\to 0^+} \leqslant 0 \,. \tag{1.6}$$

The main interpretation is as follows: if one adds species k with a very low abundance in the system, it will not be able to invade the system as a result of condition (1.6).

As a consequence, we will now focus on the following set of conditions:

$$\begin{cases} x_k^*\,(1 - x_k^* + (B\boldsymbol{x}^*)_k) &= 0 \quad \text{for } k \in [n]\,, \\ 1 - x_k^* + (B\boldsymbol{x}^*)_k &\leqslant 0 \quad \text{for } k \in [n]\,, \\ \boldsymbol{x}^* &\geqslant 0 \quad \text{componentwise}\,. \end{cases} \tag{1.7}$$

This casts the problem of finding a non negative equilibrium into the class of Linear Complementarity Problems (LCP), which we describe hereafter.

Linear Complementarity Problem (LCP). LCP is a class of problems from mathematical optimization which in particular encompasses linear and quadratic programs; standard references are [Mur88, CPS09]. Given a $n \times n$ matrix M and a $n \times 1$ vector

$\boldsymbol{q}$, the associated LCP denoted by $LCP(M, \boldsymbol{q})$ consists in finding two $n \times 1$ vectors $\boldsymbol{z}, \boldsymbol{w}$ satisfying the following set of constraints:

$$\left\{ \begin{array}{lcll} \boldsymbol{z} & \geqslant & 0\,, & \\ \boldsymbol{w} = M\boldsymbol{z} + \boldsymbol{q} & \geqslant & 0\,, & \\ \boldsymbol{w}^T\boldsymbol{z} & = & 0 \quad \Leftrightarrow \quad w_k z_k = 0 & \text{for all } k \in [n]\,. \end{array} \right. \tag{1.8}$$

Since $\boldsymbol{w}$ can be inferred from $\boldsymbol{z}$, we denote $\boldsymbol{z} \in LCP(M, \boldsymbol{q})$ if $(\boldsymbol{w}, \boldsymbol{z})$ is a solution of (1.8).

A theorem by Murty [Mur72] states that the $LCP(M, \boldsymbol{q})$ has a unique solution $(\boldsymbol{w}, \boldsymbol{z})$ iff M is a P-matrix, that is:

$$\det(M_{\mathcal{I}}) > 0\,, \quad \forall \mathcal{I} \subset [n]\,, \quad M_{\mathcal{I}} = (M_{k\ell})_{k,\ell \in \mathcal{I}}\,.$$

In view of (1.7), we look for $\boldsymbol{x}^* \in LCP(I - B, -\mathbf{1})$.

The equilibrium $\boldsymbol{x}^*$ and its stability. For a generic LV system

$$\frac{d\,y_k(t)}{dt} = y_k(r_k + (C\boldsymbol{y})_k)\,, \quad k \in [n]\,, \tag{1.9}$$

Takeuchi and Adachi (see for instance [Tak96, Th. 3.2.1]) provide a criterion for the existence of a unique equilibrium $\boldsymbol{y}^*$ and the global stability of the LV system.

Theorem 1.1 (Takeuchi and Adachi [TA80])**.** *If there exists a positive diagonal matrix Δ such that $\Delta C + C^T\Delta$ is negative definite, then $LCP(-C, \boldsymbol{r})$ admits a unique solution. In particular, for every $\boldsymbol{r} \in \mathbb{R}^n$, there is a unique equilibrium $\boldsymbol{y}^*$ to (1.9), which is globally stable in the sense that for every $\boldsymbol{y}_0 > 0$, the solution to (1.9) which starts at $\boldsymbol{y}(0) = \boldsymbol{y}_0$ satisfies*

$$\boldsymbol{y}(t) \xrightarrow[t\to\infty]{} \boldsymbol{y}^*\,.$$

Combining this result (setting $C = -(I-B)$) with results from RMT, we can guarantee the existence of a globally stable equilibrium $\boldsymbol{x}^*$ of (1.1) for a wide range of the set (α, μ). Denote by

$$\mathcal{A} = \left\{ (a, m) \in \mathbb{R}_+^* \times \mathbb{R} :\ a > \sqrt{2},\ m < \frac{1}{2} + \frac{1}{2}\sqrt{1 - \frac{2}{a^2}} \right\} \tag{1.10}$$

the set of *admissible parameters.*

Theorem 1.2. *Let $(\alpha, \mu) \in \mathcal{A}$, then a.s. matrix $(I - B) + (I - B)^T$ is eventually positive definite: with probability one, for a given realization ω, there exists $N(\omega)$ such that for $n \geqslant N(\omega)$, $(I - B^\omega) + (I - B^\omega)^T$ is positive definite. In particular, there exists a unique (random) globally stable equilibrium $\boldsymbol{x}^* \in LCP(I - B^\omega, -\mathbf{1})$ to (1.7).*

Optimal conditions for the stability of $\boldsymbol{x}^*$. In Theorem 1.2, we provided a sufficient condition, namely $(\alpha, \mu) \in \mathcal{A}$, which guarantees the stability of the LV system. If $\mu = 0$, this condition simply writes $\alpha > \sqrt{2}$.

A natural question is to find the optimal condition over α (we consider the extra condition $\mu = 0$) to have stability. Such a question is challenging and we explore a possible phase transition via numerical simulations. These simulations indicate that

$$\alpha^* = \frac{1}{\sqrt{2}}$$

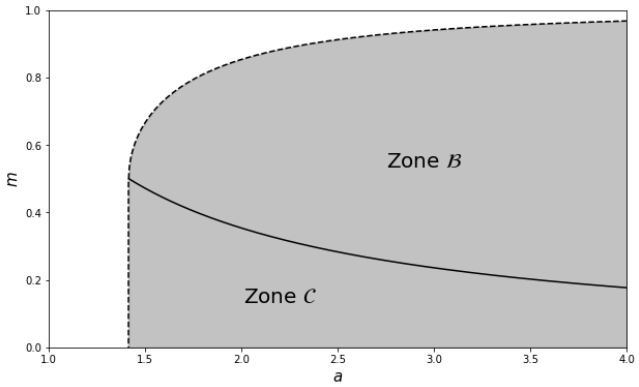


Figure 1.2: The shaded area represents the set $\mathcal{A}$ given by (1.10) yielding the existence of a unique (random) globally stable equilibrium $\boldsymbol{x}^*$. Area $\mathcal{A}$ is divided into two zones $\mathcal{B}$ and $\mathcal{C}$. Both zones correspond to parameters (α, μ) for which matrix $2I - (B + B^T)$ is definite positive, as stated in Theorem 1.2. In zone $\mathcal{B}$, $\lambda_{\max}(B + B^T)$ corresponds to a spiked eigenvalue (μ above the critical threshold $(\alpha\sqrt{2})^{-1}$). In zone $\mathcal{C}$, $\lambda_{\max}(B + B^T)$ corresponds to the right edge of the semi-circle law. Notice that zone $\mathcal{C}$ extends to negative values along the y-axis.

seems to be a threshold above which stability is granted and below which some species' abundances explode. This threshold already appears in [Bun17, Fig. 2].

To support this conjecture, we proceed with a Monte-Carlo experiment to compute the standard deviation, one of the many measures which can characterize early warning signals preceding a transition in an ecosystem [KDS+13].

Let α be fixed and P be the number of Monte-Carlo repetitions. Draw randomly the initial abundances and the interaction matrix B. Then let the dynamics of the LV system (1.2) run for a "sufficiently long" time T (to observe either an explosion of a species abundance or the convergence of the abundance vector). Let $s \in \mathbb{N}$ be the precision of the numerical scheme and $t_s = T$. Denote by

$$x_k^d(t_0),\ x_k^d(t_1), \cdots, x_k^d(t_s)$$

the time discretization of the dynamics of the abundance $x_k(t)$ of species k for $t \in [0, T]$ as obtained by the Runge–Kutta methods. Define by

$$SD_p = \frac{1}{n}\sum_{k=1}^{n}\left[\frac{1}{50}\sum_{i=s-50}^{s}(x_k^d(t_i))^2 - \left(\frac{1}{50}\sum_{i=s-50}^{s}x_k^d(t_i)\right)^2\right]^{1/2}$$

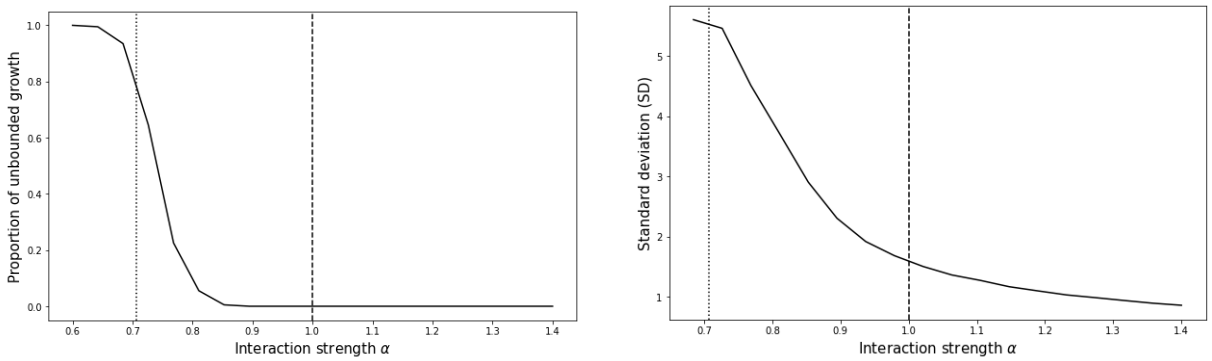
the standard deviation of experiment number p. Numerically, we say that the abundance of a species has unbounded growth if for some $k \in [n]$ there exists $s_0 \in [0, s]$ such that $x_k^d(t_{s_0})$ is no longer a number (NaN) indicating the explosion of the abundance of species k and thus the explosion of SD_p. In this case, we set

$$SD_p = \infty\,.$$

Consider now the two following indicators.

1. The global standard deviation:

$$SD = \frac{1}{P}\sum_{p=1}^{P} SD_p \mathbf{1}_{\{SD_p < \infty\}}\,.$$



(a) Proportion of unbounded growth. (b) Standard deviation

Figure 1.3: The dotted line represent the physicist threshold at $1/\sqrt{2}$ and the dashed line the expected P-matrix threshold at 1. In Fig. 1.3a, the proportion of unbounded growth as a function of the interaction strength (α) is displayed. In Fig. 1.3b, the standard deviation defined in (1.2) is illustrated as a function of the interaction strength (α).

2. The proportion of unbounded growth dynamics corresponding to

$$\frac{1}{P}\sum_{p=1}^{P}\mathbf{1}_{\{SD_p=\infty\}}\,.$$

The results of the numerical simulations are displayed in Figure 1.3. We observe that the proportion of unbounded growth significantly increases as it approaches the $\alpha^* = 1/\sqrt{2}$ threshold. The simulations show a non-linear increase of the standard deviation SD when α gets closer to α^* which can be seen as a early warning signal. Notice that when explosion occurs, the equilibrium remains unique and stable if $\alpha > \alpha^*$.

1.3 A heuristic approach to the proportion and distribution of the surviving species

1.3.1 Proportion of surviving species

In Section 1.2, we have presented conditions on parameters α, μ for the existence of a globally stable equilibrium $\boldsymbol{x}^*$ to (1.1) under the non-invadability condition. As $\boldsymbol{x}^*$ depends on the realization of matrix B, it is a random vector. Moreover since $\alpha > 0$ is fixed and does not depend on n, the equilibrium $\boldsymbol{x}^*$ will feature vanishing components (see the original argument in [DVR+18] and the discussion in [BN21]). In an ecological context, we shall refer to these non-vanishing components as the surviving species, the vanishing components corresponding to the species going to extinction with $x_k^* = 0$ and $x_k(t) \xrightarrow[t\to\infty]{} 0$.

In this section, we assume that the A_{ij}'s are $\mathcal{N}(0,1)$-distributed and describe the proportion of non-vanishing components of the equilibrium $\boldsymbol{x}^*$; we also describe the distribution of the surviving species $x_i^* > 0$ which turns out to be a truncated Gaussian.

Remark 1.3. The Gaussianity assumption facilitates the explanation of the heuristics but does not seem necessary for the result to hold. In Fig. 1.6b, the entries are not consid-

ered Gaussian but the distribution of the surviving species still matches the truncated Gaussian.

Given the random equilibrium $\boldsymbol{x}^*$, recall the definition of $\mathcal{S}$ in (1.3). We introduce the following quantities:

$$\hat{p} = \frac{|\mathcal{S}|}{n} , \qquad \hat{m} = \frac{1}{|\mathcal{S}|} \sum_{i \in [n]} x_i^* , \qquad \hat{\sigma}^2 = \frac{1}{|\mathcal{S}|} \sum_{i \in [n]} (x_i^*)^2 .$$

Notice that in the definitions of $\hat{m}$ and $\hat{\sigma}^*$ we can replace $\sum_{i\in[n]}$ by $\sum_{i\in\mathcal{S}}$.

Denote by $Z \sim \mathcal{N}(0,1)$ a standard Gaussian random variable and by Φ the cumulative Gaussian distribution function:

$$\Phi(x) = \int_{-\infty}^{x} \frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}} \, du .$$

Recall the definition of the set $\mathcal{A}$ in (1.10).

Heuristics 1.1. *Let $(\alpha, \mu) \in \mathcal{A}$. The following system of three equations and three unknowns (p, m, σ)*

$$\begin{aligned} \sigma\sqrt{p}\Phi^{-1}(1-p) + \alpha(1 + \mu\, p\, m) &= 0 , && (1.11)\\ 1 + \mu\, p\, m + \frac{\sigma\sqrt{p}}{\alpha}\mathbb{E}(Z \mid Z > -\delta) &= m , && (1.12)\\ (1 + \mu\, p\, m)^2 + (1 + \mu\, p\, m)\frac{2\sigma\sqrt{p}}{\alpha}\mathbb{E}(Z \mid Z > -\delta) & \\ + \frac{\sigma^2 p}{\alpha^2}\mathbb{E}(Z^2 \mid Z > -\delta) &= \sigma^2 && (1.13) \end{aligned}$$

where

$$\delta = \delta(p, m, \sigma) = \frac{\alpha}{\sigma\sqrt{p}}(1 + \mu p m) , \qquad (1.14)$$

admits a unique solution (p^*, m^*, σ^*) and

$$\hat{p} \xrightarrow[n\to\infty]{a.s.} p^* , \qquad \hat{m} \xrightarrow[n\to\infty]{a.s.} m^* \qquad \text{and} \qquad \hat{\sigma} \xrightarrow[n\to\infty]{a.s.} \sigma^* .$$

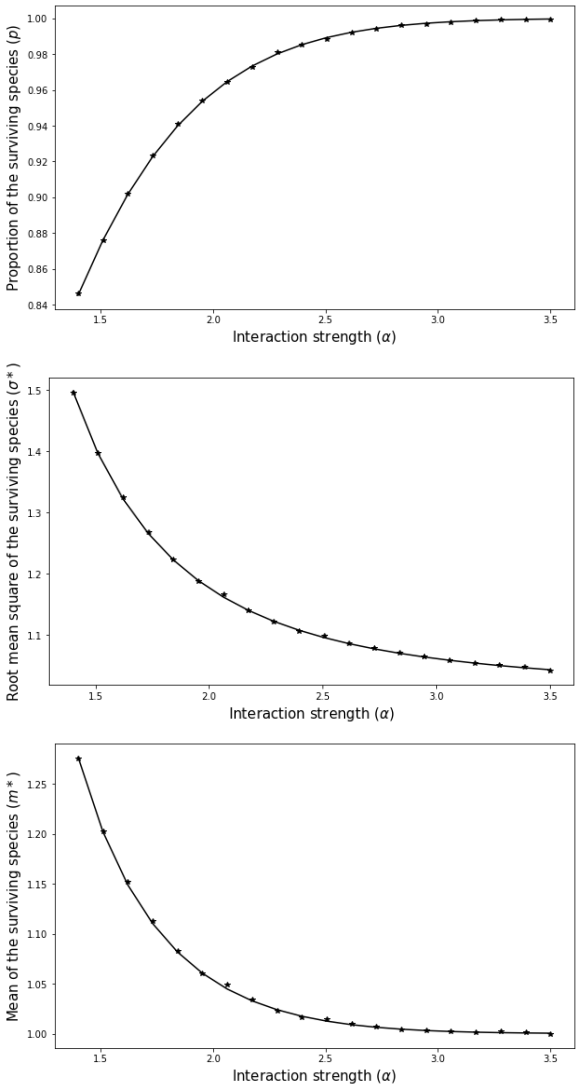
Associated to this solution (p^*, m^*, σ^*) is $\delta^* = \delta(p^*, m^*, \sigma^*)$.

There is a strong matching between the parameters obtained by solving (1.11)-(1.13) and their empirical counterparts obtained by Monte-Carlo simulations. This is illustrated in Fig. 1.4. In Fig. 1.5, we illustrate the sensitivity of σ^* to the parameters (α, μ).

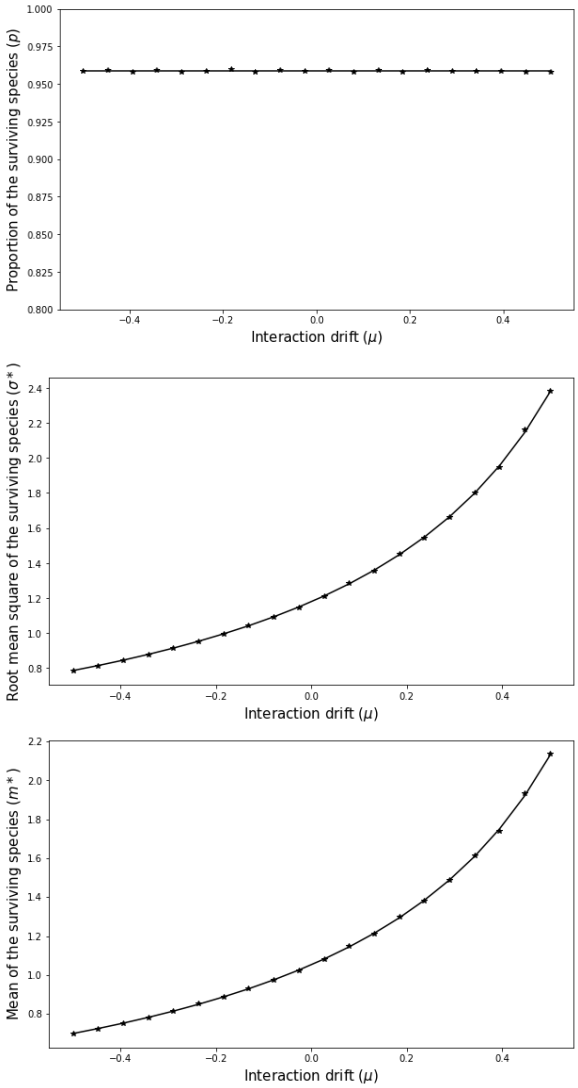
Remark 1.4. The heuristics above substantially simplifies in the centered model case, where $\mu = 0$ and $B = \frac{A}{\alpha\sqrt{n}}$. Following (1.20) – see Appendix 1.B, assume that $\alpha > \sqrt{2}$. Then the system with two unknowns (p, σ)

$$\begin{cases} \sigma\sqrt{p}\Phi^{-1}(1-p) + \alpha &= 0 \\ 1 + \frac{2\sigma\sqrt{p}}{\alpha}\mathbb{E}(Z \mid Z > -\delta) + \frac{\sigma^2 p}{\alpha^2}\mathbb{E}(Z^2 \mid Z > -\delta) &= \sigma^2 \end{cases} \quad \text{where} \quad \delta = \frac{\alpha}{\sigma\sqrt{p}}$$

admits a unique solution (p^*, σ^*). Moreover, $\hat{p} \xrightarrow[n\to\infty]{a.s.} p^*$ and $\hat{\sigma} \xrightarrow[n\to\infty]{a.s.} \sigma^*$.



(a) Parameters (p^*, σ^*, m^*) versus α.



(b) Parameters (p^*, σ^*, m^*) versus μ.

Figure 1.4: The plots represent a comparison between the theoretical solutions (p^*, σ^*, m^*) of (1.11)-(1.13) and their empirical Monte Carlo counterpart (the star marker) as functions of the interaction strength α (left) and the interaction drift μ (right). Matrix B has size $n = 500$ and the number of Monte Carlo experiments is 200. In Column (1.4a), $\mu = 0$ and $\alpha > \sqrt{2}$ on the x-axis (which guarantees a unique and stable equilibrium $\boldsymbol{x}^*$). When interaction α^{-1} increases, the number of surviving species p^* decrease but their variance σ^* and mean m^* increase. In Column (1.4b), $\alpha = 2$ and $\mu \in (-0.5, 0.5)$ on the x-axis. The interaction drift appears to have no impact on the proportion p^* of surviving species, whereas it influences their variances and means.

1.3.2 Distribution of surviving species

In the previous section, the proportion $\hat{p}$ of the surviving species, their mean $\hat{m}$ and second moment $\hat{\sigma}^2$ have been described as empirical counterparts of the solutions $p^*, m^*, (\sigma^*)^2$ of a system of equations. While establishing this system of equations, we will provide the following representation (see (1.23) in Appendix 1.C) of the abundance x_k^* of a surviving

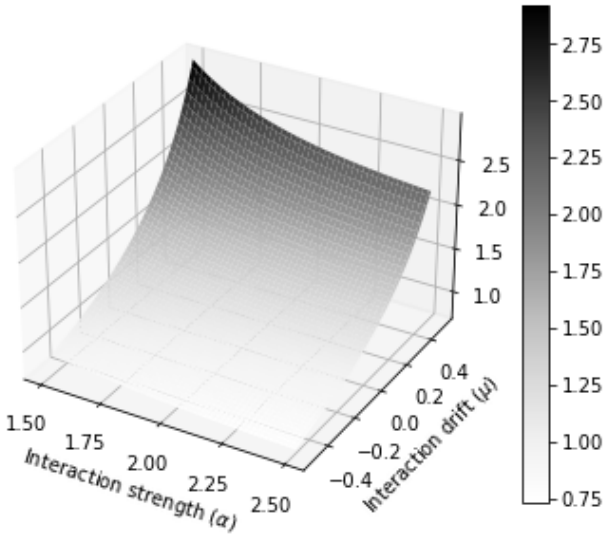


Figure 1.5: The 3D plot represents $\sigma^* = \sigma^*(\alpha, \mu)$, solution of the system (1.11)-(1.13). In contrast to the proportion of persisting species p^*, we observe that μ has a major influence over σ^*. The graph for the theoretical value of m^* has approximately the same behavior with respect to μ and α.

species:

$$x_k^* \;=\; 1 + \mu\, p^* m^* + \frac{\sigma^* \sqrt{p^*}}{\alpha} Z_k \,,$$

where $Z_k \sim \mathcal{N}(0,1)$ and $Z_k > -\delta^* = -\delta(p^*, m^*, \sigma^*)$, δ being defined in (1.14). We take here advantage of this representation to characterize x_k's distribution, which turns out to be a truncated Gaussian.

Heuristics 1.2. *Let $(\alpha, \mu) \in \mathcal{A}$, $\boldsymbol{x}^*$ the solution of (1.7) and let (p^*, m^*, σ^*) the solution of the system (1.11)-(1.13). Recall the definition (1.14) of δ and denote by $\delta^* = \delta(m^*, p^*, \sigma^*)$. Let $x_k^* > 0$ a positive component of $\boldsymbol{x}^*$, then:*

$$\mathcal{L}(x_k^*) \xrightarrow[n\to\infty]{} \mathcal{L}\left(1 + \mu p^* m^* + \frac{\sigma^* \sqrt{p^*}}{\alpha} Z \quad \middle| \quad Z > -\delta^*\right) \,,$$

where $Z \sim \mathcal{N}(0,1)$. Otherwise stated, asymptotically x_k^* admits the following density

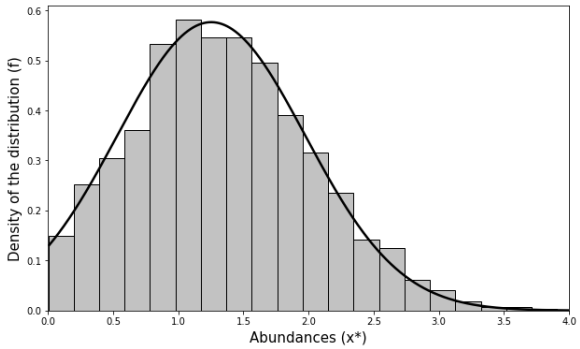
$$f(y) = \frac{\mathbf{1}_{\{y>0\}}}{\Phi(\delta^*)} \frac{\alpha}{\sigma^* \sqrt{2\pi\, p^*}} \exp\left\{-\frac{1}{2}\left(\frac{\alpha}{\sigma^* \sqrt{p^*}} y - \delta^*\right)^2\right\} \,.$$

The heuristics simply follows from the fact that if x_k^* is a surviving species then

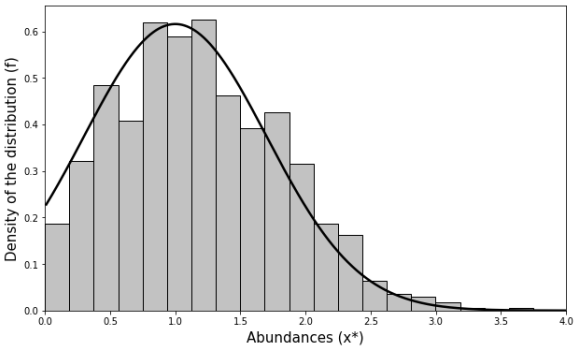
$$x_k^* = 1 + \mu\, p^* m^* + \frac{\sigma^* \sqrt{p^*}}{\alpha} Z_k$$

conditionally on the fact that the right hand side of the equation is positive, that is $Z_k > -\delta^*$. A simple change of variable yields the density - details are provided in Appendix 1.C.

Fig.1.6 illustrates the matching between the theoretical distribution obtained in Heuristics 1.2 and a histogram obtained by Monte-Carlo simulations. It also illustrates the validity of the heuristics beyond the gaussiannity assumption of the entries.



(a) Gaussian entries.



(b) Uniform entries.

Figure 1.6: Distribution of surviving species. The x-axis represents the value of the abundances and the histogram is built upon the positive components of equilibrium $\boldsymbol{x}^*$. The solid line represents the theoretical distribution for parameters (α, μ) as given by Heuristics 1.2. In Fig. (1.6a), the entries are Gaussian $\mathcal{N}(0,1)$ and the parameters are set to $(n = 2000, \alpha = 2, \mu = 0.2)$. In Fig. (1.6b), the entries are uniform $\mathcal{U}(-\sqrt{3}, \sqrt{3})$ with variance 1 and the parameters are set to $(n = 2000, \alpha = \sqrt{3}, \mu = 0)$. Notice in particular that the theoretical distribution matches with non-Gaussian entries.

1.4 Switching between equilibria: changing interaction strength

In the previous sections, the strength α of the interactions was fixed, cf. equation (1.2). However, in nature interactions between species are constantly changing due to e.g. abiotic factors such as temperature, which affect the rate at which individuals forage for prey, etc. Our contribution is rooted in the framework of asymptotic dynamics, but many recent ecological studies highlight the importance of taking into account both transient dynamics (out-of-equilibrium abundance values due to frequent perturbations) and shifts between equilibria due to changing environmental conditions [Has01, FN11, NA16]. In the sequel, we discuss a more general framework.

The model and intuition. If we restrict ourselves to a specific environment, a possible ecological interpretation of the fluctuation of interaction strength corresponds to the relationship between the size of the habitat and the probability of contact between individuals from two interacting species (e.g. think of a pool of freshwater containing piscivorous fishes and their prey species – interactions, be them competition or predation, would be potentially more frequent if the volume of water was reduced). In physics, think of particles in motion in a given volume: if the number of particles and the temperature stay constant, reducing the volume should increase the number of interactions between particles.

From a model standpoint, let μ be fixed, $\alpha = \alpha(t) : \mathbb{R}^+ \to (\sqrt{2}, \infty)$. Consider

$$\frac{dx_k}{dt} = x_k \left(1 - x_k + (B_t \boldsymbol{x})_k\right) \ , \quad k \in [n] \ , \tag{1.15}$$

where matrix B_t admits the following representation

$$B_t = \frac{A}{\alpha(t)\sqrt{n}} + \frac{\mu}{n}\mathbf{1}\mathbf{1}^T \quad \text{and} \quad (\alpha(t), \mu) \in \mathcal{A} \, .$$

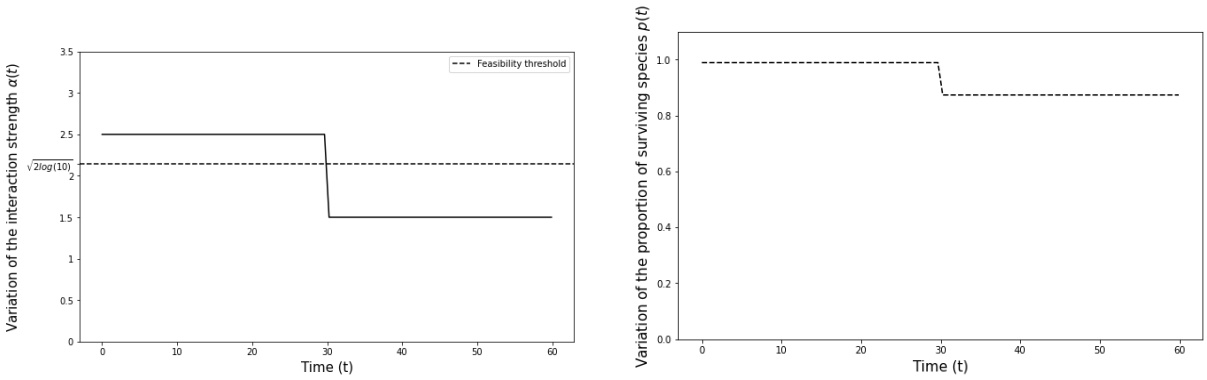
Remark 1.5. Following Theorem 1.2, condition $(\alpha(t), \mu) \in \mathcal{A}$ guarantees that there exists a unique equilibrium $\boldsymbol{x}^*(t)$ for every $t \in \mathbb{R}^+$.

We focus on the case of a system that fluctuates between two equilibrium points (Figure 1.7a). We consider a sudden incident, most often irreversible in the short term, which reduces a portion of habitat, e.g. forest fires. The system transits from a feasible state to a state with vanishing species due to the change of the strength of interactions, modelled by the following *step function*:

$$\alpha(t) \;=\; \alpha_1 \mathbf{1}_{[0,t_0)} + \alpha_2 \mathbf{1}_{[t_0,+\infty)}, \quad (\alpha_1, \alpha_2, t_0) \in (\sqrt{2}, +\infty)^2 \times \mathbb{R}_+ \tag{1.16}$$

The change of model parameter occurs at t_0 which causes a change in the strength of the interactions going from a value α_1 to α_2. One may choose $\alpha_2 < \alpha_1$ and the difference (or ratio) between the two values represents the intensity of the incident.

In large dimension, it is possible to characterize this change by its impact on the number of surviving species in the system (1.15). At a given time t, the proportion of surviving species $p = p(t) \in [0, 1]$ can be computed by resolving the system in Heuristics 1.1. This function, associated to the step function α given in (1.16), is represented in Figure 1.7b.



(a) Step function $\alpha(t)$ (b) Proportion of surviving species $p(t)$

Figure 1.7: (a) Variation of the interaction strength through time, used in the dynamics of a ten-species system depicted in Figure 1.8 ($\alpha_1 = 2.5$, $\alpha_2 = 1.5$). The dotted line represents the feasibility threshold associated to the system.
(b) Variation of the proportion of surviving species depending of the variation of $\alpha(t)$ in (a).

In the case of a sudden incident, the proportion of surviving species predicted by the heuristics has a form similar to $\alpha(t)$ i.e. a step response. In the feasible state, $p(t)$ is close to 1 (i.e. all species coexist); after the transition occurs, some species vanish and here $p(t) \approx 0.87$. Beware that the heuristics results follow instantaneously the change of α; however, there is a smoother transition in the dynamics between the two equilibria (respectively corresponding to α_1 and α_2) due to the return rate to equilibrium, see for instance [NC97], [ABLH18]). This transition is illustrated in Figure 1.8.

Simulations. We provide hereafter a simulation-based analysis of the impact of the sudden incident on a given ecosystem: Define a ten-species system (1.15) with a fixed

matrix of interactions A with Gaussian entries $\mathcal{N}(0,1)$ and consider $\alpha = \alpha(t)$ as in Fig. 1.7a.

This scenario has a mixed impact on the community, see Fig. 1.8. While some species benefit from this change through an increase in their abundances, others are severely affected by this shift, some of which become extinct. This phenomenon can be understood as follows: at first ($t \leqslant t_0$), the system admits a feasible equilibrium state with $\alpha = \alpha_1 = 2.5 > \sqrt{2\log(10)} \simeq 2,14$ and the abundances converge to this equilibrium (see Figure 1.7a). When the transition occurs at $t = t_0$, Theorem 1.2 ensures the convergence to a new equilibrium defined by parameter α_2. Since $\alpha_2 = 1.5$ is below the feasibility threshold $\sqrt{2\log(10)} \simeq 2.14$, some species vanish. In other words, this sudden change of model parameter causes an increase of interaction strengths, which has a negative impact on species diversity.

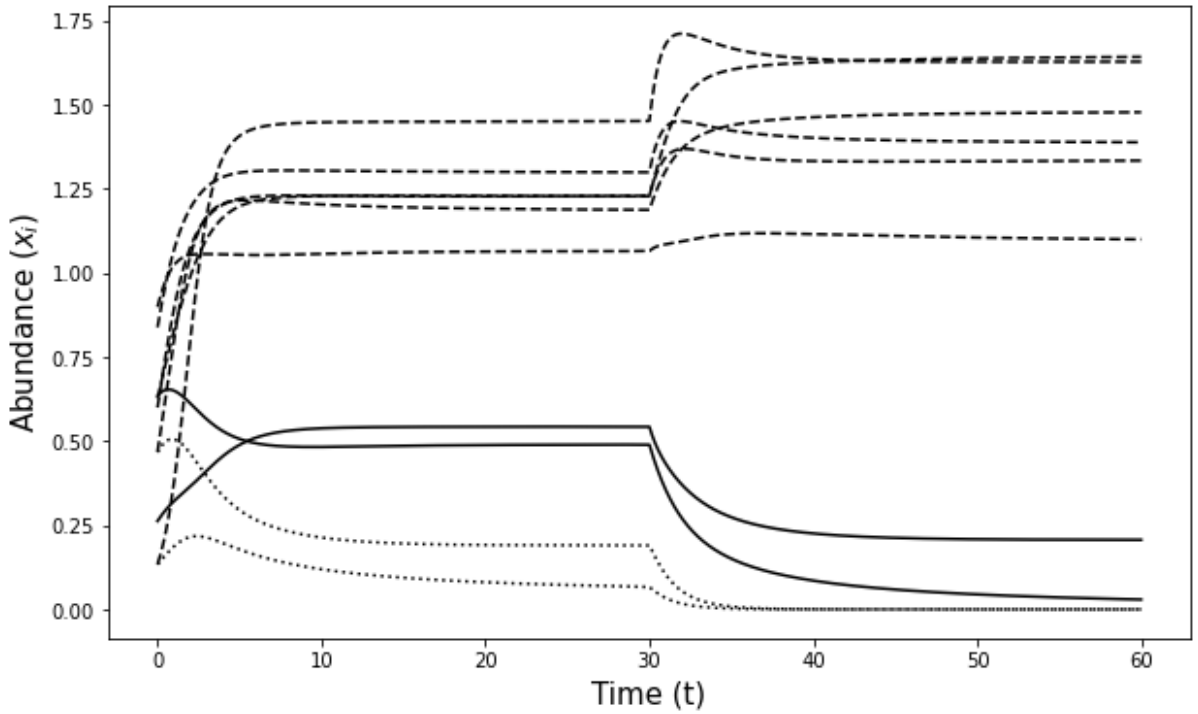


Figure 1.8: Abundance dynamics in the case of a community of ten species. The matrix of interactions A and the initial conditions are common and we apply the function of variation $\alpha(t)$ given in Figure 1.7a. The dashed lines represents species which benefit from habitat variation; solid lines represent species suffering from the change. Dotted lines represent species undergoing extinction.

Evolution of diversity Finally, we illustrate the evolution of diversity using diversity indicators more suited to the description of changes such as the one represented in Fig. 1.8 [Jos06]. We introduce here Shannon diversity H', a standard measure of biodiversity in ecology, which is given by

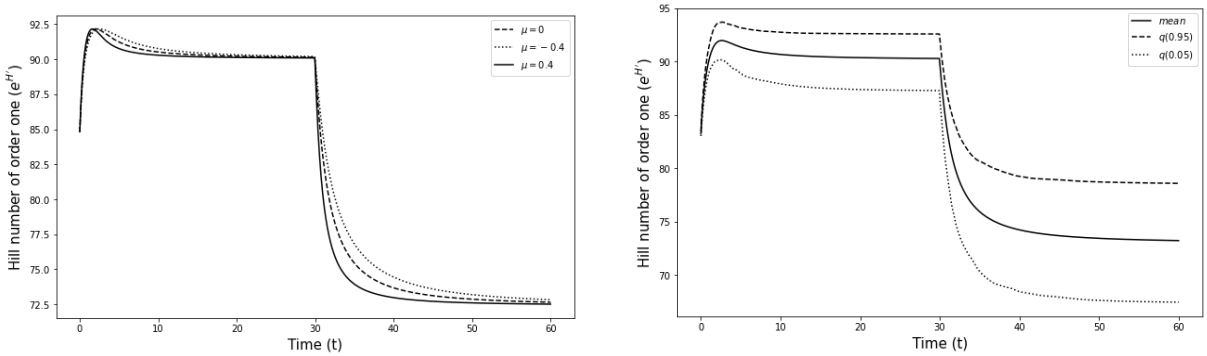
$$H' = -\sum_i \frac{x_i}{\sum_j x_j} \log\left(\frac{x_i}{\sum_j x_j}\right) \tag{1.17}$$

and ranges from 0 (one species completely dominates the community) to $\log(n)$, when each species is equally abundant. When many species become rare while others become more abundant, H' drops. Because H' varies before species actually vanish, it is a more sensitive index of community diversity than species richness.

The Hill number of order 1, defined as $e^{H'}$, is a diversity measure that takes into account species abundances and varies between 1 and n, i.e. it behaves like an "effective species richness", see e.g. [Jos07].

In Fig. 1.9a, we represent the mean of this diversity measure over time for a hundred-species system and observe a negative impact of the variation of the strength of the interactions on diversity. Parameter μ has no impact on diversity at equilibrium (similarly, μ has no impact on the proportion of persistent species), but the lower the value of μ, the slower the transition to a new equilibrium. In other words, the more generally "competitive" the ecosystem is (i.e. very negative values of μ), the longer it takes for transient dynamics to settle near equilibria.

The evolution of the Hill number of order 1 complements the evolution of species richness: when α decreases, the expected number of surviving species decreases (Fig. 1.7); at the same time, $e^{H'}$ decreases even more drastically as the abundance distribution of surviving species gets more heterogeneous. Figure 1.9b also shows that the variability of the Hill number of order 1 among simulations increases drastically when α decreases. The conclusion is that the more species are lost, the more difficult it is to predict the diversity index as σ^* depends on α and strongly influences $e^{H'}$.



(a) Impact of μ

(b) Variability of Hill number $e^{H'}$

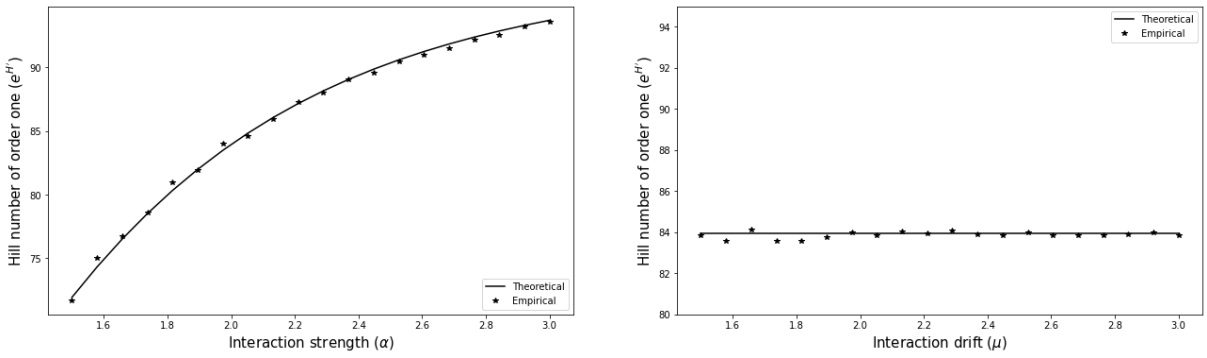
Figure 1.9: Dynamics of the Hill number of order 1 in the case of an ecosystem of a hundred species. The initial conditions are similar for each species. We define an interaction matrix A and let the dynamics of Lotka-Volterra evolve according to model (1.15) and we apply the function of variation $\alpha(t)$ of Figure 1.7a. For each time step, we compute $e^{H'}$. We repeat this scheme a large number of times (here 500), and we average the time series. In (a), we apply this procedure for different values of μ. In (b), we apply this procedure for a fixed $\mu = 0$ and compute the quantiles of the 500 trajectories.

Theoretical estimation of diversity. Standard mathematical methods (Taylor's theorem) can be used to obtain a theoretical approximation of the Hill number of order 1 (details are provided in Appendix 1.D.1):

$$e^{H'} \approx np^* \left(\frac{3}{2} - \frac{1}{2} \frac{(\sigma^*)^2}{(m^*)^2} \right) . \tag{1.18}$$

This estimator is based on the properties of the persistent species (p^*, m^*, σ^*) calculated by solving the fixed point equation of the heuristics (1.1). These three properties depend on the type of the interactions, as indicated by parameters (α, μ) (Figure 1.4). We compare the accuracy of this estimator with two examples in which the strength of the interactions (α) and the interaction drift (μ) vary (Figure 1.10).

On the one hand, a shift of the interaction drift μ does not affect the proportion of surviving species. Furthermore, the impact of μ on σ^* and m^* is proportional i.e. $\frac{\sigma^*}{m^*}$ is equal to a constant. For these reasons, μ does not affect the Hill number (Figure 1.10b). On the other hand, if α increases, then $p^*, m^*, \sigma^* \xrightarrow[\alpha \to +\infty]{} 1$ and $e^{H'} \to n$ which is makes sense because when α becomes very large, all abundances converge to 1. If α decreases: p^* decreases, and σ^* increases faster than m^*. This confirms that $e^{H'}$ decreases even more drastically as the abundance distribution of surviving species gets more heterogeneous (Figure 1.10a).



(a) Impact of α

(b) Impact of μ

Figure 1.10: Evolution of the Hill number of order 1 as a function of α (a) and μ (b). The theoretical solutions (solid line) are computed by resolving (1.1) and integrating the parameter (p^*, σ^*, m^*) in equation (1.18). The empirical solutions (star marker) are computed by a Monte-Carlo experiment (100 experiments): we define a matrix B of size 100×100, solve the LCP problem and calculate the associated Hill number $e^{H'}$ using (1.17).

1.5 Discussion

In this paper, our main interest was to describe the impact of the strength α and mean μ of interactions in large LV models on the conditions of coexistence of interacting species. Combining results from Takeuchi [Tak96] with standard RMT results, we have provided insights into the study of stability of large random ecosystems - see [Sto18, GGRA18].

We have characterized the unique equilibrium properties of the surviving species by resolving a system of equations. From a physicist point of view, similar equations were obtained by Opper and Diederich [OD92] and studied in a more general framework by Bunin using the dynamical cavity method [Bun17] and Galla [Gal18] using generating functional techniques.

The coexistence of many species in random ecosystems was also studied by Servan *et al.* [SCG+18] and Pettersson *et al.* [PSNJ20], where a more generic case was analyzed

with different growth rates. Grilli *et al.* [GAS+17] identified the key quantities regulating the parameter space leading to feasible communities. In contrast to previous approaches, an important feature of our model is the monitoring of the impact of interactions by a normalization factor ($\alpha\sqrt{n}$). From an ecological point of view, one might expect that the larger the number of species, the weaker the interactions will be due to some dilution of interactions among potential interaction partners, which would justify the use of such normalization factors. From a mathematical viewpoint, the normalizing parameter α captures the range of a unique equilibrium and the threshold for feasibility.

In nature, interactions between species are constantly changing and affected by the environment. Under the assumptions that environmental conditions influence interaction strengths, we have endeavoured in Section 1.4 to study the consequences of a sudden change of environmental conditions, expressed through a decrease in parameter α. Solving numerically the Lotka-Volterra system confirms the predictions given by heuristics, i.e. that a decrease in α negatively affects equilibrium species richness. A more precise representation of biodiversity dynamics can be obtained through Hill numbers of order 1 which also decreases after the sudden change in α. The dynamics of this diversity measure suggests that the mean of interaction coefficients, μ, affects the duration of transient dynamics, with shorter transient dynamics being associated with more mutualistic interactions (i.e. higher positive values of μ).

Many questions naturally arise as a follow-up. First, a mathematical proof of the heuristics presented here is a challenging prospect because the LCP procedure induces a statistical dependence that is a priori difficult to handle. However, looking for this proof will certainly help extend the results to other underlying assumptions on the parameters of the LV system.

Regarding the extension of the heuristics to other assumptions, two situations could be of particular interest: non-centered elliptical matrix models as in Bunin [Bun17] and LV models in which species growth rates are also controlled as in [SCG+18]. We are confident that such extensions are possible, but might hinge on more sophisticated developments, in particular to include growth rates in the calculus of order statistics.

In this paper we have only considered the case of a full interaction matrix with parameters (α, μ). However, food webs are often structured in compartments [BDB+11] and/or obey hierarchies (e.g. larger species eat smaller ones) [BAB+19]. By a numerical analysis of LV systems, one could use the same tools to study more patterned matrices [AT12]. Recent studies emphasize the sparsity of real food webs [BSHM17]. Beyond the feasibility studied by Akjouj and Najim [AN21], one could also study the existence and stability of a unique equilibrium in a sparse context.

Finally, variations of the interaction strength highlight the impact of habitat destruction. Many theoretical studies provide mathematical formulas for the return rate to equilibrium [NC97, ABLH18]. A further theoretical study of model (1.15) could provide a more quantitative answer. In this article, we have limited the analysis to the case of a single sudden incident, but other types of fluctuations for the interaction strength could be considered for a better understanding of habitat conservation phenomena. For example, a seasonal model could be appropriate to describe the evolution of the dynamics over the seasons.

Appendix

1.A Simulation details

Simulations were performed in Python. All the figures and the code are available on Github [Cle22a].

Simulations on the properties of surviving species are performed in two different ways. The theoretical solutions are obtained resolving numerically the system of equations of heuristics 1.1. We use a solver (cf. scipy.optimize) to find a local minimum of the function defined by the system of equations (a modification of the Powell hybrid method). The empirical solutions are computed using a Monte Carlo experiment. We simulate a large number of matrix matrix B, we resolve the associated LCP problem using the Lemke's algorithm. Then, we use the LCP solution to calculate the properties of the surviving species: proportion of survivors, etc. Finally, we make an average on the ensemble of experiments. The Lemke algorithm is implemented in the lemkelcp package and can be found on Github [Lam19]. The dynamics of the Lotka-Volterra are achieved by a Runge-Kutta of order 4 (RK4) implemented in the code.

1.B Proof of Theorem 1.2

We have

$$I - B + I - B^T = 2I - (B + B^T) = 2I - \left(\frac{A + A^T}{\alpha\sqrt{n}} + \frac{2\mu}{n}\mathbf{1}\mathbf{1}^* \right) .$$

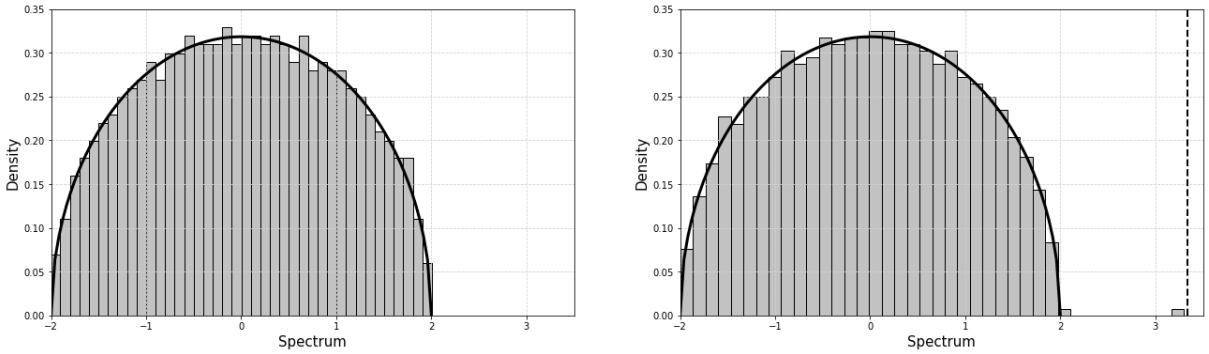
Notice that $2I - (B + B^T)$ is positive definite iff the top eigenvalue of $B + B^T$ is lower than 2:

$$\lambda_{\max}(B + B^T) < 2 \,. \tag{1.19}$$

We first focus on the random part $(A + A^T)/\alpha$ which is a symmetric matrix with independent $\mathcal{N}(0, 2/\alpha^2)$ entries above the diagonal (note that the distribution of the diagonal entries is different from the off-diagonal entries, with no asymptotic effect). In this case, it is well known that the largest eigenvalue of the normalized matrix (or equivalently its spectral norm since the matrix is symmetric) a.s. converges to the right edge of the support of the semi-circle law (see [BS10, Th. 5.2]):

$$\lambda_{\max}\left(\frac{A + A^T}{\alpha\sqrt{n}} \right) \xrightarrow[n\to\infty]{a.s.} \frac{2\sqrt{2}}{\alpha} \,. \tag{1.20}$$

In the centered case ($\mu = 0$), condition (1.19) occurs if $\alpha > \sqrt{2}$.



(a) No outlier if $\mu \leqslant (\alpha\sqrt{2})^{-1}$.

(b) Outlier if $\mu > (\alpha\sqrt{2})^{-1}$.

Figure 1.B.1: Spectrum (histogram) of the Hermitian random matrix $B + B^T$ ($n = 1000$, $\alpha = \sqrt{2}$). The solid line represents the semi-circular law. In Fig. 1.B.1a, $\mu = 0$ and there is no oulier. In Fig. 1.B.1b, $\mu = 1.5$ and one can notice the presence of an eigenvalue outside the bulk of the semi-circular law. The dashed line indicates its theoretical value.

We now consider the general case where $\mu \neq 0$. Notice that the trend matrix $P = \frac{2\mu}{n}\mathbf{1}\mathbf{1}^*$ admits a unique non zero eigenvalue 2μ. Denote by $\check{A} = \frac{A+A^T}{\alpha\sqrt{n}}$. We are interested in the top eigenvalue of the symmetric matrix $\check{A} + P$. Based on a result by Capitaine *et al.* [CDMF09, Th. 2.1], we have:

$$\lambda_{\max}(\check{A} + P) \xrightarrow[n\to\infty]{a.s} \begin{cases} 2\mu + \frac{1}{\alpha^2\mu} & \text{if } \mu > \frac{1}{\sqrt{2}\alpha}\,, \\ \frac{2\sqrt{2}}{\alpha} & \text{else.} \end{cases}$$

This result is illustrated in Figure 1.B.1.

Assume first that $\mu \leqslant \frac{1}{\alpha\sqrt{2}}$ (corresponding to zone $\mathcal{C}$ in Fig. 1.2), then $\lambda_{\max}(\check{A} + P) \xrightarrow[n\to\infty]{a.s.} \frac{2\sqrt{2}}{\alpha}$, which is strictly lower than 2 (cf. condition (1.19)) if $\alpha > \sqrt{2}$. Hence $\lambda_{\max}(\check{A} + P)$ is eventually strictly lower than 2 under this condition.

Assume now that $\mu > \frac{1}{\alpha\sqrt{2}}$ (corresponding to zone $\mathcal{B}$ in Fig. 1.2), then

$$\lambda_{\max}(\check{A} + P) \xrightarrow[n\to\infty]{a.s.} 2\mu + \frac{1}{\alpha^2\mu}\,.$$

We are interested in the conditions for which $2\mu + \frac{1}{\alpha^2\mu} < 2$ or equivalently

$$2\alpha^2\mu^2 - 2\alpha^2\mu + 1 < 0\,. \tag{1.21}$$

An elementary study of the polynomial $\xi(X) = 2\alpha^2X^2 - 2\alpha^2X + 1$ yields that ξ's discriminant is positive if $\alpha > \sqrt{2}$,

$$\xi(\mu^\pm) = 0 \quad \Leftrightarrow \quad \mu^\pm = \frac{1}{2} \pm \frac{1}{2}\sqrt{1 - \frac{2}{\alpha^2}}\,,$$

and $\xi\left(\frac{1}{\alpha\sqrt{2}}\right) < 0$, so that $\frac{1}{\alpha\sqrt{2}} \in (\mu^-, \mu^+)$. In particular condition (1.21) is fulfilled if

$$\mu \in \left(\frac{1}{\alpha\sqrt{2}}\,,\ \frac{1}{2} + \frac{1}{2}\sqrt{1 - \frac{2}{\alpha^2}}\right)\,.$$

Under this condition, (1.21) is fulfilled and a.s. $\limsup_{n\to\infty} \lambda_{\max}(\check{A} + P) < 2$, which completes the proof: we can then rely on Theorem 1.1 to conclude.

1.C Construction of the heuristics

We first discuss Heuristics 1.1 and establish Equations (1.11), (1.12) and (1.13).

1.C.1 Equation (1.11).

We first recall a result on order statistics of a Gaussian sample. Consider a family $(Z_k)_{k\in[n]}$ of i.i.d. random variables $\mathcal{N}(0,1)$ and the associated order statistics

$$Z_1^* \leqslant Z_2^* \leqslant \cdots \leqslant Z_n^* .$$

Consider an index $\lfloor n\alpha \rfloor \in [n]$ where $\alpha \in (0,1)$ is fixed, then the typical location of $Z^*_{\lfloor n\alpha\rfloor}$ is $\Phi^{-1}(\alpha)$:

$$Z^*_{\lfloor n\alpha\rfloor} \simeq \Phi^{-1}(\alpha) \quad \text{as} \quad n \to \infty , \tag{1.22}$$

see for instance [Smi49, BDH78].

Let $\boldsymbol{x}^*$ be the equilibrium of (1.1) and consider the random variable

$$\check{Z}_k = \sum_{i\in\mathcal{S}} B_{ki} x_i^* = (B\boldsymbol{x}^*)_k .$$

We assume that asymptotically the x_i^*'s are independent from the B_{ki}'s, an assumption supported by the chaos hypothesis, see for instance Geman and Hwang [GH82]. Denote by $\mathbb{E}_{\boldsymbol{x}^*} = \mathbb{E}(\,\cdot\,|\,\boldsymbol{x}^*)$ the conditional expectation with respect to $\boldsymbol{x}^*$. Notice that conditionally to $\boldsymbol{x}^*$, the $\check{Z}_k$'s are independent Gaussian random variables, whose two first moments can easily be computed, see Section 1.C.2 below for the details:

$$\mathbb{E}_{\boldsymbol{x}^*}\check{Z}_k = \mu\,\hat{p}\,\hat{m} \quad \text{and} \quad \mathrm{var}_{\boldsymbol{x}^*}(\check{Z}_k) = \frac{\hat{p}\hat{\sigma}^2}{\alpha^2} .$$

Notice that the fact that $\mathbb{E}_{\boldsymbol{x}^*}$ and $\mathrm{var}_{\boldsymbol{x}^*}(\check{Z}_k)$ only depend on $\hat{p},\hat{\sigma}$ and $\hat{m}$ which are (supposedly) converging quantities supports the idea that $\check{Z}_k$ is unconditionally a Gaussian random variable with moments:

$$\mathbb{E}\check{Z}_k = \mu\,p^*\,m^* \quad \text{and} \quad \mathrm{var}(\check{Z}_k) = \frac{p^*(\sigma^*)^2}{\alpha^2} ,$$

where p^*, m^*, σ^* are resp. the limits of $\hat{p},\hat{m},\hat{\sigma}$. We now introduce the standard Gaussian random variables $(Z_k)_{k\in[n]}$ where

$$Z_k = \frac{\check{Z}_k - \mathbb{E}\check{Z}_k}{\sqrt{\mathrm{var}(\check{Z}_k)}} = \alpha\frac{\check{Z}_k - \mu\,p^*\,m^*}{\sigma^*\sqrt{p^*}} .$$

Consider the equilibrium $\boldsymbol{x}^* = (x_k^*)_{k\in[n]}$. If $k \in \mathcal{S}$, that is $x_k^* > 0$, we have

$$1 - x_k^* + (B\boldsymbol{x}^*)_k = 0 \quad \Rightarrow \quad 1 + (B\boldsymbol{x}^*)_k > 0 .$$

This identity has two implications:

$$x_k^* = 1 + (B\boldsymbol{x}^*)_k \quad \text{and} \quad 1 + (B\boldsymbol{x}^*)_k > 0 \qquad \text{if } k \in \mathcal{S} .$$

Relying on the representation $(B\boldsymbol{x}^*)_k = \check{Z}_k$, we obtain the representation

$$x_k^* = 1 + (B\boldsymbol{x}^*)_k \;=\; 1 + \mu\, p^*\, m^* + \frac{\sigma^*\sqrt{p^*}}{\alpha} Z_k \qquad \text{if} \quad k \in \mathcal{S}\,. \tag{1.23}$$

and the condition:

$$1 + (B\boldsymbol{x}^*)_k \;=\; 1 + \mu\, p^*\, m^* + \frac{\sigma^*\sqrt{p^*}}{\alpha} Z_k \;>\; 0\,.$$

If $k \notin \mathcal{S}$ then

$$1 + (B\boldsymbol{x}^*)_k \;=\; 1 + \mu\, p^*\, m^* + \frac{\sigma^*\sqrt{p^*}}{\alpha} Z_k \;\leqslant\; 0$$

by the non invadability condition. Otherwise stated,

$$\begin{cases} Z_k \;\leqslant\; -\frac{\alpha(1+\mu\, p^* m^*)}{\sigma^*\sqrt{p^*}} & \text{if } k \notin \mathcal{S}\,, \\ Z_k \;>\; -\frac{\alpha(1+\mu\, p^* m^*)}{\sigma^*\sqrt{p^*}} & \text{if } k \in \mathcal{S}\,. \end{cases}$$

Considering the order statistics of the Z_k's we obtain:

$$Z_1^* \leqslant \cdots \leqslant Z_{\boldsymbol{i}}^* \leqslant -\frac{\alpha(1+\mu\, p^* m^*)}{\sigma^*\sqrt{p^*}} \leqslant Z_{\boldsymbol{i}+1}^* \leqslant \cdots \leqslant Z_n^*\,.$$

Now, there are exactly $n - |\mathcal{S}| = n(1-\hat{p})$ indices before the threshold corresponding to the components of $\boldsymbol{x}^*$ equal to zero. In particular, index $\boldsymbol{i} = n(1-\hat{p})$ corresponds to the value

$$Z_{\boldsymbol{i}}^* \simeq -\frac{\alpha(1+\mu\, p^*\, m^*)}{\sigma^*\sqrt{p^*}}$$

Relying on (1.22), we finally obtain

$$\Phi^{-1}(1-\hat{p}) = -\frac{\alpha(1+\mu\, p^*\, m^*)}{\sigma^*\sqrt{p^*}}\;.$$

It remains to replace $\hat{p}$ by its limit p^* to obtain (1.11).

1.C.2 Details on Equation (1.11): Moments of $\check{Z}_k$.

We compute hereafter the conditional mean and variance of $\check{Z}_k = (B\boldsymbol{x}^*)_k$ with respect to $\boldsymbol{x}^*$. We rely on the following identities:

$$\mathbb{E}B_{ki} = \frac{\mu}{n}\,, \quad \mathbb{E}(B_{ki})^2 = \frac{1}{\alpha^2 n} + \frac{\mu^2}{n^2} \simeq \frac{1}{\alpha^2 n}\,, \quad \mathbb{E}B_{ki}B_{kj} = \frac{\mu^2}{n^2} \quad (i \neq j)\,.$$

We first compute the conditional mean:

$$\begin{aligned} \mathbb{E}_{\boldsymbol{x}^*}(\check{Z}_k) = \sum_{i\in[n]} \mathbb{E}(B_{ki})x_i^* = \sum_{i\in\mathcal{S}} \mathbb{E}(B_{ki})x_i^* &= \frac{\mu}{n}\sum_{i\in\mathcal{S}} x_i^*, \\ &= \mu\frac{|\mathcal{S}|}{n}\frac{1}{|\mathcal{S}|}\sum_{i\in\mathcal{S}} x_i^*, \\ &= \mu\,\hat{p}\,\hat{m}\,. \end{aligned}$$

We now compute the second moment:

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{x}^*}(\check{Z}_k^2) &= \mathbb{E}_{\boldsymbol{x}^*}\left(\sum_{i\in\mathcal{S}} B_{ki}x_i^*\right)^2 = \mathbb{E}_{\boldsymbol{x}^*}\sum_{i,j\in\mathcal{S}} B_{ki}B_{kj}x_i^*x_j^*\,,\\
&= \sum_{i\in\mathcal{S}}\mathbb{E}(B_{ki}^2)x_i^{*2} + \sum_{i\neq j}\mathbb{E}(B_{ki}B_{kj})x_i^*x_j^*,\\
&= \frac{1}{\alpha^2 n}\sum_{i\in\mathcal{S}} x_i^{*2} + \frac{\mu^2}{n^2}\sum_{i\neq j} x_i^*x_j^*,\\
&\overset{(a)}{\simeq} \frac{\hat{p}\hat{\sigma}^2}{\alpha^2} + \frac{\mu^2\,\hat{p}^2}{|\mathcal{S}|^2}\sum_{i,j\in\mathcal{S}} x_i^*x_j^* = \frac{\hat{p}\hat{\sigma}^2}{\alpha^2} + \frac{\mu^2\,\hat{p}^2}{|\mathcal{S}|^2}\left(\sum_{i\in\mathcal{S}} x_i^*\right)^2 = \frac{\hat{p}\hat{\sigma}^2}{\alpha^2} + \mu^2\,\hat{p}^2\hat{m}^2\,,
\end{aligned}$$

where the approximation in (a) follows from the fact that

$$\frac{1}{|\mathcal{S}|^2}\sum_{i,j\in\mathcal{S}} x_i^*x_j^* = \frac{1}{|\mathcal{S}|^2}\sum_{i\neq j} x_i^*x_j^* + \mathcal{O}\left(\frac{1}{|\mathcal{S}|}\right).$$

We can now compute the variance:

$$\mathrm{var}_{\boldsymbol{x}^*}\left(\check{Z}_k\right) = \mathbb{E}_{\boldsymbol{x}^*}\left(\check{Z}_k^2\right) - \left(\mathbb{E}_{\boldsymbol{x}^*}\check{Z}_k\right)^2 = \frac{\hat{p}\,\hat{\sigma}^2}{\alpha^2}\,.$$

1.C.3 Equation (1.12).

Our starting point is the following generic representation of an abundance at equilibrium (either of a surviving or vanishing species):

$$\begin{aligned}
x_k^* &= \left(1 + \mu\,p^*m^* + \frac{\sigma^*\sqrt{p^*}}{\alpha}Z_k\right)\mathbf{1}_{\{Z_k>-\delta^*\}}\\
&= (1+\mu\,p^*m^*)\,\mathbf{1}_{\{Z_k>-\delta^*\}} + \left(\frac{\sigma^*\sqrt{p^*}}{\alpha}Z_k\right)\mathbf{1}_{\{Z_k>-\delta^*\}}
\end{aligned}$$

Summing over $\mathcal{S}$ and normalizing,

$$\begin{aligned}
\frac{1}{|\mathcal{S}|}\sum_{k\in\mathcal{S}} x_k^* &= (1+\mu\,p^*m^*)\frac{1}{|\mathcal{S}|}\sum_{k\in\mathcal{S}}\mathbf{1}_{\{Z_k>-\delta^*\}} + \frac{\sigma^*\sqrt{p^*}}{\alpha}\frac{1}{|\mathcal{S}|}\sum_{k\in\mathcal{S}} Z_k\mathbf{1}_{\{Z_k>-\delta^*\}},\\
\hat{m} &\overset{(a)}{=} (1+\mu\,p^*m^*) + \frac{\sigma^*\sqrt{p^*}}{\alpha}\frac{n}{|\mathcal{S}|}\frac{1}{n}\sum_{k\in[n]} Z_k\mathbf{1}_{\{Z_k>-\delta^*\}},\\
\hat{m} &\overset{(b)}{\simeq} (1+\mu\,p^*m^*) + \frac{\sigma^*\sqrt{p^*}}{\alpha}\frac{1}{\mathbb{P}(Z>-\delta^*)}\mathbb{E}(Z\mathbf{1}_{\{Z>-\delta^*\}}),\\
\hat{m} &\simeq (1+\mu\,p^*m^*) + \frac{\sigma^*\sqrt{p^*}}{\alpha}\mathbb{E}(Z\mid Z>-\delta^*).
\end{aligned}$$

where (a) follows from the fact that $|\mathcal{S}| = \sum_{k\in\mathcal{S}}\mathbf{1}_{\{Z_k>-\delta^*\}}$ (by definition of $\mathcal{S}$), (b) from the law of large numbers $\frac{1}{n}\sum_{k\in[n]} Z_k\mathbf{1}_{\{Z_k>-\delta\}} \xrightarrow[n\to\infty]{} \mathbb{E}Z\mathbf{1}_{\{Z>-\delta\}}$ and $\frac{|\mathcal{S}|}{n} \xrightarrow[n\to\infty]{} \mathbb{P}(Z>-\delta^*)$ with $Z\sim\mathcal{N}(0,1)$. It remains to replace $\hat{m}$ by its limit m^* to obtain (1.12).

Eq. (1.13) can be obtained similarly. Details are provided in Appendix 1.C, see Section 1.C.4.

1.C.4 Equation (1.13).

As for the proof of (1.12), we start from the generic representation of x_k^*:

$$\begin{aligned} x_k^* &= \left(1+\mu\, p^* m^* + \frac{\sigma^*\sqrt{p^*}}{\alpha} Z_k\right) \mathbf{1}_{\{Z_k > -\delta^*\}} \\ &= (1+\mu\, p^* m^*)\, \mathbf{1}_{\{Z_k > -\delta^*\}} + \frac{\sigma^*\sqrt{p^*}}{\alpha} Z_k \mathbf{1}_{\{Z_k > -\delta^*\}}\,. \end{aligned}$$

Taking the square, we get

$$\begin{aligned} x_k^{*2} = (1+\mu\, p^* m^*)^2\, \mathbf{1}_{\{Z_k > -\delta^*\}} \\ + 2(1+\mu\, p^* m^*)\frac{\sigma^*\sqrt{p^*}}{\alpha} Z_k \mathbf{1}_{\{Z_k > -\delta^*\}} + \frac{(\sigma^*)^2 p^*}{\alpha^2} Z_k^2 \mathbf{1}_{\{Z_k > -\delta^*\}}\,. \end{aligned}$$

Summing over $\mathcal{S}$ and normalizing, we get

$$\begin{aligned} \frac{1}{|\mathcal{S}|}\sum_{k\in\mathcal{S}} (x_k^*)^2 = (1+\mu\, p^* m^*)^2 \frac{1}{|\mathcal{S}|}\sum_{k\in\mathcal{S}} \mathbf{1}_{\{Z_k > -\delta^*\}} \\ + 2(1+\mu\, p^* m^*)\frac{\sigma^*\sqrt{p^*}}{\alpha}\frac{1}{|\mathcal{S}|}\sum_{k\in\mathcal{S}} Z_k \mathbf{1}_{\{Z_k > -\delta^*\}} \\ + \frac{(\sigma^*)^2 p^*}{\alpha^2}\frac{1}{|\mathcal{S}|}\sum_{k\in\mathcal{S}} Z_k^2 \mathbf{1}_{\{Z_k > -\delta^*\}}\,. \end{aligned}$$

Finally, we conclude by replacing the empirical means by their limits

$$\frac{1}{|\mathcal{S}|}\sum_{k\in\mathcal{S}} Z_k^i \mathbf{1}_{\{Z_k > -\delta^*\}} \;=\; \mathbb{E}(Z^i \mid Z > -\delta^*)\;, \quad i = 1, 2\,.$$

and get

$$\begin{aligned} \hat{\sigma}^2 = (1+\mu\, p^* m^*)^2 + 2(1+\mu\, p^* m^*)\frac{\sigma^*\sqrt{p^*}}{\alpha}\mathbb{E}(Z \mid Z > -\delta^*) \\ + \frac{(\sigma^*)^2 p^*}{\alpha^2}\mathbb{E}(Z^2 \mid Z > -\delta^*)\,. \end{aligned}$$

It remains to replace $\hat{\sigma}$ by its limit σ^* to obtain (1.13).

1.D Density of the distribution of the persistent species.

Assume that $x^* > 0$, and let $f = \mathbb{R} \to \mathbb{R}$ be a bounded continuous test function. We have

$$\begin{aligned} \mathbb{E}f(x_k^*) &= \mathbf{E}\left[f\left(1 + \frac{\sigma^*\sqrt{p^*}}{\alpha} Z_k + \mu\, p^* m^*\right) \;\middle|\; Z_k > -\delta^* \right]\,, \\ &= \int_{-\infty}^{\infty} f\left(1+\mu\, p^* m^* + \frac{\sigma^*\sqrt{p^*}}{\alpha} u\right) \frac{\mathbf{1}_{\{u > -\delta^*\}}}{1-\Phi(-\delta^*)} \frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}} du\,, \\ &= \int_0^{\infty} f(y) e^{-\frac{1}{2}\left(\frac{\alpha}{\sigma^*\sqrt{p^*}} y - \delta^*\right)^2} \frac{\alpha}{\sqrt{2\pi}\Phi(\delta^*)\, p^*\, \sigma^*}\, dy\,, \end{aligned}$$

hence the density of x_k^*.

1.D.1 Theoretical estimation of the diversity index

Recall that $|\mathcal{S}| = n\hat{p}$ is the number of surviving species and that

$$p_i = \frac{x_i}{\sum_{j\in\mathcal{S}} x_j}$$

is the frequency of (surviving) species i.

To find a theoretical estimate of Hill number of order 1, we proceed by expansion and set

$$p_i = \frac{1}{|\mathcal{S}|} + \delta_i\,, \quad |\delta_i| \ll \frac{1}{|S|}$$

where δ_i represents the deviation of species i from the standard frequency if all surviving species have the same frequency. Notice that $\sum_{i\in\mathcal{S}} \delta_i = 0$.

$$H' = -\sum_{i\in\mathcal{S}} p_i \log(p_i) = -\sum_{i\in\mathcal{S}} \left(\frac{1}{|\mathcal{S}|} + \delta_i\right) \log\left(\frac{1}{|\mathcal{S}|} + \delta_i\right)\,.$$

We use the Taylor-Young formula of order 2 to decompose the log:

$$\begin{aligned}\log\left(\frac{1}{|\mathcal{S}|} + \delta_i\right) &= \log\left(\frac{1}{|\mathcal{S}|}\right) + |\mathcal{S}|\delta_i - \frac{\delta_i^2|\mathcal{S}|^2}{2} + \delta_i^3\varepsilon(\delta_i)\,,\\ &\approx \log\left(\frac{1}{|\mathcal{S}|}\right) + |\mathcal{S}|\delta_i - \frac{\delta_i^2|\mathcal{S}|^2}{2}\,.\end{aligned}$$

$$\begin{aligned}H' &\approx -\sum_{i\in\mathcal{S}} \left(\frac{1}{|\mathcal{S}|} + \delta_i\right)\left(\log\left(\frac{1}{|\mathcal{S}|}\right) + |\mathcal{S}|\delta_i - \frac{\delta_i^2|\mathcal{S}|^2}{2}\right)\,,\\ &= -\sum_{i\in\mathcal{S}} \left[\frac{1}{|\mathcal{S}|}\log\left(\frac{1}{|\mathcal{S}|}\right) + \delta_i - \frac{\delta_i^2|\mathcal{S}|}{2} + \delta_i\log\left(\frac{1}{|\mathcal{S}|}\right) + |\mathcal{S}|\delta_i^2 - \frac{\delta_i^3|\mathcal{S}|^2}{2}\right]\,,\\ &= \log(|\mathcal{S}|) - \sum_{i\in\mathcal{S}} \frac{\delta_i^2|\mathcal{S}|}{2} + \sum_{i\in\mathcal{S}} \frac{\delta_i^3|\mathcal{S}|^2}{2}\,.\end{aligned}$$

Notice that $\sum_{i=1}^{|\mathcal{S}|} \frac{\delta_i^3|\mathcal{S}|^2}{2}$ is negligible since $|\delta_i| \ll |S|^{-1}$. The term 1 corresponds to the maximum value that the Shannon diversity index can take if $|\mathcal{S}|$ are present in the system.

It remains to develop the second term of the r.h.s.

$$\begin{aligned}
-\frac{1}{2}\sum_{i\in\mathcal{S}}\delta_i^2|\mathcal{S}| &= -\frac{|\mathcal{S}|}{2}\sum_{i\in\mathcal{S}}\left(\frac{x_i}{\sum_{j\in\mathcal{S}}x_j}-\frac{1}{|\mathcal{S}|}\right)^2 , \\
&= -\frac{|\mathcal{S}|}{2}\sum_{i\in\mathcal{S}}\left(\frac{x_i^2}{(\sum_{j\in\mathcal{S}}x_j)^2}-\frac{2}{|\mathcal{S}|}\frac{x_i}{\sum_{j\in\mathcal{S}}x_j}+\frac{1}{|\mathcal{S}|^2}\right) , \\
&= -\frac{|\mathcal{S}|}{2}\sum_{i\in\mathcal{S}}\left(\frac{x_i^2}{(\sum_{j\in\mathcal{S}}x_j)^2}\right)+\frac{1}{2} , \\
&= -\frac{|\mathcal{S}|}{2}\frac{\sum_{i\in\mathcal{S}}x_i^2}{|\mathcal{S}|^2(\frac{1}{|\mathcal{S}|}\sum_{j\in\mathcal{S}}x_j)^2}+\frac{1}{2} , \\
&= -\frac{1}{2}\frac{\frac{1}{|\mathcal{S}|}\sum_{i\in\mathcal{S}}x_i^2}{(\frac{1}{|\mathcal{S}|}\sum_{j\in\mathcal{S}}x_j)^2}+\frac{1}{2} , \\
&= -\frac{1}{2}\frac{\hat{\sigma}^2}{(\hat{m})^2}+\frac{1}{2} , \\
&= -\frac{1}{2}\left(\frac{\hat{\sigma}^2}{\hat{m}^2}-1\right) .
\end{aligned}$$

Finally the Hill number of order 1 can be computed as:

$$\begin{aligned}
e^{H'} &\approx e^{\log(|\mathcal{S}|)-\frac{|\mathcal{S}|}{2}\sum_{i=1}^{|\mathcal{S}|}\delta_i^2} , \\
&\approx |\mathcal{S}|\left(1-\frac{|\mathcal{S}|}{2}\sum_{i=1}^{|\mathcal{S}|}\delta_i^2\right) = |\mathcal{S}|\left(1-\frac{1}{2}\frac{\hat{\sigma}^2}{(\hat{m})^2}+\frac{1}{2}\right) = \frac{|\mathcal{S}|}{2}\left(3-\frac{\hat{\sigma}^2}{(\hat{m})^2}\right) .
\end{aligned}$$

Replacing $|\mathcal{S}|$ by np^* and $\hat{\sigma}$ and $\hat{m}$ by their limits, we get the desired result:

$$e^{H'} \approx \frac{np^*}{2}\left(3-\frac{(\sigma^*)^2}{(m^*)^2}\right) .$$

Chapter 2

Equilibrium in a large Lotka-Volterra system with pairwise correlated interactions

Abstract

Consider a Lotka-Volterra (LV) system of coupled differential equations:

$$\dot{x}_k = x_k(r_k - x_k + (B\boldsymbol{x})_k)\,, \quad \boldsymbol{x} = (x_k)\,, \quad 1 \leqslant k \leqslant n\,,$$

where $\boldsymbol{r} = (r_k)$ is a $n \times 1$ vector and B a $n \times n$ matrix. Assume that the interaction matrix B is random and follows the elliptic model:

$$B = \frac{1}{\alpha\sqrt{n}}A + \frac{\mu}{n}\mathbf{1}_n\mathbf{1}_n^\top\,,$$

where $A = (A_{ij})$ is a $n \times n$ matrix with $\mathcal{N}(0,1)$ entries satisfying the following dependence structure (i) the entries A_{ij} on and above the diagonal are i.i.d., (ii) for $i < j$ each vector (A_{ij}, A_{ji}) is standard gaussian with covariance ρ, and independent from the other entries; vector $\mathbf{1}_n$ stands for the $n \times 1$ vector of ones. Parameters α, μ are deterministic and may depend on n.

Leveraging on Random Matrix Theory, we analyse this LV system as $n \to \infty$ and study the existence of a positive equilibrium. This question boils down to study the existence of a (componentwise) positive solution to the linear equation:

$$\mathbf{x}_n = \boldsymbol{r}_n + B_n\mathbf{x}_n\,,$$

depending on B's parameters (α, μ, ρ), a problem of independent interest in linear algebra.

In the case where no positive equilibrium exists, we provide sufficient conditions for the existence of a unique stable equilibrium (with vanishing components), and following Bunin [Bun17], present a heuristics estimating the number of positive components of the equilibrium and their distribution.

The existence of positive equilibria for large Lotka-Volterra systems has been raised in Dougoud *et al.* [DVR+18], and addressed in various contexts by Najim *et al.* [AN21, BN21].

Such LV systems are widely used in mathematical biology to model populations with interactions, and the existence of a positive equilibrium known as a *feasible equilibrium* corresponds to the survival of all the species x_k within the system.

2.1 Introduction

2.1.1 Lotka-Volterra system of coupled differential equations.

Lotka-Volterra (LV) systems are widely used in mathematical biology, ecology, chemistry to model populations with interactions or chemical reactions [Gop84, HS98, KK08, Her90]. In the context of theoretical ecology (that we shall adopt hereafter without loss of generality), consider a given foodweb and denote by $\mathbf{x}_n^t = (x_k(t))_{1\leqslant k\leqslant n}$ the vector of abundances[1] of the various species at time $t \geqslant 0$. In a LV system, the abundances are connected via the following coupled equations:

$$\frac{dx_k(t)}{dt} = x_k(t)\left(r_k - x_k(t) + \sum_{\ell=1}^{n} B_{k\ell}x_\ell(t)\right) \qquad \text{for} \quad k \in [n] := \{1, \cdots, n\}\,, \tag{2.1}$$

where $B_n = (B_{k\ell})$ stands for the interaction matrix, and r_k stands for the intrinsic growth of species k. Notice that standard results yield that if the initial condition $\boldsymbol{x}_n^0 = \boldsymbol{x}_n|_{t=0}$ is componentwise positive, then $\boldsymbol{x}_n^t$ remains positive for every $t > 0$.

At the equilibrium $\frac{d\mathbf{x}_n}{dt} = 0$, the abundance vector $\mathbf{x}_n = (x_k)_{k\in[n]}$ is solution of the system:

$$x_k\left(r_k - x_k + \sum_{\ell\in[n]} B_{k\ell}x_\ell\right) = 0 \qquad \text{for} \quad x_k \geqslant 0 \quad \text{and} \quad k \in [n]\,. \tag{2.2}$$

An important question, which motivated recent developments [DVR+18, BN21], is the existence of a *feasible* solution $\mathbf{x}_n$ to (2.2), that is a solution where all the x_k's are positive, corresponding to a scenario where no species disappears. Notice that in this latter case, the system (2.2) takes the much simpler form:

$$\mathbf{x}_n = \boldsymbol{r}_n + B_n\mathbf{x}_n\,, \qquad \boldsymbol{r}_n = (r_k)_{k\in[n]}\,.$$

In this article, we will investigate the existence of an equilibrium, potentially feasible, for a large foodweb ($n \to \infty$) whenever the interaction matrix B_n is random. In various models of interest for B_n, Random Matrix Theory (RMT) provides an accurate description of the asymptotic properties of a large random matrix (its spectrum, spectral norm, etc.). We will leverage on RMT to infer the existence of an equilibrium in the case where B_n follows a random elliptic model, to be described hereafter.

To simplify the analysis, we will mostly consider the case where $\boldsymbol{r}_n = \mathbf{1}_n$, except for the stability where this extra assumption is not needed.

2.1.2 Random elliptic model for the interaction matrix

In the spirit of May[2], we model the interaction matrix B_n as a non-centered random matrix with pairwise correlated entries:

$$B_n = \frac{A_n}{\alpha_n\sqrt{n}} + \frac{\mu}{n}\mathbf{1}_n\mathbf{1}_n^\top\,, \tag{2.3}$$

[1] A species abundance is a quantity proportional to the number of individuals for this species.

[2] Beware that May did not consider LV systems but rather used a random matrix model for the Jacobian at equilibrium of a generic system of coupled differential equations.

where $A_n = (A_{ij})_{i,j\in[n]}$ is a random matrix satisfying the two conditions (i) $(A_{ij}, i \leqslant j)$ are standard Gaussian $\mathcal{N}(0,1)$ independent and identically distributed (i.i.d.) random variables (ii) for $i < j$ the vector (A_{ij}, A_{ji}) is a standard bivariate Gaussian vector, independent from the remaining random variables, with covariance $\mathrm{cov}(A_{ij}, A_{ji}) = \rho$ with $|\rho| \leqslant 1$. The sequence of positive numbers (α_n) is either fixed or goes to infinity. Parameter μ is a fixed real number. As a consequence, the Gaussian entries of the interaction matrix B_n admit the following moments:

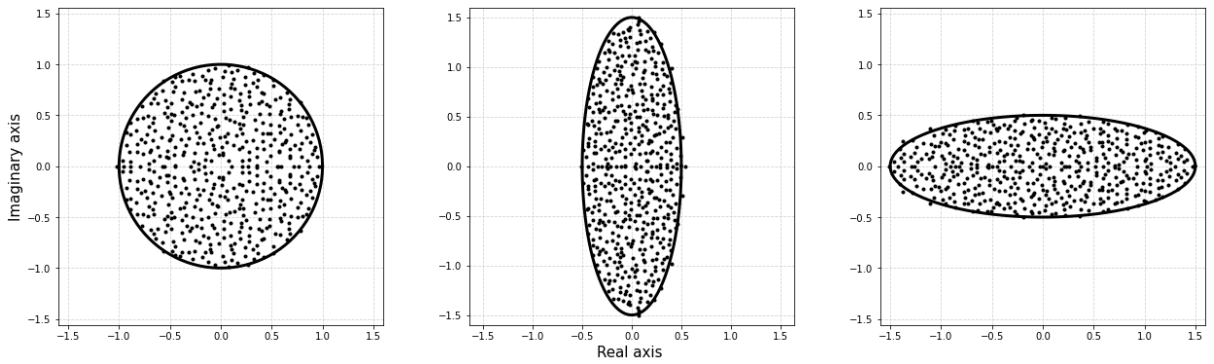
$$\mathbb{E}(B_{ij}) = \frac{\mu}{n}, \quad \mathrm{var}(B_{ij}) = \frac{1}{\alpha^2 n}, \quad \mathrm{cov}(B_{ij}, B_{ji}) = \frac{\rho}{\alpha^2 n} \ (i \neq j).$$

Such a matrix model is often called a random elliptic model for $|\rho| < 1$ since the spectrum of matrix $A_n/\sqrt{n}$ is asymptotically an ellipse, see Fig.2.1.1, in the sense that the empirical distribution of the eigenvalues of $A_n/\sqrt{n}$ converges towards the uniform distribution on the ellipsoid

$$\mathcal{E}_\rho = \left\{ z \in \mathbb{C}, \ \frac{\mathrm{Re}^2(z)}{(1+\rho)^2} + \frac{\mathrm{Im}^2(z)}{(1-\rho)^2} \leqslant 1 \right\}.$$

Originally introduced by Girko [Gir86], this model has since been widely studied [Gir95, Nau12, NO15, OR14, AK22] from a mathematical perspective.

From a theoretical ecology viewpoint, the random elliptic model is interesting [AT12, Bun17], [Ros13, Section 18.3] for its flexibility as it introduces a correlation parameter between the pairwise entries (A_{ij}, A_{ji}). Positive correlations will be used to model mutualistic interactions while negative ones will model predator/prey interactions. This model interpolates between the Wigner model ($\rho = 1$), the full i.i.d. model ($\rho = 0$ with Gaussian entries) and the antisymetric model ($\rho = -1$).



(a) $\rho = 0$ (b) $\rho = -0.5$ (c) $\rho = 0.5$

Figure 2.1.1: Spectrum of non-Hermitian matrix B_n ($n = 500$) in the centered case ($\mu = 0$) with distinct parameter $\rho \in \{-0.5,\ 0,\ 0.5\}$. The solid line represents the ellipse $\{z = x + \boldsymbol{i}y \in \mathbb{C},\ \frac{x^2}{(1+\rho)^2} + \frac{y^2}{(1-\rho)^2} = 1\}$ which is the boundary of the support of the limiting spectral distribution for an elliptic model.

The spectral norms of A_n and $\mathbf{1}_n\mathbf{1}_n^\top$ satisfy

$$\left\| \frac{A_n}{\sqrt{n}} \right\| = \mathcal{O}(1) \quad \text{and} \quad \left\| \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top \right\| = 1$$

hence both the random and deterministic parts of the interaction matrix B_n may have an impact as $n \to \infty$.

2.1.3 Presentation of the main results

In this article, we address the following issues.

Feasibility.

We first describe the conditions over parameters (ρ, α_n, μ) for which system (2.2) admits a unique feasible equilibrium. We prove that feasibility is reached whenever $\alpha_n \gg \sqrt{2\log(n)}$ and $\mu < 1$, and that there is no feasibility otherwise, see Theorem 2.1. Notice that the correlation parameter ρ has no influence since the phase transition threshold is the same as in the i.i.d. case [BN21]: the induced correlations between components x_k's of solution $\boldsymbol{x}_n$ are too weak. Pushing this remark further, we prove that the same phase transition holds if we consider a covariance profile $(\rho_{ij},\, i < j)$ where $\rho_{ij} = \mathrm{cov}(A_{ij}, A_{ji})$ instead of a fixed covariance parameter ρ.

In [BN21], Bizeul and Najim established the conditions for feasibility in the centered $(\mu = 0)$ model with i.i.d interactions (A_{ij}). In [AN21], Akjouj and Najim studied a sparse model of interactions where only $d_n \geqslant \log(n)$ interactions are non-null in each row and column of A_n. The study of the feasibility for an elliptic model completes this picture.

Stability without feasibility.

If α is fixed, Dougoud *et al.* [DVR+18] showed that no feasible solution can arise. Under this assumption, we establish in Proposition 2.3 sufficient conditions for the existence of a unique stable equilibrium to system (2.2). In this case, some species will vanish (some of the components x_k's of solution $\boldsymbol{x}_n$ are equal to zero). In order to proceed we combine results by Takeuchi [Tak96] on stability of LV systems with Random Matrix Theory (RMT) results.

Estimating the number of surviving species.

We finally conclude with an important question: given a set of parameters (ρ, α, μ) which yields to a unique stable equilibrium, is it possible to estimate the proportion of surviving species? From a mathematical point of view, this is an open question. At a physical level of rigor, Bunin [Bun17] (relying on the cavity method) and Galla [Gal18] (relying on generating functionals techniques) provide a closed-form system of equations to compute the proportion of surviving species. We state the open problem, recall Bunin's and Galla's equations and provide simulations.

In [CMN22], equations and simulations are provided in the simpler case where $\rho = 0$, together with heuristics supporting these equations.

Organisation of the article

Feasibility and stability results together with the open question on the estimation of the number of surviving species are presented in Section 2.2. Section 2.3 is devoted to the proof of the feasibility result, Theorem 2.1. Proof of the stability result, Proposition 2.3, is provided in Section 2.4. Simulations were performed in Python. All the figures and the code are available on Github [Cle22b].

Notations

If A is a matrix $A^\top$ stands for its transpose. We denote by $\log(x)$ the natural logarithm. If $\boldsymbol{x} = (x_i)_{i\in[n]}$ is a vector, we denote by $\boldsymbol{x} > 0$ (resp. $\boldsymbol{x} \geqslant 0$) the componentwise positivity (resp. non-negativity), that is the fact that $x_i > 0$ (resp. $x_i \geqslant 0$) for every $i \in [n]$. We denote by $\xrightarrow{a.s.}$ (resp. $\xrightarrow{\mathcal{P}}$) almost sure convergence (resp. convergence in probability).

2.2 Main results: Feasibility, stability and surviving species

2.2.1 Feasibility

To simplify the analysis, we consider the case where $r_k = 1$ ($k \in [n]$). Hence, the LV system takes the following form in the sequel:

$$\frac{dx_k(t)}{dt} = x_k(t)\left(1 - x_k(t) + \sum_{\ell\in[n]} B_{k\ell}x_\ell(t)\right) \qquad \text{for} \quad k \in [n]\,. \tag{2.4}$$

In the next theorem, we describe the conditions to reach a feasible equilibrium. We either assume that matrix B is given by the elliptic model or has a more general covariance profile.

Theorem 2.1 (Feasibility for the elliptic model)**.** *Assume that matrix B_n is given by the elliptic model* (2.3)*, or that B_n has a covariance profile, i.e.*

$$B_n = \frac{\tilde{A}_n}{\alpha_n\sqrt{n}} + \frac{\mu}{n}\mathbf{1}_n\mathbf{1}_n^\top\,, \tag{2.5}$$

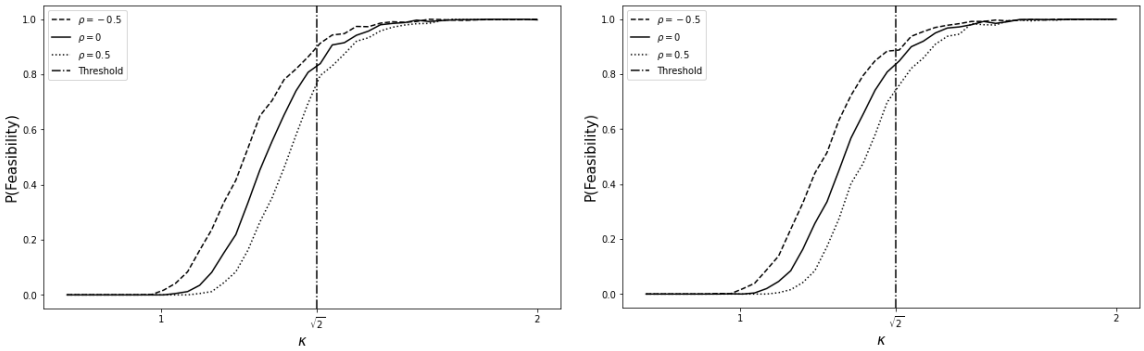
where $\tilde{A}_n$ is a $n \times n$ matrix with entries $(\tilde{A}_{ij}, i \leqslant j)$ i.i.d. $\mathcal{N}(0,1)$ and where $(\tilde{A}_{ij}, \tilde{A}_{ji})$ is a standard bivariate gaussian vector for $i < j$, independent from the remaining random variables, with covariance $\mathrm{cov}(\tilde{A}_{ij}, \tilde{A}_{ji}) = \rho_{ij}^{(n)}$, where $(\rho_{ij}^{(n)}; i < j; n \geqslant 1)$ is a collection of deterministic real numbers in $[-1, 1]$.

Let $\alpha_n \xrightarrow[n\to\infty]{} \infty$ and denote by $\alpha_n^* = \sqrt{2\log n}$. If $\mu \neq 1$ then the following equation

$$\boldsymbol{x}_n = \mathbf{1}_n + B_n\boldsymbol{x}_n$$

almost surely admits a unique solution $\boldsymbol{x}_n = (x_k)_{k\in[n]}$.

1. *(feasibility) If $\mu < 1$ and there exists $\varepsilon > 0$ such that, for n large enough, $\alpha_n \geqslant (1+\varepsilon)\alpha_n^*$ then*
$$\mathbb{P}\left\{\min_{k\in[n]} x_k > 0\right\} \xrightarrow[n\to\infty]{} 1\,.$$
2. *If $\mu > 1$ or there exists $\varepsilon > 0$ such that, for n large enough, $\alpha_n \leqslant (1-\varepsilon)\alpha_n^*$ then*
$$\mathbb{P}\left\{\min_{k\in[n]} x_k > 0\right\} \xrightarrow[n\to\infty]{} 0\,.$$



(a) Gaussian entries, (b) Uniform $[-\sqrt{3}, \sqrt{3}]$ entries.

Figure 2.2.1: Transition towards feasibility for the elliptic model (2.3) on the left subfigure (Gaussian entries), and for a model with uniform entries and same pairwise dependence structure on the right one. For each κ on the x-axis, we simulate 1000 matrices B_n of size $n = 1000$, compute the solution $\boldsymbol{x}_n$ of Theorem 2.1 at the scaling $\alpha_n(\kappa) = \kappa\sqrt{\log(n)}$ and then plot the proportion of feasible solutions obtained for the 1000 simulations. Each curve represents the proportion of feasible solutions $\boldsymbol{x}_n$ for three distinct values $\rho \in \{-0.5,\ 0,\ 0.5\}$. The dot-dashed vertical line corresponds to $\kappa = \sqrt{2}$ i.e. the critical scaling $\alpha_n^* = \sqrt{2\log(n)}$.

Proof of Theorem 2.1 is established in Section 2.3 under the assumption that B_n follows the elliptic model. The adaptations needed to cover the covariance profile case are provided in Appendix 2.A.

Remark 2.1. If one considers the system (2.1) instead of (2.4), that is $\boldsymbol{r}_n \neq \mathbf{1}_n$, the sharp phase transition at $\alpha_n^* = \sqrt{2\log(n)}$ does not hold any more. The transition takes place on a wider region for the α_n's (transition buffer) which upper and lower bounds depend on $\boldsymbol{r}_n$'s characteristics: Absence of feasibility[3] if $\alpha_n \ll \frac{\sigma_{\boldsymbol{r}}(n)}{\boldsymbol{r}_{\max}(n)}\sqrt{2\log(n)}$ and feasibility if $\alpha_n \gg \frac{\sigma_{\boldsymbol{r}}(n)}{\boldsymbol{r}_{\min}(n)}\sqrt{2\log(n)}$ where

$$\sigma_{\boldsymbol{r}}(n) = \left\|\boldsymbol{r}/\sqrt{n}\right\|, \quad \boldsymbol{r}_{\max}(n) = \max_{i\in[n]} r_i(n), \quad \boldsymbol{r}_{\min}(n) = \min_{i\in[n]} r_i(n),$$

with the assumption that $0 < \kappa \leqslant \boldsymbol{r}_{\min}(n) \leqslant \boldsymbol{r}_{\max}(n) \leqslant K < \infty$. This can be established following the general lines of the proof (Normal Comparison Lemma - Th. 2.4, representation lemma - Lemma 2.9, etc.) and adapting the arguments from [BN21, Section 4.2].

Remark 2.2 (Structural stability). In the context of the previous remark, notice that if $\alpha_n \gg \frac{\sigma_{\boldsymbol{r}}(n)}{\boldsymbol{r}_{\min}(n)}\sqrt{2\log(n)}$, a small change in $\boldsymbol{r}_n$, say $\boldsymbol{r}_n \to \boldsymbol{r}_n + \boldsymbol{\delta}_n$ such that $\sigma_{\min}(\boldsymbol{r}_n + \boldsymbol{\delta}_n)$ remains to close to $\sigma_{\min}(\boldsymbol{r}_n)$ will not affect the feasibility (and stability, see for instance Remark 2.4). The fixed point is therefore *structurally stable* in the sense of Grilli *et al.* [GAS+17].

Remark 2.3 (Non-Gaussian random variables). A natural question is whether Theorem 2.1 remains true if B_n's entries are no longer Gaussian. Simulations with non-Gaussian pairwise correlated random variables support this idea, see Fig. 2.2.1b where uniform (centered with variance one) entries are used. However an effective proof will require to

[3]Here $a_n \ll b_n$ means that there exists $\varepsilon > 0$ such that $a_n \leqslant (1-\varepsilon)b_n$ eventually.

overcome some of the arguments based on the Gaussiannity of the entries, such as the Normal Comparison Lemma (Theorem 2.4) and a representation lemma (Lemma 2.9). We do not pursue in this direction here.

2.2.2 No feasibility but a unique stable equilibrium.

Aside from the question of feasibility arises the question of *stability*: for a complex system, how likely a perturbation of the solution $\mathbf{x}_n$ at equilibrium will return to the equilibrium? Gardner and Ashby [GA70] considered stability issues of complex systems connected at random. Based on the circular law for large random matrices with i.i.d. entries, May [May72] provided a complexity/stability criterion and motivated the systematic use of large random matrix theory in the study of foodwebs, see for instance Allesina *et al.* [AT15]. Recently, Stone [Sto18] and Gibbs *et al.* [GGRA18] revisited the relation between feasibility and stability.

For a generic LV system

$$\frac{d\, y_k(t)}{dt} = y_k(r_k + (C\boldsymbol{y})_k)\,, \quad k \in [n]\,, \tag{2.6}$$

Takeuchi and Adachi provide a criterion for the existence of a unique equilibrium $\boldsymbol{y}^*$ and the global stability of LV systems, see Theorem 3.2.1 in [Tak96].

Theorem 2.2 (Takeuchi and Adachi 1980)**.** *If there exists a positive diagonal matrix Δ such that $\Delta C + C^\top \Delta$ is negative definite, there is a unique non-negative equilibrium $\boldsymbol{y}^*$ to (2.6), which is globally stable:*

$$\forall \boldsymbol{y}_0 > 0\,, \quad \begin{cases} \boldsymbol{y}(0) = \boldsymbol{y}_0 \\ \boldsymbol{y}(t) \text{ satisfies (2.6)} \end{cases}, \quad \boldsymbol{y}(t) \xrightarrow[t\to\infty]{} \boldsymbol{y}^*\,.$$

Combining this result (setting $I - B = -C$) with results from Random Matrix Theory, we can guarantee the existence of a globally stable equilibrium $\boldsymbol{x}^*$ of (2.1) for a wide range of parameters (ρ, α, μ). Denote by

$$\mathcal{A} = \Big\{ (\rho,\alpha,\mu) \in (-1,1)\times(0,\infty)\times\mathbb{R}\,, \\ \alpha > \sqrt{2(1+\rho)}, \quad \mu < \frac{1}{2} + \frac{1}{2}\sqrt{1 - \frac{2(1+\rho)}{\alpha^2}} \Big\} \tag{2.7}$$

the set of admissible parameters.

Proposition 2.3. *Let $\check{A}_n$ be a $n \times n$ matrix with entries $(\check{A}_{ij}, i \leqslant j)$ i.i.d., centered with unit variance, with a fourth finite moment and where $(\check{A}_{ij}, \check{A}_{ji})$ is a standard bivariate vector for $i < j$, independent from the remaining random variables, with covariance $\mathrm{cov}(\check{A}_{ij}, \check{A}_{ji}) = \rho$. Denote by*

$$B_n = \frac{\check{A}_n}{\alpha\sqrt{n}} + \frac{\mu}{n}\mathbf{1}_n\mathbf{1}_n^\top\,. \tag{2.8}$$

Consider the system (2.1) and let $(\rho, \alpha, \mu) \in \mathcal{A}$, then almost surely, matrix

$$(I - B_n) + (I - B_n)^\top$$

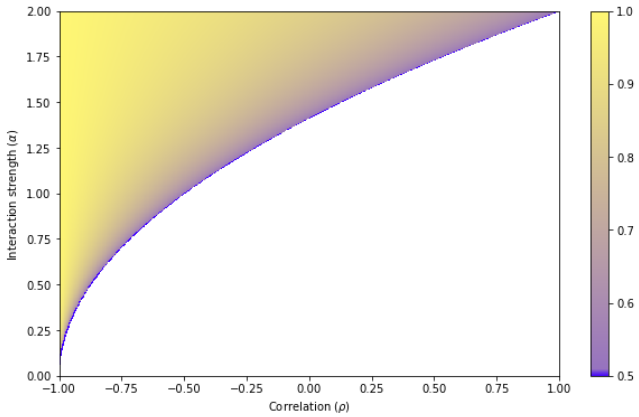
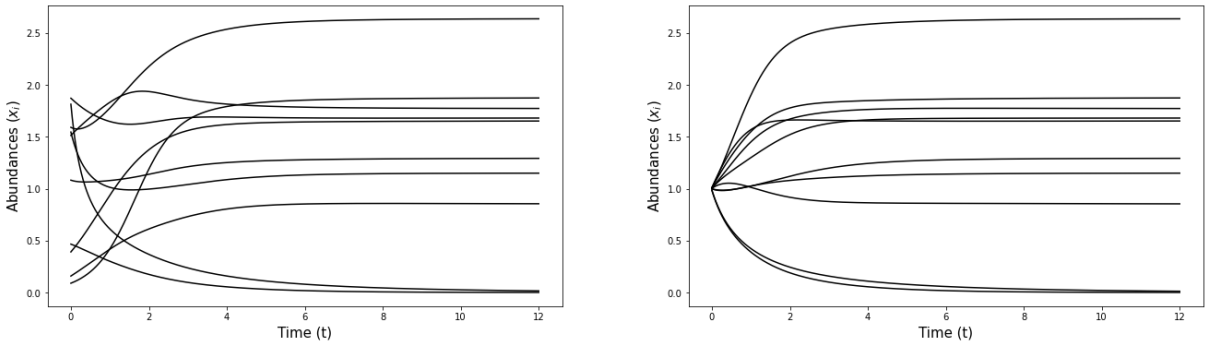


Figure 2.2.2: Representation of the set of admissible parameters $\mathcal{A}$ by a heat map. The set $\mathcal{A}$ given by (2.7) yields the existence of a unique (random) globally stable equilibrium $\boldsymbol{x}^*$. The x-axis corresponds to ρ, the y-axis to α and the intensity of the color μ.

is eventually positive definite: with probability one, for a given realization ω, there exists $N(\omega)$ such that for $n \geqslant N(\omega)$, $(I - B_n^\omega) + (I - B_n^\omega)^\top$ is positive definite. In particular, there exists a unique globally stable non-negative equilibrium $\boldsymbol{x}^*$.

Proof of Proposition 2.3 is provided in Section 2.4.

Remark 2.4. Notice that this result applies to a general vector $\boldsymbol{r}_n$ and to non-Gaussian random variables with a finite fourth moment. In particular, the assumption does not involve vector $\boldsymbol{r}_n$ and a small change $\boldsymbol{r}_n \to \boldsymbol{r}_n + \delta \boldsymbol{r}_n$ will not affect the stability of the system. The finite fourth moment assumption is necessary to control $\lambda_{\max}(B_n + B_n^\top)$.



(a) Initial conditions drawn in $(0, 2)$, (b) Initial conditions equal to 1.

Figure 2.2.3: Representation of the dynamics of a ten-species system. For a fixed matrix of interactions B_{10} with parameters $(\rho = 0, \alpha = 2, \mu = 0) \in \mathcal{A}$, we consider two distinct initial conditions. Simulations show that the abundances converge in both cases toward the unique globally stable equilibrium x^* predicted by Proposition 2.3. Notice that since $\alpha < \sqrt{2\log(10)} \simeq 2.14$, we witness vanishing species.

2.2.3 Estimating the number of surviving species: Towards Bunin and Galla's equations.

After giving conditions for the realization of a feasible equilibrium and investigating the existence and uniqueness of a stable sub-population (i.e some species vanish), we address the question of estimating the proportion of surviving species as a function of the model parameters (ρ, α, μ).

To our knowledge, this question has not received yet an answer at a mathematical level of rigor and remains open. However theoretical physicists/ecologists provided a solution to this problem supported by simulations. Tools from physics to study population dynamics in the context of Lotka-Volterra equations were first introduced by Opper *et al.* [DO89, OD92]. In 2017, Bunin [Bun17] precisely answers the question of estimating the proportion of surviving species for the model under investigation (non-centered elliptic model B). He uses the dynamical cavity method (a review of which can be found in [BA17]). The key concept consists of assuming that a unique fixed point exists and introducing a new species with new interactions in the existing system. Provided the coherence of the assumption, an analogy between the properties of the solutions with n and $n+1$ species yields closed-form equations that we present hereafter.

Notice that recently, similar equations were obtained by Galla [Gal18] using generating functional techniques.

The system of equations presented hereafter is a version of Bunin's equations without the carrying capacity. It is similar to the equations obtained by the replicator equations [DO89, OD92]. Notice that we mention but do not discuss the many implicit assumptions yielding the system of equations (see Appendix 2.B for more details on the system of equations (2.10)-(2.13)).

Let $(\rho, \alpha, \mu) \in \mathcal{A}$ and $\boldsymbol{x}^*$ given by Proposition 2.3. We first introduce the following quantities:

$$\phi = \frac{\mathrm{Card}\{x_i^* > 0, i \in [n]\}}{n} , \quad \langle \boldsymbol{x} \rangle = \frac{1}{n}\sum_{j=1}^{n} x_j^* , \quad \langle \boldsymbol{x}^2 \rangle = \frac{1}{n}\sum_{j=1}^{n} (x_j^*)^2 . \tag{2.9}$$

Denote by $Z \sim \mathcal{N}(0,1)$ and set

$$\Delta = (1 + \langle \boldsymbol{x} \rangle \mu) \frac{\alpha}{\sqrt{\langle \boldsymbol{x}^2 \rangle}} .$$

The following system of 4 equations has 4 unknowns, among which the (supposedly existing) asymptotic limits of $\phi, \langle \boldsymbol{x} \rangle, \langle \boldsymbol{x}^2 \rangle$, denoted (by abuse of notations) by the same notations. The fourth unknown v is a parameter essentially related to the dynamical

cavity method. This system is supposed to admit a unique solution:

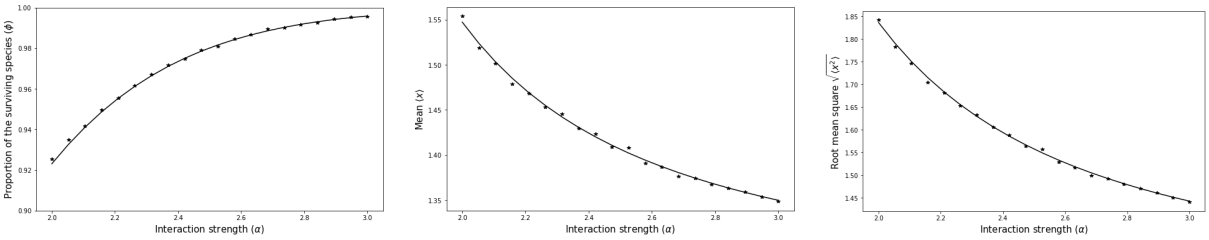
$$\phi = \frac{1}{\sqrt{2\pi}} \int_{-\Delta}^{+\infty} e^{\frac{-z^2}{2}} dz \tag{2.10}$$

$$\langle \boldsymbol{x} \rangle = \frac{\phi}{1 - \frac{\rho v}{\alpha}} \left((1 + \langle \boldsymbol{x} \rangle \mu) + \frac{\sqrt{\langle \boldsymbol{x}^2 \rangle}}{\alpha} \mathbb{E}(Z | Z > -\Delta) \right) \tag{2.11}$$

$$\begin{aligned} \langle \boldsymbol{x}^2 \rangle = & \left(\frac{\sqrt{\phi}}{1 - \frac{\rho v}{\alpha}} \right)^2 \left((1 + \langle \boldsymbol{x} \rangle \mu)^2 + 2(1 + \langle \boldsymbol{x} \rangle \mu) \frac{\sqrt{\langle \boldsymbol{x}^2 \rangle}}{\alpha} \mathbb{E}(Z | Z > -\Delta) \right. \\ & \left. + \frac{\langle \boldsymbol{x}^2 \rangle}{\alpha^2} \mathbb{E}(Z^2 | Z > -\Delta) \right) \end{aligned} \tag{2.12}$$

$$v = \phi \left(\frac{1}{\alpha - \rho v} \right) \tag{2.13}$$

The theoretical solutions of system (2.10)-(2.13) are compared with the empirical values obtained by Monte-Carlo experiments. As shown in Fig. 2.2.4, the matching is remarkable.



(a) ϕ versus α, (b) $\langle x \rangle$ versus α, (c) $\langle x^2 \rangle$ versus α.

Figure 2.2.4: Theoretical values of ϕ, $\langle \boldsymbol{x} \rangle$ and $\langle \boldsymbol{x}^2 \rangle$ (solid line) obtained by solving the system (2.10)-(2.13) given the parameters $(\mu = 0.2, \rho = 0.5)$, compared to the empirical values (dots) obtained by Monte-Carlo simulations (size of matrix $n = 500$, number of random samples $P = 200$). The x-axis corresponds to the interaction strength α.

The impact of the correlation ρ on the proportion of the surviving species is shown in Figure 2.2.5.

Remark 2.5. From a theoretical ecology point of view, notice that a negative correlation (prey-predator) seems to slow down the decline of the surviving species, whereas a positive correlation (mutualism and competition) reverses the trend. These types of results are similar to Allesina and Tang [AT12] where they notice that prey-predator interactions seem to stabilize the system.

2.3 Feasibility: Proof of Theorem 2.1

We assume that matrix B_n is given by (2.3) (elliptic model). The case where matrix B_n is given by (2.5) (covariance profile model) needs extra arguments which are provided in Appendix 2.A.

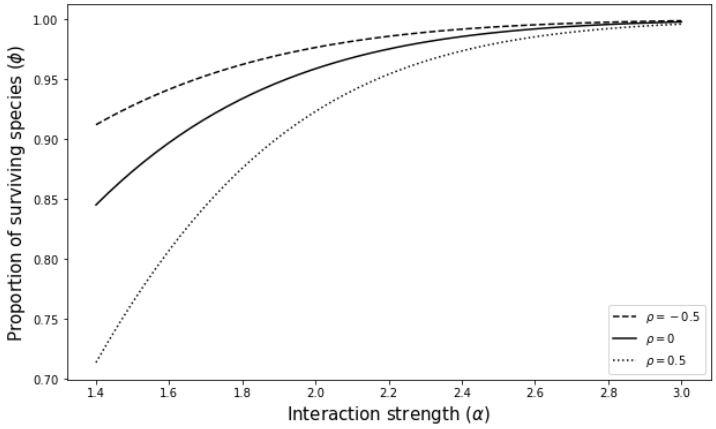


Figure 2.2.5: Effect of the correlation ρ and the interaction strength α on the proportion of surviving species ϕ. Each curve is plotted by resolving the system (2.10)-(2.13) in the centered case $\mu = 0$.

2.3.1 Preliminary results

Extreme Value Theory (EVT) and the Normal Comparison Lemma

Let $(Z_k)_{k\in\mathbb{N}}$ be a sequence of i.i.d. $\mathcal{N}(0,1)$ random variables and denote:

$$\begin{cases} M_n = \max_{k\in[n]} Z_k \\ \widetilde{M}_n = \min_{k\in[n]} Z_k \end{cases} , \quad \alpha_n^* = \sqrt{2\log n}\,, \quad \beta_n^* = \alpha_n^* - \frac{1}{2\alpha_n^*}\log(4\pi\log n)\,. \tag{2.14}$$

Let $G(x) = e^{-e^{-x}}$ be the Gumbel cumulative distribution function, then classical EVT results (see for instance [LLR83, Theorem 1.5.3]) yield that for every $x \in \mathbb{R}$,

$$\mathbb{P}\left\{\alpha_n^*(M_n - \beta_n^*) \leqslant x\right\} \xrightarrow[n\to\infty]{} G(x)\,, \quad \mathbb{P}\left\{\alpha_n^*(\widetilde{M}_n + \beta_n^*) \geqslant -x\right\} \xrightarrow[n\to\infty]{} G(x)\,. \tag{2.15}$$

We consider the following dependent framework: Let $(\mathcal{Z}_{k,n})_{k\in[n]}$ be a Gaussian vector whose components are $\mathcal{N}(0,1)$ with covariance

$$\operatorname{cov}\left(\mathcal{Z}_{k,n}; \mathcal{Z}_{\ell,n}\right) = \frac{\rho}{n}\,, \quad |\rho| \leqslant 1\,, \quad k \neq \ell\,.$$

We are interested in the behaviour of $\mathcal{M}_n = \max_{k\in[n]} \mathcal{Z}_{k,n}$ and $\widetilde{\mathcal{M}}_n = \min_{k\in[n]} \mathcal{Z}_{k,n}$, and shall prove the counterpart of (2.15) with the help of the Normal Comparison Lemma (NCL):

Theorem 2.4 (Theorem 4.2.1, [LLR83]). *Suppose that $(\xi_i,\, i \in [n])$ is a gaussian vector where the ξ_i's are standard normal variables, with covariance matrix $\Lambda^1 = \left(\Lambda^1_{ij}\right)$. Similarly, let $(\eta_i,\, i \in [n])$ be a gaussian vector where the η_i's are standard normal, with covariance matrix $\Lambda^0 = \left(\Lambda^0_{ij}\right)$. Denote by $\rho_{ij} = \max\left\{|\Lambda^0_{ij}|, |\Lambda^1_{ij}|\right\}$ and let $(u_i,\, i \in [n])$ be real numbers. Then:*

$$\begin{aligned} &\left|P\left\{\xi_j \leqslant u_j\,,\ j \in [n]\right\} - P\left\{\eta_j \leqslant u_j\,,\ j \in [n]\right\}\right| \\ &\qquad \leqslant \frac{1}{2\pi} \sum_{1\leqslant i<j\leqslant n} \left|\Lambda^1_{ij} - \Lambda^0_{ij}\right| \left(1-\rho_{ij}^2\right)^{-1/2} \exp\left(-\frac{\frac{1}{2}(u_i^2+u_j^2)}{1+\rho_{ij}}\right)\,. \end{aligned} \tag{2.16}$$

Corollary 2.5. *Recall the definition of* $(\mathcal{Z}_{k,\ell})_{k\in[n]}$, $\mathcal{M}_n$ *and* $\widetilde{\mathcal{M}}_n$ *above, then*

$$\mathbb{P}\left\{\alpha_n^*(\mathcal{M}_n-\beta_n^*)\leqslant x\right\}\xrightarrow[n\to\infty]{} G(x)\,,\quad \mathbb{P}\left\{\alpha_n^*(\widetilde{\mathcal{M}}_n+\beta_n^*)\geqslant -x\right\}\xrightarrow[n\to\infty]{} G(x)\,. \tag{2.17}$$

Proof. We apply the NCL to $(Z_k)_{k\in[n]}$ and $(\mathcal{Z}_{k,n})_{k\in[n]}$. Let $\rho_{ij}=\frac{|\rho|}{n}$ and $u_n(x)=\frac{x}{\alpha_n^*}+\beta_n^*$, then

$$\begin{aligned}
&|\mathbb{P}\{\alpha_n^*(M_n-\beta_n^*)\leqslant x\}-\mathbb{P}\{\alpha_n^*(\mathcal{M}_n-\beta_n^*)\leqslant x\}|\\
&\quad=\quad |\mathbb{P}\{Z_j\leqslant u_n(x)\,,\ j\in[n]\}-\mathbb{P}\{\mathcal{Z}_{j,n}\leqslant u_n(x)\,,\ j\in[n]\}|\ ,\\
&\quad\leqslant\quad \frac{1}{2\pi}\frac{n(n-1)}{2}\frac{|\rho|}{n}\left(1-\frac{\rho^2}{n^2}\right)^{-\frac{1}{2}}\exp\left(-\frac{u_n^2(x)}{1+\frac{|\rho|}{n}}\right)\quad\leqslant\quad K\,n\exp\left(-\frac{u_n^2(x)}{1+\frac{1}{n}}\right)\,.
\end{aligned}$$

Now eventually $u_n(x)=\alpha_n^*(1+o(1))\geqslant\kappa\alpha_n^*$ for any $\kappa<1$ and

$$n\exp\left(-\frac{u_n^2(x)}{1+\frac{1}{n}}\right)\quad\leqslant\quad n\exp\left(-\frac{2\kappa^2\log(n)}{1+\frac{1}{n}}\right)\quad=\quad n^{-\left(\frac{2\kappa^2}{1+\rho/n}-1\right)}\,.$$

This last term goes to zero as $n\to\infty$ for a well-chosen κ sufficiently close to one. This concludes the proof for $\mathcal{M}_n$. The proof for $\widetilde{\mathcal{M}}_n$ can be handled similarly with minor modifications. □

Random Matrix Theory

Let B_n be given by model (2.3).

Lemme 2.6. *Let* A_n *a* $n\times n$ *matrix with i.i.d.* $\mathcal{N}(0,1)$ *entries for* $i\leqslant j$ *and* (A_{ij},A_{ji}) *a standard bivariate Gaussian vector with covariance* ρ *for* $i<j$*, then the following estimate holds true: almost surely,*

$$\limsup_{n\to\infty}\left\|\frac{A_n}{\sqrt{n}}\right\|\quad\leqslant\quad\sqrt{2}\left(\sqrt{1+\rho}+\sqrt{1-\rho}\right)\quad\leqslant\quad 2\sqrt{2}\,.$$

Proof. The proof relies on two arguments: the classical estimate of the asymptotic spectral norm of a Wigner matrix [BS10, Th. 5.1] and the following decomposition of matrix $A_n/\sqrt{n}$ as linear combination of Hermitian Wigner matrices:

$$\frac{A_n}{\sqrt{n}}=\frac{A_n+A_n^\top}{2\sqrt{n}}-\boldsymbol{i}\frac{\left[\boldsymbol{i}\left(A_n-A_n^\top\right)\right]}{2\sqrt{n}}\,,\qquad(\boldsymbol{i}^2=-1)\,. \tag{2.18}$$

Notice that both matrices $W_n^1=\frac{A_n+A_n^\top}{2\sqrt{n}}$ and $W_n^2=\frac{\left[\boldsymbol{i}\left(A_n-A_n^\top\right)\right]}{2\sqrt{n}}$ are Wigner matrices, with off-diagonal variances $(i<j)$:

$$\operatorname{var}\left(\left[\frac{A_n+A_n^\top}{2}\right]_{ij}\right)=\frac{1+\rho}{2}\qquad\text{and}\qquad\operatorname{var}\left(\left[\frac{\boldsymbol{i}\left(A_n-A_n^\top\right)}{2}\right]_{ij}\right)=\frac{1-\rho}{2}\,.$$

Hence,

$$\limsup_n\left\|\frac{A_n}{\sqrt{n}}\right\|\leqslant\limsup_n\|W_n^1\|+\limsup_n\|W_n^2\|=2\left(\sqrt{\frac{1+\rho}{2}}+\sqrt{\frac{1-\rho}{2}}\right)$$

An elementary analysis yields $\sqrt{2}(\sqrt{1+\rho}+\sqrt{1-\rho})\leqslant 2\sqrt{2}$ for $|\rho|\leqslant 1$. □

2.3.2 Proof of Theorem 2.1 - the centered case $\mu = 0$

Some preparation and strategy of proof

We first prove Theorem 2.1 in the case where $\mu = 0$ and focus on the equation

$$\boldsymbol{x}_n = \mathbf{1}_n + \frac{A_n}{\alpha_n\sqrt{n}}\boldsymbol{x}_n\,. \tag{2.19}$$

By Lemma 2.6, $\limsup_n \|A_n/\sqrt{n}\|$ is a.s. bounded hence

$$\left\|\frac{A_n}{\alpha_n\sqrt{n}}\right\| \xrightarrow[n\to\infty]{a.s.} 0\,.$$

As a consequence, the resolvent $Q_n = (I_n - A_n/(\alpha_n\sqrt{n}))^{-1}$ is a.s. eventually well-defined and the solution $\boldsymbol{x}_n = (x_k)_{k\in[n]}$ of (2.19) writes $\boldsymbol{x}_n = Q_n\mathbf{1}_n$. Denote by $\boldsymbol{e}_k$ the kth canonical vector of $\mathbb{R}^n$. The following representation holds true (we shall often drop index n in the following):

$$\begin{aligned} x_k &= \boldsymbol{e}_k^\top \boldsymbol{x} = \boldsymbol{e}_k Q\mathbf{1} = \sum_{\ell=0}^{\infty} \boldsymbol{e}_k^\top \left(\frac{A}{\alpha_n\sqrt{n}}\right)^\ell \mathbf{1}\,, \\ &= 1 + \frac{1}{\alpha_n}\boldsymbol{e}_k^\top\left(\frac{A}{\sqrt{n}}\right)\mathbf{1} + \frac{1}{\alpha_n^2}\boldsymbol{e}_k^\top\left(\frac{A}{\sqrt{n}}\right)^2 Q\,\mathbf{1}\,. \end{aligned} \tag{2.20}$$

Denote by

$$Z_{k,n} = \boldsymbol{e}_k^\top\left(\frac{A}{\sqrt{n}}\right)\mathbf{1} = \frac{1}{\sqrt{n}}\sum_i A_{ki} \quad \text{and} \quad R_{k,n}(A) = \boldsymbol{e}_k^\top\left(\frac{A}{\sqrt{n}}\right)^2 Q\,\mathbf{1}\,. \tag{2.21}$$

Notice that the $Z_{k,n}$'s are standard $\mathcal{N}(0,1)$ however they are not independent as

$$\mathrm{cov}(Z_{k,n}, Z_{\ell,n}) = \frac{1}{n}\mathrm{cov}(A_{k\ell}, A_{\ell k}) = \frac{\rho}{n}\,, \quad k \neq \ell\,.$$

Introducing $M_n = \max_{k\in[n]} Z_{k,n}$ and $\breve{M}_n = \min_{k\in[n]} Z_{k,n}$, we proved in Corollary 2.5 that

$$\mathbb{P}\left\{\alpha_n^*(M_n - \beta_n^*) \leqslant x\right\}\ ,\ \mathbb{P}\left\{\alpha_n^*(\breve{M}_n + \beta_n^*) \geqslant -x\right\} \xrightarrow[n\to\infty]{} G(x)\,. \tag{2.22}$$

In the sequel, we often drop n and simply write $R_k(A)$ instead of $R_{k,n}(A)$. Following the same strategy as in [BN21], we notice that (2.20) yields

$$\begin{cases} \min_{k\in[n]} x_k &\geqslant 1 + \frac{1}{\alpha_n}\breve{M} + \frac{1}{\alpha_n^2}\min_{k\in[n]} R_k(A) \\ \min_{k\in[n]} x_k &\leqslant 1 + \frac{1}{\alpha_n}\breve{M} + \frac{1}{\alpha_n^2}\max_{k\in[n]} R_k(A) \end{cases},$$

which we can rewrite

$$\begin{aligned} \min_{k\in[n]} x_k &\geqslant 1 + \frac{\alpha_n^*}{\alpha_n}\left(\frac{\breve{M}+\beta_n^*}{\alpha_n^*} - \frac{\beta_n^*}{\alpha_n^*} + \frac{\min_{k\in[n]} R_k(A)}{\alpha_n^*\alpha_n}\right)\,, \\ &= 1 + \frac{\alpha_n^*}{\alpha_n}\left(-1 + o_P(1) + \frac{\min_{k\in[n]} R_k(A)}{\alpha_n^*\alpha_n}\right)\,, \end{aligned}$$

where we have used the fact that $\frac{\breve{M}+\beta_n}{\alpha_n^*} = o(1)$, cf. (2.22). Similarly, we have:

$$\min_{k\in[n]} x_k \leqslant 1 + \frac{\alpha_n^*}{\alpha_n}\left(-1 + o_P(1) + \frac{\max_{k\in[n]} R_k(A)}{\alpha_n^*\alpha_n}\right) \tag{2.23}$$

The proof in the centered case follows then from the following lemma:

Lemme 2.7. *Let $R_{k,n}(A)$ be defined as in (2.21) and recall that $\alpha_n \xrightarrow[n\to+\infty]{} +\infty$, then:*

$$\frac{\max_{k\in[n]} R_{k,n}(A)}{\alpha_n\sqrt{2\log n}} \xrightarrow[n\to\infty]{\mathcal{P}} 0 \qquad \text{and} \qquad \frac{\min_{k\in[n]} R_{k,n}(A)}{\alpha_n\sqrt{2\log n}} \xrightarrow[n\to\infty]{\mathcal{P}} 0 \ .$$

The remaining of the section is devoted to the proof of Lemma 2.7.

Lipschitziannity and Gaussian concentration

We first introduce a truncated version of $R_{k,n}(A)$. Let $\eta \in (0,1)$ and $\varphi : \mathbb{R}^+ \to [0,1]$ a smooth function satisfying:

$$\varphi(x) = \begin{cases} 1 & \text{if } x \in [0, 2\sqrt{2}+\eta] \\ 0 & \text{if } x \geqslant 4 \end{cases} , \tag{2.24}$$

decreasing from 1 to 0 gradually as x goes from $2\sqrt{2}+\eta$ to 4. Let

$$\widetilde{R}_{k,n}(A) = \varphi_n R_{k,n}(A) \qquad \text{where} \qquad \varphi_n = \varphi\left(\left\|\frac{A_n}{\sqrt{n}}\right\|\right) . \tag{2.25}$$

Notice that $\widetilde{R}_k(A)$ differs from $R_k(A)$ if $\varphi_n < 1$ which happens with vanishing probability as $\mathbb{P}\{\varphi_n < 1\} = \mathbb{P}\left\{s_n > 2\sqrt{2}+\eta\right\} \xrightarrow[n\to\infty]{} 0$ by Lemma 2.6. The following lemma is a first step towards Gaussian concentration.

Lemme 2.8. *Let $\widetilde{R}_k$ defined by (2.25) and M an $n \times n$ matrix. Then the function*

$$M \mapsto \widetilde{R}_k(M) = \boldsymbol{e}_k^\top \left(\frac{M}{\sqrt{n}}\right)^2 \left(I - \frac{M}{\alpha_n\sqrt{n}}\right)^{-1} \mathbf{1}$$

is K-Lipschitz, i.e.

$$\left|\widetilde{R}_k(M) - \widetilde{R}_k(N)\right| \leqslant K \left\|M - N\right\|_F \tag{2.26}$$

where M, N are $n \times n$ matrices, $\|M\|_F = \sqrt{\sum_{ij} |M_{ij}|^2}$ is the Frobenius norm and K a constant independent from k and n.

The second step is to notice that $\widetilde{R}_k(A)$ (where A has Gaussian entries but with off-diagonal pairwise correlations) can be in fact expressed as a Lipschitz function of i.i.d. $\mathcal{N}(0,1)$ entries.

Lemme 2.9. *Consider the linear function $\Gamma : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ defined by*

$$\Gamma_{ii}(X) = X_{ii} \qquad \text{and} \qquad \begin{cases} \Gamma_{ij}(X) = \sqrt{\frac{1+\rho}{2}} X_{ij} + \sqrt{\frac{1-\rho}{2}} X_{ji} & (i<j) \ , \\ \Gamma_{ji}(X) = \sqrt{\frac{1+\rho}{2}} X_{ij} - \sqrt{\frac{1-\rho}{2}} X_{ji} & (i<j) \ . \end{cases}$$

Then

1. *We have $\|\Gamma(X)\|_F \leqslant K_\rho \|X\|_F$ where $K_\rho = 2\sqrt{1+|\rho|}$ hence Γ is K_ρ-Lipschitz.*
2. *If matrix $X_n = (X_{ij})$ has i.i.d. $\mathcal{N}(0,1)$ entries, then $A_n = \Gamma(X_n)$ has i.i.d. $\mathcal{N}(0,1)$ entries on and above the diagonal ($i \leqslant j$) and each vector (A_{ij}, A_{ji}) is a standard bivariate Gaussian vector with covariance ρ for $i < j$.*

The proof is straightforward and is thus omitted.

A consequence of this lemma is that $\widetilde{R}_k(A) = \widetilde{R}_k(\Gamma(X))$ is $K \times K_\rho$-Lipschitz. Applying Tsirelson-Ibragimov-Sudakov inequality [BLM13, Theorem 5.5] finally yields:

Proposition 2.10. *Let K the Lipschitz constant of Lemma 2.8 and $K_\rho = 2\sqrt{1+|\rho|}$. Then*

$$\mathbb{E} \max_{k\in[n]} \left(\widetilde{R}_k(A) - \mathbb{E}\widetilde{R}_k(A) \right) \leqslant 2\, K_\rho\, K \sqrt{\log n}\,.$$

Details of the proof are similar to those in [BN21] and are thus omitted.

Remark 2.6. Notice that $\varphi_n \leqslant 1$ and that $\varphi_n = 0$ if $\|A/\sqrt{n}\| \geqslant 4$. In particular,

$$\varphi_n \left\| \frac{A}{\sqrt{n}} \right\| \;\leqslant\; 4 \qquad \text{and} \qquad \varphi_n \|Q\| \;\leqslant\; \frac{1}{1-4\alpha_n^{-1}} \;\leqslant\; 2$$

for n large enough. For the latter estimate, write $Q = \left(I - \frac{A}{\alpha_n\sqrt{n}}\right)^{-1}$, $Q^{-1}Q = I$ and $Q = I + \frac{A}{\alpha_n\sqrt{n}}Q$, then apply the triangular inequality.

Proposition 2.11. *The following estimate $\mathbb{E}\widetilde{R}_k(A_n) = \mathcal{O}(1)$ holds true, uniformly for $k \in [n]$.*

Proof. We shall prove that the variables $\widetilde{R}_k$ have a common distribution for $k \in [n]$, which in particular implies that

$$\mathbb{E}\widetilde{R}_k = \mathbb{E}\widetilde{R}_i\,, \quad \forall k,i \in [n] \qquad \text{and} \qquad \mathbb{E}\widetilde{R}_k = \frac{1}{n}\sum_{i\in[n]} \mathbb{E}\widetilde{R}_i\,. \tag{2.27}$$

Once this fact is established, the proof is straightforward:

$$\left|\mathbb{E}\widetilde{R}_k\right| = \left|\frac{1}{n}\sum_{i\in[n]} \mathbb{E}\widetilde{R}_i\right| = \left|\frac{1}{n}\mathbb{E}\varphi_n \mathbf{1}^\top \left(\frac{A}{\sqrt{n}}\right)^2 Q\mathbf{1}\right| \leqslant \left\|\frac{\mathbf{1}}{\sqrt{n}}\right\|^2 \mathbb{E}\varphi_n \left\|\frac{A}{\sqrt{n}}\right\|^2 \|Q\| = \mathcal{O}(1)\,,$$

where the last equality follows from the arguments developed in Remark 2.6.

Let us now establish (2.27).

Denote by Δ_σ the matrix associated to the permutation $\sigma : [n] \mapsto [n]$ and defined by

$$[\Delta_\sigma]_{ij} = \begin{cases} 1 & \text{if } i = \sigma(j) \\ 0 & \text{else} \end{cases}\,.$$

Notice in particular that $\Delta_\sigma \boldsymbol{e}_i = \boldsymbol{e}_{\sigma(i)}$, $\Delta_\sigma\Delta_\tau = \Delta_{\sigma\tau}$ for σ,τ two permutations and $\Delta_{\sigma^{-1}} = \Delta_\sigma^\top$. Denote by (ij) the transposition swapping i and j, i.e. $(ij)i = j$, $(ij)j = i$ and $(ij)\ell = \ell$ otherwise. We consider $\breve{A} = \Delta_{(ij)}A\Delta_{(ij)}$, that is $\breve{A}$ is obtained by swapping A's ith and jth column, then the ith and jth row. Observe that A and $\breve{A}$ have the same distribution and so is the case for $R_k(A)$ and $R_k(\breve{A})$.

We have $\Delta_{(ij)}^2 = I_n$, implying that $\breve{A}^k = \Delta_{(ij)}A^k\Delta_{(ij)}$ and then

$$\begin{aligned} R_i(\breve{A}) &= \boldsymbol{e}_i^\top \sum_{k\geqslant 2} \frac{1}{\alpha_n^{k-2}} \left(\frac{\breve{A}}{\sqrt{n}}\right)^k \mathbf{1} &= \boldsymbol{e}_i^\top \Delta_{(ij)} \sum_{k\geqslant 2} \frac{1}{\alpha_n^{k-2}} \left(\frac{A}{\sqrt{n}}\right)^k \Delta_{(ij)}\mathbf{1} \\ &= \boldsymbol{e}_j^\top \sum_{k\geqslant 2} \frac{1}{\alpha_n^{k-2}} \left(\frac{A}{\sqrt{n}}\right)^k \mathbf{1} &= R_j(A)\,. \end{aligned}$$

This proves that $R_i(A), R_i(\check{A}), R_j(A)$ have the same law, hence the same expectation. Eq.(2.27) is established, which concludes the proof. □

We are now in position to prove Lemma 2.7.

Proof of lemma 2.7. Recall that $\mathbb{E}\widetilde{R}_k(A) = \mathbb{E}\widetilde{R}_1$. Since $\max_{k\in[n]} \widetilde{R}_k(A) - \widetilde{R}_1(A) \geqslant 0$, Markov inequality yields:

$$\begin{aligned}
\mathbb{P}\left\{ \frac{\max_{k\in[n]} \widetilde{R}_k(A) - \widetilde{R}_1(A)}{\alpha_n\sqrt{2\log n}} \geqslant \varepsilon \right\} &\leqslant \frac{\mathbb{E}\left(\max_{k\in[n]} \widetilde{R}_k(A) - \widetilde{R}_1(A)\right)}{\varepsilon\alpha_n\sqrt{2\log n}}, \\
&= \frac{\mathbb{E}\left(\max_{k\in[n]} \widetilde{R}_k\left(A\right) - \mathbb{E}\widetilde{R}_k(A)\right)}{\varepsilon\alpha_n\sqrt{2\log n}}, \\
&= \frac{\mathbb{E}\left(\max_{k\in[n]} \left(\widetilde{R}_k\left(A\right) - \mathbb{E}\widetilde{R}_k(A)\right)\right)}{\varepsilon\alpha_n\sqrt{2\log n}}, \\
&\leqslant \frac{\sqrt{2}K \times K_\rho}{\varepsilon\alpha_n},
\end{aligned}$$

where the last inequality follows from Proposition 2.10.
This implies that

$$\frac{\max_{k\in[n]} \widetilde{R}_k\left(A\right) - \widetilde{R}_1\left(A\right)}{\alpha_n\sqrt{2\log n}} \xrightarrow[n\to\infty]{\mathcal{P}} 0\,.$$

It remains to prove that

$$\frac{\widetilde{R}_1\left(A\right)}{\alpha_n\sqrt{2\log n}} \xrightarrow[n\to\infty]{\mathcal{P}} 0 \qquad \text{and} \qquad \frac{\max_{k\in[n]} R_k\left(A\right)}{\alpha_n\sqrt{2\log n}} \xrightarrow[n\to\infty]{\mathcal{P}} 0\,.$$

The arguments are similar to those in [BN21, Section 2.3]. Proof of the second assertion of Lemma 2.7 can be done similarly. This concludes the proof. □

2.3.3 Proof of Theorem 2.1 - the non centered case.

Recall that $\alpha_n \to \infty$ as $n \to \infty$. Denote by $\boldsymbol{u}_n = \frac{1}{\sqrt{n}}\mathbf{1}_n$ and notice that the spectrum of $I_n - \mu\boldsymbol{u}_n\boldsymbol{u}_n^\top$ is $\{1-\mu, 1\}$, the eigenvalue 1 with multiplicity $n-1$. Notice in particular that if $\mu \neq 1$, then $I - \mu\boldsymbol{u}\boldsymbol{u}^\top$ is invertible. So is (eventually) $I - \frac{A}{\alpha_n\sqrt{n}} - \mu\boldsymbol{u}\boldsymbol{u}^\top$ as $\|A/(\alpha_n\sqrt{n})\| \to 0$ a.s. We shall also rely on the fact that $\|Q - I\| \xrightarrow[n\to\infty]{} 0$ a.s. As a consequence,

$$\boldsymbol{u}^\top Q\boldsymbol{u} \xrightarrow[n\to\infty]{a.s.} 1\,.$$

Denote by $\tilde{\boldsymbol{x}}$ and $\boldsymbol{x}$ the vectors solutions of the equations:

$$\tilde{\boldsymbol{x}} = \mathbf{1} + B\tilde{\boldsymbol{x}} = \mathbf{1} + \left(\frac{A}{\alpha_n\sqrt{n}} + \mu\boldsymbol{u}\boldsymbol{u}^\top\right)\tilde{\boldsymbol{x}} \qquad \text{and} \qquad \boldsymbol{x} = \mathbf{1} + \frac{A}{\alpha_n\sqrt{n}}\boldsymbol{x}\,.$$

The following representations hold:

$$\tilde{\boldsymbol{x}} = (I - B)^{-1}\mathbf{1} \qquad \text{and} \qquad \boldsymbol{x} = \left(I - \frac{A}{\alpha_n\sqrt{n}}\right)^{-1}\mathbf{1}\,.$$

Recall that $Q = (I - A/(\alpha_n\sqrt{n}))^{-1}$. By rank one perturbation identity (Woodbury), we have:

$$(I-B)^{-1} = Q + \frac{Q\boldsymbol{u}\boldsymbol{u}^\top Q}{1-\mu\boldsymbol{u}^\top Q\boldsymbol{u}}$$

and

$$\tilde{\boldsymbol{x}} \quad = \quad \frac{Q\mathbf{1}(1-\mu\boldsymbol{u}^\top Q\boldsymbol{u}) + \mu Q\boldsymbol{u}\boldsymbol{u}^\top Q\mathbf{1}}{1-\mu\boldsymbol{u}\top Q\boldsymbol{u}} \quad = \quad \frac{\boldsymbol{x}}{1-\mu\boldsymbol{u}^\top Q\boldsymbol{u}}\,.$$

If $\mu < 1$ and $\alpha_n \geqslant (1+\varepsilon)\alpha_n^*$ then eventually, $\tilde{\boldsymbol{x}}$ has positive components. This is no longer the case if $\mu > 1$ or $\alpha_n \leqslant (1-\varepsilon)\alpha_n^*$. This concludes the proof of Theorem 2.1.

2.4 Stability: Proof of Proposition 2.3

Proof. We have

$$I - B + I - B^\top = 2I - (B + B^\top) = 2I - \left(\frac{\check{A} + \check{A}^\top}{\alpha\sqrt{n}} + \frac{2\mu}{n}\mathbf{1}\mathbf{1}^\top\right).$$

We will rely on the following condition:

$$2I - (B + B^\top) \text{ is positive definite} \quad \Leftrightarrow \quad \lambda_{\max}(B + B^\top) < 2\,. \tag{2.28}$$

Notice that $(\check{A} + \check{A}^\top)/\alpha$ is a symmetric matrix with independent centered entries with variance $2(1+\rho)/\alpha^2$) above the diagonal (the diagonal entries have a different distribution from the off-diagonal entries, with no asymptotic effect). Notice that by assumption, these entries have a finite fourth moment. In this case, it is well known that the largest eigenvalue of the normalized matrix (or equivalently its spectral norm since the matrix is symmetric) almost surely converges to the right edge of the support of the semi-circle law (see [BS10, Theorem 5.2]):

$$\lambda_{\max}\left(\frac{\check{A} + \check{A}^\top}{\alpha\sqrt{n}}\right) \quad \xrightarrow[n\to\infty]{a.s.} \quad \frac{2\sqrt{2(1+\rho)}}{\alpha}\,. \tag{2.29}$$

Suppose that $(\rho, \alpha, \mu) \in \mathcal{A}$. Notice that in this case,

$$\frac{\sqrt{1+\rho}}{\alpha\sqrt{2}} \quad < \quad \frac{1}{2} \quad < \quad \frac{1}{2} + \frac{1}{2}\sqrt{1 - \frac{2(1+\rho)}{\alpha^2}}\,.$$

We consider three subcases

(i) $\mu = 0$,

(ii) $\mu \leqslant \frac{\sqrt{1+\rho}}{\alpha\sqrt{2}}$,

(iii) $\mu \in \left(\frac{\sqrt{1+\rho}}{\alpha\sqrt{2}}\ ,\ \frac{1}{2} + \frac{1}{2}\sqrt{1 - \frac{2(1+\rho)}{\alpha^2}}\right)$.

In the centered case (i), condition (2.28) asymptotically occurs whenever $\alpha > \sqrt{2(1+\rho)}$.

Before studying subcases (ii) and (iii), we recall a result on small rank perturbations of large random matrices.

Notice that the trend matrix $P = \frac{2\mu}{n}\mathbf{1}\mathbf{1}^\top$ admits a unique non zero eigenvalue 2μ. Denote by $H = \frac{A+A^\top}{\alpha\sqrt{n}}$. We are concerned with the top eigenvalue of the symmetric matrix $H + P$. Based on a result by Capitaine *et al.* [CDMF09, Theorem 2.1], we have:

$$\lambda_{\max}(H+P) \xrightarrow[n\to\infty]{a.s} \begin{cases} 2\mu + \frac{1+\rho}{\alpha^2\mu} & \text{if } \mu > \frac{\sqrt{1+\rho}}{\sqrt{2}\alpha}\,, \\ \frac{2\sqrt{2(1+\rho)}}{\alpha} & \text{else.} \end{cases}$$

Consider now subcase (ii), then $\lambda_{\max}(H+P) \xrightarrow[n\to\infty]{a.s.} \frac{2\sqrt{2(1+\rho)}}{\alpha}$, which is strictly lower than 2 since $(\rho,\alpha,\mu) \in \mathcal{A}$. Hence $\lambda_{\max}(H+P)$ is eventually strictly lower than 2 in this case.

We finally consider subcase (iii). In this case,

$$\lambda_{\max}(H+P) \xrightarrow[n\to\infty]{a.s.} 2\mu + \frac{1+\rho}{\alpha^2\mu}\,.$$

We shall prove that $2\mu + \frac{1+\rho}{\alpha^2\mu} < 2$ or equivalently

$$2\alpha^2\mu^2 - 2\alpha^2\mu + 1 + \rho < 0\,. \tag{2.30}$$

An elementary study of the polynomial $\Omega(X) = 2\alpha^2X^2 - 2\alpha^2X + 1 + \rho$ yields that Ω's discriminant is positive if $\alpha > \sqrt{2(1+\rho)}$ and Ω's roots are given by

$$\Omega(\mu^\pm) = 0 \quad \Leftrightarrow \quad \mu^\pm = \frac{1}{2} \pm \frac{1}{2}\sqrt{1 - \frac{2(1+\rho)}{\alpha^2}}\,.$$

Also remark that $\Omega\left(\frac{\sqrt{1+\rho}}{\alpha\sqrt{2}}\right) < 0$, so that $\frac{\sqrt{1+\rho}}{\alpha\sqrt{2}} \in (\mu^-,\mu^+)$. In particular condition (2.30) is fulfilled for $\mu \in \left(\frac{\sqrt{1+\rho}}{\alpha\sqrt{2}}, \mu^+\right)$, which is precisely subcase (iii). Hence a.s. $\limsup_{n\to\infty} \lambda_{\max}(H+P) < 2$. We can then rely on Theorem 2.2 to conclude.

□

Appendix

2.A Proof of Theorem 2.1: adaptations to the case of a covariance profile

In this section, we provide the arguments to prove Theorem 2.1 in the case where matrix B follows the model (2.5), i.e.

$$B = \frac{\tilde{A}_n}{\alpha_n \sqrt{n}} + \frac{\mu}{n} \mathbf{1}_n \mathbf{1}_n^\top ,$$

where $\tilde{A}_n$'s entries are i.i.d. $\mathcal{N}(0,1)$ on and above the diagonal ($i \leqslant j$), and $(\tilde{A}_{ij}, \tilde{A}_{ji})$ is a standard bivariate Gaussian vector ($i < j$) with covariance $\mathrm{cov}(\tilde{A}_{ij}, \tilde{A}_{ji}) = \rho_{ij}$, and independent from the remaining random variables.

There are essentially 3 issues to resolve, to fully adapt the proof developed in Section 2.3 to the covariance profile case:

1. The decomposition (2.18) yields $\frac{\tilde{A}_n}{\sqrt{n}} = W_n^1 + \boldsymbol{i} W_n^2$, where W_n^1, W_n^2 are Hermitian matrices with

$$\mathrm{var}\left([W_n^1]_{ij}\right) = \frac{1+\rho_{ij}}{2n} \qquad \text{and} \qquad \mathrm{var}\left([W_n^2]_{ij}\right) = \frac{1-\rho_{ij}}{2n} .$$

   Since W_n^1, W_n^2 are no longer Wigner matrices, but rather matrices with a variance profile, an extra argument is needed to obtain an almost-sure upper bound for $\limsup_n \|W_n^1\| + \limsup_n \|W_n^2\|$.

2. The Lipschitz property for $\widetilde{R}_{k,n}(\tilde{A}_n)$. Essentially, we need the counterpart of Lemma 2.9 to the context of a covariance profile.

3. The control of the term $\mathbb{E}\, \widetilde{R}_{k,n}(\tilde{A})$.

2.A.1 Proof of issue 1: Control of the spectral norm of a Hermitian matrix with a variance profile

Applying Latała's theorem [Lat05], we easily show that

$$\mathbb{E}\|W_n^1\| + \mathbb{E}\|W_n^2\| \leqslant C$$

where C is a constant independent from n.

Now write

$$W_n^1 = \frac{\Upsilon_n \circ X_n}{\sqrt{n}} \qquad \text{where} \qquad \Upsilon_n = (\Upsilon_{ij}) \ , \quad \Upsilon_{ij} = \sqrt{\frac{1+\rho_{ij}}{2}},$$

matrix $X_n = (X_{ij})$ is a Wigner matrix with i.i.d. $\mathcal{N}(0,1)$ entries on and above the diagonal, and $\circ$ stands for the Hadamard product, i.e. $\Upsilon_n \circ X_n = (\Upsilon_{ij} X_{ij})$. Notice that $\sqrt{n} W_n^1$ is 1-Lipschitz with respect to the Frobenius norm

$$\|X_n\|_{\text{Frob}} = \sqrt{\sum_{ij} |X_{ij}|^2} \, .$$

Hence by Gaussian concentration, we have

$$\mathbb{P}\left\{ \left| \sqrt{n} \|W_n^1\| - \sqrt{n} \mathbb{E}\|W_n^1\| \right| > \delta \right\} \leqslant 2e^{-\frac{\delta^2}{2}} \, .$$

Taking $\delta = \varepsilon\sqrt{n}$, we obtain

$$\mathbb{P}\left\{ \left| \|W_n^1\| - \mathbb{E}\|W_n^1\| \right| > \varepsilon \right\} \leqslant 2e^{-\frac{n\varepsilon^2}{2}} \, .$$

The same holds for W_n^2, hence the upper control:

$$\limsup_n \left(\|W_n^1\| + \|W_n^2\| \right) \quad \leqslant \quad \limsup_n \left(\mathbb{E}\|W_n^1\| + \mathbb{E}\|W_n^2\| \right) \quad \leqslant \quad C$$

almost surely. It remains to replace the truncation function φ in (2.24) by the smooth function

$$\psi(x) = \begin{cases} 1 & \text{if } x \leqslant C + \eta, \\ 0 & \text{else.} \end{cases}$$

to proceed.

2.A.2 Proof of issue 2: $\widetilde{R}_k(\tilde{A})$ is a Lipschitz function of Gaussian i.i.d. random variables

It suffices to replace function Γ in Lemma 2.9 by

$$\widetilde{\Gamma} : \mathbb{R}^{n \times n} \to \mathbb{R}^{n \times n}$$

where

$$\widetilde{\Gamma}_{ii}(X) = X_{ii} \qquad \text{and} \qquad \begin{cases} \widetilde{\Gamma}_{ij}(X) = \sqrt{\frac{1+\rho_{ij}}{2}} X_{ij} + \sqrt{\frac{1-\rho_{ij}}{2}} X_{ji} & (i < j) \, , \\ \widetilde{\Gamma}_{ji}(X) = \sqrt{\frac{1+\rho_{ij}}{2}} X_{ij} - \sqrt{\frac{1-\rho_{ij}}{2}} X_{ji} & (i < j) \, . \end{cases}$$

and to modify accordingly the Lipschitz constant by $\widetilde{K} = 2\sqrt{2} \geqslant 2\sqrt{1 + \max_{ij} |\rho_{ij}|}$.

2.A.3 Proof of issue 3: Magnitude of $\mathbb{E}\, \widetilde{R}_{k,n}(\tilde{A}_n)$

To address this issue, we provide a quick argument which relies on Isserlis' theorem also called Wick's formula (see [Jan97, Th. 1.28]), highly dependent on the Gaussiannity of the entries.

Theorem 2.12 (Isserlis Theorem). *if $(X_1, \cdots, X_n)$ is a centered normal vector, then*

$$\mathbb{E}(X_1 X_2 \cdots X_n) = \sum_{\Pi} \prod_{\{i,j\} \in \Pi} \mathbb{E}(X_i X_j) \tag{2.31}$$

where the sum is over all the partitions Π of $[n]$ into pairs $\{i,j\}$, and the product over all the pairs contained in Π.

Recall that:

$$\mathbb{E}\widetilde{R}_k(\tilde{A}_n) = \sum_{\ell\geqslant 2}\mathbb{E}\left[\frac{\boldsymbol{e}_k^\top}{\alpha_n^{\ell-2}}\left(\frac{\tilde{A}}{\sqrt{n}}\right)^\ell \mathbf{1}\right] = \sum_{\ell\geqslant 2}\frac{\mathbb{E}\boldsymbol{e}_k^\top \tilde{A}^\ell \mathbf{1}}{\alpha_n^{\ell-2}n^{\frac{\ell}{2}}} =: \sum_{\ell\geqslant 2}\frac{C_\ell}{\alpha_n^{\ell-2}n^{\frac{\ell}{2}}}\,.$$

Consider a matrix $\overline{A}_n$ where the pairwise covariance $\mathrm{cov}(\overline{A}_{ij},\overline{A}_{ji}) = 1$. Denote by $\overline{C}_\ell = \mathbb{E}\boldsymbol{e}_k^\top \overline{A}^\ell \mathbf{1}$. We will show that each quantity $|C_\ell|$ is bounded by $\overline{C}_\ell$. Notice that:

$$C_\ell = \sum_{i_1,\dots,i_{\ell+1}} \mathbb{E}(\tilde{A}_{i_1 i_2}\tilde{A}_{i_2 i_3}\dots\tilde{A}_{i_\ell,i_{\ell+1}})\,. \tag{2.32}$$

By Isserlis' theorem, we have:

$$\begin{aligned}\left|\mathbb{E}(\tilde{A}_{i_1,i_2}\tilde{A}_{i_2,i_3}\dots\tilde{A}_{i_\ell,i_{\ell+1}})\right| &\leqslant \left|\sum_{\Pi}\prod_{\{j,k\}\in\Pi}\mathbb{E}(\overline{A}_{i_j i_{j+1}}\overline{A}_{i_k i_{k+1}})\right| \\ &\leqslant \sum_{\Pi}\prod_{\{j,k\}\in\Pi}\mathbb{E}(\overline{A}_{i_j i_{j+1}}\overline{A}_{i_k i_{k+1}}) \quad = \quad \mathbb{E}\left(\overline{A}_{i_1 i_2}\cdots\overline{A}_{i_\ell i_{\ell+1}}\right)\,.\end{aligned}$$

From this, we deduce that $|C_\ell| \leqslant \overline{C}_\ell$, hence $|\mathbb{E}\widetilde{R}_k(\tilde{A})| \leqslant \mathbb{E}\widetilde{R}_k(\overline{A})$. This gives the desired bound since $\mathbb{E}\widetilde{R}_k(\overline{A}) = \mathcal{O}(1)$.

2.B Details on the system of equations (2.10)-(2.13)

The system of equations (2.10)-(2.13) is a re-interpretation of the final form of Bunin's equations to estimate the proportion of the surviving species. Note that for specific details on the dynamical cavity method, see Bunin [Bun17]. First recall the following quantities:

$$\phi = \frac{\mathrm{Card}\{x_i^* > 0, i\in[n]\}}{n}\,,\quad \langle \boldsymbol{x}\rangle = \frac{1}{n}\sum_{j=1}^n x_j^*\,,\quad \langle \boldsymbol{x}^2\rangle = \frac{1}{n}\sum_{j=1}^n (x_j^*)^2\,,$$

and v a parameter essentially related to the dynamical cavity method.

Starting from (2.4), the following change of variables is used

$$\forall k\in[n],\quad \overline{x}_k = \frac{x_k^*}{\frac{1}{n}\sum_{j=1}^n x_j^*} = \frac{x_k^*}{\langle \boldsymbol{x}\rangle}\,,$$

and expose nontrivial symmetries of the problem. In particular, this change induces the study of a new system of equations where the fixed point is precisely those of the replicator equations. Suppose that the system is at a stable equilibrium point resistant to invasions i.e. the reintroduction of an extinct species at small abundance, automatically implies its decay to zero.

To analyze this system, add a new species $\overline{x}_0$ and study the analogy between the properties of the solutions with n and $n+1$ species. After some computations (see Bunin [Bun17]), one remark that if $\overline{x}_0 > 0$ the distribution of the new invading species is a Gaussian variable i.e.

$$\overline{x}_0^+ = \frac{1}{\alpha-\rho v}\left[\alpha\left(\frac{1}{\langle \boldsymbol{x}\rangle}+\mu\right)+\frac{\sqrt{\langle \boldsymbol{x}^2\rangle}}{\langle \boldsymbol{x}\rangle}\right]Z \quad,\quad Z\sim\mathcal{N}(0,1)\,.$$

Given $\overline{x}_0$ the abundance of the new species and using the argument of the resistance of invasion, Bunin finally obtain the expression of abundance of the new species:

$$\overline{x}_0 = \max(0, \overline{x}_0^+)\,.$$

By an identification argument, when the new species is added to the system, it follow the same distribution as the other. We may generalized the distribution to all the species and after a revision of the change of variables, we recover an explicit distribution of each species included in vector of abundances $\boldsymbol{x}$:

$$\forall k \in [n],\ x_k^* = \max\left[0, x_k^{*+}\right] = \max\left[0, \frac{1}{1-\frac{\rho v}{\alpha}}(1 + \langle \boldsymbol{x}\rangle \mu + \frac{\sqrt{\langle \boldsymbol{x}^2\rangle}}{\alpha} Z_k)\right] \quad , \quad Z_k \sim \mathcal{N}(0,1)\,.$$

Notice that the distribution of the vector of abundances $\boldsymbol{x}$ is a truncated normal distribution on 0 (see [CMN22]).

The last step is to find a system of 4 equations, 4 unknowns: v, $\langle \boldsymbol{x}^2\rangle$ and $\langle \boldsymbol{x}\rangle$ and the proportion of persistent species ϕ. Given $\Delta = (1 + \langle \boldsymbol{x}\rangle \mu)\alpha/\sqrt{\langle \boldsymbol{x}^2\rangle}$, we obtain equation (2.10):

$$\begin{aligned}
\phi &= \mathbb{P}(x_k^* > 0)\,,\\
&= \mathbb{P}\left(\frac{1}{1-\frac{\rho v}{\alpha}}\left(1 + \langle \boldsymbol{x}\rangle \mu + \frac{\sqrt{\langle \boldsymbol{x}^2\rangle}}{\alpha} Z_k\right) > 0\right)\,,\\
&= \mathbb{P}\left(Z_k > -\left(1 + \langle \boldsymbol{x}\rangle \mu\right)\frac{\alpha}{\sqrt{\langle \boldsymbol{x}^2\rangle}}\right)\,,\\
&= \mathbb{P}(Z_k > -\Delta) = \frac{1}{\sqrt{2\pi}}\int_{-\Delta}^{+\infty} e^{\frac{-z^2}{2}}\,dz\,.
\end{aligned}$$

Then equation (2.11):

$$\begin{aligned}
\langle \boldsymbol{x}\rangle &\approx \mathbb{E}(x_k^*)\,,\\
&= \mathbb{E}\left(\max\left[0, x_k^{*+}\right]\right)\,,\\
&= \mathbb{E}(x_k^{*+}\mathbf{1}_{x_k^{*+}>0})\,,\\
&= \mathbb{E}(x_k^{*+} \mid x_k^{*+} > 0)\mathbb{P}(x_k^{*+} > 0)\,,\\
&= \mathbb{E}\left(\frac{1}{1-\frac{\rho v}{\alpha}}\left(1 + \langle \boldsymbol{x}\rangle \mu + \frac{\sqrt{\langle \boldsymbol{x}^2\rangle}}{\alpha} Z_k\right) \mid Z_k > -\Delta\right)\phi\,,\\
&= \frac{\phi}{1-\frac{\rho v}{\alpha}}(1 + \langle \boldsymbol{x}\rangle \mu) + \left(\frac{\phi}{1-\frac{\rho v}{\alpha}}\right)\frac{\sqrt{\langle \boldsymbol{x}^2\rangle}}{\alpha}\mathbb{E}(Z_k \mid Z_k > -\Delta)\,.
\end{aligned}$$

And then equation (2.12):

$$\begin{aligned}
\langle \boldsymbol{x}^2 \rangle &\approx \mathbb{E}((x_k^*)^2)\,, \\
&= \mathbb{E}(\max(0, x_k^{*+})^2)\,, \\
&= \mathbb{E}((x_k^{*+})^2 \mathbf{1}_{x_k^{*+}>0})\,, \\
&= \mathbb{E}((x_k^{*+})^2 \mid x_k^{*+} > 0)\mathbb{P}(x_k^{*+} > 0)\,, \\
&= \mathbb{E}\left(\left(\frac{1}{1-\frac{\rho v}{\alpha}}(1+\langle \boldsymbol{x}\rangle \mu + \frac{\sqrt{\langle \boldsymbol{x}^2\rangle}}{\alpha} Z_k)\right)^2 \mid Z_k > -\Delta\right)\phi\,, \\
&= \left(\frac{\sqrt{\phi}}{1-\frac{\rho v}{\alpha}}\right)^2 \left[(1+\langle \boldsymbol{x}\rangle \mu)^2 + 2(1+\langle \boldsymbol{x}\rangle \mu)\frac{\sqrt{\langle \boldsymbol{x}^2\rangle}}{\alpha}\mathbb{E}(Z_k \mid Z_k > -\Delta) + \frac{\langle \boldsymbol{x}^2\rangle}{\alpha^2}\mathbb{E}(Z_k^2 \mid Z_k > -\Delta)\right]\,,
\end{aligned}$$

where

$$\mathbb{E}(Z\mathbf{1}_{Z>-\Delta}) = \frac{1}{2\pi}\int_{-\Delta}^{+\infty} e^{\frac{-z^2}{2}} z dz \quad , \quad \mathbb{E}(Z^2\mathbf{1}_{Z>-\Delta}) = \frac{1}{2\pi}\int_{-\Delta}^{+\infty} e^{\frac{-z^2}{2}} z^2 dz\,.$$

The last equation (2.13) is intrinsically related to the dynamical cavity method and can be incorporated in the other three equations by resolving the 2nd order polynomial of v. Parameter v represents a major "turning point" in the understanding of the system of equations (2.10)-(2.13) in the elliptic model.

$$v = \phi\left(\frac{1}{\alpha - \rho v}\right)\,.$$

A system of four equations and four unknowns is obtained and can be solved numerically by evaluating the integrals. We refer to Bunin's article [Bun17] for the uniqueness of the solution of the system.

Chapter 3

Impact of a block structure in large systems of Lotka-Volterra

Abstract

The Lotka-Volterra (LV) system represents a simple, robust and versatile model used to describe large interacting systems such as food webs or microbiomes. This model consists of n coupled differential equations linking the abundances of the different species present in the system. When the number of species becomes very large, the true value of each interaction is difficult to observe or estimate, therefore the interactions between the different species can be modeled as random variables in order to understand the system dynamics. In this paper, we extend the LV model to describe the properties of a multi-community model. By adding a block structure to the matrix of interactions, we study the properties (feasibility, existence of an attrition phenomenon within each community) of distinct communities by adjusting the inter- and intra-community interactions. In particular, we analyze the properties and dynamics that emerge with two communities of interacting species. The interplay between the two communities affects their respective equilibrium and their resistance to small perturbations (stability).

Introduction

Understanding large ecosystems and the underlying mechanisms that enable high species diversity is a major challenge in theoretical ecology. Motivated by the seminal work of May [May72], the emergence of random matrices has been a key mathematical step to model high dimensional ecosystems. A better understanding of these tools has expanded our understanding of the nature of interactions and food webs to achieve stability [AT12, TPA14]. The Lotka-Volterra model [Lot25, Vol26] plays a key role in the study of population dynamics over time. This model has been studied both in ecology and in mathematics. The study of the stability of this model has raised much interest [GGRA18] as well as its feasibility i.e. conditions under which all species persist [BN21, GAS+17, Sto18].

In nature, ecological networks are rather structured and many studies have assessed the network structures that contribute to the stability of a given community [TF10, AGB+15]. A widespread network structure in nature is the compartmentalization of food webs, also called modularity. The underlying concept is that the network is structured in the form of groups of nodes that interact more strongly within their group and more weakly between groups. A mathematical formulation of modularity has been defined by Newman [New06]. Subsequently, modularity has been of great importance in ecology [GSSP+10], in complex networks, see Variano *et al.* [VML04] and Fortunato [For10] for a complete review of community detection (compartmentalization in ecology). May had already mentioned that a multi-community structure should improve stability [May72], a hypothesis that was later investigated by Pimm [Pim79]. Although studies show that modularity improves the persistence (:= non-extinction of the species, generally related to their resistance to external perturbations) of the species in the dynamical system [SB11]. Grilli *et al.* [GRA16] studied the impact of modularity on the stability of the Jacobian of a system, the so called "community matrix".

In this article, we consider a block structure network representing the inter- and intra-community interactions. Each block is identified by its interaction parameters, on the one hand the strength of the interactions i.e. the standard deviation of the random part of interactions and on the other hand a mean interaction parameter controlling on average the type of interactions (mutualism, competitiveness). The idea that the interaction strength plays a key role in the stability of ecosystems was brought by May [May72]. Rooney *et al.* [RMGM06] showed that real food webs are structured such that fast and slow channels convey stability to food webs. In this paper, we study some properties of distinct communities in interaction, in particular feasibility (i.e. whether there exists an equilibrium with all species at non-zero abundances) and the existence of an attrition phenomenon (some species may vanish) within each community. In particular, we analyze the properties and dynamics that emerge with two communities of interacting species. The interplay between the two communities affects their respective equilibrium and their resistance to small perturbations (stability). Finally, we investigate the similarities between the strength of interactions and the connectance in the Lotka-Volterra model with respect to the stability-complexity threshold of May [May72].

Model and assumptions. The Lotka-Volterra model is a standard model in ecology to study the dynamics of a community of species over time. This model describes the population dynamics of a n-species system. It is defined by a system of n differential

equations

$$\frac{dx_k(t)}{dt} = x_k(t)\left(r_k - \theta x_k(t) + \sum_{\ell\in[n]} B_{k\ell}x_\ell(t)\right) , \tag{3.1}$$

where $k \in [n] = \{1, \cdots, n\}$. The abundance of species k at time t is represented by $x_k(t)$ with $\boldsymbol{x} = (x_1, \cdots, x_n)$ the vector of abundances. Parameter θ is the self regulation coefficient or intra-specific interaction of each species. Parameter r_k corresponds to the growth rate of species k. The coefficient $B_{k\ell}$ represents the impact of species ℓ on species k. The matrix B, representing the interaction network, is decomposed in a block structure. This structure differentiates different groups of species in the form of communities that interact with each other. On the one hand, the diagonal blocks of B correspond to interactions within each community. Each community has its own strength and type of interaction. On the other hand, the off-diagonal blocks correspond to the impact of the communities on each other. Within the framework of two communities, the matrix $B = (B_{k\ell})_{n,n}$ is defined as

$$B = \frac{1}{\sqrt{n}}\begin{pmatrix} \frac{A_{11}}{\alpha_{11}} & \frac{A_{12}}{\alpha_{12}} \\ \frac{A_{21}}{\alpha_{21}} & \frac{A_{22}}{\alpha_{22}} \end{pmatrix} + \frac{1}{n}\begin{pmatrix} \mu_{11}\mathbf{1}_{\mathcal{I}_1}\mathbf{1}_{\mathcal{I}_1}^\top & \mu_{12}\mathbf{1}_{\mathcal{I}_1}\mathbf{1}_{\mathcal{I}_2}^\top \\ \mu_{21}\mathbf{1}_{\mathcal{I}_2}\mathbf{1}_{\mathcal{I}_1}^\top & \mu_{22}\mathbf{1}_{\mathcal{I}_2}\mathbf{1}_{\mathcal{I}_2}^\top \end{pmatrix} , \tag{3.2}$$

where:

- $\boldsymbol{\beta} = (\beta_1, \beta_2)$, $\sum_{i=1}^2 \beta_i = 1$ is the size by proportion of each of the blocks,
- $\mathcal{I}_1$ (resp $\mathcal{I}_2$) is a subset of $[n]$ of size $|\mathcal{I}_1| := \beta_1 n$ (resp $|\mathcal{I}_2| := \beta_2 n$) matching the index of species belonging to Community 1 (resp Community 2) where $\mathcal{I}_1 = [\beta_1 n]$ and $\mathcal{I}_2 = \{\beta_1 n + 1, \cdots, n\}$,
- $\mathbf{1}_{\mathcal{I}_i}$ is a vector whose entries are 1's of size $\beta_i n$.
- A_{ij} is a non hermitian random matrix of size $(\beta_i n, \beta_j n)$ with standard Gaussian entries i.e. $\mathcal{N}(0, 1)$.

To standardize the model, we define a matrix associated to the coefficient α_{ij} of each block as

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} .$$

The associated matrix representing the strength of the interactions, i.e. their standard deviation, is defined by

$$\boldsymbol{s} = \begin{pmatrix} \frac{1}{\alpha_{11}} & \frac{1}{\alpha_{12}} \\ \frac{1}{\alpha_{21}} & \frac{1}{\alpha_{22}} \end{pmatrix} .$$

The diagonal terms represent the interaction strength in each community. The off-diagonal terms s_{ij} represent the interaction strength of the impact of the community j on the community i. The lower the value of α_{ij}, the stronger the interactions. If α_{ij} is very large then the interactions are absent. A mean matrix $\boldsymbol{\mu}$ defined by

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_{11} & \mu_{12} \\ \mu_{21} & \mu_{22} \end{pmatrix} ,$$

is a matrix describing the average value of interaction coefficients. By extension, it also defines the dominant type of interactions within and among communities. A negative

value results in interactions to be on average detrimental to the species i.e. competition, while a positive value means the interactions are on average beneficial to the species i.e. mutualism.

In Figure 3.0.1, an interaction matrix of a 20-species ecosystem with two communities of different sizes is represented. The large number of parameters makes it possible to model many ecological situations. The results of the article can be mostly extended to the case of b blocks. In Appendix 3.C, we define the model with several (more than two) communities.

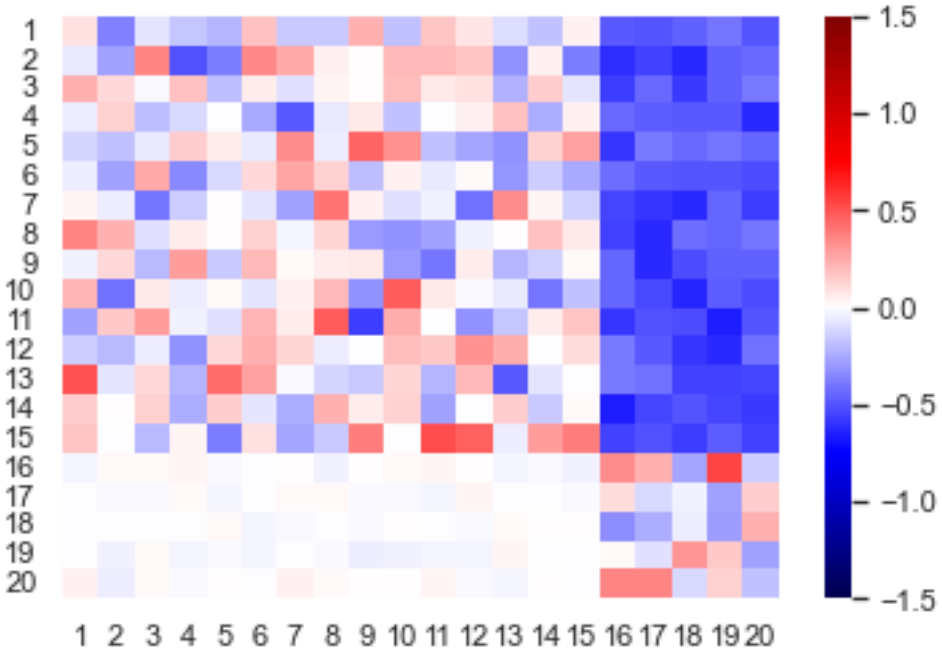


Figure 3.0.1: Representation of an interaction matrix (3.2) of a 20 species system in the form of a heatmap (the color tone gives information on the value of each entry of the matrix). There are two distinct communities of different size, their characteristics are given by the following parameters

$$n = 20 \; ; \; \beta_1 = .75 \; ; \; \boldsymbol{s} = \begin{pmatrix} 1 & 1/3 \\ 1/10 & 1 \end{pmatrix} \; ; \; \boldsymbol{\mu} = \begin{pmatrix} 0 & -10 \\ 0 & 0 \end{pmatrix} .$$

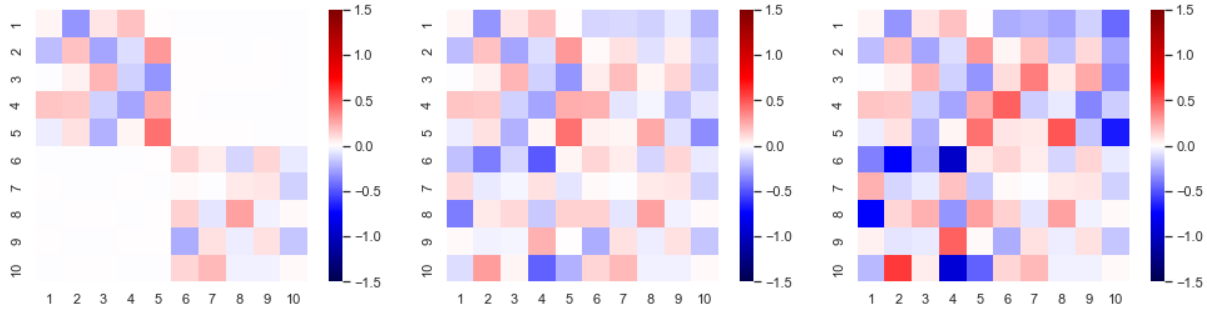
Community 1 has a low impact on Community 2 whereas Community 2 has a competitive (negative) impact on community one.

A typical example: assume two distinct groups of species following the dynamics of the Model (3.3) (see Figure 3.0.3). Initially, the two communities are not interacting

$$\boldsymbol{s} = \begin{pmatrix} 1/2 & \varepsilon \\ \varepsilon & 1/2 \end{pmatrix}, \varepsilon > 0 ,$$

and both are feasible in the sense that all species survive (see Figure 3.0.3 for $t < 5$). Starting from $t = 5$, the two communities begin to interact more and more with each other (α_{12}, α_{21} decrease with a linear growth, see Figure 3.0.2). The persistence of species

(i.e. the number of persisting species) decreases in both communities. The ranking of abundances does not seem to be respected, the species that have the highest abundance in Community 1 are not the ones that have the highest abundances when the interactions increase. In our example, the species with the highest abundance in Community 2 becomes extinct when the interactions between the two groups are increased.



(a) $t = 5$ (b) $t = 10$ (c) $t = 15$

Figure 3.0.2: Representation of the evolution of the interaction matrix (3.2) when two communities of 5-species start to interact with each other (see Figure 3.0.3). It is illustrated in the form of a heatmap (the color tone gives information on the value of each entry of the matrix). In Fig. (a), at $t = 5$ the two feasible communities are not interacting. In Fig. (b), at $t = 10$ the interaction strength are equal in and between the communities. In Fig. (c), $t = 15$ the inter-community interactions are strong, the communities reach a new equilibrium at $t = 20$ with vanishing species from both communities.

Properties of the dynamical system. We are interested in the impact of a block structure of the food web on its persistence and stability. We limit the study to the 2-block case i.e. B defined in (3.2) and we focus on the model where $r_k = 1, \forall k \in [n]$ and $\theta = 1$:

$$\frac{dx_k}{dt} = x_k \left(1 - x_k + (B\boldsymbol{x})_k\right) \ , \quad k \in [n] \, . \tag{3.3}$$

A key element to understand the dynamics of the LV system (3.3) is the existence of an equilibrium $\boldsymbol{x}^* = (x_k^*)_{k\in[n]}$ such that

$$\begin{cases} x_k^* \left(1 - x_k^* + (B\boldsymbol{x}^*)_k\right) = 0 \, , \quad \forall k \in [n] \, , \\ x_k^* \geqslant 0. \end{cases} \tag{3.4}$$

and the study of its stability, that is the convergence of a solution $\boldsymbol{x}$ to the equilibrium $\boldsymbol{x}^*$: $\boldsymbol{x}(t) \xrightarrow[t\to\infty]{} \boldsymbol{x}^*$ if $\boldsymbol{x}(0)$ is sufficiently close to $\boldsymbol{x}^*$.

The Lotka-Volterra system is an autonomous differential system. If the initial conditions are positive i.e. $\boldsymbol{x}(0) > 0$ (componentwise), it implies $\boldsymbol{x}(t) > 0$ for every $t > 0$. However, some of the components $x_k(t)$ may converge to zero if the equilibrium $\boldsymbol{x}^*$ has components equal to zero. In this article, we are considering two related behaviors. On the one hand, under a feasible equilibrium $\boldsymbol{x}^* > 0$ of (3.3), the equilibrium set of equations becomes a linear equation:

$$\boldsymbol{x}^* = \mathbf{1} + B\boldsymbol{x}^* \, . \tag{3.5}$$

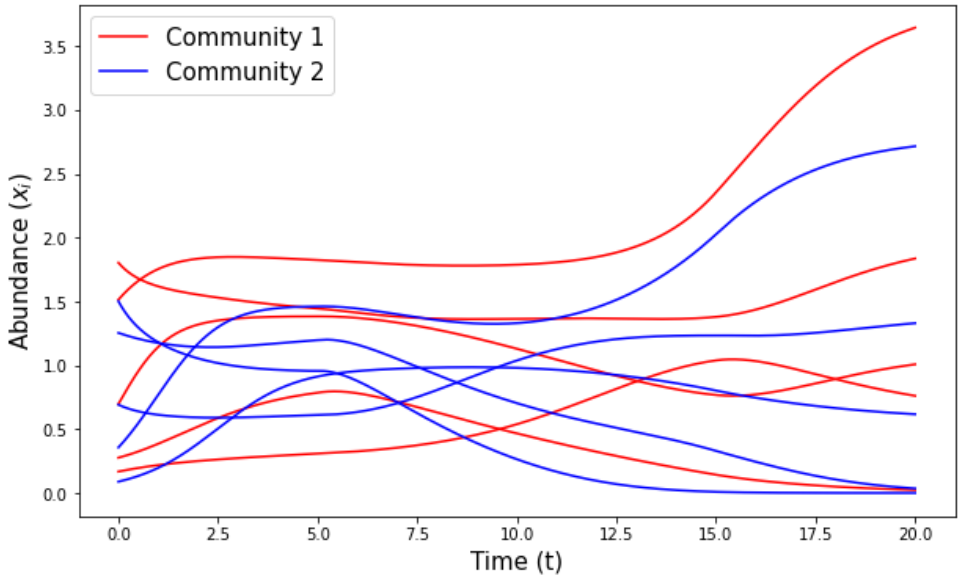


Figure 3.0.3: Dynamics of model (3.1) of 2 distinct communities of 5 species with interaction matrix (3.2). At $t = 0$, the two communities converge to their feasible equilibrium point and are not interacting. At $t = 5$, the two communities start to interact i.e. α_{12} and α_{21} increase with a linear growth until $t = 15$. Then, the two communities converges to their new equilibrium point with persisting and extinct species in both communities.

In the context of a single community, the existence of a positive solution has been studied by Bizeul and Najim [BN21] and extended for more complex food webs in [AN21, CEFN22]. On the other hand, if $\boldsymbol{x}^*$ has vanishing components, the equilibrium equations are no longer linear, and do not satisfy (3.5) anymore. We can then associate the equilibrium with a nonlinear optimization problem which has been studied by Clenet *et al.* [CMN22] in the single community case.

3.1 Feasibility

Denote by α the interaction (normalisation) parameter in the case of a unique community. According to the work of Dougoud *et al.*. [DVR+18], if α is fixed (i.e. does not depend on n) then there can be no feasible equilibrium at large n. Following this work, Bizeul and Najim [BN21] provided the right normalization of α to have a feasible equilibrium. The threshold corresponds to $\alpha \sim \sqrt{2\log(n)}$ above which the equilibrium is feasible almost surely. Some extensions of these results have been made in the sparse case [AN21] and with a mean and pairwise correlated entries [CEFN22]. In this section, conditions on the matrices $\boldsymbol{s}$ and $\boldsymbol{\mu}$ are given to get a feasible equilibrium in each community. We then provide some ecological interpretations.

3.1.1 Theoretical analysis of the threshold

Recall the notation $\boldsymbol{x} = (x_k)_{k\in[n]}$. Denote by $\|\boldsymbol{x}\|_\infty = \max_{k\in[n]} |x_k|$. We are interested in the existence of a feasible solution of the fixed point problem associated with the model (3.3).

Consider $\boldsymbol{s}$ and $\boldsymbol{\mu}$ such that $I - B$ is invertible. The problem is defined by

$$\boldsymbol{x}^* = \mathbf{1} + B\boldsymbol{x}^* \;\Leftrightarrow\; \boldsymbol{x}^* = (I - B)^{-1}\mathbf{1} = Q\mathbf{1}\,, \tag{3.6}$$

where $Q = (I - B)^{-1}$ is the resolvent of the matrix B. The problem (3.6) admits a unique solution. An extension of the computations of Bizeul and Najim is carried out in the framework of a block structure network. In the sequel, one considers the centered case $\boldsymbol{\mu} = \mathbf{0}$ (see Subsection 3.1.4 for the non-zero case). Recall the Hadamard product $\circ$ of two matrices $A \circ B = (A_{ij}B_{ij})$ and consider

$$\boldsymbol{s} \circ \boldsymbol{s} = \begin{pmatrix} \frac{1}{\alpha_{11}^2} & \frac{1}{\alpha_{12}^2} \\ \frac{1}{\alpha_{21}^2} & \frac{1}{\alpha_{22}^2} \end{pmatrix}.$$

We consider a matrix $\boldsymbol{s}$ which depends on n, i.e. $\boldsymbol{s} = \boldsymbol{s}_n$ such that:

$$\boldsymbol{s}_n \xrightarrow[n\to\infty]{} \mathbf{0} \quad \Leftrightarrow \quad \forall\, i,j \in \{1,2\}, \alpha_{ij} \xrightarrow[n\to\infty]{} \infty\,.$$

Let B_n a matrix defined by

$$B_n = V\boldsymbol{s}_n V^\top \circ \frac{1}{\sqrt{n}} \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \tag{3.7}$$

where

$$V \in \mathcal{M}_{n\times 2},\; V = \begin{pmatrix} \mathbf{1}_{\mathcal{I}_1} & 0 \\ 0 & \mathbf{1}_{\mathcal{I}_2} \end{pmatrix}.$$

The spectral radius of $\frac{1}{\sqrt{n}} \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$ a.s. converges to 1. So as long as $\boldsymbol{s}_n$ is close to zero, the matrix $I - B_n$ is eventually invertible.

Theorem 3.1 (Feasibility for the 2-blocks model)**.** *Assume that matrix B_n is defined by the 2-blocks model* (3.7)*, $\boldsymbol{\mu} = \mathbf{0}$. Let $\boldsymbol{\beta} = (\beta_1, \beta_2)$, $\beta_1 + \beta_2 = 1$ represents the proportion of each community. Let $\boldsymbol{s}_n \xrightarrow[n\to\infty]{} 0$ and denote by $s_n^* = 1/\sqrt{2\log n}$ the critical threshold. Let $x_n = (x_k)_{k\in[n]}$ be the solution of* (3.6)*.*

1. *If there exists $\varepsilon > 0$ such that eventually $\left\|(\boldsymbol{s_n} \circ \boldsymbol{s_n})\boldsymbol{\beta}^\top\right\|_\infty \geqslant (1+\varepsilon)(s_n^*)^2$ then*

$$\mathbb{P}\left\{\min_{k\in[n]} x_k > 0\right\} \xrightarrow[n\to\infty]{} 0\,.$$

2. *If there exists $\varepsilon > 0$ such that eventually $\left\|(\boldsymbol{s_n} \circ \boldsymbol{s_n})\boldsymbol{\beta}^\top\right\|_\infty \leqslant (1-\varepsilon)(s_n^*)^2$ then*

$$\mathbb{P}\left\{\min_{k\in[n]} x_k > 0\right\} \xrightarrow[n\to\infty]{} 1\,.$$

Sketch of proof. The first step consists in decomposing the equilibrium $\boldsymbol{x}^*$:

$$\begin{aligned} x_k^* = e_k^\top \boldsymbol{x}^* &= e_k^\top Q\mathbf{1}\,, \\ &= \sum_{\ell=0}^{\infty} e_k^\top B^\ell \mathbf{1}\,, \\ &= 1 + e_k^\top B\mathbf{1} + e_k^\top B^2 Q\mathbf{1}\,, \\ &= 1 + Z_k + R_k\,, \end{aligned}$$

where $Z_k = \sum_{\ell=1}^{n} B_{k\ell}$, $\forall k \in [n]$.

For the moment, we suppose $\forall k \in [n]$, R_k is a negligible term if n is sufficiently large. This part of the proof has not been treated yet and relies on Gaussian concentration of Lipschitz functionnals. However, we are confident that the techniques applied in [BN21] will succeed in handling R_k.

The feasibility of the two communities is studied independently. Using Gaussian addition properties, a simpler form of Z_k is deduced. Consider a family $(\check{Z}_k)_{k\in[n]}$ of i.i.d. random variables $\mathcal{N}(0,1)$.

$$\begin{aligned} \text{If } k \in \mathcal{I}_1,\ Z_k &= \sum_{\ell\in\mathcal{I}_1} B_{k\ell} + \sum_{\ell\in\mathcal{I}_2} B_{k\ell}\,, \\ &\sim \mathcal{N}\left(0\,,\,\frac{\beta_1}{\alpha_{11}^2}\right) + \mathcal{N}\left(0\,,\,\frac{\beta_2}{\alpha_{12}^2}\right)\,, \\ &\sim \sqrt{\frac{\beta_1}{\alpha_{11}^2} + \frac{\beta_2}{\alpha_{12}^2}}\check{Z}_k\,. \end{aligned}$$

Similarly

$$\text{if } k \in \mathcal{I}_2,\ Z_k \sim \sqrt{\frac{\beta_1}{\alpha_{21}^2} + \frac{\beta_2}{\alpha_{22}^2}}\check{Z}_k\,.$$

Given $\boldsymbol{\beta} = (\beta_1, \beta_2)$, conditions on the matrix $\boldsymbol{\alpha}$ are inferred to have:

$$\mathbb{P}(\min_{k\in[n]} x_k > 0) = 1 \Leftrightarrow \mathbb{P}(\min_{k\in[n]} Z_k > -1) = 1\,. \tag{3.8}$$

In order to compute a tractable form of $\min_{k\in[n]} Z_k$, an additional approximation is made, if n is large enough

$$\min_{k\in\mathcal{I}_i} \check{Z}_k \sim -\sqrt{2\log(\beta_i n)} \approx -\sqrt{2\log(n)} \tag{3.9}$$

$$\begin{aligned} \min_{k\in[n]} Z_k &= \min\left(\sqrt{\frac{\beta_1}{\alpha_{11}^2} + \frac{\beta_2}{\alpha_{12}^2}}\min_{k\in\mathcal{I}_1}\check{Z}_k, \sqrt{\frac{\beta_1}{\alpha_{21}^2} + \frac{\beta_2}{\alpha_{22}^2}}\min_{k\in\mathcal{I}_2}\check{Z}_k\right)\,, \\ &\simeq \min\left(\sqrt{\frac{\beta_1}{\alpha_{11}^2} + \frac{\beta_2}{\alpha_{12}^2}}\left(-\sqrt{2\log(n)}\right), \sqrt{\frac{\beta_1}{\alpha_{21}^2} + \frac{\beta_2}{\alpha_{22}^2}}\left(-\sqrt{2\log(n)}\right)\right)\,, \\ &= \min\left(-\sqrt{\frac{2\beta_1\log(n)}{\alpha_{11}^2} + \frac{2\beta_2\log(n)}{\alpha_{12}^2}}, -\sqrt{\frac{2\beta_1\log(n)}{\alpha_{21}^2} + \frac{2\beta_2\log(n)}{\alpha_{22}^2}}\right)\,, \\ &= -\max\left(\sqrt{\frac{2\beta_1\log(n)}{\alpha_{11}^2} + \frac{2\beta_2\log(n)}{\alpha_{12}^2}}, \sqrt{\frac{2\beta_1\log(n)}{\alpha_{21}^2} + \frac{2\beta_2\log(n)}{\alpha_{22}^2}}\right)\,. \end{aligned}$$

Following the approximation (3.9), the condition $\min_{k\in[n]} Z_k > -1$ asymptotically boils down

to

$$\begin{aligned}
& \max\left(\sqrt{\frac{2\beta_1\log(n)}{\alpha_{11}^2}+\frac{2\beta_2\log(n)}{\alpha_{12}^2}},\sqrt{\frac{2\beta_1\log(n)}{\alpha_{21}^2}+\frac{2\beta_2\log(n)}{\alpha_{22}^2}}\right)<1\,,\\
\Leftrightarrow\quad & \max\left(\frac{2\beta_1\log(n)}{\alpha_{11}^2}+\frac{2\beta_2\log(n)}{\alpha_{12}^2},\frac{2\beta_1\log(n)}{\alpha_{21}^2}+\frac{2\beta_2\log(n)}{\alpha_{22}^2}\right)<1\,,\\
\Leftrightarrow\quad & \max\left(\frac{\beta_1}{\alpha_{11}^2}+\frac{\beta_2}{\alpha_{12}^2},\frac{\beta_1}{\alpha_{21}^2}+\frac{\beta_2}{\alpha_{22}^2}\right)<\frac{1}{2\log(n)}\,,\\
\Leftrightarrow\quad & \left\|(\boldsymbol{s_n}\circ\boldsymbol{s_n})\boldsymbol{\beta}^\top\right\|_\infty<\frac{1}{2\log(n)}:=(s_n^*)^2\,.
\end{aligned}$$

□

The extension to the b-block case can be found in Appendix 3.C.
Since $\alpha\propto\sqrt{\log(n)}$ is the critical regime, we introduce the matrix $\boldsymbol{\kappa}$ defined by

$$\boldsymbol{\kappa}=\frac{\boldsymbol{\alpha}}{\sqrt{\log(n)}}\,,$$

where $\boldsymbol{\kappa}$ will be $O(1)$ at criticality, this will be convenient for ecological interpretations. Using (3.9), the condition on $\boldsymbol{\kappa}$ writes

$$\max\left(\frac{2\beta_1}{\kappa_{11}^2}+\frac{2\beta_2}{\kappa_{12}^2},\frac{2\beta_1}{\kappa_{21}^2}+\frac{2\beta_2}{\kappa_{22}^2}\right)<1\,. \tag{3.10}$$

Remark 3.1.
1. If $\forall i\in\{1,2\}\,,\beta_i=\frac{1}{2}$ and the entry of the matrix $\boldsymbol{\kappa}$ are equal, then condition (3.10) gives the threshold $\kappa_{ij}>\sqrt{2}$, we find again the threshold $\sqrt{2\log(n)}$ in [BN21].

2. If $\kappa_{12}=\kappa_{21}=\infty$, then condition (3.10) gives the feasibility conditions for each community:
$$\alpha_{11}>\sqrt{2\beta_1\log(n)}\quad\text{and}\quad\alpha_{22}>\sqrt{2\beta_2\log(n)}\,.$$
For the same α, it means $\alpha>\sqrt{2\log(n)}\max(\beta_1,\beta_2)$.

Remark 3.2. Assume $\kappa_{11}=\kappa_{22}=\nu_1$ and $\kappa_{12}=\kappa_{21}=\nu_2$, condition (3.10) is reformulated as:

$$\max\left(\frac{2\beta_1}{\nu_1^2}+\frac{2\beta_2}{\nu_2^2},\frac{2\beta_1}{\nu_2^2}+\frac{2\beta_2}{\nu_1^2}\right)<1\,.$$

If β_1,β_2 and ν_2 are fixed, then the phase transition on the intra-community interactions occurs at

$$\nu_1>\min\left(\sqrt{\frac{\beta_1}{\frac{1}{2}-\frac{\beta_2}{\nu_2^2}}},\sqrt{\frac{\beta_2}{\frac{1}{2}-\frac{\beta_1}{\nu_2^2}}}\right)\,.$$

In Figure 3.1.1, the phase transition is represented for a chosen set of parameters. Note that the transition is rather smooth. However, an increase in the size of the model would allow a smoother curve. The threshold depends on ν_2, its increase (decrease of the inter-block interactions) lowers the feasibility threshold to reach at least 1 (for communities of the same size).

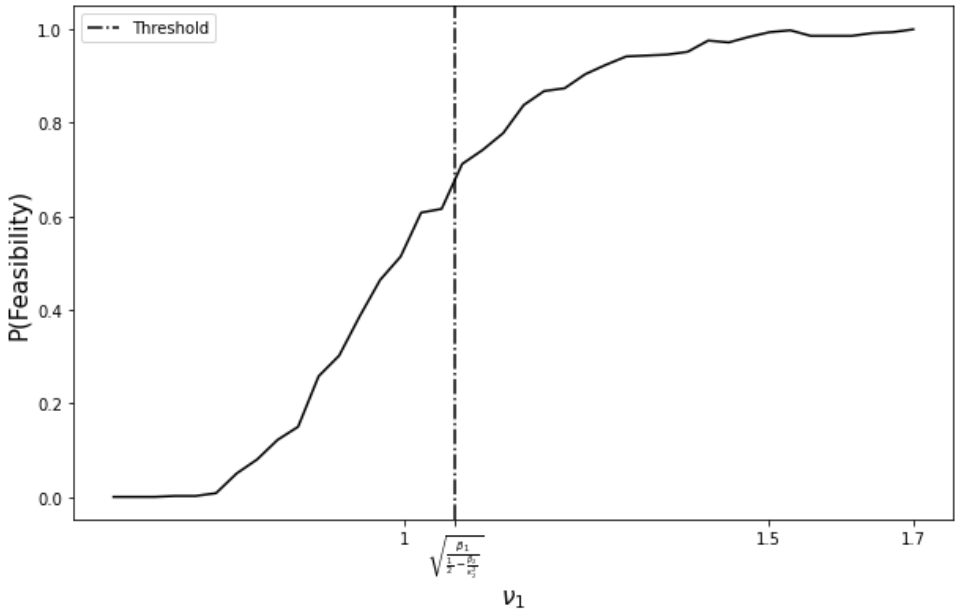


Figure 3.1.1: Transition towards feasibility for the 2-blocks model (3.2). For each value ν_1 on the x-axis, we simulate 500 matrices B of size $n = 500$ with two communities of the same size ($\beta_1 = \beta_2 = 0.5$) with the inter-block interactions fixed at $\alpha_{21}(\nu_2) = \alpha_{21}(\nu_2) = 2\sqrt{\log(n)}$ and compute the solution $\boldsymbol{x}$ of Theorem (3.1) at the scaling for the intra-block interactions $\alpha_{11}(\nu_1) = \alpha_{22}(\nu_1) = \nu_1\sqrt{\log(n)}$. The curve represents the proportion of feasible solutions $\boldsymbol{x}$ obtained for the 500 simulations. The dotdashed vertical line corresponds to $\nu_1 = \sqrt{\frac{\beta_1}{\frac{1}{2}-\frac{\beta_2}{\nu_2^2}}} = 2/\sqrt{3}$.

3.1.2 Preservation of feasibility between two groups

Equation (3.10) defines a "feasibility domain" and gives a constraint in five dimensions. The two communities of species can be studied independently i.e. the two components of equation (3.10) respectively give the feasibility condition for each community.

- If $\frac{2\beta_1}{\kappa_{11}^2} + \frac{2\beta_2}{\kappa_{12}^2} < 1$, then Community 1 is feasible.
- If $\frac{2\beta_1}{\kappa_{21}^2} + \frac{2\beta_2}{\kappa_{22}^2} < 1$, then Community 2 is feasible.

The first community will be affected by the modification of the coefficient: κ_{11}, κ_{12}, the second by: κ_{21}, κ_{22}. From a general point of view, an increase in the inter- or intra-interaction strength will decrease the probability to have a feasible equilibrium.

An example of application: suppose we start with two feasible communities of the same size ($\beta_1 = \beta_2 = 0.5$) and add interactions between these two groups, the feasibility of the general model decreases. The feasibility domain is represented in Figure 3.1.2. It shows a feasibility threshold where above the curve the feasible property is satisfied. This means that the lower the values of κ_{11} and κ_{22}, i.e. the stronger the interactions within the groups, the more likely the feasibility property is lost. We can deduce that an independent group structure is more likely to be feasible and therefore stable which supports previous work on compartments models [SB11].

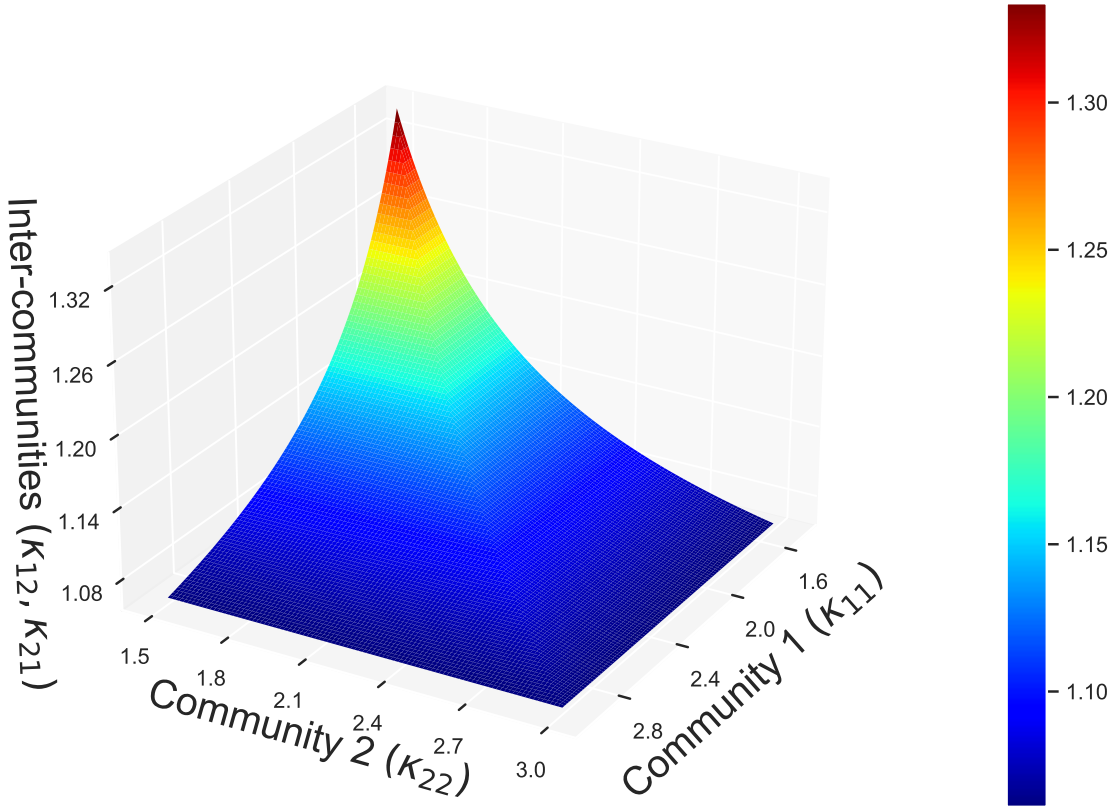


Figure 3.1.2: Representation of the feasibility phase diagram. The feasible domain is above the surface. The z-axis (resp x-axis) is the strength of interaction within Community 1 - κ_{11} (resp Community 2 - κ_{22}). The y-axis is the inter-community interactions $\kappa_{12} = \kappa_{21}$. The colored surface illustrate the threshold between the feasibility and non-feasibility domain in the system (3.3).

3.1.3 Impact of the community size

From another point of view, for a fixed matrix $\boldsymbol{\kappa}$, the condition to have a feasible fixed point domain can be computed as a function of the size of each community i.e. the pair $\boldsymbol{\beta} = (\beta_1, \beta_2)$. Starting from (3.10):

$$\max\left(\sqrt{\frac{2\beta_1}{\kappa_{11}^2}+\frac{2\beta_2}{\kappa_{12}^2}}, \sqrt{\frac{2\beta_1}{\kappa_{21}^2}+\frac{2\beta_2}{\kappa_{22}^2}}\right) < 1\,,$$

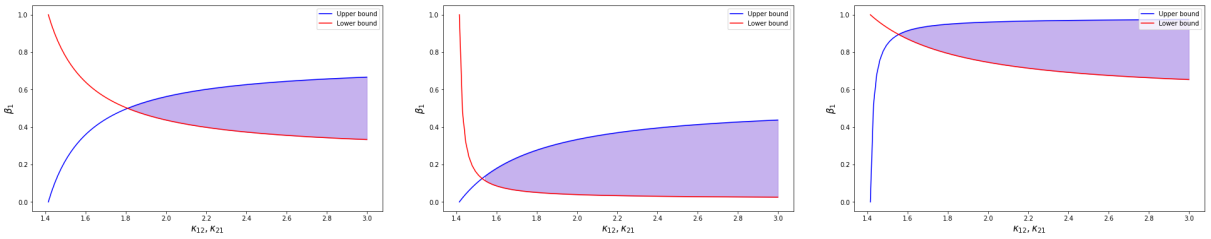
the two components are studied independently,

$$\begin{aligned}\sqrt{\frac{2\beta_1}{\kappa_{11}^2}+\frac{2(1-\beta_1)}{\kappa_{12}^2}} < 1 \Leftrightarrow\ & \frac{2\beta_1}{\kappa_{11}^2}+\frac{2(1-\beta_1)}{\kappa_{12}^2} < 1\,,\\ \Leftrightarrow\ & \beta_1\left(\frac{2}{\kappa_{11}^2}-\frac{2}{\kappa_{12}^2}\right) < 1-\frac{2}{\kappa_{12}^2}\,,\\ \Leftrightarrow\ & \beta_1 < \frac{1-\frac{2}{\kappa_{12}^2}}{\left(\frac{2}{\kappa_{11}^2}-\frac{2}{\kappa_{12}^2}\right)},\ \kappa_{11} < \kappa_{12}\,.\end{aligned}$$

In an identical manner, one has

$$\sqrt{\frac{2\beta_1}{\kappa_{21}^2}+\frac{2(1-\beta_1)}{\kappa_{22}^2}}<1 \;\Leftrightarrow\; \beta_1 > \frac{1-\frac{2}{\kappa_{22}^2}}{\left(\frac{2}{\kappa_{21}^2}-\frac{2}{\kappa_{22}^2}\right)},\; \kappa_{22}<\kappa_{21}\,.$$

In the case where the intra-community interactions $(\kappa_{11}\,,\,\kappa_{22})$ are smaller than the inter-community interactions $(\kappa_{12}\,,\,\kappa_{21})$, we obtain an upper and a lower bound for the admissible size of each community β_1, β_2 to have a feasible equilibrium. In Figure 3.1.3, different cases of the feasibility zone are represented according to the inter-community interactions $(\kappa_{12}\,,\,\kappa_{21})$. If the intra-community interactions are different, the community with the lowest interaction κ_{ii} is advantaged i.e. the size of the community can be larger.



(a) $\kappa_{11} = 1.2$, $\kappa_{22} = 1.2$ (b) $\kappa_{11} = 1.2$, $\kappa_{22} = 1.4$ (c) $\kappa_{11} = 1.4$, $\kappa_{22} = 0.9$

Figure 3.1.3: Representation of the feasibility domain depending on the fixed intra-community interaction. In (a), (b), (c), a different scenario of intra-community interaction is presented. Each panel represents the upper-bound (blue curve) and the lower-bound (red curve) of the size of Community 1 as a function of the interaction between two communities $(\kappa_{12}, \kappa_{21})$. The blue area is the admissible zone to have a feasible fixed point in (3.3). The size of community 2 is equal to $\beta_2 = 1 - \beta_1$.

3.1.4 Extension to the non-centered case

Theorem 3.1 is extended to the non centered case $\boldsymbol{\mu} \geqslant \mathbf{0}$. The matrix of interaction B can be decomposed into the sum of a main matrix $\check{A}$ and a mean matrix P,

$$B = \underbrace{\frac{1}{\sqrt{n}}\begin{pmatrix} \frac{A_{11}}{\alpha_{11}} & \frac{A_{12}}{\alpha_{12}} \\ \frac{A_{21}}{\alpha_{21}} & \frac{A_{22}}{\alpha_{22}} \end{pmatrix}}_{\check{A}} + \underbrace{\frac{1}{n}\begin{pmatrix} \mu_{11}\mathbf{1}_{\mathcal{I}_1}\mathbf{1}_{\mathcal{I}_1}^\top & \mu_{12}\mathbf{1}_{\mathcal{I}_1}\mathbf{1}_{\mathcal{I}_2}^\top \\ \mu_{21}\mathbf{1}_{\mathcal{I}_2}\mathbf{1}_{\mathcal{I}_1}^\top & \mu_{22}\mathbf{1}_{\mathcal{I}_2}\mathbf{1}_{\mathcal{I}_2}^\top \end{pmatrix}}_{P}.$$

Recalling the fixed point equation (3.6),

$$\boldsymbol{x}^* = \mathbf{1} + \check{A}\boldsymbol{x}^* \;\Leftrightarrow\; \boldsymbol{x}^* = (I-\check{A})^{-1}\mathbf{1} = Q\mathbf{1}\,.$$

Denote by

$$\tilde{\boldsymbol{x}} = \mathbf{1} + B\tilde{\boldsymbol{x}} = \mathbf{1} + (\check{A}+P)\tilde{\boldsymbol{x}}\,,$$

the equilibrium of the non-centered system. Consider $\boldsymbol{\alpha}$ and $\boldsymbol{\mu}$ such that $(I-\check{A})$ and $(I-\check{A}-P)$ are invertibles. Recall that $Q = (I-\check{A})^{-1}$, by using a special case of the Woodbury formula

$$(I-\check{A}-P)^{-1} = Q + QP(I-\check{A}-P)^{-1}\,,$$

multiplying both side by $\mathbf{1}$,

$$\tilde{\boldsymbol{x}} = \boldsymbol{x}^* + QP\tilde{\boldsymbol{x}} = (I - QP)^{-1}\boldsymbol{x}^* .$$

From a heuristic point of view, when the dimension of the system becomes very large $Q \approx I$ because each entry of $\boldsymbol{\alpha}$ converge to infinity.

The second step consists in estimating the matrix $(I - P)^{-1}$. If matrix $(I - P)^{-1}$ has only non negative components and $(I - P)^{-1}$ has no zero rows (:= a row with only 0 entries), then $\tilde{x}$ is positive. The second condition is impossible because if $(I-P)^{-1}$ exists, its determinant is non-zero.

Here, we only consider the case where the mean has non-negative elements $\boldsymbol{\mu} \geqslant 0$. In this case, the properties of class of M-matrices can help. For a survey on characterizations of non-singular M-matrices, see [Ple77].

Definition 3.1 (M-matrix)**.** An $n \times n$ matrix that can be expressed in the form $sI - C$, where $C = (c_{ij})$ with $c_{ij} \geqslant 0, 1 \leqslant i, j \leqslant n$, and $s > \rho(C)$, the spectral radius of C, is called an M-matrix.

Proposition 3.2. *$I - P$ is an M-matrix if and only if $(I - P)^{-1}$ exists and $(I - P)^{-1} \geqslant 0$ component-wise.*

Remark 3.3. The class of M-matrices is included in the class of P-matrices (see section 3.2) and of inverse-positive matrices i.e. matrix inverses with non-negative elements.

By identification, we are interested in the specific setting where $C = P$. If all the coefficients of the matrix P are non-negative i.e. $\boldsymbol{\mu} \geqslant 0$ and $\rho(P) < 1$, then $(I - P)$ is a M-matrix. It is then sufficient to a give a condition on the spectral radius of the matrix P.

The non-zero eigenvalues of the matrix P are the eigenvalues of the matrix $\mathrm{diag}(\boldsymbol{\beta})\boldsymbol{\mu}$. Suppose $\theta \neq 0$ is an eigenvalue of $P = U\boldsymbol{\mu}U^\top$ with $U \in \mathcal{M}_{n\times 2}(\mathbb{R})$ defined by

$$U = \frac{1}{\sqrt{n}} \begin{pmatrix} \mathbf{1}_{\mathcal{I}_1} & 0 \\ 0 & \mathbf{1}_{\mathcal{I}_2} \end{pmatrix} ,$$

then there exists u such that

$$\begin{aligned} Pu = \theta u &\Leftrightarrow U\boldsymbol{\mu}U^\top u = \theta u , \\ &\Leftrightarrow U^\top U\boldsymbol{\mu}U^\top u = \theta U^\top u , \\ &\Leftrightarrow \mathrm{diag}(\boldsymbol{\beta})\boldsymbol{\mu}U^\top u = \theta U^\top u . \end{aligned} \tag{3.11}$$

θ is an eigenvalue of the matrix $\mathrm{diag}(\boldsymbol{\beta})\boldsymbol{\mu}$ and $U^\top u$ its associated eigenvector.
If $\boldsymbol{\mu} \geqslant 0$, $\tilde{\boldsymbol{x}}$ is feasible if:

$$\rho\left(\mathrm{diag}(\boldsymbol{\beta})\boldsymbol{\mu}\right) < 1 \Leftrightarrow \rho\left(\begin{pmatrix} \beta_1\mu_{11} & \beta_1\mu_{12} \\ \beta_2\mu_{21} & \beta_2\mu_{22} \end{pmatrix}\right) < 1 .$$

For a mean matrix of rank 1, $P = \frac{\mu}{n}\mathbf{1}_n\mathbf{1}_n^\top$, the condition is $\mu < 1$ which is similar to [CEFN22].

Remark 3.4. If the mean is the same for every block $\boldsymbol{\mu} = \mu\mathbf{1}\mathbf{1}^\top$, $\mu \neq 0$, one can use the identical arguments as in Clenet *et al.* [CEFN22].

3.2 Existence of a unique equilibrium

In Section 3.1, the case of a feasible equilibrium i.e. $\boldsymbol{x}^* > 0$ has been analysed. However, in Figure 3.0.3, for $t > 15$, we notice that the system can also converge to an equilibrium with vanishing species. Here, the conditions of existence and uniqueness of an equilibrium with vanishing species are described. Moreover, the next step will be the investigation of the properties of the persisting species in each communities.

3.2.1 Theoretical requirement

The research of fixed points of (3.3) is equivalent to the identification of solutions of the system (3.4). However, the number of potential solutions can be extremely large. The works of Takeuchi [Tak96, Theorem 3.2.5] have made it possible to reduce this number by relying on standard properties of dynamical systems. In particular, a necessary condition for the equilibrium $\boldsymbol{x}^*$ to be stable is that

$$1 - x_k^* + (B\boldsymbol{x}^*)_k \leqslant 0\,. \tag{3.12}$$

Condition (3.12) is better known in ecology as the non-invadability condition [LM96]. In reference to the ODE (3.3), the requirement for a given species $k \in [n]$ to be non-invasive is equivalent to:

$$\left(\frac{1}{x_k}\frac{dx_k}{dt}\right)_{x_k \to 0^+} \leqslant 0\,. \tag{3.13}$$

The condition 3.13 describes the fact that if we add a species to the system at very low abundance, then it will not be able to invade the system. As a consequence, we will now focus on the following set of conditions:

$$\left\{\begin{array}{rcll} x_k^*\,(1 - x_k^* + (B\boldsymbol{x}^*)_k) & = & 0 & \text{for } k \in [n]\,, \\ 1 - x_k^* + (B\boldsymbol{x}^*)_k & \leqslant & 0 & \text{for } k \in [n]\,, \\ \boldsymbol{x}^* & \geqslant & 0 & \text{componentwise}\,. \end{array}\right. \tag{3.14}$$

This casts the problem of finding a non negative equilibrium into the class of Linear Complementarity Problems (LCP), which we describe hereafter.

Linear Complementarity Problem (LCP). The linear complementarity problem is a class of problems in mathematical optimization theory. It was proposed by Cottle and Danzig in 1968 [CD68] and appears frequently in computational mechanics problems. In particular LCP encompasses linear and quadratic programs; standard references are [Mur88, CPS09]. Given a $n \times n$ matrix M and a $n \times 1$ vector $\boldsymbol{q}$, the associated LCP denoted by $LCP(M, \boldsymbol{q})$ consists in finding two $n \times 1$ vectors $\boldsymbol{z}, \boldsymbol{w}$ satisfying the following set of constraints:

$$\left\{\begin{array}{rcll} \boldsymbol{z} & \geqslant & 0\,, & \\ \boldsymbol{w} = M\boldsymbol{z} + \boldsymbol{q} & \geqslant & 0\,, & \\ \boldsymbol{w}^\top \boldsymbol{z} & = & 0 \quad \Leftrightarrow \quad w_k z_k = 0 & \text{for all } k \in [n]\,. \end{array}\right. \tag{3.15}$$

Since $\boldsymbol{w}$ can be inferred from $\boldsymbol{z}$, we denote $\boldsymbol{z} \in LCP(M, \boldsymbol{q})$ if $(\boldsymbol{w}, \boldsymbol{z})$ is a solution of (3.15).

A theorem by Murty [Mur72] states that the $LCP(M, \boldsymbol{q})$ has a unique solution $(\boldsymbol{w}, \boldsymbol{z})$ iff M is a P-matrix, that is:

$$\det(M_{\mathcal{I}}) > 0\,, \quad \forall\, \mathcal{I} \subset [n]\,, \quad M_{\mathcal{I}} = (M_{k\ell})_{k,\ell \in \mathcal{I}}\,.$$

In view of (3.14), we look for $\boldsymbol{x}^* \in LCP(I - B, -\mathbf{1})$.

The equilibrium x^* and its stability.

Definition 3.2 (Lyapunov diagonal stability)**.** Matrix M is called Lyapunov diagonally stable, denoted by $M \in S_\omega$, if and only if there exists a diagonal matrix D with positive elements such that $DM + M^\top D$ is negative definite i.e. all eigenvalues are negative.

Proposition 3.3 (Takeuchi *et al.* , 1978)**.** *If $M \in S_\omega$ then $-M$ is a P-matrix.*

Recall the system (3.1) with different growth rates for each species and consider matrix B is arbitrary,

$$\frac{d\, y_k(t)}{dt} = y_k(r_k + ((-\theta I + B)\boldsymbol{y})_k)\,, \quad k \in [n]\,. \tag{3.16}$$

Takeuchi and Adachi (see for instance [Tak96, Th. 3.2.1]) provide a criterion for the existence of a unique equilibrium $\boldsymbol{y}^*$ and the global stability of the LV system.

Theorem 3.4 (Takeuchi and Adachi, [TA80])**.** *If $-\theta I + B \in S_\omega$, then LCP($\theta I - B$,$\boldsymbol{r}$) admits a unique solution. In particular, for every $\boldsymbol{r} \in \mathbb{R}^n$, there is a unique equilibrium $\boldsymbol{y}^*$ to* (3.16)*, which is globally stable*

$$\forall \boldsymbol{y}_0 > 0,\ \begin{cases} \boldsymbol{y}(0) = \boldsymbol{y}_0 \\ \boldsymbol{y}(t) \text{ satisfies} \end{cases},\ \boldsymbol{y}(t) \underset{t\to\infty}{\longrightarrow} \boldsymbol{y}^*\,.$$

Combining this result with RMT, we can guarantee the existence of a globally stable equilibrium $\boldsymbol{x}^*$ of (3.1) for a wide range of the parameters $(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\mu})$.

3.2.2 Centered case: $\boldsymbol{\mu} = 0$

Before going into details, we make some reminders about Stieltjes transforms, a key element of proofs in random matrix theory. We denote by

$$\mathbb{C}^+ := \{z \in \mathbb{C} \,:\, \mathrm{Im}(z) > 0\}$$

the upper half of the complex plane.

Definition 3.3 (Stieltjes transform)**.** Let $\nu \in \mathcal{P}(\mathbb{R})$ a probability measure. The Stieltjes transform of ν denoted by $g_\nu : \mathbb{C}^+ \to \mathbb{C}$ is defined by

$$g_\nu(z) = \int_{\mathbb{R}} \frac{1}{\lambda - z} \nu(d\lambda)\,, z \in \mathbb{C}^+\,.$$

Remark 3.5. Let ν_A the empirical measure of the eigenvalues $\lambda_1, \cdots, \lambda_n$ of the symmetric matrix A $\in \mathcal{M}_n(\mathbb{C})$ define by

$$\nu_A := \frac{1}{n} \sum_{k=1}^{n} \delta_{\lambda_k(A)}.$$

then the associated Stieltjes transform is given by

$$g_{\nu_A}(z) = \int \frac{1}{\lambda - z} \nu_A(d\lambda) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\lambda_i - z} = \frac{1}{n} Tr((A - zI)^{-1}).$$

where $Q = (A - zI)^{-1}$ is the resolvent of the matrix A.

Proposition 3.5 (Stieltjes inversion)**.** *Let g_ν the Stieltjes transform of the measure ν of finite mass $\nu(\mathbb{R})$. If $a, b \in \mathbb{R}$ and $\nu(\{a\}) = \nu(\{b\}) = 0$, then*

$$\nu(a,b) = \frac{1}{\pi} \lim_{y \to 0^+} \operatorname{Im} \int_a^b g_\nu(x+iy)dx\,,$$

and

$$\forall x \in \mathbb{R},\ \nu(\{x\}) = \frac{1}{\pi} \lim_{y \to 0^+} \operatorname{Im}(g_\nu(x+iy))\,.$$

Theorem 3.6. *Let $\boldsymbol{\mu} = \mathbf{0}$ and assume that*

$$\left\| \operatorname{diag}(\boldsymbol{\beta})^{1/2} \left((\boldsymbol{s} \circ \boldsymbol{s}) + (\boldsymbol{s} \circ \boldsymbol{s})^\top \right) \operatorname{diag}(\boldsymbol{\beta})^{1/2} \right\|_2 < 1\,,$$

then a.s. matrix $(I-B)+(I-B)^T$ is eventually positive definite: with probability one, for a given realization ω, there exists $N(\omega)$ such that for $n \geqslant N(\omega)$, $(I-B^\omega)+(I-B^\omega)^\top$ is positive definite. In particular, there exists a unique (random) globally stable equilibrium $\boldsymbol{x}^* \in LCP(I-B^\omega, -\mathbf{1})$ to (3.14).

Sketch of proof. From Theorem 3.4, we need to verify the Lyapunov diagonally stable condition of the matrix $(-I+B) \in S_\omega$ by analyzing its largest eigenvalue.

$$(-I+B)+(-I+B^\top) = -2I + \frac{1}{\sqrt{n}} \begin{pmatrix} \frac{A_{11}+A_{11}^\top}{\alpha_{11}} & \frac{A_{12}}{\alpha_{12}} + \frac{A_{21}^\top}{\alpha_{21}} \\ \frac{A_{21}}{\alpha_{21}} + \frac{A_{12}^\top}{\alpha_{12}} & \frac{A_{22}+A_{22}^\top}{\alpha_{22}} \end{pmatrix}.$$

Denote by H the symmetric matrix

$$H = \frac{1}{\sqrt{n}} \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix} = \frac{1}{\sqrt{n}} \begin{pmatrix} \frac{A_{11}+A_{11}^\top}{\alpha_{11}} & \frac{A_{12}}{\alpha_{12}} + \frac{A_{21}^\top}{\alpha_{21}} \\ \frac{A_{21}}{\alpha_{21}} + \frac{A_{12}^\top}{\alpha_{12}} & \frac{A_{22}+A_{22}^\top}{\alpha_{22}} \end{pmatrix},$$

where $\forall\, i,j \in \{1,2\}$, H_{ij} is a matrix of size $\beta_i n \times \beta_j n$ and each off-diagonal entries follow a Gaussian distribution $\mathcal{N}\left(0, 1/\alpha_{ij}^2 + 1/\alpha_{ji}^2\right)$.

A matrix is negative definite if and only if all its eigenvalues are negatives. Here, notice that $-2I+H$ is negative definite iff the top eigenvalues of H is lower than 2. The aim of the proof is to give condition on the parameter $\boldsymbol{\alpha}$ such that

$$\lambda_{\max}(H) < 2\,.$$

The matrix H has a variance profile, such a model has been studied in great details by Erdös *et al.* and is linked to the theory of the Quadratic Vector Equation (QVE, see [AEK17, AEK19] for more technical information). Given $\boldsymbol{m}(z) = (m_1(z), \cdots, m_n(z))$, the QVE associated to the matrix H is decomposed as

$$\begin{aligned} k \in \mathcal{I}_1, -\frac{1}{m_k(z)} &= z + \sum_{\ell \in \mathcal{I}_1} \frac{2}{\alpha_{11}^2 n} m_\ell(z) + \sum_{\mathcal{I}_2} \frac{1}{n} \left(\frac{1}{\alpha_{12}^2} + \frac{1}{\alpha_{21}^2} \right) m_\ell(z)\,, \\ k \in \mathcal{I}_2, -\frac{1}{m_k(z)} &= z + \sum_{\ell \in \mathcal{I}_1} \frac{1}{n} \left(\frac{1}{\alpha_{12}^2} + \frac{1}{\alpha_{21}^2} \right) m_\ell(z) + \sum_{\mathcal{I}_2} \frac{2}{\alpha_{22}^2 n} m_\ell(z)\,. \end{aligned}$$

Denote by $1/\boldsymbol{m}(z) = (1/m_1(z), \cdots, 1/m_n(z))^\top$ and

$$S = \frac{1}{n} \begin{pmatrix} \frac{2}{\alpha_{11}^2} \mathbf{1}_{\mathcal{I}_1} \mathbf{1}_{\mathcal{I}_1}^\top & (\frac{1}{\alpha_{12}^2} + \frac{1}{\alpha_{21}^2}) \mathbf{1}_{\mathcal{I}_1} \mathbf{1}_{\mathcal{I}_2}^\top \\ (\frac{1}{\alpha_{12}^2} + \frac{1}{\alpha_{21}^2}) \mathbf{1}_{\mathcal{I}_2} \mathbf{1}_{\mathcal{I}_1}^\top & \frac{2}{\alpha_{22}^2} \mathbf{1}_{\mathcal{I}_2} \mathbf{1}_{\mathcal{I}_2}^\top\,, \end{pmatrix}$$

the QVE can be written in the standard form

$$-\frac{1}{\boldsymbol{m}(z)} = z + S\boldsymbol{m}(z)\,. \tag{3.17}$$

Following Theorem 2.1 in Ajanki *et al.* [AEK19], $\forall z \in \mathbb{C}^+$, Equation (3.17) has a unique solution $\boldsymbol{m} = \boldsymbol{m}(z)$ where $n^{-1}\sum m_i$ is a Stieltjes transform of a probability measure and the support of the associated measure is included in $[-\Sigma, \Sigma]$, where $\Sigma = 2\,\|S\|_2^{1/2}$.

This information gives an asymptotic bound on the support of the matrix H associated with (3.17) i.e. asymptotically $\forall \varepsilon > 0$ for a given realization ω, there exists $N(\omega)$; $\forall\, n \geqslant N(\omega)$

$$\lambda_{\max}(H) \leqslant 2\,\|S\|_2^{1/2} + \varepsilon\,.$$

Recall that $-2I + H$ is negative definite iff $\lambda_{\max}(H) < 2$. This condition is fulfilled if

$$2\,\|S\|_2^{1/2} < 2\,,$$

or equivalently

$$\|S\|_2 < 1\,.$$

Notice that this condition is sufficient but not necessary. Given the particular shape of the matrix S, computing its norm is equivalent to computing the norm of a matrix of size 2×2

$$\begin{aligned}\|S\|_2 &= \left\|\mathrm{diag}(\boldsymbol{\beta})^{1/2}\left((\boldsymbol{s}\circ\boldsymbol{s}) + (\boldsymbol{s}\circ\boldsymbol{s})^\top\right)\mathrm{diag}(\boldsymbol{\beta})^{1/2}\right\|_2 \\ &= \left\|\begin{pmatrix}\beta_1 & 0\\ 0 & \beta_2\end{pmatrix}^{1/2}\begin{pmatrix}\frac{2}{\alpha_{11}^2} & \frac{1}{\alpha_{12}^2}+\frac{1}{\alpha_{21}^2}\\ \frac{1}{\alpha_{12}^2}+\frac{1}{\alpha_{21}^2} & \frac{2}{\alpha_{22}^2}\end{pmatrix}\begin{pmatrix}\beta_1 & 0\\ 0 & \beta_2\end{pmatrix}^{1/2}\right\|_2.\end{aligned}$$

which completes the proof, we can then rely on Theorem 3.4 to conclude.

Remark 3.6.

1. In the context of a unique community, suppose that $\boldsymbol{\alpha} = \alpha\mathbf{1}\mathbf{1}^\top$, then the previous condition takes the simpler form $\alpha > \sqrt{2}$ which was already stated in [CMN22]. Indeed, starting from the condition of Theorem 3.6, the condition on the matrix is

$$\left\|\begin{pmatrix}1/2 & 0\\ 0 & 1/2\end{pmatrix}^{1/2}\begin{pmatrix}2/\alpha^2 & 2/\alpha^2\\ 2/\alpha^2 & 2/\alpha^2\end{pmatrix}\begin{pmatrix}1/2 & 0\\ 0 & 1/2\end{pmatrix}^{1/2}\right\|_2 = \left\|\begin{pmatrix}1/\alpha^2 & 1/\alpha^2\\ 1/\alpha^2 & 1/\alpha^2\end{pmatrix}\right\|_2,$$

   which has eigenvalues $2/\alpha^2$ and 0. The same sufficient condition is obtained $2/\alpha^2 < 1 \;\Leftrightarrow\; \alpha > \sqrt{2}$.

2. The proof in the b-blocks case is provided in Appendix 3.C. The condition given there is similar to Theorem 3.6.

3. The condition given in Theorem 3.6 is sufficient to guarantee a unique solution to $LCP(I - B, \mathbf{1})$ but not necessary, although it provides more information and guarantees the global stability. This condition might be relaxed finding the bound associated to the P-matrix property of $I - B$ (see Appendix 3.D).

Spectrum: a computer based approach. Theorem 3.6 only provides sufficient conditions for the existence of a unique stable equilibrium and is based on the rough asymptotic upper bound estimation $\Sigma = 2\,\|S\|_2^{1/2}$. We can assess the sharpness of this bound by comparing it to the limiting spectrum of matrix H, which can be plotted via numerical simulations. An efficient way to compute numerically the spectrum of the matrix H comes from the system of non linear equations (3.17).

Starting from the QVE (3.17) associated to the matrix H, the system takes the simpler form

$$\begin{cases} -\frac{1}{\mu} &= z + \frac{2\beta_1}{\alpha_{11}^2}\mu + \beta_2(\frac{1}{\alpha_{12}^2} + \frac{1}{\alpha_{21}^2})\nu\,, \\ -\frac{1}{\nu} &= z + \beta_1(\frac{1}{\alpha_{12}^2} + \frac{1}{\alpha_{21}^2})\mu + \frac{\beta_2}{\alpha_{22}^2}\nu\,, \end{cases}$$

where $\forall k \in \mathcal{I}_1$, $m_k(z) = \mu(z)$ and $\forall k \in \mathcal{I}_2$, $m_k(z) = \nu(z)$. All the knowledge of (3.17) rely on $\mu(z)$ and $\nu(z)$. Then, using RMT theory, the resolvent G of the symmetric matrix H can be approximated by

$$G(z) = (H - zI)^{-1} \simeq \mathrm{diag}(\mu(z)\mathbf{1}_{\mathcal{I}_1}^\top, \nu(z)\mathbf{1}_{\mathcal{I}_2}^\top)\,.$$

From Remark 3.5, the trace of the resolvent is equal to the Stieltjes transform

$$g(z) = \frac{1}{n}Tr(G) \simeq \beta_1\mu(z) + \beta_2\nu(z)$$

of the spectral measure. Finally, the spectral density can be obtained using a the Stieltjes inversion (Prop. 3.5). The spectral density of the matrix H can be computed numerically by an iterative scheme. The initial condition of the two measurements (μ, ν) is $\mu_0 = \nu_0 = -\frac{1}{z}$. Then, the iterative scheme

$$\begin{cases} -\frac{1}{\mu_p} &= z + \frac{2\beta_1}{\alpha_{11}^2}\mu_{p-1} + \beta_2(\frac{1}{\alpha_{12}^2} + \frac{1}{\alpha_{21}^2})\nu_{p-1} \\ -\frac{1}{\nu_p} &= z + \beta_1(\frac{1}{\alpha_{12}^2} + \frac{1}{\alpha_{21}^2})\mu_{p-1} + \frac{2\beta_2}{\alpha_{22}^2}\nu_{p-1} \end{cases},$$

converge to $\mu_\infty = \lim_{p\to+\infty}\mu_p$ and $\nu_\infty = \lim_{p\to+\infty}\nu_p$. The last step consist of using the property of the Stieltjes inversion (Prop.3.5).

Remark 3.7. To handle the Stieltjes inversion (Prop.3.5) numerically, it is similar as starting with $z = x + \epsilon i, \epsilon \approx 10^{-3}$.

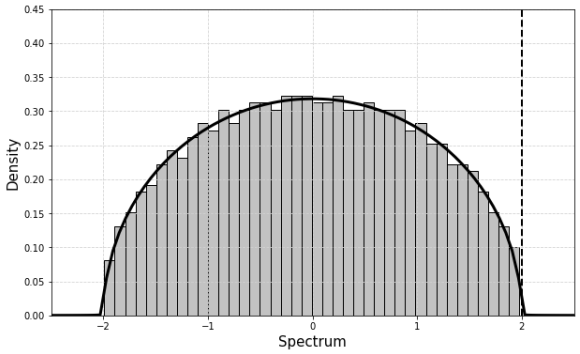
In Figure 3.2.1, the estimation of the spectral density by the numerical method is represented for different types of interactions in the blocks of the matrix H. The threshold given by $2\,\|S\|_2^{1/2}$ appears to be sharp.

3.2.3 Non centered case ($\boldsymbol{\mu} \neq 0$)

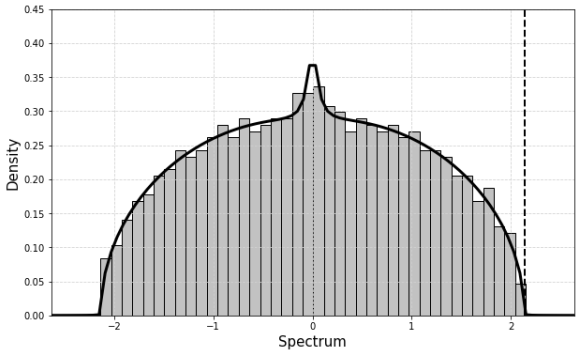
The extension to the non-centered case ($\boldsymbol{\mu} \neq \mathbf{0}$) may be mathematically tractable using the QVE theory. However, at this point we shall consider the case where all the coefficients α_{ij} are equal to a constant α.

Proposition 3.7. *Let* $\theta_{\max} = \lambda_{\max}(\mathrm{diag}(\boldsymbol{\beta})(\boldsymbol{\mu} + \boldsymbol{\mu}^\top))$. *Let* $\boldsymbol{\alpha} = \alpha\mathbf{1}\mathbf{1}^\top$, $\alpha > \sqrt{2}$ *and assume that if* $\theta_{\max} > \sqrt{2}/\alpha$,

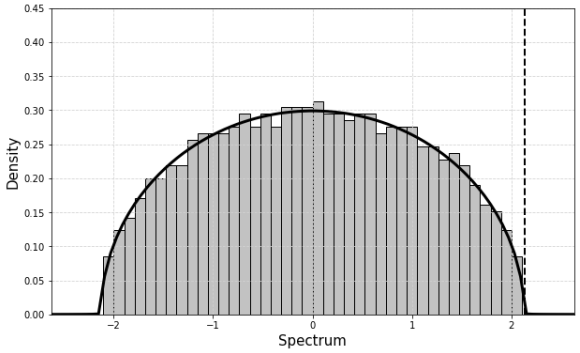
$$\theta_{\max} \in \left(\frac{\sqrt{2}}{\alpha}\ ,\ 1 + \sqrt{1 - \frac{2}{\alpha^2}}\right),$$



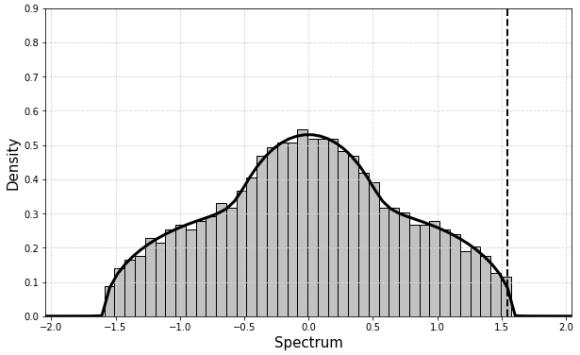
(a) $\boldsymbol{\beta} = [1/2, 1/2]$, $\boldsymbol{\alpha} = \begin{pmatrix} \sqrt{2} & \sqrt{2} \\ \sqrt{2} & \sqrt{2} \end{pmatrix}$



(b) $\boldsymbol{\beta} = [1/2, 1/2]$, $\boldsymbol{\alpha} = \begin{pmatrix} 2 & 1 \\ 1 & 8 \end{pmatrix}$



(c) $\boldsymbol{\beta} = [1/2, 1/2]$, $\boldsymbol{\alpha} = \begin{pmatrix} 1 & 2 \\ 8 & 1 \end{pmatrix}$



(d) $\boldsymbol{\beta} = [3/4, 1/4]$, $\boldsymbol{\alpha} = \begin{pmatrix} 1/3 & 1/5 \\ 1 & 1/2 \end{pmatrix}$

Figure 3.2.1: Spectrum (histogram) of the Hermitian random matrix H ($n = 1000$), conditions on $(\boldsymbol{\beta}, \boldsymbol{\alpha})$ are given in each panel. The solid line represents the distribution of the spectrum computed by the numerical approach. The dashed vertical line indicates the upper bound of the largest eigenvalue of H given by $2\,\|S\|_2^{1/2}$.

then a.s. matrix $(I-B)+(I-B)^T$ *is eventually positive definite: with probability one, for a given realization* ω*, there exists* $N(\omega)$ *such that for* $n \geqslant N(\omega)$*,* $(I - B^\omega) + (I - B^\omega)^\top$ *is positive definite. In particular, there exists a unique (random) globally stable equilibrium* $\boldsymbol{x}^* \in LCP(I - B^\omega, -\mathbf{1})$ *to* (3.14)*.*

Proof. In order to have the existence of a unique globally stable equilibrium, it is sufficient that matrix

$$-I + B + (-I + B^\top) = -2I + H + \tilde{P}\,,$$

is negative definite where

$$H = \frac{1}{\alpha\sqrt{n}} \begin{pmatrix} A_{11} + A_{11}^\top & A_{12} + A_{21}^\top \\ A_{21} + A_{12}^\top & A_{22} + A_{22}^\top \end{pmatrix},$$

$$\tilde{P} = \frac{1}{n} \begin{pmatrix} 2\mu_{11}\mathbf{1}_{\mathcal{I}_1}\mathbf{1}_{\mathcal{I}_1}^\top & (\mu_{12} + \mu_{21})\mathbf{1}_{\mathcal{I}_1}\mathbf{1}_{\mathcal{I}_2}^\top \\ (\mu_{21} + \mu_{12})\mathbf{1}_{\mathcal{I}_2}\mathbf{1}_{\mathcal{I}_1}^\top & 2\mu_{22}\mathbf{1}_{\mathcal{I}_2}\mathbf{1}_{\mathcal{I}_2}^\top \end{pmatrix}.$$

Recall that $-2I + (B + B^\top)$ is negative definite if and only if the largest eigenvalue of $B + B^\top$ is lower than 2 i.e.

$$\lambda_{\max}(B + B^\top) < 2\,. \tag{3.18}$$

Remark 3.8. In [CMN22], an admissible zone is described for a rank 1 mean matrix where the condition on the strength of the interactions is $\alpha > \sqrt{2}$.

The matrix H is a GOE (Gaussian Orthogonal Ensemble) matrix, its spectrum converges almost surely to the semi-circular law of support $[-\frac{2\sqrt{2}}{\alpha}, \frac{2\sqrt{2}}{\alpha}]$.

The theory of random matrices gives precise information about the spectrum of the matrix $H + \tilde{P}$. In particular, matrix $\tilde{P}$ may create outliers whose positions depend on α and the eigenvalues of the matrix $\tilde{P}$.

In Capitaine *et al.* [CDMF09, Th. 2.1], they describes the impact of a mean matrix of a finite rank symmetric matrix on the spectrum. In particular, they give the exact location of the spikes in the case of a symmetric matrix H and a mean symmetric matrix of finite rank which is the case here.

First of all, matrix $\tilde{P}$ is a deterministic symmetric matrix of fixed finite rank r and has J eigenvalues $\theta_1 > \cdots > \theta_J$ independent of N. To recover the largest eigenvalues, it is sufficient to study the matrix of dimension 2 (see (3.11) for details)

$$\mathrm{diag}(\boldsymbol{\beta})(\boldsymbol{\mu} + \boldsymbol{\mu}^\top)\,.$$

Denote by the largest eigenvalue by

$$\theta_{\max} = \lambda_{\max}(\mathrm{diag}(\boldsymbol{\beta})(\boldsymbol{\mu} + \boldsymbol{\mu}^\top))$$

Using the results of Capitaine *et al.* [CDMF09], the largest eigenvalue of matrix $H + \tilde{P}$ converges to

$$\lambda_{\max}(H + \tilde{P}) \xrightarrow[n\to\infty]{a.s} \begin{cases} \theta_{\max} + \frac{2}{\alpha^2\theta_{\max}} & \text{if } \theta_{\max} > \frac{\sqrt{2}}{\alpha}\,, \\ \frac{2\sqrt{2}}{\alpha} & \text{else.} \end{cases}$$

Assume first that $\theta_{\max} \leqslant \frac{\sqrt{2}}{\alpha}$, then $\lambda_{\max}(H + \tilde{P}) \xrightarrow[n\to\infty]{a.s.} \frac{2\sqrt{2}}{\alpha}$, which is strictly lower than 2 if $\alpha > \sqrt{2}$. Hence $\lambda_{\max}(H + \tilde{P})$ is eventually strictly lower than 2 under this condition.

Assume now that $\theta_{\max} > \frac{\sqrt{2}}{\alpha}$, then

$$\lambda_{\max}(H + \tilde{P}) \xrightarrow[n\to\infty]{a.s.} \theta_{\max} + \frac{2}{\alpha^2\theta_{\max}}\,.$$

We are interested in the conditions for which $\theta_{\max} + \frac{2}{\alpha^2\theta_{\max}} < 2$ or equivalently

$$\alpha^2\theta_{\max}^2 - 2\alpha^2\theta_{\max} + 2 < 0\,. \tag{3.19}$$

An elementary study of the polynomial $\xi(X) = \alpha^2\theta_{\max}^2 - 2\alpha^2\theta_{\max} + 2$ yields that ξ's discriminant is positive if $\alpha > \sqrt{2}$,

$$\xi(\theta_{\max}^{\pm}) = 0 \quad \Leftrightarrow \quad \theta_{\max}^{\pm} = 1 \pm \sqrt{1 - \frac{2}{\alpha^2}}\,,$$

and $\xi\left(\frac{\sqrt{2}}{\alpha}\right) < 0$, so that $\frac{\sqrt{2}}{\alpha} \in (\theta_{\max}^-, \theta_{\max}^+)$. In particular condition (3.19) is fulfilled if

$$\theta_{\max} \in \left(\frac{\sqrt{2}}{\alpha}\,,\ 1 + \sqrt{1 - \frac{2}{\alpha^2}}\right)\,.$$

Under this condition, (3.19) is fulfilled and a.s. $\limsup_{n\to\infty} \lambda_{\max}(H + \tilde{P}) < 2$. □

Remark 3.9. In the simplest case where the two communities are of equal size $\beta = \left(\frac{1}{2}, \frac{1}{2}\right)$, the analytical formula for the two eigenvalues are

$$\theta_1 = \frac{1}{2}\left[\mu_{11} + \mu_{22} + \sqrt{(\mu_{12}+\mu_{21})^2 + (\mu_{11}-\mu_{22})^2}\right],$$
$$\theta_2 = \frac{1}{2}\left[\mu_{11} + \mu_{22} - \sqrt{(\mu_{12}+\mu_{21})^2 + (\mu_{11}-\mu_{22})^2}\right].$$

Ecological interpretation of the condition. In this section, an analysis of the impact of the mean parameter $\boldsymbol{\mu}$ is achieved. We are interested in the sensitivity of mutualism and competition to the uniqueness and existence of an equilibrium when the two interacting communities are of the same size. One defines the function ψ by

$$\psi(\mu_{11}, \mu_{12}, \mu_{21}, \mu_{22}) = \frac{1}{2}\left[\mu_{11} + \mu_{22} + \sqrt{(\mu_{12}+\mu_{21})^2 + (\mu_{11}-\mu_{22})^2}\right],$$

corresponding to the largest eigenvalue θ_1 of $B + B^\top$ if $\theta_1 > \sqrt{2}/\alpha$.

The impact of the mean interaction coefficient μ_{12} on the largest eigenvalue is quantified by

$$\frac{\partial \psi}{\partial \mu_{12}} = \frac{\mu_{12}+\mu_{21}}{2\sqrt{(\mu_{12}+\mu_{21})^2 + (\mu_{11}-\mu_{22})^2}},$$
$$\frac{\partial \psi}{\partial \mu_{12}} > 0 \;\Leftrightarrow\; \mu_{12} > -\mu_{21}.$$

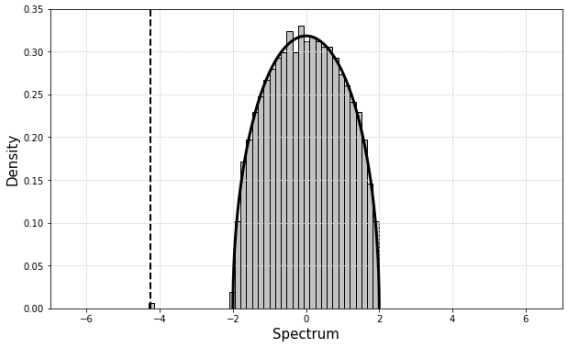
If the type of effect of Community 2 on Community 1 (μ_{12}) has the opposite behavior than the type of effect of the Community 1 on 2 (μ_{21}) i.e. competition/mutualism or mutualism/competition, then the equilibrium is likely to be unique globally stable. Conversely, this helps explain the following rather counter-intuitive phenomenon: when the competition between the communities increases, the left support of the spectrum is strongly affected. However, if $\boldsymbol{\mu}$ is of rank 2 then the right edge of the spectrum will also be affected when the second eigenvalue exceeds a certain threshold. Suppose that μ_{21} is negative, then the stronger the competition of μ_{12} will be, the more chance of losing the stability condition. This phenomenon does not appear in the case of a single community where the increase in competition does not affect stability (see [CMN22]). This phenomenon is shown in Figure 3.2.2. In this example, we increase the competition of Community 2 on Community 1, stability is lost when exceeding a certain threshold [AT12].

In Figure 3.2.3, the behavior of ψ as a function of the variable μ_{12} is represented by a parabola whose minimum is obtained in $\mu_{12} = -\mu_{21} = 2$. The general conclusion is that opposite inter-communities interactions behaviors between two community stabilize the ecosystem whereas a mostly competitive or mutualistic type of relationship has the opposite behavior.

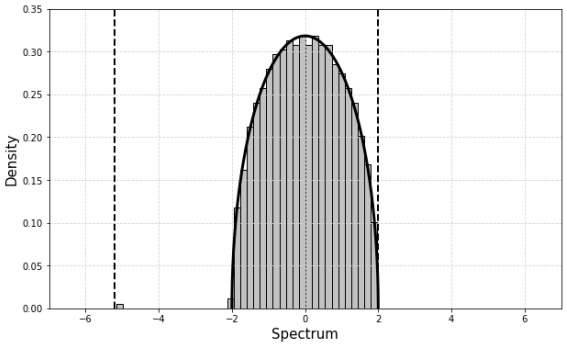
The same arguments can be performed on the diagonal terms, in particular to Community 1 μ_{11}.

$$\frac{\partial \psi}{\partial \mu_{11}} = \frac{1}{2} + \frac{\mu_{11}-\mu_{22}}{2\sqrt{(\mu_{12}+\mu_{21})^2 + (\mu_{11}-\mu_{22})^2}}.$$

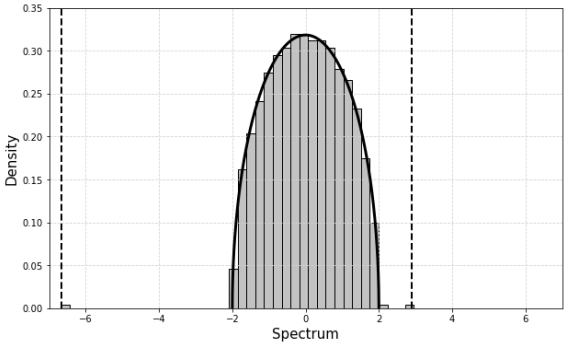
$$\frac{\partial \psi}{\partial \mu_{11}} > 0 \;\Leftrightarrow\; \sqrt{(\mu_{12}+\mu_{21})^2 + (\mu_{11}-\mu_{22})^2} + \mu_{11} - \mu_{22} > 0,$$
$$\Leftrightarrow\; \sqrt{(\mu_{12}+\mu_{21})^2 + (\mu_{11}-\mu_{22})^2} > \mu_{22} - \mu_{11}.$$



(a) $\alpha = \sqrt{2}$, $\boldsymbol{\mu} = \begin{pmatrix} -2 & -2 \\ -2 & -2 \end{pmatrix}$



(b) $\alpha = \sqrt{2}$, $\boldsymbol{\mu} = \begin{pmatrix} -2 & -4 \\ -2 & -2 \end{pmatrix}$



(c) $\alpha = \sqrt{2}$, $\boldsymbol{\mu} = \begin{pmatrix} -2 & -7 \\ -2 & -2 \end{pmatrix}$

Figure 3.2.2: Spectrum (histogram) of the Hermitian random matrix $B + B^\top$ ($n = 1000$, $\alpha = \sqrt{2}$). The solid line represents the semi-circular law. In Fig. 3.2.2a, $\boldsymbol{\mu} = -2$ and there is a unique negative outlier. In Fig. 3.2.2b, $\mu_{12} = -4$ and one can notice the presence of one outlier and the other outlier is contained in the bulk of the semi-circular law. In Fig. 3.2.2c, $\mu_{12} = -7$, the competition has reach a threshold with 2 outliers outside the bulk including on the right-side disturbing the stability. The dashed line indicates its theoretical value.

On the one hand, if $\mu_{11} > \mu_{22}$ this is always true, on the other hand by taking the square, one has $(\mu_{12} + \mu_{21})^2 > 0$ which is also true for any matrix $\boldsymbol{\mu}$. The behavior of intra-community mean is the same as in the one-community case: a negative mean does not affect stability, but a positive mean will have a very strong impact on the top eigenvalue and therefore stability.

Remark 3.10. The most general case could be treated by complex computations leading to implicit equations on the conditions and therefore difficult to interpret. A non-sharp bound by the triangular inequality

$$\left\| H + \tilde{P} \right\|_2 \leqslant \|H\|_2 + \left\| \tilde{P} \right\|_2 = \|H\|_2 + \max(|\theta_1|, |\theta_2|) .$$

This is not optimal in the sense that when competitive interactions are added the largest positive eigenvalue is not always influenced. However, if the mean matrix $\tilde{P}$ is large, it is anticipated that the eigenvalues that come out of the bulk are still "not too far" from the expected value.

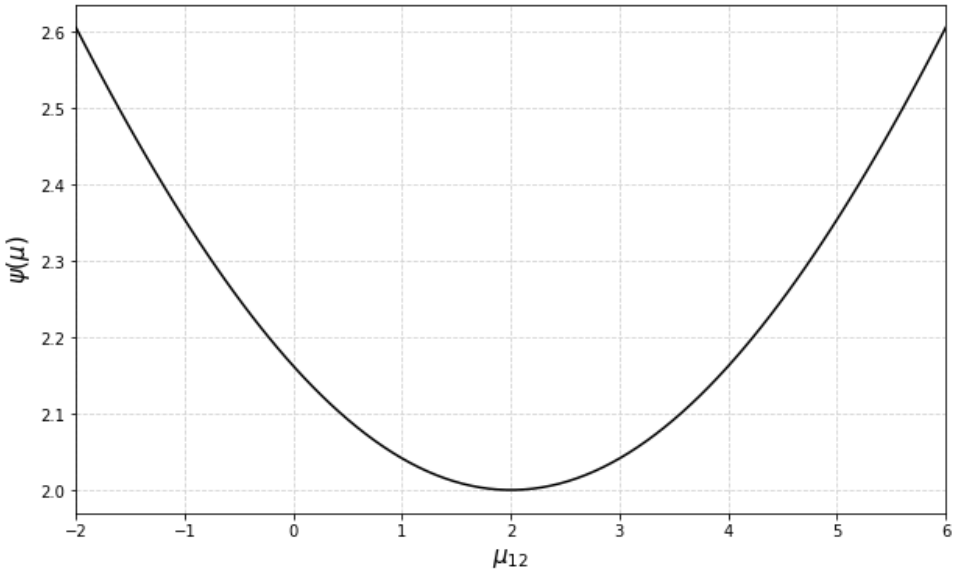


Figure 3.2.3: Representation of the function $\psi(\cdot, \mu_{12})$ by a parabola. The perturbed matrix is set $\boldsymbol{\mu} = \begin{pmatrix} -2 & \mu_{12} \\ -2 & -2 \end{pmatrix}$.

3.3 Persisting species

In Section 3.2, we have presented conditions on the matrix $\boldsymbol{\alpha}, \boldsymbol{\mu}$ and the proportion of community $\boldsymbol{\beta}$ for the existence of a globally stable equilibrium $\boldsymbol{x}^*$ to (3.1) under the non-invadability condition. The vector $\boldsymbol{x}^*$ is random and depends on the realization of matrix B. Moreover since $\boldsymbol{\alpha}$ has fixed components and does not depend on n, the equilibrium $\boldsymbol{x}^*$ will feature vanishing components in the case of many communities (see the original argument for a unique community in [DVR+18] and the discussion in [BN21]). In an ecological context, we define two categories of species in the vector $\boldsymbol{x}^*$, the persisting species (non-vanishing components $x_k^* > 0$) and the vanishing components corresponding to the species going to extinction with $x_k^* = 0$ and $x_k(t) \xrightarrow[t\to\infty]{} 0$.

In this section, we are interested in the properties (proportion, variance, distribution) of non-vanishing components of the equilibrium $\boldsymbol{x}^*$ for each community; we also describe the distribution of the persisting species $x_k^* > 0$ which turns out to be a truncated Gaussian.

Remark 3.11. The Gaussianity assumption facilitates the explanation of the heuristics but does not seem necessary for the result to hold. In Figure 3.3.3, the entries are not considered Gaussian anymore but the two first moment $\mathbb{E}(B_{k\ell})$ and $\mathbb{E}(|B_{k\ell}|^2)$ coïncide. In this case, the distribution of the persisting species still matches the truncated Gaussian.

3.3.1 A heuristics of the number of persisting species

Assume that the considered ecosystem has two distinct interacting communities. The set of persisting species in community $i \in \{1, 2\}$ is defined as

$$\mathcal{S}_1 = \{k \in \mathcal{I}_1,\ x_k^* > 0\} \quad ; \quad \mathcal{I}_1 = [1, \beta_1 n]\,,$$

$$\mathcal{S}_2 = \{k \in \mathcal{I}_2,\ x_k^* > 0\} \quad ; \quad \mathcal{I}_2 = [\beta_1 n + 1, n]\,.$$

Given the random equilibrium $\boldsymbol{x}^*$, we introduce the following quantities for each community $i \in \{1,2\}$

$$\hat{p}_i = \frac{|\mathcal{S}_i|}{|\mathcal{I}_i|}, \quad \hat{m}_i = \frac{1}{|\mathcal{S}_i|}\sum_{k\in\mathcal{I}_i} x_k^*, \quad \hat{\sigma}_i^2 = \frac{1}{|\mathcal{S}_i|}\sum_{k\in\mathcal{I}_i}(x_k^*)^2 .$$

Denote by $Z \sim \mathcal{N}(0,1)$ a standard Gaussian random variable and by Φ the cumulative Gaussian distribution function:

$$\Phi(x) = \int_{-\infty}^{x} \frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}}\, du .$$

Heuristics 3.1. *Let $(\boldsymbol{\alpha}, \boldsymbol{\mu})$ and assume that either condition of Theorem 3.6 or Proposition 3.7 holds, then the following system of six equations and six unknowns $(p_1, p_2, m_1, m_2, \sigma_1, \sigma_2)$*

$$\begin{aligned}
p_1 &= 1 - \Phi(\delta_1) , && (3.20)\\
p_2 &= 1 - \Phi(\delta_2) , && (3.21)\\
m_1 &= 1 + \lambda_1 + \Delta_1 \mathbb{E}(Z|Z>\delta_1) , && (3.22)\\
m_2 &= 1 + \lambda_2 + \Delta_2 \mathbb{E}(Z|Z>\delta_2) , && (3.23)\\
(\sigma_1)^2 &= (1+\lambda_1)^2 + 2(1+\lambda_1)\Delta_1 \mathbb{E}(Z|Z>\delta_1) + \Delta_1^2 \mathbb{E}(Z^2|Z>\delta_1) , && (3.24)\\
(\sigma_2)^2 &= (1+\lambda_2)^2 + 2(1+\lambda_2)\Delta_2 \mathbb{E}(Z|Z>\delta_2) + \Delta_2^2 \mathbb{E}(Z^2|Z>\delta_2) , && (3.25)
\end{aligned}$$

where

$$\Delta_i = \sqrt{p_1(\sigma_1)^2 \frac{\beta_1}{\alpha_{i1}^2} + p_2(\sigma_2)^2 \frac{\beta_2}{\alpha_{i2}^2}} \ ; \ \lambda_i = p_1 m_1 \beta_1 \mu_{i1} + p_2 m_2 \beta_2 \mu_{i2} \ ; \ \delta_i = \frac{-1-\lambda_i}{\Delta_i}, \quad (3.26)$$

admits a unique solution $(p_1^*, p_2^*, m_1^*, m_2^*, \sigma_1^*, \sigma_2^*)$ and $\forall\ i \in \{1,2\}$

$$\hat{p}_i \xrightarrow[n\to\infty]{a.s.} p_i^* , \qquad \hat{m}_i \xrightarrow[n\to\infty]{a.s.} m_i^* \qquad \text{and} \qquad \hat{\sigma}_i \xrightarrow[n\to\infty]{a.s.} \sigma_i^* .$$

There is a strong matching between the parameters obtained by solving (3.20)-(3.25) and their empirical counterparts obtained by Monte-Carlo simulations. This is illustrated in Fig. 3.3.1.

3.3.1.1 Construction of the heuristics

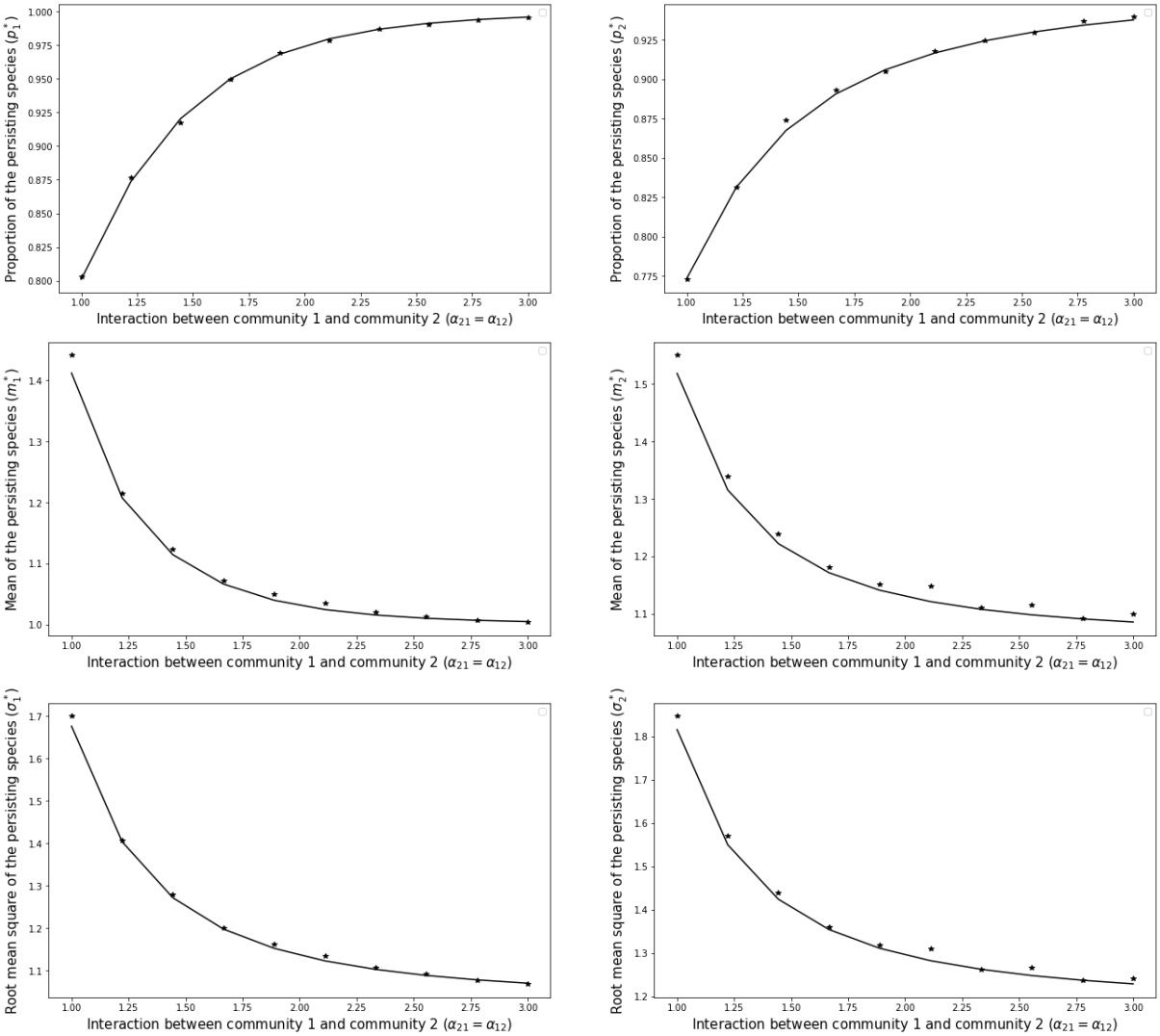
Getting information about the fixed point is equivalent to solving the LCP problem

$$x_k^* \left(1 - x_k^* + \sum_{\ell=1}^{n} B_{k\ell} x_\ell^*\right) = 0 \ , \ \forall k \in [n] .$$

Consider the random variables:

$$\forall k \in [n], \ \check{Z}_k = \sum_{\ell \in \mathcal{S}_1 \cup \mathcal{S}_2} B_{k\ell} x_\ell^* .$$

We assume that asymptotically the x_ℓ^*'s are independent from the $B_{k\ell}$'s, an assumption supported by the chaos hypothesis, see for instance Geman and Hwang [GH82]. Denote by $\mathbb{E}_{\boldsymbol{x}^*} = \mathbb{E}(\cdot \mid \boldsymbol{x}^*)$ the conditional expectation with respect to $\boldsymbol{x}^*$. Notice that conditionally



(a) Parameters $(p_1^*, \sigma_1^*, m_1^*)$ versus α_{21}/α_{12}.

(b) Parameters $(p_2^*, \sigma_2^*, m_2^*)$ versus α_{21}/α_{12}.

Figure 3.3.1: Comparison between the theoretical solutions $(p_1^*, p_2^*, \sigma_1^*, \sigma_2^*, m_1^*, m_2^*)$ of (3.20)-(3.25) and their empirical Monte Carlo counterpart (the star marker) as functions of the off-diagonal block interaction strength α_{12}/α_{21}. The left column is associated to the properties of Community 1. The rigth column is associated to the properties of Community 2. Matrix B has size $n = 100$ and the number of Monte Carlo experiments is 500. The parameters are $\boldsymbol{\mu} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$, $\boldsymbol{\alpha} = \begin{pmatrix} 3 & \alpha_{12} \\ \alpha_{21} & \sqrt{2} \end{pmatrix}$, $\boldsymbol{\beta} = \left(\frac{1}{2}, \frac{1}{2}\right)$. When off-diagonal block interaction $\alpha_{12}^{-1}, \alpha_{21}^{-1}$ increases, the number of persisting species p^* decrease but their variance σ^* and mean m^* increase.

to $\boldsymbol{x}^*$, the $\check{Z}_k$'s are independent Gaussian random variables, whose two first moments can easily be computed, see Appendix 3.B for the details

$$\forall k \in \mathcal{I}_i,\ \mathbb{E}_{\boldsymbol{x}^*}(\check{Z}_k) = \hat{p}_1 \hat{m}_1 \beta_1 \mu_{i1} + \hat{p}_2 \hat{m}_2 \beta_2 \mu_{i2}\,,$$

$$\forall k \in \mathcal{I}_i, \mathrm{var}_{\boldsymbol{x}^*}(\check{Z}_k) = \hat{p}_1 \hat{\sigma}_1^2 \frac{\beta_1}{\alpha_{i1}^2} + \hat{p}_2 \hat{\sigma}_2^2 \frac{\beta_2}{\alpha_{i2}^2}\,.$$

Notice that the fact that $\mathbb{E}_{\boldsymbol{x}^*}$ and $\mathrm{var}_{\boldsymbol{x}^*}(\check{Z}_k)$ only depend on $\hat{p}_1$, $\hat{p}_2$, $\hat{m}_1$, $\hat{m}_2$, $\hat{\sigma}_1$, $\hat{\sigma}_2$ (that are converging quantities) supports the idea that $\check{Z}_k$ is unconditionally a Gaussian random variable with moments:

$$\mathbb{E}\check{Z}_k = p_1^* m_1^* \beta_1 \mu_{i1} + p_2^* m_2^* \beta_2 \mu_{i2} \quad \text{and} \quad \mathrm{var}(\check{Z}_k) = p_1^* (\sigma_1^*)^2 \frac{\beta_1}{\alpha_{i1}^2} + p_2^* (\sigma_2^*)^2 \frac{\beta_2}{\alpha_{i2}^2}\,,$$

where p_1^*, p_2^*, m_1^*, m_2^* σ_1^*, σ_2^* are resp. the limits of $\hat{p}_1$, $\hat{p}_2$, $\hat{m}_1$, $\hat{m}_2$, $\hat{\sigma}_1$, $\hat{\sigma}_2$. We can introduce two families of standard Gaussian random variables $(Z_k)_{k\in\mathcal{I}_1}$ and $(Z_k)_{k\in\mathcal{I}_2}$:

$$\forall k \in \mathcal{I}_i, \quad Z_k = \frac{\check{Z}_k - \mathbb{E}(\check{Z}_k)}{\sqrt{\mathrm{var}(\check{Z}_k)}} = \frac{\check{Z}_k - p_1^* m_1^* \beta_1 \mu_{i1} + p_2^* m_2^* \beta_2 \mu_{i2}}{\sqrt{p_1^* (\sigma_1^*)^2 \frac{\beta_1}{\alpha_{i1}^2} + p_2^* (\sigma_2^*)^2 \frac{\beta_2}{\alpha_{i2}^2}}}.$$

To simplify the following computations, we denote:

$$\Delta_i^* = \sqrt{p_1^* (\sigma_1^*)^2 \frac{\beta_1}{\alpha_{i1}^2} + p_2^* (\sigma_2^*)^2 \frac{\beta_2}{\alpha_{i2}^2}}\ ;\ \lambda_i^* = p_1^* m_1^* \beta_1 \mu_{i1} + p_2^* m_2^* \beta_2 \mu_{i2}\ ;\ \delta_i^* = \frac{-1-\lambda_i^*}{\Delta_i^*}\,.$$

Here, λ_i^* (resp Δ_i^*) corresponds to the average effect of the interactions on the community i (resp average variance of the interactions on the community i).

Consider the equilibrium $\boldsymbol{x}^* = (x_k^*)_{k\in[n]}$, the definition of the LCP equilibrium implies if $k \in \mathcal{S}_1 \cup \mathcal{S}_2$:

$$x_k^*(1 - x_k^* + (Bx^*)_k) = 0 \text{ and } 1 + (B\boldsymbol{x}^*)_k = 1 + \check{Z}_k > 0\,.$$

We finally obtain the following relationship for the persisting species:

$$x_k^* = 1 + \lambda_i^* + \Delta_i^* Z_k \text{ if } k \in \mathcal{S}_i\,. \tag{3.27}$$

Heuristics (3.20)-(3.21)**.** We can write the first two equations:

$$\mathbb{P}(x_k^* \geqslant 0 | k \in \mathcal{S}_1) = 1 - \Phi(\delta_1^*)\,,$$

and

$$\mathbb{P}(x_k^* \geqslant 0 | k \in \mathcal{S}_2) = 1 - \Phi(\delta_2^*)\,.$$

We finally obtain (3.20) and (3.21):

$$p_1^* = 1 - \Phi(\delta_1^*)\,.$$

$$p_2^* = 1 - \Phi(\delta_2^*)\,.$$

Heuristics (3.22)-(3.23)**.** Our starting point is the following generic representation of an abundance at equilibrium (either of a persisting or vanishing species) in the case $k \in \mathcal{S}_i$:

$$x_k^* = (1 + \lambda_i^* + \Delta_i^* Z_k)\, \mathbf{1}_{\{Z_k > \delta_i^*\}} \\ = (1 + \lambda_i^*)\, \mathbf{1}_{\{Z_k > \delta_i^*\}} + (\Delta_i^* Z_k)\, \mathbf{1}_{\{Z_k > \delta_i^*\}}\,.$$

Summing over $\mathcal{S}_i$ and normalizing,

$$\begin{aligned}
\frac{1}{|\mathcal{S}_i|} \sum_{k \in \mathcal{S}_i} x_k^* &= (1 + \lambda_i^*) \frac{1}{|\mathcal{S}_i|} \sum_{k \in \mathcal{S}_i} \mathbf{1}_{\{Z_k > \delta_i^*\}} + \Delta_i^* \frac{1}{|\mathcal{S}_i|} \sum_{k \in \mathcal{S}_i} Z_k \mathbf{1}_{\{Z_k > \delta_i^*\}}, \\
\hat{m}_i &\overset{(a)}{=} (1 + \lambda_i^*) + \Delta_i^* \frac{|\mathcal{I}_i|}{|\mathcal{S}_i|} \frac{1}{|\mathcal{I}_i|} \sum_{k \in \mathcal{I}_i} Z_k \mathbf{1}_{\{Z_k > \delta_i^*\}}, \\
\hat{m}_i &\overset{(b)}{\simeq} (1 + \lambda_i^*) + \Delta_i^* \frac{1}{\mathbb{P}(Z > \delta_i^*)} \mathbb{E}(Z \mathbf{1}_{\{Z > \delta_i^*\}}), \\
\hat{m}_i &\simeq (1 + \lambda_i^*) + \Delta_i^* \mathbb{E}(Z \mid Z > \delta_i^*).
\end{aligned}$$

where (a) follows from the fact that $|\mathcal{S}_i| = \sum_{k \in \mathcal{S}_i} \mathbf{1}_{\{Z_k > \delta_i^*\}}$ (by definition of $\mathcal{S}_i$), (b) from the law of large numbers $\frac{1}{|\mathcal{I}_i|} \sum_{k \in \mathcal{I}_i} Z_k \mathbf{1}_{\{Z_k > \delta_i\}} \xrightarrow[n \to \infty]{} \mathbb{E} Z \mathbf{1}_{\{Z > \delta_i^*\}}$ and $\frac{|\mathcal{S}_i|}{|\mathcal{I}_i|} \xrightarrow[n \to \infty]{} \mathbb{P}(Z > \delta_i^*)$ with $Z \sim \mathcal{N}(0,1)$. It remains to replace $\hat{m}_i$ by its limit m_i^* to obtain heuristics 2. We finally obtain the 3rd and 4th equations:

$$m_1^* = 1 + \lambda_1^* + \Delta_1^* \mathbb{E}(Z | Z > \delta_1^*)$$
$$m_2^* = 1 + \lambda_2^* + \Delta_2^* \mathbb{E}(Z | Z > \delta_2^*)$$

Heuristics (3.24)-(3.25)**.** By similar computations, one can obtain similarly the 5th and 6th equations:

$$(\sigma_1^*)^2 = (1 + \lambda_1^*)^2 + 2(1 + \lambda_1^*)\Delta_1^* \mathbb{E}(Z | Z > \delta_1^*) + (\Delta_1^*)^2 \mathbb{E}(Z^2 | Z > \delta_1^*)$$
$$(\sigma_2^*)^2 = (1 + \lambda_2^*)^2 + 2(1 + \lambda_2^*)\Delta_2^* \mathbb{E}(Z | Z > \delta_2^*) + (\Delta_2^*)^2 \mathbb{E}(Z^2 | Z > \delta_2^*)$$

General properties of the ecosystem Information on the properties of the total proportion of the block can be computed.

1. Proportion of persisting species.

$$\mathbb{P}(x_k^* \geqslant 0) = \mathbb{P}(x_k^* \geqslant 0 | k \in \mathcal{I}_1)\mathbb{P}(k \in \mathcal{I}_1) + \mathbb{P}(x_k^* \geqslant 0 | k \in \mathcal{I}_2)\mathbb{P}(k \in \mathcal{I}_2)\,,$$
$$p^* = p_1^* \beta_1 + p_2^* \beta_2\,.$$

2. Mean square of the abundance of the persisting species.

$$\mathbb{E}((x_k^*)^2) = \mathbb{E}((x_k^*)^2 | k \in \mathcal{I}_1)\mathbb{P}(k \in \mathcal{I}_1) + \mathbb{E}((x_k^*)^2 | k \in \mathcal{I}_2)\mathbb{P}(k \in \mathcal{I}_2)\,,$$
$$(\sigma^*)^2 = (\sigma_1^*)^2 \beta_1 + (\sigma_2^*)^2 \beta_2\,.$$

3. Mean of the abundance of the persisting species

$$\mathbb{E}(x_k^*) = \mathbb{E}(x_k^* | k \in \mathcal{I}_1)\mathbb{P}(k \in \mathcal{I}_1) + \mathbb{E}(x_k^* | k \in \mathcal{I}_2)\mathbb{P}(k \in \mathcal{I}_2)\,,$$
$$m^* = m_1^* \beta_1 + m_2^* \beta_2\,.$$

3.3.2 Distribution of the persisting species

We may recall the following representation of the abundance x_k^* of a persisting species when $k \in \mathcal{S}_i$:

$$x_k^* = 1 + \lambda_i^* + \Delta_i^* Z_k \text{ if } k \in \mathcal{S}_i,$$

where $Z_k \sim \mathcal{N}(0,1)$ and $Z_k > \delta_i^* = \delta_i(p_i^*, m_i^*, \sigma_i^*)$ defined in (3.27). This representation allows to characterize the distribution of x_k^* of each community. It turns out that the persisting species of each community follow a truncated Gaussian distribution.

Heuristics 3.2. *Let $(\boldsymbol{\alpha}, \boldsymbol{\mu})$ and assume that either condition of Theorem 3.6 or Proposition 3.7 holds and let $(p_1^*, p_2^*, m_1^*, m_2^*, \sigma_1^*, \sigma_2^*)$ be the solution of the system (3.20)-(3.25). Recall the definition (3.26) of λ_i, Δ_i, δ_i and denote by $\delta_i^* = \delta_i(p_i^*, m_i^*, \sigma_i^*)$. Let $x_k^* > 0$ be a positive component of $\boldsymbol{x}^*$ belonging to the community i, then:*

$$\mathcal{L}(x_k^*) \xrightarrow[n\to\infty]{} \mathcal{L}\left(1 + \lambda_i^* + \Delta_i^* Z \quad \middle| \quad Z > \delta_i^*\right) ,$$

where $Z \sim \mathcal{N}(0,1)$. Otherwise stated, asymptotically $\forall k \in \mathcal{S}_i$, x_k^* admits the following density

$$f_k(y) = \frac{\mathbf{1}_{\{y>0\}}}{\Phi(-\delta_i^*)} \frac{1}{\Delta_i^* \sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{y}{\Delta_i^*} + \delta_i^*\right)^2\right\} . \tag{3.28}$$

The heuristics simply follows from the fact that if x_k^* is a persisting species and $k \in \mathcal{S}_i$ then

$$x_k^* = 1 + \lambda_i^* + \Delta_i^* Z_k,$$

conditionally on the fact that the right hand side of the equation is positive, that is $Z_k > \delta_i^*$. A simple change of variable yields the density - details are provided in Appendix 3.B.

Fig. 3.3.2 illustrates the matching between the theoretical distribution obtained by 3.28 and a histogram obtained by generating the interaction matrix for 2 communities. In Figure 3.3.3, the validity of heuristics in the case of non-Gaussian entries is illustrated.

3.3.3 Toward a general case

In the heuristics 3.1 and 3.2, we are restricted to conditions to have unique and globally stable equilibrium on $(\boldsymbol{\alpha}, \boldsymbol{\mu})$ given by Theorem 3.6 and Proposition 3.7.

(i) On the one hand, in Theorem 3.6, we assume that $\boldsymbol{\mu} = \mathbf{0}$ and give condition on $\boldsymbol{\alpha}$.

(ii) On the other hand, in Proposition 3.7, we assume $\boldsymbol{\alpha} = \alpha \mathbf{1}\mathbf{1}^\top$. The first condition is $\alpha > \sqrt{2}$ and we give conditions on $\boldsymbol{\mu}$ depending of α.

However, in Figure 3.3.4, we notice that the results remain convincing for couples $(\boldsymbol{\alpha}, \boldsymbol{\mu})$ beyond these assumptions. For this reason, we shall use in the sequel the heuristics beyond the conditions of Theorem 3.6 and Proposition 3.7. We remain confident that by using the QVE theory for the non-centered case, we could extend the conditions over $(\boldsymbol{\alpha}, \boldsymbol{\mu})$ to have a unique and globally stable equilibrium beyond conditions (i) and (ii).

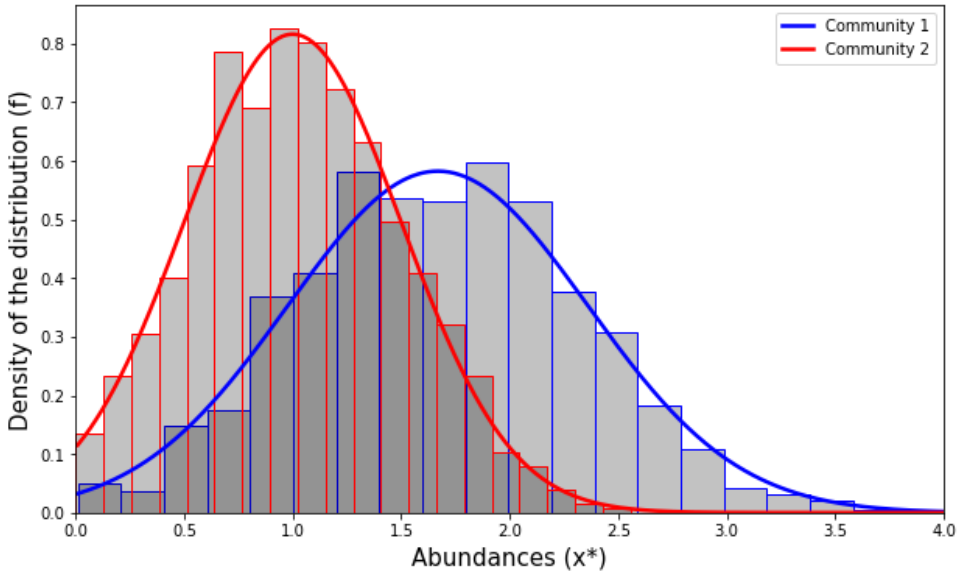


Figure 3.3.2: Distribution of persisting species in each community. The x-axis represents the value of the abundances and the histogram is built upon the positive components of equilibrium $\boldsymbol{x}^*$ associated to each community. The blue-solid line (resp. red-solid line) represents the theoretical distribution of Community 1 (resp. Community 2) for parameters $(\boldsymbol{\alpha}, \boldsymbol{\mu})$ as given by Heuristics 3.2. The entries are Gaussian $\mathcal{N}(0,1)$ and the parameters are set to

$$n = 1000\ ,\ \boldsymbol{\mu} = \begin{pmatrix} 0.5 & 0.5 \\ 0 & 0 \end{pmatrix}\ ,\ \boldsymbol{\alpha} = \begin{pmatrix} 2 & 3 \\ 3 & 3 \end{pmatrix}\ ,\ \boldsymbol{\beta} = \begin{pmatrix} \frac{1}{2}, \frac{1}{2} \end{pmatrix}\ .$$

3.3.4 Diversity is contagious

In Figure 3.0.3, adding interactions between two feasible communities disturb the persistence of the ecosystem i.e. beyond a given threshold, some species go extinct. Using the heuristic equations (3.20)-(3.25), the evolution of the proportion of persisting species in each communities can be quantified. Understanding the effect of species richness in each communities becomes an important endeavour. In particular, the effect of the species richness of Community 1 on Community 2 when interactions are added (see Figure 3.3.5). In Figure 3.3.6, we represent the impact of the species richness of Community 1 (which depends mainly on α_{11}) on Community 2. When Community 1 has more persisting species (larger α_{11}), its impact is smaller on the persistence decay of Community 2. We conclude that there is a contagion of diversity: the higher the persistence of a community, the less its impact will be harmful on the other communities. This can be compared to a spatial averaging or insurance effect: the more persistent species there are in the active community, the more interactions the passive community will receive i.e. functional complementarity in the communities (see Loreau *et al.* [LNI+01, LMG03]). From an evolutionary standpoint, contagious diversity has been studied by Calcagno *et al.* [CJL+17] where diversity favours diversification.

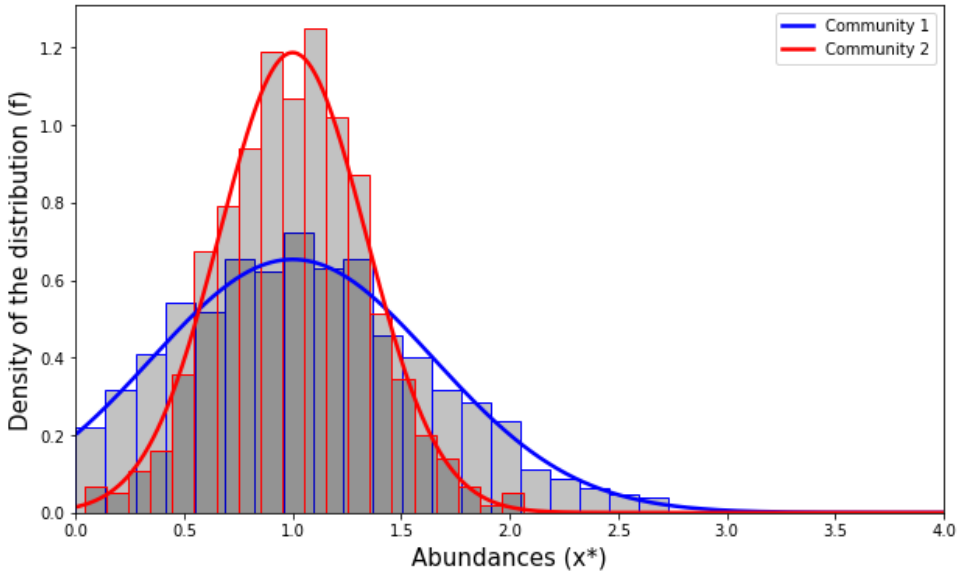


Figure 3.3.3: Distribution of persisting species in each community. The x-axis represents the value of the abundances and the histogram is built upon the positive components of equilibrium $\boldsymbol{x}^*$ associated to each community. The solid lines (blue for Community 1, red for Community 2) represents the theoretical distribution for parameters $(\boldsymbol{\alpha}, \boldsymbol{\mu})$ as given by Heuristics 3.2. The entries are uniform $\mathcal{U}(-\sqrt{3}, \sqrt{3})$ with variance 1 and the parameters are set to

$$n = 1000\ ,\ \boldsymbol{\mu} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}\ ,\ \boldsymbol{\alpha} = \begin{pmatrix} 2 & 1.5 \\ 3 & 4 \end{pmatrix}\ ,\ \boldsymbol{\beta} = \begin{pmatrix} \frac{1}{2}, \frac{1}{2} \end{pmatrix}\ .$$

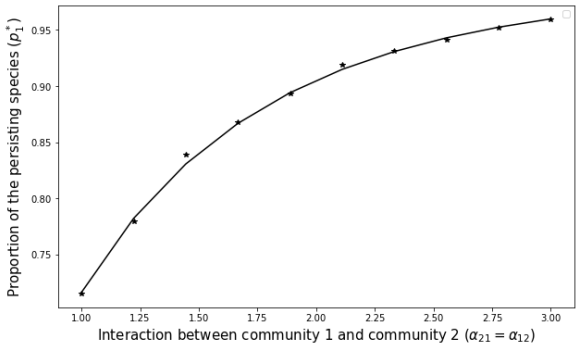
Notice in particular that the theoretical distribution matches even with non-Gaussian entries.

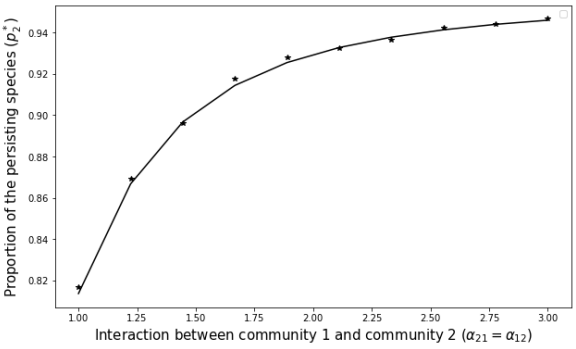
3.3.5 Feedback effect

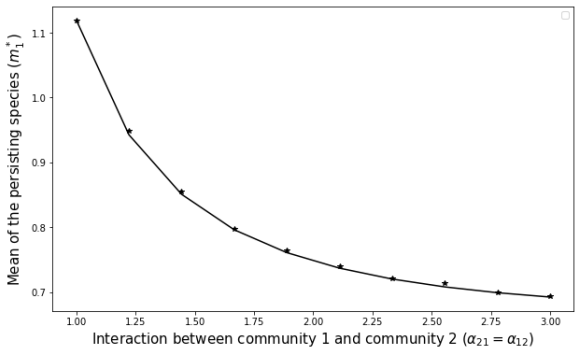
A natural question that emerges is the existence of a feedback effect. In Figure 3.3.7, a diagram of the situation is represented, two communities are interacting, the impact of Community 1 on Community 2 is increasing. When this increase occurs, if Community 1 has higher persistence then Community 2 is less severely affected (see section 3.3.4). However, when Community 2 is affected, there is a feedback loop affecting Community 1. The evolution of the persistence is represented in Figure 3.3.8. A decrease of the persistence p_1^* is observed when the interaction increases i.e. α_{21} decreases. The feedback effect is negative, which produces a detrimental cycle: if Community 1 is less persistent, it affects Community 2 more negatively, which in turn affects Community 1 more strongly. The decline in persistence between two interacting communities is not linear but has a double negative effect, hence the importance of maintaining persistent communities and not neglecting feedback phenomena when dealing with ecosystems (see Loreau [LNI+01]).

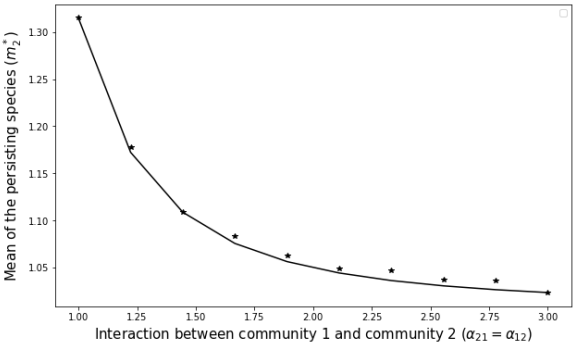
3.3.6 Type of food web interactions

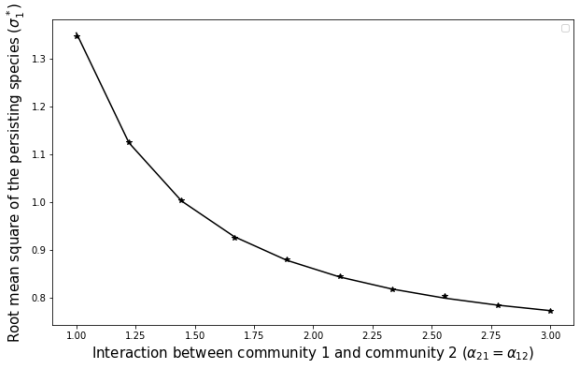
In the previous sections, the impact of the strength of interactions $\boldsymbol{\alpha}$ was investigated without the mean interaction parameter $\boldsymbol{\mu} = 0$. However, particular types of food webs

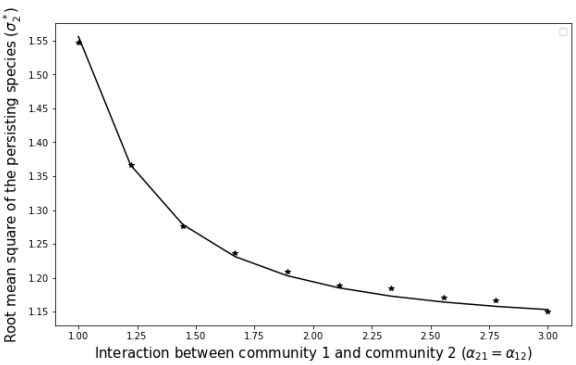












(a) Parameters $(p_1^*, \sigma_1^*, m_1^*)$ versus α_{21}/α_{12}.

(b) Parameters $(p_2^*, \sigma_2^*, m_2^*)$ versus α_{21}/α_{12}.

Figure 3.3.4: Comparison between the theoretical solutions $(p_1^*, p_2^*, \sigma_1^*, \sigma_2^*, m_1^*, m_2^*)$ of (3.20)-(3.25) and their empirical Monte Carlo counterpart (the star marker) as functions of the off-diagonal block interaction strength α_{12}/α_{21}. The left column is associated to the properties of Community 1. The right column is associated to the properties of Community 2. Matrix B has size $n = 100$ and the number of Monte Carlo experiments is 500. The parameters are $\boldsymbol{\mu} = \begin{pmatrix} -0.3 & -0.5 \\ 0.3 & -0.3 \end{pmatrix}$, $\boldsymbol{\alpha} = \begin{pmatrix} 2 & \alpha_{12} \\ \alpha_{21} & \sqrt{2} \end{pmatrix}$, $\boldsymbol{\beta} = \left(\frac{1}{2}, \frac{1}{2}\right)$. When off-diagonal block interaction $\alpha_{12}^{-1}, \alpha_{21}^{-1}$ increases, the number of persisting species p^* decrease but their variance σ^* and mean m^* increase.

(mutualism, competition, antagonistic) arise in nature and their consequences could be related to specific patterns. The use of the mean parameter $\boldsymbol{\mu}$ of the matrix B allows to control the sign of interactions (on average) in each block, and therefore the nature of the food web.

Inter-community interactions. Let us consider two communities whose interactions are initially mostly competitive ($\mu_{12} = \mu_{21} < 0$), mutualistic ($\mu_{12} = \mu_{21} > 0$) or antago-

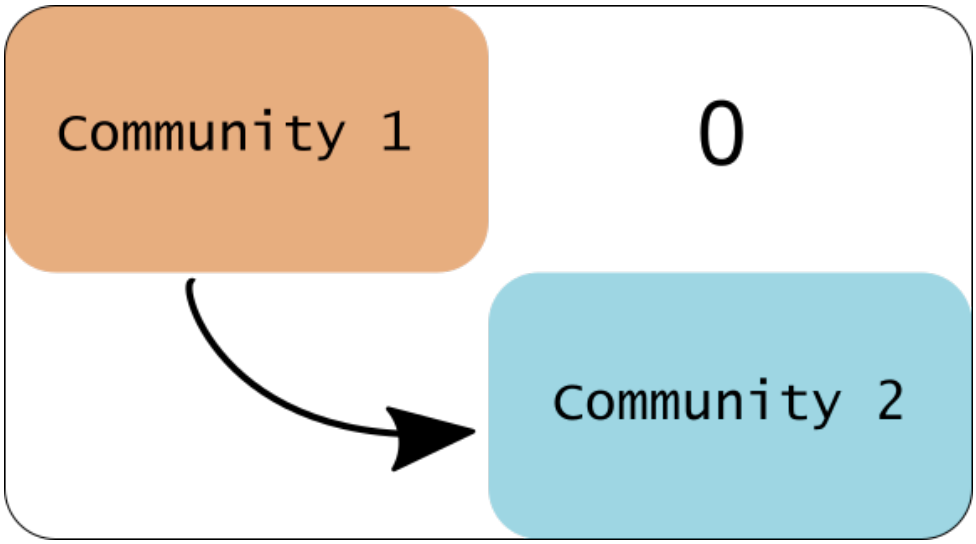


Figure 3.3.5: Diagram of the experience of the impact of the species richness of Community 1 on Community 2. The off-diagonal block A_{12} is set to 0, $\alpha_{12} \to \infty$.

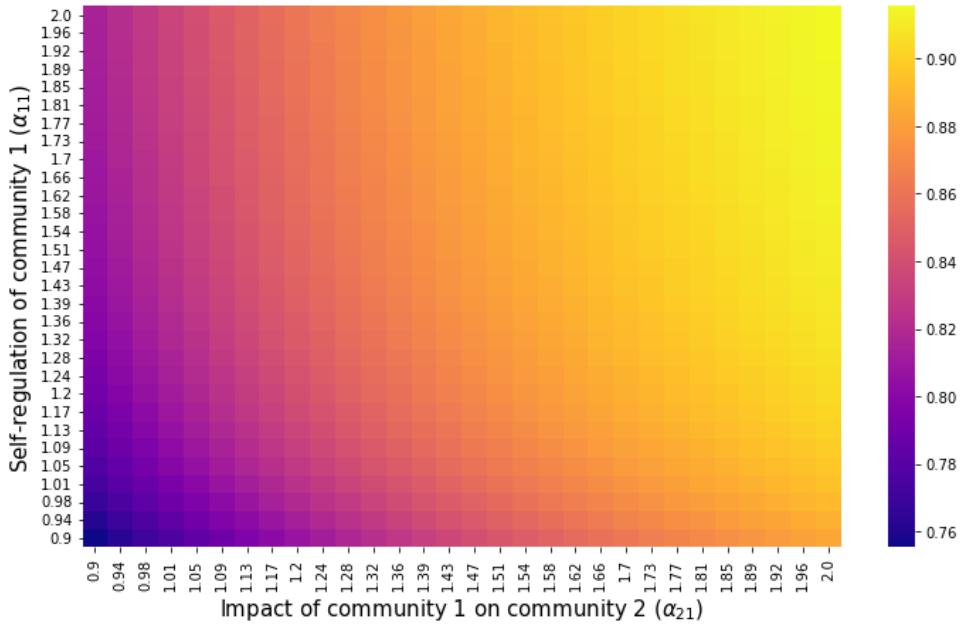


Figure 3.3.6: Heatmap of the proportion of persisting species p_2^* in Community 2. The x-axis represents the impact of Community 1 on Community 2 by increasing the off-diagonal interaction strength α_{21}. The y-axis represents the interaction strength in Community 1 which is directly related to the proportion of persistent species p_1^*.

nistic ($\mu_{12} = -\mu_{21}$).

Remark 3.12. From a theoretical standpoint, in order to have pairwise antagonistic interactions within the community, we would have to add a correlation parameter ρ between the pairs of interactions. Antagonistic relationships appears when $\rho < 0$. Here, an inverse relationship between the inter-community parameter $\mu_{12} = -\mu_{21}$ increases the number of antagonistic interactions on average.

The previous results (see Figure 3.1.2) indicate a decrease in persistence in both com-

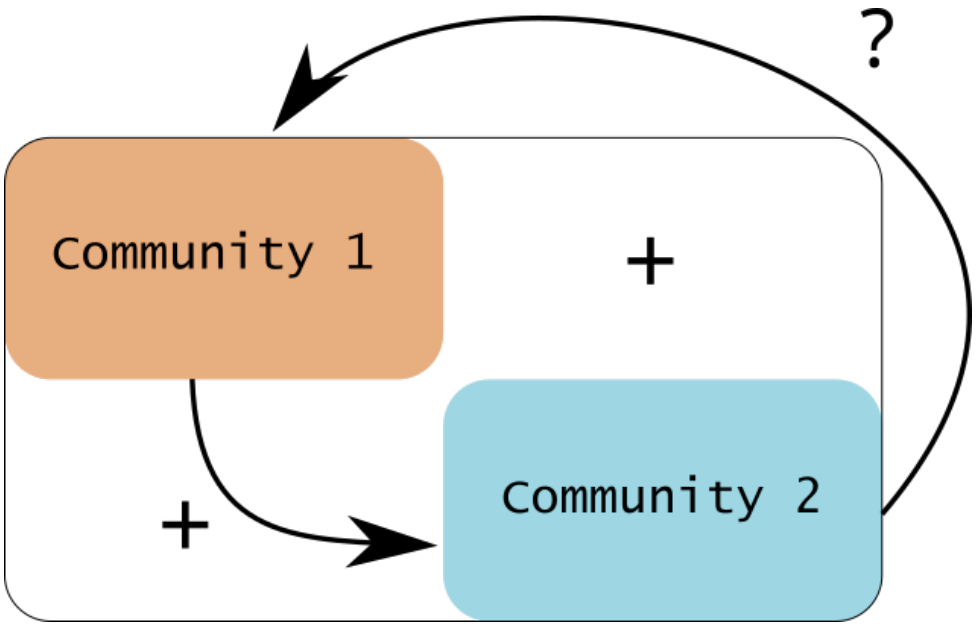


Figure 3.3.7: Diagram of the experience of a feedback effect. The off-diagonal interaction strength α_{12} is set to a constant. The interaction of Community 1 on Community 2 α_{21} is enhanced.

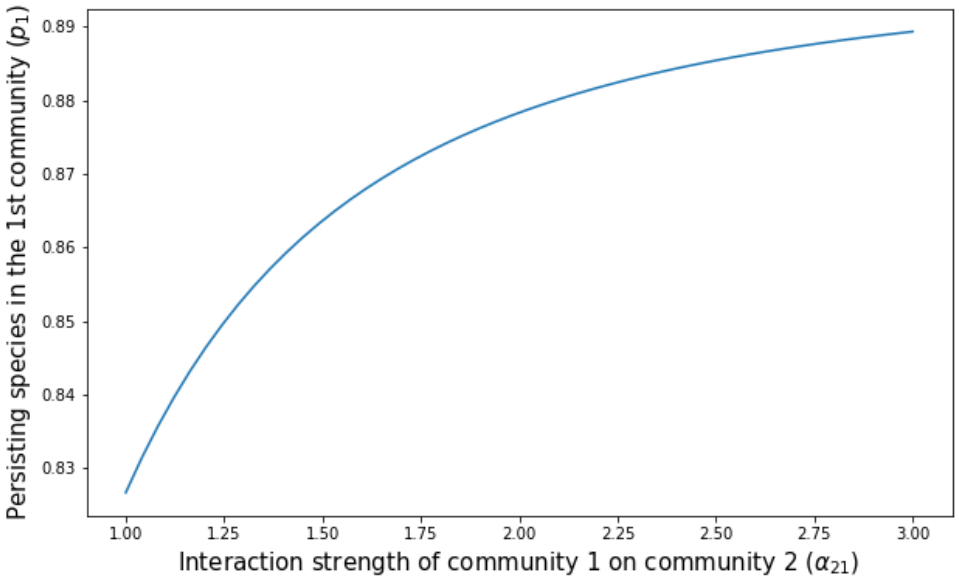
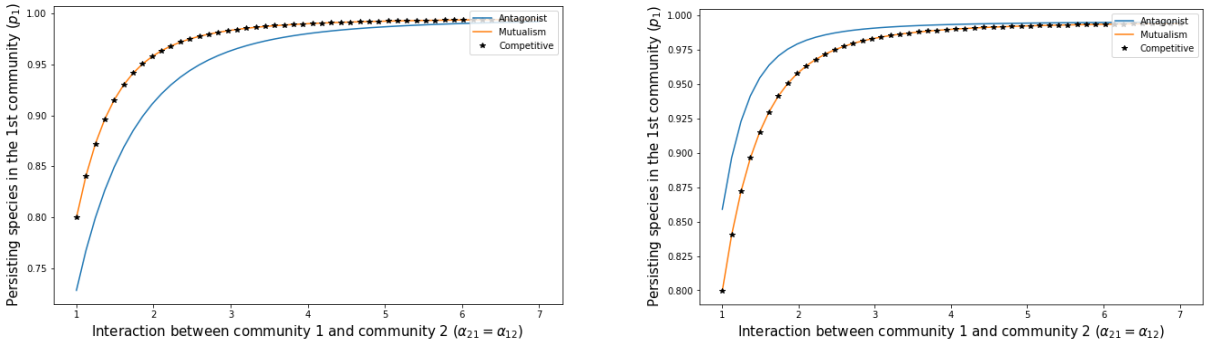


Figure 3.3.8: Representation of the feedback effect. Proportion of the persisting species p_1^* in Community 1 when increasing the interaction strength of Community 1 on Community 2 (α_{21} decreases).

munities with no mean ($\boldsymbol{\mu} = 0$) when there is an increase in inter-community interactions. In Figure 3.3.9, we consider these three different cases. The impact on persistence is the same for mutualistic and competitive interactions. However, for an antagonistic relationship, the community benefiting from the interaction (here Community 2) will have a better persistence [TF10, AT12]. Inversely Community 1 will be deficient of the interaction.



(a) Community 1

(b) Community 2

Figure 3.3.9: Representation of the impact of the type of inter-community interactions. The two figures illustrates the persisting species in each community (Community 1 in Fig. 3.3.9a, Community 2 in Fig. 3.3.9b) as a function of the inter-community interaction strength ($\alpha_{12} = \alpha_{21}$). Three types of inter-community are investigated:

$$\text{Antagonist } \boldsymbol{\mu}_a = \begin{pmatrix} 0 & -0.4 \\ 0.4 & 0 \end{pmatrix},$$

$$\text{Mutualism } \boldsymbol{\mu}_m = \begin{pmatrix} 0 & 0.4 \\ 0.4 & 0 \end{pmatrix},$$

$$\text{Competition } \boldsymbol{\mu}_c = \begin{pmatrix} 0 & -0.4 \\ -0.4 & 0 \end{pmatrix}.$$

Intra-community interactions. The type of inter-community interactions has an effect on persistence within communities. The type of intra-community interactions can also have a significant effect. For this purpose, we compare the persistence of Community 2 under the condition that Community 1 is mostly competitive or mutualistic. In Figure 3.3.10, we notice that when interactions are added (α_{21}), and the type of interaction is mostly competitive in Community 1 then the persistence of Community 2 is less affected. The idea is that if competition is strong within Community 1, then the abundances will be more homogeneous in Community 1 and therefore the effect will be weaker on Community 2. Conversely, a mutualistic community will have a greater and more heterogeneous abundance and its effect will be strongly increased over the other communities.

More resilient community. A natural question is which type of community is better able to withstand the impact of other communities. In Figure 3.3.11, we observe that the larger the mean of Community 2, the better the community's persistence. Upon comparison with Figure 3.3.10, the two heatmaps appear similar with the exception of a permutation. A mostly mutualistic community $\mu_{22} = 0.4$ is as resistant to the impact of any community as a standard community which is resistant to a mostly competitive community $\mu_{11} = -0.4$ and therefore less affected.

Some additional summary graphs. Given the number of parameters of the model, it is possible to test many possibilities. See appendix 3.E for other types of effects.

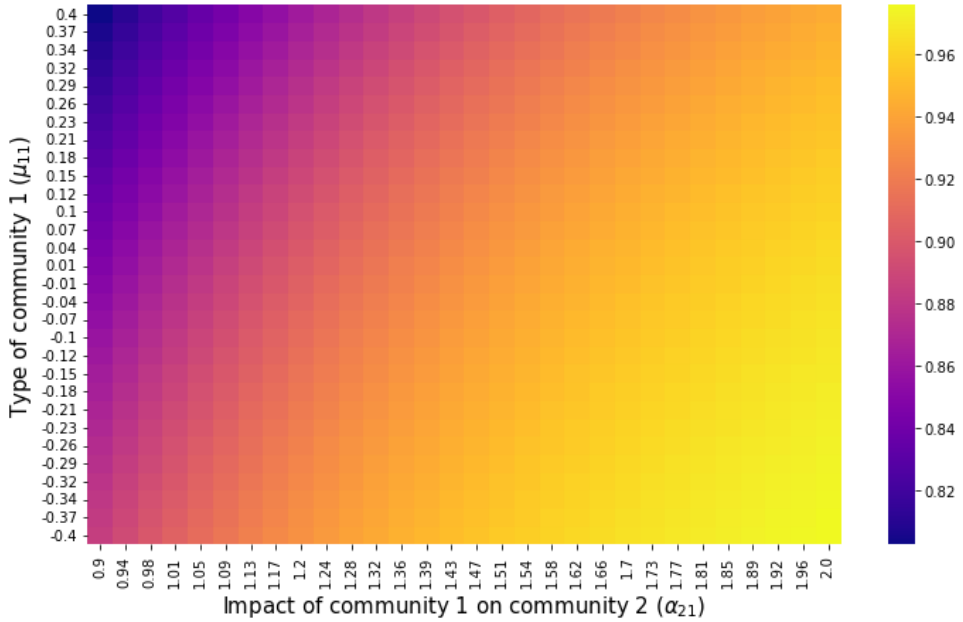


Figure 3.3.10: Heatmap of the proportion of persisting species p_2^* in Community 2. The x-axis corresponds to the impact of Community 1 on Community 2 by increasing the off-diagonal interaction strength α_{21}. The y-axis corresponds to the disturbance μ_{11} in Community 1 which is directly related to the average type of interactions (mutualism/competition). The parameters are

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_{11} & 0 \\ 0 & 0 \end{pmatrix} \ , \ \boldsymbol{\alpha} = \begin{pmatrix} 2 & \infty \\ \alpha_{21} & 2 \end{pmatrix} \ , \ \boldsymbol{\beta} = \left(\frac{1}{2}, \frac{1}{2}\right) .$$

3.4 Parallel between connectance and interaction strength

In this article, we have mainly focused on the impact of the parameter $\boldsymbol{\alpha}$ i.e. on the observation of the behavior of the model when the strength of the interactions varies. In his work [May72], May provided a stability threshold depending on three factors (n, σ, C) in the case of a single size community. n is the dimension of the system, σ is the variance of the interactions, and C is the average proportion of non-zero values in the community matrix. These three factors summarize the properties of the community matrix (the Jacobian) which represents the impact between species around the equilibrium. The stability condition is given by

$$\sigma\sqrt{nC} < 1 .$$

When compared to the matrix model (3.2) as a Jacobian matrix, the factor n is absorbed by σ which is of order $O(n^{-1/2})$ by construction. The second factor σ is identified with our $1/\alpha$ - the inverse of the interaction strength and the last is the connectance C. Following these remarks, May's criterion with (3.2) is rewritten with our notations as

$$\frac{\sqrt{C}}{\alpha} < 1 .$$

The objective of this section is to discuss through simulations whether the parameters of connectance C and strength of interactions α may exactly offset or mimic each other in

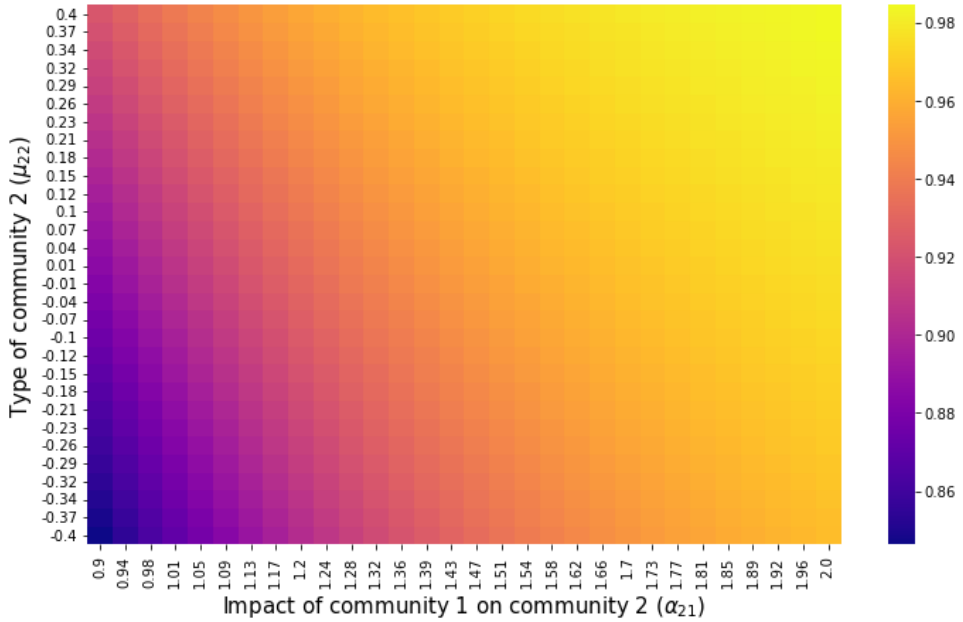


Figure 3.3.11: Heatmap of the proportion of persisting species p_2^* in Community 2. The x-axis corresponds to the impact of Community 1 on Community 2 by increasing the off-diagonal interaction strength α_{21}. The y-axis corresponds to the disturbance μ_{11} in Community 1 which is directly related to the average type of interactions (mutualism/competition). The parameters are

$$\boldsymbol{\mu} = \begin{pmatrix} -0.4 & 0 \\ 0 & \mu_{22} \end{pmatrix} , \ \boldsymbol{\alpha} = \begin{pmatrix} 2 & \infty \\ \alpha_{21} & 2 \end{pmatrix} , \ \boldsymbol{\beta} = \left(\frac{1}{2}, \frac{1}{2}\right) .$$

their effects on feasibility, stability and persistent species properties. In physical terms, is there a composite parameter that completely determines the feasibility in the Lotka-Volterra model? In particular, can we make a parallel between the connectance and the strength of interactions?

In the sequel, we can define the connectance as a function (similarity) on the strength of the interactions by

$$C = f(\alpha) = \frac{1}{\alpha^2}, \forall \alpha \in [1, \infty) . \tag{3.29}$$

Remark 3.13. The similarity relation is only defined for $\alpha \geqslant 1$ whereas for $1/\sqrt{2} < \alpha < 1$, the Lotka-Volterra model (3.3) admits a unique equilibrium [Bun17].

Similar interaction matrix To compare the similarities between connectance and interaction strength in the model, we define a new interaction matrix B^S that does not depend on interaction strength. From (3.2), one can define $\tilde{B}$ whose interaction strength is the same for all matrices i.e. $\boldsymbol{\alpha} = \mathbf{1}\mathbf{1}^\top$ and S the adjacency matrix of a specific stochastic block model, then

$$B^S = S \circ \tilde{B} ,$$

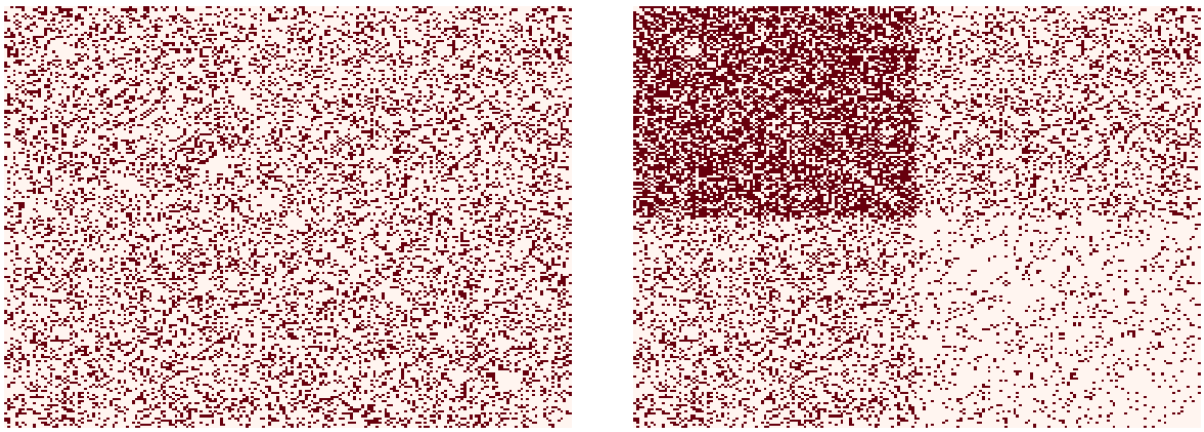
where $\circ$ corresponds to the Hadamard product between two matrices.

Remark 3.14. In the rest of the section, we assume that all graphs are undirected. If species k interacts with species ℓ, reciprocally species ℓ interacts with species k. This specific choice reduces the type of relationship that we can have in our model, "absolute" commensalism and amensalism (in the sense that $B^S_{k\ell} \neq 0$ and $B^S_{\ell k} = 0$) are not represented. However, we can consider that this type of relationship is taken into account if we have a significantly greater relationship of one species on another than the reverse.

A brief reminder about Erdös-Rényi (ER) graphs denoted by $\mathcal{G}(n,p)$. It is a graph with n vertices. It is assumed that there is an edge between two vertices with probability p independent from every other edge. (see [Bol98] for a review on random graphs). The adjacency matrix associated to this graph is symmetric and contains on average $pn(n-1)$ entries equal to 1. We associate the connectance $C = p$, we say that if $p = O(1)$, then the graph is dense which will be the study case. The case of an interaction matrix B^S where S is an adjacency graph of an Erdös-Rényi matrix has been already considered by May in the context of the Jacobian.

Let consider a subclass of stochastic block model (see [Abb18, LW19] for review on SBM). Let an ecosystem have b communities $\mathcal{C}_1, ..., \mathcal{C}_b$, let $\boldsymbol{\beta} = (\beta_1, ..., \beta_b)$ be the sizes of each community such that $\sum_{i=1}^{b} \beta_i = 1$. A SBM is a random graph whose vertices are partitioned into b communities. Given $P := (p_{ij}, 1 \leqslant i, j \leqslant b,\ p_{ij} \in [0,1])$ a symmetric matrix, there exists an edge between vertex $u \in \mathcal{C}_i$ and $v \in \mathcal{C}_j$ with probability p_{ij} independent from every other edge. To summarize, each block corresponds to an ER graph where the final adjacency matrix S is symmetric. We associate a connectivity matrix $\boldsymbol{C} = P$ with $\forall 1 \leqslant i, j \leqslant b,\ p_{ij} = O(1)$ i.e. each graph associated to each block is dense. In ecology and in general, SBMs are used to cluster the species in communities (see [BDB+11, MM17].

In Figure 3.4.1, the adjacency matrices of an Erdös Renyi graph and an SBM are illustrated.



(a) Erdös Renyi (b) Stochastic Block Model

Figure 3.4.1: Representation of an adjacency matrix of the interactions of an ecosystem of size $n = 200$. In Fig (a), an Erdös Renyi graph of parameter $\mathcal{G}(200, 0.25)$ is illustrated. In Fig (b) a graph of a symmetric Stochastic Block model of parameter $P = \begin{pmatrix} 0.6 & 0.25 \\ 0.25 & 0.1 \end{pmatrix}$. A red colored cell indicates $S_{k\ell} = 1$, on the contrary, a white colored cell indicates that there is no interaction $S_{k\ell} = 0$.

3.4.1 Standard case: a unique community

Two types of matrices are compared if there is a single community: B^S vs. B where S is the adjacency matrix of an ER graph, B is defined in (3.2) with $\boldsymbol{\alpha} = \frac{1}{\alpha}\mathbf{1}\mathbf{1}^\top$ and $\boldsymbol{\mu} = \mathbf{0}$ for both models.

Feasibility Following the work of Bizeul and Najim [BN21], the feasibility threshold is given by $\alpha_n = \sqrt{2\log(n)}$. From (3.29),

$$C_n = f(\alpha_n) = \frac{1}{2\log(n)}$$

corresponds to the feasibility transition for the matrix B^S. In Figure 3.4.2, we observe that the feasibility threshold is consistent. Even if the normalization term $1/\sqrt{n}$ is always

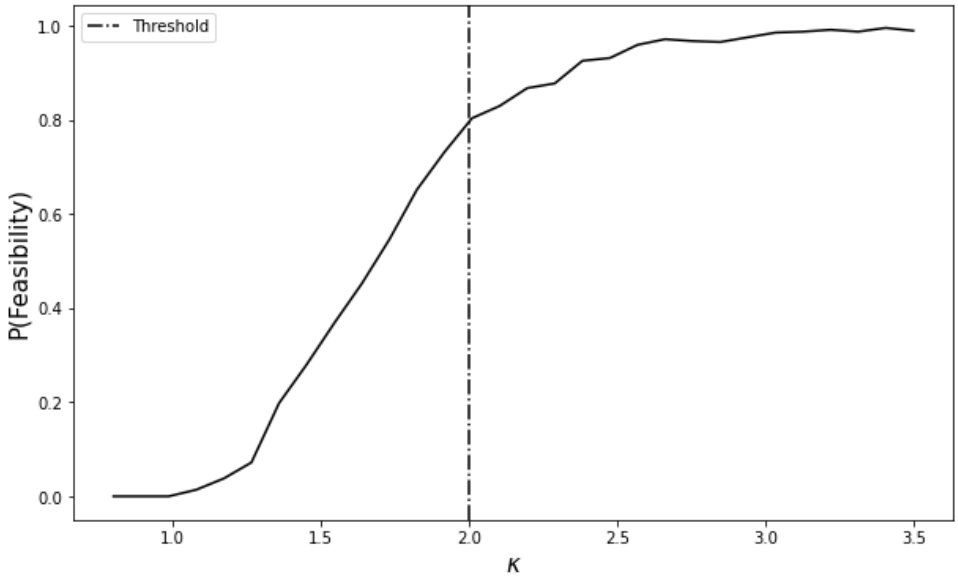
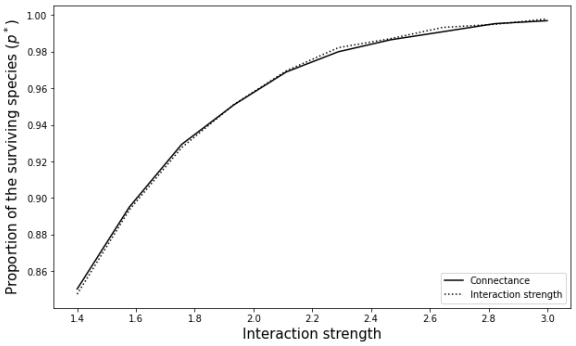


Figure 3.4.2: Transition towards feasibility for the model B^S. For each value $\kappa = \alpha^2/\log(n)$ on the x-axis, we simulate 500 matrices B^S of size $n = 500$ and compute the solution $\boldsymbol{x}$ of Theorem (3.1) adapted to B^S at the scaling $C_n(\kappa) = 1/\kappa\log(n)$. The curve represents the proportion of feasible solutions $\boldsymbol{x}$ obtained for the 500 simulations. The dotdashed vertical line corresponds to $\kappa = 2$.

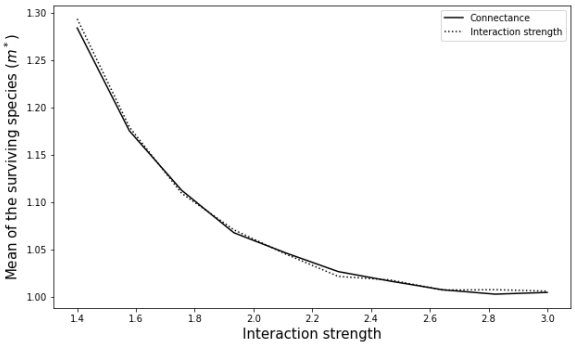
present, the transition threshold in $C_n = 1/2\log(n)$ expresses a rather large proportion of interactions in an ecosystem. For example, the critical connectance threshold for a system with 1000 species is $C = 1/4$. For $C > 1/4$, the probability that a random LV system with 1000 species admits a feasible equilibrium decreases quickly.

Persisting species The second step consists of testing the properties of the persisting species in the model. In the case of a single community, the heuristics has been established by Clenet *et al.* [CMN22]. In two separate procedures, we compute empirically the properties of the two models B and B^S. In Figure 3.4.3, we observe that the two models seem to match for the 3 properties: proportion of persisting species, mean and root mean square of persisting species. The remark 3.11 on the Gaussian assumption can be

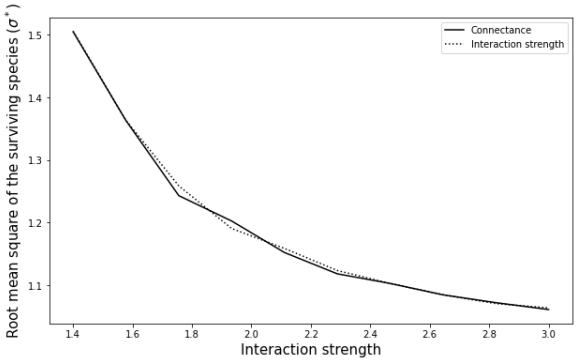
completed when the connectance between species is inferior than 1. There exists a trade-off between the strength of the interactions (α) and the connectance C in the LV model with a single community.



(a) Proportion of persisting species



(b) Mean of the persisting species



(c) Root mean square of the persisting species

Figure 3.4.3: Comparison between the empirical solutions (properties) (p^*, m^*, σ^*) of matrix B and B^S as functions of the interaction strength α and connectance $C = f(\alpha)$. Matrix B and B^S has size $n = 100$ and the number of Monte Carlo experiments is 300. The mean is fixed to zero $\mu = 0$.

3.4.2 Many communities: Stochastic Block Model

The results for a single community can be extended to several communities using the framework of this paper. Suppose, there are two communities where B is defined by (3.2) with interactions between two symmetric communities

$$\boldsymbol{\beta} = (\beta_1, \beta_2)\ ,\ \boldsymbol{\alpha} = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{12} & \alpha_{22} \end{pmatrix}\ ,\ \boldsymbol{\mu} = \mathbf{0}\,.$$

The connectance model is defined by the matrix B^S where S is an adjacency matrix of a symmetric Stochastic Block Model defined by

$$\boldsymbol{\beta} = (\beta_1, \beta_2)\ ,\ \boldsymbol{C} = f(\boldsymbol{\alpha}) = \begin{pmatrix} 1/\alpha_{11}^2 & 1/\alpha_{12}^2 \\ 1/\alpha_{12}^2 & 1/\alpha_{22}^2 \end{pmatrix}\,.$$

Two specific properties are studied, the global stability and the properties of persisting species in each community (proportion of persisting species, mean and root mean square of the persisting species).

Global stability properties

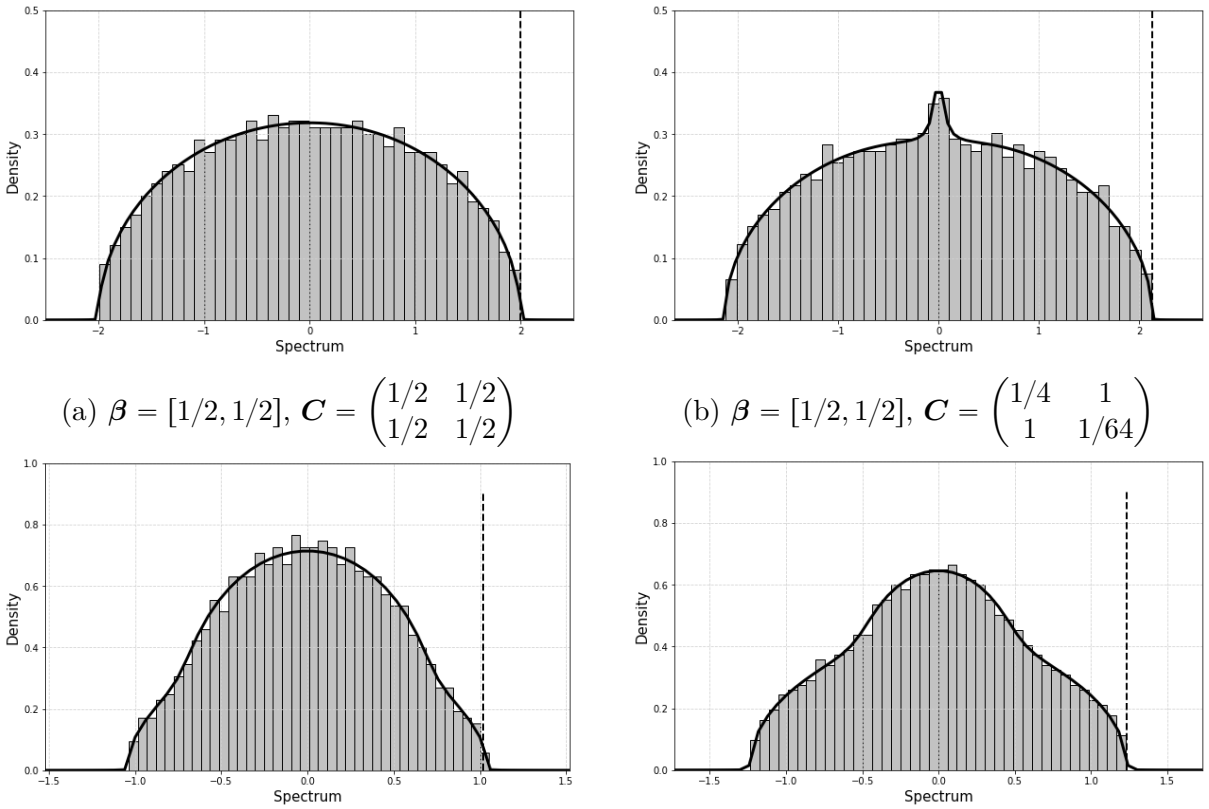
Conjecture 3.1. Let $\boldsymbol{\mu} = \mathbf{0}$ and

$$\left\|\mathrm{diag}(\boldsymbol{\beta})^{1/2}\left(\boldsymbol{C}+\boldsymbol{C}^\top\right)\mathrm{diag}(\boldsymbol{\beta})^{1/2}\right\|_2 < 1\,,$$

then a.s. matrix $(I-B^S)+(I-B^S)^\top$ is eventually positive definite: with probability one, for a given realization ω, there exists $N(\omega)$ such that for $n \geqslant N(\omega)$, $(I-(B^S)^\omega)+(I-(B^S)^\omega)^\top$ is positive definite. In particular, there exists a unique (random) globally stable equilibrium $\boldsymbol{x}^* \in LCP(I-(B^S)^\omega, -\mathbf{1})$ to (3.14).

The significance of the conjecture inequality is simple to formalize. Assume that there are two communities of the same size $\beta_1 = \beta_2 = 1/2$ and recall that the matrix $\boldsymbol{C}$ is symmetric. The condition then becomes $\|\boldsymbol{C}\| < 1$.

The histogram of the real eigenvalues of the $B+B^\top$ matrix seems similar to that of the $B^S+(B^S)^\top$ matrix. In figure 3.4.4 the histogram of the matrix $B^S+(B^S)^\top$ is compared to the distribution and the theoretical bound of the matrix $B+B^\top$. The result motivates the conjecture.



(a) $\boldsymbol{\beta} = [1/2, 1/2]$, $\boldsymbol{C} = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$

(b) $\boldsymbol{\beta} = [1/2, 1/2]$, $\boldsymbol{C} = \begin{pmatrix} 1/4 & 1 \\ 1 & 1/64 \end{pmatrix}$

(c) $\boldsymbol{\beta} = [1/2, 1/2]$, $\boldsymbol{C} = \begin{pmatrix} 1/9 & 1/25 \\ 1/25 & 1/4 \end{pmatrix}$

(d) $\boldsymbol{\beta} = [1/4, 3/4]$, $\boldsymbol{C} = \begin{pmatrix} 1/9 & 1/25 \\ 1/25 & 1/4 \end{pmatrix}$

Figure 3.4.4: Spectrum (histogram) of the Hermitian random matrix $B^S+(B^S)^\top$ ($n = 1000$), condition on $(\boldsymbol{\beta}, \boldsymbol{C})$ are given in each sub-figures. The solid line represents the distribution of the spectrum computed by the numerical approach. The dashed vertical line indicates the upper bound of the largest eigenvalue of $B+B^\top$ given by $2\,\|S\|_2^{1/2}$.

Properties of the persisting species We compare the two models for the properties of persisting species with two communities. The heuristics have been established in the section 3.3. We compute empirically the properties of the models B and B^S. In figure 3.4.5, we look at these properties for each community: proportion of persisting species, mean and root mean square. The results show curves that merge into each others and confirm the similarity between connectance C and interaction strength α when there are several communities.

3.4.3 Many communities with mean interaction parameter $\mu \neq 0$

In the framework of a community matrix, we consider a certain interaction trend $\mu \neq 0$. We distinguish two types of model:

$$B = \frac{A}{\alpha\sqrt{n}} + \frac{\mu}{n}\mathbf{1}\mathbf{1}^\top$$

and

$$B^S = S \circ \tilde{B},\ \tilde{B} = \frac{A}{\sqrt{n}} + \frac{\mu}{n}\mathbf{1}\mathbf{1}^\top$$

where S is the adjacency matrix of an ER graph $\mathcal{G}(n, C)$, $C = f(\alpha)$.

We have seen in section 3.2, that when we change the trend μ of the matrix, the stability can be affected by a "spike" eigenvalue that goes out of the bulk. We observe that for the matrix B^S, we apply the graph structure on matrix $\tilde{B}$. This implies that the spike eigenvalue of the matrix will be less pronounced because only a proportion C of components are affected. If we want a perturbation equivalent to that of the matrix B we should add a trend $\tilde{\mu} = \alpha\mu$ in matrix B^S.

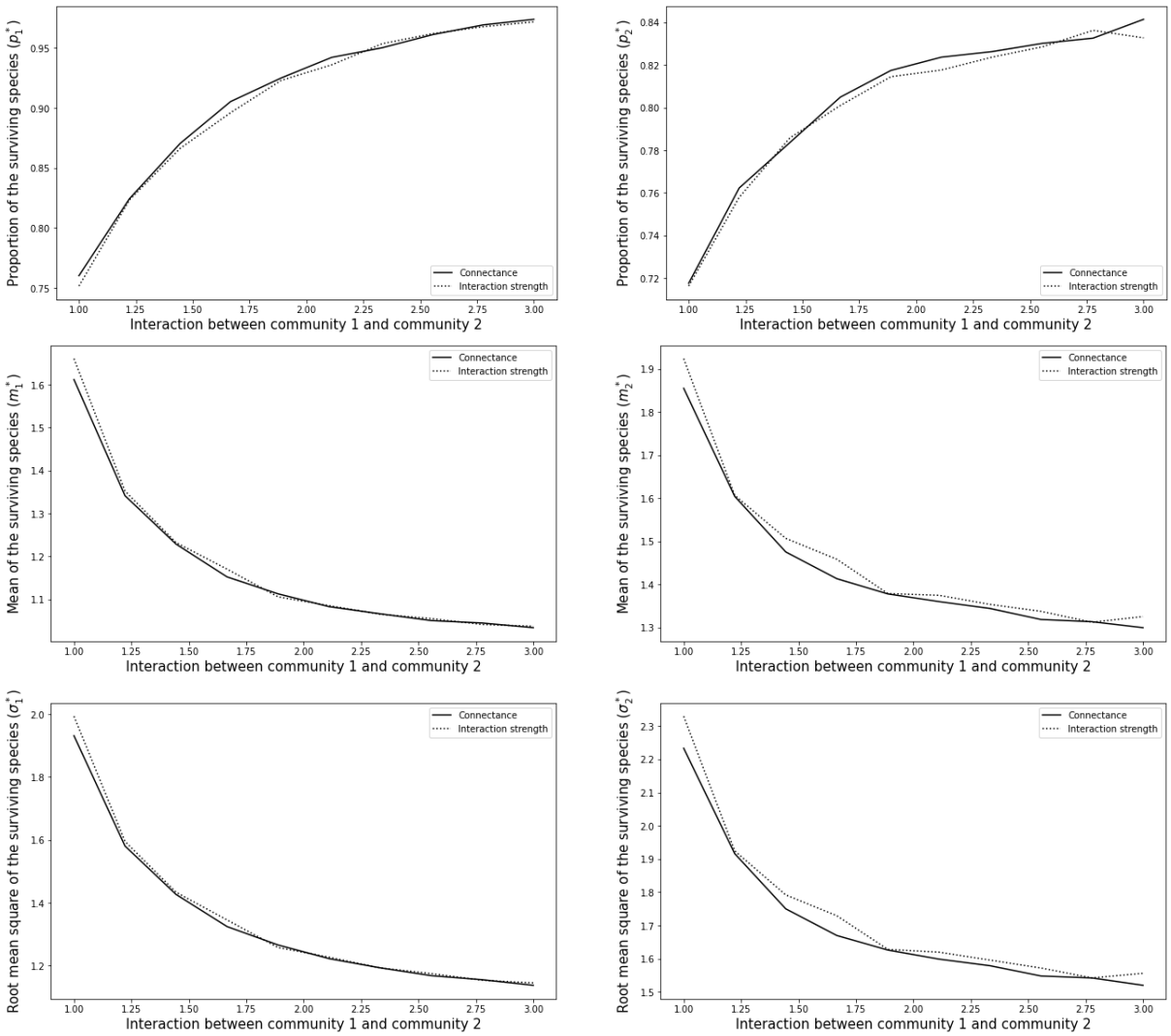
Individually, each species can undergo a larger mean interaction strength while remaining stable. We fall back on a complexity result in the sense that the B^S matrix is less complex in terms of the number of interactions.

3.4.4 Partial conclusion

The SBM model is a type of food web that has been widely used in ecological models. We choose a symmetric SBM, this allows to give a reasonable ecological condition so that the interaction between two species is always reciprocal. We show numerically that in the framework of the Lotka-Volterra model (3.3) choosing a type of network with connectivity is equivalent to being interested in a model where all species are connected with a variable interaction strength. In the Lotka-Volterra multi-community model, the relation between connectance and interaction strength is $\boldsymbol{C} = f(\boldsymbol{\alpha})$. Even if the meaning is different, we fall back on the same stability criteria as for the Jacobian in the paper of May [May72].

However, when the interaction matrix has a interaction trend, its effect is weaker in the case of a matrix with a lower connectance because it affects fewer interactions. It is possible to find an equivalent by increasing the trend and finding spikes of the same size.

A numerical study has been performed, a challenge would be to compare theoretically the difference between the two models and find the same similarities.



(a) Parameters $(p_1^*, \sigma_1^*, m_1^*)$ versus α_{21}/α_{12}.

(b) Parameters $(p_2^*, \sigma_2^*, m_2^*)$ versus α_{21}/α_{12}.

Figure 3.4.5: Comparison between the empirical solutions of matrix B and B^S as functions of the off-diagonal block interaction strength α_{12}/C_{12}. The left column is associated to the properties of Community 1 $(p_1^*, m_1^*, \sigma_1^*)$. The right column is associated to the properties of Community 2 $(p_2^*, m_2^*, \sigma_2^*)$. Matrix B and B^S has size $n = 100$ and the number of Monte Carlo experiments is 300. The parameters are

$$\boldsymbol{\mu} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \ , \ \boldsymbol{\alpha} = \begin{pmatrix} 2 & \alpha_{12} \\ \alpha_{12} & 1 \end{pmatrix} \ , \ \boldsymbol{C} = f(\boldsymbol{\alpha}), \ , \boldsymbol{\beta} = \left(\frac{1}{2}, \frac{1}{2}\right) .$$

3.5 Discussion

In this paper, we describe a model of the dynamics of species abundances when interaction among species is structured in multiple communities. The main interest is to outline the effect of a block structure on the stability and persistence of the species. Specifically, we describe the dynamics and properties of each community in the system (proportion of persisting species, mean and root mean square of the abundances of persisting species) and their effect on each other. We define an interaction matrix per block which has several characteristics such as the strength of the interactions $1/\boldsymbol{\alpha}$, the mean interaction trend matrix $\boldsymbol{\mu}$ and the size of the community $\boldsymbol{\beta}$. In this context, we focused most of our analysis to the case of two interacting communities. However, our results are scalable to multiple communities.

At first, we extend the feasibility results found by Bizeul and Najim [BN21] in the case of a block structure. A feasibility threshold was found in the form of an inequality that must be verified to have a feasible community set. This complements the recent results on interactions with a sparse structure [AN21] and interactions with a correlation profile [CEFN22]. We notice that to maintain the feasibility of two communities, we have to minimize interactions between the communities. Moreover, the community with weaker interactions will be able to display a larger total abundance in the ecosystem while maintaining the feasibility threshold. We extend the feasibility result in the case of a strongly mutualistic community $\boldsymbol{\mu} \geqslant \mathbf{0}$. However, for our model not to have exploding abundances, it is necessary to keep the mean interaction trend relatively low, as is already well known for mutualistic Lotka-Volterra models.

Subsequently, we studied what happens below the feasibility threshold where species can become extinct. Theoretical conditions were given for a unique globally stable equilibrium in the model (3.3) with persisting and vanishing species. This result is given by Lyapunov conditions related to a result of Takeuchi and Adachi [TA80] and random matrix theory. These stability results had been found in the case of a single community by Clenet *et al.* [CMN22]. This complement the properties of stability in the Lotka-Volterra system that has been studied by Stone [Sto18] and Gibbs *et al.* [GGRA18]. Recent random matrix methods allow us to describe the spectrum of a block matrix and plot it numerically. Furthermore, we extended the result with the addition of an interaction trend in each community. Communities with with opposite dominant interactions (e.g. mutualistic vs. competitive) are more likely to result in a unique globally stable equilibrium. Later on, we showed that whereas in the single community case where a competition trend does not affect the stability [CMN22], when there are communities have mostly competition interactions, it destabilize the system [AT12]. An antagonistic network between two communities is more stable than a mutualistic or competitive one.

In a last open sub-section, hints to give sufficient and necessary conditions to obtain a unique equilibrium are given. These conditions provided by Murty [Mur72] are related to the results on the P-matrices associated with the LCP problem.

In a third part we give heuristics on the persisting species (proportion, mean and root mean square of their abundances). These heuristics have also been found in the case of a single community by Clenet *et al.* [CMN22]. From a physicist point of view and using the methods of Bunin [Bun17] and Galla [Gal18], Barbier *et al.* [BABL18] and Poley *et al.* [PBG22] have extended the heuristics in the block and cascade model. Previously, obtaining properties on persisting species in the LV model (not normalized by $\sqrt{n}$) has

already been done by Servan *et al.* [SCG+18] where they consider a different growth rate for each species. The study of the stability and properties of persisting species in the LV system was also carried out by Pettersson *et al.* [PSNJ20, PSJ20]. From an ecological point of view, heuristics are deduced from the properties of interactions between multiple communities. A first consequence is that diversity is contagious: a feasible community has less negative effect on the persistence of the communities it interacts with. The larger the fraction of persistent species in a community, the less harmful its effect on other communities. The second is the existence of a feedback between the effects of communities on one another. Indeed, this causes a detrimental cycle between communities.

The average type of interaction in the food web plays a major role (competition, mutualism, antagonism). For example, when two communities interact mainly through mutualism or mainly through competition between them, the persistence of species in both communities ends up being the same (see Figure 3.3.9). However, for an antagonistic inter-community interaction ($\mu_{12} = -\mu_{21}$), the community benefiting from the interaction will have a larger fraction of persisting species. In the case of the type of intra-community interaction will also affect species persistence: if a community is mostly competitive, it will have less effect on other communities, while a mutualistic community will have a stronger negative effect on other communities (see Figure 3.3.10). Overall, the most resilient communities (i.e. with the highest fraction of persisting species) are those mutualistic communities which benefit from inter-communities antagonistic interactions

In a last open section, we discuss the similarities between the strength of interactions and the connectance in the Lotka-Volterra model. The general conclusion is that the same equivalence between connectance and strength of interaction is found as in the works of May [May72] on the stability-complexity threshold in the case of several communities. The analysis of a symmetric SBM network that has been studied a lot in ecology [MM17] can be done through the prism of a block model with varying interaction strength.

Many mathematical and ecological questions remain unanswered in this type of model.

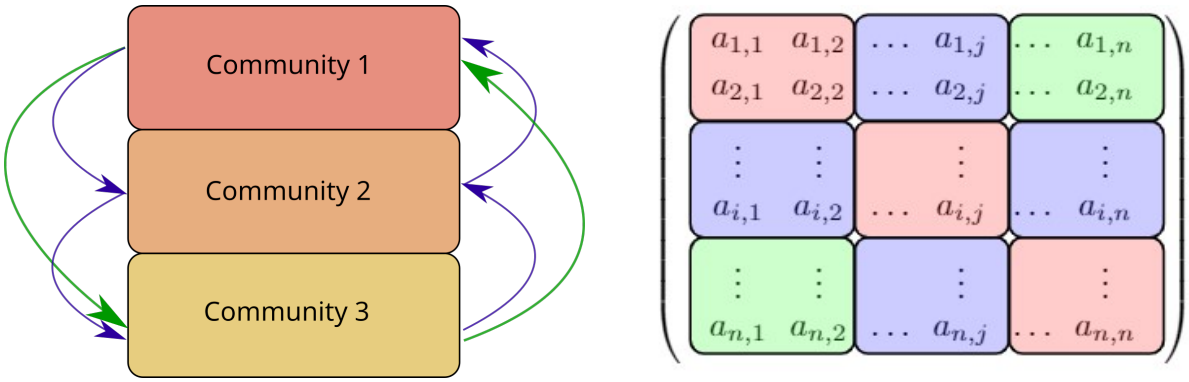
First, a rigorous mathematical proof of the heuristics presented here would be of interest although the LCP procedure induces a statistical dependence a priori difficult to handle. This issue is still pending in the single community case [CMN22] and appears to be challenging to address.

Second, we could extend the heuristics for two different scenarios. On the one hand, it would be interesting to add pairwise correlation between species. This has already been done by physicists, see [BABL18, PBG22]. In the study of feasibility, it has been shown that a correlation profile does not play a role on the feasibility threshold [CEFN22]. On the other hand, for the sake of complexity, we have chosen to set the growth rates equal to the same value $r_k = 1, \forall k \in [n]$. It would be relevant to control the distribution of the growth rate as in [SCG+18] or to consider structural stability as Saavedra *et al.* [SRB+17], i.e. by how much can the growth rates be perturbed (initially all equal to 1) without changing the type of equilibrium $\boldsymbol{x}^*$ obtained.

Two subsections of the paper are based on simulations: the extension of the conditions for a single globally stable equilibrium using the equivalence between P-matrix and the LCP problem [Mur72] and the analysis of the trade-off of May connectance versus interaction strength in the Lotka-Volterra model. In both cases, both problems are based on RMT results.

The applications in ecology are numerous on this type of model. We could consider a

spatial structure accounting for spatial proximity in the sense that two close communities tend to be more strongly connected. For example, in an aquatic environment, we could imagine the existence of an up/down gradient in a water column. In figure 3.5.1, a situation where three communities are involved is illustrated



(a) (b)

Figure 3.5.1: In (a), a representation of the gradient of interaction between three communities in a water column is represented. The blue arrows correspond to strong interaction strength due to their spatial proximity. On the opposite, the communities 1 and 3 are separated, the green arrow represents a weaker interaction. In (b), the block matrix associated with this type of model is displayed. The colors of the blocks match the colors of the arrows. The red color block corresponds to intra-community interactions.

Originally introduced by R.T. Paine [Pai66, Pai69], the concept of keystone species in ecology is widespread i.e. one species controls the coexistence of the others, if we remove it, we lose species with whom it was interacting. Mouquet *et al.* [MGMC13] suggested to extend the concept of keystone species to communities. One could analyze in the block system the existence of a keystone community that would have disproportionately large effect on other communities. In a metacommunity dynamic, Resetarits *et al.* [RCL18] have studied the keystone community concept where patches have a strong effect on other patches.

One could imagine that the same species is present several times in the system, but in different blocks, see Gravel *et al.* [GML16]. In this case, the inter-blocks represent interactions between spatially isolated communities (so should be less strong). If each diagonal or non-diagonal block is a copy of the same interaction pattern (possibly slightly perturbed) and we can add linear effect to the system to represent emigration and immigration, then we could study the feasibility properties of this system. In [GML16], they found that it works best when the dispersion is intermediate.

Last but not least, it would be relevant to compare the patterns obtained with data in ecology as in the recent article by Hu *et al.* [HAB+21] in the case of a single community.

Appendix

3.A Numerical methods

Simulations were performed in Python. All the figures and the code will be available on Github.

Simulations on the properties of persisting species are performed in two different ways. The theoretical solutions are obtained resolving numerically the system of equations of heuristics 3.1. We use a solver (cf. scipy.optimize) to find a local minimum of the function defined by the system of equations (a modification of the Powell hybrid method). The empirical solutions are computed using a Monte Carlo experiment. We simulate a large number of matrix matrix B, we resolve the associated LCP problem using the Lemke's algorithm. Then, we use the LCP solution to calculate the properties of the persisting species: proportion of survivors, etc. Finally, we make an average on the ensemble of experiments. The Lemke algorithm is implemented in the lemkelcp package and can be found on Github [Lam19]. The dynamics of the Lotka-Volterra are achieved by a Runge-Kutta of order 4 (RK4) implemented in the code.

3.B Remaining computations

3.B.1 Moments of $\check{Z}_k$

We compute hereafter the conditional mean and variance of $\check{Z}_k = (B\boldsymbol{x}^*)_k$ with respect to $\boldsymbol{x}^*$. We rely on the following identities $\forall k \in \mathcal{I}_i$, $\forall \ell \in \mathcal{I}_j$ and $\forall o \in \mathcal{I}_q$:

$$\mathbb{E}B_{k\ell} = \frac{\mu_{ij}}{n}\,, \quad \mathbb{E}(B_{k\ell}^2) = \frac{1}{\alpha_{ij}^2 n} + \frac{\mu_{ij}^2}{n^2} \simeq \frac{1}{\alpha_{ij}^2 n}\,, \quad \mathbb{E}B_{k\ell}B_{ko} = \frac{\mu_{ij}\mu_{iq}}{n^2} \quad (\ell \neq o)\,.$$

We first compute the conditional mean:

$$\begin{aligned}
\forall k \in \mathcal{I}_i,\ \mathbb{E}_{\boldsymbol{x}^*}(\check{Z}_k) &= \sum_{\ell\in[n]} \mathbb{E}(B_{k\ell})x_\ell^* = \sum_{\ell\in\mathcal{S}_1} \mathbb{E}(B_{k\ell})x_\ell^* + \sum_{\ell\in\mathcal{S}_2} \mathbb{E}(B_{k\ell})x_\ell^* \\
&= \frac{\mu_{i1}}{n}\sum_{\ell\in\mathcal{S}_1} x_\ell^* + \frac{\mu_{i2}}{n}\sum_{\ell\in\mathcal{S}_2} x_\ell^*, \\
&= \mu_{i1}\frac{|\mathcal{I}_1|}{n}\frac{|\mathcal{S}_1|}{|\mathcal{I}_1|}\frac{1}{|\mathcal{S}_1|}\sum_{\ell\in\mathcal{S}_1} x_\ell^* + \mu_{i2}\frac{|\mathcal{I}_2|}{n}\frac{|\mathcal{S}_2|}{|\mathcal{I}_2|}\frac{1}{|\mathcal{S}_2|}\sum_{\ell\in\mathcal{S}_2} x_\ell^*, \\
&= \mu_{i1}\,\beta_1\,\hat{p}_1\,\hat{m}_1 + \mu_{i2}\,\beta_2\,\hat{p}_2\,\hat{m}_2\,.
\end{aligned}$$

We now compute the second moment:

$$\begin{aligned}
\forall k \in \mathcal{I}_i,\ \mathbb{E}_{\boldsymbol{x}^*}(\check{Z}_k^2) &= \mathbb{E}_{\boldsymbol{x}^*}\left(\sum_{\ell\in[n]} B_{k\ell}x_\ell^*\right)^2 = \mathbb{E}_{\boldsymbol{x}^*}\sum_{\ell,o\in[n]} B_{k\ell}B_{ko}x_\ell^*x_o^*\,,\\
&= \sum_{\ell\in\mathcal{S}_1\cup\mathcal{S}_2}\mathbb{E}(B_{k\ell}^2)x_\ell^{*2} + \sum_{\ell\neq o}\mathbb{E}(B_{k\ell}B_{ko})x_\ell^*x_o^*,\\
&= \frac{1}{\alpha_{i1}^2 n}\sum_{\ell\in\mathcal{S}_1} x_\ell^{*2} + \frac{1}{\alpha_{i2}^2 n}\sum_{\ell\in\mathcal{S}_2} x_\ell^{*2} + \sum_{\ell\neq o}\frac{\mu_{ij}\mu_{iq}}{n^2}x_\ell^*x_o^*,\\
&\overset{(a)}{\simeq} \frac{\beta_1\hat{p}_1\hat{\sigma}_1^2}{\alpha_{i1}^2} + \frac{\beta_2\hat{p}_2\hat{\sigma}_2^2}{\alpha_{i2}^2} + \sum_{\ell,o\in\mathcal{S}_1\cup\mathcal{S}_2}\frac{\mu_{ij}\mu_{iq}}{n^2}x_\ell^*x_o^*\,,\\
&\overset{(b)}{=} \frac{\beta_1\hat{p}_1\hat{\sigma}_1^2}{\alpha_{i1}^2} + \frac{\beta_2\hat{p}_2\hat{\sigma}_2^2}{\alpha_{i2}^2} + \left(\mu_{i1}\,\beta_1\,\hat{p}_1\,\hat{m}_1 + \mu_{i2}\,\beta_2\,\hat{p}_2\,\hat{m}_2\,\right)^2\,,
\end{aligned}$$

where the approximation in (a) follows from the fact that

$$\frac{1}{n^2}\sum_{\ell,o\in[n]} x_\ell^*x_o^* = \frac{1}{n^2}\sum_{\ell\neq o} x_\ell^*x_o^* + \mathcal{O}\left(\frac{1}{n}\right)\,,$$

and (b) follows from the fact that

$$\begin{aligned}
\sum_{\ell,o\in\mathcal{S}_1\cup\mathcal{S}_2}\frac{\mu_{ij}\mu_{iq}}{n^2}x_\ell^*x_o^* &= \sum_{\ell\in\mathcal{S}_1,o\in\mathcal{S}_1}\frac{\mu_{ij}\mu_{iq}}{n^2}x_\ell^*x_o^* + \sum_{\ell\in\mathcal{S}_1,o\in\mathcal{S}_2}\frac{\mu_{ij}\mu_{iq}}{n^2}x_\ell^*x_o^*\\
&+ \sum_{\ell\in\mathcal{S}_2,o\in\mathcal{S}_1}\frac{\mu_{ij}\mu_{iq}}{n^2}x_\ell^*x_o^* + \sum_{\ell\in\mathcal{S}_2,o\in\mathcal{S}_2}\frac{\mu_{ij}\mu_{iq}}{n^2}x_\ell^*x_o^*\,,\\
&= \frac{|\mathcal{I}_1|^2}{n^2}\frac{|\mathcal{S}_1|^2}{|\mathcal{I}_1|^2}\frac{1}{|\mathcal{S}_1|^2}\mu_{i1}^2\left(\sum_{\ell\in\mathcal{S}_1}x_\ell^*\right)^2\\
&+ \frac{|\mathcal{I}_1|}{n}\frac{|\mathcal{I}_2|}{n}\frac{|\mathcal{S}_1|}{|\mathcal{I}_1|}\frac{|\mathcal{S}_2|}{|\mathcal{I}_2|}\mu_{i1}\mu_{i2}\frac{1}{|\mathcal{S}_1|}\left(\sum_{\ell\in\mathcal{S}_1}x_\ell^*\right)\frac{1}{|\mathcal{S}_2|}\left(\sum_{o\in\mathcal{S}_2}x_o^*\right) + ...\\
&= \beta_1^2\hat{p}_1^2\mu_{i1}^2\hat{m}_1^2 + 2\beta_1\beta_2\hat{p}_1\hat{p}_2\mu_{i1}\mu_{i2}\hat{m}_1\hat{m}_2 + \beta_2^2\hat{p}_2^2\mu_{i2}^2\hat{m}_2^2\,,\\
&= \left(\mu_{i1}\,\beta_1\,\hat{p}_1\,\hat{m}_1 + \mu_{i2}\,\beta_2\,\hat{p}_2\,\hat{m}_2\,\right)^2\,.
\end{aligned}$$

We can now compute the variance:

$$\forall\, k\in\mathcal{I}_i,\ \mathrm{var}_{\boldsymbol{x}^*}\left(\check{Z}_k\right) = \mathbb{E}_{\boldsymbol{x}^*}\left(\check{Z}_k^2\right) - \left(\mathbb{E}_{\boldsymbol{x}^*}\check{Z}_k\right)^2 = \frac{\beta_1\hat{p}_1\hat{\sigma}_1^2}{\alpha_{i1}^2} + \frac{\beta_2\hat{p}_2\hat{\sigma}_2^2}{\alpha_{i2}^2}\,.$$

3.B.2 Details of Heuristics 3

In the case $k\in\mathcal{S}_i$:

$$\begin{aligned}
x_k^* &= (1+\lambda_i^*+\Delta_i^*Z_k)\,\mathbf{1}_{\{Z_k>\delta_i^*\}}\\
&= (1+\lambda_i^*)\,\mathbf{1}_{\{Z_k>\delta_i^*\}} + (\Delta_i^*Z_k)\,\mathbf{1}_{\{Z_k>\delta_i^*\}}\,.
\end{aligned}$$

Taking the square, we get:

$$(x_k^*)^2 = (1+\lambda_i^*)^2 \mathbf{1}_{\{Z_k>\delta_i^*\}} = 2(1+\lambda_i^*)\Delta_i^* \mathbf{1}_{\{Z_k>\delta_i^*\}} + \left((\Delta_i^*)^2 Z_k^2\right)\mathbf{1}_{\{Z_k>\delta_i^*\}}.$$

Summing over $\mathcal{S}_i$ and normalizing, we get

$$\frac{1}{|\mathcal{S}_i|}\sum_{k\in\mathcal{S}_i}(x_k^*)^2 = (1+\lambda_i^*)^2 \frac{1}{|\mathcal{S}_i|}\sum_{k\in\mathcal{S}_i}\mathbf{1}_{\{Z_k>-\delta_i^*\}} + 2(1+\lambda_i^*)\Delta_i^*\frac{1}{|\mathcal{S}_i|}\sum_{k\in\mathcal{S}_i} Z_k \mathbf{1}_{\{Z_k>-\delta_i^*\}} + (\Delta_i^*)^2\frac{1}{|\mathcal{S}_i|}\sum_{k\in\mathcal{S}_i} Z_k^2\mathbf{1}_{\{Z_k>\delta_i^*\}}.$$

Finally, we conclude by replacing the empirical means by their limits

$$\frac{1}{|\mathcal{S}_i|}\sum_{k\in\mathcal{S}_i} Z_k^j \mathbf{1}_{\{Z_k>\delta_i^*\}} \quad = \quad \mathbb{E}(Z^j \mid Z>\delta_i^*)\,, \quad j=1,2\,,$$

and get

$$\hat{\sigma}_i^2 = (1+\lambda_i^*)^2 + 2(1+\lambda_i^*)\Delta_i^*\mathbb{E}(Z\mid Z>\delta_i^*) + \Delta_i^2\mathbb{E}(Z^2\mid Z>\delta_i^*).$$

It remains to replace $\hat{\sigma}_i$ by its limit σ_i^* to obtain (3.24)-(3.25).

3.B.3 Density of the distribution of the persistent species.

Assume that $x^*>0$, and let $f=\mathbb{R}\to\mathbb{R}$ be a bounded continuous test function. We have $\forall k\in\mathcal{S}_i$

$$\begin{aligned}
\mathbb{E}f(x_k^*) &= \mathbf{E}\left[f(1+\lambda_i^*+\Delta_i^* Z_k) \;\middle|\; Z_k>\delta_i^*\right], \\
&= \int_{-\infty}^{\infty} f(1+\lambda_i^*+\Delta_i^* u)\frac{\mathbf{1}_{\{u>\delta_i^*\}}}{1-\Phi(\delta_i^*)}\frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}}du\,, \\
&= \int_0^{\infty} f(y) e^{-\frac{1}{2}\left(\frac{y}{\Delta_i^*}+\delta_i^*\right)^2}\frac{1}{\sqrt{2\pi}\Phi(-\delta_i^*)\Delta_i^*}\,dy\,,
\end{aligned}$$

hence the density of x_k^*, $\forall k\in\mathcal{S}_i$.

3.C Extension to the b-blocks model

3.C.1 Model

Within the framework of b communities,the matrix $B=(B_{k\ell})_{n,n}$ is defined as

$$B = \frac{1}{\sqrt{n}} V \boldsymbol{s} V^\top \circ A + \frac{1}{n} V \boldsymbol{\mu} V^\top , \tag{3.30}$$

where

$$V \in \mathcal{M}_{n\times b},\ V = \begin{pmatrix} \mathbf{1}_{\mathcal{I}_1} & 0 & \cdots & 0 \\ 0 & \mathbf{1}_{\mathcal{I}_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{1}_{\mathcal{I}_b} \end{pmatrix},\ A = \begin{pmatrix} A_{11} & \cdots & A_{1b} \\ \vdots & \ddots & \vdots \\ A_{b1} & \cdots & A_{bb} \end{pmatrix},$$

$$\boldsymbol{s} = \begin{pmatrix} 1/\alpha_{11} & \cdots & 1/\alpha_{1b} \\ \vdots & \ddots & \vdots \\ 1/\alpha_{b1} & \cdots & 1/\alpha_{bb} \end{pmatrix},\ \boldsymbol{\mu} = \begin{pmatrix} \mu_{11} & \cdots & \mu_{1b} \\ \vdots & \ddots & \vdots \\ \mu_{b1} & \cdots & \mu_{bb} \end{pmatrix},$$

- $\boldsymbol{\beta} = (\beta_1, \beta_2, .., \beta_b)$, $\sum_{i=1}^b \beta_i = 1$ is the size by proportion of each of the blocks.
- $\mathcal{I}_i$ is a subset of $[n]$ of size $|\mathcal{I}_i| := \beta_i n$ matching the index of species belonging to Community i.
- $\mathbf{1}_{\mathcal{I}_i}$ is a entry wise vector of 1 of size $\beta_i n$.
- A_{ij} is a non-Hermitian random matrix of size $(\beta_i n, \beta_j n)$ with reduced centered Gaussian entries i.e. $\mathcal{N}(0,1)$.

3.C.2 Feasibility

We consider a growing scaling matrix

$$\boldsymbol{s}_n \xrightarrow[n\to\infty]{} \mathbf{0} \quad \Leftrightarrow \quad \forall i,j \in \{1,b\}, \alpha_{ij} \xrightarrow[n\to\infty]{} \infty .$$

Let B_n a matrix defined by

$$B_n = V \boldsymbol{s}_n V^\top \circ \frac{1}{\sqrt{n}} A . \tag{3.31}$$

The spectral radius of $\frac{1}{\sqrt{n}}A$ a.s. converges to 1. So as long as $\boldsymbol{s}_n$ is close to zero, the matrix $I - B_n$ is eventually invertible.

Theorem 3.8 (Feasibility for the b-blocks model)**.** *Assume that matrix B_n is defined by the b-blocks model* (3.31)*, $\boldsymbol{\mu} = \mathbf{0}$. Let $\boldsymbol{\beta} = (\beta_1, \beta_2, .., \beta_b)$, $\sum_{i=1}^b \beta_i = 1$ represents the proportion of each community. Let $\boldsymbol{s}_n \xrightarrow[n\to\infty]{} 0$ and denote by $s_n^* = 1/\sqrt{2\log n}$. Let $x_n = (x_k)_{k\in[n]}$ be the solution of* (3.6)*.*

1. *If there exists $\varepsilon > 0$ such that eventually $\left\|(\boldsymbol{s_n} \circ \boldsymbol{s_n})\boldsymbol{\beta}^\top\right\|_\infty \geqslant (1+\varepsilon)(s_n^*)^2$ then*

$$\mathbb{P}\left\{\min_{k\in[n]} x_k > 0\right\} \xrightarrow[n\to\infty]{} 0 .$$

2. *If there exists $\varepsilon > 0$ such that eventually $\left\|(\boldsymbol{s_n} \circ \boldsymbol{s_n})\boldsymbol{\beta}^\top\right\|_\infty \leqslant (1-\varepsilon)(s_n^*)^2$ then*

$$\mathbb{P}\left\{\min_{k\in[n]} x_k > 0\right\} \xrightarrow[n\to\infty]{} 1 .$$

Sketch of proof. Starting from the decomposition the equilibrium $\boldsymbol{x}^*$:

$$x_k^* = 1 + Z_k + R_k\,,$$

where $Z_k = \sum_{\ell=1}^{n} B_{k\ell}$, $\forall k \in [n]$ and we suppose $\forall k \in [n]$, R_k is a negligible term if n is sufficiently large.

The feasibility of the b communities is studied independently. Using Gaussian addition properties, a simpler form of Z_k is deduced. Consider a family $(\check{Z}_k)_{k\in[n]}$ of i.i.d. random variables $\mathcal{N}(0,1)$.

$$\begin{aligned}
\text{If } k \in \mathcal{I}_i,\ Z_k &= \sum_{j=1}^{b} \sum_{\ell\in\mathcal{I}_j} B_{k\ell}\ , \\
&\sim \sum_{j=1}^{b} \mathcal{N}\left(0\,,\ \frac{\beta_j}{\alpha_{ij}^2}\right)\ , \\
&\sim \sqrt{\sum_{j=1}^{b} \frac{\beta_j}{\alpha_{ij}^2}}\check{Z}_k\,.
\end{aligned}$$

Given $\boldsymbol{\beta} = (\beta_1, \beta_2, .., \beta_b)$, conditions on the matrix $\boldsymbol{\alpha}$ are inferred to have

$$\mathbb{P}(\min_{k\in[n]} x_k > 0) = 1 \Leftrightarrow \mathbb{P}(\min_{k\in[n]} Z_k > -1) = 1\,.$$

In order to compute a tractable form of $\min\limits_{k\in[n]} Z_k$, an additional approximation is made, if n is large enough

$$\min_{k\in\mathcal{I}_i} \check{Z}_k \sim -\sqrt{2\log(\beta_i n)} \simeq -\sqrt{2\log(n)}\,. \tag{3.32}$$

$$\begin{aligned}
\min_{k\in[n]} Z_k &= \min_{i\in[b]} \left(\sqrt{\sum_{j=1}^{b} \frac{\beta_j}{\alpha_{ij}^2}} \min_{k\in\mathcal{I}_i} \check{Z}_k \right), \\
&\simeq \min_{i\in[b]} \left(\sqrt{\sum_{j=1}^{b} \frac{\beta_j}{\alpha_{ij}^2}} \left(-\sqrt{2\log(n)}\right) \right), \\
&= \min_{i\in[b]} \left(-\sqrt{\sum_{j=1}^{b} \frac{2\beta_j \log(n)}{\alpha_{ij}^2}} \right), \\
&= -\max_{i\in[b]} \left(\sqrt{\sum_{j=1}^{b} \frac{2\beta_j \log(n)}{\alpha_{ij}^2}} \right).
\end{aligned}$$

Following the approximation (3.32), the condition $\min\limits_{k\in[n]} Z_k > -1$ asymptotically boils

down to

$$\begin{aligned}
&\max_{i\in[b]}\left(\sqrt{\sum_{j=1}^{b}\frac{2\beta_j\log(n)}{\alpha_{io}^2}}\right)<1\,,\\
\Leftrightarrow\quad&\max_{i\in[b]}\left(\sum_{j=1}^{b}\frac{2\beta_j\log(n)}{\alpha_{ij}^2}\right)<1\,,\\
\Leftrightarrow\quad&\max_{i\in[b]}\left(\sum_{j=1}^{b}\frac{\beta_j}{\alpha_{ij}^2}\right)<\frac{1}{2\log(n)}\,,\\
\Leftrightarrow\quad&\left\|(\boldsymbol{s_n}\circ\boldsymbol{s_n})^2\boldsymbol{\beta}^\top\right\|_\infty<\frac{1}{2\log(n)}:=(\check{\alpha}_n^*)^2\,.
\end{aligned}$$

□

Notice that the non-centered case can be treated in the same principle for several communities.

3.C.3 Existence of a unique equilibrium

3.C.3.1 Centered case

Denote by H symmetric the matrix

$$H=\begin{pmatrix}H_{11}&\cdots&H_{1b}\\\vdots&\ddots&\vdots\\H_{b1}&\cdots&H_{bb}\end{pmatrix},$$

where $\forall\, i,j\in[b]$, H_{ij} is a matrix of size $\beta_i n\times\beta_j n$ and each off-diagonal entries follow a Gaussian distribution $\mathcal{N}(0,1/\alpha_{ij}^2+1/\alpha_{ji}^2)$.

The Quadratic Vector Equation (QVE) associated to the matrix H is decomposed as

$$k\in\mathcal{I}_i,\ -\frac{1}{m_k(z)}=z+\sum_{j=1}^{b}\sum_{\ell\in\mathcal{I}_j}\frac{1}{n}\left(\frac{1}{\alpha_{ij}^2}+\frac{1}{\alpha_{ji}^2}\right)m_\ell(z)\,.$$

Given $\boldsymbol{m}(z)=(m_1(z),\cdots,m_n(z))$, denote by $1/\boldsymbol{m}(z)=(1/m_1(z),\cdots,1/m_n(z))$ and $S=\frac{1}{n}V(\boldsymbol{s}+\boldsymbol{s}^\top)V^\top$ the QVE can be written in the standard form

$$-\frac{1}{\boldsymbol{m}(z)}=z+S\boldsymbol{m}(z)\,. \tag{3.33}$$

Following Theorem 2.1 in Ajanki *et al.* [AEK19], $\forall z\in\mathbb{C}$, Equation (3.33) has a unique solution $\boldsymbol{m}=\boldsymbol{m}(z)$ and the support of the associated measure is included in $[-\Sigma,\Sigma]$, $\Sigma=2\left\|S\right\|_2^{1/2}$. This information gives a bound to the support of the matrix H associated with (3.33). In particular, $\lambda_{\max}(H)\leqslant 2\left\|S\right\|_2^{1/2}$. Recall that $-2I+H$ is negative definite if $\lambda_{\max}(H)\leqslant 2$. The final condition, which is non-sharp, to have a unique globally stable equilibrium is

$$2\left\|S\right\|_2^{1/2}<2\ \Leftrightarrow\ \left\|S\right\|_2<1\,,$$

Given the particular shape of the matrix S, computing its norm is equivalent to computing the norm of a matrix of size b

$$\|S\|_2 = \left\|\mathrm{diag}(\boldsymbol{\beta})^{1/2}\left((\boldsymbol{s}\circ\boldsymbol{s}) + (\boldsymbol{s}\circ\boldsymbol{s})^\top\right)\mathrm{diag}(\boldsymbol{\beta})^{1/2}\right\|_2 .$$

which completes the proof: we can then rely on Theorem 3.4 to conclude.

3.C.4 Persisting species

Let $(\boldsymbol{\alpha}, \boldsymbol{\mu})$ satisfying the condition for $\boldsymbol{x}^*$ to be a unique stable equilibrium. The following system of 3b equations and 3b unknowns $\boldsymbol{p} = (p_1, p_2, .., p_b)$, $\boldsymbol{m} = (m_1, m_2, .., m_b)$, $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, .., \sigma_b)$

$$\begin{aligned}
\forall i \in [b],\ p_i &= 1 - \Phi(\delta_i) ,\\
\forall i \in [b],\ m_i &= 1 + \lambda_i + \Delta_i \mathbb{E}(Z|Z > \delta_i) ,\\
\forall i \in [b],\ (\sigma_i)^2 &= (1+\lambda_i)^2 + 2(1+\lambda_i)\Delta_i \mathbb{E}(Z|Z>\delta_i) + \Delta_i^2 \mathbb{E}(Z^2|Z>\delta_i) ,
\end{aligned}$$

where

$$\Delta_i = \sqrt{\sum_{j=1}^{b} p_j (\sigma_j)^2 \frac{\beta_j}{\alpha_{ij}^2}} \ ;\ \lambda_i = \sum_{j=1}^{b} p_j m_j \beta_j \mu_{ij} \ ;\ \delta_i = \frac{-1-\lambda_i}{\Delta_i},$$

admits a unique solution $(\boldsymbol{p}^*, \boldsymbol{m}^*, \boldsymbol{\sigma}^*)$ and $\forall\ i \in [b]$

$$\hat{p}_i \xrightarrow[n\to\infty]{a.s.} p_i^* , \qquad \hat{m}_i \xrightarrow[n\to\infty]{a.s.} m_i^* \qquad \text{and} \qquad \hat{\sigma}_i \xrightarrow[n\to\infty]{a.s.} \sigma_i^* .$$

3.C.5 Distribution of the persisting species

Let $(\boldsymbol{\alpha}, \boldsymbol{\mu})$ satisfying the condition for $\boldsymbol{x}^*$ to be a unique stable equilibrium. $\boldsymbol{x}^*$ the solution of (3.14) and let $(\boldsymbol{p}^*, \boldsymbol{m}^*, \boldsymbol{\sigma}^*)$ the solution of the Heuristics. Recall the definition (3.26) of λ_i, Δ_i, δ_i and denote by $\delta_i^* = \delta_i(p_i^*, m_i^*, \sigma_i^*)$. Let $x_k^* > 0$ a positive component of $\boldsymbol{x}^*$ belonging to the community i, then:

$$\mathcal{L}(x_k^*) \xrightarrow[n\to\infty]{} \mathcal{L}\left(1 + \lambda_i^* + \Delta_i^* Z \ \middle|\ Z > \delta_i^*\right) ,$$

where $Z \sim \mathcal{N}(0,1)$. Otherwise stated, asymptotically $\forall k \in \mathcal{S}_i$, x_k^* admits the following density

$$f_k(y) = \frac{\mathbf{1}_{\{y>0\}}}{\Phi(-\delta_i^*)} \frac{1}{\Delta_i^*\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{y}{\Delta_i^*} + \delta_i^*\right)^2\right\} . \tag{3.34}$$

3.D Necessary and sufficient condition of P-matrix

As a reminder, Murty [Mur72] states that the $LCP(I-B, -\mathbf{1})$ has a unique solution iff $I - B$ is a P-matrix. The condition given in Theorem 3.6 is sufficient to guarantee a unique solution to $LCP(I-B, \mathbf{1})$ but not necessary, although it provides more information and guarantees the global stability. This condition might be relaxed finding the bound associated to the P-matrix property of $I - B$. In this section, we present, from a heuristic

point of view, the information on the potential bound to have the existence of a unique equilibrium to the LCP problem.

First, recall the definition of P-matrix with its equivalent given by Fiedler and Pták [FP66].

Theorem 3.9 (Theorem 1.3 [FP66]). *The following properties of a square matrix M are equivalent:*

1. *All principal minors of M are positive*

$$\det(M_{\mathcal{I}}) > 0\ , \quad \forall\, \mathcal{I} \subset [n]\ , \quad M_{\mathcal{I}} = (M_{k\ell})_{k,\ell \in \mathcal{I}}\,;$$

2. *All real eigenvalues of M and its principal submatrices are positive.*

The main issue is to understand the spectrum of the centered matrix B i.e. $\boldsymbol{\mu} = \mathbf{0}$. According to the Theorem 3.9, one has the implication

$$\sup_{\mathcal{I}\subset[n]} \rho(B^{\mathcal{I}}) < 1 \Rightarrow \forall\, \mathcal{I} \subset [n],\ \det\left(I - B^{\mathcal{I}}\right) > 0$$

where $B^{\mathcal{I}}$ is the main sub-matrix of index $\mathcal{I} \times \mathcal{I}$. If we conjecture that the spectral radius of any main sub-matrix of B is smaller than the radius of B i.e.

$$\sup_{\mathcal{I}\subset[n]} \rho(B^{\mathcal{I}}) < \rho(B)\,,$$

then in the single community case i.e. $\boldsymbol{\alpha} = \alpha \mathbf{1}\mathbf{1}^{\top}$, one has $\rho(B) < 1 \Leftrightarrow \ \alpha > 1$. To conclude $\alpha > 1$ implies the existence of a unique equilibrium to the LCP problem. This sufficient condition is possibly necessary.

When considering many communities, an open question is the conjecture on the existence of a similar bound depending on the parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

Given $V^{\mathcal{I}}$ a variance profile of matrix B associated to $\mathcal{I} \subset [n]$. In the case of two communities, $V^{\mathcal{I}}$ is a main sub-matrix of

$$V = \frac{1}{n}\begin{pmatrix} \frac{1}{\alpha_{11}^2}\mathbf{1}_{\mathcal{I}_1}\mathbf{1}_{\mathcal{I}_1}^T & \frac{1}{\alpha_{12}^2}\mathbf{1}_{\mathcal{I}_1}\mathbf{1}_{\mathcal{I}_2}^T \\ \frac{1}{\alpha_{21}^2}\mathbf{1}_{\mathcal{I}_2}\mathbf{1}_{\mathcal{I}_1}^T & \frac{1}{\alpha_{22}^2}\mathbf{1}_{\mathcal{I}_2}\mathbf{1}_{\mathcal{I}_2}^T \end{pmatrix}.$$

By RMT properties, Najim *et al.* [CHNR21] prove the convergence of the spectral measure for a non-Hermitian matrix with variance profile. The convergence of the spectral radius of B is not established but is expected. The spectral radius of B is computed from the variance profile V

$$\rho(B) = \rho(V) = \rho\left(\begin{pmatrix}\beta_1 & 0\\ 0 & \beta_2\end{pmatrix}^{1/2}\begin{pmatrix}\frac{1}{\alpha_{11}^2} & \frac{1}{\alpha_{12}^2}\\ \frac{1}{\alpha_{21}^2} & \frac{1}{\alpha_{22}^2}\end{pmatrix}\begin{pmatrix}\beta_1 & 0\\ 0 & \beta_2\end{pmatrix}^{1/2}\right).$$

Finally, if we establish the same conjecture as in the single community case, we would have

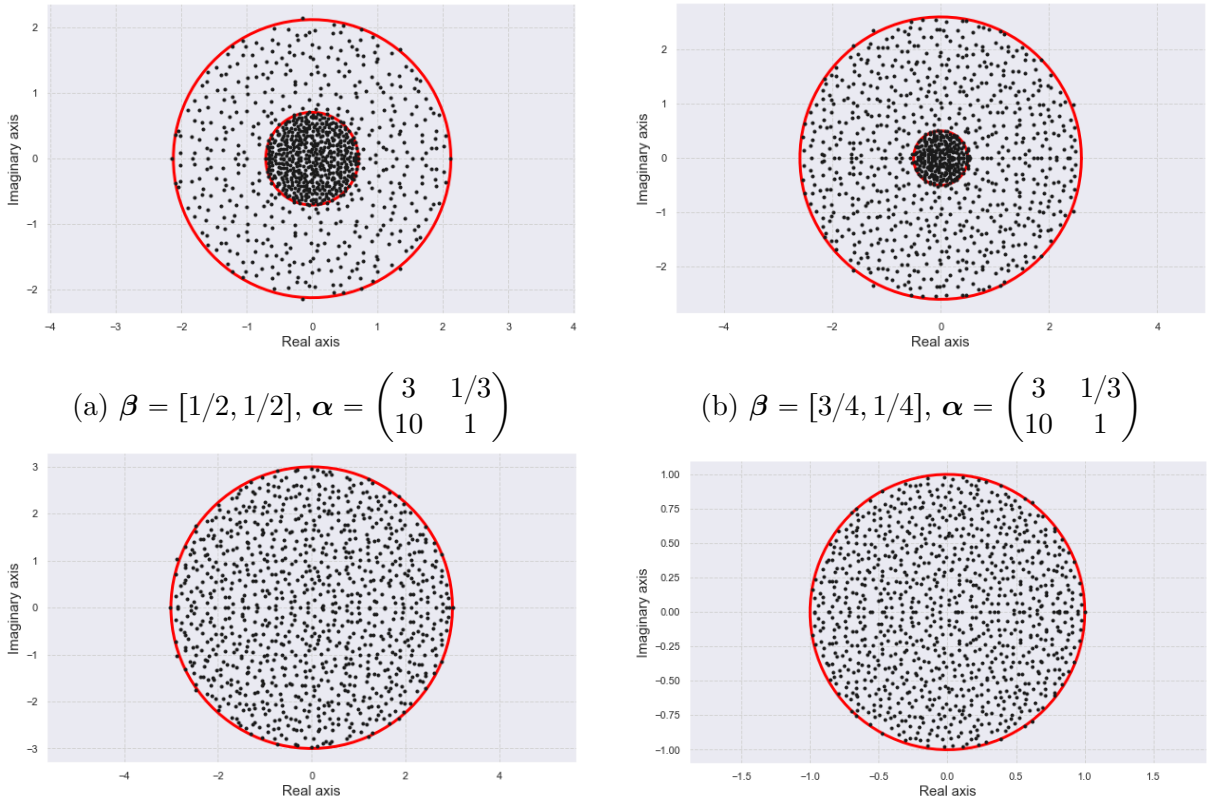
$$\sup_{I\subset[n]} \rho(V^I) < \rho(V) < 1\,.$$

To conclude $\rho(V) < 1$ is a sufficient (and possibly necessary) condition to the existence of a unique equilibrium to the LCP problem.

Remark 3.15. The spectral radius of matrix V

$$\left(\begin{pmatrix}\beta_1 & 0\\ 0 & \beta_2\end{pmatrix}^{1/2}\begin{pmatrix}\frac{1}{\alpha_{11}^2} & \frac{1}{\alpha_{12}^2}\\ \frac{1}{\alpha_{21}^2} & \frac{1}{\alpha_{22}^2}\end{pmatrix}\begin{pmatrix}\beta_1 & 0\\ 0 & \beta_2\end{pmatrix}^{1/2}\right)$$

represents the radius of the disc which supports the limiting eigenvalue distribution. The module of the second eigenvalue is the radius of an inner disc where we can witness higher concentration of the eigenvalues, see Figure 3.D.1.



(a) $\boldsymbol{\beta} = [1/2, 1/2]$, $\boldsymbol{\alpha} = \begin{pmatrix}3 & 1/3\\ 10 & 1\end{pmatrix}$

(b) $\boldsymbol{\beta} = [3/4, 1/4]$, $\boldsymbol{\alpha} = \begin{pmatrix}3 & 1/3\\ 10 & 1\end{pmatrix}$

(c) $\boldsymbol{\beta} = [1, 0]$, $\boldsymbol{\alpha} = \begin{pmatrix}3 & 1/3\\ 10 & 1\end{pmatrix}$

(d) $\boldsymbol{\beta} = [0, 1]$, $\boldsymbol{\alpha} = \begin{pmatrix}3 & 1/3\\ 10 & 1\end{pmatrix}$

Figure 3.D.1: Spectrum of non-Hermitian matrix B in the complex plan ($n = 1000$), see caption in each sub-figure for the setting of $(\boldsymbol{\beta}, \boldsymbol{\alpha})$. The solid line circle represents the boundary of the circular law and concentration circles associated to the eigenvalues of the variance profile matrix V. In plot (c)-(d), the conditions of the circular law are met.

Remark 3.16. A comparison with the bound when there is a single community shows that the same result is obtained. The spectrum of the matrix of rank 1

$$\frac{1}{2\alpha^2}\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix},$$

has two eigenvalues 0 and $\frac{1}{\alpha^2}$. The condition for a single-community is recovered $\frac{1}{\alpha^2} < 1 \Leftrightarrow 1 < \alpha$.

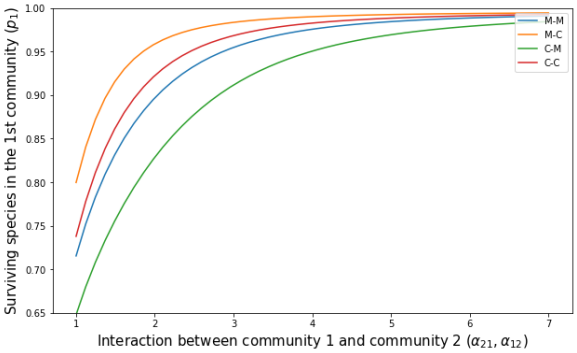
3.E Additional graphs: type of food web interactions

Here we summarize with 6 graphs (see Figure 3.E.1) the effects on the persistence of each community according to the type of interactions. For this purpose we define 4 disturbance matrices

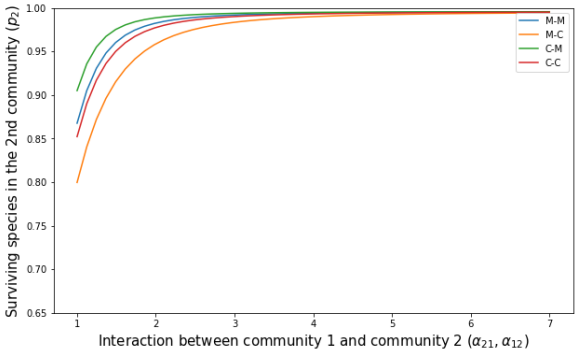
$$\boldsymbol{\mu_{MM}} = \begin{pmatrix} 0.4 & \mu_{12} \\ \mu_{21} & 0.4 \end{pmatrix}, \ \boldsymbol{\mu_{MC}} = \begin{pmatrix} 0.4 & \mu_{12} \\ \mu_{21} & -0.4 \end{pmatrix},$$

$$\boldsymbol{\mu_{CM}} = \begin{pmatrix} -0.4 & \mu_{12} \\ \mu_{21} & 0.4 \end{pmatrix}, \ , \boldsymbol{\mu_{CC}} = \begin{pmatrix} -0.4 & \mu_{12} \\ \mu_{21} & 0.4 \end{pmatrix}.$$

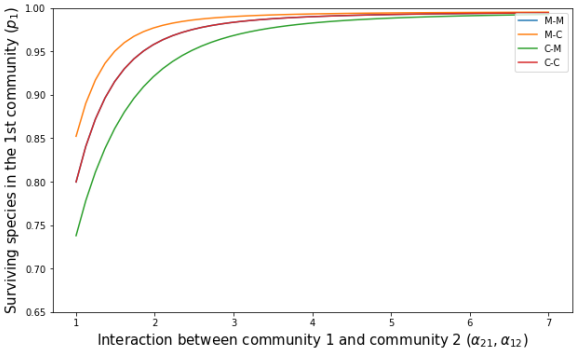
For each of these matrices, an analysis of the antagonistic ($\mu_{12} = -\mu_{21} = 0.4$), mutualistic ($\mu_{12} = \mu_{21} = 0.4$) and competitive ($\mu_{12} = \mu_{21} = -0.4$) case is performed.



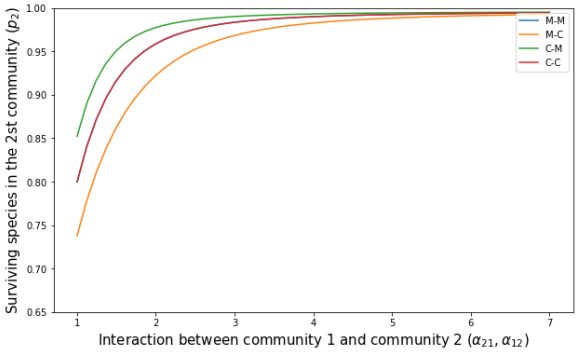
(a) Persistence of Community 1 (p_1), antagonistic ($\mu_{12} = -\mu_{21} = 0.4$).



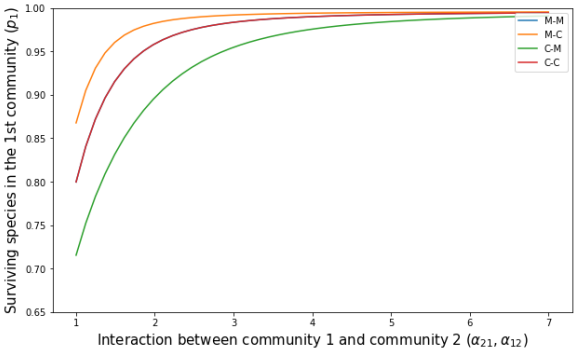
(b) Persistence of Community 2 (p_1), antagonistic ($\mu_{12} = -\mu_{21} = 0.4$).



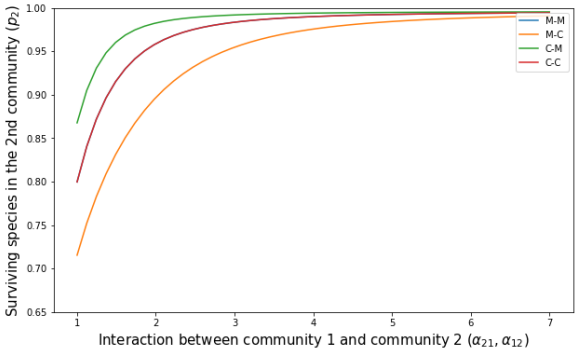
(c) Persistence of Community 1 (p_1), mutualistic ($\mu_{12} = \mu_{21} = 0.4$).



(d) Persistence of Community 2 (p_2), mutualistic ($\mu_{12} = \mu_{21} = 0.4$).



(e) Persistence of Community 1 (p_1), competitive ($\mu_{12} = \mu_{21} = -0.4$).



(f) Persistence of Community 2 (p_1), competitive ($\mu_{12} = \mu_{21} = -0.4$).

Figure 3.E.1: Representation of the effect of the type of foodweb interactions. Each plot illustrate the proportion of the persisting species as a function of the interaction strength between Community 1 and Community 2.

Chapter 4

A probabilistic perspective of the hierarchical competition-colonization trade-off model

Abstract

Introduced and analyzed by Tilman, the hierarchical competition-colonization trade-off model represents a system of species competing for a set of habitat patches. In this model, the competition is hierarchical: the dynamics of each species' occupancy within the metacommunity depends on its colonization and extinction rates. For a better understanding of the restrictions induced on the colonization rate in a large-dimension system, we propose a probabilistic interpretation of the model by looking at colonization parameters following a given probability distribution. Our aim is to determine the distribution maximizing the coexistence between the species. Based on this information, we can assess species occupancies and characterize the assembly process of the ecosystem. To answer this question, we first carried out analytical and simulation-based work to investigate the optimal distribution, persistence and stability. Second, we analyzed two different types of assembly processes: a "all-at-once approach" starting from a pool of species by letting the dynamics elapse, and a "invasion sequence approach" developing an invasion sequence that involves a historical contingency effect.

The hierarchical competition-colonization trade-off model represents a first step in our understanding of species-rich metacommunities. From a mathematical point of view, we find information on the stability and persistence of the model allowing the whole or a sub-population of species to coexist. We continue this investigation by providing insights into the shape of the distribution of the colonization rate. On the one hand, for a wide range of distributions, we found that if species are thrown together all at once, on average one in two species persists indefinitely. On the other hand, when species try to invade in a random sequence, the heavier the distribution "tail", the higher the probability of coexistence. Subsequently, the comparison of the two assembly processes shows us that the invasion sequence approach seems to be much more restrictive in terms of the number of persisting species due to historical contingencies and extinction cascades.

To conclude, this probabilistic perspective of the hierarchical competition-colonization trade-off model allows to put forward and compare two different types of distinct assemblages and gives conditions for many species to coexist under the competition-colonization trade-off.

Introduction

Motivations

Understanding the functioning of large ecosystems represents an important challenge in theoretical ecology. The complexity of these systems makes it necessary to use mathematical modeling. Regarding the Lotka-Volterra model [Lot25, Vol26], early mathematical models did not take into account the spatial structure of habitats to understand the mechanisms underlying species coexistence. However, under the impulse of the research of MacArthur [Mac84] in population biology and geography and the formulation of the theory of island biodiversity by MacArthur and Wilson in 1967 [MW67], the consideration of spatial dynamics became a major issue in ecology. Subsequently, a theory derived from different concepts (patch dynamics, species sorting, mass effect, neutral theory) emerged: the metacommunity. More recently, the metacommunity framework has subsumed several concepts, notably niche theory, ecological filtering, dispersal limitation, patch dynamics and the importance of stochasticity, under a unified theoretical umbrella, with the aim of pushing the field further in the direction of scientific, refutable community ecology (see Leibold *et al.* [LHM+04] for a review).

One of the most popular model of spatially structured environment is the competition-colonization (C-C) trade-off model. It belongs to the class of patch dynamics models where the interest is in modeling the occupancy of a species in the landscape and not its abundance. The considerable advantage of this model is its simplicity, allowing for both theoretical and empirical research. The origin of this model goes back to the work of Levins *et al.* [Lev69, LC71] who introduced a simple model of colonization and extinction dynamics. This model assumes a patch-occupancy where the main driver is the ability of the population to disperse between different patches while being subject to extinction. This metapopulation model has subsequently been extended and integrated into many empirical and theoretical works (see Hanski [Han99] for a review). Later on, Levins' model has been extended to a multiple species framework. The C-C trade-off model has emerged as an important object where the main idea is to keep few parameters, but to find a simple rule to make species coexist, the competition-colonization trade-off.

In a first step, this model was studied in the framework of a hierarchical competition i.e. the most competitive species is the worst colonizer. In early work, Hastings [Has80] was interested in studying disturbance in the model where he demonstrated the so-called "Intermediate Disturbance Hypothesis". Then, Nee and May [NM92] added habitat destruction to the model. Further research was carried out for two competing species by Hanski [Han83] and for a general form of the hierarchical competition-colonization trade-off model by Tilman [Til94] and including the impact of habitat destruction by Tilman *et al.* [TMLN94]. Kinzig *et al.* [KLD+99] finally analyze the high diversity limit in the hierarchical C-C model.

In the case of non-hierarchical competition, the classical paradigm between competition and colonization was studied by Amarasekare *et al.* [Ama03] and Yu and Wilson [YW01] in the replacement competition case. Subsequent analyses are also found in [PR98]. Calcagno *et al.* [CMJD06a] gave a new impulse in response to Yu and Wilson by taking into account competition and a preemption parameter. Empirical research has also been conducted by Cadotte [Cad07] to study the intermediate disturbance hypothesis (IDH) related to the C-C trade-off model. In fact, the IDH is not supposed to be a metapopulation scale assumption but rather a local scale assumption. Fox [Fox13] has

argued extensively that this assumption is shown (in a mathematical sense) not to be possible i.e. no model of local dynamics predicts the IDH under realistic assumptions. This type of model has been extended in other fields than ecology, in particular on host-parasite interactions by May and Nowak [MN94, NM94, MN95].

More recently, the C-C model is still the subject of considerable attention with extensions for higher-order competition [LBL20], dispersal network [ZBN+21], multimodality in diversity–disturbance relationships [ML02, LBB22].

In this article and to clarify the restrictions induced on the colonization rate in a large-dimension system, we propose a probabilistic interpretation of the hierarchical C-C model. We consider communities that emerge through the model dynamics from an initial random pool where the colonization parameters are sampled from a specific probability distribution. A link can be formally established between the C-C model and the Lotka-Volterra model (see appendix of [CMJD06a]) providing a new proof of the global stability of the system. This probabilistic approach aims at understanding the distribution and characteristic outcomes that maximizes coexistence between species. Surprisingly, a universality result appears for the distribution of number of persisting species from a pool of many species. Furthermore, we challenge this universality result when communities are assembled one-at-a-time from a regional pool [SA21]. This different assembly process shows different dynamics properties going from linear to logarithmic growth. Patterns of competition colonization trade-off appears showing the importance of finding a balance between competitors and colonizers.

Model

Commonly called the competition-colonization trade-off model, it is an extension of the well-known Levins [Lev69] model to n-species. Its most general form is provided in Calcagno *et al.* [CMJD06a]

$$\begin{aligned}\frac{dp_i(t)}{dt} = c_i p_i(t)\left(1-\sum_{j=1}^{n} p_j(t)\right) - m_i p_i(t) \\ + c_i p_i(t)\sum_{j\neq i} p_j(t)\eta_{ij} - p_i(t)\sum_{j\neq i} c_j p_j(t)\eta_{ji}\,, \forall\, i \in [n]\,,\end{aligned} \tag{4.1}$$

where $p_i(t) \in [0,1]$ represents the occupancy of species i at a given time t, m_i is the extinction rate of species i, c_i represents the colonization rate of species i, η_{ij} corresponds to the probability that species i takes over a patch already occupied by species j if it lands there.

The first term of the equation represents the colonization of species i into empty patches, the second term of the equation represents the extinction of patches where species i is currently present. The third term corresponds to the colonization of species i into patches occupied by other species, the weakness of species j present in a patch is represented by the competition term η_{ij}. Conversely, the last term of the equation represents the patches of species i that are subject to colonization by other species. The species i is weakened by species j according to the competition factor η_{ji}.

In this paper, we focus on its simplified version which corresponds to Hastings' and Tilman's version where we assume that the competition is hierarchical (referred as the

HT model). The HT model represents a system of species competing for a set of habitats (patch). It is a metacommunity model where competition is hierarchical: the dynamics of each species within the population depends only on its colonization rate (c) and its extinction rate (m).

Hastings [Has80] was concerned about the effects of disturbance $m_i = m, \forall\, i \in [n]$ on species richness and considered the extinction rate of each species the same. The HT model in its more similar form to the C-C model was introduced by Tilman and represented as

$$\frac{dp_i(t)}{dt} = c_i p_i(t)\left(1 - \sum_{j=1}^{i} p_j(t)\right) - m_i p_i(t) - \left(\sum_{j=1}^{i-1} c_j p_j(t) p_i(t)\right), \forall\, i \in [n]\,, \tag{4.2}$$

where $p_i(t) \in [0,1]$ corresponds to the occupancy of species i at a given time t, $c_i \in \mathbb{R}_+^*$ corresponds to the colonization rate of the species i where species are arranged in increasing order of c i.e. $c_1 < \cdots < c_n$, $m_i \in \mathbb{R}_+$ corresponds to the extinction rate of species i and

$$\eta_{ij} = \begin{cases} 1 & \text{if } i < j\,, \\ 0 & \text{otherwise.} \end{cases}$$

The term η_{ij} reflects the additional requirement intrinsically related to the model: the establishment of a competitive hierarchy between species. The main idea consists in having the first species as the best competitor but the poorest colonizer. Each species added to the system must be a better colonizer and a poorer competitor. The key idea to have a stable coexistence of a maximum number of species is that high colonization rates could offset the competition due to other species by invading the patch (sites) that are not occupied.

Denote by $\boldsymbol{c} = (c_1, ..., c_n)^\top$ the vector of colonization rates. Without loss of generality, it is assumed that the colonization rates are sorted in increasing order i.e.

$$c_1 < c_2 < ... < c_n\,.$$

This means that species 1 is the most competitive and species n the least competitive. From a probability theory viewpoint, a sorted independent identically distributed (i.i.d.) random sample from a continuous distribution is an order statistic of a statistical sample. For standard properties of order statistics see Appendix 4.B.2 (for a more detailed review, see Arnold *et al.* [ABN08]).

The purpose of this article is to improve our understanding of spatially structured communities comprising many species. However, when the number of species becomes very large, it is challenging and costly to collect data. In general, it is easier to have an estimate of the ratio c/m for a Levins model (on the one hand, it is linked to the equilibrium, on the other hand, if we can estimate a number of colonization events, we will necessarily relate it to a number of extinctions to be able to make something of it). Here, we suppose that the colonization rates $\boldsymbol{c}$ follow a certain probability distribution with continuous positive density. From an empirical point of view, this is justified insofar as we wish to know if there are distributions independent of the exact values of the colonization parameters. We note Bertrand's paradox on distributions that Calcagno

et al. [CMJD06b] take up in their answer to Adler [Adl06]. The thinking process is similar to May's seminal paper [May72] on the community matrix where he assumes that around the equilibrium point the effect of one species on another has a certain statistical distribution. and to the numerous works that followed on the effect of different distributions. Subsequently, many works have followed to understand the effect of the diverse distributions, in particular Allesina *et al.* [AT12, AT15, GRA16].

Many questions arise as follow-up:

- Is it possible to find a statistical distribution that maximizes the coexistence between the species?
- Given a certain distribution, can we predict the expected number of coexisting species?
- Given the statistical distribution maximizing the coexistence, what is the occupancy of the persistent species?

Notation

See Appendix 4.B for more details on the probability distribution.

- Uniform distribution of support $[1, 2]$ is $\mathcal{U}([1, 2])$ or 'Uniform 1'.
- Uniform distribution of support $[1, c_{\max}]$ is $\mathcal{U}([1, c_{\max})$ or 'Uniform $c_{\max} - 1$'.
- Pareto distribution of support $[1, \infty)$ with parameter a is $\mathcal{P}(a)$ or 'Pareto a'.
- Exponential distribution of support $[1, \infty)$ with parameter λ is $\mathcal{E}(\lambda)$ or 'Exponential λ'.
- Log-Cauchy distribution of support $[0, \infty)$ with parameter μ, σ is $LC(0, 1)$ or 'Log-Cauchy'.

4.1 Dynamics and connection with the Lotka-Volterra model

4.1.1 Set of admissible solutions

At this point, without concern for the dynamics of the system (4.2), many properties of the model can be obtained in a simple form [Til94]. In detail, equilibrium occupancy of each species $\boldsymbol{p}^* = (p_1^*, ..., p_n^*)$ can be computed with iterative equations. Given $\boldsymbol{p}^*$, one can deduce the proportion of empty space when i species are present

$$h_i = 1 - \sum_{j=1}^{i} p_j^* \,.$$

The results obtained by Tilman for the occupancies, empty space and colonization rate are recalled in section 4.A.1.

Assume the extinction rates are all equal i.e. $\forall i \in [n]$, $m_i = m$. At this point, the conditions for species coexistence only relate to the choice of the parameters $\boldsymbol{c}$. The

invading condition of a species depends only on its colonization rate and the state of the system. At equilibrium with n species present in the system, a species invades and persists if and only if

$$c_{n+1} > \frac{m}{h_n^2}\,,$$

where c_{n+1} corresponds to the colonization rate of the invading species and h_n to the proportion of empty patches before invasion. This condition forces the new species to be a better colonizer than the previous ones. These conditions can be written for all species and make up the admissible set for the colonization rate.

An example of a two-species system Consider the two-species version of the model (4.2), species 1 is the most competitive, species 2 the least competitive. Assume that both species have the same extinction rate m. The dynamics is governed by a two-equation system:

$$\begin{aligned}\frac{dp_1(t)}{dt} &= c_1 p_1(t)(1-p_1(t)) - m p_1(t) = p_1(t)(c_1(1-p_1(t)) - m)\,,\\ \frac{dp_2(t)}{dt} &= c_2 p_2(t)(1-p_1(t)-p_2(t)) - m p_2(t) - c_1 p_1(t) p_2(t)\,.\end{aligned}$$

At equilibrium, the occupancy of species 1 is

$$p_1^* = 1 - \frac{m}{c_1}\,,$$

which corresponds to the fixed point value of the Levins model [Lev69]. For species 2, the occupancy is

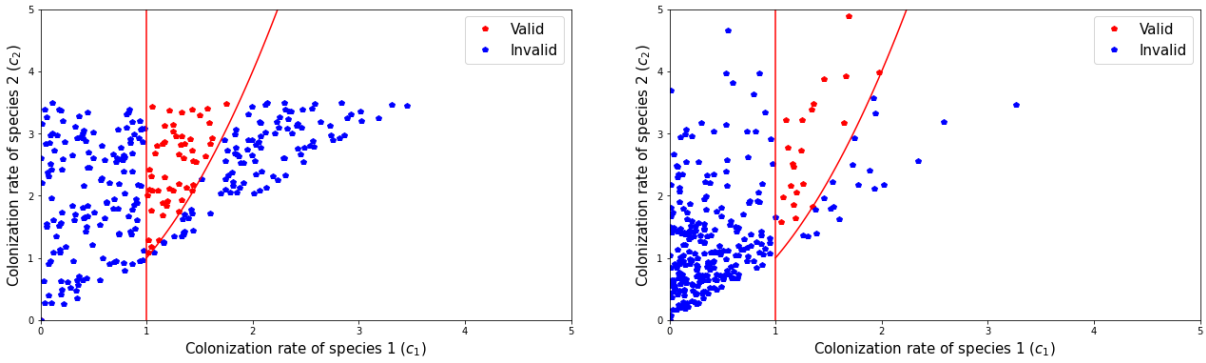
$$p_2^* = 1 - \frac{m}{c_2} - \left(1 + \frac{c_1}{c_2}\right) p_1^*\,.$$

Using the two equilibrium equations, we can deduce the constraints of coexistence between the species

$$c_1 > m\,;\; c_2 > \frac{c_1^2}{m}\,.$$

The constraints for a two-species system in the space of admissible solutions is illustrated by a density plot for two different probability distributions in Figure 4.1.1. We observe that each distribution gives a different number of red admissible points.

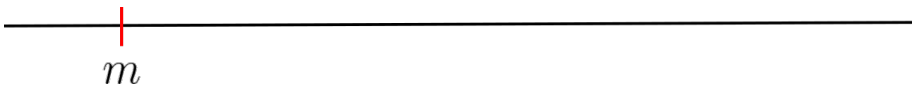
Graphical representation in the n-species system A natural question concerns the distribution of the vector $\boldsymbol{c}$ having a maximum of point in the admissible set. From a graphical perspective, we can see the sequence of conditions on an interval. Given a sample of colonization rates $\boldsymbol{c}$.



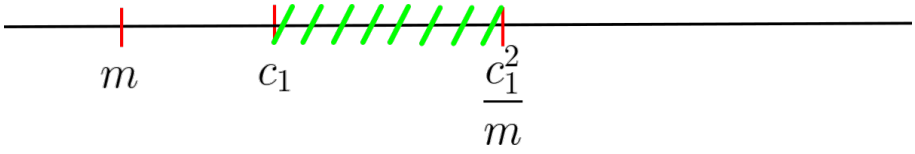
(a) Uniform $\mathcal{U}([0, 3.5])$ (b) Exponential $\mathcal{E}(1)$

Figure 4.1.1: Given two species with the same extinction rate $m = 1$, we sample $\boldsymbol{c} = (c_1, c_2)$ by a given distribution (in (a) uniform, in (b) exponential) and sort them $c_1 < c_2$. Each sample (c_1, c_2) is represented by a point in space, the number of samples is 300. If the couple coexist, the point is red. Otherwise, the point is blue. The set of conditions $\mathcal{C}$ is represented as red solid lines.

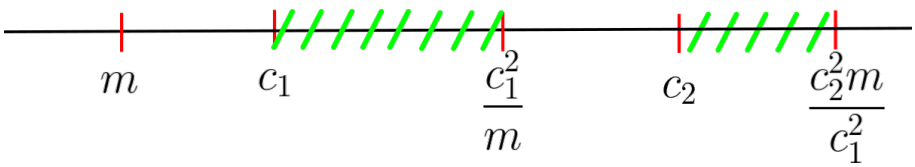
- As a simple requirement, the colonization rate must be greater than the extinction rate, otherwise the species goes to extinction. Is $c_1 > m$?



- If no, species 1 does not invade and the new condition is $c_2 > m$?
- If $c_1 > m$, species 1 invades the system and add a new condition. This condition has been referred to as the niche shadow in [KLD+99]. According to Meszéna *et al.* [MGPM06], this is similar to limiting similarity. Is $c_2 > c_1^2/m$?



- If yes, species 2 invades the system and adds a new niche shadow etc.



Remark 4.1. From a theoretical standpoint, we can compute the probability of coexistence by simple integral computations. However, without fixing the density function f of the colonization rates $\boldsymbol{c}$, the computations are untractable. Given a fixed distribution f, we can compute analytically the integral (in small dimension). In the contrary case, we can estimate the integral numerically by using standard Python library or Monte Carlo experiments.

Admissible sets for the colonization rates Consider a species pool of n species in which the colonization rates are randomly drawn from a given distribution following the dynamics of (4.2). Assume the extinction rate is the same for all species $\forall i \in [n]$, $m_i = m$. Let $\mathcal{S} \subset [n]$ be the subset of indices of the persisting species. The components of the equilibrium vector $\boldsymbol{p}^*$ can be classified into two classes: the persisting species, $i \in \mathcal{S}$ such that $p_i^* > 0$ and the vanishing species, $i \in \mathcal{S}^c$ such that $p_i^* = 0$.

The conditions on the colonization rate vector $\boldsymbol{c}$ are formulated in order to have coexistence between the species in the form of a set. The set of admissible solutions represents a series of algebraic conditions and depends on $\boldsymbol{c}$ and m

$$\mathcal{C}_m = \left\{ \boldsymbol{x} \in \mathbb{R}_+^n : x_{2i} > \frac{\left(\prod_{j=1}^{i-1} x_{2j+1}\right)^2}{m\left(\prod_{j=1}^{i-1} x_{2j}\right)^2} \quad ; \quad x_{2i+1} > \frac{m\left(\prod_{j=1}^{i} x_{2j}\right)^2}{\left(\prod_{j=1}^{i-1} x_{2j+1}\right)^2} \right\} . \tag{4.3}$$

The standard case ($m = 1$) will be denoted $\mathcal{C}$. These conditions are obtained using the recursion formula in the Appendix 4.A.2. The equations for the occupancy $\boldsymbol{p}^*$ for each species as a function of the colonization rate $\boldsymbol{c}^*$ and the fraction of empty patches formulas (already obtained in [Til94]) are re-computed.

How to maximize the coexistence between the species in the HT model? From a down-to-earth point of view, one could just choose a vector of colonization rates satisfying $\boldsymbol{c} \in \mathcal{C}$ corresponding to the case where all species coexist and survive. However, this condition is very stringent. To simplify this restriction, it is assumed that only a subset fulfills the conditions $\boldsymbol{c}_\mathcal{S} \in \mathcal{C}$ where $\boldsymbol{c}_\mathcal{S} : \{c_i, i \in \mathcal{S}\}$.

The coexistence problem can be reinterpreted as: given a finite number of species with colonization rates taken from a given finite support, how many species can be chosen so that they all persist?

$$\text{If } c_\mathcal{S} \in \mathcal{C}, \quad \mathbb{P}(|\mathcal{S}| = k \mid n) .$$

This problem is equivalent to check 2^n possibilities, either the presence or the absence of the species. However, since the problem includes a competitive hierarchy, if the first species has the opportunity to invade, it cannot be displaced by the next species. The most competitive species will always have priority. We end up testing only n conditions by following a decision tree (see Figure 4.1.2).

To sum up, the system follows the dynamics given by equation (4.2) starting from a pool of n species. The vector of colonization rate $\boldsymbol{c}$ is sampled from a positive probability distribution and we sort the $\boldsymbol{c}$ in increasing order. The aim is to assess if $c_\mathcal{S} \in \mathcal{C}$, $\mathbb{P}(|\mathcal{S}| = k \mid n)$, the remaining colonization rates for a fixed n. A schematic way to solve this problem is to browse a tree represented in Figure 4.1.2. From an algorithmic standpoint, a version is presented in pseudo-algorithm 1 which keeps the persistent species at the equilibrium point. We define this algorithm or ecosystem construction as all-at-once metacommunity process.

4.1.2 Relation with the Lotka-Volterra model and global stability of the equilibrium

Introduced at the beginning of the 20th century by Lotka [Lot25] and Volterra [Vol26], the Lotka-Volterra (LV) model is one of the most popular models in ecology. One of its

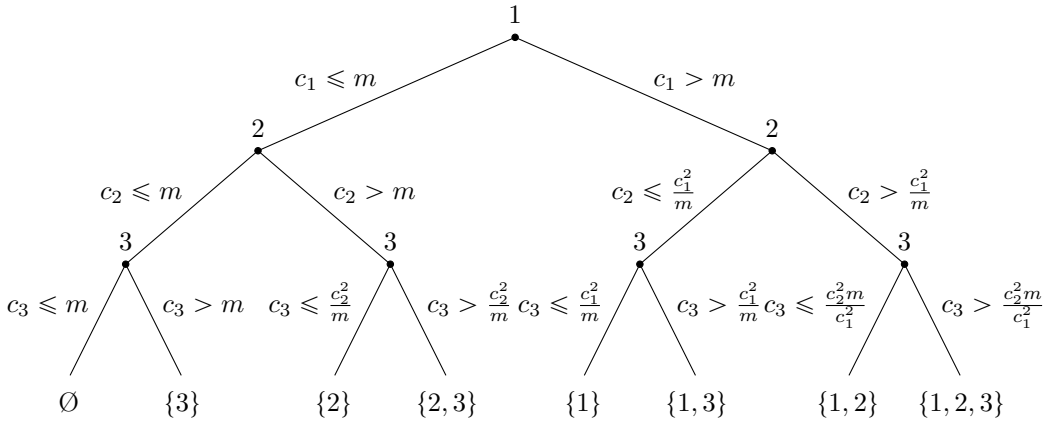


Figure 4.1.2: Decision tree for a 3-species system. The path of the binary tree selects the persisting species. At each node, the path on the right corresponds to the survival of the species, the one on the left to the extinction of the species. At the end, the root indicates the indexes of the persisting species.

strengths is its versatility: many models can be related to an LV model including, in particular, the extension of the Levins' metapopulation model to several species. We revised the HT model as a Lotka-Volterra model and gives a demonstration of the uniqueness and global stability of the equilibrium point of (4.2). Notice that Tilman [Til94] and Hastings [Has80] also carries out this problem in appendix of their articles.

First, we reformulate the HT model as a Lotka-Volterra model

$$\frac{dp_i}{dt} = c_i p_i \left(1 - \sum_{j=1}^{i} p_j\right) - m_i p_i - \left(\sum_{j=1}^{i-1} c_j p_j p_i\right) , \forall\, i \in [n] ,$$

$$\Leftrightarrow \frac{dp_i}{dt} = p_i (r_i - (Ap)_i) = p_i \left(r_i - \sum_{j=1}^{n} A_{ij} p_j\right) , \forall\, i \in [n] , \tag{4.4}$$

$$\Leftrightarrow \frac{d\boldsymbol{p}}{dt} = \mathrm{diag}(\boldsymbol{p})(\boldsymbol{r} - A\boldsymbol{p}) ,$$

where

$$r_i = c_i - m_i , \forall\, i \in [n] \text{ and } A = \begin{pmatrix} c_1 & 0 & 0 & \cdots & 0 \\ c_1 + c_2 & c_2 & 0 & & \\ c_1 + c_3 & c_2 + c_3 & c_3 & \ddots & \\ \vdots & & & \ddots & 0 \\ c_1 + c_n & & & & c_n \end{pmatrix} .$$

Generally, $\boldsymbol{r} = (r_1, ..., r_n)$ is understood as a growth rate and corresponds here to the dynamics of the species i without interactions. If the colonization rate is superior to the extinction rate, the species survives and grows indefinitely, otherwise, the species vanishes. However, this must be interpreted carefully because $p_i \in [0, 1]$, therefore $\boldsymbol{r}$ can not be clearly understood without A. The matrix A corresponds to a matrix of interactions, it is a competitive interaction matrix because $-A_{ij} < 0$, $\forall i, j \in [n]$. The impact of species $j < i$ on species i is $c_i + c_j$. On the one hand, this interaction coefficient depends on the colonization of c_j, the higher the colonization rates of better competitors, the less easy

Algorithm 1 All-at-once metacommunity

Require: $n \geqslant 0$
  $c \leftarrow list$
  for $l \in [1, n]$ **do** ▷ Creation of a random vector
    randomly choose c_{new} from a probability distribution;
    $c \leftarrow [c, c_{new}]$;
  end for
  $c \leftarrow Sorted(c)$; ▷ Ascending sorting algorithm
  for $j \in [1, len(c)]$ **do** ▷ Selection by the tree of the persisting species
    $S \leftarrow m$;
    if $c[j] \leqslant S$ **then**
      $del(c[j])$;
    else
      $S \leftarrow c[j]^2/S$;
    end if
  end for

it is for a species to persist. On the other hand, the presence of the c_i coefficient is less intuitive and intrinsically related to the model (4.2). The c_i comes out of the matrix term and states that we do not recolonize the patches where we are present.

Feasible fixed point To study the behavior of $\boldsymbol{p}(t), t \to +\infty$, we characterize the equilibrium of the system (4.4). An equilibrium $\boldsymbol{p}^*$ is defined as a vector satisfying

$$\forall i \in [n], \quad \frac{dp_i^*}{dt} = 0 \quad \Leftrightarrow \quad p_i^*(r_i - (A\boldsymbol{p}^*)_i) = 0\,.$$

If A is non-singular and a feasible fixed point exists i.e. $p_i^* > 0,\ \forall i \in [n]$, then the equilibrium $\boldsymbol{p}^*$ can be explicitly determined by

$$\boldsymbol{p}^* = A^{-1}\boldsymbol{r}\,.$$

The condition on the vector $\boldsymbol{c}$ to have all species coexisting is $\boldsymbol{c} \in \mathcal{C}_m$. These conditions are very restrictive. We are rather interested to determine a subsystem of species which coexist at equilibrium.

Fixed point with vanishing species In general, we consider cases in which there is no feasible equilibrium. A particular attention is given to the fixed point where some species may vanish i.e $p_i(t) \xrightarrow[t\to+\infty]{} 0$. In the following, we show that equation (4.4) has a unique globally stable equilibrium.

A unique equilibrium $\boldsymbol{p}^*$ to (4.4) is globally stable if for every $\boldsymbol{p}_0 > 0$, the solution to (4.4) which starts at $\boldsymbol{p}(0) = \boldsymbol{p}_0$ satisfies

$$\boldsymbol{p}(t) \xrightarrow[t\to\infty]{} \boldsymbol{p}^*\,.$$

Definition 4.1 (P-matrix)**.** A square matrix M is said to be a P-matrix if and only

1. All principal minors of M are positive

$$\det(M_{\mathcal{I}}) > 0\,, \quad \forall\, \mathcal{I} \subset [n]\,, \quad M_{\mathcal{I}} = (M_{k\ell})_{k,\ell\in\mathcal{I}}\,.$$

2. All real eigenvalues of M and its principal submatrices are positive.

Proposition 4.1. *The matrix A defined in* (4.4) *is a P-matrix.*

Proof. $\boldsymbol{c}$ is positive and A is a lower triangular matrix. The proof rely on two properties of the triangular matrix.

First, the eigenvalues of a lower triangular matrix are the diagonal entries of the matrix. Second, each principal submatrices of A are a lower triangular matrix and there eigenvalues correspond to a subset of $\boldsymbol{c} > 0$. Using definition 4.1 ends the proof. □

In graph theory, the matrix A represents a directed graph.

Definition 4.2 (Directed cycle)**.** A directed cycle in a directed graph A is a non-empty directed path in which only the first and last vertices are equal.

Since the matrix A is triangular, A represents a directed acyclic graph i.e. there is no cycle in the graph. Consequently, $-A$ is composed of cycles of length one (a cycle on himself).

Theorem 4.2 (Takeuchi *et al.* [TAT78])**.** *Suppose that $-A$ has only cycles of length one. Then the system* (4.4) *and every reduced system of* (4.4) *have a nonnegative and globally stable equilibrium point for each $\boldsymbol{r} \in \mathbb{R}^n$ iff A is a P-matrix.*

To conclude, relying on Theorem 4.2, the system (4.4), equivalent to (4.2), has a unique globally stable equilibrium point $\boldsymbol{p}^*$ independently of the parameter values i.e. for any initial condition $\boldsymbol{p}_0$, colonization rate $\boldsymbol{c}$ and extinction rate $\boldsymbol{m}$.

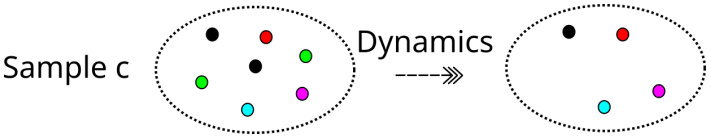
4.1.3 Dynamics of the model

In a given landscape of patches, suppose there is a pool of species whose dynamics is (4.2). It is assumed that for $i \in [n]$, c_i are i.i.d. random variables on $\mathbb{R}_+^*$. After drawing, sort the c_i in increasing order. The continuous dynamics is given by the differential equations (4.2) and the behavior is simple. Independently of the parameters $(\boldsymbol{c}, \boldsymbol{m})$ of the model and the initial condition, the dynamics converges to a unique equilibrium $\boldsymbol{p}^* = (p_1^*, ..., p_n^*)$.

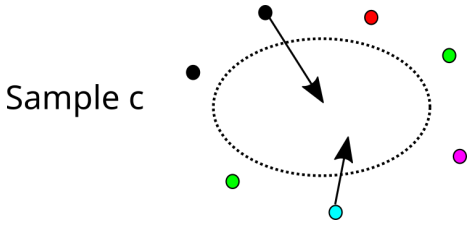
In fact, direct information on the persistent species is given by the conditions (4.3). If one waits long enough, the equilibrium is reached. This equilibrium is saturated because it is resistant against invasion of absent species [HS98]. Let $i \in \mathcal{S}^c$, the indices of the extinct species, then:

$$\left(\frac{1}{p_i}\frac{dp_i}{dt}\right)_{p_i \to 0^+} \leqslant 0\,.$$

By construction, the dynamics of the model (4.2) can be understood in two equivalent directions (see Figure 4.1.3). On the one hand, there is a primary ecosystem with a pool of n species that have different initial occupancies. The ecosystem changes continuously according to the ODE of the model. On the other hand, the ecosystem is assumed to be initially empty and species try to invade it sequentially in a random order. When a species tries to invade the system at a certain time t, it can cause the extinction of other species, but also the expansion of other species that were already present in the system. Indeed, the particularity is that a species that has invaded the system, even if it becomes extinct, it can invade again at any time.



(a) All-at-once metacommunity

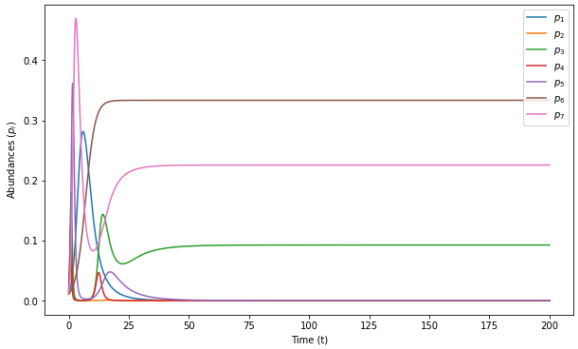


(b) Sequential invasion without exclusions

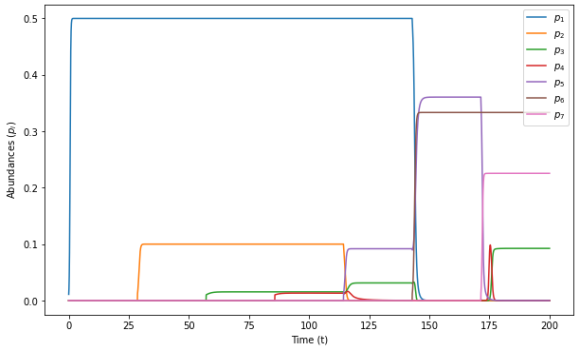
Figure 4.1.3: Diagrams of the two types of construction of the model (4.2) dynamics are represented for a ecosystem. The ellipse corresponds to the habitat and the points corresponds to the species. In Fig. (a), standard dynamics is shown. In Fig. (b), the assembly dynamics is represented with an invasion of each species.

Remark 4.2. A permanent extinction never occurs in the dynamics of the model. However, the vanishing components corresponding to the species going to extinction with $p_i^* = 0$ and $p_i(t) \xrightarrow[t\to\infty]{} 0$.

Both types of construction are represented in Figure 4.1.4. The ecosystem is composed of 7 species. We observe a convergence of the two dynamics towards the same equilibrium which is true for infinitely long periods. We notice that the disturbances generated by the first method are mainly located at the beginning of the dynamics whereas in the second case, each invasion generates disturbances.



(a) All-at-once metacommunity



(b) Sequential invasion without exclusions

Figure 4.1.4: The two types of construction of the model (4.2) dynamics are represented for a ecosystem of 7 species ($n = 7$). In Fig. (a), all-at-once metacommunity is shown with a starting pool of 7 species with random initial conditions. The abundance or habitat proportion of each species as a function of time is displayed. In Fig. (b), the sequential invasion without exclusions is represented with an invasion of each species at a regular time interval.

4.1.4 Choice of the extinction rate

In this paper, we focus on a specific HT model form (4.2) similar to the study of Hastings [Has80] . The extinction rate is equal for each species i.e. $\forall i \in [n]$, $m_i = m$. Without this condition, the set of admissible solutions $\mathcal{C}_m$ in (4.3) cannot be defined. In this paper, the

main interest is on the colonization rate $\boldsymbol{c}$. However, studying the model with different extinction rates is an intriguing perspective.

Without loss of generality and in a framework where the interest is on $\boldsymbol{c}$, we could have chosen an extinction rate equal to $m = 1$ for all species. However, we will maintain the term m (when possible) to have common results. We begin with equation (4.2) where we divide the equation by m

$$\frac{dp_i}{dt} = p_i c_i \left(1 - \sum_{j=1}^{i} p_j\right) - m p_i - \left(\sum_{j=1}^{i-1} c_j p_j p_i\right) , \tag{4.5}$$

$$\Leftrightarrow \quad \frac{dp_i}{dt}\frac{1}{m} = p_i \frac{c_i}{m}\left(1 - \sum_{j=1}^{i} p_j\right) - p_i - \left(\sum_{j=1}^{i-1} \frac{c_j}{m} p_j p_i\right) .$$

One can define a new vector $\boldsymbol{c}'$ where $c'_i = \frac{c_i}{m}\ \forall i \in [n]$. Considering m fixed, the analyses can proceed equivalently on $\boldsymbol{c}$ or $\boldsymbol{c}'$:

$$\frac{dp_i}{dt}\frac{1}{m} = p_i c'_i \left(1 - \sum_{j=1}^{i} p_j\right) - p_i - \left(\sum_{j=1}^{i-1} c'_j p_j p_i\right) . \tag{4.6}$$

The equilibrium of equation (4.6) is similar to (4.5) with $m = 1$. The factor $1/m$ has only an impact on the speed of the convergence. We incorporate the extinction rate m into the colonization rate, this corresponds to sample the ratio $c'_i = c_i/m$ from a positive probability distribution. Denote by $\mathcal{C}'$ the series of algebraic conditions associated with the vector $\boldsymbol{c}'$, then

$$\boldsymbol{c}' \in \mathcal{C}' \quad \Leftrightarrow \quad \boldsymbol{c} \in \mathcal{C}_m .$$

4.2 All-at-once metacommunity dynamics: persistent species in the hierarchical competition-colonization trade-off model

In the section 4.1, the set of admissible solutions $\mathcal{C}$ has been defined, this gives information about the number of persistent species in the model (4.2). Information on the dynamics was given to narrow the focus of the analysis to $\mathcal{C}$. We have all the necessary mathematical conditions to find a statistical distribution that maximizes the probability of coexistence between the species and to have information about the species richness which can coexist on average.

4.2.1 An invariance of species richness

Evolution of the species richness Given $\boldsymbol{c}$ sampled from a positive probability distribution, the species richness k after a sufficiently long time and after removal of species that are going to extinction as a function of the number of species in the initial pool n is on average $n/2$,

$$\mathbb{E}(|\mathcal{S}| = k |\, n) = \frac{n}{2} .$$

This result is shown by heuristic analysis and numerical simulations in the next section. Notice that the number of species in the initial pool n is equivalent to the number of species that can invade the system at any time. The function of the species richness according to the number of invasions is represented in Figure 4.2.1. All the supports of the distributions start at m which implies that no species is repressed by lack of colonization. Notice that on average half of the species coexist independently of the distribution of the colonization rate. For example, if we have 1000 species at the beginning, on average 500 will remain in the end, whether the $\boldsymbol{c}$ is drawn from a uniform or a power law distribution.

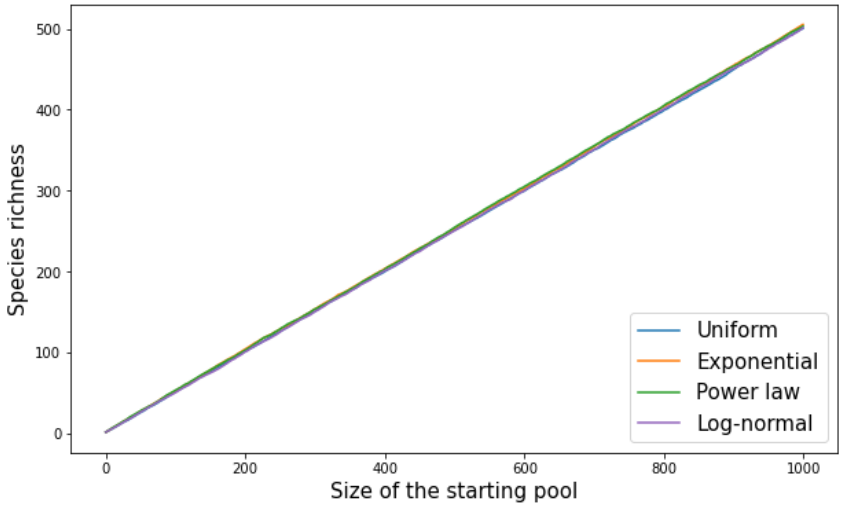


Figure 4.2.1: Representation of the species richness of the persistent species k as a function of the number of species in the initial pool n for different distributions. The curve is derived using Monte Carlo simulations by computing $P = 300$ times the algorithm 1 and averaging the number of persistent species. Independently of the distribution, the behavior of the curve is $f : x \to x/2$.

A subsidiary issue is the distribution of the persistent species. Is it possible to describe the distribution of the persistent species $\mathbb{P}(|\mathcal{S}| = k \mid n)$?

Distribution of the number of persistent species On average we observe that half of the persisting species remain, what is the distribution of the number of persisting species i.e. if the pool of species is n what is the probability that there is $|\mathcal{S}| = k$ persistent species?

The result is non-trivial. Independently of the distribution of the colonization rates $\boldsymbol{c}$, the distribution of the number of persistent species is close to the binomial distribution $\mathcal{B}(n, \frac{1}{2})$ if n is large i.e.

$$\mathbb{P}(|\mathcal{S}| = k \mid n) \underset{n\to\infty}{\sim} \binom{n}{k} \frac{1}{2^n}.$$

This is consistent with $\mathbb{E}(|\mathcal{S}| \mid n) = n/2$.

In Figure 4.2.2, the vicinity of the curves for different distributions is shown, the result is obvious. The symmetry of the binomial distribution indicates that the probability of having very few persistent species is the same as almost all persistent species.

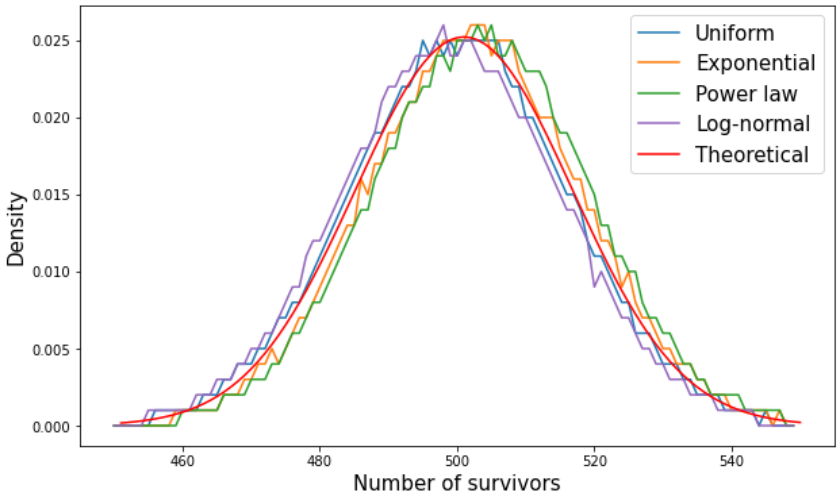


Figure 4.2.2: Representation of the distribution of the number of persistent species for a initial pool of $n = 1000$ species and for different distribution of colonization rate. Each curve is derived using Monte Carlo experiments by computing $P = 100000$ times the algorithm 1 and store the values obtained to form the outline of a histogram. The red curve corresponds to the density function of the binomial distribution $\mathcal{B}(n, \frac{1}{2})$.

4.2.2 Elements of proof of this invariance

Niche shadows From the definition of the admissible set $\mathcal{C}$ for the colonization rate, the odd condition is

$$c_{2i+1} > \frac{\left(\prod_{j=1}^{i} c_{2j}\right)^2}{\left(\prod_{j=1}^{i-1} c_{2j+1}\right)^2}.$$

Given l_{2i+1} the minimum threshold required for the colonization rate of species $2i+1$ to settle in the system, we have by definition:

$$l_{2i+1} = \frac{\left(\prod_{j=1}^{i} c_{2j}\right)^2}{\left(\prod_{j=1}^{i-1} c_{2j+1}\right)^2}.$$

Notice a recursive formula for the l_i

$$c_{2i+1} > l_{2i+1} = \frac{c_{2i}^2}{l_i}.$$

In an equivalent way, the same computations can be performed for the even condition, then we obtain a general recurrence formula:

$$c_{i+1} > l_{i+1} = \frac{c_i^2}{l_i}. \tag{4.7}$$

By definition, the interval (c_i, l_{i+1}) is called the niche shadow of species c_i where l_{i+1} correspond to the minimum threshold required for the colonization rate of species $i+1$ to settle in the system i.e. $c_{i+1} > l_{i+1}$.

Remark 4.3. If c_i has a narrow niche shadow i.e. $c_i/l_i \approx 1$, then c_{i+1} can be close to c_i and consequently many species might coexist in the final community (the so-called "infinite niche packing" [KLD+99]).

Given X_i a random variable describing the amount by which c_i exceeds its bound i.e. $X_i = c_i - l_i$. If the number of species in the initial pool n is large and the distribution of the colonization rates is sufficiently dense (close to each other), in the case of a finite support X_i will be small $X_i \sim O(n^{-1})$. From (4.7)

$$\begin{aligned} l_{i+1} &= \frac{c_i^2}{l_i}\,, \\ &= \frac{(l_i + X_i)^2}{l_i}\,, \\ &= l_i + 2X_i + \frac{X_i^2}{l_i}\,, \\ l_{i+1} &\approx l_i + 2X_i\,. \end{aligned} \tag{4.8}$$

Remark 4.4. The relation $l_{i+1} \approx l_i + 2X_i$ is important and already gives indications on the number of species present in the system on average. Indeed, the interval $[l_i, l_{i+1}]$ contains one species and is of size $2X_i$ where X_i is a random variable of unknown distribution of order $1/n$.

Proof in the uniform case Assume that n is sufficiently large, for convenience $m = 1$ and $\boldsymbol{c} \sim \mathcal{U}([1,2])$. Under the assumption (4.8), the assembly process is invariant to shifting the c_i and the problem is equivalent to $m = 0$ and $\boldsymbol{c} \sim \mathcal{U}([0,1])$. The aim consists in proving that the probability to have k persistent species from a pool of n species is $\mathbb{P}(|\mathcal{S}| = k|\, n) \sim \mathcal{B}(n, 1/2)$ according to the conditions (4.3).

Given $c_1 = x$ i.e. the probability that the first species falls in the interval $(x, x + dx)$ is

$$P(c_1 \in (x, x + dx)) = f(x)dx\,.$$

The probability that the species $2, ..., b + 1$ fall in the interval $(x, 2x)$ (i.e. these species are excluded by species 1) is

$$P(c_i \in (x, 2x), i \in [2, b]) = (F(2x) - F(x))^b\,.$$

The probability that the remaining $n-b-1$ species fall in the interval $(2x, 1)$ (distributed according to the original distribution left-truncated at $2x$ *independently* of the first $b + 1$ species) is

$$P(c_i \in (2x, 1), i \in [b + 1, n]) = (1 - F(2x))^{n-b-1}\,.$$

A combinatorial argument allows to compute the number of possibilities: $n!/(b!(n-1-b)!)$.

Let β_1 be a random variable corresponding to the number of species excluded by the first species. To compute the probability that $\beta_1 = b$ i.e. excludes the subsequent $b < n-1$ species, we integrate over the possible values of x (note that x runs only up to $1/2$ when $b < n - 1$, because $x > 1/2$ implies that all subsequent $n - 1$ species are excluded). We

will deal with that case separately below.

$$\begin{aligned}
P(\beta_1 = b) &= \int_0^{1/2} \frac{n!}{b!(n-1-b)!} f(x)(F(2x)-F(x))^b(1-F(2x))^{n-1-b}dx \\
&= \int_0^{1/2} \frac{n!}{b!(n-1-b)!}(2x-x)^b(1-2x)^{n-1-b}dx \\
&= \binom{n}{b}(n-b)\int_0^{1/2} x^b(1-2x)^{n-1-b}dx \\
&= \binom{n}{b}(n-b)\int_0^1 \frac{1}{2^b}u^b(1-u)^{n-1-b}\frac{1}{2}du\,,\ (u=2x)\,, \\
&= \binom{n}{b}(n-b)\frac{1}{2^{b+1}}\int_0^1 u^b(1-u)^{n-1-b}du\,, \\
&= \binom{n}{b}(n-b)\frac{1}{2^{b+1}}\mathrm{B}(b+1,n-b)\,.
\end{aligned}$$

where we have recognized the beta function in the last line and because b and n are integers, the beta function

$$\mathrm{B}(b+1,n-b) = \frac{b!(n-b-1)!}{n!}\,.$$

This cancels with the earlier combinatorial factor and we finally find

$$P(\beta_1 = b) = \frac{1}{2^{b+1}}\,.$$

In the case $b = n-1$, we can use a standard geometric series argument

$$P(\beta_1 = n-1) = 1 - \sum_{b=0}^{n-2}\frac{1}{2^{b+1}} = 1 - \sum_{b=1}^{n-1}\frac{1}{2^b} = 1 - \frac{1-\frac{1}{2^n}}{\frac{1}{2}} = \frac{1}{2^{n-1}}\,.$$

If the truncated distribution is a rescaled instance of the original distribution as for the uniform distribution, then we can find the probability that a particular subset of species $\alpha_1, \alpha_2, ...\alpha_k$ survives by simply multiplying the probabilities that the first species excludes $b_1 = \alpha_2 - \alpha_1 - 1$ species, the remaining $n - b_1 - 1$ are distributed uniformly on $(x, 1)$. By rescaling, we find that the probability that the second survivor (the minimum among the remaining $n-b_1-1$ species) excludes $b_2 = \alpha_3 - \alpha_2 - 1$ species, and so on. For convenience, we specify the subset of coexisting species by the sequence of gaps: $b_i = \alpha_{i+1} - \alpha_i - 1$. Ultimately, we find the probability of a particular sequence of k survivors by computing

$$\begin{aligned}
\mathbb{P}\left(\bigcap_{i=1}^{k}\{\beta_i = b_i\}\right) &= \prod_{i=1}^{k}\mathbb{P}(\beta_i = b_i)\,, \\
&= \prod_{i=1}^{k-1}\frac{1}{2^{b_i+1}} \times \frac{1}{2^{b_k}}\,, \\
&= \frac{1}{2^{\sum_{i=1}^{k} b_k + k - 1}}\,.
\end{aligned}$$

The final factor $(1/2^{b_k})$ has a reduced power because the kth (final) survivor must exclude all remaining species. By definition, $\sum_{i=1}^{k} b_i = n - k$, so the probability of *any* particular set of survivors is

$$\mathbb{P}\left(\bigcap_{i=1}^{k}\{\beta_i = b_i\}\right) = \frac{1}{2^{n-1}}\,.$$

This probability does not depend on the particular sequence $\beta_1, ...\beta_k$, or even on k. This implies that the probability of finding k survivors is just the number of possible sets of k survivors, multiplied by the probability of each. Since the first species always survives, we have

$$P(|\mathcal{S}| = k \mid n) = \binom{n-1}{k-1}\frac{1}{2^{n-1}}\,,$$

which is the binomial distribution $\mathcal{B}(n-1, 1/2)$.

Insight in the general case In this section, we present a non-rigorous intuition to extend this result to a certain class of distributions. Assume that $\boldsymbol{c}$ is distributed according to an arbitrary distribution with support on $[m, \infty)$, PDF $f(x)$ and CDF $F(x)$. As above, we shift to obtain support on $[0, \infty)$ and fix $m = 0$. The probability that the first species excludes b species is

$$P(\beta_1 = b) = \int_0^{\infty} \frac{n!}{b!(n-1-b)!} f(x)(F(2x) - F(x))^b (1 - F(2x))^{n-1-b} dx\,. \tag{4.9}$$

The assumption comes from the fact that the density of species 1 is close to zero when n becomes large. Consequently, the integral (4.9) can be approximated, first we truncate the integral at the median, as in the uniform case. This choice avoids integrating over negative probabilities. When $b = n - 1$, we instead integrate up to ∞.

$$P(\beta_1 = b) \approx \int_0^{F^{-1}(\frac{1}{2})} \frac{n!}{b!(n-1-b)!} f(x)(F(2x) - F(x))^b (1 - F(2x))^{n-1-b} dx\,.$$

The second argument consider the density of the colonization rate of species 1. The integral (4.9) is dominated by the behavior near $x = 0$ because the colonization rate c_1 has high probability to be close between $[0, \epsilon], \epsilon > 0$, then

$$P(\beta_1 = b) \approx \int_0^{\epsilon} \frac{n!}{b!(n-1-b)!} f(x)(F(2x) - F(x))^b (1 - F(2x))^{n-1-b} dx\,.$$

The last argument is the less rigorous and gives an insight of the behavior in general distribution, we use a linear approximation to write $F(2x) = 2\,F(x)$ which is poor as x becomes large, but for distributions with high density at 0 (m, in the original frame) and sufficiently light tails, the integral will be dominated by the behavior near $x = 0$, where the linear approximation is good.

$$P(\beta_1 = b) \approx \int_0^{F^{-1}(\frac{1}{2})} \frac{n!}{b!(n-1-b)!} f(x)(F(x))^b (1 - 2F(x))^{n-1-b} dx\,.$$

The rest of the proof consists in pursuing the uniform case example. We use the substitution $u = 2\,F(x)$. This implies $du = 2\,f(x)dx$, so we have

$$P(\beta_1 = b) = \binom{n}{b}(n-b)\int_0^1 \frac{1}{2^b}u^b(1-u)^{n-1-b}\frac{1}{2}du = \frac{1}{2^{b+1}}\,,$$

exactly as in the uniform case and $\frac{1}{2^b}$ for $b = n+1$. Consequently, we can find the probability of an particular sequence of survivors by multiplying the probabilities of $\bigcap_{i=1}^{k}\{\beta_i = b_i\}$.

4.2.3 Properties of the persisting species

In the previous section, we showed a precise description of the number of persisting species in the model (4.2) as a function of the starting n pool. A natural question is to understand the properties of the persisting species (proportion of habitat, empty space left, impact of the extinction rate) as in the work of Tilman [Til94] or Kinzig *et al.* [KLD+99].

Evolution of the fraction of empty patches

The fraction of empty patches features the first interesting property as a consequence of the "all-at-once" equilibrium described above. In a counter-intuitive way even when the size of the ecosystem is very large at the beginning $n \to +\infty$, h_n, it does not necessarily converge towards 0. In the single-species Levins model with parameter (c, m), the fraction of empty patches is given by m/c i.e. the higher the colonization rate, smaller the fraction of empty patches. In the n-species model, the conclusion is similar, we have the relation $h_i \sim (m/c_i)^{1/2}$

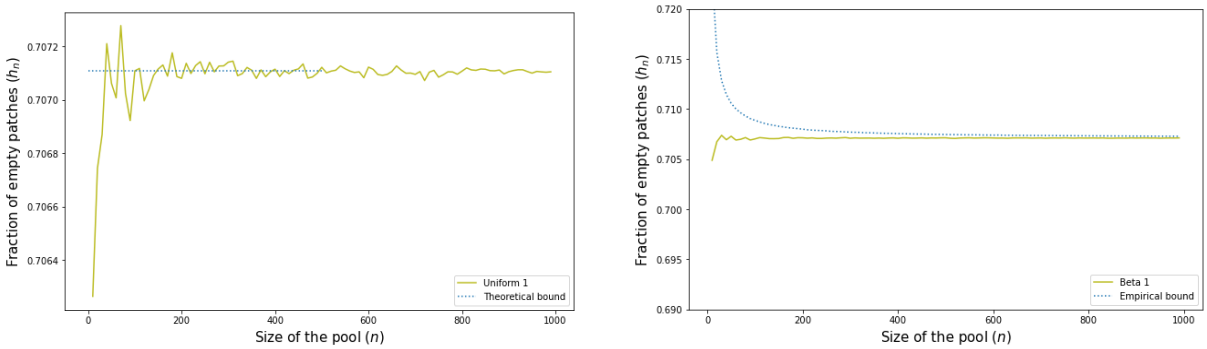
$$\begin{aligned} h_i &= 1 - \sum_{j=1}^{i} p_j^*\,, \\ &= \frac{mh_{i-1} + \sum_{j=1}^{i-1} mp_j^*}{c_i h_{i-1}}\,, \\ \frac{h_i}{m} &= \frac{h_{i-1} + \sum_{j=1}^{i-1} p_j^*}{c_i h_{i-1}}\,, \\ &= \frac{1}{c_i h_{i-1}}\,. \end{aligned}$$

using the fact that $h_i < h_{i-1}$, one obtain the same equality as in Kinzig *et al.* [KLD+99],

$$\sqrt{\frac{m}{c_{i+1}}} \leqslant h_i \leqslant \sqrt{\frac{m}{c_i}}\,. \tag{4.10}$$

This equation shows that to have an idea of the empty space, it is sufficient to have a good approximation of $c_n = \max_{i\in[n]} c_i$. In the appendix 4.B.3, we recall the maximum distributions of different distributions (uniform, Exponential, Pareto) and we give a heuristic to compute numerically the maximum of a family of r.v for a finite n.

The simplest example is that of a finite support. In figure 4.2.3, we compare the relation $h_i \sim (m/c_i)^{1/2}$ for a uniform distribution and a Beta distribution. Theoretically in the case of the uniform distribution, consider n standard i.i.d. random variables $U_i \sim$



(a) $\mathcal{U}([1,2])$ (b) $Beta(1,1)$

Figure 4.2.3: Comparison between the empirical and theoretical bounds of the fraction of empty spaces (h_n) as a function of the size of the initial pool of species. In panel (a), this relationship is illustrated in the case where $\boldsymbol{c}$ follows a uniform distribution (of max = 2) and the theoretical estimation is $1/\sqrt{2}$. In panel (b), the same relationship is given when $\boldsymbol{c}$ follows a Beta distribution of parameters $Beta(1,1)$ and compared to the empirical estimation of the maximum of the Beta function as a function of n applied to the expression $1/\sqrt{c_n}$.

$\mathcal{U}([1, c_{\max}])$, $\mathbb{E}(\max U_i) \to c_{\max}$. In the case of the Beta distribution, we use the heuristic of the empirical approximation.

When the distribution of $\boldsymbol{c}$ has infinite support, theory states that when the number of species tends to infinity, the fraction of empty spaces tends to zero. However, when n is too small, the prediction is not accurate because of the gap between c_{i+1} and c_i. The standard example is the Pareto distribution which is shown in Figure 4.2.4 with the upper and lower bound given in (4.10).

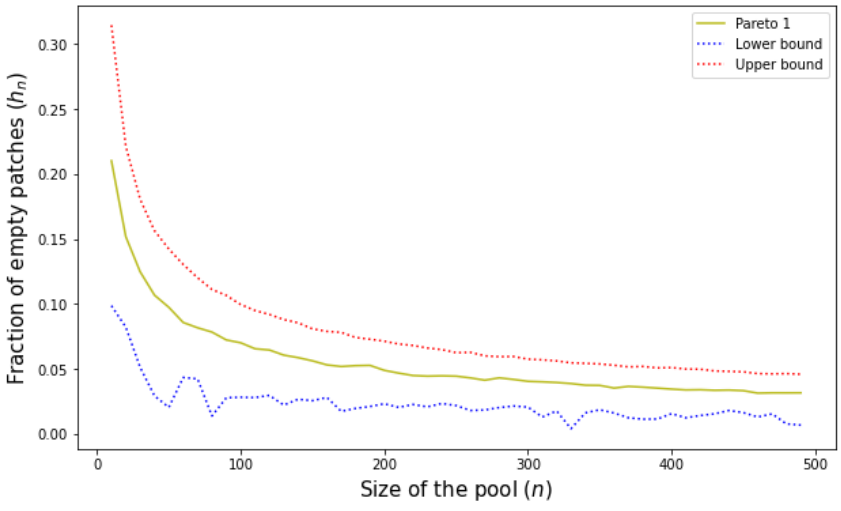


Figure 4.2.4: Comparison between the empirical and theoretical bounds of the fraction of empty spaces (h_n) as a function of the size of the initial pool for the Pareto distribution $\mathcal{P}(1)$. The dotted lines represent the upper bound (red) and lower bound (blue) of (4.10) computed using the heuristics to obtain an estimation of the maximum for a fixed n.

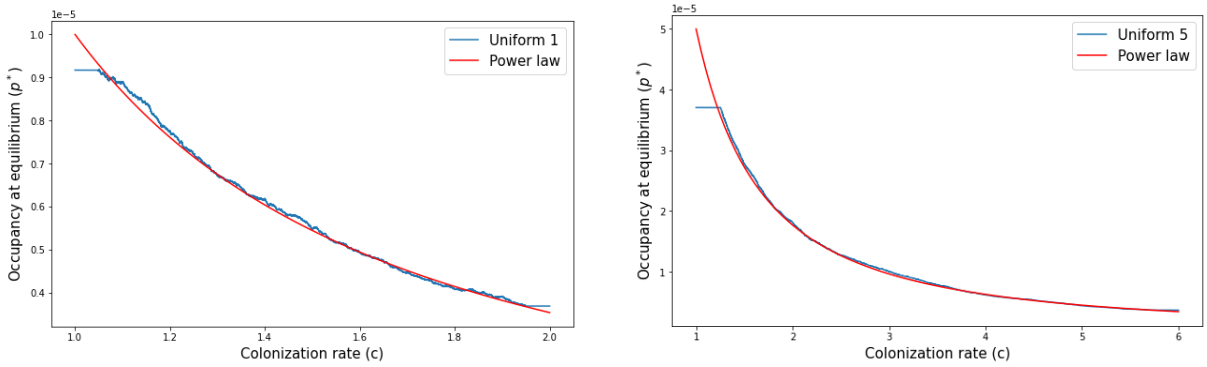
From an ecological point of view, the empty space in the habitat depends mostly on

the colonization capacity of the best colonizing species.

Occupancies

The distribution of the occupancies has been studied by Kinzig *et al.* [KLD+99]. In this paragraph, the main result is illustrated. They give a relationship between occupancies and colonization rates, they show that $p^*(c) \propto c^{-3/2}$ when extinction rate is equal to 1. The most competitive species tends to occupy a larger share of the landscape than the most colonizing species.

In figure 4.2.5 we plot the power distribution relationship between the occupancies and the colonization rate $p^*(c) \sim c^{-3/2}$ for an ecosystem with $n = 100000$ species for two different uniform distributions. The matching is quite remarkable.



(a) $\mathcal{U}([1,2])$ (b) $\mathcal{U}([1,6])$

Figure 4.2.5: Representation of the occupancies at equilibrium $\boldsymbol{p}^*$ as a function of the colonization rates $\boldsymbol{c}$. The occupancy of a species is associated to its colonization rate is computed for a ecosystem of $n = 100000$ species. A moving average is used to smooth out all the points. In Fig.(a), the uniform distribution of support $[1, 2]$ is plotted and the red curve represents the power distribution $1/c^{3/2}$. In Fig.(b), the uniform distribution of support $[1, 6]$ is plotted and the red curve represents the power distribution $5/c^{3/2}$.

In Figure 4.2.6, the empirical distribution in the case of an exponential distribution $\mathcal{E}(1)$ is displayed. The result is quite intriguing and different. The beginning of the curve seems to decrease and then there is an increase in the proportion of habitat. This is the effect of the tail of the distribution. Indeed, the exponential distribution has an infinite support, very colonizing species can be present in the system and recover a large proportion of habitat.

Impact of the extinction rate

The impact of the extinction rate or the diversity disturbance relationship has been largely studied and introduced in Hastings [Has80] in the HT model. The number of species in the system differs according to the extinction rate and the number of persistent species as a function of the extinction rate gives an optimal threshold i.e intermediate disturbance hypothesis. The number of persisting species reaches a peak when the left support of the probability distribution is at m. Let $c_{\min}$, the left edge of the support of the distribution.

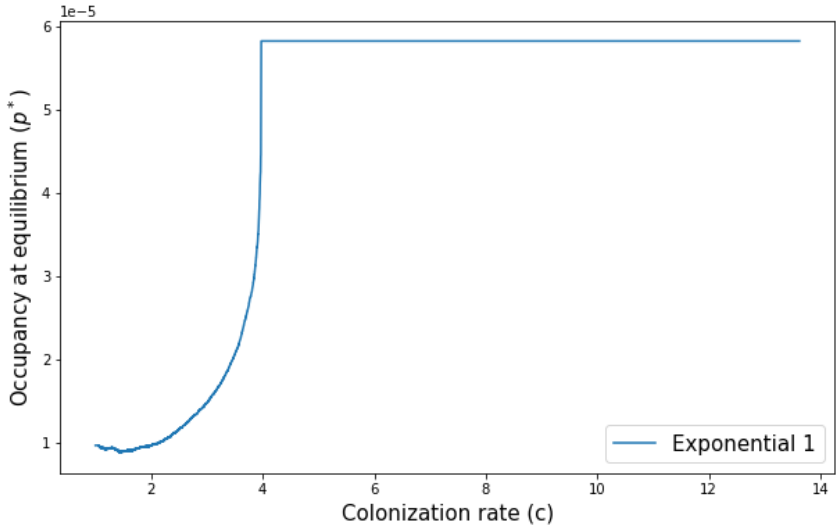


Figure 4.2.6: Representation of the occupancies at equilibrium $\boldsymbol{p}^*$ as a function of the colonization rates $\boldsymbol{c}$. The occupancy of a species is associated to its colonization rate is computed for a ecosystem of $n = 100000$ species. A moving average is used to smooth out all the points. The case of an exponential distribution $\mathcal{E}(1)$ of colonization rates is represented.

The standard example of a value $c_{\min} = 1$ is studied. At $m = c_{\min}$, it is the convergence to the binomial distribution, half of the species survive independently of the chosen distribution (see Figure 4.2.1).

The edge effect is more accentuated when the distribution has high density close to $c_{\min}$.

Mathematically, the causes of the loss of species richness are explained when $m \neq c_{\min}$ by

- if $m \gg c_{\min}$, a large part of the density of the distribution is truncated with the first condition $c_i > m, \forall i \in [n]$. Species that have a lower colonization rate than the extinction rate cannot survive.

- $m \ll c_{\min}$: a large part of the density of the distribution is truncated with the second condition, if we suppose that c_1 is very close to $m_{min} = 1$, then $c_2 > \frac{c_1^2}{m} \approx \frac{1}{m}$. Species whose colonization rate is less than the inverse of the extinction rate cannot survive. If c_{min} starts far from m, it implies larger and larger niche shadows, because the first species takes up a large part of the patches (and so the next one must be considerably better colonizer, etc. by domino effect).

From an ecological point of view, an increased extinction rate of all species is generally associated with a decrease in species richness. However, it is not intuitive that a decrease in extinction rate implies a decrease in species richness. In the context of abundances, Fox [Fox13] argues that the intermediate disturbance hypothesis should be abandoned. In the HT model, a decrease in extinction rate implies a supremacy of the most competitive species at the expense of the others, hence the decrease in extinction.

In the case of $m_{min} = 1$, we represent in figure 4.2.7 the impact of the extinction rate on the species richness in the case of two uniform distributions of different support. We

can observe that if we take two extinction rates: $m_1 = 1.3$ and $m_2 = 1/m_1 = 0.77$, the species richness is the same. A larger distribution support will tend to reduce the impact of the extinction rate, as a higher density on the left side of the support increases the deterioration of the ecosystem.

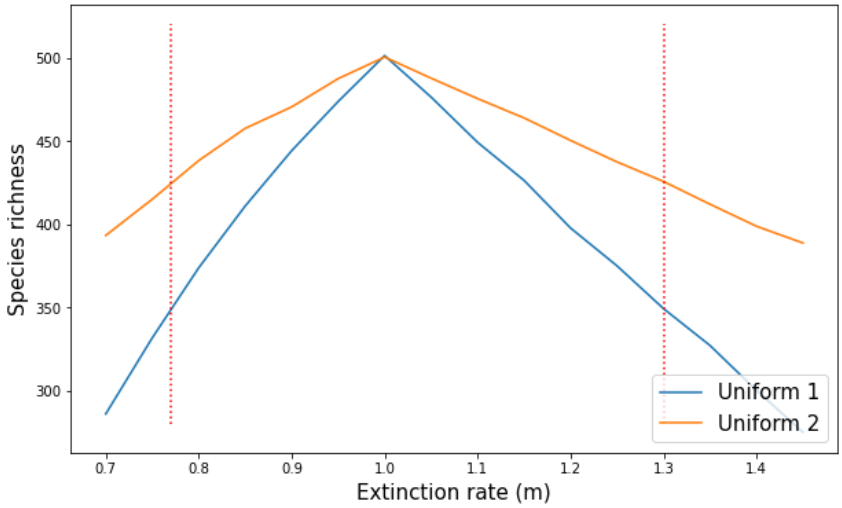


Figure 4.2.7: Representation of species richness as a function of extinction. Two distributions are identified $\mathcal{U}([1,2])$ (blue line) and $\mathcal{U}([1,3])$ (orange line) whose left edge $c_{\min} = 1$. The right red vertical lines represent $m_1 = 1.3$ and the left $m_2 = 1/m_1 = 0.77$.

4.3 Sequences of invasions and community assembly

In section 4.2, the all-at-once metacommunity dynamics was presented in an ecosystem with a starting species pool where the model evolves according to the dynamics (4.2). We have seen that this approach is equivalent to considering an empty ecosystem at the beginning, adding species one by one and considering that they can invade the system again at any time. Here we want to consider what would happen if an extinct species was condemned to never return to the system. This has potential realism (when compared to a real assembly process) and allows to answer new questions such as the effects of invasions in communities. This approach is called the assembly process or invasion sequences [MDC+17, RR85, Cas90, RZB+09, RBL+19, Tok04]. It consists in starting from an empty system and filling it through a series of invasions. We define an iterative process where at each time step a species tries to invade the system and is then confronted with an invasion condition. If this invasion condition is satisfied, it can enter the system and potentially displace species from the system. These displaced species will not be able to re-enter the system afterward so the composition of the system at any point in time does not only depend on which species have tried to invade but also on the order in which they have tried to do so. The invasion condition related to $\mathcal{C}$ evolves each time a new species enters the system. The invasion of a species depends on the current state of the system and therefore depends on the invasions that have been carried out in the past. This implementation of the model has a historical contingency that was not present in the all-at-once metacommunity dynamics and reveals different outcomes.

4.3.1 Description of the invasion process and extinction cascades

Dynamics of the invasion process

The sequential invasion process follows the dynamics described by the system (4.2). We start from an empty system and we inject one species after the other into the system. Each species is characterized by its colonization rate which is randomly drawn following the same probability distribution. The invasion of a new species can cause the extinction of one or more species. In Figure 4.3.1, a diagram of the situation is illustrated. Each species is injected into the system at very low occupancy $p_l \sim \varepsilon > 0$ and we let the dynamics elapse. If the occupancy of a species converges to 0, the species is considered extinct and is removed from the system. According to Theorem (4.2), there exists a unique globally stable equilibrium.

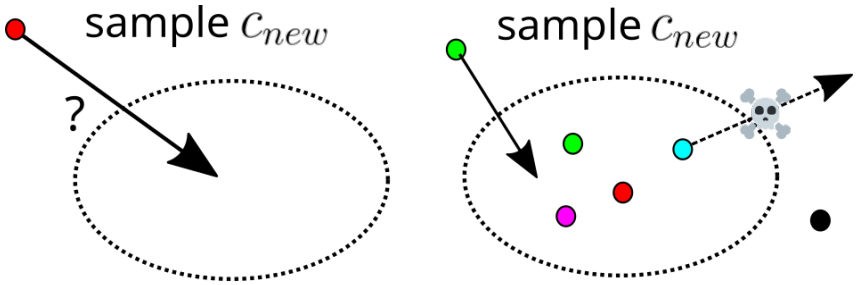
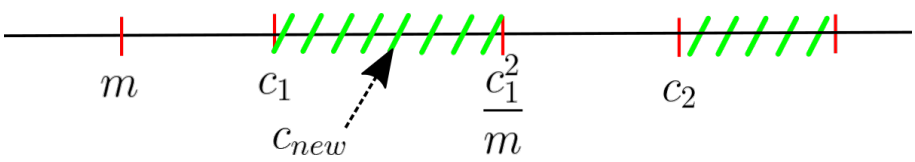


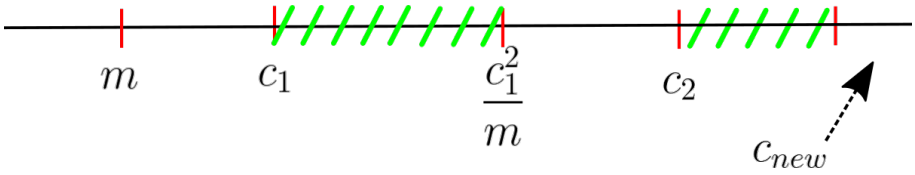
Figure 4.3.1: Sequential invasion process in an ecosystem. The ellipse corresponds to the habitat and the points corresponds to the species. For each new species, a new colonization rate c_{new} is drawn. The left diagram represents the possibility (or not) of invasion of a new species in an ecosystem. The right diagram represents the situation of invasion of a new species leading to the extinction of a species present in the ecosystem.

Three types of situations can be observed when new species attempts to invade. Consider the situation of a system by an interval with the presence of the niche shadows (i.e. the interval of exclusion created by each species in the system by the condition defined by $\mathcal{C}$) of the different species of the system. The niche shadows are represented by a shaded green zone.

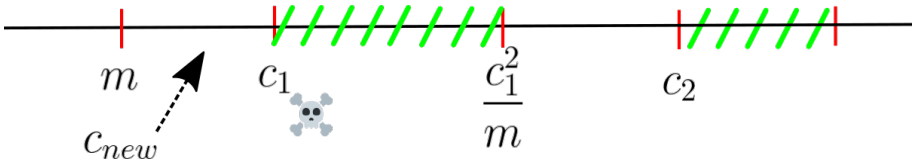
- The new species cannot invade if its colonization rate falls into the niche shadow of a species in the community. In this case the equilibrium of the system remains the same (same species, same occupancies).



- The new species invades and the system converges to a new equilibrium that includes the new species.



- The new species invades and generates an extinction cascade when its niche shadow affects other species and challenges all the shadow niches in the system. The system converges to a new equilibrium that includes the new species but without some of the more colonizing species, depending on the location of the new species niche shadow and the shifts it implies in the niche shadows of all other species.



In the figure 4.3.2, we represent the dynamics of the system as a result of the successive invasions. A first observation is the fast convergence of the system towards its new equilibrium. The invasion of a less competitive species does not affect the abundance of more competitive species. However, it does change the abundance of less competitive species and can cause extinctions. We note the extinction of species 3 and 5 during the invasion of species 6.

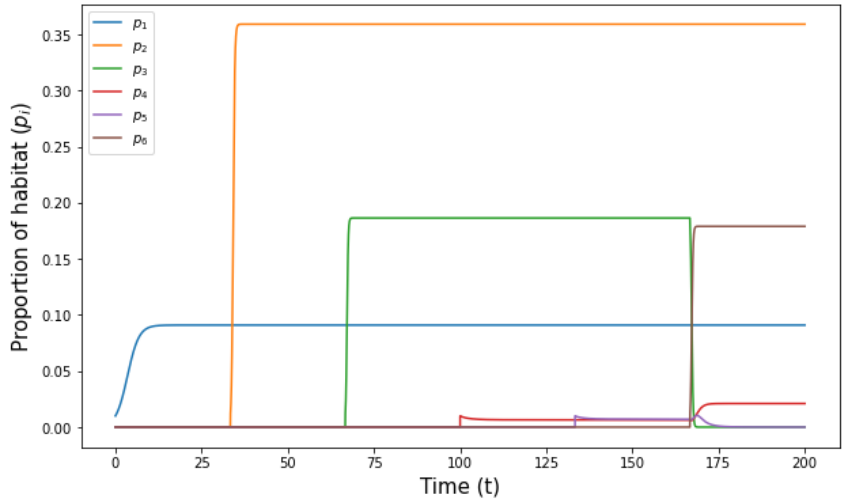
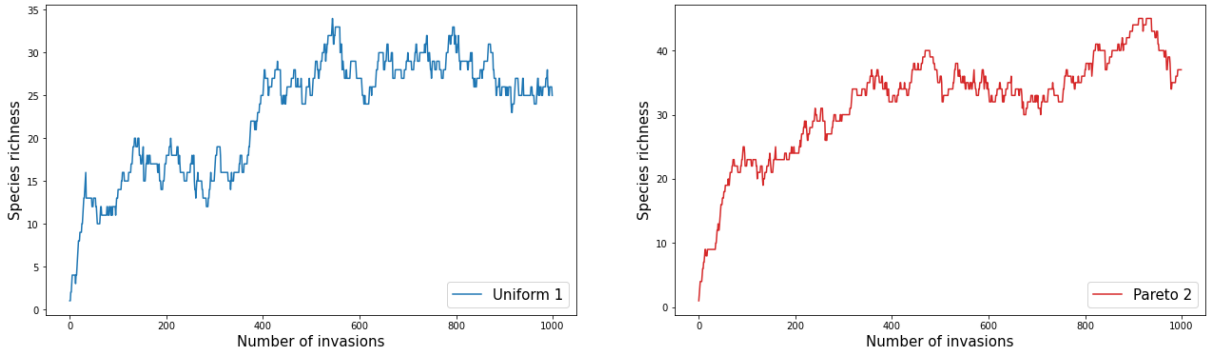


Figure 4.3.2: Representation of the sequential invasion dynamics with an invasion of each species (6 in total) at a regular time interval. The dynamics of the system between each invasion corresponds to (4.2).

The dynamics of the sequential invasion process indicates convergence to a new unique globally stable equilibrium at each invasion (see Theorem 4.2). Finally, in the study of the sequential invasion process, the restriction to analyze the conditions of invasions by reducing the dynamic phase between invasions (and species removed) is sufficient. Without loss of generality, we can describe the dynamics where each iteration corresponds

to an invasion. At each iteration, suppose we wait long enough for our system to reach an equilibrium. The dynamics of the system by iteration is represented for different distributions in Figure 4.3.3. The growth of the species richness is no longer linear (30 species in the system after 1000 invasions), the system evolves in an irregular way with invasions and extinctions of species.



(a) $\mathcal{U}([1,2])$ (b) $\mathcal{P}(2)$

Figure 4.3.3: Dynamics of the species richness as a function of the invasion trials in the sequential invasion process. In both figures the extinction rate is $m = 1$.

Historical contingencies

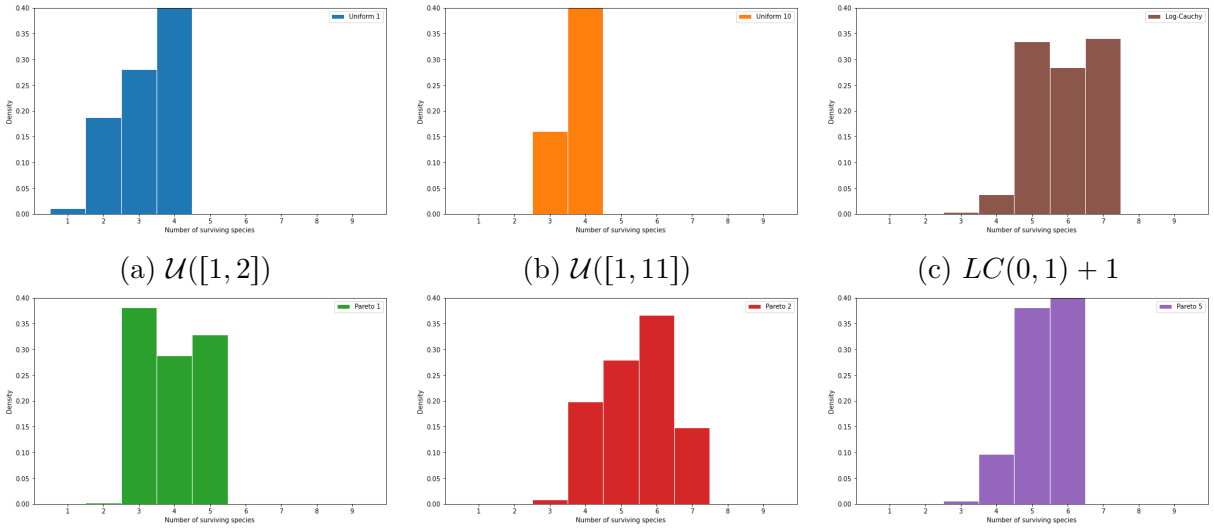
The impact of the order of invasions is important in the sequential invasion process, which gives rise to historical contingencies i.e. the probability of invasion of a new species depends strongly on the past. This is the main difference with the all-at-once metacommunity which does not admit this phenomenon because all species can invade the system at any time. In Figure 4.3.4, all orders of invasions are tested in a unique 9 species framework. It highlights the historical contingencies phenomenon for different distributions. We observe that the distribution seems to have a rather small variance (around ~ 3 possible values of species richness).

To avoid the bias of a single realization and have a general idea of the distribution, we observe a histogram of the number of persisting species in a very simple case with 6 species with a sample of 500 realisations (see Figure 4.3.5). We observe that depending on the order of invasion, the range of values goes from a single species persisting to all species persisting. We also notice that the distribution of the number of persisting species depends strongly on the distribution of colonization rates. However, the behavior of the distribution looks like a unimodal distribution. This is related to the sequential invasion process which is similar to the all-at-once metacommunity because there are not many extinctions at the beginning of an invasion sequence.

Formalization of the sequential invasion process

The purpose of this section consists in formalizing the iterative sequential invasion process. When a species wants to invade the system, there are two main scenarios (we combine the invasions with 0+ extinctions):

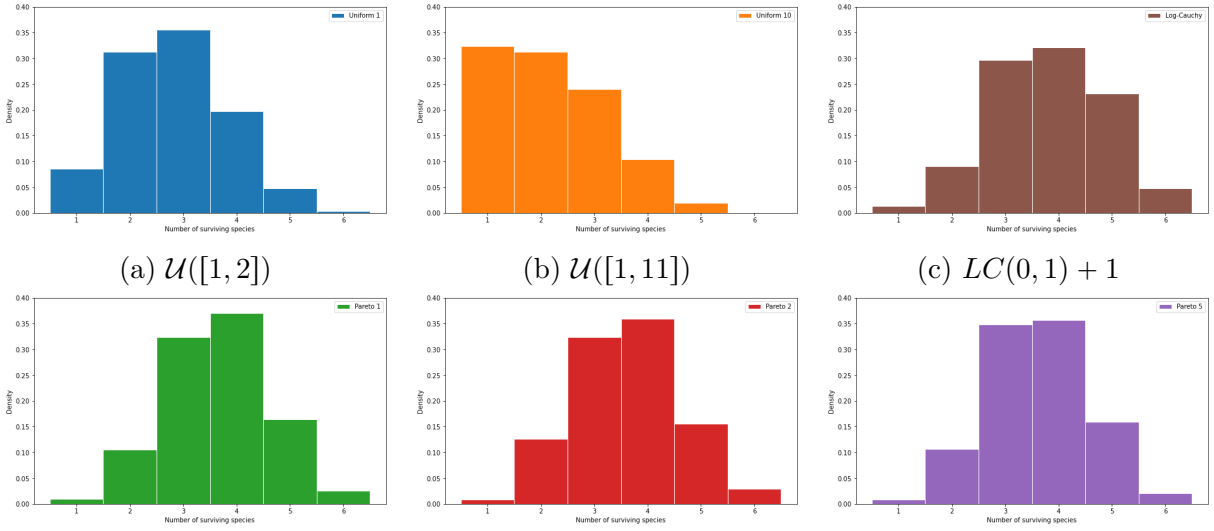
1. failure of the invasion of the new colonizer,



(a) $\mathcal{U}([1,2])$ (b) $\mathcal{U}([1,11])$ (c) $LC(0,1)+1$

(d) $\mathcal{P}(1)$ (e) $\mathcal{P}(2)$ (f) $\mathcal{P}(5)$

Figure 4.3.4: Histogram of the distribution of the number of persisting species after the invasion of 9 species. The extinction rate is $m = 1$. All order of invasions are tested and each plot corresponds to a different distribution for the drawing of the new values of colonization rate.



(a) $\mathcal{U}([1,2])$ (b) $\mathcal{U}([1,11])$ (c) $LC(0,1)+1$

(d) $\mathcal{P}(1)$ (e) $\mathcal{P}(2)$ (f) $\mathcal{P}(5)$

Figure 4.3.5: Histogram of the distribution of the number of persistent species after 6 invasions for a large number of realizations $P = 500$. The extinction rate is $m = 1$. All 720 orders of invasions are tested and each plot corresponds to a different distribution for the drawing of the new values of colonization rate.

2. success of the invasion: zero, one or more species vanish.

Consider an initial failure zone or niche shadow zone (if the colonization rate of the invading species falls in this zone, it does not invade) depending on the extinction rate $\mathcal{F}^0 = [0, m]$. The colonization rate of a new species c_{new} trying to invade the system is drawn from a chosen positive probability distribution. There exists two main options

1. $c_{new} \in \mathcal{F}^0$, the species fails,

2. $c_{new} \notin \mathcal{F}^0$, the species invades.

If the species invades, a new failure zone is defined $\mathcal{F}^1 = [0, m] \cup [c_{new}, c_{new}^2/m]$ as well as a vector of colonization of the species present after one iteration $\boldsymbol{c_1^1} = (c_{new})$. If the species fails, the empty vector after one iteration $\boldsymbol{c_1^0}$ is defined.

Given $\boldsymbol{c_l^k} = (c_1, ..., c_k)$ the vector of colonization rate of the k species in the system after l iteration. The failure zone $\mathcal{F}^k$ associated to the k species in the system corresponds to an union of intervals

$$\begin{aligned}\mathcal{F}^k = [0, m] \cup &\left(\bigcup_{i=1}^{\lfloor \frac{k}{2} \rfloor} \left[x_{2i}, \frac{\left(\prod_{j=1}^{i} x_{2j} \right)^2}{\left(\prod_{j=1}^{i-1} x_{2j+1} \right)^2} \right] \right) \\ \bigcup &\left(\bigcup_{i=1}^{\lfloor \frac{k+1}{2} \rfloor} \left[x_{2i-1}, \frac{\left(\prod_{j=1}^{i-1} x_{2j+1} \right)^2}{\left(\prod_{j=1}^{i-1} x_{2j} \right)^2} \right] \right).\end{aligned} \tag{4.11}$$

This union of intervals represents a non-admissible invasion set evolving through the sequential invasion process. By construction, they are disjoint intervals. On the opposite, we define by $\mathcal{A}^k = [0, +\infty) \backslash \mathcal{F}^k$ the admissible set.

Given a new species trying to invade the system with a colonization rate c_{new} drawn from a chosen probability distribution, there are two options:

1. $c_{new} \in \mathcal{F}^k$, the species fails,

2. $c_{new} \notin \mathcal{F}^k$, the species invades.

In case of failure, we can define the vector at the next iteration: $\boldsymbol{c_l^k} = \boldsymbol{c_{l+1}^k}$. In the case of success, the problem is more complex, the next paragraph deals with its description.

Update of the non-admissible zone The second case is more sophisticated (the species invades). The invasion of a new species into the system can have important consequences on the conditions of the failure zone and the species present in the system. Two sub-scenarios are distinguished.

At iteration l, after the invasion, the new colonization rate c_{new} is added to the vector of colonization rate of the species present $\boldsymbol{c_l^k} = (c_1, .., c_k)$. We seek the index i such that $c_i < c_{new} < c_{i+1}$ and define a new vector: $\boldsymbol{c_l^{k+1}} = (c_1, ..., c_i, c_{new}, c_{i+1}, .., c_k)$. Recall the set of admissible conditions

$$\mathcal{C} = \left\{ \boldsymbol{x} \in \mathbb{R}_+^n : x_{2i} > \frac{\left(\prod_{j=1}^{i-1} x_{2j+1} \right)^2}{\left(\prod_{j=1}^{i-1} x_{2j} \right)^2} \quad ; \quad x_{2i+1} > \frac{\left(\prod_{j=1}^{i} x_{2j} \right)^2}{\left(\prod_{j=1}^{i-1} x_{2j+1} \right)^2} \right\}.$$

Two corresponding cases depending on the impact on the species present in the system are considered

- if $\boldsymbol{c_l^{k+1}} \in \mathcal{C}$, the species invades without causing extinction, update $\mathcal{F}^k \to \mathcal{F}^{k+1}$ using the new vector of colonization $\boldsymbol{c_l^{k+1}} \to \boldsymbol{c_{l+1}^{k+1}}$.

- if $\boldsymbol{c_l^{k+1}} \notin \mathcal{C}$, up to the value c_{new}, the conditions are fulfilled because the invasion of a less competitive species does not affect the more competitive species. Then, we go through the rest of the vector, each time a condition is not met, we delete the associated species and re-compute the niche shadows. At the end of the process, d species are extinct, we obtain a new vector $\boldsymbol{c_{l+1}^{k+1-d}}$.

If $d > 0$, this is called an extinction cascade.

Algorithm As for the all-at-once metacommunity process, the sequential invasion process algorithm is described by the pseudo code given in the Algorithm 2. It is possible to optimize the algorithm by keeping in memory the right bounds of the intervals of $\mathcal{F}^k$. This avoids having to recompute the set $\mathcal{F}^k$ at each iteration and allows to check that a species does not invade faster at each iteration i.e. the selection by the tree start at the index of c_{new}. Recall that in contrast to the all-at-once metacommunity process, the

Algorithm 2 Sequential invasion process

Require: $n \geqslant 0$
  $c \leftarrow list$
  $S \leftarrow m$
  for $l \in [1, n]$ **do**
    randomly choose c_{new} from a probability distribution;
    $c \leftarrow [c, c_{new}]$;
    $c \leftarrow Sorted(c)$; ▷ Ascending sorting algorithm
    for $j \in [1, len(c)]$ **do** ▷ Selection by the tree of the remaining species
      $S \leftarrow m$;
      if $c[j] \leqslant S$ **then**
        $del(c[j])$;
      else
        $S \leftarrow c[j]^2/S$;
      end if
    end for
  end for

order of appearance of species is important due to historical contingencies. This effect can be illustrated by a small example: $c_1 = 4$, $c_2 = 5$, $c_3 = 20$.

- $c_2 \rightarrow c_3 \rightarrow c_1$, only species 1 survives (species 3 does not invade).
- $c_2 \rightarrow c_1 \rightarrow c_3$, species 1 and species 3 survive (species 2 collapses but species 3 invades because $4^2 = 16 < 20$).

Analytical insights

At iteration l, assume that k species are present in the system. Given c_{new}, the colonization rate of a new species drawn from a chosen probability distribution with density function f, then the probability that the new species invades is

$$\mathbb{P}(c_{new} \notin \mathcal{F}^k) = 1 - \int_{\mathcal{F}^k} f(x)dx\,. \tag{4.12}$$

The probability that there is no extinction cascades i.e. $c_l^{k+1} \in \mathcal{C}$ is difficult to express due to its strong dependence on the past events. The arrival of the new species depends on the state of the system just as the species present in the system depend on the past states of the system. However, insights on the probability of the number of extinction cascades are addressed in the next paragraph.

Extinction cascades

Extinction cascades [CBB+11, PA13, VAN15, RZVM17] are the key events that disrupt the system. Unlike the all-at-once metacommunity process, which allows a species to invade the system at any time, the sequential invasion process does not provide for transient species. Let x_l be the number of species in the system at iteration l, the following recurrence equation determines the behaviour of the species richness in the ecosystem

$$x_{l+1} = \begin{cases} x_l + 1 - e_l \text{ if } c_{new} \notin \mathcal{F}\,, \\ x_l \text{ if } c_{new} \in \mathcal{F}\,, \end{cases} \tag{4.13}$$

where c_{new} corresponds to the colonization rate of the invader and e_k the number of extinctions if the new species invades (see Witting *et al.* [WTL00] for a practical application of probability to extinction cascades).

If an invader manages to invade by being a very good colonizer then $\mathcal{F}$ will not change significantly. However, if an invader is very competitive with a very low colonization rate, then the set $\mathcal{F}$ will be severely affected.

Remark 4.5. This sequence indicates that the growth of the number of species in the system is a combination of two phenomena:

- the probability of invasion,
- the magnitude of the extinction cascades.

In the following section, we will see that the speed of dynamics of the richness of the system seems logarithmic (see Figure 4.3.9a). Either the probability of invasion decreases with time, or the number of extinction cascades increases.

Theoretical toy model

From a theoretical standpoint, the recursion equation (4.13) can be presented as a toy model in order to understand the underlying mechanisms of the sequential invasion process. Given a starting condition $x_0 = 0$ at the beginning of the sequential invasion process i.e. there is initially no species in the system and x_l the number of persisting species at iteration l. We define a recurrence relation depending on random variables

$$x_{l+1} = X_l(x_l + Y_l - (1 - Y_l)(Z_l - 1)) + (1 - X_l)x_l\,, \tag{4.14}$$

where $X_l \sim \mathcal{B}(1, p_l^X)$ are independent random variables, p_l^X correspond to the probability of an invasion at iteration l. The random variables $Y_l \sim \mathcal{B}(1, p_l^Y)$ are independent, p_l^Y is the probability that there is no extinction after an invasion occurring at iteration l. The random variables $Z_l \sim \mathcal{L}(\lambda_l^Z)$ are independent and $\mathcal{L}$ is an unknown distribution and $\lambda_l^Z \in [1, \infty)$ (the "1" comes from the fact that there is a minimum of one extinction). The

mean of the distribution λ_l^Z corresponds to the number of extinctions after an invasion at iteration l.

The 'toy model' is an approximation of the sequential invasion process. We seek estimates of the three parameters defining the r.v. given in equation (4.14). The estimators of the parameters for the binomials X, Y are the classical estimators of the mean. For the parameter of Z, we assume an underlying Poisson model, we compute the classical estimator of the mean.

In Figure 4.3.6 when $c_{new} \sim \mathcal{U}([0,1])$, the dynamics of the quantities $(p_l^X, p_l^X, \lambda_l^Z - 1)$ is illustrated as a function of interval of invasions i.e. we estimate each parameter in an interval $\forall\, i \in \mathbb{N}^*$, $[100(i-1), 100i] \subset [0, n]$ by averaging the values every 100 time steps. We notice that the probability of invasion is slightly higher than 0.5. When a species invades, it has a probability close to 1/3 of avoiding extinction and if it does, on average 1.5 species will be ejected from the landscape. The recurrence formula (4.14) can be

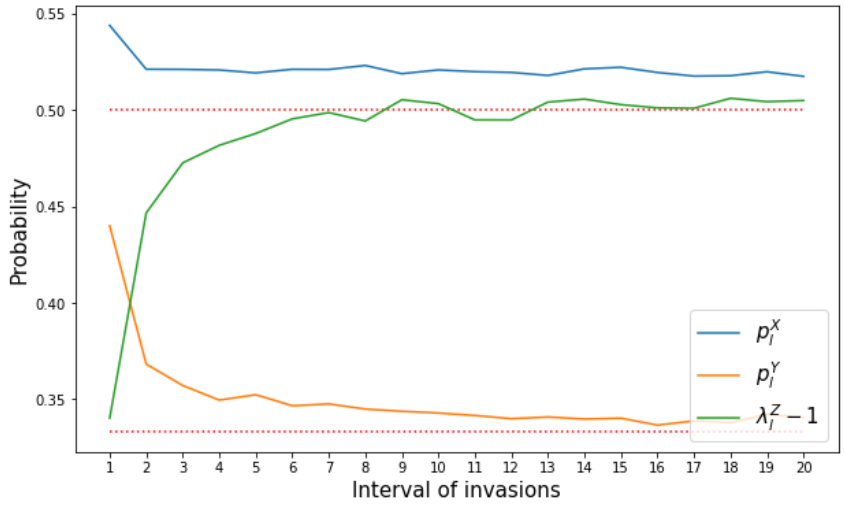


Figure 4.3.6: Representation of the parameters $(p_l^X, p_l^X, \lambda_l^Z - 1)$ as a function of interval of invasions. A interval of invasions $i \in [1, 20]$ corresponds to $[100(i-1), 100i]$. In this example, the sequential invasion process count $n = 2000$ invasions in total and c_{new} is sampled from a uniform distribution $\mathcal{U}([0,1])$. Each interval l provides an estimator of the parameters. The red dotted horizontal lines are indicators plotted respectively at 1/3 and 1/2.

rewritten in a simpler form

$$\begin{aligned} x_{l+1} &= x_l + X_l(1 - Z_l + Y_l Z_l)\,, \\ &= \sum_{k=1}^{l} X_k(1 - Z_k + Y_k Z_k)\,. \end{aligned}$$

Recall that $\forall k \in \mathbb{N}^*$, X_k, Y_k, Z_k are independent by construction (Z is the r.v. of the number of extinctions conditional on there being one after an invasion, Y is the r.v. that says there are no extinctions conditional on there being an invasion, X is the r.v. that

says there is an invasion), then

$$\begin{aligned}\mathbb{E}(x_{l+1}) &= \sum_{k=1}^{l} \mathbb{E}(X_k)\left(1 - \mathbb{E}(Z_k) + \mathbb{E}(Y_k)\mathbb{E}(Z_k)\right) , \\ &= \sum_{k=1}^{l} p_k^X (1 - \lambda_k^Z + p_k^Y \lambda_k^Z) .\end{aligned}$$

Given, $v_k = p_k^X(1 - \lambda_k^Z + p_k^Y \lambda_k^Z)$ the increment to the expected species richness between invasion k and $k+1$. From a heuristics standpoint, the habitat is non-saturated if

$$v_k \sim O(k^{-a}), a \leqslant 1 \Rightarrow \mathbb{E}(x_{l+1}) \xrightarrow[l\to\infty]{} \infty .$$

A naive computation is to use the values of Figure 4.3.6, we notice the value $v_k \xrightarrow[k\to\infty]{} 0$, but how fast?

In Figure 4.3.7, we compare v_k to k^{-1}, we notice that v_k seems to decrease slowly. The value of the slope gives -a where $v_k \sim k^{-a}$. From a mathematical point of view, this is a promising way to understand rigorously the dynamics of x_l and to show that the habitat is not saturated in the sequential invasion process.

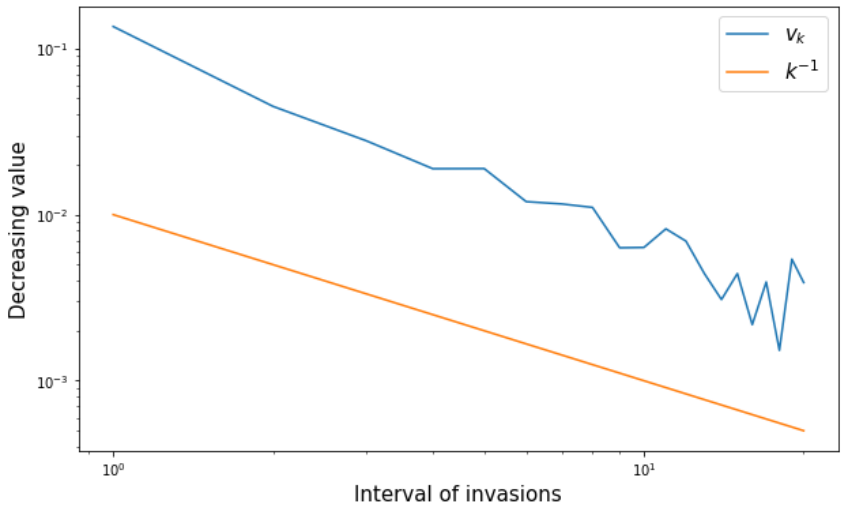
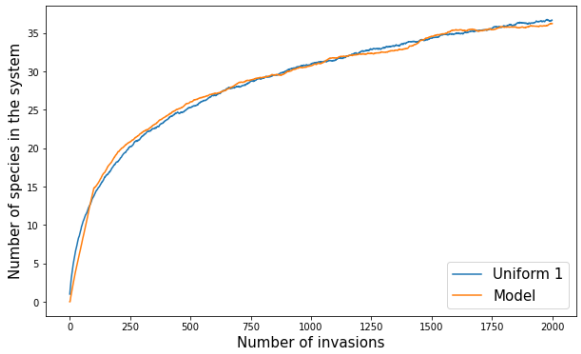


Figure 4.3.7: Comparison of the the increment to the expected species richness v_k and k^{-1} the threshold to get an infinite sum of v_k.

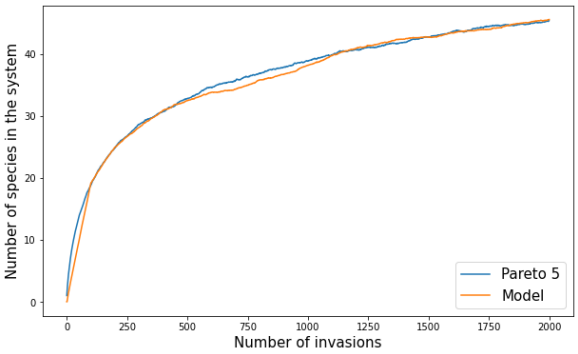
To motivate the model (4.14), by estimating the parameters $(p_l^X, p_l^X, \lambda_l^Z)$ as a function of interval of invasions, we reproduce the (mean) of the growth of the sequential invasion process for two different distributions in Figure 4.3.8.

4.3.2 Dynamics of the system over time and final composition

As in the all-at-once metacommunity process, we are interested in understanding the dynamics of the sequential invasion process over time and the properties of the species that inhabit the system to grasp the trade-off between competition and colonization.



(a) $\mathcal{U}([1,2])$



(b) $\mathcal{P}(5)$

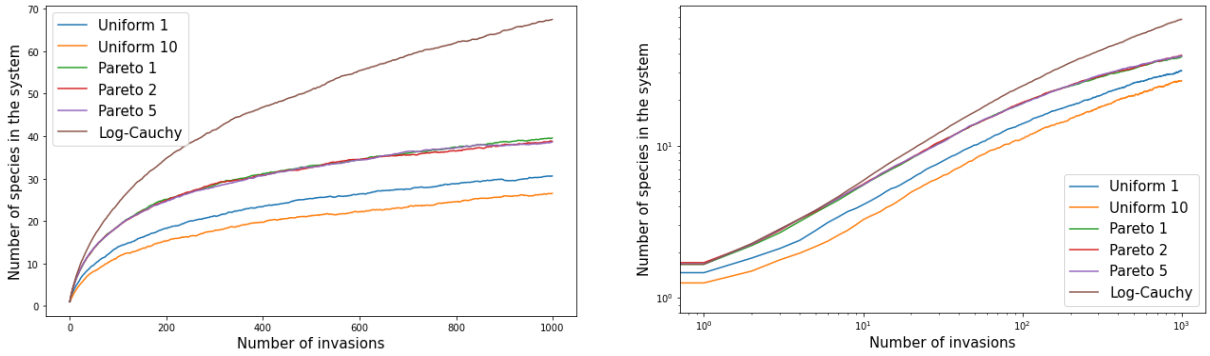
Figure 4.3.8: Dynamics of the species richness as a function of the number of invasions in the sequential invasion process. The extinction rate is $m = 1$. In both figures, the blue curve represents the empirical mean of $P_1 = 1000$ sequential invasion process. The orange curve represents mean of $P_2 = 1000$ curve of toy model (4.14) where the parameters $(p_l^X, p_l^X, \lambda_l^Z)$ are estimated in each interval of invasions of size 100.

Dynamics of the species richness

In the context of the dynamics of the sequential invasion process over invasions, we notice that regardless of the distribution chosen, the system tends to saturate or at least seems to be growing very slowly. In Figure 4.3.9a, we represent the dynamics of the species richness over the iterations (invasion trials) for some benchmark distributions (uniform, Pareto/Power, Log-Cauchy). At the beginning, we note a very fast growth reminiscent of the all-at-once metacommunity dynamics, then a consequent slowing down. If we compare the performances of each distribution, colonization rates sampled by a Log-Cauchy has higher species richness than Pareto and the uniform distributions. Regular distributions seem to have a weaker performance than heavy-tail distributions which seem to perform equally well in the Pareto case. An increase in the maximum value of the support of regular distributions does not improve the situation such as the uniform $\mathcal{U}([1,11])$, which has worse species richness than uniform $\mathcal{U}([1,2])$.

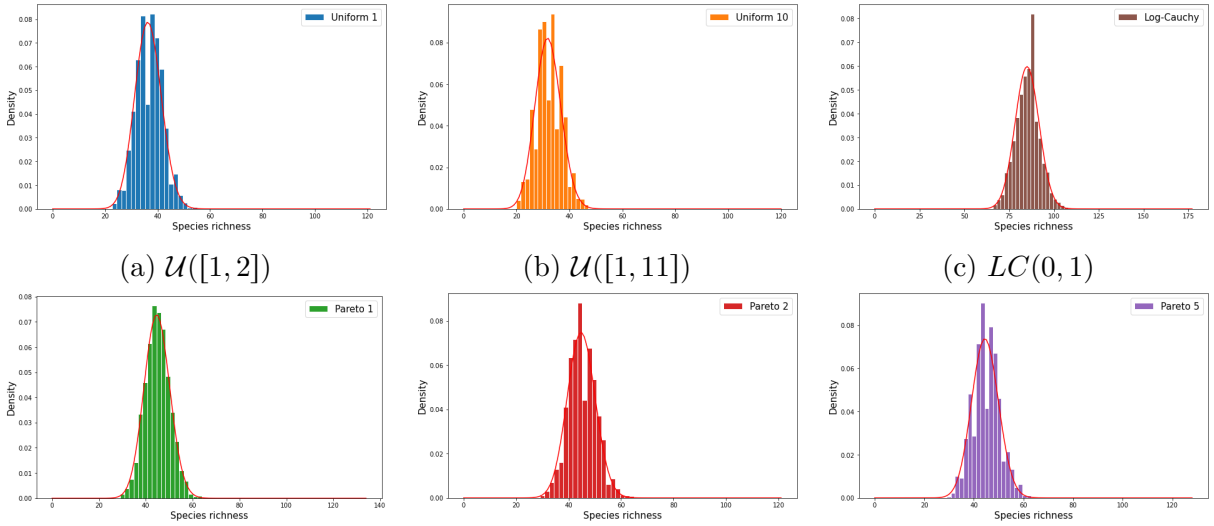
The shape of the curves suggest a logarithmic relationship between the species richness and the number of invasion trials (see Figure 4.3.9b). One could think of habitat size variation where more habitat tends to contain more species with a form of saturation caused by extinctions, dispersion, prey-predator interaction, etc.

Distribution of species richness One of the major results in the all-at-once metacommunity process is the convergence of the distribution of the realized species richness to a binomial distribution. For a given number of invasions, we want to have an idea of the distribution of the richness of the system. In Figure 4.3.10, we represent a histogram of the number of species for a large sample of system. A comparison is established with a Binomial distribution with an estimator of the mean and variance. The species richness distribution appears to be symmetric, but it is close to the binomial distribution. Many "peaks" appear in the density shape (see Fig. 4.3.10-b).



(a) Standard scale (b) Log-log scale

Figure 4.3.9: Representation of the species richness of the assembly as a function of the invasions for different distributions. The curve is derived using Monte Carlo simulations by computing $P = 2000$ times the algorithm 2 and averaging the number of persistent species. In panel (a) the plot corresponds to the standard case, whereas panel (b) is the same plot using log-log scale.



(a) $\mathcal{U}([1,2])$ (b) $\mathcal{U}([1,11])$ (c) $LC(0,1)$

(d) $\mathcal{P}(1)$ (e) $\mathcal{P}(2)$ (f) $\mathcal{P}(5)$

Figure 4.3.10: Histogram of the species richness for a fixed number of invasions ($n = 2000$). The histogram is derived using Monte Carlo simulations by computing $P = 2000$ times the algorithm 2 which corresponds to the size of the sample for the histogram. In each plots, the red curve corresponds to a Binomial mass function adjusted to the mean and variance estimators of the sample.

Impact and dynamics of extinction cascades

In section 4.3.1, the phenomenon of successive extinctions taking place when a certain type of species invades the system has been carefully defined. It was deduced that these extinctions are a major cause of the phenomenon of habitat saturation. The purpose of this section is to quantify numerically the extinction cascades in order to describe precisely the sequential invasion process.

In Figure 4.3.11, the frequency of occurrence of the cascade is represented for a very

large number of systems. To take into account the effect of extinction on the state of the system i.e. the species richness is not similar for every distribution, the size of the cascade is studied in proportion to the size of the final system. Here, we decide to normalize by the state of the final system. However, we could normalize by the size of the system at the time the cascade takes place.

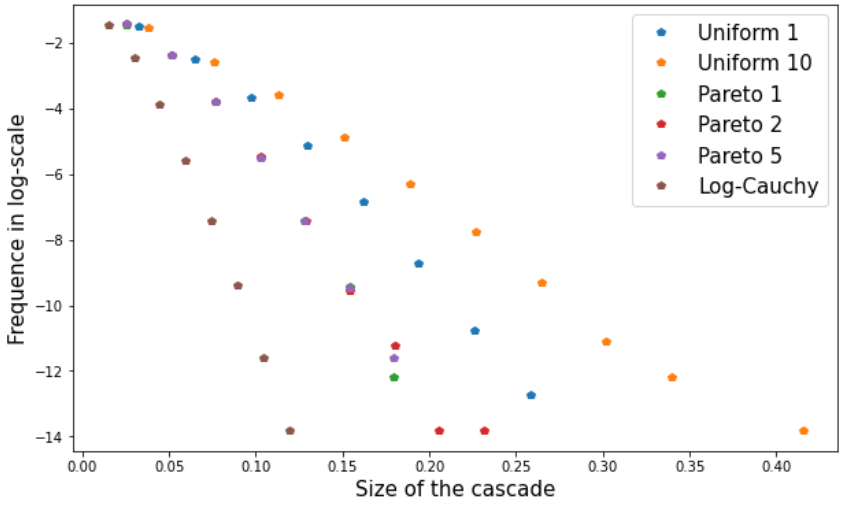


Figure 4.3.11: Distribution of the size of extinction cascades as a function of the size of the cascade normalized by the final number of species in the system. We realize a Monte Carlo experiment for a system of size $n = 1000$ and perform the sequential invasion process $P = 1000$ times.

On the one hand, by comparing the different distributions, the regular distributions undergoes much more important cascades in relation to the final size of the system. On the contrary, the Log-Cauchy distribution seems to be much less affected by the cascade phenomenon. Finally, if we look at the Pareto distribution, we observe the heavier the tail, the more extinction cascades occur. For regular distributions, a similar phenomenon is observed, a larger support implies a larger cascade.

On the other hand, the frequency to have at least one extinction is higher for the sequential invasion process with heavy-tailed distribution (see Figure 4.3.12)

The invasion of a competitor species is more likely to affect the species present in the system (a more detailed study is made in section 4.3.3) and cause extinction cascades because species invasions cannot ever cause extinctions of species with lower colonization rates. This phenomenon depends mainly on the density of the distribution of colonization rates, a trade-off between a strong density on the left and a strong tail on the right must be found.

We are also interested in understanding the distribution of the number of extinctions following an invasion for a fixed number of invasions. In a rather naive approach, we compare the distribution of the number of extinctions following an invasion (via sampling of many systems) in the figure 4.3.13 to a Poisson distribution. The match is not perfect but gives an idea of the distribution to be explored. The identification works much better in the case of a Pareto distribution than in the case of a uniform distribution where the distribution of the number of extinctions is more strongly right-skewed than expected.

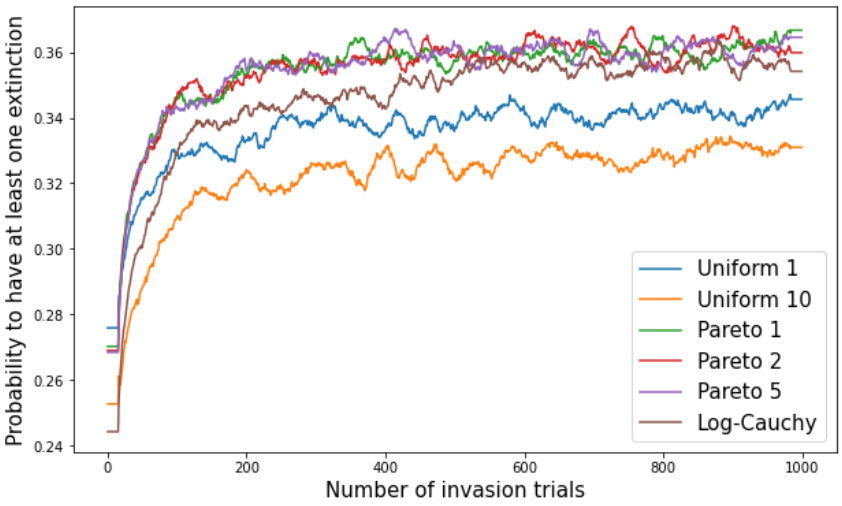
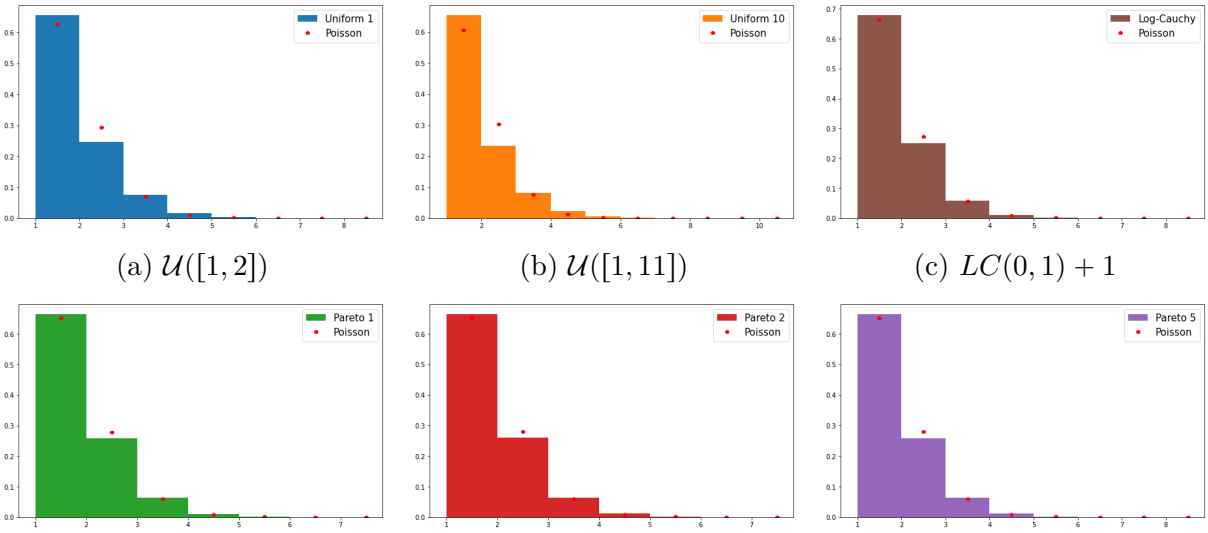


Figure 4.3.12: Probability to have at least one extinction as a function of the invasion trials. We realize a Monte Carlo experiment for a system of size $n = 1000$ and perform the sequential invasion process $P = 1000$ times.



(a) $\mathcal{U}([1,2])$ (b) $\mathcal{U}([1,11])$ (c) $LC(0,1)+1$

(d) $\mathcal{P}(1)$ (e) $\mathcal{P}(2)$ (f) $\mathcal{P}(5)$

Figure 4.3.13: Histogram of the number of extinctions for a fixed number of invasion $n = 1000$ computed for a large number of sequential invasion process $P = 500$. The red point corresponds to the Poisson distribution with a estimated mean.

Another important issue is the dynamics of the amplitude of the cascades. The saturation phenomenon indicates the increase of the regularity of the inputs/outputs of the species in the system. Figure 4.3.12 confirms this hypothesis with a saturation of the cascades phenomena with each iteration. The behavior of the Pareto distributions are similar. Like the general system, and independently of the distribution, the probability to have at least one extinction seems to stabilize or converge to a state that is self-regulating.

Impact of the extinction rate

The impact of the extinction rate is similar as what was observed under the "all-at-once" implementation of the model. Assume $c_{\min} = 1$, a representation of the species richness as a function of the extinction rate is displayed in Figure 4.3.14. Independently of the distribution of the colonization rate of the new invaders, the threshold remains at $m = c_{\min}$.

The most striking example is the effect of the extinction rate on the Log-Cauchy distribution, it is a distribution which tends to take values very close to $c_{\min} = 1$, the change of species richness around the extinction rate $m = 1$ is important. For the Pareto distribution, at the extinction rate $m = 1$, the dynamics is the same independently of the parameter that we choose. However, we observe that a lower parameter a increases coexistence of the species. The left density of the Pareto function is weaker when the parameter is small and the tail of the distribution is increasingly heavy. Consequently, a change in extinction near $c_{\min}$ will less affect a heavier-tailed Pareto distribution. To conclude, the intermediate disturbance hypothesis [Has80] is satisfied in the sequential invasion process as in the Top down process.

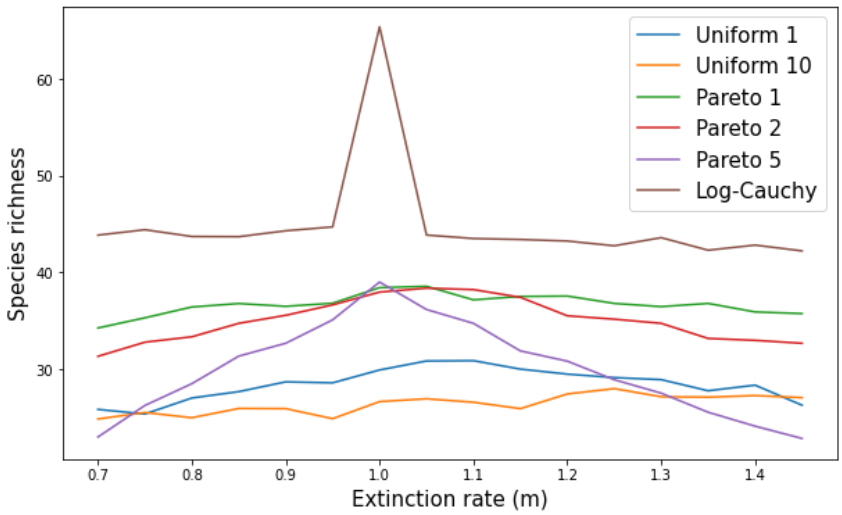
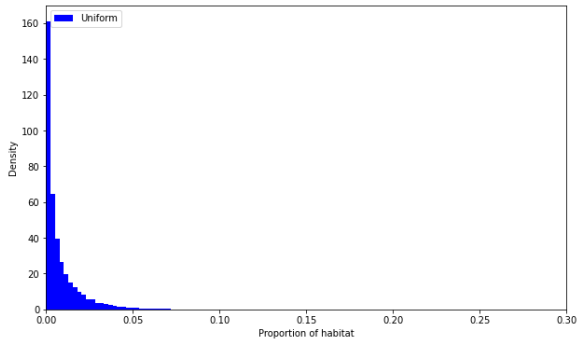


Figure 4.3.14: Representation of species richness as a function of extinction for standard distributions. The threshold of the extinction rate is at $m = c_{\min} = 1$.

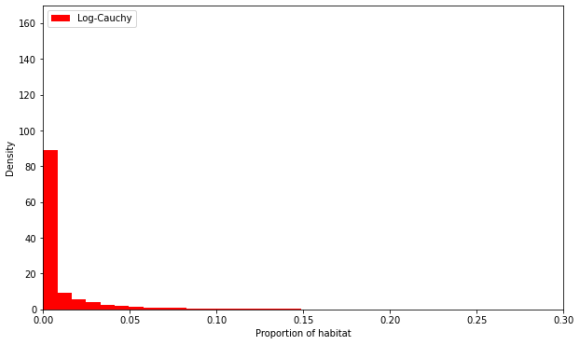
Species occupancies

Complementary information to the number of species in the system is their occupancies $\boldsymbol{p}^*$. The distribution of the colonization rate has an effect on species richness, we notice that it can also have an effect on species occupancies.

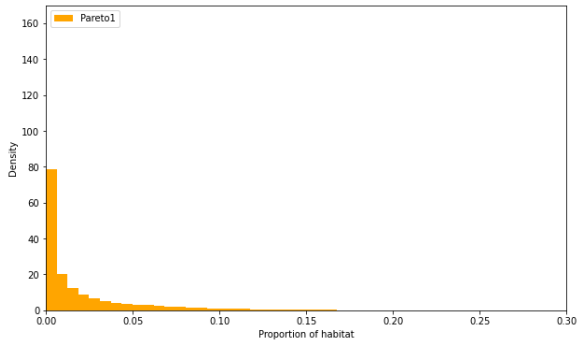
A first noteworthy observation (see Figure 4.3.15) is the difference between the uniform distribution and the 'heavy-tail' distributions. On the one hand, in Figure 4.3.15(a), the distribution is heavily positively skewed whereas few species have a large proportion of habitat. On the other hand, in Figure 4.3.15(b-d), (heavy-tail distributions) some species occupy a large part of the habitat. However, we do not notice any difference between the different distributions.



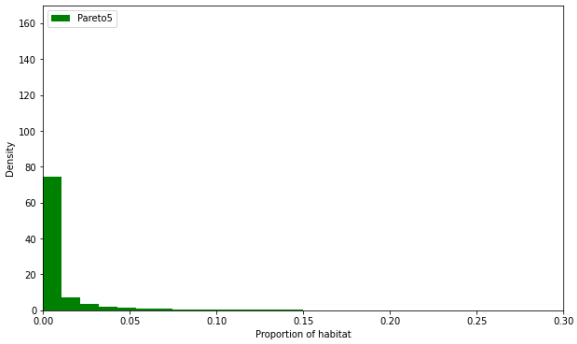
(a) $\mathcal{U}([1,2])$



(b) $LC(0,1)+1$



(c) $\mathcal{P}(1)$



(d) $\mathcal{P}(5)$

Figure 4.3.15: Representation of a histogram of the occupancy for a fixed number of invasions $n = 1000$. The histogram is derived using Monte Carlo simulations by computing $P = 100$ times the algorithm 2 and recovering the associated proportion of habitat of each final system.

In the all-at-one metacommunity process, Kinzig *et al.* [KLD+99] proved that species occupancies were related to their colonization rates through a power law.

In the sequential invasion implementation of the model, the problem appears more complex. To analyze species occupancies, we observe in Figure 4.3.16 the amount of space taken by a species according to species rank i.e. species are ranked according to their colonization rate. For the Log-Cauchy distribution, we observe a phenomenon of "hump" which seems also present in the uniform distribution. This "hump" phenomenon indicates that species that are good competitors and good colonizers at the same time occupy the most space. The comparison of the different parameters of the Pareto distribution is also very interesting: the weaker the parameter (heavier the tail), the weaker the density on the left. A low density on the left indicates a faster increase in species occupancy with species colonization rank while conversely, a late increase is due to a higher density on the left. The remarks on the Pareto example can be transposed to the case of the uniform distribution.

Remark 4.6. A related problem is the distance of each colonization rate and its below niche shadows threshold. The density of the spacing could give a more precise idea of this hump.

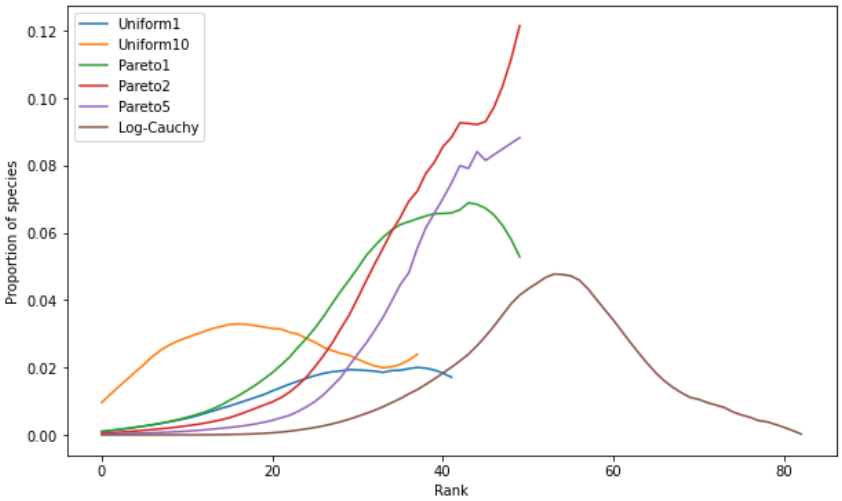


Figure 4.3.16: Plot of the occupancy as a function of the rank of colonization rate at iteration $n = 1000$. We use a Monte Carlo experiment $P = 1000$ and average the occupancy values for each rank. For the last ranks, we truncate the graph if we do not have a sufficient number of values (i.e. ¡10) to obtain an average.

Dynamics of the fraction of empty patches

Previously, we were interested in species occupancies according to their colonization rates. However, an important question is the amount of space that the whole community will occupy. In Figure 4.3.17, we represent the dynamics of the fraction of empty patches during the sequential invasion process. A first observation is that the fraction of empty patches seems to converge to a given limit for any distribution of colonization rate. Furthermore, the convergence towards the final fraction of empty patches seems fast, in 200 invasions, we have a precise idea of the remaining fraction of empty patches.

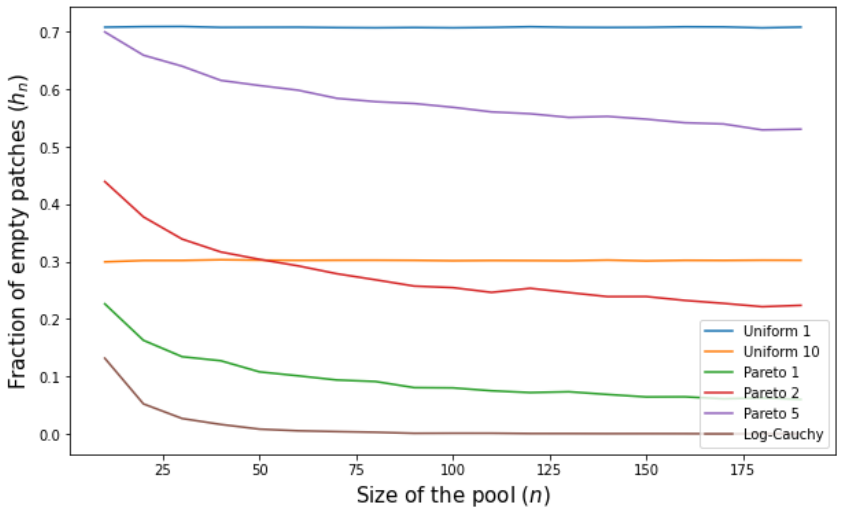
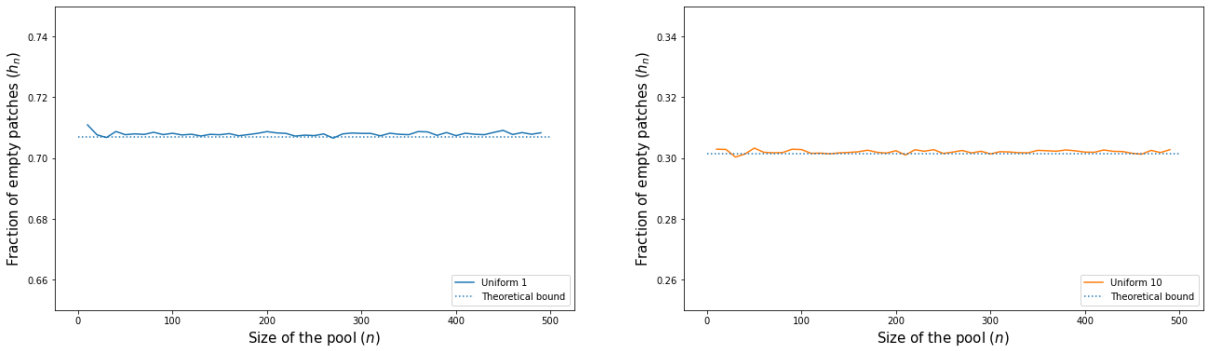


Figure 4.3.17: Representation of the fraction of empty patches as a function of the number of invasions. We use a Monte Carlo experiment $P = 1000$ and average the fraction of empty patches values at each iteration.

In the all-in-once metacommunity process, equation (4.10) describes precisely the frac-

tion of empty patches h_n when n becomes large. From Figure 4.3.18, independently of the distribution, the estimate seems to be correct in the sequential invasion process. The estimate obtained analytically under the all-at-once implementation of the model (h_n) seems to be correct also when the model is implemented using sequential invasions.



(a) $\mathcal{U}([1,2])$ (b) $\mathcal{U}([1,11])$

Figure 4.3.18: Comparison between the empirical and theoretical bound of the fraction of empty patches (h_n) as a function of the size of the initial pool. In Fig. (a) the uniform distribution $\mathcal{U}([1,2])$ is plotted, the theoretical estimation is computed by $1/\sqrt{2}$. In Fig. (b) the uniform distribution $\mathcal{U}([1,2])$ is plotted, the theoretical estimation is computed by $1/\sqrt{11}$.

If we compare the fraction of empty patches to the species richness in the system, the Pareto distribution of the colonization rate is a good example of less fraction of empty patches does not necessarily imply richer communities: the increase in the number of species in the Pareto case $a = 1$ and $a = 5$ is similar and yet the fraction of empty patches is very different. The uniform distribution with a larger support has less fraction of empty patches. Note again that in the sequential invasion process, what matters is the colonization rate of the largest species present in the system.

Development of the habitat diversity

In addition to species richness, we want to describe the dynamics of the system with other measures of diversity. We introduce two standard measures of diversity [Whi72, Jos06], the Shannon index defined by

$$H' = -\sum_{i=1}^{n} \frac{p_i}{1-h_n} \log\left(\frac{p_i}{1-h_n}\right), \tag{4.15}$$

where h_n corresponds to the empty space $h_n = 1 - \sum_{i=1}^{n} p_i$. For comparison purpose with species richness, the Hill number of order 1 will be considered $e^{H'}$. The second diversity index is the inverse Simpson index (:= Hill number of order 2)

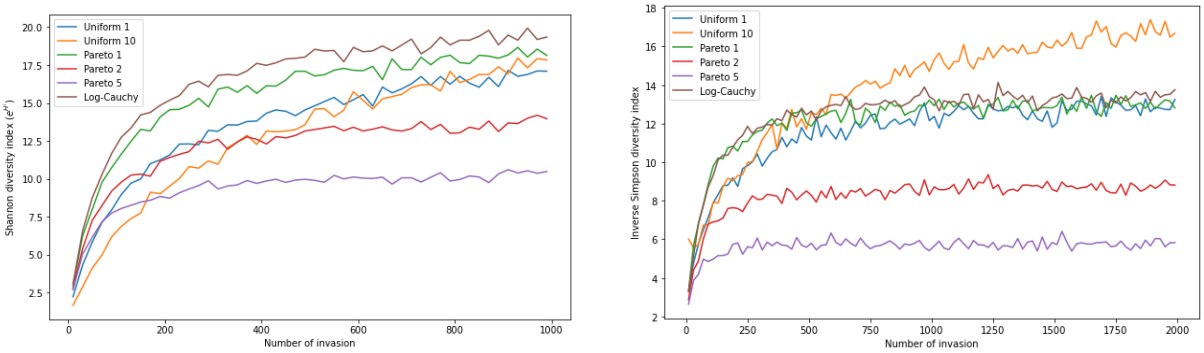
$$\frac{1}{\lambda} = \frac{1}{\sum_{i=1}^{n} \left(\frac{p_i}{1-h_n}\right)^2}. \tag{4.16}$$

These two measures of diversity are Hill numbers (Shannon's exponential version and Simpson's inverse), their maximum value is equal to the species richness. These are more

refined measures of diversity than species richness because they take into account the (evenness) of occupancy between species.

Diversity indices do not respond similarly to changes in the distribution of colonization rates. An important marker is that the diversity is greater in the case of the uniform distribution when the right-hand support is larger, whereas in Figure 4.3.9a, the final species richness of the system is lower when the right-hand edge of the uniform distribution support grows. The intuition of this phenomenon is the strong impact of the density on the right. We also observe that Log-Cauchy distribution seems less "out of the box" than for species richness. The behavior of its diversity index is equivalent to Pareto distribution. However, if we compare the different Pareto distributions, the result is clear: the heavier the tail, the higher the diversity.

The uniform distribution appears to have a smaller gap between its species richness and diversity than the heavy-tailed distributions. The heavier the tail of the distribution, the more similar the competitors will be and the larger the colonizers will be and stay in the system.



(a) Hill number of order 1

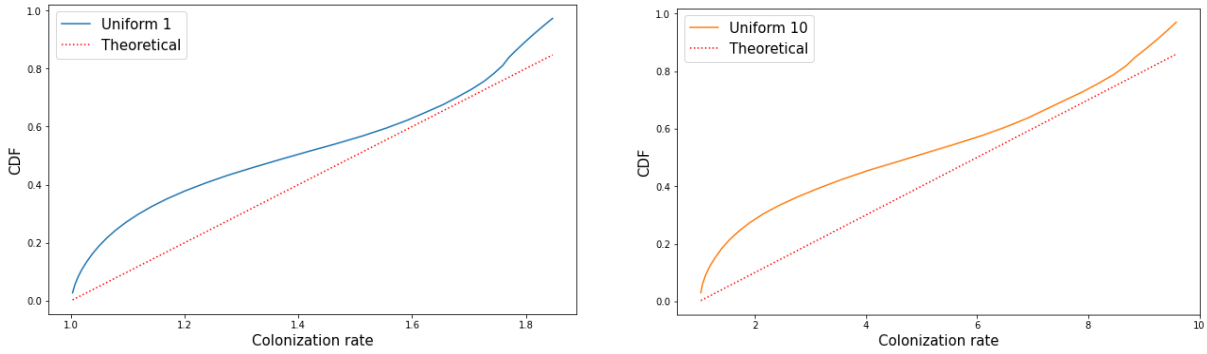
(b) Inverse Simpson index

Figure 4.3.19: Representation of the diversity index of the assembly as a function of the invasions for different distributions. The curve is derived using Monte Carlo simulations by computing $P = 2000$ times the algorithm 2 and averaging the number of persistent species. In Fig. (a), the Hill number of order 1 is plotted using the exponential of H' defined in (4.15). In Fig. (b), the Inverse Simpson index is illustrated from equation (4.16).

Distribution of the final vector of colonization rate

For a fixed number of invasions, we are interested in the distribution of the colonization rates of the persistent species. This information provides an indication of the impact of niche shadows on the distribution of $\boldsymbol{c}$. In Figure 4.3.20, for 500 invasions, the cumulative distribution function (CDF) of the final colonization rates $\boldsymbol{c}$ is represented when the colonization rates of species attempting to invade is drawn from a uniform distribution compared to the theoretical CDF of the uniform distribution. We observe that independently of the support of the uniform law the curvature of the CDF is the same. Moreover, we observe three distinct phases. In a first phase, a rapid increase of the CDF is observed, showing the persistence of many competitive species (with low c). In a second phase, we

have a slowing down of the curve growth, then in a last phase an increase of the slope of the CDF which shows an accumulation of good colonizers.



(a) $\mathcal{U}([1,2])$ (b) $\mathcal{U}([1,11])$

Figure 4.3.20: Representation of the empirical CDF after the sequential invasion of the colonization rate of the persistent species after 500 invasion trials. The curve is derived using Monte Carlo simulations $P = 1000$. In both plots, the red dotted line represents the respective CDF of the initial distribution of the colonization rate.

The example of the Pareto distribution is shown in Appendix 4.D. The CDF is characterized by initially rapid growth and then significantly slow growth. The shape is similar to a CDF of the standard Pareto distribution.

4.3.3 Properties of the invaders

Previously, we focused on the study of the general dynamics of the system over time: richness, diversity, extinction cascades, etc. This system depends on invasions of new species. In this section we are specifically addressing the properties of invading species over time.

Probability of invading the system

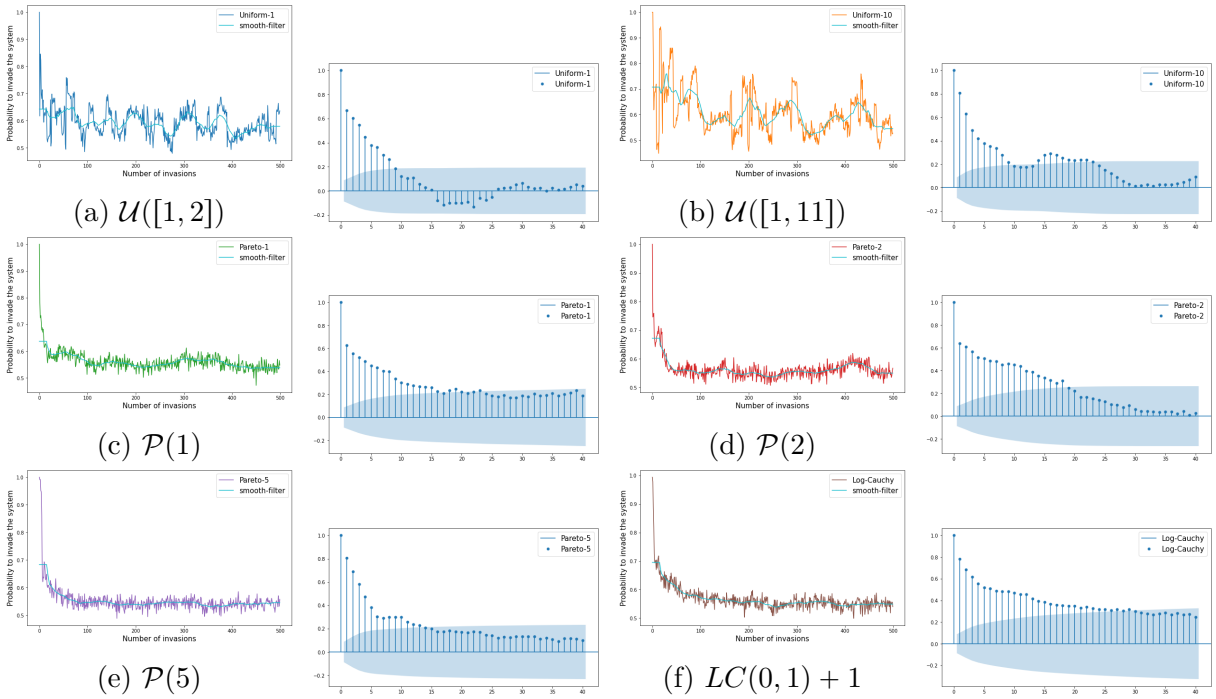
A first natural question is whether it is increasingly difficult to invade the system as it builds up. Moreover, considering the lth invasion of the system, will it be more difficult to invade the system at the next iteration (fluctuation of the invasion probability)?

The probability to invade the system at an iteration l is defined in equation (4.12). This probability changes with time depending on the state of the system $\mathcal{F}$ and the distribution of the colonization rates of the persistent species. The complexity of the random failure zone $\mathcal{F}$ restricts us to a numerical computation of this integral over the iterations. We represent the numerical results for the different distributions in figure 4.3.21 as well as the auto-correlation of each time series.

In general, unlike the all-at-once metacommunity process, we observe a probability that seems slightly higher than 0.5. This may be due to a size effect since the richness of the system is small and asymptotically, we may expect a convergence toward 0.5. The invasions remain frequent during the dynamics of the system and it is mainly the extinction cascades which "drive" the dynamics of the general system. We have seen that the extinction cascades are frequent.

In Figure 4.3.21, two distinct behaviors appear between heavy-tailed variables (Pareto, Log-Cauchy) and uniform variables. In the case of heavy-tailed variables, we observe a convergence of the invasion probability with little noise. Conversely, in the uniform case, the dynamics of the system does not seem completely clear with a lot of noise, even though a trend is emerging.

However, a common feature of the sequential invasion processes is the positive correlation between the number of invasion attempts and the invasion probability.



(a) $\mathcal{U}([1,2])$ (b) $\mathcal{U}([1,11])$

(c) $\mathcal{P}(1)$ (d) $\mathcal{P}(2)$

(e) $\mathcal{P}(5)$ (f) $LC(0,1)+1$

Figure 4.3.21: Dynamics of the probability of a new species to invade the system as a function of the number of invasions. We distinguish the 6 cases of standard distributions (a)-(f). The noisy curve corresponds to an average of several sequential invasion processes. In each graph, the blue curve corresponds to a smoothing filter. In each case, the right plot is an autocorrelation function of the left plot.

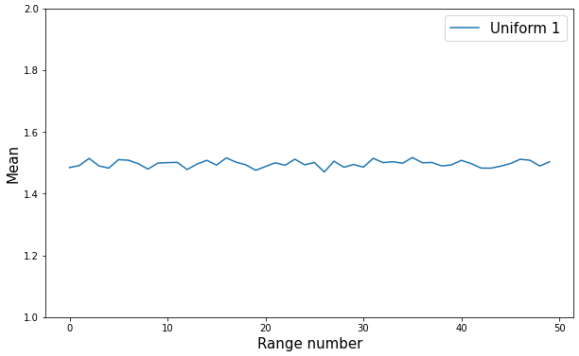
Remark 4.7. See appendix 4.D for an alternative method of computing the probability of invasion.

Feature of an invader

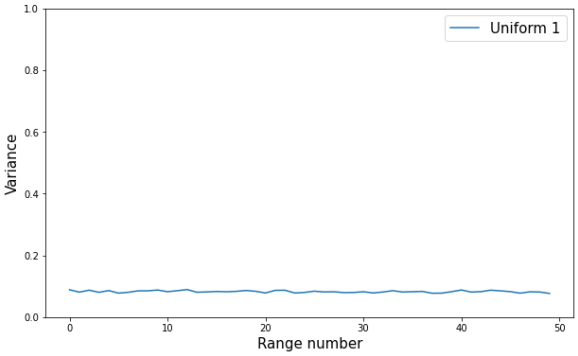
We are interested in the properties of the invaders. In particular, we would like to know as the number of invasion trials becomes large and the number of species in the system increases whether the invaders have particular features. The numerical method used is for a given system, we compare the mean and variance of the colonization rates $\boldsymbol{c}$ of the invaders at different time intervals (see Figure 4.3.22).

Remark 4.8. A second numerical method is presented in Appendix 4.D by performing a moving average on the realized colonization rates and average the time series for different sequential invasion processes.

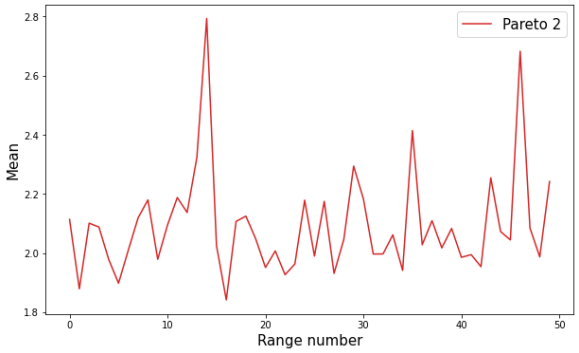
With both methods, we observe similar trends. We can classify the results in two categories: the distributions of initial c values with finite mean and variance and the



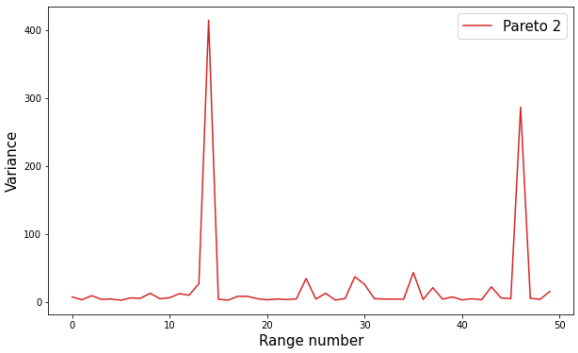
(a) Mean $\mathcal{U}([1,2])$



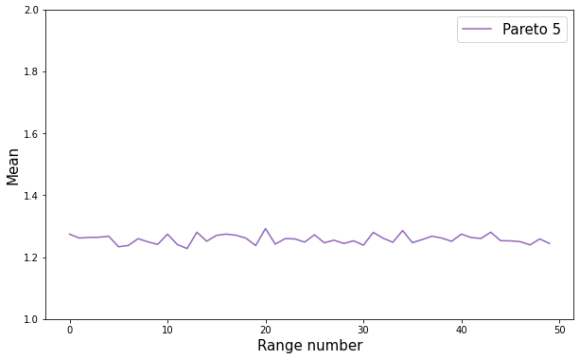
(b) Variance $\mathcal{U}([1,2])$



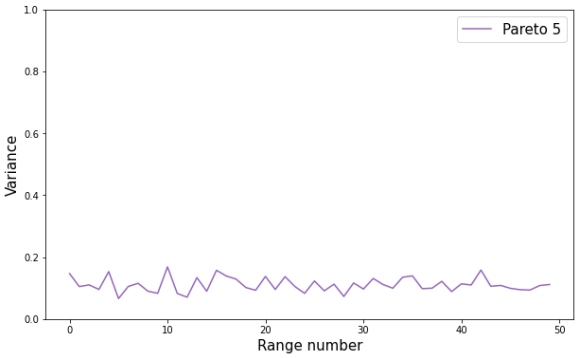
(c) Mean $\mathcal{P}(2)$



(d) Variance $\mathcal{P}(2)$



(e) Mean $\mathcal{P}(5)$



(f) Variance $\mathcal{P}(5)$

Figure 4.3.22: Representation of the mean and variance of the colonization rates of the invaders at different time interval. We consider an sequential invasion process of size $n = 50000$, which we split into a set of intervals of size 50 $[0, 1000] \cup [1000, 2000] \cup [2000, 3000] ... \cup [49000, 50000]$. In each interval, the mean and variance of the colonization rates of the invaders are computed. The plots (a)-(b) correspond to the uniform distribution $\mathcal{U}([1,2])$, the plots (c)-(d) correspond to Pareto $a = 2$, the plots (e)-(f) correspond to Pareto $a = 5$.

distributions with infinite variance i.e. Pareto with parameter $a \leqslant 2$ or the Log-Cauchy distribution.

In the case of finite-variance distributions (see Figure 4.3.22-a,c), the properties of the invaders remain constant with little noise over the iterations. The mean and variance of the colonization rates of the invaders remain constant over time, the profile of the invaders

does not change over the iterations.

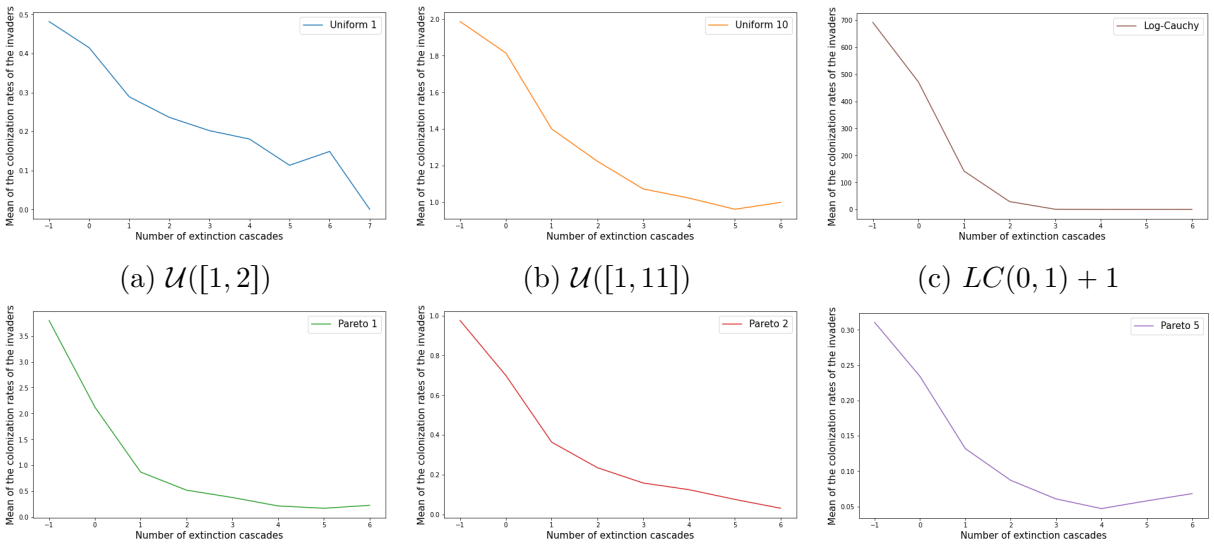
In the case of a distribution with infinite variance (see Figure 4.3.22-b), the mean and the variance the colonization rates of the invaders fluctuate over time with "peaks" which correspond to the cases where species with very large colonization rate invade the system. We observe that the fluctuation remains very strong despite the peaks. This fluctuation will be even stronger in cases where the mean also tends to infinity (for example Log-Cauchy). However, there is no "trend" in the sense that the peaks seem to appear randomly and the variations behave like noise.

To conclude, regardless of the distribution of the initial colonization rates, the profile of the invaders seems to remain constant over the iterations. However, during an invasion, the system tends to change, the type of invader can play an important role. We quantify its impact in the next section.

Impact of an invader

When a species invades the system, we have seen in section 4.3.2 it can involve extinction cascades. These extinctions will depend on the general colonization parameter distribution but also on the colonization parameter of the invader. In Figure 4.3.23 the mean value of the colonization rate of the invader is displayed for each extinction cascade effect.

Regardless of the distribution, we notice simply that the lower the colonization rate, the greater the chance of implying a extinction cascade effect. From an ecological point of view, the idea is that if the colonization rate is low, the species is very competitive and will displace the species already present in the system. We notice that the difference in scale is due to the distribution of the colonization rate, the bending of the curve will be greater than the right tail of the probability density is light (difference between U1 and U10 and P1, P2, P5).



(a) $\mathcal{U}([1,2])$ (b) $\mathcal{U}([1,11])$ (c) $LC(0,1)+1$

(d) $\mathcal{P}(1)$ (e) $\mathcal{P}(2)$ (f) $\mathcal{P}(5)$

Figure 4.3.23: Mean of the colonization rates of the invaders as a function of the number of extinctions cascades it creates. This experiment is done only once in a large sequential invasion process $n = 200000$.

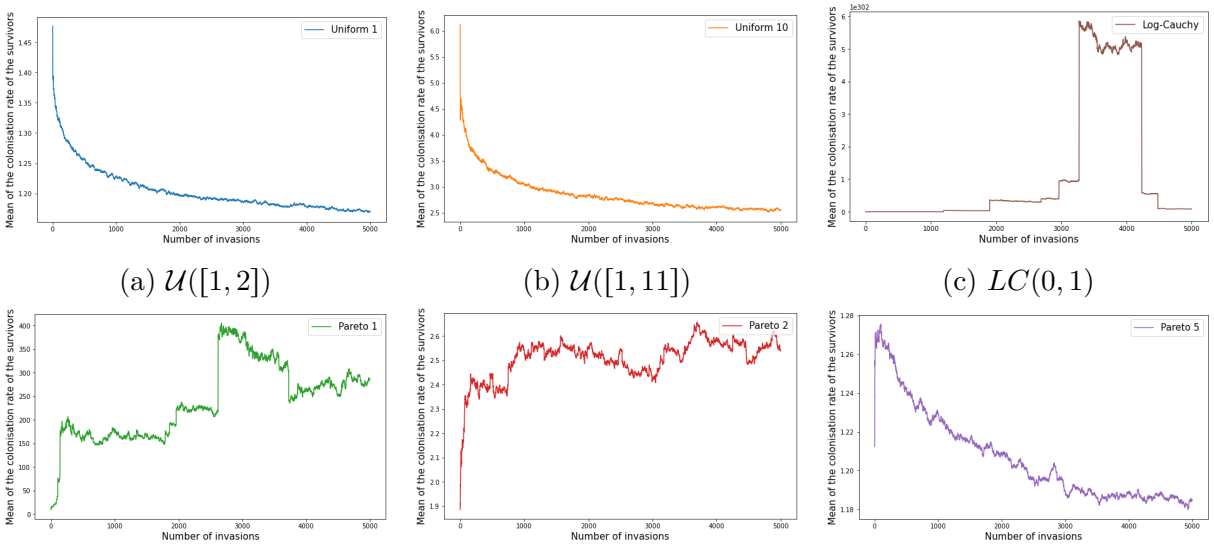
4.3.4 Properties of resistant species

The main focus has been on describing the properties of the invaders. When a species invasion occurs, a possible scenario is the extinction cascade: one or more species become extinct as a result of the invasion. In this type of iterative process, what is the ability of a species to resist to invasions? In this part, we quantify empirically, which species (in terms of colonization value) are more likely to "resist" these invasions. We know that competitors are more resilient, we expect high c values to come and go faster than low values. However, we suspect that it is difficult to stack up a lot of small values of c, so we would finally like to quantify the distribution of the c's that resist as a function of time, in order to answer the question: is there a growing fraction of the colonization rate support that resists invasions?

Properties of the colonization rate of the persistent species

General metrics for studying the dynamics of persisting species is the average (see Figure 4.3.24) and median (see Figure 4.3.25-b) of the colonization rates of persisting species as a function of invasion trials.

We distinguish three different behaviors: when the distribution of colonization rates of the species attempting to invade admits a finite variance, the mean converges to 1. This indicates that the proportion of highly competitive species tends to grow in the system and that colonizing species do not seem to stay in the system "on average". When the mean and variance of the distribution is infinite (Pareto 1, Log-Cauchy), we have jumps in the mean of the survivors. This suggests that a highly colonizing species enters the system and is not displaced by more competitive species for a significant amount of time. The last case is infinite variance but finite mean distribution (Pareto 2), the trend seems to be increasing even if the curve is noisy.



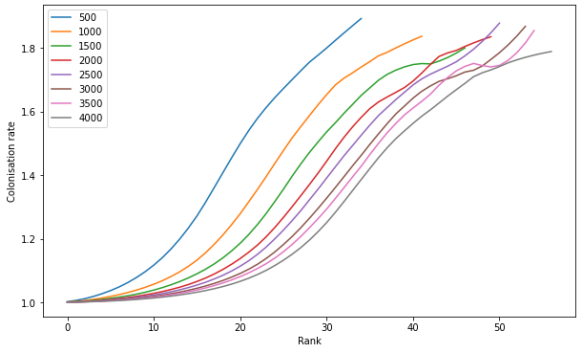
(a) $\mathcal{U}([1,2])$ (b) $\mathcal{U}([1,11])$ (c) $LC(0,1)$

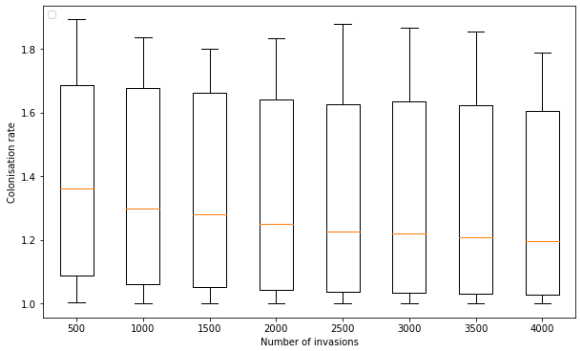
(d) $\mathcal{P}(1)$ (e) $\mathcal{P}(2)$ (f) $\mathcal{P}(5)$

Figure 4.3.24: Mean of the colonization rate of the persisting species as a function of the number of invasions trials. The curve is derived using Monte Carlo simulations $P = 1000$ of many sequential invasion process.

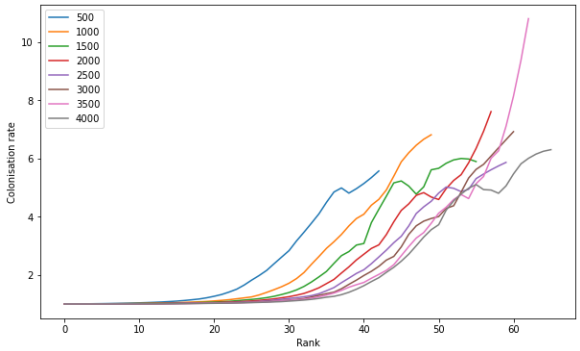
In a second step, we look at the curve of colonization rates as a function of rank for a different number of invasions trials (see Figure 4.3.25). For each of these curves, we look at a box plot to see the distribution of colonization rates.

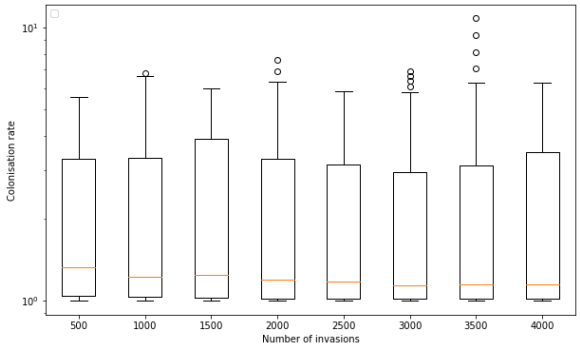
We distinguish again two cases: the case of the regular distribution (uniform 1) and the case of a heavy-tailed distribution with a finite mean and variance (Pareto 3). In the case of the uniform distribution, we again observe a decrease in the colonization rates of the persistent species as the number of invasions trials increase, the mass of colonization rates approaches 1. However, we do not observe a reduction of the values of colonization present in the system. We observe a similar behavior for the heavy-tailed distribution with a decrease of the median without much change for the colonizing values (we remain in a case where the variance and the mean are finite).





(a) Uniform $\mathcal{U}([1,2])$





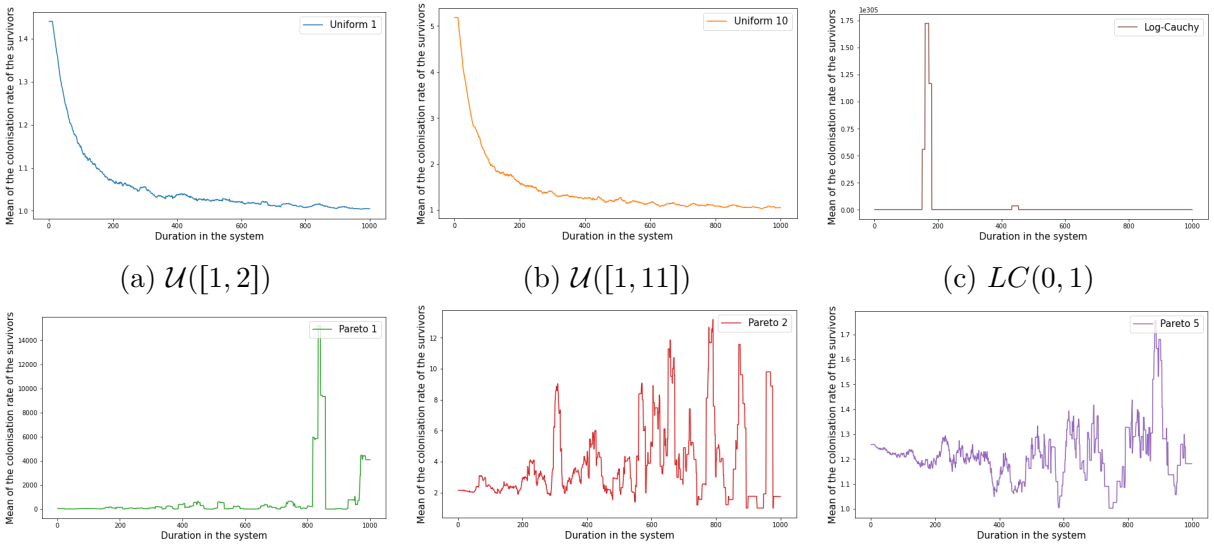
(b) Pareto $\mathcal{P}(3)$

Figure 4.3.25: The left plots are representations of the colonization rates as a function of colonization rank for a different number of invasions. For each of these curves, on the right plots, we compute a box plot which gives additional information on the distribution of colonization rates. A Monte Carlo experiment is realized for $P = 1000$ sequential invasion processes to obtain the functions on the left plots.

Lifetime of a species in the system

It was observed that under some distributions, competitive species seem to become a majority in the system while in other cases a pool of colonizing species tends to remain in the system. Another way to illustrate this phenomenon is to study the average time a species survives in the system.

In Figure 4.3.26, we represent the mean colonization rate $\overline{c_{age}}$ of species persisting till age. In the case of regular distributions, we observe that long-lived species have a low average c while short-lived species have a higher average c. In the case of heavy-tailed distributions, colonizers seem to persist longer in the system. Numerically, the large values taken by the heavy-tailed distributions can cause some issues in the graphical representations, there is not really a relation between lifetime and colonization rates for heavy-tailed distribution.



(a) $\mathcal{U}([1,2])$ (b) $\mathcal{U}([1,11])$ (c) $LC(0,1)$

(d) $\mathcal{P}(1)$ (e) $\mathcal{P}(2)$ (f) $\mathcal{P}(5)$

Figure 4.3.26: Representation of the mean value of colonization rate as a function of its lifetime in the system for $n = 1000$ invasions.

Discussion

Analyzed by Tilman [Til94], the HT model (competition-colonisation trade-off with hierarchical competition) represents a first step in our understanding of species-rich metacommunities. In this paper, we bring a new perspective to the HT model (4.2) by studying the behavior of the model when colonization rates are randomly distributed and a large number of species is considered. The question is whether species can coexist to form a metacommunity with many species under the strict conditions of this model. To carry out this investigation, we identified two different types of sequential invasion processes:

1. a all-at-once metacommunity process starting from a initial pool of species in the landscape or from an equivalent point of view starting from an empty system with an invasion sequence where it is assumed that all species can invade at any time,

2. a sequential invasion process developing an invasion sequence that involves a historical contingency effect.

First, from a mathematical standpoint, we are interested in the behavior of the system of n equations (4.2). As already mentioned by Calcagno *et al.* [CMJD06a], there is an equivalence between the C-C trade-off model and the Lotka-Volterra model (4.4). From

a new perspective, using the Theorem of Takeuchi *et al.* [TAT78], we show that there is a unique globally stable equilibrium of system (4.2). This important aspect makes it possible to overlook transient dynamics in the study of model (4.2) for both types of implementations. We recall the conditions of coexistence depending on the colonization rates in the form of a set of admissible conditions (4.3). The number of persistent species at equilibrium $\boldsymbol{p}^*$ is computed simply using this set. Like Hastings [Has80], we restrict our analysis to a fixed extinction rate equal for all the species, which simplifies the computations.

All-at-once metacommunity

Intuitively, in light of the quadratic conditions of the set of admissible solutions (4.3), one could assume that the coexistence of many species is complicated. However, we show that when n is large enough, then the number of persistent species is distributed according to a binomial distribution (Figure 4.2.2), and that half of the species coexist (Figure 4.2.1). Furthermore, we show that coexistence does not depend on the distribution of colonization rates in the all-at-once metacommunity. In particular, we can consider very different colonization distributions: uniform, Exponential, Pareto, Log-Cauchy, etc.

In a second part, we study the properties of persistent species by recalling some results of Kinzig *et al.* [KLD+99]. The dynamics of the fraction of empty patches depends only on the colonization rate of the best colonizer in the system. A distribution with bounded support may let a greater fraction of empty patches, whereas if the support of the distribution is large, the empty space will be filled. In addition, the distribution of the occupancies of each species as a function of its colonization rates follow a power law $p \propto c^{-3/2}$ when c is drawn from any uniform distribution. However, for heavier-tailed distributions, the best colonizers will also be able to occupy a large fraction of the available patches. As a final point, we clarify the intermediate disturbance hypothesis that had been observed by Hastings [Has80] where optimal coexistence between species occurs when $m = c_{\min}$ i.e. the left bound of the support of the different distributions (Figure 4.2.7).

Sequential invasion process and extinction cascades

Compared to the all-at-once approach, the sequential invasion process is much more restrictive i.e. diversity accumulates much more slowly when species cannot re-invade without a concomitant invader to help the new species. The number of persisting species saturate with a logarithmic growth due to historical contingencies and extinction cascades (Figure 4.3.9). In contrast to the all-at-once metacommunity process, there is no universality result under this implementation, every result depends on the distribution chosen to draw the colonization rates of species attempting to invade. The dynamics of the species richness depends on the distribution of the colonization rate of the new invaders. Regular distributions of colonization rates such as the uniform distribution have a worse coexistence probability than heavier-tailed distributions. The phenomenon of extinction cascades is a key element of the saturation phenomenon (Figure 4.3.11). Its study allows us to understand the difference in performance between the distributions. With each iteration, the phenomenon of extinction cascades also seems to stabilize. We reach a system where the inputs/outputs seem to compensate one another independently of the distribution. The impact of the extinction rate is not the most important element,

however the intermediate disturbance hypothesis still drives the impact of the extinction rate (Figure 4.3.14). Species occupancies are affected by the the distribution of the colonization rates. On the one hand, a major difference is observed between regular and heavy-tailed distributions. On the other hand, when we study the occupancy of species by their colonization rank, we observe that the greater the density of the distribution on the left, i.e. there are more competitive species trying to invade the system, the lower the proportion of competitive species. In the extreme case of the Log-Cauchy distribution, we observe that there is a balance of species importance with small and large colonization rates to occupy a larger part of the patches. An additional point on the distribution of landscape is the fraction of empty patches, the conclusion is unambiguous: the heavier the tail the smaller the fraction of empty patches (Figure 4.3.17). The fraction of empty patches is similar to the all-at-once metacommunity process (Figure 4.3.18). In addition to the study of species richness, two diversity indices (Hill number of order 1, Inverse Simpson index := Hill number of order 2) are studied. The behavior of diversity is the same as species richness. On the one hand, the diversity of species that persist when colonization rates are drawn into the uniform distribution is high; on the other hand the diversity of species that persist when colonization rates are pulled into the Log-Cauchy distribution is significantly lower. In general, the heavier the tail, the greater the diversity (Figure 4.3.19).

Independently of the distribution, we can make several general observations, first of all the probability of invasion decreases with time to converge to a value close to $\frac{1}{2}$ (Figure 4.3.21). The properties (mean, variance) of the colonization rates of the invaders is constant over the iterations and depends intrinsically on the initial distribution (Figure 4.3.22). Ultimately, the extinction cascade is greater when the invader has a colonization rate that is low and therefore the more competitive a species, the stronger its impact on the extinction of other species (Figure 4.3.23). However, there are some notable differences between regular (uniform) and heavy-tailed (Pareto, Log-Cauchy) distributions. First, the probability of invasion is much less noisy in the case of heavy-tailed distributions. This can be explained by the lower number of extinction cascades and thus the system changes less over time. Secondly, the distribution of invaders seems to remain constant when colonization rates are drawn following a regular distribution while in the other case, the properties of the invaders fluctuate a lot over time even if the trend remains constant. The reason is the very small and very large values that the colonization rates can reach and this has an effect on the mean colonization rate which implies extinction cascades.

In the case of regular distributions of colonization rates, the species that appear to remain in the system are the most competitive (Figure 4.3.24). There is a system that becomes more and more competitive and few colonizing species manage to stay in the system (Figure 4.3.25). This could explain the small increase of species in the system. Finally the community of persistent species keeps a smaller and smaller fraction of the species that try to invade the system. Conversely, heavy-tailed distributions behave differently, with colonizing species managing to maintain themselves even if the competing species remain key species in the system (Figure 4.3.26).

To conclude, a partial answer is that the species that will tend to stay in the system the longest are the competitive species because they are less subject to extinction cascades. On the one hand, if there are many competitive species in the system, the competitive species will have more difficulty to invade the system. On the other hand, the colonizing species will succeed in invading the system without changing (much) its structure (which

is facilitated by a heavy-tailed density function). However, when a competitive species succeeds in invading the system, it will tend to push out some of the best colonizing species. Finally, we find that a good balance between competition and colonization promote coexistence i.e. if too many species are competitive, then no one can invade the system and conversely too many colonizers gives a more volatile system.

Perspectives

While keeping a probabilistic structure of the $\boldsymbol{c}$ colonization rates, a first natural perspective would be to introduce more complexity to model 4.2, such as new parameters, to understand their impact on the all-at-once metacommunity and the convergence to the binomial distribution or the sequential invasion process. In this paper, a fixed extinction rate has been assumed for all species $m_i = m$, $\forall i \in [n]$. However, one could consider a different extinction rate as in Tilman's historical paper [Til94] and observe its effect. The impact of an multimodal disturbance factor as in Liao *et al.* [LBB22] could give different results. One could also study the impact of habitat loss as in the paper by Nee and May [NM92], Tilman [TMLN94] or more recently Li *et al.* [LBL20]. In this paper, explicit spatial structure linking patches is not considered, we assume a global dispersal ability. In nature, habitats are structured [HO00, OH01] i.e. paths from one patch to another can be indirect, adding spatial structure would undoubtedly affect our quantitative results [GB12, WBvN$^+$21], as in Zhang *et al.* [ZBN$^+$21] where they showed that network heterogeneity (i.e. variation in the number of links between patches) promotes species coexistence in hierarchical competitive communities.

The major extension would be to monitor the effects of a more general version of the hierarchical competition-colonization trade-off model (4.2). Indeed, this model has been criticized for being based on unrealistic assumptions. In particular, this model assumes that worse colonizers always out-compete better ones, regardless of how similar their colonization rates i.e. perfectly asymmetric competition. Calcagno *et al.* [CMJD06a] studied the competition-colonization trade-off model 4.1 relaxing the hierarchical competition assumption assuming that there is preemption (:= a species present in a patch has an ability to resist invasion of a new species). They found that the effects of these changes were more nuanced, but did not fundamentally alter the ability of the trade-off to maintain coexistence. Calcagno *et al.* showed that relaxing this assumption can actually favor coexistence. This is especially likely when competition is not perfectly asymmetric, suggesting that these assumptions might "cancel out" to some extent.

Although in the study of the sequential invasion process, we consider only the colonization rate of the invader sampled from the same distribution. We could consider that the distribution of colonization rates change due to environmental conditions. If the distribution of invaders change, habitat and persistent species in the system will be subject to new constraints. We could measure the "resistance" of the different distributions against others and therefore give arguments to an optimal distributions. We could choose a metric between the two distributions and see how it changes over time: speed of growth, extinction cascades increase and arguments of the kind (competitor/colonizer) has a better resistance in the system.

Similarity with the Lotka-Volterra model has been already suggested by Calcagno *et al.* [CMJD06a]. Surprisingly, the behavior of a standard Lotka-Volterra random system is the same. Servan *et al.* [SCG$^+$18] showed that when interaction coefficients are sampled

independently from a symmetric distribution with mean zero, the number of coexisting species from a pool of size n is also distributed as $B(n, \frac{1}{2})$. They considered a scenario where all species are statistically equivalent and interactions are essentially unstructured, while we consider communities with strongly hierarchical interactions. The fact that the final distribution of the persistent species are identical in these highly dissimilar cases suggests the possibility that this behavior is a more general result of ecological dynamics where species possess some symmetry by emerging from ecological mechanisms.

Stability in model (4.1) is still an open-question. In the LV form (4.4), the interaction matrix is similar to an anti-symmetric predator-prey matrix [YA73, RW84]. From a theoretical standpoint, the diversity-stability debate has not found a clear mechanism to propose an alternative to May's results [May72]. Spatial dynamics and the study of patch dynamics could be a possible alternative for the stability of ecosystem dynamics and provide an answer to the stability paradox.

Last but not least, a rigorous proof in the case of the all-at-once metacommunity process could help to improve our understanding of the universality mechanisms associated with the model. However, the task is not simple due to the inerrant nonlinear conditions associated with the set of conditions of an admissible solution (4.3).

Appendix

Numerical methods

Simulations were performed in Python.

4.A Theoretical background

4.A.1 Reminder of Tilman's article

The following equations are reminders of the ones obtained by Tilman (1994). At equilibrium, the species i will occupy p_i^* proportion of habitat

$$p_i^* = 1 - \frac{m_i}{c_i} - \sum_{j=1}^{i-1} \left[p_j^* (1 + \frac{c_j}{c_i}) \right] .$$

Remark 4.9. Denote by h_i the proportion of empty space left when i species are present in the model.

- If the extinction rate of every species is equal to $m_i = 0,\ \forall i \in [n]$, regardless of the colonization rate of the other species, it is always the most competitive one that will remain in the end and colonize the whole space i.e $p_1(t) \xrightarrow[t \to +\infty]{} 1$.
- If we develop the above equality, we obtain an upper bound for each species: $\forall i \in [n],\ p_i^* \leqslant h_{i-1} - \frac{m_i}{c_i}$. Even if a species is very colonizing, it will never be able to colonize more than the space it has been given. This condition illustrates the regulation of competition.

Apart from the number of species and whether the extinction rate is superior than one, at equilibrium there is always empty space. It may be briefly recalled in Levins one-species model, the fraction of empty patches corresponds to $1 - \frac{m}{c}$. The fraction of empty patches when i species are present is given by

$$h_i = 1 - \sum_{j=1}^{i} p_j^* = \frac{m_i h_{i-1} + \sum_{j=1}^{i-1} p_j^* m_j}{c_i h_{i-1}} .$$

The fraction of empty patches is necessarily greater than $\frac{m_i}{c_i}$ and therefore the landscape is necessary occupied by $1 - \frac{m}{c_n}$ (see section 4.2.3 for precise estimation of the fraction of the empty patches).

Selection of the parameter Throughout the paper, the extinction rate of all species is equal to $m_i = m,\ \forall i \in [n]$. The relevance of the paper relates to the choice of the parameter $\boldsymbol{c}$. The possibility of a species to invade the system depends only on its colonization rate and the state of the system. At equilibrium, a species will be able to invade the system and persist if and only if

$$c_n > \frac{m}{h_{n-1}^2}\,, \tag{4.17}$$

where c_n corresponds to the colonization rate of the invading species and h_{n-1} to the fraction of empty patches of the system.

Condition (4.17) is very strong and assumes that each species wanting to invade the system is always less competitive and always more invasive. The condition will be reworded based on the colonization rates of other species. These strong constraints will allow the creation of the all-at-once metacommunity model.

4.A.2 Condition on the colonization rate and occupancy

In this part, we find Tilman's conditions by another approach. Assume the c_i are arranged in increasing order and the dynamics of the species is

$$\frac{dp_i}{dt} = p_i \left[c_i - m - c_i \sum_{j=1}^{i} p_j - \sum_{j=1}^{i-1} c_j p_j \right].$$

The equation that gives the non-trivial equilibrium ($p_i > 0$) can be written as follows

$$1 - \frac{m}{c_i} - \sum_{j=1}^{i} p_j - \sum_{j=1}^{i-1} \frac{c_j p_j}{c_i} = 0\,. \tag{4.18}$$

One way to solve this equation in the case where $p_i > 0,\ \forall i \in [n]$ is to introduce the fraction of empty patches h_i defined by

$$h_i = 1 - \sum_{j=1}^{i} p_j\,. \tag{4.19}$$

Equation (4.18) can then be rewritten as

$$-m + c_i h_{i-1} - c_i p_i - \sum_{j=1}^{i-1} c_j p_j = 0\,. \tag{4.20}$$

For the iteration $i-1$, we can replace $c_{i-1}h_{i-2} - c_{i-1}p_{i-1}$ by $c_{i-1}h_{i-1}$, which gives

$$-m + c_{i-1} h_{i-1} - \sum_{j=1}^{i-2} c_j p_j = 0. \tag{4.21}$$

We then remove (4.20) from (4.21) to obtain a recurrence equation linking p and h

$$p_i = \left(1 - \frac{c_{i-1}}{c_i}\right) h_{i-1} - \frac{c_{i-1}}{c_i} p_{i-1}\,, \tag{4.22}$$

then we can complete using equation (4.19) rewritten

$$h_i = h_{i-1} - p_i = \frac{c_{i-1}}{c_i}(h_{i-1} + p_{i-1})\,. \tag{4.23}$$

The equations (4.22) and (4.23) can be written as vector equations using the following notations

$$X_i = \begin{pmatrix} p_i \\ h_i \end{pmatrix}, \quad A = \begin{pmatrix} -1 & -1 \\ 1 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix},$$

(4.22) and (4.23) are then expressed as

$$X_i = \left(\frac{c_{i-1}}{c_i}A + B\right) X_{i-1}. \tag{4.24}$$

$A^2 = 0$, $B^2 = 0$, $AB = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$ $BA = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$ $ABA = A$ $BAB = B$

Suppose we have

$$X_{2i+1} = (\alpha_i A + \beta_i B)X_0, \tag{4.25}$$

with $p_0 = 0$ and $h_0 = 1$. Applying twice in a row the recurrence (4.24) and the simple computations, we obtain

$$X_{2i+3} = \left(\frac{c_{2i+2}}{c_{2i+3}}A + B\right) X_{2i+2} = \left(\frac{c_{2i+2}}{c_{2i+3}}AB + \frac{c_{2i+1}}{c_{2i+2}}BA\right) X_{2i+1}\,. \tag{4.26}$$

By injecting proposition (4.25) into equation (4.26), we find that

$$X_{2i+3} = \left(\frac{c_{2i+2}}{c_{2i+3}}AB + \frac{c_{2i+1}}{c_{2i+2}}BA\right) (\alpha_i A + \beta_i B)\, X_0 = \left(\frac{c_{2i+2}}{c_{2i+3}}\alpha_i A + \frac{c_{2i+1}}{c_{2i+2}}\beta_i B\right) X_0\,. \tag{4.27}$$

That is, the following recurrence relations:

$$\alpha_{i+1} = \frac{c_{2i+2}}{c_{2i+3}}\alpha_i \tag{4.28}$$

$$\beta_{i+1} = \frac{c_{2i+1}}{c_{2i+2}}\beta_i \tag{4.29}$$

$$\tag{4.30}$$

Given $\alpha_0 = m/c_1$, $c_0 = m$ and $\beta_0 = 1$, we derive the general expressions of the coefficients for $i > 0$

$$\alpha_i = \frac{\prod_{j=0}^{i} c_{2j}}{\prod_{j=0}^{i} c_{2j+1}}\,,$$

$$\beta_i = \frac{\prod_{j=0}^{i-1} c_{2j+1}}{\prod_{j=1}^{i} c_{2j}}\,.$$

In the same way, if we suppose

$$X_{2i} = (\gamma_i A + \delta_i B)\, X_1, \quad p_1 = 1 - m/c_1, \quad h_1 = m/c_1\,.$$

The same method of calculation shows

$$\begin{aligned} X_{2i+2} &= \left(\frac{c_{2i+1}}{c_{2i+2}} A + B\right) X_{2i+1}, \\ &= \left(\frac{c_{2i+1}}{c_{2i+2}} AB + \frac{c_{2i}}{c_{2i+1}} BA\right) X_{2i}, \\ &= \left(\frac{c_{2i+1}}{c_{2i+2}} \gamma_i A + \frac{c_{2i}}{c_{2i+1}} \delta_i B\right) X_1. \end{aligned}$$

This leads to the fact that

$$\gamma_i = \frac{\prod_{j=0}^{i-1} c_{2j+1}}{\prod_{j=1}^{i} c_{2j}}\,, \tag{4.31}$$

$$\delta_i = \frac{\prod_{j=1}^{i-1} c_{2j}}{\prod_{j=1}^{i-1} c_{2j+1}}\,. \tag{4.32}$$

$$\tag{4.33}$$

It has therefore been shown, with equations (4.28)-(4.29) and (4.31)-(4.32), that the recurrence (4.24) can be decomposed and solved between even and odd indices as follows

$$X_{2i+1} = \left(\frac{\prod_{j=0}^{i} c_{2j}}{\prod_{j=0}^{i} c_{2j+1}} A + \frac{\prod_{j=0}^{i-1} c_{2j+1}}{\prod_{j=1}^{i} c_{2j}} B\right) . \begin{pmatrix} 0 \\ 1 \end{pmatrix}\,, \tag{4.34}$$

$$X_{2i} = \left(\frac{\prod_{j=0}^{i-1} c_{2j+1}}{\prod_{j=1}^{i} c_{2j}} A + \frac{\prod_{j=1}^{i-1} c_{2j}}{\prod_{j=1}^{i-1} c_{2j+1}} B\right) . \begin{pmatrix} 1 - \frac{m}{c_1} \\ \frac{m}{c_1} \end{pmatrix}\,. \tag{4.35}$$

In terms of p_i, this yields

$$p_{2i+1} = \frac{\prod_{j=0}^{i-1} c_{2j+1}}{\prod_{j=1}^{i} c_{2j}} - \frac{\prod_{j=0}^{i} c_{2j}}{\prod_{j=0}^{i} c_{2j+1}}\,, \tag{4.36}$$

$$p_{2i} = \frac{m}{c_1} \frac{\prod_{j=1}^{i-1} c_{2j}}{\prod_{j=1}^{i-1} c_{2j+1}} - \frac{\prod_{j=0}^{i-1} c_{2j+1}}{\prod_{j=1}^{i} c_{2j}} = \frac{\prod_{j=0}^{i-1} c_{2j}}{\prod_{j=0}^{i-1} c_{2j+1}} - \frac{\prod_{j=0}^{i-1} c_{2j+1}}{\prod_{j=1}^{i} c_{2j}}\,. \tag{4.37}$$

It may also be noted that the equations (4.34) and (4.35) provide expressions for h_i (the unoccupied habitat by species 1 to i) at equilibrium. As long as all species persist:

$$h_{2i+1} = \frac{\prod_{j=0}^{i} c_{2j}}{\prod_{j=0}^{i} c_{2j+1}}, \tag{4.38}$$

$$h_{2i} = \frac{\prod_{j=0}^{i-1} c_{2j+1}}{\prod_{j=1}^{i} c_{2j}}. \tag{4.39}$$

By manipulating these two equations, we see that: $h_i = \left(\frac{c_{i-1}}{c_i}\right) h_{i-2}$, which gives a decrease rate of h_i for every two species added.

For species i to persist in the system, its equilibrium given by (4.36) or (4.37) must be positive. This implies that the conditions for species persistence can be given as

$$p_{2i+1} \text{ persist } \Leftrightarrow c_{2i+1} > m\left(\prod_{j=1}^{i} c_{2j}\right)^2 / \left(\prod_{j=0}^{i-1} c_{2j+1}\right)^2, \tag{4.40}$$

$$p_{2i} \text{ persist } \Leftrightarrow c_{2i} > \left(\prod_{j=0}^{i-1} c_{2j+1}\right)^2 / \left[m\left(\prod_{j=1}^{i-1} c_{2j}\right)^2\right]. \tag{4.41}$$

We can reorder the inequalities (4.40)-(4.41), so that the conditions of coexistence can be written as ($c_0 = m$)

$$1 < \frac{c_1}{c_0} < \frac{c_2}{c_1} < \frac{c_1}{c_0}\frac{c_3}{c_2} < \frac{c_2}{c_1}\frac{c_4}{c_2} < \frac{c_1}{c_0}\frac{c_3}{c_2}\frac{c_5}{c_4} < ...$$

stopping inequalities with on the right a fraction where the largest index of the numerator is equal to the index of the last persistent species.

4.B Reminders of probability

4.B.1 Definition of standard distribution

For reproductibility purposes, we recall some standard positive probability distribution. Each distribution is defined by a random variables X following the probability distribution function (PDF) f and the cumulative distribution function (CDF) F. We denote by

$$\mathbf{1}_{[a,b]}(x) = \begin{cases} 1 \text{ if } x \in [a,b], \\ 0 \text{ else,} \end{cases}$$

the characteristic function.

Definition 4.3 (Continuous uniform)**.** The continuous uniform distribution $\mathcal{U}([a,b])$ describes an experiment where there is an arbitrary outcome that lies between certain bounds: $(a,b) \in \mathbb{R}_+^2$,

$$f(x;a,b) = \frac{1}{b-a}\mathbf{1}_{[a,b]}(x) \ , \ F(x;a,b) = \frac{x-a}{b-a}\mathbf{1}_{[a,b]}(x) + \mathbf{1}_{[b,+\infty)}(x).$$

Given X a random variable following the distribution $\mathcal{U}([a,b])$

$$\mathbb{E}(X) = \frac{a+b}{2} \ , \ \mathrm{Var}(X) = \frac{(b-a)^2}{12}$$

Definition 4.4 (Pareto)**.** The Pareto distribution $\mathcal{P}(a)$ is a power distribution distribution with shape a, support $[1,+\infty)$,

$$f(x;a) = \frac{a}{x^{a+1}}\mathbf{1}_{[1,+\infty)}(x) \ , \ F(x;a) = \left(1 - \frac{1}{x^a}\right)\mathbf{1}_{[1,+\infty)}(x).$$

Given X a random variable following the distribution $\mathcal{P}(a)$:

$$\mathbb{E}(X) = \begin{cases} \frac{a}{a-1} \text{ if } a > 1, \\ +\infty \text{ otherwise} \end{cases} \ , \ \mathrm{Var}(X) = \begin{cases} \frac{a}{(a-1)^2(a-2)} \text{ if } a > 2, \\ +\infty \text{ otherwise.} \end{cases}$$

The Pareto PDF f_{par} can be expressed as a function of the exponential PDF $f_{\exp}$:

$$f_{par}(x;a) = f_{\exp}(\log(x);a)$$

Definition 4.5 (Log-normal)**.** The log-normal distribution $Lognormal(\mu,\sigma^2)$, $\mu \in \mathbb{R}$ and $\sigma > 0$, is a probability distribution of a random variable whose logarithm is distributed in accordance with a Normal distribution.

$$f(x,\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\log(x)-\mu)^2}{2\sigma^2}\right)\mathbf{1}_{(0,+\infty)}(x),$$

$$F(x,\mu,\sigma) = \Phi\left(\frac{\log(x)-\mu}{\sigma}\right)\mathbf{1}_{(0,+\infty)}(x),$$

where Φ corresponds to the CDF of the normal distribution $\mathcal{N}(0,1)$.
Given X a random variable following the distribution $Lognormal(\mu,\sigma^2)$:

$$\mathbb{E}(X) = e^{\mu+\sigma^2/2} \ , \ \mathrm{Var}(X) = (e^{\sigma^2}-1)e^{2\mu+\sigma^2}$$

Definition 4.6 (Log-Cauchy)**.** The log-Cauchy distribution LogCauchy(μ,σ),$\mu \in \mathbb{R}$ and $\sigma > 0$, is a probability distribution of a random variable whose logarithm is distributed in accordance with a Cauchy distribution.

$$f(x;\mu,\sigma) = \frac{1}{x\pi}\left[\frac{\sigma}{(\log(x)-\mu)^2+\sigma^2}\right]\mathbf{1}_{(0,+\infty)}(x),$$

$$F(x,\mu,\sigma) = \frac{1}{2} + \frac{1}{\pi}\arctan\left(\frac{\log(x)-\mu}{\sigma}\right).$$

Mean and variance are infinite.

Definition 4.7 (Exponential)**.** The exponential distribution $\mathcal{E}(\lambda)$, $\lambda > 0$ is a continuous analogue of the geometric distribution.

$$f(x,\lambda) = \lambda e^{-\lambda x}\mathbf{1}_{[0,+\infty)}(x)\ ,\ F(x,\lambda) = 1 - e^{-\lambda x}\mathbf{1}_{[0,+\infty)}(x).$$

Given X a random variable following the distribution $\mathcal{E}(\lambda)$:

$$\mathbb{E}(X) = \frac{1}{\lambda}\ ,\ \mathrm{Var}(X) = \frac{1}{\lambda^2}$$

4.B.2 Reminder on order statistics

Given a sample of iid random variables $X = (X_1, \cdots, X_n)$ distributed by the density function f and F stand for the cumulative distribution function. We denote the order statistics: $X_{(1)}, \cdots, X_{(n)}$ by sorting the realization of X by increasing order. In particular, on the one hand $X_{(1)}$ correspond to the smallest value and the minimum of X and on the other hand $X_{(n)} = \max(X_1, ..., X_n)$.

- One can defined each order statistics by its probability density function

$$f_{X_{(k)}}(x) = \frac{n!}{(k-1)!(n-k)!} f(x)F(x)^{k-1}(1-F(x))^{n-k}\,.$$

- The joint density function of the order statistics is

$$f(x_{(1)}, ..., x_{(n)}) = n!\left(\prod_{i=1}^{n} f(x_{(i)})\right)\mathbb{1}_{x_{(1)}<....<x_{(n)}}\,.$$

- The joint density function of $X_{(k)}$ and $X_{(l)}$, $1 \leqslant k < l \leqslant n$ is

$$f_{X_{(k)},X_{(l)}}(x,y) = \frac{n!}{(k-1)!(l-1-k)!(n-l)!} f(x)f(y)\times \\ [F(x)]^{k-1}[F(y)-F(x)]^{l-1-k}[1-F(y)]^{n-l}\,.$$

- The joint density function of $X_{(k)}$, $X_{(l)}$ and $X_{(m)}$, $1 \leqslant k < l < m \leqslant n$ is

$$f_{X_{(k)},X_{(l)},X_{(m)}}(x,y,z) = \frac{n!}{(k-1)!(l-1-k)!(m-l-1)!(n-m)!} f(x)f(y)f(z)\times \\ [F(x)]^{k-1}[F(y)-F(x)]^{l-1-k}[F(z)-F(y)]^{m-1-l}[1-F(z)]^{n-m}\,.$$

- In the case of a large sample of n order statistics, if F is continuous non zero, an application of the multivariate central limit theorem and the delta method yields

$$X_{(\lfloor np \rfloor)} \sim \mathcal{N}\left(F^{-1}(p), \frac{p(1-p)}{n[f(F^{-1}(p))]^2}\right)\,.$$

  where $p \in (0,1)$ and $\lfloor \dots \rfloor$ denote the integer part.

For more information on order statistics, I suggest the reader to look at Arnold *et al.* [ABN08].

4.B.3 Properties of the maximum

To be able to compare the distributions in the different sequential invasion processes, we obtain information on $X_{(n)} = \max(X_1, \cdots, X_n)$. In particular, we get information about the mean of this random variable. One can explicit the mean of a positive random variable in terms of the survival function

$$\mathbb{E}(X_{(n)}) = \int_0^{+\infty} \mathbb{P}(X_{(n)} > x)dx\,.$$

Continuous uniform Since the uniform distribution has a finite support, given $X = (X_1, \cdots, X_n) \sim \mathcal{U}([a, b])$:

$$\mathbb{E}(X_{(n)}) \xrightarrow[n\to+\infty]{} b\,.$$

Exponential Given $X = (X_1, \cdots, X_n) \sim \mathcal{E}(\lambda)$, the right edge of the support is infinite, we have

$$\begin{aligned}
\mathbb{E}(X_{(n)}) &= \int_0^{+\infty} 1 - (1 - e^{-\lambda x})^n dx\,,\\
&= \int_0^{+\infty} 1 - \sum_{k=0}^{n} \binom{n}{k} (-1)^k e^{-\lambda k x} dx\,,\\
&= \int_0^{+\infty} \sum_{k=1}^{n} \binom{n}{k} (-1)^{k-1} e^{-\lambda k x} dx\,,\\
&= \sum_{k=1}^{n} (-1)^{k-1} \binom{n}{k} \int_0^{+\infty} e^{-\lambda k x} dx\,,\\
&= \sum_{k=1}^{n} (-1)^{k-1} \binom{n}{k} \frac{1}{\lambda k}\,.
\end{aligned}$$

Pareto Given $X = (X_1, \cdots, X_n) \sim \mathcal{P}(a)$, the right edge of the support is infinite and $a > 2$ to have a finite variance:

$$\begin{aligned}
\mathbb{E}(X_{(n)}) &= \int_0^{+\infty} 1 - \left(\left(1 - \frac{1}{x^a}\right) \mathbf{1}_{[1,+\infty)}(x) \right)^n dx\,,\\
&= 1 + \int_1^{+\infty} 1 - \left(1 - \frac{1}{x^a}\right)^n dx\,,\\
&= 1 + \int_1^{+\infty} 1 - \sum_{k=0}^{n} (-1)^k \frac{1}{x^{ak}} \binom{n}{k} dx\,,\\
&= 1 + \sum_{k=1}^{n} (-1)^{k-1} \binom{n}{k} \int_1^{+\infty} \frac{1}{x^{ak}} dx\,,\\
&= 1 + \sum_{k=1}^{n} (-1)^{k-1} \binom{n}{k} \left[-\frac{1}{(ak-1)x^{ak-1}} \right]_1^{+\infty}\,,\\
&= 1 + \sum_{k=1}^{n} (-1)^{k-1} \binom{n}{k} \frac{1}{(ak-1)}\,.
\end{aligned}$$

Log-Cauchy Since the mean and variance of the Log-Cauchy are infinite, we deduce the mean of the maximum is as well infinite.

An empirical point of view For a fixed n, it is always possible to compute empirically the value of the maximum provided that the mean does not tend to infinity. A Monte Carlo (MC) method can be used, the maximum of a random vector of the desired distribution is computed a large number of times and its empirical mean is computed. Assuming that the mean of the maximum is increasing with respect to n (which is not completely absurd).

If a desired maximum value is given, it is possible by a fixed point method to find the parameters of the distribution giving this maximum. This way of proceeding can be a bit shaky because our function only returns an "approximation" of the maximum of the distribution. However, if the MC process is repeated a large number of times, a reasonable approximation can be obtained.

4.B.4 Reminder on heavy-tailed distribution

A heavy-tailed distribution is a probability distribution whose tail is not exponentially bounded i.e. given X a random variables and F its CDF the moment generating function of X, $M_X(t)$, is infinite for all $t > 0$ which mean

$$\int e^{tx} dF(x) = \infty, \, \forall t > 0 \, .$$

In others words, given $G(x) = \mathbb{P}(X > x)$ the survival function of X, the distribution is heavy-tailed if

$$\lim_{x \to \infty} e^{tx} G(x) = \infty, \, \forall \, t > 0 \, .$$

In this paper, common heavy-tailed distribution are Pareto distribution and Log-Cauchy distribution is considered as "super heavy-tailed" distribution because it exhibits a logarithmic decay which is heavier than the Pareto distribution.

The reference we are working with is Foss *et al.* [FKZ13].

4.C An optimal density function

Is it possible to find a continuous density function f which optimize the probability of coexistence of n species in the process (beyond a universality result)? This section is an open ended part to understand the intuitive idea of a heavy tailed distribution. In the sequel, the extinction rate is $m = 1$.

The 2-species case

Let begin with a system of two species with colonization rates c_1 and c_2, two random variables with density function f. We would like to maximize

$$P = \max_{f \in \mathcal{D}} 2 \int_1^{+\infty} \int_{x^2}^{+\infty} f(x) f(y) \mathbb{1}_{x<y} dy dx \, ,$$

where $\mathcal{D}$ is the class of density function.

This kind of problem seems mathematically intractable even less when the dimension will increase. To circumvent this problem, we can express the density function as a sum of Dirac

$$f(x) = \frac{1}{n}\sum_{i=1}^{n} \delta_{a_i}(x)\,.$$

The idea behind this is have some insights about the distribution of the mass. The vector of the mass is given by $\boldsymbol{a} = (a_1, \cdots, a_n) \in \mathbb{R}_+^n$. We could approximate the Dirac by a selection of well chosen functions or get information about the cumulative distribution function F. By replacing f by the sum of Dirac

$$P = \max_{\boldsymbol{a}\in\mathbb{R}_+^n} \frac{2}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{1}_{\{a_i>1\}}\mathbb{1}_{\{a_j>a_i^2\}}\mathbb{1}_{\{a_j>a_i\}}$$

To maximize P, we have to choose $\forall i$, $a_i > 1$, i.e. colonization rates cannot be less than 1. There is a major loss of the mass if the support of the density is below the extinction rate. In addition, there is the following implication $a_j > a_i^2 \Rightarrow a_j > a_i$. Let $a_i, i \in [n]$ in a increasing order

$$1 < a_1 < a_2 < \cdots < a_n\,.$$

Then

$$P = \max_{\boldsymbol{a}\in(1,+\infty)^n} \frac{2}{n^2}\sum_{i=1}^{n}\sum_{j>i}^{n} \mathbb{1}_{\{a_j>a_i^2\}}\,.$$

Finally, the condition on the vector of mass $\boldsymbol{a}$ to maximize coexistence between two species is

$$a_n > a_{n-1}^2 > a_{n-2}^4 > \cdots > a_1^{2^{n-1}}\,.$$

The ratio between two mass is getting bigger and bigger

$$\frac{a_i}{a_{i-1}} > a_1^{2^{i-2}} > exp(2^{i-2}\log(a_1)) > e^{exp((i-2)\log(2))}\,.$$

The main idea is that if $n \to \infty$ we pick two colonization rate (c_1, c_2) from the set $\boldsymbol{a}$ we are a.s. that the condition of coexistence is satisfied.

Remark 4.10. Given a_i, $i \in [n]$ in a increasing order with $1 < a_1 < a_2 < \cdots < a_n$ and $a_1 = 1+\epsilon$, then the minimal value of a_2 must be $a_2 = (1+\epsilon)^2$. With the same arguments, we find

$$a_3 = (1+\epsilon)^4\ ,\ a_4 = (1+\epsilon)^8,\ a_i = (1+\epsilon)^{2^{i-1}} = \sum_{k=0}^{2^{i-1}} \binom{2^{i-1}}{k}\epsilon^k\,.$$

The 3-species case

Using the same arguments as in the 3-species case and considering a vector of mass $\boldsymbol{a} = (a_1, \ldots, a_n)$,

$$P = \max_{f\in\mathcal{D}} 6\int_1^{+\infty}\int_{x^2}^{+\infty}\int_{y^2/x^2}^{+\infty} f(x)f(y)f(z)\mathbb{1}_{x<y}\mathbb{1}_{y<z}dzdydx\,,$$

$$P = \max_{\boldsymbol{a}\in\mathbb{R}_+^n} \frac{6}{n^3} \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} \mathbb{1}_{\{a_i>1\}}\mathbb{1}_{\{a_j>a_i^2\}}\mathbb{1}_{\left\{a_k>\frac{a_j^2}{a_i^2}\right\}}\mathbb{1}_{\{a_j>a_i\}}\mathbb{1}_{\{a_k>a_j\}} .$$

Following the same assumptions of the 2-species case and with the arguments

$$a_j > a_i^2 \Leftrightarrow a_j^2 > a_i^2 a_j \quad , \quad a_k > a_j^2/a_i^2 \Rightarrow a_k > a_j \, ,$$

$$P = \max_{\boldsymbol{a}\in(1,+\infty)^n} \frac{6}{n^3} \sum_{i=1}^{n}\sum_{j>i}^{n}\sum_{k>j}^{n} \mathbb{1}_{\{a_j>a_i^2\}}\mathbb{1}_{\left\{a_k>\frac{a_j^2}{a_i^2}\right\}} .$$

Finally if $\forall j > i, a_j > a_i^2$ then for $k > j$, we have $a_k > a_j^2 > \frac{a_j^2}{a_i^2}$. This implies that the first condition controls the sum. One can count

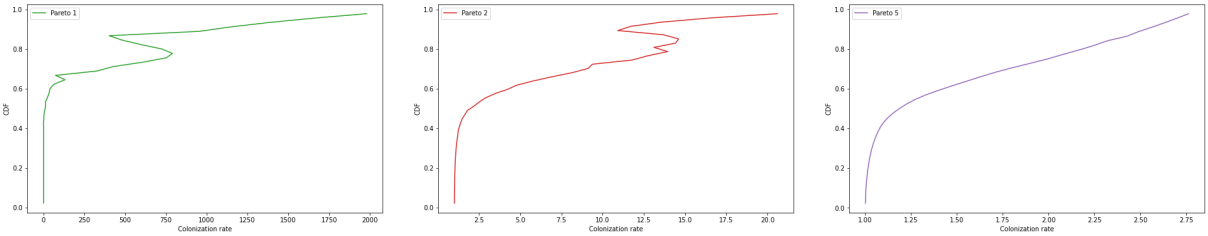
$$\frac{1}{n^3} \sum_{i=1}^{n}\sum_{j>i}^{n}\sum_{k>j}^{n} 1 \xrightarrow[n\to+\infty]{} \frac{1}{6} .$$

Finally, by switching to a continuous distribution, we would have a heavy density on the left side and a heavy-tail on the right side. These conditions are sufficient to find an optimal density for a finite number of species. If the number of species becomes infinite, then the probability of falling on the same mass would no longer be 0. This section gives an idea of the balance between competition and colonization found in the sequential invasion process.

4.D Additional graphics

Sequential invasion process

Distribution of the colonization rate

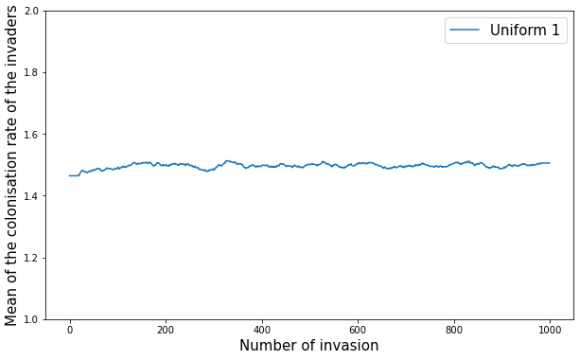


(a) $\mathcal{P}(1)$ (b) $\mathcal{P}(2)$ (c) $\mathcal{P}(5)$

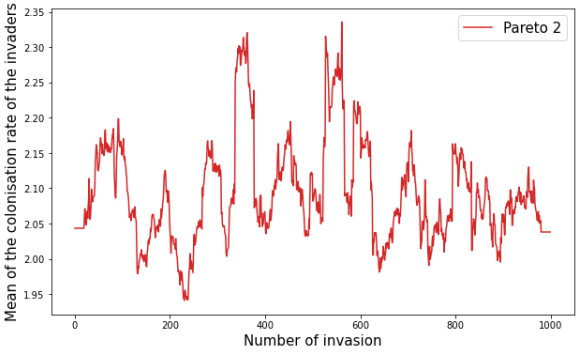
Figure 4.D.1: Representation of the empirical CDF after the sequential invasion of the colonization rate of the persistent species after 500 invasion trials. The curve is derived using Monte Carlo simulations $P = 1000$.

Feature of an invader

We perform a moving average on the realized colonization rates and average the time series for different sequential invasion processes (see Figure 4.D.2).



(a) $\mathcal{U}([1,2])$

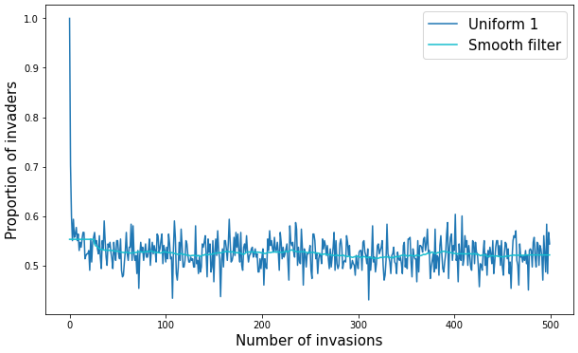


(b) $\mathcal{P}(2)$

Figure 4.D.2: Representation of a moving average on the colonization rates of the invaders for a sequential invasion process of size $n = 1000$. A Monte Carlo experiment is realized $P = 100$ times by getting times series of the moving average curve on the colonization values and average the time series of P sequential invasion process. In Fig. (a), the plot corresponds to the uniform distribution $\mathcal{U}([1,2])$. In Fig. (b), the plot corresponds to the Pareto distribution with parameter $a = 2$.

Properties of the invaders

We can also compute a good intuition of the probability by keeping in memory for all invasion trials if a species invades (or not) the system. (see Figure 4.D.3) One can repeat the process many times and have an "average" at time l that a species invades the system. One obtains a series of values which are decorrelated from each other.



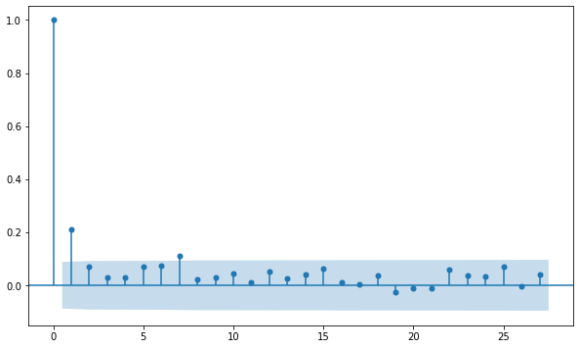


Figure 4.D.3: Representation of the probability of an invader as a function of the invasion trials (left) and its auto-correlation curve (right). The probability is computed by a Monte-Carlo experiment. We run P sequential invasion process and keep a binary vector where the value is 1 if the species invades at invasion trial l, 0 otherwise. We average all the binary vectors and obtain for each invasion trial l a probability of invasion. A smoothing function is also drawn. The right graph corresponds to an auto-correlation plot of the left curve.

Conclusion and Perspectives

Conclusion

From a general perspective, this thesis is based on the development of mathematical and numerical techniques for the analysis of theoretical models of ecosystems. In particular, I am interested in developing a better understanding of the diversity of species and the complexity of their interactions. This represents a huge challenge in theoretical ecology and the complexity of the system requires mathematical models. The aim is to improve our understanding of their properties (feasibility, stability, persistence) in order to preserve biodiversity, save endangered species or manage exploited species in a sustainable way. Future disturbances will heavily affect ecosystems and preventing and understanding these changes may be one of the major challenges of ecological research.

In order to explain the coexistence of species, three issues need to be considered, the relative importance of which is still unknown... The abiotic constraints and the niche of the species i.e. species sorting [LHM+04] shows that we find species adapted to their environment. The dispersion of species and the spatial structure of the environments i.e. mass effect explain that clumsy species will coexist as long as they are able to disperse well and that their habitats of predilection are not far. The role of interactions between species as a driver of species coexistence remains to this day a major issue in ecology. We are thinking mostly about competition, but we need to broaden this to all types of interactions. The idea is that phenomena like Paine's keystone predation [Pai66] are probably important elements in understanding species coexistence (= if we better understand interactions and their impact on the number of persistent species, we should understand cases where neither species sorting nor mass effect give us the keys).

From a more specific point of view, the objective of my thesis was to elaborate a quantitative analysis of large Lotka-Volterra models based on Random Matrix Theory. The theory of random matrices is a powerful mathematical tool still in full expansion with many unsolved problems so far. Numerous matrix structures have been studied creating a direct link to the many types of interaction or food-web matrices between species within the same community.

Starting from the beginning, it is May [May72] who introduced the use of random matrices in ecology by emphasizing a paradox between diversity and stability. Following May's work, many efforts have been made to study particular structures of the random Jacobian matrix also called "community matrix", in particular the work of Allesina *et al.* [AT12, AGB+15, GAS+17]. However, the study of the Jacobian only gives information on the stability of a system around the equilibrium. The Lotka-Volterra model considering a random interaction matrix is a first step towards a more sophisticated system. This model is simple (the A_{ij} has few distribution parameters), versatile because many models can be written in the form of an LV model like the competition-colonization trade-off model and

robust in its dynamic behavior (cyclic, chaos, equilibrium). The large-scale version of the LV model has undergone many studies in recent years, on topics such as the feasibility of an equilibrium [BN21], the stability of the Jacobian of the LV model with dependence between the interaction matrix and the vector of abundances [GGRA18, Sto18] and structural stability [RSB14, GAS+17, SRB+17, RSB14]. Servan *et al.* [SCG+18] have given results on coexistence in a feasible sub-population model. My thesis work is positioned at the center of gravity between the work of Bizeul and Najim, Servan *et al.* [SCG+18] and results using physics-based tools such as the dynamical cavity method physicists use to study the LV model [Bun17, Gal18].

In Chapter 1, we give sufficient conditions for the existence and uniqueness of a globally stable equilibrium with vanishing species in a community. This study is carried out in the framework of a random interaction matrix with two main parameters, α controlling the interaction strength and μ controlling the dominant type of interaction. We give heuristics to find properties of this equilibrium such as the proportion of persistent species, as well as the variance and mean of their equilibrium abundances. We observed that an increase in the strength of interaction between species increases the number of extinctions. In nature, interactions between species are constantly changing and affected by the environment. Under the assumptions that environmental conditions influence interaction strengths, we have endeavored to study the consequences of a sudden change of environmental conditions, expressed through a decrease in parameter α. Solving numerically the Lotka-Volterra system confirms the predictions given by heuristics, i.e. that a increase of the interaction strength negatively affects equilibrium species richness. In this context, a analysis of diversity measure suggests that mutualistic interactions imply shorter transient dynamics.

In Chapter 2, we extend the feasibility results of Bizeul and Najim [BN21] to the case of pairwise correlated interactions. We consider a correlation profile where the correlation can be different between the different pairs of interactions. The main outcome is that the correlation between the interactions does not influence the feasibility threshold. Moreover, we give sufficient conditions for the existence of a unique globally stable equilibrium for a sub-population of persisting species. In order to proceed we combine results by Takeuchi and Adachi [TA80] the stability of LV systems with Random Matrix Theory results. We finally conclude on estimating the proportion of persisting species. At a physical level of rigor, we state the open problem, recall Bunin's and Galla's equations and provide simulations of a closed-form system of equations to compute the proportion of persisting species.

In Chapter 3, we investigate the LV model to describe the properties of a multi-community model. By adding a block structure to the matrix of interactions, we study the properties (feasibility, existence of an attrition phenomenon within each community) of distinct communities by adjusting the inter- and intra-community interactions. In particular, we analyze the properties and dynamics that emerge with two communities of interacting species. New patterns emerge when considering interacting communities. The interplay between the two communities affect their respective equilibrium and their resistance to small perturbations (stability). From an ecosystem point of view, this means

understanding the coexistence between communities with different properties that interact with each other. A numerical similarity study is carried out between a model where the strength of the interactions varies and a model where the connectance in each of the communities varies, which gives an interaction matrix known as the Stochastic Block Model (Bernouilli version). This similarity is analyzed through the stability condition given historically by May [May72].

In Chapter 4, we analyze a multi-species occupancy model representing a system of hierarchical competition between species where the dynamics of each species' occurrences within the metacommunity depends mainly on its colonization rate and its extinction rate. For a better understanding of the restrictions induced on the colonization rate in a large-dimension system, we propose a probabilistic interpretation of the model by looking at colonization parameters following a given probability distribution. We carried out analytical and simulation-based work to investigate the optimal distribution. From a mathematical point of view, we elucidate the stability and persistence of the model allowing the whole or a sub-population of species to coexist using its similarity with the Lotka-Volterra model. Second, we analyzed two different types of approaches: a all-at-once metacommunity process starting from a pool of species by letting the dynamics elapse, and a sequential invasion process developing an invasion sequence that involves a historical contingency effect. On the one hand, surprisingly in the all-at-once metacommunity process we find a universality result on the distribution of persistent species for a wide range of distributions. We find that on average the proportion of persistent species is one half. On the other hand, in the sequential invasion process, the heavier the right-hand side "tail" of the distribution of the colonization rate of the invaders, the higher the probability of coexistence. Subsequently, the comparison of the two approaches shows us that the sequential invasion process seems to be much more restrictive in terms of the number of persisting species due to historical contingencies and extinction cascades. To conclude, this probabilistic perspective of the hierarchical competition-colonization trade-off model allows to put forward and compare two different types of distinct assemblages and gives conditions for many species to coexist under the competition-colonization trade-off.

Take-home messages

The bottom line of this thesis is that some details are important for understanding the dynamics and diversity of large Lotka-Volterra systems or related models.. In particular, the strength of the interactions (Chapter 1) and the block structure (Chapter 3) have an important impact on the properties and stability of an equilibrium where some species vanish. However, I have shown that pairwise correlation has no impact on the feasibility of the equilibrium (Chapter 2).

I found a heuristic to improve the understanding and relationship between important properties of the equilibrium such as the proportion of persistent species, mean and variance of persistent species for a general Lotka-Volterra model (Chapter 1) and its extensions (Chapter 3). However, if we follow the same approach for the competition-colonization trade-off model, we would have a much simpler result for the proportion of persistent species i.e. half of the species persist (Chapter 4).

From an ecological perspective, the species richness and stability of ecosystems are properties that cannot be studied without taking into account the interactions between

species (Chapters 1 and 2) and the spatial structure (Chapters 3 and 4).

Perspectives

Many perspectives have been discussed at the end of each chapter, I have decided to select five of them whose future issues are important for the understanding of community dynamics and also from the mathematical standpoint.

In the model of Lotka-Volterra in the context of this thesis, we identified three types of interaction: mutualism ($\mu > 0, \rho > 0$), competition ($\mu < 0, \rho > 0$) and predation ($\mu \approx 0, \rho < 0$). In chapter 2, we considered the feasibility problem in a system with pairwise correlated entries. However, in chapter 1 and 3, we only deal with the cases where there is no correlation between the interactions because it is mathematically more difficult to handle these cases. This reduces the general nature of the model. Moreover, with tools coming from physics, one may obtain a number of results (see Bunin [Bun17] in the unique community case and Poley *et al.* [PBG22] in the block case). Using mathematical (not physicists') methods, we are confident that such extensions are possible, but might hinge on more sophisticated developments. If an interaction matrix has a block structure, one could consider correlations in the broader sense of community instead of individual. In this system, we would have similarities between communities according to their degree of correlation.

Recently, many researchers have been interested in exploring higher-order interactions i.e. interactions among three or more species [MS17, GBMSA17, GLL22]. A more distant perspective would be to extend work on pairwise correlations in random matrices to higher-order correlations. Korkmazhan *et al.* [KD22] show that high-order correlations in species interactions lead to complex diversity-stability relationships when studying the "community matrix" or random Jacobian. This type of interactions leads to particular random matrix spectra. Aceituno *et al.* [ARS19] studied matrix spectra in the case where there are higher-order cyclic correlations between k-tuples matrix entries.

A promising method to prove heuristics in larger cases would be the use of the Approximate Message Passing (AMP) algorithm which has distant similarities with the dynamical cavity method. AMP has the advantage of having many mathematically rigorous results [FVRS21, MR16, CT21, JM12].

In this thesis, we were mainly interested in equilibrium properties. However, the Lotka-Volterra model can exhibit other behaviors such as chaos or multi-equilibrium dynamics. In nature, we also find out of equilibrium dynamics. To study this type of behavior, many theories have been developed such as the permanence [Jan87, LM96, JS98, HS98]. A system is called permanent if no species become extinct where the boundaries act as repeller. Another category of non-equilibrium dynamics is the transient species (i.e. species that may persist over a long time period followed by rapid changes in dynamics) investigated by Hastings *et al.* [Has01, HAC+18].

In Chapter 1, a model was defined where the strength of the interactions vary, which could be represented by climate disruption acting on the variability of ecosystems. However, no quantitative indication of this rate of change was given. In the context of the periodic model with $\alpha(t)$, we would like to explain in detail the phenomenon of return

to equilibrium, in particular the speed of convergence or resilience. The rate of the return to equilibrium used by the "standard" theorists depends on the real part of the largest eigenvalue of the Jacobian. Eventually what we observe is that if the strength of the interactions is weak, then the largest eigenvalue will be small and consequently the return to equilibrium will be fast. If the strength of the interactions increases, the dominant eigenvalue will be closer to 0 and the return to equilibrium slower. The second phenomenon is that the rate of return to equilibrium depends directly on the abundance of the species. We observe in the seasonal dynamics that the greater the abundance of a species, the faster it will return to an equilibrium. This is a good explanation to say that when a species converges too fast to 0 then it may never return to the system over a short period of time (in the sense that it will never regain its initial abundance). Arnoldi *et al.* [ALH16, ABLH18] have extensively studied measures of stability such as the resilience or reactivity of a system. A further theoretical study of the LV model could provide a more quantitative answer. In Chapter 1, we have limited the analysis to the case of a single sudden incident, but other types of fluctuations for the interaction strength could be considered for a better understanding of habitat conservation phenomena. For example, a seasonal model could be appropriate to describe the evolution of the dynamics over the seasons.

It is known that if the trajectory stays in a compact space where $x_k > 0, \forall k \in [n]$ (i.e. in some cases of chaotic attractors or cyclic dynamics), then the long-term average of the vector of abundances i.e.

$$\breve{x} = \frac{1}{T}\int_0^T \boldsymbol{x}(t)dt\,.$$

is equal to $\breve{x} = B^{-1}\mathbf{1}$, the same formula to determine a feasible equilibrium.

Could the LCP theory be used in the same way to describe a cyclic phase where some species vanish? In line with this problem, one could be interested in the distribution of species over time and their variations.

A paper released in 2013 by Säterberg *et al.* [SSE13] has shown that the disappearance of ecological interactions or functional extinctions precedes population extinctions. In other words, a species already loses links with its neighbors at the beginning of the decline, making other species around it disappear long before it has itself disappeared. A naive answer would be to study the impact of sparsity (i.e. the average number of interactions per species) on the model. Paradoxically, in chapter 3, we have shown that decreasing the number of interactions increases the stability. However, this only happens when $\boldsymbol{r} > 0$ and each species can survive independently of the others. It would be interesting, but probably difficult, to study a model where some species are strongly dependent on interactions with other species for their survival i.e. $\boldsymbol{r} < 0$. The robustness of the $LCP(M, q)$ results to the q value would be a starting point to analyze this type of dynamics. One could possibly use the structural stability as the work of Saavedra *et al.* [SRB+17] on results with positive r and see how far one can move the vector r without changing the attractor.

In chapter 4, a probabilistic perspective on the competition-colonization trade-off model with hierarchical competition was given, i.e. the competition parameter is intrinsic to the model and the coexistence of species depends only on the colonization and extinction rates. The study of stability in the model designed in Calcagno *et al.* [CMJD06a] could lift the veil on the stability-complexity debate. Spatial dynamics and the study of patch

dynamics may be a possible answer to the stability-diversity paradox. As the model can be rewritten as a Lotka-Volterra model, we could consider random competitions parameter and a random colonization rates.

Recall model in [CMJD06a],

$$\frac{dp_i}{dt} = c_i p_i \left(1 - \sum_{j=1}^{n} p_j\right) - e p_i + c_i p_i \sum_{j \neq i} p_j \eta_{ij} - p_i \sum_{j \neq i} c_j p_j \eta_{ji} \,, \; \forall\, i \in [n] \,,$$

where p_i represents the occupancy of species i. Parameter e is the extinction parameter of all species. Parameter c_i represents the colonization parameter of the species i and η_{ij} corresponds to probability that a population of species i can locally overtake an established population of species j

This system can be represented in the form of a Lotka Volterra competition scheme with asymmetric interactions.

$$\frac{dp_i}{dt} = p_i \left(c_i - e + \sum_{j=1}^{n} p_j \left(c_i \eta_{ij} - c_j \eta_{ji} - c_i\right)\right) \,, \; \forall\, i \in [n] \,.$$

Given $\eta = (\eta_{ij})_{n,n}$ be a square matrix of size n and

$$\boldsymbol{c} = (c_i)_n \,, \; \boldsymbol{p} = (p_i)_n \,, \; \boldsymbol{e} = \mathbf{1} e \,, \; Dp^\top = (\frac{dp_1}{dt}, \frac{dp_2}{dt}, \cdots, \frac{dp_n}{dt}) \,.$$

We can then rewrite the equation in a matrix form

$$\begin{aligned} Dp &= \mathrm{diag}(\boldsymbol{p}) \left[\boldsymbol{c} - \boldsymbol{e} + (\mathrm{diag}(\boldsymbol{c})\eta - \eta^\top \mathrm{diag}(\boldsymbol{c}) - \boldsymbol{c}\mathbf{1}^\top)\boldsymbol{p}\right] \,, \\ &= \mathrm{diag}(\boldsymbol{p})(\boldsymbol{c} - \boldsymbol{e} + A\boldsymbol{p}) \,, \end{aligned}$$

with $A = B - C = \mathrm{diag}(\boldsymbol{c})\eta - \eta^\top \mathrm{diag}(\boldsymbol{c}) - \boldsymbol{c}\mathbf{1}^\top$,

$$C = \begin{pmatrix} c_1 & c_1 & \cdots & c_1 \\ c_2 & c_2 & & c_2 \\ \vdots & & & \\ \vdots & & & \\ c_N & c_N & & c_N \end{pmatrix} \,,$$

$$B = \begin{pmatrix} 0 & c_1\eta_{12} - c_2\eta_{21} & \cdots & c_1\eta_{1n} - c_n\eta_{n1} \\ c_2\eta_{21} - c_1\eta_{12} & 0 & & c_2\eta_{2n} - c_n\eta_{n2} \\ \vdots & & & \\ \vdots & & & \\ c_n\eta_{n1} - c_1\eta_{1n} & & & 0 \end{pmatrix} \,.$$

Matrix B is a real antisymmetric matrix, the nonzero eigenvalues of B are non-real and C is a rank one matrix, therefore it admits 0 as an eigenvalue of multiplicity $n-1$.

In an effort to extend the study of the intermediate disturbance hypothesis, Liao *et al.* [LBB22] analyzed the competition-colonization trade-off model in the case of multimodality in diversity disturbance hypothesis. At the end, they discuss the storage effect as a temporal niche segregation i.e. species are specialized in different phases of environmental fluctuations [MSS22]. If we understand the impact of the η_{ij} matrix in the system, we could imagine varying a matrix $\eta(t)$ over time and observe the behavior of the system as in Chapter 1.

Last but not least, in this thesis, no relation with data has been carried out. Instead, the use of models and the theory of random matrices helps understand what could theoretically happen in the absence of information. The development of this theory provides a better understanding of the underlying mechanisms that ensure the coexistence of species in large ecosystems. A major challenge for future research in this field would be to work with large datasets and determine whether they are consistent with what we observe theoretically. Recently, Hu *et al.* [HAB+21] compared the theoretical Lotka-Volterra phase diagram with experimental data from a laboratory trial. Regarding the Tilman-Hastings model, there are probably ways to apply it to more epidemiological data (see the application of Slatkin [Sla74] by Madec and Gjini [MG20]). With the rise of artificial intelligence tools to collect data and machines in laboratories that are increasingly sophisticated to work with many species, we might be able to better understand the functioning of microbiota [CSF15] or large ecosystems. Finally, I think that the complementarity between theoretical and empirical ecology is very important and represents a beneficial cycle for both fields. In view of the recent climate emergency and the stakes of maintaining ecosystems, which is a major issue in ecology, we must act now!

Bibliography

[AB22] Ada Altieri and Giulio Biroli. Effects of intraspecific cooperative interactions in large ecosystems. *SciPost Physics*, 12(1):013, January 2022.

[Abb18] Emmanuel Abbe. Community Detection and Stochastic Block Models: Recent Developments. *Journal of Machine Learning Research*, 18(177):1–86, 2018.

[ABC20] Ada Altieri, Giulio Biroli, and Chiara Cammarota. Dynamical mean-field theory and aging dynamics. *Journal of Physics A: Mathematical and Theoretical*, 53(37):375006, August 2020.

[ABLH18] J.-F. Arnoldi, A. Bideault, M. Loreau, and B. Haegeman. How ecosystems recover from pulse perturbations: A theory of short- to long-term responses. *Journal of Theoretical Biology*, 436:79–92, January 2018.

[ABN08] Barry C. Arnold, Narayanaswamy Balakrishnan, and Haikady Navada Nagaraja. *A first course in order statistics*. SIAM, 2008.

[Adl06] Frederick R. Adler. Commentary on Calcagno et al. (2006): Coexistence in a metacommunity: the competition–colonization trade-off is not dead. *Ecology Letters*, 9(8):907–909, 2006.

[AEK17] Oskari Ajanki, Laszlo Erdos, and Torben Krüger. Universality for general Wigner-type matrices. *arXiv:1506.05098 [math]*, August 2017.

[AEK19] Oskari Ajanki, Laszlo Erdos, and Torben Krüger. Quadratic vector equations on complex upper half-plane. *Memoirs of the American Mathematical Society*, 261(1261):0–0, September 2019.

[AGB+15] Stefano Allesina, Jacopo Grilli, György Barabás, Si Tang, Johnatan Aljadeff, and Amos Maritan. Predicting the stability of large structured food webs. *Nature Communications*, 6(1):7842, November 2015.

[AK22] Johannes Alt and Torben Krüger. Local elliptic law. *Bernoulli*, 28(2):886–909, May 2022.

[ALH16] J-F. Arnoldi, M. Loreau, and B. Haegeman. Resilience, reactivity and variability: A mathematical comparison of ecological stability measures. *Journal of Theoretical Biology*, 389:47–59, January 2016.

[Alt22] Ada Altieri. Glassy features and complex dynamics in ecological systems, August 2022.

[Ama03] Priyanga Amarasekare. Competitive coexistence in spatially structured environments: a synthesis. *Ecology Letters*, 6(12):1109–1122, 2003.

[AN21] Imane Akjouj and Jamal Najim. Feasibility of sparse large Lotka-Volterra ecosystems. Technical Report arXiv:2111.11247, arXiv, November 2021.

[ARCB21] Ada Altieri, Felix Roy, Chiara Cammarota, and Giulio Biroli. Properties of Equilibria and Glassy Phases of the Random Lotka-Volterra Model with Demographic Noise. *Physical Review Letters*, 126(25):258301, June 2021.

[ARS19] Pau Vilimelis Aceituno, Tim Rogers, and Henning Schomerus. Universal hypotrochoidic law for random matrices with cyclic correlations. *Physical Review E*, 100(1):010302, July 2019.

[AT12] Stefano Allesina and Si Tang. Stability criteria for complex ecosystems. *Nature*, 483(7388):205–208, March 2012.

[AT15] Stefano Allesina and Si Tang. The stability–complexity relationship at age 40: a random matrix perspective. *Population Ecology*, 57(1):63–75, January 2015.

[BA17] Matthieu Barbier and Jean-François Arnoldi. The cavity method for community ecology. preprint, Ecology, June 2017.

[BAB+19] Ulrich Brose, Phillippe Archambault, Andrew D. Barnes, Louis-Felix Bersier, Thomas Boy, João Canning-Clode, Erminia Conti, Marta Dias, Christoph Digel, Awantha Dissanayake, Augusto A. V. Flores, Katarina Fussmann, Benoit Gauzens, Clare Gray, Johanna Häussler, Myriam R. Hirt, Ute Jacob, Malte Jochum, Sonia Kéfi, Orla McLaughlin, Muriel M. MacPherson, Ellen Latz, Katrin Layer-Dobra, Pierre Legagneux, Yuanheng Li, Carolina Madeira, Neo D. Martinez, Vanessa Mendonça, Christian Mulder, Sergio A. Navarrete, Eoin J. O'Gorman, David Ott, José Paula, Daniel Perkins, Denise Piechnik, Ivan Pokrovsky, David Raffaelli, Björn C. Rall, Benjamin Rosenbaum, Remo Ryser, Ana Silva, Esra H. Sohlström, Natalia Sokolova, Murray S. A. Thompson, Ross M. Thompson, Fanny Vermandele, Catarina Vinagre, Shaopeng Wang, Jori M. Wefer, Richard J. Williams, Evie Wieters, Guy Woodward, and Alison C. Iles. Predator traits determine food-web architecture across ecosystems. *Nature Ecology & Evolution*, 3(6):919–927, June 2019.

[BABL18] Matthieu Barbier, Jean-François Arnoldi, Guy Bunin, and Michel Loreau. Generic assembly patterns in complex ecological communities. *Proceedings of the National Academy of Sciences*, 115(9):2156–2161, February 2018.

[Bai97] Z. D. Bai. Circular Law. *The Annals of Probability*, 25(1):494–529, 1997.

[BBC18] Giulio Biroli, Guy Bunin, and Chiara Cammarota. Marginally stable equilibria in critical ecosystems. *New Journal of Physics*, 20(8):083051, August 2018.

[BBP78] G. P. Barker, A. Berman, and R. J. Plemmons. Positive diagonal solutions to the Lyapunov equations. *Linear and Multilinear Algebra*, 5(4):249–256, January 1978.

[BC12] Charles Bordenave and Djalil Chafaï. Around the circular law. *Probability Surveys*, 9, January 2012.

[BDB+11] Edward B. Baskerville, Andy P. Dobson, Trevor Bedford, Stefano Allesina, T. Michael Anderson, and Mercedes Pascual. Spatial Guilds in the Serengeti Food Web Revealed by a Bayesian Group Model. *PLoS Computational Biology*, 7(12):e1002321, December 2011.

[BDH78] A. A. Balkema and L. De Haan. Limit Distributions for Order Statistics. I. *Theory of Probability & Its Applications*, 23(1):77–92, December 1978.

[BFPG+09] Ugo Bastolla, Miguel A. Fortuna, Alberto Pascual-García, Antonio Ferrera, Bartolo Luque, and Jordi Bascompte. The architecture of mutualistic networks minimizes competition and increases biodiversity. *Nature*, 458(7241):1018–1020, April 2009.

[BG20] Joseph W. Baron and Tobias Galla. Dispersal-induced instability in complex ecosystems. *Nature Communications*, 11(1):6032, December 2020.

[BGR16] Florent Benaych-Georges and Jean Rochet. Outliers in the Single Ring Theorem. *Probability Theory and Related Fields*, 165(1):313–363, June 2016.

[BGR+18] Andreas Brechtel, Philipp Gramlich, Daniel Ritterskamp, Barbara Drossel, and Thilo Gross. Master stability functions reveal diffusion-driven pattern formation in networks. *Physical Review E*, 97(3):032307, March 2018.

[BJMO03] J. Bascompte, P. Jordano, C. J. Melian, and J. M. Olesen. The nested assembly of plant-animal mutualistic networks. *Proceedings of the National Academy of Sciences*, 100(16):9383–9387, August 2003.

[BJRG22a] Joseph W. Baron, Thomas Jun Jewell, Christopher Ryder, and Tobias Galla. Eigenvalues of Random Matrices with Generalized Correlations: A Path Integral Approach. *Physical Review Letters*, 128(12):120601, March 2022.

[BJRG22b] Joseph W. Baron, Thomas Jun Jewell, Christopher Ryder, and Tobias Galla. Non-Gaussian random matrices determine the stability of Lotka-Volterra communities. *arXiv:2202.09140 [cond-mat, q-bio]*, February 2022.

[BLM13] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, February 2013.

[BMSA17] György Barabás, Matthew J. Michalska-Smith, and Stefano Allesina. Self-regulation and the stability of large ecological networks. *Nature Ecology & Evolution*, 1(12):1870–1875, December 2017.

[BN21] Pierre Bizeul and Jamal Najim. Positive solutions for large random linear systems. *Proceedings of the American Mathematical Society*, 149(6):2333–2348, June 2021.

[Bol98] Béla Bollobás. Random Graphs. In Béla Bollobás, editor, *Modern Graph Theory*, Graduate Texts in Mathematics, pages 215–252. Springer, New York, NY, 1998.

[BS10] Zhidong Bai and Jack W. Silverstein. *Spectral analysis of large dimensional random matrices*. Springer series in statistics. Springer, New York ; London, 2nd ed edition, 2010.

[BSHM17] Daniel M. Busiello, Samir Suweis, Jorge Hidalgo, and Amos Maritan. Explorability and the origin of network sparsity in living systems. *Scientific Reports*, 7(1):12323, December 2017.

[BSY88] Z. D Bai, Jack W Silverstein, and Y. Q Yin. A note on the largest eigenvalue of a large dimensional sample covariance matrix. *Journal of Multivariate Analysis*, 26(2):166–168, August 1988.

[BTH06] Michael Begon, Colin R. Townsend, and John L. Harper. *Ecology: from individuals to ecosystems*. Blackwell Pub, Malden, MA, 4th ed edition, 2006.

[Bun16] Guy Bunin. Interaction patterns and diversity in assembled ecological communities, July 2016.

[Bun17] Guy Bunin. Ecological communities with Lotka-Volterra dynamics. *Physical Review E*, 95(4):042414, April 2017.

[Bun21] Guy Bunin. Directionality and community-level selection. *Oikos*, 130(4):489–500, 2021.

[BY88] Z. D. Bai and Y. Q. Yin. Necessary and Sufficient Conditions for Almost Sure Convergence of the Largest Eigenvalue of a Wigner Matrix. *The Annals of Probability*, 16(4):1729–1741, 1988.

[Cad07] Marc William Cadotte. Competition-Colonization Trade-offs and Disturbance Effects at Multiple Scales. *Ecology*, 88(4):823–829, 2007.

[Cas90] T J Case. Invasion resistance arises in strongly interacting species-rich model competition communities. *Proceedings of the National Academy of Sciences*, 87(24):9610–9614, December 1990.

[CBB+11] Alva Curtsdotter, Amrei Binzer, Ulrich Brose, Francisco de Castro, Bo Ebenman, Anna Eklöf, Jens O. Riede, Aaron Thierry, and Björn C. Rall. Robustness to secondary extinctions: Comparing trait-based sequential deletions in static and dynamic food webs. *Basic and Applied Ecology*, 12(7):571–580, November 2011.

[CBN90] Joel E. Cohen, Frédéric Briand, and Charles M. Newman. *Community Food Webs*, volume 20 of *Biomathematics*. Springer, Berlin, Heidelberg, 1990.

[CD68] Richard W. Cottle and George B. Dantzig. Complementary pivot theory of mathematical programming. *Linear Algebra and its Applications*, 1(1):103–125, January 1968.

[CDMF09] Mireille Capitaine, Catherine Donati-Martin, and Delphine Féral. The largest eigenvalues of finite rank deformation of large Wigner matrices: Convergence and nonuniversality of the fluctuations. *The Annals of Probability*, 37(1), January 2009.

[CEFN22] Maxime Clenet, Hafedh El Ferchichi, and Jamal Najim. Equilibrium in a large Lotka–Volterra system with pairwise correlated interactions. *Stochastic Processes and their Applications*, 153:423–444, November 2022.

[CHNR21] Nicholas Cook, Walid Hachem, Jamal Najim, and David Renfrew. Non-Hermitian Random Matrices with a Variance Profile (II): Properties and Examples. *Journal of Theoretical Probability*, November 2021.

[CJL+17] Vincent Calcagno, Philippe Jarne, Michel Loreau, Nicolas Mouquet, and Patrice David. Diversity spurs diversification in ecological communities. *Nature Communications*, 8(1):15810, June 2017.

[Cle22a] Maxime Clenet. Equilibrium and surviving species in a large Lotka-Volterra system of differential equations, 2022. Published: *https://github.com/maxime-clenet/Equilibrium-and-surviving-species-in-a-large-Lotka-Volterra-system*.

[Cle22b] Maxime Clenet. Feasibility in a large Lotka-Volterra system with pairwise correlated interactions, 2022. Published: *https://github.com/maxime-clenet/Feasibility-in-a-large-Lotka-Volterra-system-with-pairwise-correlated-interactions*.

[CMJD06a] V. Calcagno, N. Mouquet, P. Jarne, and P. David. Coexistence in a metacommunity: the competition–colonization trade-off is not dead. *Ecology Letters*, 9(8):897–907, August 2006.

[CMJD06b] Vincent Calcagno, Nicolas Mouquet, Philippe Jarne, and Patrice David. Rejoinder to Calcagno et al. (2006): Which immigration policy for optimal coexistence? *Ecology Letters*, 9(8):909–911, 2006.

[CMN22] Maxime Clenet, François Massol, and Jamal Najim. Equilibrium and surviving species in a large Lotka-Volterra system of differential equations, May 2022.

[CN88] Joel E. Cohen and Charles M. Newman. Dynamic Basis of Food Web Organization. *Ecology*, 69(6):1655–1664, 1988.

[Cox94] Gregory E. Coxson. The P-matrix problem is co-NP-complete. *Mathematical Programming*, 64(1):173–178, March 1994.

[CPS09] Richard Cottle, Jong-Shi Pang, and Richard E. Stone. *The linear complementarity problem*. Number 60 in Classics in applied mathematics. Society

for Industrial and Applied Mathematics, Philadelphia, siam ed., [classics ed.] edition, 2009.

[CS18] Simone Cenci and Serguei Saavedra. Structural stability of nonlinear population dynamics. *Physical Review E*, 97(1):012401, January 2018.

[CSF15] K. Z. Coyte, J. Schluter, and K. R. Foster. The ecology of the microbiome: Networks, competition, and stability. *Science*, 350(6261):663–666, November 2015.

[CT21] Wei-Kuo Chen and Si Tang. On Convergence of the Cavity and Bolthausen's TAP Iterations to the Local Magnetization. *Communications in Mathematical Physics*, 386(2):1209–1242, September 2021.

[DO89] S. Diederich and M. Opper. Replicators with random interactions: A solvable model. *Physical Review A*, 39(8):4333–4336, April 1989.

[DVR+18] Michaël Dougoud, Laura Vinckenbosch, Rudolf P. Rohr, Louis-Félix Bersier, and Christian Mazza. The feasibility of equilibria in large ecosystems: A primary but neglected concept in the complexity-stability debate. *PLOS Computational Biology*, 14(2):e1005988, 2018.

[DWM02] Jennifer A. Dunne, Richard J. Williams, and Neo D. Martinez. Food-web structure and network theory: The role of connectance and size. *Proceedings of the National Academy of Sciences*, 99(20):12917–12922, October 2002.

[Dys62] Freeman J. Dyson. Statistical Theory of the Energy Levels of Complex Systems. I. *Journal of Mathematical Physics*, 3(1):140–156, January 1962.

[Ede97] Alan Edelman. The Probability that a Random Real Gaussian Matrix haskReal Eigenvalues, Related Distributions, and the Circular Law. *Journal of Multivariate Analysis*, 60(2):203–232, February 1997.

[EJK+13] Anna Eklöf, Ute Jacob, Jason Kopp, Jordi Bosch, Rocío Castro-Urgal, Natacha P. Chacoff, Bo Dalsgaard, Claudio de Sassi, Mauro Galetti, Paulo R. Guimarães, Silvia Beatriz Lomáscolo, Ana M. Martín González, Marco Aurelio Pizo, Romina Rader, Anselm Rodrigo, Jason M. Tylianakis, Diego P. Vázquez, and Stefano Allesina. The dimensionality of ecological networks. *Ecology Letters*, 16(5):577–583, 2013.

[ER60] P. Erdős and A. Renyi. On the evolution of random graphs, 1960.

[FBM22] Jules Fraboul, Giulio Biroli, and Silvia De Monte. Artificial selection of communities drives the emergence of structured interactions, August 2022.

[FK81] Z. Füredi and J. Komlós. The eigenvalues of random symmetric matrices. *Combinatorica*, 1(3):233–241, September 1981.

[FKZ13] Sergey Foss, Dmitry Korshunov, and Stan Zachary. *An Introduction to Heavy-Tailed and Subexponential Distributions*. Springer Series in Operations Research and Financial Engineering. Springer New York, New York, NY, 2013.

[FN11] Tadashi Fukami and Mifuyu Nakajima. Community assembly: alternative stable states or alternative transient states? *Ecology Letters*, 14(10):973–984, 2011.

[For10] Santo Fortunato. Community detection in graphs. *Physics Reports*, 486(3):75–174, February 2010.

[Fox13] Jeremy W. Fox. The intermediate disturbance hypothesis should be abandoned. *Trends in Ecology & Evolution*, 28(2):86–92, February 2013.

[FP62] Miroslav Fiedler and Vlastimil Pták. On matrices with non-positive off-diagonal elements and positive principal minors. *Czechoslovak Mathematical Journal*, 12(3):382–400, 1962.

[FP66] Miroslav Fiedler and Vlastimil Pták. Some generalizations of positive definiteness and monotonicity. *Numerische Mathematik*, 9(2):163–172, December 1966.

[FVRS21] Oliver Y. Feng, Ramji Venkataramanan, Cynthia Rush, and Richard J. Samworth. A unifying tutorial on Approximate Message Passing. *arXiv:2105.02180 [cs, math, stat]*, May 2021.

[GA70] Mark R. Gardner and W. Ross Ashby. Connectance of Large Dynamic (Cybernetic) Systems: Critical Values for Stability. *Nature*, 228(5273):784–784, November 1970.

[Gal18] Tobias Galla. Dynamically evolved community size and stability of random Lotka-Volterra ecosystems. *EPL (Europhysics Letters)*, 123(4):48004, September 2018.

[GAS+17] Jacopo Grilli, Matteo Adorisio, Samir Suweis, György Barabás, Jayanth R. Banavar, Stefano Allesina, and Amos Maritan. Feasibility and coexistence of large ecological communities. *Nature Communications*, 8(1):14389, April 2017.

[GB12] Luis J. Gilarranz and Jordi Bascompte. Spatial network structure and metapopulation persistence. *Journal of Theoretical Biology*, 297:11–16, March 2012.

[GBMSA17] Jacopo Grilli, György Barabás, Matthew J. Michalska-Smith, and Stefano Allesina. Higher-order interactions stabilize dynamics in competitive network models. *Nature*, 548(7666):210–213, August 2017.

[GGRA18] Theo Gibbs, Jacopo Grilli, Tim Rogers, and Stefano Allesina. Effect of population abundances on the stability of large random ecosystems. *Physical Review E*, 98(2):022410, August 2018.

[GH82] Stuart Geman and Chii Ruey Hwang. A chaos hypothesis for some large systems of random equations. *Zeitschrift fr Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 60(3):291–314, 1982.

[Gin65] Jean Ginibre. Statistical Ensembles of Complex, Quaternion, and Real Matrices. *Journal of Mathematical Physics*, 6(3):440–449, March 1965.

[Gir85] V. L. Girko. Circular Law. *Theory of Probability & Its Applications*, 29(4):694–706, January 1985.

[Gir86] V. L. Girko. Elliptic Law. *Theory of Probability & Its Applications*, 30(4):677–690, December 1986.

[Gir95] V. L. Girko. The Elliptic Law: ten years later I. 3(3):257–302, January 1995.

[GJ77] B. S. Goh and L. S. Jennings. Feasibility and stability in randomly assembled Lotka-Volterra models. *Ecological Modelling*, 3(1):63–71, February 1977.

[GLL22] Theo Gibbs, Simon A. Levin, and Jonathan M. Levine. Coexistence in diverse communities with higher-order interactions, March 2022.

[GML16] Dominique Gravel, François Massol, and Mathew A. Leibold. Stability and complexity in model meta-ecosystems. *Nature Communications*, 7(1):12457, November 2016.

[Goh77] B. S. Goh. Global Stability in Many-Species Systems. *The American Naturalist*, 111(977):135–143, 1977.

[Gop84] K. Gopalsamy. Global asymptotic stability in Volterra's population systems. *Journal of Mathematical Biology*, 19(2):157–168, April 1984.

[GRA16] Jacopo Grilli, Tim Rogers, and Stefano Allesina. Modularity and stability in ecological communities. *Nature Communications*, 7(1):12031, November 2016.

[GSSP+10] R. Guimerà, D. B. Stouffer, M. Sales-Pardo, E. A. Leicht, M. E. J. Newman, and L. A. N. Amaral. Origin of compartmentalization in food webs. *Ecology*, 91(10):2941–2951, October 2010.

[HAB+21] Jiliang Hu, Daniel R. Amor, Matthieu Barbier, Guy Bunin, and Jeff Gore. Emergent phases of ecological diversity and dynamics mapped in microcosms. preprint, Biophysics, October 2021.

[HAC+18] Alan Hastings, Karen C. Abbott, Kim Cuddington, Tessa Francis, Gabriel Gellner, Ying-Cheng Lai, Andrew Morozov, Sergei Petrovskii, Katherine Scranton, and Mary Lou Zeeman. Transient phenomena in ecology. *Science*, 361(6406):eaat6412, September 2018.

[Han83] Ilkka Hanski. Coexistence of Competitors in Patchy Environment. *Ecology*, 64(3):493–500, June 1983.

[Han99] Ilkka Hanski. *Metapopulation Ecology*. OUP Oxford, March 1999.

[Has80] Alan Hastings. Disturbance, coexistence, history, and competition for space. *Theoretical Population Biology*, 18(3):363–373, December 1980.

[Has01] Alan Hastings. Transient dynamics and persistence of ecological systems. *Ecology Letters*, 4(3):215–220, 2001.

[Her90] Roger H. Hering. Oscillations in Lotka-Volterra systems of chemical reactions. *Journal of Mathematical Chemistry*, 5(2):197–202, June 1990.

[HMS16] Jan O. Haerter, Namiko Mitarai, and Kim Sneppen. Food Web Assembly Rules for Generalized Lotka-Volterra Equations. *PLOS Computational Biology*, 12(2):e1004727, 2016.

[HO00] Ilkka Hanski and Otso Ovaskainen. The metapopulation capacity of a fragmented landscape. *Nature*, 404(6779):755–758, April 2000.

[HS98] Josef Hofbauer and Karl Sigmund. *Evolutionary Games and Population Dynamics*. Cambridge University Press, May 1998.

[HSD74] Morris W. Hirsch, Stephen Smale, and Robert L. Devaney. *Differential Equations, Dynamical Systems, and Linear Algebra*. Academic Press, June 1974.

[Huf58] C. Huffaker. Experimental studies on predation: Dispersion factors and predator-prey oscillations. *Hilgardia*, 27(14):343–383, August 1958.

[IC07] Anthony R. Ives and Stephen R. Carpenter. Stability and Diversity of Ecosystems. *Science*, 317(5834):58–62, July 2007. Publisher: American Association for the Advancement of Science.

[Jan87] Wolfgang Jansen. A permanence theorem for replicator and Lotka-Volterra systems. *Journal of Mathematical Biology*, 25(4):411–422, September 1987.

[Jan97] Svante Janson. *Gaussian Hilbert Spaces*. Cambridge Tracts in Mathematics. Cambridge University Press, Cambridge, 1997.

[JM12] Adel Javanmard and Andrea Montanari. State Evolution for General Approximate Message Passing Algorithms, with Applications to Spatial Coupling. *arXiv:1211.5164 [cs, math, stat]*, December 2012. arXiv: 1211.5164.

[JMM+16] Claire Jacquet, Charlotte Moritz, Lyne Morissette, Pierre Legagneux, François Massol, Philippe Archambault, and Dominique Gravel. No complexity–stability relationship in empirical ecosystems. *Nature Communications*, 7(1):12573, November 2016.

[Jos06] Lou Jost. *Entropy and diversity. Oikos*, 113(2):363–375, May 2006.

[Jos07] Lou Jost. Partitioning Diversity into Independent Alpha and Beta Components. *Ecology*, 88(10):2427–2439, 2007.

[JS98] Vincent A. A. Jansen and Karl Sigmund. Shaken Not Stirred: On Permanence in Ecological Communities. *Theoretical Population Biology*, 54(3):195–201, December 1998.

[KD22] Elgin Korkmazhan and Alexander R. Dunn. High-order correlations in species interactions lead to complex diversity-stability relationships for ecosystems. *Physical Review E*, 105(1):014406, January 2022.

[KDS+13] Sonia Kéfi, Vasilis Dakos, Marten Scheffer, Egbert H. Van Nes, and Max Rietkerk. Early warning signals also precede non-catastrophic transitions. *Oikos*, 122(5):641–648, 2013.

[KK08] Krisztina Kiss and Sándor Kovács. Qualitative behavior of n-dimensional ratio-dependent predator–prey systems. *Applied Mathematics and Computation*, 199(2):535–546, June 2008.

[KLD+99] A. P. Kinzig, S. A. Levin, J. Dushoff, S. Pacala, and Associate Editor: Nicholas J. Gotelli. Limiting Similarity, Species Packing, and System Stability for Hierarchical Competition-Colonization Models. *The American Naturalist*, 153(4):371–383, 1999.

[KS15] David A. Kessler and Nadav M. Shnerb. Generalized model of island biodiversity. *Physical Review E*, 91(4):042705, April 2015.

[Lam19] A. Lamperski. Lemke's algorithm for linear complementarity problems, 2019. Publication Title: GitHub repository.

[Lat05] Rafał Latała. Some Estimates of Norms of Random Matrices. *Proceedings of the American Mathematical Society*, 133(5):1273–1282, 2005.

[LB92] Richard Law and Jerry C. Blackford. Self-Assembling Food Webs: A Global Viewpoint of Coexistence of Species in Lotka-Volterra Communities. *Ecology*, 73(2):567–578, April 1992.

[LBB22] Jinbao Liao, György Barabás, and Daniel Bearup. Competition–colonization dynamics and multimodality in diversity–disturbance relationships. *Ecology*, 103(5), May 2022.

[LBL20] Yinglin Li, Daniel Bearup, and Jinbao Liao. Habitat loss alters effects of intransitive higher-order competition on biodiversity: a new metapopulation framework. *Proceedings of the Royal Society B: Biological Sciences*, 287(1940):20201571, December 2020.

[LC71] R. Levins and D. Culver. Regional Coexistence of Species and Competition between Rare Species. *Proceedings of the National Academy of Sciences*, 68(6):1246–1248, June 1971.

[Lem65] Carlton Lemke. Bimatrix Equilibrium Points and Mathematical Programming. *Management Science*, 11(7):681–689, May 1965.

[Lev69] R. Levins. Some Demographic and Genetic Consequences of Environmental Heterogeneity for Biological Control. *Bulletin of the Entomological Society of America*, 15(3):237–240, September 1969.

[LH63] R. Levins and Harold Heatwole. On the distribution of organisms on islands. *Caribbean Journal of Science*, 3:173–177, January 1963.

[LH64] C. E. Lemke and Jr. Howson, J. T. Equilibrium Points of Bimatrix Games. *Journal of the Society for Industrial and Applied Mathematics*, 12(2):413–423, June 1964.

[LHM+04] M. A. Leibold, M. Holyoak, N. Mouquet, P. Amarasekare, J. M. Chase, M. F. Hoopes, R. D. Holt, J. B. Shurin, R. Law, D. Tilman, M. Loreau, and A. Gonzalez. The metacommunity concept: a framework for multi-scale community ecology: The metacommunity concept. *Ecology Letters*, 7(7):601–613, June 2004.

[Lia07] A. Liapounoff. Problème général de la stabilité du mouvement. *Annales de la Faculté des sciences de Toulouse : Mathématiques*, 9:203–474, 1907.

[LIPJ+06] Thomas M. Lewinsohn, Paulo Inácio Prado, Pedro Jordano, Jordi Bascompte, and Jens M. Olesen. Structure in plant–animal interaction assemblages. *Oikos*, 113(1):174–184, 2006.

[LLR83] M. R. Leadbetter, Georg Lindgren, and Holger Rootzén. *Extremes and Related Properties of Random Sequences and Processes*. Springer Series in Statistics. Springer New York, New York, NY, 1983.

[LM96] Richard Law and R. Daniel Morton. Permanence and the Assembly of Ecological Communities. *Ecology*, 77(3):762–775, April 1996.

[LMB+18] Pietro Landi, Henintsoa O. Minoarivelo, Åke Brännström, Cang Hui, and Ulf Dieckmann. Complexity and stability of ecological networks: a review of the theory. *Population Ecology*, 60(4):319–345, October 2018.

[LMG03] Michel Loreau, Nicolas Mouquet, and Andrew Gonzalez. Biodiversity as spatial insurance in heterogeneous landscapes. *Proceedings of the National Academy of Sciences*, 100(22):12765–12770, October 2003.

[LNI+01] M. Loreau, S. Naeem, P. Inchausti, J. Bengtsson, J. P. Grime, A. Hector, D. U. Hooper, M. A. Huston, D. Raffaelli, B. Schmid, D. Tilman, and D. A. Wardle. Biodiversity and Ecosystem Functioning: Current Knowledge and Future Challenges. *Science*, 294(5543):804–808, October 2001.

[Log93] Dmitriĭ Olegovich Logofet. *Matrices and graphs: stability problems in mathematical ecology*. CRC Press, Boca Raton, 1993.

[Log05] Dmitrii O. Logofet. Stronger-than-Lyapunov notions of matrix stability, or how "flowers" help solve problems in mathematical ecology. *Linear Algebra and its Applications*, 398:75–100, March 2005.

[Lot25] Lotka. *Elements of Physical Biology*. Williams and Wilkins Company, 1925.

[LW19] Clement Lee and Darren J. Wilkinson. A review of stochastic block models and extensions for graph clustering. *Applied Network Science*, 4(1):122, December 2019.

[Mac55] Robert MacArthur. Fluctuations of Animal Populations and a Measure of Community Stability. *Ecology*, 36(3):533–536, 1955.

[Mac70] Robert MacArthur. Species packing and competitive equilibrium for many species. *Theoretical Population Biology*, 1(1):1–11, May 1970.

[Mac84] Robert H. MacArthur. *Geographical Ecology: Patterns in the Distribution of Species*. Princeton University Press, July 1984.

[May72] Robert M. May. Will a Large Complex System be Stable? *Nature*, 238(5364):413–414, August 1972.

[May73] Robert M. May. *Stability and complexity in model ecosystems*. Number 6 in Monographs in population biology. Princeton University Press, Princeton, N.J, 1973.

[May76] Robert M. May. Simple mathematical models with very complicated dynamics. *Nature*, 261(5560):459–467, June 1976.

[McC00] Kevin Shear McCann. The diversity–stability debate. *Nature*, 405(6783):228–233, May 2000.

[MDC+17] F. Massol, M. Dubart, V. Calcagno, K. Cazelles, C. Jacquet, S. Kéfi, and D. Gravel. Chapter Four - Island Biogeography of Food Webs. In David A. Bohan, Alex J. Dumbrell, and François Massol, editors, *Advances in Ecological Research*, volume 56 of *Networks of Invasion: A Synthesis of Concepts*, pages 183–262. Academic Press, January 2017.

[Meh67] M. L. Mehta. *Random Matrices and the Statistical Theory of Energy Levels*. Academic Press, 1967.

[MG20] Sten Madec and Erida Gjini. Predicting N-Strain Coexistence from Cocolonization Interactions: Epidemiology Meets Ecology and the Replicator Equation. *Bulletin of Mathematical Biology*, 82(11):142, October 2020.

[MGMC13] Nicolas Mouquet, Dominique Gravel, François Massol, and Vincent Calcagno. Extending the concept of keystone species to communities and ecosystems. *Ecology Letters*, 16(1):1–8, 2013.

[MGPM06] Géza Meszéna, Mats Gyllenberg, Liz Pásztor, and Johan A.J. Metz. Competitive exclusion and limiting similarity: A unified theory. *Theoretical Population Biology*, 69(1):68–87, February 2006.

[ML02] Nicolas Mouquet and Michel Loreau. Coexistence in Metacommunities: The Regional Similarity Hypothesis. *The American Naturalist*, 159(4):420–426, April 2002.

[MM17] Catherine Matias and Vincent Miele. Statistical clustering of temporal networks through a dynamic stochastic block model. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 79(4):1119–1141, 2017.

[MN94] Robert M. May and Martin A. Nowak. Superinfection, Metapopulation Dynamics, and the Evolution of Diversity. *Journal of Theoretical Biology*, 170(1):95–114, September 1994.

[MN95] Robert M. May and Martin A. Nowak. Coinfection and the evolution of parasite virulence. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 261(1361):209–215, August 1995.

[MP67] V. A. Marčenko and L. A. Pastur. Distribution of eigenvalues in certain sets of random matrices. *Mathematics of the USSR-Sbornik*, 1(4):457, April 1967.

[MPV86] M Mezard, G Parisi, and M Virasoro. *Spin Glass Theory and Beyond: An Introduction to the Replica Method and Its Applications*, volume 9 of *World Scientific Lecture Notes in Physics*. World Scientific, November 1986.

[MR16] Andrea Montanari and Emile Richard. Non-Negative Principal Component Analysis: Message Passing Algorithms and Sharp Asymptotics. *IEEE Transactions on Information Theory*, 62(3):1458–1484, March 2016.

[MS17] Margaret M. Mayfield and Daniel B. Stouffer. Higher-order interactions capture unexplained complexity in diverse communities. *Nature Ecology & Evolution*, 1(3):0062, March 2017.

[MSS22] Immanuel Meyer, Bnaya Steinmetz, and Nadav M. Shnerb. How the storage effect and the number of temporal niches affect biodiversity in stochastic and seasonal environments. *PLOS Computational Biology*, 18(3):e1009971, March 2022.

[Mur72] Katta G. Murty. On the number of solutions to the complementarity problem and spanning properties of complementary cones. *Linear Algebra and its Applications*, 5(1):65–108, January 1972.

[Mur88] Katta G. Murty. *Linear complementarity, linear and nonlinear programming*. Number 3 in Sigma series in applied mathematics. Heldermann, Berlin, 1988.

[MW63] Robert H. MacArthur and Edward O. Wilson. An Equilibrium Theory of Insular Zoogeography. *Evolution*, 17(4):373–387, 1963.

[MW67] Robert H. MacArthur and Edward O. Wilson. *The Theory of Island Biogeography*. Princeton University Press, 1967.

[NA16] Ben C. Nolting and Karen C. Abbott. Balls, cups, and quasi-potentials: quantifying stability in stochastic systems. *Ecology*, 97(4):850–864, 2016.

[Nau12] Alexey Naumov. Elliptic law for real random matrices, August 2012. arXiv:1201.1639 [math].

[NC97] Michael G. Neubert and Hal Caswell. Alternatives to Resilience for Measuring the Responses of Ecological Systems to Perturbations. *Ecology*, 78(3):653–665, 1997. Publisher: Ecological Society of America.

[New06] M. E. J. Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582, June 2006.

[NM92] Sean Nee and Robert M. May. Dynamics of Metapopulations: Habitat Destruction and Competitive Coexistence. *Journal of Animal Ecology*, 61(1):37–40, 1992.

[NM94] Martin A. Nowak and Robert Mccredie May. Superinfection and the evolution of parasite virulence. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 255(1342):81–89, January 1994.

[NO15] Hoi H. Nguyen and Sean O'Rourke. The Elliptic Law. *International Mathematics Research Notices*, 2015(17):7620–7689, January 2015.

[OD92] Manfred Opper and Sigurd Diederich. Phase transition and 1/f noise in a game dynamical model. *Physical Review Letters*, 69(10):1616–1619, September 1992.

[OH01] Otso Ovaskainen and Ilkka Hanski. Spatially Structured Metapopulation Models: Global and Local Assessment of Metapopulation Capacity. *Theoretical Population Biology*, 60(4):281–302, December 2001.

[OR14] Sean O'Rourke and David Renfrew. Low rank perturbations of large elliptic random matrices. *arXiv:1309.5326 [math-ph]*, May 2014.

[Ost56] Alexander Ostrowski. Determinanten mit überwiegender Hauptdiagonale und die absolute Konvergenz von linearen Iterationsprozessen. *Commentarii mathematici Helvetici*, 30:175–210, 1956.

[PA13] Ian S. Pearse and Florian Altermatt. Extinction cascades partially estimate herbivore losses in a complete Lepidoptera—plant food web. *Ecology*, 94(8):1785–1794, 2013.

[Pai66] Robert T. Paine. Food Web Complexity and Species Diversity. *The American Naturalist*, 100(910):65–75, 1966.

[Pai69] Robert T. Paine. The Pisaster-Tegula Interaction: Prey Patches, Predator Food Preference, and Intertidal Community Structure. *Ecology*, 50(6):950–961, 1969.

[PBG22] Lyle Poley, Joseph W. Baron, and Tobias Galla. Generalised Lotka-Volterra model with hierarchical interactions, August 2022.

[PBHM19] Clàudia Payrató-Borràs, Laura Hernández, and Yamir Moreno. Breaking the Spell of Nestedness: The Entropic Origin of Nestedness in Mutualistic Systems. *Physical Review X*, 9(3):031024, August 2019.

[PEM12] Michael J. O. Pocock, Darren M. Evans, and Jane Memmott. The Robustness and Restoration of a Network of Ecological Networks. *Science*, 335(6071):973–977, February 2012.

[Pim79] Stuart L. Pimm. The structure of food webs. *Theoretical Population Biology*, 16(2):144–158, 1979.

[PLC91] Stuart L. Pimm, John H. Lawton, and Joel E. Cohen. Food web patterns and their consequences. *Nature*, 350(6320):669–674, April 1991.

[Ple77] R. J. Plemmons. M-matrix characterizations.I—nonsingular M-matrices. *Linear Algebra and its Applications*, 18(2):175–188, January 1977.

[PNJ21] Susanne Pettersson and Martin Nilsson Jacobi. Spatial heterogeneity enhance robustness of large multi-species ecosystems. *PLOS Computational Biology*, 17(10):e1008899, October 2021.

[PR98] Stephen W. Pacala and Mark Rees. Models Suggesting Field Experiments to Test Two Hypotheses Explaining Successional Diversity. *The American Naturalist*, 152(5):729–737, November 1998.

[PRS13] Alessandro Pizzo, David Renfrew, and Alexander Soshnikov. On finite rank deformations of Wigner matrices. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 49(1):64–94, February 2013.

[PSJ20] Susanne Pettersson, Van M. Savage, and Martin Nilsson Jacobi. Stability of ecosystems enhanced by species-interaction constraints. *Physical Review E*, 102(6):062405, December 2020.

[PSNJ20] Susanne Pettersson, Van M. Savage, and Martin Nilsson Jacobi. Predicting collapse of complex ecological systems: quantifying the stability–complexity continuum. *Journal of The Royal Society Interface*, 17(166):20190391, May 2020.

[Pé06] S. Péché. The largest eigenvalue of small rank perturbations of Hermitian random matrices. *Probability Theory and Related Fields*, 134(1):127–173, January 2006.

[RBBB20] Felix Roy, Matthieu Barbier, Giulio Biroli, and Guy Bunin. Complex interactions can create persistent fluctuations in high-diversity ecosystems. *PLOS Computational Biology*, 16(5):e1007827, 2020.

[RBBC19] F Roy, G Biroli, G Bunin, and C Cammarota. Numerical implementation of dynamical mean field theory for disordered systems: application to the Lotka–Volterra model of ecosystems. *Journal of Physics A: Mathematical and Theoretical*, 52(48):484001, November 2019.

[RBL+19] Tamara N. Romanuk, Amrei Binzer, Nicolas Loeuille, W. Mather A. Carscallen, and Neo D. Martinez. Simulated evolution assembles more realistic food webs with more functionally similar species than invasion. *Scientific Reports*, 9(1):18242, December 2019.

[RCL18] Emlyn J. Resetarits, Sara E. Cathey, and Mathew A. Leibold. Testing the keystone community concept: effects of landscape, patch removal, and environment on metacommunity structure. *Ecology*, 99(1):57–67, 2018.

[RMGM06] Neil Rooney, Kevin McCann, Gabriel Gellner, and John C. Moore. Structural asymmetry and the stability of diverse food webs. *Nature*, 442(7100):265–269, July 2006.

[Roh12] Jiri Rohn. On Rump's characterization of P-matrices. *Optimization Letters*, 6(5):1017–1020, June 2012.

[Ros13] Axel G. Rossberg. *Food webs and biodiversity*. Wiley-Blackwell, Chichester, West Sussex, UK, 2013.

[RR85] John D. Rummel and Jonathan Roughgarden. A Theory of Faunal Buildup for Competition Communities. *Evolution*, 39(5):1009, September 1985.

[RS13] David Renfrew and Alexander Soshnikov. On finite rank deformations of wigner matrices ii: delocalized perturbations. *Random Matrices: Theory and Applications*, 02(01):1250015, January 2013.

[RSB14] Rudolf P. Rohr, Serguei Saavedra, and Jordi Bascompte. On the structural stability of mutualistic systems. *Science*, 345(6195):1253497, July 2014.

[Rum03] Siegfried M. Rump. On P-matrices. *Linear Algebra and its Applications*, 363:237–250, April 2003.

[RW84] Ray Redheffer and Wolfgang Walter. Solution of the stability problem for a class of generalized volterra prey-predator systems. *Journal of Differential Equations*, 52(2):245–263, April 1984.

[RZB+09] Tamara N. Romanuk, Yun Zhou, Ulrich Brose, Eric L. Berlow, Richard J. Williams, and Neo D. Martinez. Predicting invasion success in complex ecological networks. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1524):1743–1754, June 2009.

[RZVM17] T. N. Romanuk, Y. Zhou, F. S. Valdovinos, and N. D. Martinez. Chapter Five - Robustness Trade-Offs in Model Food Webs: Invasion Probability Decreases While Invasion Consequences Increase With Connectance. In David A. Bohan, Alex J. Dumbrell, and François Massol, editors, *Advances in Ecological Research*, volume 56 of *Networks of Invasion: A Synthesis of Concepts*, pages 263–291. Academic Press, January 2017.

[SA21] Carlos A. Serván and Stefano Allesina. Tractable models of ecological assembly. *Ecology Letters*, 24(5):1029–1037, May 2021.

[SB11] Daniel B. Stouffer and Jordi Bascompte. Compartmentalization increases food-web persistence. *Proceedings of the National Academy of Sciences*, 108(9):3648–3652, March 2011.

[SC95] Jack W. Silverstein and Sang-Il Choi. Analysis of the Limiting Spectral Distribution of Large Dimensional Random Matrices, 1995.

[SCG+05] D. B. Stouffer, J. Camacho, R. Guimerà, C. A. Ng, and L. A. Nunes Amaral. Quantitative Patterns in the Structure of Model and Empirical Food Webs. *Ecology*, 86(5):1301–1311, 2005.

[SCG+18] Carlos A. Serván, José A. Capitán, Jacopo Grilli, Kent E. Morrison, and Stefano Allesina. Coexistence of many species in random ecosystems. *Nature Ecology & Evolution*, 2(8):1237–1242, August 2018.

[SGB+15] Samir Suweis, Jacopo Grilli, Jayanth R. Banavar, Stefano Allesina, and Amos Maritan. Effect of localization on the stability of mutualistic ecological networks. *Nature Communications*, 6(1):10179, December 2015.

[SKA13] Phillip P. A. Staniczenko, Jason C. Kopp, and Stefano Allesina. The ghost of nestedness in ecological networks. *Nature Communications*, 4(1):1391, January 2013.

[Sla74] Montgomery Slatkin. Competition and Regional Coexistence. *Ecology*, 55(1):128–134, 1974.

[Smi49] N. V. Smirnov. Limit distributions for the terms of a variational series. *Akademiya Nauk SSSR. Trudy Matematicheskogo Instituta imeni V. A. Steklova*, 25:60, 1949.

[SRB+17] Serguei Saavedra, Rudolf P. Rohr, Jordi Bascompte, Oscar Godoy, Nathan J. B. Kraft, and Jonathan M. Levine. A structural approach for understanding multispecies coexistence. *Ecological Monographs*, 87(3):470–486, 2017.

[SSE13] Torbjörn Säterberg, Stefan Sellman, and Bo Ebenman. High frequency of functional extinctions in ecological networks. *Nature*, 499(7459):468–470, July 2013.

[Sto16] Lewi Stone. The Google matrix controls the stability of structured ecological and biological networks. *Nature Communications*, 7(1):12857, November 2016.

[Sto18] Lewi Stone. The feasibility and stability of large complex biological networks: a random matrix approach. *Scientific Reports*, 8(1):8246, December 2018.

[Sto20] Lewi Stone. The stability of mutualism. *Nature Communications*, 11(1):2648, December 2020.

[TA80] Yasuhiro Takeuchi and Norihiko Adachi. The existence of globally stable equilibria of ecosystems of the generalized Volterra type. *Journal of Mathematical Biology*, 10(4):401–415, December 1980.

[Tak96] Y. Takeuchi. *Global dynamical properties of Lotka-Volterra systems*. World Scientific, Singapore ; River Edge, NJ, 1996.

[Tan39] A. G. Tansley. British Ecology During the Past Quarter-Century: The Plant Community and the Ecosystem. *Journal of Ecology*, 27(2):513–530, 1939.

[Tao13] Terence Tao. Outliers in the spectrum of iid matrices with bounded rank perturbations. *Probability Theory and Related Fields*, 155(1):231–263, 2013.

[TAT78] Yasuhiro Takeuchi, Norihiko Adachi, and Hidekatsu Tokumaru. Global stability of ecosystems of the generalized volterra type. *Mathematical Biosciences*, 42(1):119–136, November 1978.

[Tay88] Peter J. Taylor. Consistent scaling and parameter choice for linear and Generalized Lotka-Volterra models used in community ecology. *Journal of Theoretical Biology*, 135(4):543–568, December 1988.

[TF10] Elisa Thébault and Colin Fontaine. Stability of ecological communities and the architecture of mutualistic and trophic networks. *Science*, 329(5993):853–856, 2010.

[Til94] David Tilman. Competition and Biodiversity in Spatially Structured Habitats. *Ecology*, 75(1):2–16, January 1994.

[TMLN94] David Tilman, Robert M. May, Clarence L. Lehman, and Martin A. Nowak. Habitat destruction and the extinction debt. *Nature*, 371(6492):65–66, September 1994.

[Tok04] Kei Tokita. Species Abundance Patterns in Complex Evolutionary Dynamics. *Physical Review Letters*, 93(17):178102, October 2004.

[TPA14] Si Tang, Samraat Pawar, and Stefano Allesina. Correlation between interaction strengths drives stability in large ecological networks. *Ecology Letters*, 17(9):1094–1100, September 2014.

[TVK10] Terence Tao, Van Vu, and Manjunath Krishnapur. Random matrices: Universality of ESDs and the circular law. *The Annals of Probability*, 38(5):2023–2065, September 2010.

[VAN15] Marcos Costa Vieira and Mário Almeida-Neto. A simple stochastic model for complex coextinctions in mutualistic networks: robustness decreases with connectance. *Ecology Letters*, 18(2):144–152, 2015.

[VML04] Evan A. Variano, Jonathan H. McCoy, and Hod Lipson. Networks, Dynamics, and Modularity. *Physical Review Letters*, 92(18):188701, May 2004. Publisher: American Physical Society.

[Vol26] Vito Volterra. Fluctuations in the Abundance of a Species considered Mathematically1. *Nature*, 118(2972):558–560, October 1926.

[VPNJ22] Ankit Vikrant, Susanne Pettersson, and Martin Nilsson Jacobi. Spatial coherence and the persistence of high diversity in spatially heterogeneous landscapes. *Ecology and Evolution*, 12(6):e9004, 2022.

[Wan78] Peter J. Wangersky. Lotka-Volterra Population Models. *Annual Review of Ecology and Systematics*, 9:189–218, 1978.

[WBvN+21] Shaopeng Wang, Ulrich Brose, Saskya van Nouhuys, Robert D. Holt, and Michel Loreau. Metapopulation capacity determines food chain length in fragmented landscapes. *Proceedings of the National Academy of Sciences*, 118(34):e2102733118, August 2021.

[Whi72] R. H. Whittaker. Evolution and Measurement of Species Diversity. *Taxon*, 21(2/3):213–251, 1972.

[Wig55] Eugene P. Wigner. Characteristic Vectors of Bordered Matrices With Infinite Dimensions. *Annals of Mathematics*, 62(3):548–564, 1955. Publisher: Annals of Mathematics.

[Wig67] Eugene P. Wigner. Random Matrices in Physics. *SIAM Review*, 9(1):1–23, January 1967.

[Wil92] David Sloan Wilson. Complex Interactions in Metacommunities, with Implications for Biodiversity and Higher Levels of Selection. *Ecology*, 73(6):1984–2000, 1992.

[Wis28] John Wishart. The Generalised Product Moment Distribution in Samples from a Normal Multivariate Population. *Biometrika*, 20A(1/2):32–52, 1928.

[WM00] Richard J. Williams and Neo D. Martinez. Simple rules yield complex food webs. *Nature*, 404(6774):180–183, March 2000.

[WTL00] Lars Witting, Jürgen Tomiuk, and Volker Loeschcke. Modelling the optimal conservation of interacting species. *Ecological Modelling*, 125(2):123–144, January 2000.

[YA73] J. A. Yorke and W. N. Anderson. Predator-Prey Patterns. *Proceedings of the National Academy of Sciences*, 70(7):2069–2071, July 1973.

[Yod81] P. Yodzis. The stability of real ecosystems. *Nature*, 289(5799):674–676, February 1981.

[YW01] Douglas W. Yu and Howard B. Wilson. The Competition-Colonization Trade-off Is Dead; Long Live the Competition-Colonization Trade-off. *The American Naturalist*, 158(1):49–63, July 2001.

[ZBN+21] Helin Zhang, Daniel Bearup, Ivan Nijs, Shaopeng Wang, György Barabás, Yi Tao, and Jinbao Liao. Dispersal network heterogeneity promotes species coexistence in hierarchical competitive communities. *Ecology Letters*, 24(1):50–59, 2021.

[Zee95] Mary Lou Zeeman. Extinction in Competitive Lotka-Volterra Systems. *Proceedings of the American Mathematical Society*, 123(1):87–96, 1995.

Étude des grands systèmes de Lotka-Volterra : l'écologie théorique à travers les matrices aléatoires

Mots clés : Écologie théorique, Systèmes dynamiques, Matrice aléatoire, Lotka-Volterra, Écologie des communautés, Stabilité, Diversité, Faisabilité, Métacommunauté.

La diversité des espèces et la complexité de leurs interactions représentent un défi important en écologie théorique. La difficulté à analyser ces systèmes rend nécessaire le recours à la modélisation mathématique. Le système de Lotka-Volterra forme un modèle simple, robuste et polyvalent utilisé pour décrire de grands systèmes en interaction tels que les réseaux trophiques ou les microbiomes. Ce modèle est constitué de n équations différentielles couplées reliant les abondances des différentes espèces présentes dans le système. Lorsque le nombre d'espèces devient très important, les paramètres du modèle sont trop nombreux pour pouvoir être observés ou estimés correctement. Par conséquent, les interactions entre les différentes espèces peuvent être modélisées comme des variables aléatoires afin de comprendre la dynamique du système. Dans cette thèse, je développe une analyse quantitative des grands systèmes de Lotka-Volterra en m'appuyant sur la théorie des matrices aléatoires et des simulations numériques. Je me focalise d'abord sur l'existence d'une sous-population stable dont je décris les propriétés à l'équilibre. Ensuite, j'étudie l'existence d'un seuil critique permettant la faisabilité de l'équilibre (:= toutes les espèces du système survivent) lorsque les interactions sont corrélées par paires. Une meilleure compréhension de ces phénomènes permet d'étendre ces propriétés à une structure d'interaction par blocs décrivant un modèle multi-communautés. J'analyse les propriétés (faisabilité, existence d'un phénomène d'attrition au sein de chaque communauté) de communautés distinctes en ajustant les interactions inter- et intra-communautaires. Dans une dernière partie, je propose une interprétation probabiliste d'un modèle de compétition multi-espèces dans lequel on modélise non pas l'abondance des espèces à un niveau local mais leurs occurrences par site au niveau du paysage. J'examine dans ce modèle le compromis compétition-colonisation lorsque les paramètres de colonisation suivent une distribution de probabilité donnée. Les résultats obtenus dans ces différents chapitres démontrent l'existence de "lois asymptotiques" régissant le comportement des modèles écologiques lorsque le nombre d'espèces devient très grand.

Large Lotka-Volterra model: when random matrix theory meets theoretical ecology

Keywords: Theoretical ecology, Dynamical systems, Random matrix, Lotka-Volterra, Community ecology, Stability, Diversity, Feasibility, Metacommunity.

The diversity of species and the complexity of their interactions represent a huge challenge in theoretical ecology. The system's complexity requires mathematical modelisation. The Lotka-Volterra system forms a simple, robust and versatile model used to describe large interacting systems such as food webs or microbiomes. This model consists of n coupled differential equations linking the abundances of the different species present in the system. When the number of species becomes very large, the model parameters are too large to be observed or estimated precisely. Therefore, the interactions between the different species can be modeled as random variables to understand the dynamics of the system. In this thesis, I develop a quantitative analysis of large Lotka-Volterra systems based on random matrix theory and numerical simulations. I first focus on the existence of a stable subpopulation for which I describe the equilibrium properties. Then, I study the existence of a critical threshold that allows the feasibility of equilibrium (:= all species in the system survive) when the interactions are pairwise correlated. A better understanding of these phenomena allows to extend these properties to a block interaction structure describing a multi-community model. I analyze the properties (feasibility, existence of attrition within each community) of distinct communities by adjusting the inter- and intra-community interactions. In a final section, I propose a probabilistic interpretation of a hierarchical competition-colonization trade-off model in which we model not the abundance of species at a local level but their occurrences by patch at the landscape level. I examine in this model the competition-colonization trade-off when the colonization parameters follow a given probability distribution. The results obtained in these different chapters show the existence of "asymptotic laws" governing the behavior of ecological models when the number of species becomes very large.