THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Structure, stability and macroecology of populations and ecosystems

Ankit Vikrant

Department of Space, Earth and Environment
Chalmers University of Technology
Gothenburg, Sweden, 2024

Structure, stability and macroecology of populations and ecosystems

Ankit Vikrant

ISBN 978-91-8103-072-3
Doktorsavhandlingar vid Chalmers tekniska högskola, Ny serie nr 5530
ISSN 0346-718X

Department of Space, Earth and Environment
Chalmers University of Technology
SE-412 96 Gothenburg, Sweden
www.chalmers.se

This thesis has been typeset using LaTeX.

Printed by Chalmers Reproservice
Gothenburg, Sweden, July2024

Structure, stability and macroecology of populations and ecosystems

ANKIT VIKRANT

Division of Physical Resource Theory
Department of Space, Earth and Environment
Chalmers University of Technology

Abstract

Dispersal and interaction networks underlie many ecological systems, often dictating spatio-temporal dynamics as well as emergent patterns. The mode and time scales of dispersal can vary considerably depending on whether ecosystems exist in contiguous landscapes or in island-like configurations. In particular, island systems have been extensively studied for many ecological processes, in part because of the simplicity in representing processes such as immigration of organisms from outside. This representation has also been used in many other habitats, such as mountain tops which can also be considered island-like because of their isolation and similar properties of immigration.

Immigration could result from species that already exist on the islands or those that are novel. This distinction connects to the concept of species invasions which are driven by new species not present in an ecosystem before. Invasions can be seen as an important part of how ecosystems are assembled. Invasive species can also have disruptive impacts on systems like islands, which for example can introduce novel species like predators. Invasions serve as a useful representation of various means and ends of biological processes, and hence their study has extended way beyond their initial scope.

Dispersal, interactions and invasions have implications ranging from establishment of distinct subpopulations of individual species to large scale spatial patterns that define ecosystems. Such implications and frameworks to explain them are the central themes of this thesis.

One of the systems we study consists of sites along a coastal landscape that harbours the seagrass *Halodule uninervis*. The task is to partition the entire population into subpopulations that have very little dispersal across them. We use an algorithm to aggregates sites that are more strongly connected in the same subpopulation for a range of dispersal scenarios. These scenarios

are useful since the dispersal processes and life-spans of the seagrass are not well-understood.

The other articles focus on ecological communities with a range of interaction types between species. We study how ecological communities on islands are assembled from a large pool of species that can also immigrate from the mainland source, and if it can provide mechanisms to explain the scaling of species richness with island areas – namely the species-area relationship. A drastically different setting is large contiguous landscapes.

We posit that species interactions vary across space, which helps characterize habitat patches that would be connected by high dispersal of species. This simplified picture provides a range of situations to understand how habitat heterogeneity affects species richness.

We also address the problem of how ecosystems with many interacting species can still be stable. We do this by deriving a very general mathematical bound on stability which is expressed in terms of ecologically realistic constraints.

Species identities are important with respect to biological invasions for the risk of replacement of resident species. We developed a framework to predict the structure of invasion outcomes, i.e., what set of species survive following an invasion event. This method is applicable in a range of cases characterized by high initial abundance of invaders and imperfect knowledge of interactions.

An overarching theme in investigating these complex ecological systems is to first analyze the fully random case without assuming any structure. This null case admits description in terms of a few statistics that greatly simplify the study of system level properties and processes. One can then ask if there are parameter regimes where these properties break down. The difficulty in finding such regimes might indicate universal properties but even the sudden disappearance might unravel phase transitions or unexplored new properties.

Keywords: Complex ecological systems, species area relationships, metacommunity, community assembly, ecosystem stability, biological invasions

List of Publications

This thesis is based on the following publications:

Paper 1
A. Vikrant and M. Nilsson Jacobi , "Complex ecological communities and the emergence of island species-area relationships". *Theoretical Ecology* 15, no. 4 (2022): 311- 320.

Paper 2
A. Vikrant[†] , S. Pettersson[†] , M. Nilsson Jacobi , "Spatial coherence and the persistence of high diversity in spatially heterogeneous landscapes". *Ecology and Evolution* 12, no. 6 (2022): e9004.

Paper 3
A. Vikrant , "Predicting the structure of invasion outcomes using proximate systems". *To be submitted.*

Paper 4
V. Savage, M. Nilsson Jacobi, A. Vikrant, I. Tripathi and O. Venturelli , "Ecological Constraints Beget Both Complexity and Stability: Building More General Mathematical Bounds for Real Systems". *To be submitted.*

Paper 5
R. D. Evans, K.M. McMahon, K. -J. van Dijk, K. Dawkins, M. Nilsson Jacobi, A. Vikrant , "Identification of dispersal barriers for a colonising seagrass using seascape genetic analysis". *Science of the Total Environment* (2021), 763, 143052..

Contributions

Paper 1
AV conceived and designed the study, AV performed the simulations and wrote the paper, AV and MNJ analyzed the results.

Paper 2
Conceptualisation, interpretation of results and, review and editing: MNJ, SP,

AV; Simulations, visualisation and original draft: SP, AV; Analytic results: SP

Paper 3
AV conceived and designed the study, performed the simulations, analyzed the results and wrote the paper.
Paper 4
Conceptualisation: VS, MNJ; Methodology: VS, MNJ, AV; Formal analysis: VS, AV, IT; Writing - original draft: VS; Interpretation of results - VS, AV; Writing - review and editing: VS, AV, IT, OV

Paper 5
RDE, KMM: Conceptualization, Methodology, Formal analysis, Writing - original draft, Project administration; KJvD, KD, MNJ, AV: Methodology, Formal analysis, Writing - review and editing.

Acknowledgments

Scientific research is largely an international and collaborative endeavour, and working toward my PhD has given me a peek into that. I want to thank all the people who have made these years more fun and eventful. Firstly, I'd like to thank the division of Physical Resource Theory at Chalmers for funding my PhD and allowing me to steer my research, teaching and networking with much freedom and agency.

I want to thank my supervisors, mentors and co-authors over the years. Martin for introducing me to the craft, making white-board discussions enjoyable and opening up opportunities for me to network. Susanne for being a very brilliant, kind and fun co-author and supervisor. Thank you for your time listening to my work and reading my chaotic drafts. Claes for being my main supervisor and anchoring the complex systems group at the division. You are a constant source of conversations on evolution and human cultures, and I am glad that the course we ran for the past couple of years will be renewed for the coming years. My co-author and mentor Van has guided and inspired me in many ways. Thank you for being super kind with your time. It is delightful watching you approach problems with a rare blend of passion, rigour and persistence, and I have thoroughly enjoyed working with you. Kristian, thanks for your time as my first examiner. I am glad to have taught the information theory course with you for all the discussions on the subject and your brilliant research in the field. Christian for your support as an examiner and for propelling things forward for me. Thanks for also being a fun person to be with.

Martin Persson for keeping the division alive with board game nights and for being caring and efficient in your role as the director of graduate studies. Frances, thanks for being an available and fun manager, and for inducting me in to the running group.

Many people at the division made my time super enjoyable. Avi and Çağlar, my office besties for keeping the office and my life super lively. Ahmet, you have been the go-to guy for almost everything and thanks for putting together some of the best times at work and outside, and of course the barbecues! Andreas, Caroline, David, Devon, Ella, Gavin, Hanna, Jinxi, Manu, Markus, Ricardo, Ridwan, Sofia, Sonia, Vadim, Xiaoming, Yodefia, Yuan – I have shared some great memories with all of you, thank you for giving me more reasons to come to work! A special thanks to Angelica for always being super

available, kind and loving.

My friends from BITS have had a great influence on my life for almost 15 years now. Abhiroop, I can't remember a day without fun and laughs in all the time we spent together. Thank you for opening me up to so many things in life. Shila, thanks for always putting together great plans and reminding us to have fun. Akshay, thank you for the best conversations and birding trips, especially during the last few years. I still need to reply to your messages while I write this, but thanks for being very deliberate and involved in all the things you say and do.

I want to thank my family for being a constant source of love and for everything in life. Swapna, thank you for being my greatest cheerleader and my harshest critic.

Acronyms

SAR:	Species-area relationship
gLV:	Generalized Lotka-Volterra
AHTO:	Area-heterogeneity trade-off
ESS:	Evolutionarily Stable Strategy
EER:	Energetic Equivalence Rule
MTE:	Metabolic Theory of Ecology
ODB:	Oceanographic dispersal barrier
FDD:	Fragment dispersal duration
KLD:	Kullback-Leibler divergence
ESD:	Empirical spectral density
SLM:	Stochastic Logistic Model
SLVM:	Stochastic Lotka-Volterra Model

Contents

CHAPTER 1

Introduction

Biological species interact in myriad ways with one another and their environment to generate the diversity of life that we see around us. In addition to ecology that concerns various aspects of organisms and their interactions with the environment, evolution is another cornerstone of organismal biology. Unlike physical systems – whose fundamental units are assumed to be immutable – biological entities could undergo changes through evolution, which has implications for species as well as the ecosystems they inhabit. Variability in the basic units of an ecological system could be further pruned and filtered by how individuals, species (and collections thereof) disperse in space. Processes on ecological time scales such as variation in resource abundance, temperature driven effects on metabolism affect how species disperse in space, which results in the emergence of widespread spatial patterns and laws. Abundance patterns such as persistent fluctuations can be driven endogenously by spatial dispersal and are signatures of highly diverse and complex ecological communities [1–3].

The oldest and the most prominent among such spatial laws is the species-area relationship (SAR) that describes how the number of species in a patch scale as the area of the patch increases. Two functional forms of the SAR have

been most widely observed in empirical studies – the semi-log and the power law form [4]. The power-law form has especially been the subject of many studies trying to explain its origins. Preston used log-normal abundance distributions – that are widely observed – to argue for a power-law form of the SAR [5]. This work also demonstrated differences in studying nested samples in a contiguous landscape versus a system of isolates such as islands. A qualitatively different suggestion was that islands could be at internal equilibrium that constrained patterns of species diversity on them [5].

The simplicity of island systems prompted MacArthur and Wilson to propose an equilibrium theory of island biogeography [6]. They argued that the number of species found on any island depends on an equilibrium between immigration and extinction rates. Further, since increasing island size leads to higher immigration rates but lower extinction rates, one could also relate the number of species to island area. This simple model also inspired the study of species diversity patterns in other island-like ecosystems [7–9]. The fact that species diversity patterns could be reproduced without invoking species differences, has also been the basis of more successful recent theories (cite). However, the same simplifications prevent these theories from capturing temporal patterns and processes such as those related to population dynamics [10].

Island systems give way to the idea of having discrete spatial patches representing different populations that could be connected by migration or gene-flow. The concept of a metapopulation – meaning population of populations – had some primitive analogues already in the 1950's [11], but the word was first used by Levins in a sense that is closer to its contemporary usage [12]. A metapopulation could witness local extinction of the species at one spatial patch that might be later recolonized through dispersal from other patches. Mainland-island systems are a special case of metapopulations where the mainland prevents local extinctions on the islands. Given a spatial distribution of many different populations, one could classify them into multiple metapopulations depending upon factors that impede migration or gene-flow across populations. Real ecosystems are much more complex in terms of the number of species they harbour and also the kinds of interactions between them. The idea that complex interactions varying in space can affect diversity patterns was the basis of the metacommunity idea.

Metacommunities can be studied in various paradigms that differ in the

amount of dispersal and the variation in species traits resulting from environmental heterogeneity [13, 14]. The amount of dispersal between local communities dictates the influence of environmental stochasticity on community assembly at different scales. High dispersal not only acts to homogenize local communities across patches but it also increases the pace of community assembly such that stochastic effects from the environment have minimal bearing on the fate of these communities. This thesis considers metacommunities harbouring many interacting species such that these species cannot be reduced to their positions along one or a few trait axes. This operational definition implies that habitat heterogeneity is captured through variation in interspecies interaction strengths across space, as in [1].

Across ecological studies, arguably the most common response variable is species richness defined as the number of species. Any serious characterization of biodiversity inevitably boils down to the definition of diversity, particularly how it can quantified. Many widely used diversity metrics in ecology are either agnostic of species identities or do not have a clear relationship with the functioning of ecosystems. It's not a surprise that many celebrated laws in ecology are also framed in terms of metrics like species richness. These limitations restrict the ways in which processes like biological invasions could be understood. Species identities are important with respect to biological invasions for the risk of replacement of resident species. Species replacement might preserve the number of species, but it could alter the composition of the resident community resulting in homogenization of species [15]. This alteration could eliminate certain functional aspects of ecosystems, which calls for including measures of functional diversity.

A major focus of community ecology is to characterize ecosystem function which is related to the diversity of species traits. Trait-based approaches in ecology rest on the assumption that ecosystem function and community-wide impacts of environmental change are better explained by traits of dominant species rather than the species richness of communities. It has become increasingly clear that ecosystem functioning does not always increase with species richness [16, 17], partly because the same environmental variables could influence both richness and functional aspects of ecosystems. Alternative approaches or extensions have helped shift the focus from just looking at plain diversity metrics or trait diversity.

One such leading framework connects the frequency distribution of traits

associated with metabolism to the rate of biomass growth, and eventually informs estimation of the total biomass of the community – the Trait Driver Theory [18]. This powerful tool quantifies environmental impacts on the full trait distribution and shows how mean, variance and other higher moments of this distribution affect measures of ecosystem function such as productivity. This is partly enabled by the Metabolic Theory of Ecology (MTE) that studies the dependence of population, community and ecosystem level processes on body size and temperature [19]. MTE achieves this by building on the relationship of individual metabolic rate with individual body mass [20] and temperature [21].

Metabolic scaling has also been used to study population dynamics of complex food-webs [22, 23]. This is possible since biological quantities such as consumption rates and handling time scale in predictive ways with body mass. This also constrains the prey/resource body sizes that are available to predators/consumers. While the stability of food-webs has been addressed using such constraints [23], there is much potential to combine them with the rich body of literature on general mathematical bounds on the stability of ecological systems.

1.1 Motivation and Aim

Macroecological patterns are those that show up at large spatial and temporal scales. Many of these patterns are not species specific and might hold for a range of taxa. The population dynamics of more than one ecological community is invariably a component of how these patterns emerge. At smaller spatial scales, processes like species invasion are commonplace that not only depend on the attributes of the invading species but also the mode of assembly of the resident ecological community. This PhD thesis uncovers large scale patterns resulting from species dispersal and the assembly of ecological communities that are in turn shaped by processes like invasions. The goal is to find limits within which these patterns and processes operate by disentangling the role of heterogeneity in the functional form of interactions, the parameters in space and across species, and the procedures to assemble communities.

Many approaches explain species coexistence in a low-dimensional parameter space such as when species can be described using a few traits [24–26]. However, these methods have limited success when ecological communi-

ties are extremely high-dimensional [27], especially ones with a large number of species such that species interactions cannot be ignored and there is no clear niche-partitioning. One broad aim of my recent work is to describe the emergence of spatial laws and patterns in such high-dimensional ecosystems. Specifically, is to possible to understand the emergence of the widely observed SAR for high-dimensional communities? If yes, then how can one study this relationship using simple models of such communities? Further, what ecological systems are best suited to scrutiny using such simplified models? (Paper 1)

Ecosystems are better described as spatially-extended rather than well-mixed systems. This implies that species can move around across different locations and species interactions could vary spatially. For anyone who has worked on stability of large complex ecosystems, it is natural to ask whether a fully-feasible and stable ecosystem is possible for arbitrarily complex local communities? What kind of dispersal processes and spatial patterns of species interactions could enable highly complex yet stable ecosystems, theoretically? Further, do large contiguous landscapes entail high species diversity when understood as collections of many local communities connected by dispersal? (Paper 2)

Ecological communities are not closed systems operating entirely on endogenous processes. External influences can rewire the interactions between constituent species, which can lead to the system resting at a different state that might or might not be qualitatively similar to the original state's structure. Examples including extinctions that not only reorganize the abundances of the species, but remove one or more species resulting in a different structural state. We consider the detailed characterization of such processes that fall under the umbrella of invasions. What kind of representations can capture a very general class of invasion processes while still keeping track of the species identities? In paper 3, we address such questions and more by also accounting for the noisy nature of ecosystems that precludes a perfect estimation of interspecies interactions.

An implicit assumption while finding generalizable results for high-dimensional ecosystems is the system being at an equilibrium. Even when changes in the state of a community is considered, we assume that the system starts off at some stable equilibrium and eventually rests at the same or different stable equilibrium. The patterns most often observed in nature are those that should

persist in spite of events that perturb the system in different ways. This is the key motive for the systematic study of dynamical systems. Only some kinds of stability lend themselves to rigorous mathematical analysis. Specifically, it is the study of whether a system reverts to an equilibrium after it has been slightly perturbed, that has been a concern of a very large number of papers. Some very general mathematical bounds have been found in this regard [28], which have not only shaped entire sub-fields within community ecology but also helped build intuition about trade-offs in nature mediating stability. Such results have been extended to some other types of communities, but they cannot be written in terms of ecologically-interpretable expressions and also do not account for heterogeneity in other parameters outside species interactions. Paper 4 aims to further this cause. We aim to move beyond totally random interaction structures by considering correlations that connect to well-known ecological constraints on consumption rates across species.

One objective of Paper 5 is to identify clusters in a large network of seagrass populations that are established through underlying dispersal of vegetative fragments. Clustering in networks has been a very active research area over the past couple of decades, but the spatial and ecological constraints imposed on biological populations imply that the clusters identified might be sub-optimal using generic algorithms. My aim was to use an effective algorithm [29] that identifies different metapopulations such that sites within a metapopulation are more strongly connected than sites outside it. Also, if two clusterings have the same score – in that they are equally optimal – then how to best decide the better one? Combined with genetic analysis, the clustering results answer the question – what is the right spatial scale for management such that dispersal barriers do not hamper the re-establishment of seagrass populations following catastrophic events.

1.2 Thesis outline

This thesis introduces key concepts and existing theories in Chapter 2. The relevant literature and related approaches are also surveyed to place the accompanying research in a broader context. Chapter 3 provides a summary of the appended articles.

Paper 1 uses a modification of generalized Lotka-Volterra equations to study factors that could result in different island SAR forms. Paper 2 studies large

contiguous landscapes as habitat patches connected by high-dispersal. We use a spatially-extended version of the gLV equations where dispersal is modelled through diffusion in discretized space. Paper 3 deals with outcomes of invasion events providing for a range of settings and parameter uncertainties. Paper 4 derives very general mathematical bounds on the stability of ecosystems that are also ecologically interpretable. Paper 5 identifies the most likely scenarios to explain the partition of a colonizing seagrass into distinct metapopulations along the northern and western coasts of Australia. This study uses two complementary approaches, one utilizing ocean current data to estimate transition probabilities between different sites and the other based on genetic data.

Chapter 4 reflects upon the ecological systems and results discussed in this thesis to motivate possible future directions, while also zooming out on the process and progression of my PhD research.

CHAPTER 2

Background

The spatial aspects of single-species metapopulations versus highly-diverse ecosystems differ a great deal conceptually. My contribution in paper 5 is closest in spirit to the problem of finding clusters in a large network. I therefore provide a brief background on what this means for generic large networks and specifically for spatio-ecological systems. Section 2.2 describes complex high-dimensional ecosystems that I study using random matrices where pairwise interactions between species are randomly drawn from some distribution. The dynamics of many interacting species could exhibit patterns typical of large ecosystems for a large part of the parameter space, even when more structure in the species interactions is incorporated. Such highly disordered ecosystems (and systems in general) admit description in terms of a few statistics, which provides some baseline expectations about properties and processes that might be otherwise intractable for such complex systems.

2.1 Community detection in networks

Spatially distributed populations of a species can be partitioned into distinct metapopulations. Such metapopulations would be expected to have two major

but related properties. The dispersal across metapopulations is much lower than within them. In an evolutionary sense, there is negligible gene flow between different metapopulations. This partitioning problem is analogous to finding clusters within a large network where individual populations are nodes, and the transition probabilities characterize the edges between them. One class of methods for clustering in networks is based on spectral analysis that was originally used for graph partitioning [30].

The central object of study in such methods is usually the graph Laplacian whose second eigenvector – corresponding to the second smallest eigenvalue called Fiedler value shows distinct clusters of nodes such that there are much fewer edges across these clusters [30]. The graph Laplacian is also related to the transition matrix of a random walker on a graph. Therefore, the spectral methods can be extended to identify 'lumpings' of states of a Markov process. The problem of graph partitioning can be posed as a problem of minimizing the cut size – defined as the number of edges between two distinct groups of nodes – that could be written in terms of the graph Laplacian. Using a modularity matrix that is analogous to the graph Laplacian, one can similarly maximize the modularity of a given network to find constituent modules [31].

A broad class of methods rely on minimizing (or maximizing) certain measures that quantify the degree of good or bad division of a network into communities. Network modularity [32] is a measure that takes high values if there are many edges within the communities than across them. Such maximization methods are much faster than most other methods in dividing very large networks, especially when the networks are sparse. A highly efficient algorithm proposed in [33] utilized sophisticated data structures that considerably speed up the search and insert operations in such maximization problems.

A majority of spectral as well as hierarchical partitioning methods cannot be extended directly to ecological networks. This is especially true when the underlying networks are directed and asymmetric, or if the transition matrix is not symmetric with respect to some scalar product [34]. For example, the dispersal of a species between geographical sites can be encoded in a transition matrix that contains probabilities of transition from one site to another. An efficient hierarchical partitioning algorithm for this problem was introduced in [29] to minimize a function that characterizes the connectivity between different subpopulations. Ecological constraints on dispersal could be introduced in a tuneable penalty term that allowed for different comparable

partitions. The ecological system studied in paper 5 uses this algorithm, and a heuristic is suggested to compare partitions that are equally optimal (defined by a score).

Algorithm for finding subpopulations using connectivity matrices

Paper 5 uses the algorithm proposed in [29] to aggregate geographical sites into distinct subpopulations. The analysis is based on fragment dispersal matrices – termed as connectivity matrices C here – whose elements (i.e. C_{ij}) define the probability of individuals dispersing from site i ending up at site j. A minimization problem is defined using a preliminary function that quantifies the connectivity between two subpopulations:

$$G(\beta) = 2\sum_{ij}(C_{ij} - \beta^{-1}E_{ij})\Omega_{ij} \tag{2.1}$$

where Ω is matrix with $\Omega_{ij} = 1$ if site i and j are in different subpopulations else $\Omega_{ij} = 0$. $C_{ij}\Omega_{ij}$ is the naïve connectivity that equals zero if all sites are placed in a single subpopulation. Sites across subpopulations with transition probability below β^{-1} is rewarded with a term $\beta^{-1}E_{ij}$ where the matrix E has all elements equal to 1. This term therefore prevent a trivial solution with only one subpopulation.

It is often useful to redefine Ω in terms of a vector s that takes the values ± 1. Different sites can be placed into two subpopulations depending on the sign of the respective element [31]. The relationship $\Omega_{ij} = (1 - s_i s_j)/2$ then allows the redefinition:

$$G(\beta) = -s^T C s + \beta^{-1}(\sum_i s_i)^2 \tag{2.2}$$

The equation 2.2 is written in a matrix form where terms independent of s have been removed since they do not affect the minimization. s^T is the transpose of the s matrix.

Monte-Carlo or simulated annealing techniques are computationally very expensive in trying to find the minimum for this problem. An efficient alternative is to allow s_i to take continuous values such that the final solution can be read off from the signs of s_i. To ensure that the minimization problem is

bounded, the following function is more suitable:

$$G' = G + \frac{1}{\gamma + 1} \sum_i |s_i|^{\gamma+1} \tag{2.3}$$

where $\gamma \geq 1$ prevents the function from diverging. The minimum can be found by using an iterative scheme that solves the derivative of 2.3 starting from a random guess for s. The recursion reads as:

$$s_{t+1,i} = a s_{t,i} + (1-a)\text{sign}(s_{t,i}) \times |\sum_j (C_{ij} + C_{ji}) s_{t,j} - 2\beta^{-1} \sum_j s_{t,j}|^{1/\gamma} \tag{2.4}$$

where the derivative was $-\sum_j (C_{ij} + C_{ji}) s_j + 2\beta^{-1} \sum_j s_j + |s_i|^\gamma \text{sign}(s_i)$. This also explains the choice of having $1/(\gamma+1)$ in the minimization function. Small but positive values of a guide the solution towards the minimum.

The identification of subpopulations can then be summarized as follows. Starting with some low value of $1/\beta$, a single subpopulation is split into two using the recipe described above. Then each subpopulation is recursively split as long as the global value of G decreases. Random combinations of the subpopulations are merged, and a possible merge is only accepted if G goes down. This gives the best possible partition for a given β. Sweeping over different β values gives many possible partitions that can be compared using a leakage measure:

$$Q = \langle \sum_{l \neq k} \tilde{C}_{kl} / \tilde{C}_{kk} \rangle_k \tag{2.5}$$

where $\tilde{C}_{kl} = \langle \sum_{i \in L_k} C_{ij} \rangle_{j \in L_l}$ is the mean connectivity from subpopulation l to k defined using set of sites L_k and L_l that lie in these subpopulations.

2.2 Stability and complexity of large ecological systems

Local asymptotic stability and the complexity-stability debate

Large ecosystems can be studied by stripping away the complexity of species interactions or the environment. At one end lies the niche paradigm that describes community properties by characterizing differences between species

in terms of a few traits. In contrast to this selection dominated approach, the neutral view disregards inter-species differences such that species are identical on a per capita basis in their birth, death, dispersal or speciation rates [35]. A radically different approach is to retain the complexity of species interactions but assume that any structure in this network of interactions is washed away. This is one of the simplest ways to study ecosystems where the complexity of species interactions cannot be ignored.

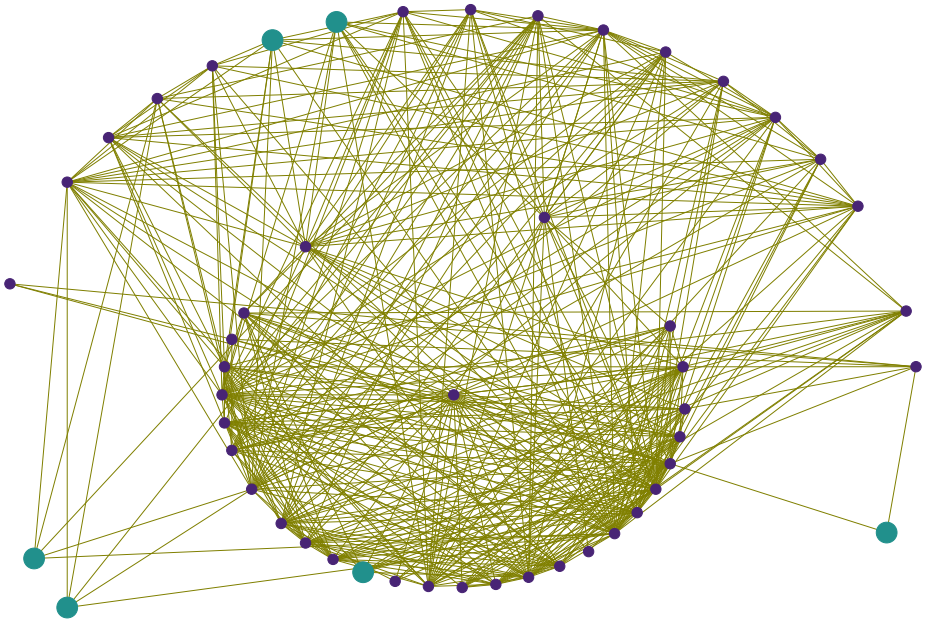

Figure 2.1: A food web with 50 species generated using the niche model in [36]. The larger nodes represent predators in a higher trophic level. The other species constitute an ecological community and lie in the same trophic level. In this sense, the community is embedded within a larger ecosystem represented by the entire food web. The given layout represents the most connected species closer to the centre. Not all possible interaction links are realized. The fraction of realized links is called the connectance.

An early example of describing such complex but random ecosystems is May's seminal work [28]. He used random matrix theory to analyse the stability of random ecosystems whose Jacobian can be written as:

$$\mathbf{J} = \mathbf{A} - \mathbf{I} \tag{2.6}$$

where $\mathbf{I}$ is an identity matrix with diagonal elements equal to 1. $\mathbf{A}$ is the interaction matrix, i.e., a random matrix where non-zero pairwise interaction strengths are drawn from a distribution symmetric around 0. Any interaction pair could have a non-zero interaction strength with probability c (i.e. the connectance). Equivalently, c is the fraction of links between species realized out of all the possible links. The eigenvalue spectrum of $\mathbf{J}$ is central to studying local asymptotic stability. This form of stability relates to the behaviour of the system in response to small perturbations around an equilibrium point (or fixed point). If the largest eigenvalue of $\mathbf{J}$ is negative, then the system is locally asymptotically stable and converges to the equilibrium point if slightly perturbed from it.

The matrix $\mathbf{A}$ can be parameterized in terms of the average interaction strength σ that is also proportional to the standard deviation of the underlying distribution. This parameter can be viewed as a scaling factor that scales all the interaction strengths by a constant factor. Equation 2.6 can be written more explicitly as:

$$\mathbf{J} = \sigma\mathbf{A} - \mathbf{I} \tag{2.7}$$

The above form helps tune the mean and standard deviation of $\mathbf{A}$ using just the σ parameter. May used this matrix to show that as the complexity of an ecosystem increases, it is less likely to be stable. This result is known as May's paradox, and it was at odds with ecological intuition until that point since ecologists expected more complex and diverse ecosystems to be more stable. It sparked-off the famous *complexity-stability debate* by suggesting that in order to guarantee stability one must cut-down complexity which is defined in terms of the number of species, σ and the connectance c of the interaction matrix.

This major result for random interaction matrices provided intuition for possible aspects of stability in real ecosystems. A simple implication is that communities with a large number of species are dominated by weak interactions between species, and strong interactions are more likely in communities with fewer connections between species. Also, for some given average interaction strength and connectance, interaction matrices with a modular structure would be more stable. This paper also raises the question: What systems have

Jacobians of the form used by May, which are independent of the equilibrium abundances?

The population dynamics of many interacting species can be described by a system of ordinary differential equations (ODEs). One of the simplest such equations is the Generalized Lotka-Volterra (GLV) equations.

$$\frac{dy_i}{dt} = r_i y_i + y_i \sum_j A_{ij} y_j \tag{2.8}$$

where r_i is the intrinsic growth rate of i^{th} species. A is the interaction matrix whose elements are pairwise interaction strengths between species. Other general models of population dynamics differ in the functional form used for the growth and interaction terms. Equation 2.8 could be written in another form by explicitly including the carrying capacity K_i in a self-interaction term.

$$\frac{dy_i}{dt} = r_i y_i (1 - \frac{y_i}{K_i}) + \sigma y_i \sum_{j \neq i} A_{ij} y_j \tag{2.9}$$

Note that equation 2.9 could have up to 2^N equilibria for a system of N species since this system admits solutions with $x_i^* = 0$. The corresponding Jacobian depends on the equilibrium abundances x_i^*. The Jacobian does have a more elaborate structure than the one studied by May, but one can draw some interesting comparisons by analysing the case without any $x_i^* = 0$. For this case, if we set all r_i and K_i equal to 1, then the Jacobian can be written as:

$$\mathbf{J} = \mathbf{X}^*(\sigma \mathbf{A} - \mathbf{I}) \tag{2.10}$$

where $\mathbf{X}^*$ is a diagonal matrix with equilibrium abundances on the diagonal. This does resemble the Jacobian that May analysed but that case did not have the $\mathbf{X}^*$ part. If the interaction strengths are weak such that all species end up with equilibrium abundances close to 1 (i.e. the carrying capacities), then the gLV equations have similar properties to the kind of systems May analysed. This also explains the fact that May's framework implicitly excludes stable equilibria with $x_i^* = 0$. There is a rather sharp transition towards instability.

Equilibrium solutions with all positive abundances are called feasible solutions. The parameter regime corresponding to feasible solutions is also struc-

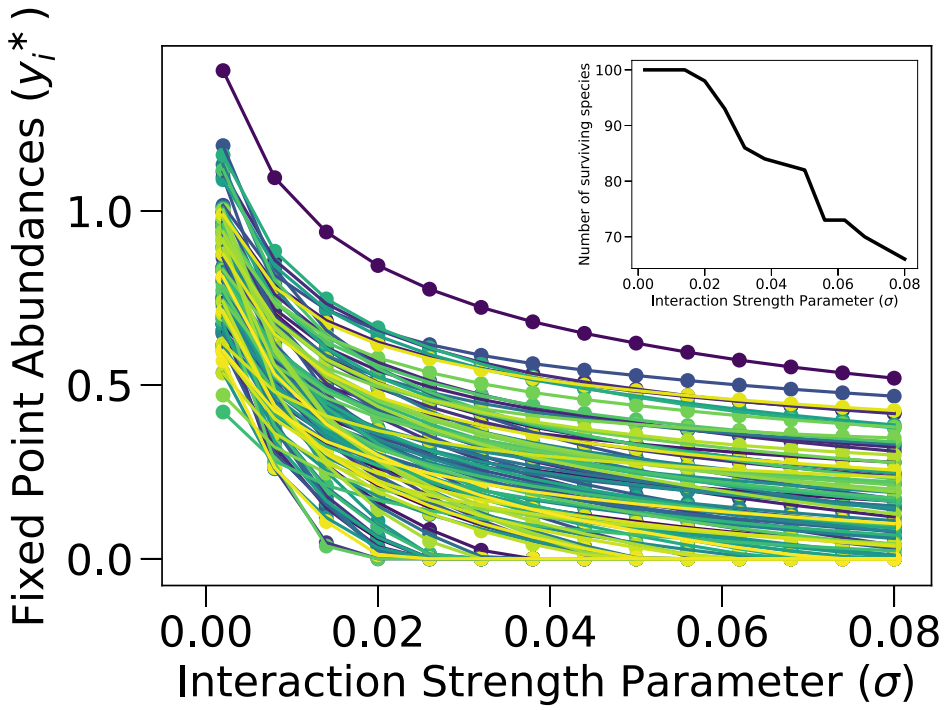

Figure 2.2: Equilibrium abundances of a competitive community of 100 species for increasing values of the interaction strength parameter (σ) (See equation 2.9). Each set of vertical dots represents an assembled community corresponding to a given value of σ. The bold black line in the inset traces the corresponding number of surviving species. Note that higher values of σ correspond to more extinctions. The interaction strengths, growth rates and carrying capacities are chosen from normal distributions with means -1, +1 and +1 respectively. The standard deviation is set to 0.2 for each of these.

turally stable, in that a small change in parameters does not change the qualitative dynamical behaviour of the system [37]. The gLV equations allow for a locally stable but structurally unstable phase characterized by single species extinctions. This phase facilitates the study of community assembly where not all species from the larger pool survive (Figure 2.2). This is an important aspect of papers 1 and 2.

Mathematical bounds on asymptotic stability of ecological systems

The key result in May's paper drew upon Wigner's semi-circle law [38] and earlier results in random-matrix theory [39, 40]. Here we go over some useful results from random-matrix theory that describe the distribution of eigenvalues for different classes of random matrices and have been instrumental in the study of stability in large ecological systems.

To state Wigner's semi-circle law, we introduce a few preliminaries.

- **Empirical spectral density(ESD)**: Given an $N \times N$ matrix A with eigenvalues $\lambda_1(A)$, $\lambda_2(A)$ ... $\lambda_N(A)$, the empirical spectral density is defined as:

$$n(\lambda; A) = \frac{1}{N} \sum_{i=1}^{N} \delta(\lambda - \lambda_i(A)) \tag{2.11}$$

  where δ is a Dirac delta. ESD is a probability measure defined in the complex plane $\mathcal{C}$. It can be interpreted as a counting function which when integrated over a subset of $\mathcal{C}$ gives the fraction of eigenvalues that lie in that subset. For simplicity we consider the case of Hermitian matrices which ensure real eigenvalues. In random matrix theory, one usually draws A from a random matrix ensemble. Then $n(\lambda)$ is a random measure whose average is more instructive.

- **Average ESD**: This quantity is the ESD averaged over the joint probability density function of the eigenvalues:

$$\rho(\lambda) = \frac{1}{N} \int P(\lambda_1, \lambda_2 ...., \lambda_N) \sum_{i=1}^{N} \delta(\lambda - \lambda_i) d\lambda_1 d\lambda_2 ... d\lambda_N \tag{2.12}$$

$\rho(\lambda)$ can be computed for some classes of random matrix ensembles in the limit $N \to \infty$. Wigner found an elegant semi-circular (or rather semi-elliptical) average ESD that holds for many random matrix ensembles. This is the famous Wigner's semi-circle law [38].

- **Wigner's semi-circle law**: Let A_N be an $N \times N$ Hermitian random matrix whose elements are independent random variables drawn from

symmetric distributions with mean zero, unit variance and bounded moments. Then in the limit $N \to \infty$, the average ESD of $A_N/\sqrt{N}$ is:

$$\rho_{SC}(\lambda) = \frac{1}{2\pi}\sqrt{4-\lambda^2}\mathbb{1}_{[-2,2]} \tag{2.13}$$

where $\mathbb{1}_{[-2,2]}$ is the Indicator function that maps elements in the range $[-2, 2]$ to 1 and rest to 0.

There was related work on non-Hermitian matrices after Wigner, which made a strong case for the *circular law* [39, 40]. The most general version of this law was finally proved by Tao et al. [41, 42].

- **Circular Law**: Let A_N be an $N \times N$ random matrix whose elements are independent random variables mean zero and unit variance. Then in the limit $N \to \infty$, the average ESD of the matrix $A_N/\sqrt{N}$ converges to the uniform distribution on the unit disk.

May's result can be recovered by using the circular law. For a given matrix A with standard deviation σ not equal to 1 and connectance C, we can divide the elements by $\sigma\sqrt{C}$ to get $\sigma\sqrt{C}$ times a matrix with unit variance. For large N, the circular law implies that the largest real part of any eigenvalue is $\sigma\sqrt{NC}$. A unit diagonal matrix would just shift the eigenvalues by -1, hence the stability condition becomes $\sigma\sqrt{NC} - 1 < 0$, identical to May's result.

Another interesting law was found in the 80's which considers correlation ϵ between elements across the diagonal A_{ij} and A_{ji}, i.e., $\epsilon = \mathcal{E}[A_{ij}A_{ji}]$. This is called Girko's Elliptic Law [43, 44].

- **Girko's Elliptic Law**: Let A_N be an $N \times N$ random matrix whose off-diagonal elements are random variables drawn in pairs with correlation $\epsilon = \mathcal{E}[A_{ij}A_{ji}]$ and marginal mean and marginal variance being zero and one respectively. Then in the limit $N \to \infty$, the average ESD of the matrix $A_N/\sqrt{N}$ converges to the uniform distribution on the unit ellipse with horizontal and vertical semi-axes of length $1 + \epsilon$ and $1 - \epsilon$ respectively.

By using analogous arguments as for May's result, we can show that the stability condition is $\sigma\sqrt{NC}(1 + \epsilon) - 1 < 0$ since $1 + \epsilon$ is the length along the real axis. In fact, for systems with such a correlation the off-diagonal element would be likely to have a non-zero mean μ. The criterion then becomes

$\sigma\sqrt{NC}(1+\epsilon) - \mu - 1 < 0$ This result is very useful for ecological communities since the underlying matrix can represent predator-prey, mutualistic or competitive interactions depending on the value of the correlation ϵ [45, 46].

Diagonal stability

The gLV equations belong to a class of models that admit diagonal stability under some very simple conditions. To define diagonal stability, we consider the system:

$$\frac{dx(t)}{dt} = \mathbf{A}x(t) \tag{2.14}$$

Since $\mathbf{A}$ is the Jacobian for this system, it follows that asymptotic stability is guaranteed if all the eigenvalues have negative real parts. Equivalently if one can find a positive definite matrix $\mathbf{P}$ then asymptotic stability follows if the following matrix is negative definite [47]:

$$\mathbf{A}^T\mathbf{P} + \mathbf{P}\mathbf{A} \tag{2.15}$$

Additionally if P is also a diagonal matrix, then the system is said to be *diagonally stable* (or more precisely Hurwitz diagonally stable) [48]. Asymptotic stability is easy to demonstrate for the case when $\mathbf{P}$ is an identity matrix. For a matrix to be negative definite, it's sufficient if only it's symmetric part, i.e. $\mathbf{A} + \mathbf{A}^T$ is negative definite. This implies that $\mathbf{A}$ would be guaranteed to be negative definite. Diagonal stability is of interest since it helps find a Lyapunov function for systems such as the gLV equations.

Such dynamical systems are called Persidskii-type systems [49]. These systems can be expressed in a vector form as:

$$\dot{x} = \mathbf{A}f(x) \tag{2.16}$$

where x is n-dimensional. Each component f_i of the function $f(x)$ depends only on the respective element of the vector x. The components $f_i(x_i)$ belong to a class of functions known as *infinite sector non-linearities* that satisfy the following conditions:

$$f_i(0) = 0 \tag{2.17}$$

$$f_i(x_i) * x_i > 0 \tag{2.18}$$

$$\int_0^{x_i} f_i(\omega)d\omega \to \infty \qquad \text{for} \qquad |x_i| \to \infty \tag{2.19}$$

Diagonal stability facilitates some very useful results on stability for Persidskii-type systems (2.16). If the matrix A in equation 2.16 is diagonally stable with respect to a diagonal positive definite matrix P, then the zero solution of this system is globally stable and admits the following diagonal-type Lyapunov function [49]:

$$V(x) = \sum_{i=1}^{n} P_i \int_0^{x_i} f_i(\omega)d\omega \tag{2.20}$$

As expected of a Lyapunov function, the time derivative of this function is negative.

$$\dot{V} = \sum_{i=1}^{n} P_i f_i(x_i)\dot{x_i} \tag{2.21}$$

$$\implies \dot{V} = \sum_{i=1}^{n} P_i f_i(x_i) \sum_{j=1}^{n} A_{ij} f_j(x_j) \tag{2.22}$$

$$\implies \dot{V} = f^T(x)(PA)f(x) \tag{2.23}$$

$f^T(x)PA$ is the transpose of the vector $A^T Pf(x)$. This follows from the fact that PA and $A^T P$ are obtained by multiplying the i^{th} diagonal element of P with the corresponding rows and columns of A and A^T respectively. This holds because P is a diagonal matrix. When the vector $A^T Pf(x)$ is multiplied by $f(x)$, it gives the same scalar as $f^T(x)(PA)f(x)$. Therefore,

$$\implies \dot{V} = \frac{1}{2} f^T(x)(PA + A^T P)f(x) \tag{2.24}$$

which is negative since $PA + A^T P$ is negative definite. Among other properties of a Lyapunov function, V is also radially bounded that follows from the property 2.19 of $f(x)$.

A simple change of variables can be used to express the gLV equations in a

Persidskii-type form. Rewriting equation 2.8 as:

$$\frac{\dot{y_i}}{y_i} = r_i + \sum_{j=1}^{n} A_{ij}\exp(\log y_j) \tag{2.25}$$

$$\implies \frac{d\log y_i}{dt} = r_i + \sum_{j=1}^{n} A_{ij}\exp(\log y_j) \tag{2.26}$$

One can then solve the system of equations $r_i + \sum_{j=1}^{n} A_{ij}\exp(\log y_j^*) = 0$ to find the feasible equilibrium $\log y^*$, assuming that this equilibrium exists. Equation 2.26 can be rewritten as:

$$\frac{d(\log y_i - \log y_i^*)}{dt} = \sum_{j=1}^{n} A_{ij}(\exp(\log y_j) - \exp(\log y_j^*)) \tag{2.27}$$

$$\implies \frac{d(\log y_i - \log y_i^*)}{dt} = \sum_{j=1}^{n} A_{ij}(\exp(\log y_j - \log y_j^* + \log y_j^*) - \exp(\log y_j^*)) \tag{2.28}$$

Defining $z_i = \log y_i - \log y_i^*$ and $g_i(z_i) = g_i(\log y_i - \log y_i^*) = \exp(\log y_i - \log y_i^* + \log y_i^*) - \exp(\log y_i^*)$, we obtain the equations in a Perdiskii-type form:

$$\frac{dz_i}{dt} = \sum_{j=1}^{n} A_{ij} g_j(z_j)) \tag{2.29}$$

The function g_i also satisfies the conditions 2.17, 2.18 and 2.19. This system admits the Lyapunov function:

$$V(z) = \sum_{i=1}^{n} P_i \int_0^{z_i} g_i(\omega)d\omega \tag{2.30}$$

whose time-derivative can be similarly computed as before.

$$\implies \dot{V(z)} = \frac{1}{2} g^T(z)(PA + A^T P)g(z) \tag{2.31}$$

The most widely used form of the Lyapunov function for the gLV system [50] can be obtained from equation 2.30 by expressing g_i and z_i in terms of y_i:

$$V(y) = \sum_i p_i(y_i - y_i^* - y_i^* \log y_i/y_i^*) \tag{2.32}$$

Structural stability

Structural stability captures the ability of a system to exhibit the same qualitative behaviour when subject to small changes in the underlying parameters or the model itself.

The concept is well-defined for dynamical systems with respect to *hyperbolic* fixed points. The Jacobian matrix computed at these fixed points does not have any purely imaginary eigenvalues. Small changes to the underlying model or its parameters are captured through the phase portrait of the system. Then the phase portrait of the dynamical system is called structurally stable if any mapping results in a topologically equivalent phase portrait, i.e., a mapping that preserves the direction of time and is homeomorphic. A homeomorphism is a continuous one-to-one mapping between two topological spaces and has a continuous inverse.

From the Andronov-Pontryagin criterion [51], it is known that structurally stable systems in two-dimensions have a finite number of fixed points, all of which are hyperbolic. Further, preserving the dynamical behaviour requires that no homoclinic or heteroclinic orbits exist, i.e., saddle points don't have self loops and are not connected to other saddle points.

The condition about hyperbolic fixed points implies that the phase portrait around centres (neutrally stable fixed points) is not structurally stable. If one considers a simple harmonic oscillator that admits such a center then its easy to argue that the dynamical behaviour changes if the underlying model is modified by adding the smallest of damping terms. The system finally relaxes to an asymptotically stable fixed point, implying structural instability of the original system. A similar structural instability occurs for heteroclinic orbits that are usually observed in conservative and reversible systems eg. a pendulum system that exhibits a full range of motion in a vertical plane. Infinitesimal amount of damping is sufficient to eliminate the heteroclinic orbits.

Structural stability in ecological systems

In ecological systems, coexistence of species requires that the underlying dynamical system has a feasible and dynamically stable equilibrium. The structural perspective argues that it is insufficient to quantify feasibility and stability of equilibria for arbitrary values of the intrinsic growth rates, and one should rather study the range of these growth rates that allow for coexistence. Rohr et al. [37] quantified the structural stability of ecological systems in terms of the proximity to structural vector(s) that lies in the middle of the feasibility domain – the domain of intrinsic growth rates that permit feasible and stable equilibria. Earlier studies have used a preliminary structural approach to study coexistence of competing species through conditions on rescaled versions of the intrinsic growth rates [52, 53].

A related metric of structural stability can be obtained for a given interaction matrix by defining a feasibility domain formed by intrinsic growth rates that are compatible with feasibility conditions [54, 55]. This domain is a convex polyhedral cone in R^n – n being the number of species – formed by growth rate vectors that satisfy the following condition with respect to the gLV equations 2.8:

$$r_i = -\sum_{j=1}^{n} A_{ij} y_j^* \tag{2.33}$$

where y^* is a feasible solution. Any point within the feasible domain can be expressed as a positive linear combination of *generators* g_i of the cone [54]:

$$r = \sum_{i=1}^{n} g_i \lambda_i \tag{2.34}$$

where $\lambda_i > 0$. The columns of the interaction matrix $\mathbf{A}$ also constitute a set of generators. In fact, rescaling one or more of the generators also results in a generator of the convex cone. One can then use the columns v_i of A and the positive abundances y^*, to define the feasibility domain [55]:

$$D_F(\mathbf{A}) = \{r = \sum_{i=1}^{n} v_i y_i^* \in R^n | y_i^* > 0\} \tag{2.35}$$

The structural stability of a system with a fixed matrix $\mathbf{A}$ can be expressed

as the fraction of the volume that the convex cone subtends on the surface of a unit sphere in R^n [54]. A simple expression for this metric was obtained for the case when the number of species is large and the interaction matrix is a negative definite random matrix. In this limit, the metric is independent of the finer details of the communities and only depends on the size of the community and the first few moments of the distribution of interaction strengths.

2.3 Species-area relationships

It is debatable whether any universal quantitative laws exist in ecology. The species-area relationship (SAR) comes close to this distinction, but more than one functional forms of this relationship have been reported in empirical studies [4]. The two simplest forms of the SAR are the power law $S \sim A^z$ [56] and the semi-log $S \sim z\ log\ A$ [57] – both forms suggest that species richness S increases monotonically with patch area A.

The power-law form has received much attention in terms of studies investigating its incidence. Its origin has been explained using empirical patterns such as lognormal species-abundance distributions [5] and clustering of conspecific (i.e. of same species) individuals [58]. The SAR could be studied for island-like systems or for nested sampling areas [5]. Island SARs differ in terms of processes such as immigration from a mainland pool, which is a key aspect of the analysis in Paper 2 (Section 3.2).

2.4 High-dimensional ecosystems in space

The past few decades have witnessed many approaches to study ecological communities distributed over spatially connected regions. A prominent approach is the metacommunity framework that considers local communities connected by dispersal of species. The spatial patches could be heterogeneous in the environmental conditions that further affect species traits [13, 14]. Paper 3 does focus on metacommunities but the term has a slightly different meaning operationally as in [1]. The focus is again on highly diverse complex communities but now these are spatially distributed over different patches that allow migration between one another.

The problem of studying complex ecosystems connected by migration draws from the tradition of embracing complexity over structure, which allows for

understanding system-wide properties using a few statistics. Studies in this direction have uncovered novel results ranging from the stability-complexity paradox to newer types of dynamics that facilitate higher diversity of species [1, 2, 59]. These studies use similar approaches but we present the general paradigm using the GLV type equations discussed in [1]. Consider M patches connected by dispersal, each of which represents a well-mixed ecosystem with unstructured inter-species interactions:

$$\frac{d}{dt}\phi_{i,u} = \phi_{i,u}\left[K_{i,u} - \phi_{i,u} - \sum_{j\neq i} A_{ij,u}\phi_{j,u}\right] + \sum_{v} D_{i,uv}(\phi_{i,v} - \phi_{i,u}) \quad (2.36)$$

The dynamics of abundance $\phi_{i,u}$ of a species i in patch u depends in a familiar way on growth and interaction terms. For simplicity, the carrying capacities $K_{i,u}$ can be set to 1. $A_{ij,u}$ denotes the interaction matrix at a patch u, which could differ from other patches. The dispersal rate between patches u and v is $D_{i,uv}$.

It was shown in [59] that increasing the number of ecologically distinct patches – characterized by variation in species interactions – increases the stability of the meta-ecosystem. In [1], it was found that such spatially-extended systems can enter a chaotic phase for intermediate dispersal rates. A key feature of this phase are abundances fluctuations whose higher strength corresponds to higher fraction of species surviving. A cut-off is set below which species are considered extinct, but the results are unchanged qualitatively for different possible values. The authors of [2] found a similar chaotic regime for interaction matrices with an anti-symmetric structure.

Paper 3 considers only stable equilibria but the ideas of metacommunity and surviving fraction of species is used in the same vein as [1] (Section 3.3).

2.5 Evolutionary game theory

Evolutionary game theory was conceived to provide a Darwinian explanation for processes like animal conflict by analyzing the dynamics of interacting strategies. Unlike classical game theory where players can choose between strategies, evolutionary game theory deals with players having fixed strategies that are passed on to the next generation. This effectively amounts to

strategies playing against one another. This framework also differs in terms of its dynamical outlook on games. We first describe the Hawk-Dove game to demonstrate the class of questions that are of interest while also discussing basic concepts in game theory and their extensions.

The Hawk-Dove game models contests which involve two possible strategies in a conflict situation. This two-player game encodes the payoff received by each player in terms of the value V of the resource being contested and the cost of escalation C. This is a symmetric game in that both players have the same payoff matrix. The players could either pick a *pure strategy* which is just a single strategy or employ a *mixed strategy* that is a distribution over the different strategies. The payoff matrix for the game is:

	Hawk	Dove
Hawk	$\frac{1}{2}(V-C)$	V
Dove	0	$\frac{V}{2}$

Table 2.1: Payoff matrix for the hawk-dove game. Each entry in row i and column j represents the payoff that strategy i receives when playing against strategy j.

Nash equilibrium

One of the central concepts in game theory is that of a Nash equilibrium. For a two player game, it is defined as the set of strategies for which either player cannot improve their payoff by changing their strategy alone. In the most general games with two players, Nash equilibrium is defined in terms of a pair of mixed strategies q_1 and q_2 for player 1 and 2 respectively, such that the following conditions hold:

$$q_1.Aq_2 \geq q_1'.Aq_2 \qquad \forall q_1' \text{ of player 1} \tag{2.37}$$

$$q_2.Aq_1 \geq q_2'.Aq_1 \qquad \forall q_2' \text{ of player 2} \tag{2.38}$$

Note that q_1, q_2 etc. are vectors containing frequencies of different strategies.

For symmetric games with a payoff matrix A, the Nash equilibrium is a mixed/pure strategy q such that:

$$q.Aq \geq p.Aq \tag{2.39}$$

which means that the payoff for q playing against itself is always higher than the payoff of another strategy p playing against q.

For the hawk-dove game with $V < C$, three Nash equilibria exist. Two of these involve pure strategies for the two players, i.e., (Hawk, Dove) and (Dove, Hawk). While none of the players can improve their payoff by changing their strategy alone, it is more interesting to consider many such pairwise contests over different rounds. Let's consider a population where one player plays hawk over many different rounds and the other player always plays dove. Evolutionary contexts also allow for chance mutations where a small fraction of the 'hawk' population could suddenly play 'dove' and vice versa. This could be thought of as an invading set of strategies that makes player 1 and 2 deviate from their usual hawk and dove strategies in some of the pairwise contests. If the two players are now made to play over multiple generations and each of the players eventually revert back to playing pure strategies, then the pure set of strategies would be stable in some sense. It's evident from the example of invading strategies that the (Hawk, Dove) Nash equilibrium does not admit such stability. Player 1 would now have some reduction in the number of hawk offspring resulting from contests with player 2's small hawk population. At the same time, player 1's initial dove population would now grow over generations owing to the dove-dove contests with player 2. The multi-generational dynamics would preclude the system going back to the pure (Hawk, Dove) strategy. We can already guess that any stable set of strategies would be a mixed strategy.

We formalize the notion of stability just described using replicator equations which is the canonical model of evolutionary game theory.

Replicator Equations

The replicator equations are used to model the dynamics of relative frequencies p_i of strategies. The continuous time version is written as:

$$\frac{dp_i}{dt} = p_i * (R_i(\mathbf{p}) - \sum_j p_j * R_j(\mathbf{p})) \tag{2.40}$$

where R_i is the fitness of each species. The second term in the bracket is the mean fitness over all species. The simplest form of the fitness is $R_i = \sum_j A_{ij} p_j$ where A_{ij} is the pairwise interaction strength between strategies i and j. It is useful to express the fitness terms as a vector $\mathbf{R}$ in terms of the interaction matrix A and the vector $\mathbf{p}$ of the frequencies: $\mathbf{R} = A\mathbf{p}$.

Evolutionary Stable Strategy

Consider a resident population at equilibrium having a mixed strategy distribution $\mathbf{q}$. If the population always reverts back to the distribution $\mathbf{q}$ whenever a small fraction ϵ of the residents is replaced by an invading population with a different distribution $\mathbf{p}$, then $\mathbf{q}$ constitutes an ESS. This requires the residents to have greater fitness than all $\mathbf{p}$, i.e.:

$$\mathbf{q} \cdot A((1-\epsilon)\mathbf{q} + \epsilon\mathbf{p}) > \mathbf{p} \cdot A((1-\epsilon)\mathbf{q} + \epsilon\mathbf{p}) \tag{2.41}$$

The following two conditions could therefore define an ESS:

$$\mathbf{q} \cdot A\mathbf{q} \geq \mathbf{p} \cdot A\mathbf{q} \tag{2.42}$$

$$\mathbf{q} \cdot A\mathbf{q} = \mathbf{p} \cdot A\mathbf{q} \implies \mathbf{q} \cdot A\mathbf{p} > \mathbf{p} \cdot A\mathbf{p} \tag{2.43}$$

where the first condition 2.42 is the same as that of a symmetric Nash equilibrium. For the hawk-dove game, the formal definition implies that a pure hawk or a pure dove strategy can't be evolutionarily stable if $V < C$. For a pure hawk strategy $\mathbf{q} = [1, 0]$, $\mathbf{q} \cdot A\mathbf{q}$ will be highly negative such that a strategy $\mathbf{p}$ with even a small proportion of doves will result in a higher payoff against $\mathbf{q}$, that is $\mathbf{p} \cdot A\mathbf{q} > \mathbf{q} \cdot A\mathbf{q}$. Similarly, a pure dove strategy $\mathbf{q} = [0, 1]$ would always have a lower payoff than any mixed strategy $\mathbf{p}$ that has even a small fraction of hawks. Any evolutionarily stable strategy for this case would necessarily be a mixed strategy.

Note that the distribution $\mathbf{r} = (1-\epsilon)\mathbf{q} + \epsilon\mathbf{p}$ is sufficiently close to $\mathbf{q}$. Then equation 2.41 follows from the following condition:

$$\mathbf{q} \cdot A\mathbf{r} > \mathbf{r} \cdot A\mathbf{r} \tag{2.44}$$

One can see this by expanding the right hand side of 2.44 such that:

$$\mathbf{q} \cdot A\mathbf{r} > \mathbf{q} \cdot A\mathbf{r} - \epsilon(\mathbf{q} \cdot A\mathbf{r}) + \epsilon(\mathbf{p} \cdot A\mathbf{r}) \tag{2.45}$$

which leads to the inequality 2.41. Therefore, a mixed strategy $\mathbf{q}$ is an ESS if $\mathbf{q} \cdot A\mathbf{r} > \mathbf{r} \cdot A\mathbf{r}$ for all $\mathbf{r}$ sufficiently close to $\mathbf{q}$.

Thomas [60] defined a stronger version of the ESS by replacing 2.43 with:

$$\mathbf{q} \cdot A\mathbf{p} \geq \mathbf{p} \cdot A\mathbf{p} \tag{2.46}$$

2.6 Correspondence between gLV and replicator equations

There exists a direct correspondence between the gLV and replicator equations. The vast array of results from evolutionary game theory specifically replicator equations [61–66] have helped guide many similar advances in ecological contexts using the gLV equations [67, 68]. To motivate the connection, we first consider the following population dynamics equations:

$$\frac{d}{dt}X_i = X_i * R_i(\mathbf{X}) \tag{2.47}$$

where X_i is the abundance of species i and R_i is its effective growth rate which could be a non-linear function of the species abundances.

If we consider frequencies $p_i = \frac{X_i}{\sum_j X_j}$, then the time derivative can be be found by using the chain rule:

$$\frac{d}{dt}p_i = \frac{\dot{X}_i \sum_j X_j - X_i \sum_j \dot{X}_j}{(\sum_j X_j)^2} \tag{2.48}$$

Note that the dot above the variables denotes a time derivative. Substituting

2.47 in 2.48 then yields the replicator equations.

$$\frac{d}{dt}p_i = p_i\Big(R_i(\mathbf{X}) - \sum_j p_j R_j(\mathbf{X})\Big) \tag{2.49}$$

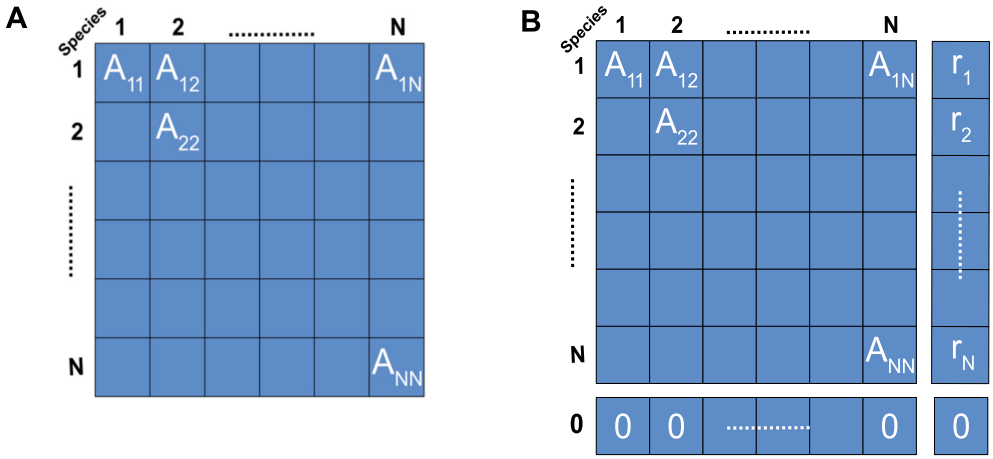

Figure 2.3: A. Interaction matrix for gLV equations. **B.** Corresponding payoff-matrix for replicator equations with one additional species.

We make the connection more precise for the special case when $R_i(\mathbf{X}) = r_i + \sum_j A_{ij}X_j$ where r_i is the intrinsic growth rate and A_{ij} is the interaction between species i and j. This corresponds to the gLV equations 2.8. An n-dimensional gLV system has a duality with $n+1$ dimensional replicator equations. The payoff matrix of the replicator equations is obtained by augmenting the interaction matrix $\mathbf{A}$ by an additional row and column such that the additional column contains the intrinsic growth rates of the species. The additional row is populated with zeros (Figure 2.3).

The dynamics of gLV equations can be obtained from the equivalent replicator system by dividing the abundances of the n species by the abundance of the additional species (Figure 2.4).

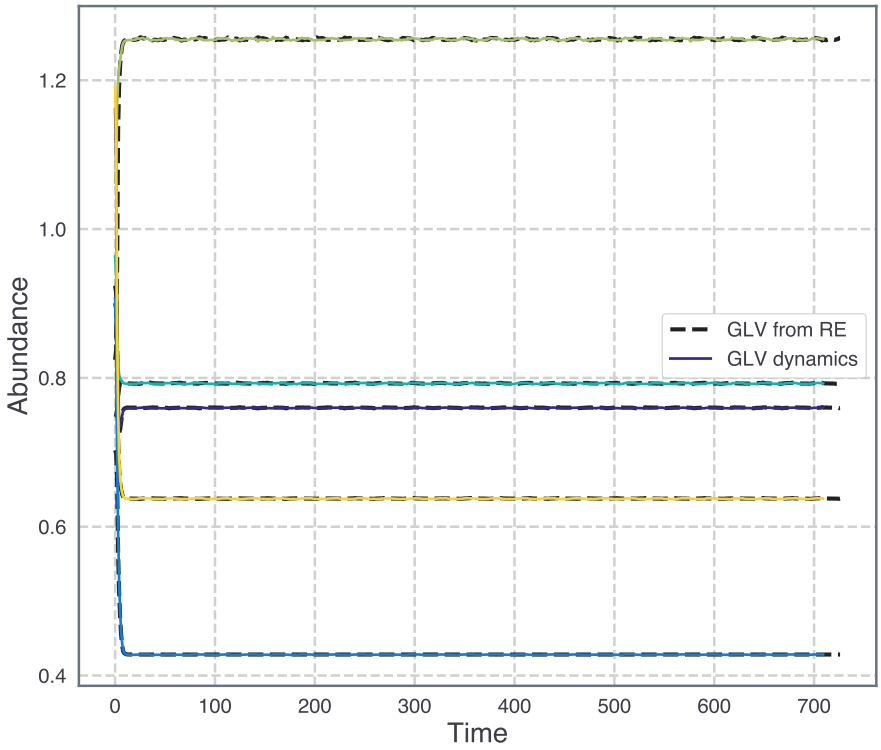

Figure 2.4: The dynamics from gLV equations (solid lines) and equivalent replicator equations (dashed lines)

Family of dualities

The gLV equations (Equation 2.8) are dual to a much larger family of replicator equations with different types of fitness functions. The simplest extension is that with an exponential functional form:

$$\frac{d}{dt}p = p * \left(e^{Ap} - p.e^{Ap}\right) \tag{2.50}$$

The above form works better if the system has many species such that any feasible solution requires the interaction to be weak. The relative frequencies p_i are always smaller than 1, so the exponential function in equation 2.50 can be expanded to obtain the replicator equations with type 1 functional response. This duality works well even with as few as 5 species (Figure 2.5).

Along the same lines, we can define many other replicator equation systems

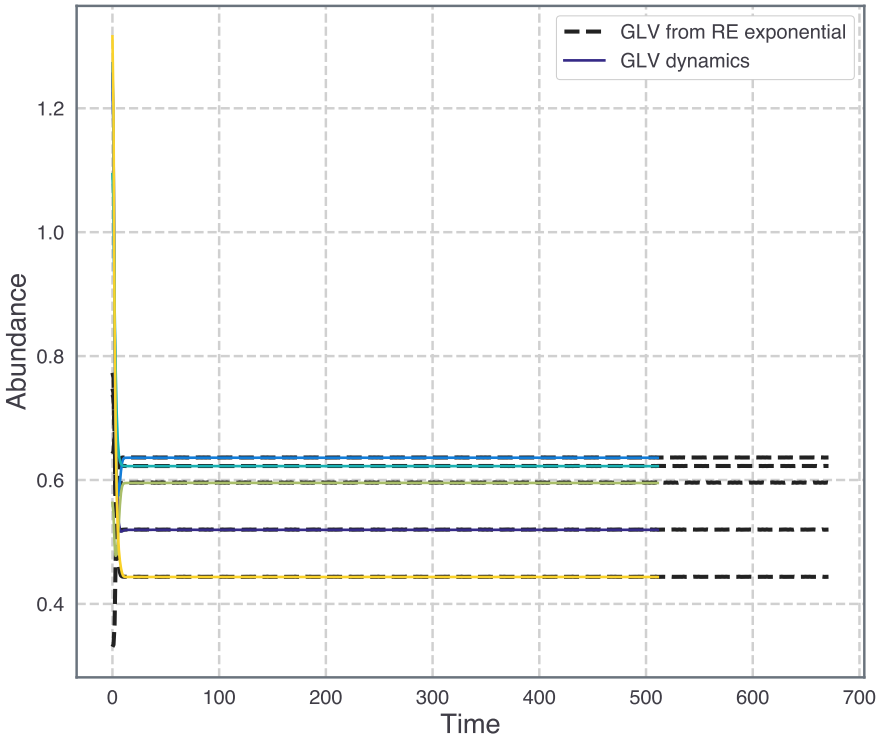

Figure 2.5: The dynamics from gLV equations (solid lines) and equivalent dual equations (dashed lines) 2.50

that are equivalent. These include equations with a logarithmic fitness function

$$\frac{d}{dt}p = p * \Big(\log(1 + Ap) - p.\log(1 + Ap)\Big) \tag{2.51}$$

and those with fitness functions containing a logarithm of relative frequencies

$$\frac{d}{dt}p = p * \Big(A\log(p) - p.A\log(p)\Big) \tag{2.52}$$

Beyond the theoretical connection, the correspondence with equation 2.50 also relates to models used for studying plant community dynamics [69]. From equations 2.47 and 2.49, it follows that equation 2.50 is also dual to:

$$\frac{d}{dt}y = y * \Big(r + e^{Ay}\Big) \tag{2.53}$$

This functional form is used in models of plant communities [69, 70].

Evolutionary stable strategies and uninvadable equilibria

The concept of an ESS has also gained traction within the field of theoretical community ecology since it encapsulates a useful kind of stability. The analogous concept in ecology is called an uninvadable equilibrium. Given a pool of n species which are put together, let's assume that k species go extinct in the equilibrium attained. This equilibrium is called uninvadable if none of the extinct species can invade when introduced again to the equilibrium state. For a general system of equations describing community dynamics, it is not straightforward to find parameter choices that guarantee uninvadable equilibria.

The gLV equations can be constrained in simple ways that ensure the existence of uninvadable equilibria. This includes making the self-interaction terms more negative such that $A + A^T$ is negative definite. Combined with feasibility, this guarantees an uninvadable equilibrium which is also known as a saturated rest point [37, 64]. It differs from an ESS in terms of being globally stable – even if invader's are reintroduced in large numbers, they cannot invade this equilibrium.

2.7 Metabolic Theory of Ecology

Processes in ecosystems are a consequence of organisms' interactions with each other and abiotic factors such as temperature that drive growth, reproduction, phenotypic plasticity etc. Some of these interactions manifest through actions such as dispersal in space which for example, determines encounter rates between consumers and resources. Metabolism is a key determinant of the limits within which consumption/loss of biomass operates, and metabolic rates determine the speed and related aspects of motion so as to minimize loss of energy [71, 72] while also shaping home ranges [73]. The most well-known relationship involving metabolic rates is the Kleiber's law which states that an individual's metabolic rate has a power law dependence on body mass with the scaling exponent 3/4 [20]. Since metabolic rates are also related to temperature via rates of biochemical reactions [21, 74], they provide a crucial link between individual, species, ecosystem level processes and global environmen-

tal change.

Metabolic theory of ecology is a mechanistic and quantitative framework that leverages the temperature and body mass dependence of an individual's metabolism to understand the flow of energy and matter in whole populations and ecosystems [19]. We briefly describe some key results that connect to ecosystem level quantities.

Kleiber's law can be combined with the temperature dependence of metabolic rates to get the following expression for individual metabolic rate B [21]:

$$B = B_0 M^{3/4} \exp^{-E/kT} \tag{2.54}$$

where M is the body mass, E is the activation energy which determines the rate of the biochemical reactions, k is the Boltzmann's constant, T is the temperature and B_0 is a normalization constant that depends on the taxon and the metabolic state. Sometimes it's useful to work with the mass-specific metabolic rate $\bar{B}$ which is $\frac{B}{M}$.

Individual biomass production P scales in the same way as B, i.e., $P \propto M^{3/4} \exp -E/kT$. One can then find the scaling for ecosystem level production per unit standing biomass, which is:

$$\bar{P} = \frac{N.P}{N.M} \propto M^{-1/4} \exp^{-E/kT} \tag{2.55}$$

At the population level, the maximum rate of exponential growth $r_{\max}$ was known to scale as $M^{-1/4}$ like the mass-specific metabolic rate $\bar{B}$ [75–78]. Starting from the general formula 2.54 for B, the dependence of $r_{\max}$ on body size and temperature was also derived in addition to a general expression relating growth rate r to the population level metabolic rate at steady state [79].

Metabolic scaling exponents

Equations like the Kleiber's law are called *Allometric equations* which express biological quantities like metabolic rate, development time etc as a power function of body mass M. The scaling exponents are defining features of these laws, such as the 3/4 metabolic scaling exponent in Kleiber's law.

While the value 3/4 is a good fit for data spanning many orders of magnitude across taxa, it is not the best fit when considering specific taxa. Like for fish

the exponent has been found to be much higher and closer to 1 [80, 81]. In general, the scaling exponents can lie between 2/3 and 1 [82–84].

2.8 Stochastic models for microbial communities

Here using the example of microbial communities, we show how stochasticity can be included in models of ecological communities.

Macroecological laws in microbial communities

Relative abundance data from microbiomes exhibit three universal laws. These were introduced in a paper by Grilli [85], and are sometimes called Grilli's laws. These include:

1. For each species, the abundance fluctuation distribution (AFD), i.e., the fluctuations around the mean abundance are described by a Gamma distribution.

2. **Taylor's law**: The variances of the distributions are proportional to their means across species.

3. The mean abundance distribution (MAD), i.e., the distribution of mean abundances is well described by a lognormal distribution.

It was shown that the dynamics of microbial systems can be described using the stochastic logistic model (SLM) which includes an environmental noise term. The time-scale of the fluctuations is much shorter than that of the population dynamics. The SLM equations for abundances x_i are:

$$\frac{dx_i}{dt} = \frac{x_i}{\tau}(1 - \frac{x_i}{K_i}) + \sqrt{\frac{\sigma_i}{\tau}} x_i \xi_i(t) \tag{2.56}$$

where τ^{-1} is the growth rate, K_i is the carrying capacity. $\xi_i(t)$ is the noise term with correlations $\langle \xi_i(t)\xi_j(t')\rangle = \delta_{ij}\delta(t-t')$ and σ is the coefficient of variation of the fluctuations around mean abundance. The three macroecological laws can be explained using the SLM at stationarity. The SLM cannot reproduce the distribution of pairwise correlations in abundances across species. This is because the abundance dynamics of each species is independent of the others.

Stochastic Lotka-Volterra Model

The Stochastic Lotka-Volterra Model (SLVM) [86] extends the SLM by also including species interaction terms in the equations. For communities with low connectance, this model not only reproduces Grilli's laws but also explains the distribution of pairwise correlations in abundances. For S species, the model is expressed as:

$$\frac{dx_i}{dt} = \frac{x_i}{\tau}(1 + \sum_{j=1}^{S} A_{ij}x_j) + x_i\xi_i \tag{2.57}$$

where A_{ij} are the pairwise interactions encoded in the interaction matrix $\mathbf{A}$. Note that the term A_{ii} contains the carrying capacity. The matrix $\mathbf{W} = (w_{ij})$ now encodes the environmental fluctuations such that the correlation $\langle \xi_i(t)\xi_j(t') \rangle = w_{ij}\delta(t - t')$. For simplicity, we can assume $\mathbf{W} = w\mathbf{I}$. In the limit of low connectance, the interactions can be considered as a small perturbation to the SLM model, which then allows agreement with Grilli's laws.

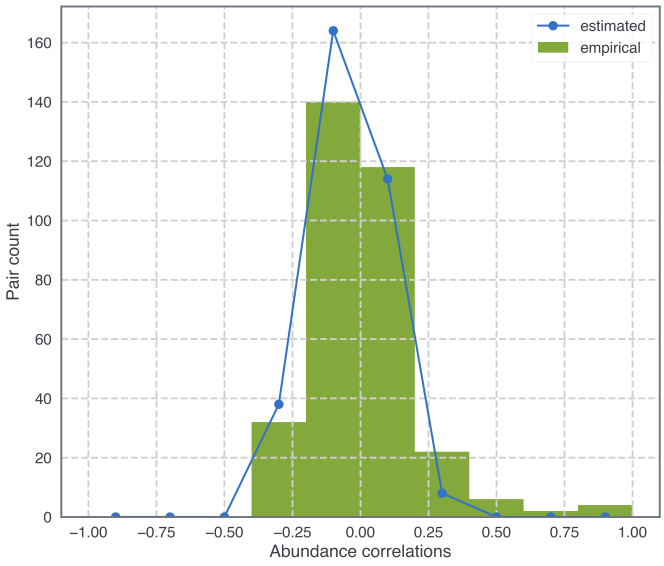

Figure 2.6: Empirical and estimated distributions of pairwise correlations in abundances. The estimated distribution corresponds to inferred interactions in a stochastic Lotka-Volterra model.

Inferring interactions from time-series data

For a given time series, we can compute the empirical distribution $\rho(x)$ of the pairwise correlations in abundances and use that to estimate the interaction matrix $\mathbf{A}$. This is achieved by using a Markov Chain Monte-Carlo approach where starting from an interaction matrix prior, we randomly pick a species pair at every step and change its pairwise interaction slightly. Changes are accepted if the likelihood of the correlation distribution $P(\rho|\mathbf{A})$ improves and the updated interaction matrix results in a stable and feasible community. In figure 2.6, we show the empirical correlation distribution against that estimated from inferred species interactions. This is based on abundance time series from gut microbiome data processed by Dr. Nandita Garud's group at UCLA.

CHAPTER 3

Present Work

3.1 Paper 1

This paper draws from the effects of the interaction strength parameter (σ) as in equation 2.9 to understand how assembled communities could be related to the patch area. The GLV equations can generate assembled communities of different sizes from a source pool through σ or equivalently the standard deviation of the interaction matrix [67]. The number of species decrease monotonically as σ increases in the regime where stable equilibria exist. This can be likened to the well-studied increase in species richness for increasing areas. It is interesting to ask if a modified version of the GLV equations could yield known forms of the species-area relationship without invoking any structure in species interactions? We address this question through a simple modification of the GLV equations that combines ecological intuition with observations from empirical studies.

Methods

The interaction strength parameter (σ) in equation 2.9 scales the strengths of all inter-species interactions. This effect is analogous to shrinking the area of a patch such that species encounter each other more frequently. Before analysing the consequences of spatial scaling, we make a few modifications to our existing system of equations. We use A to represent area of the patch now, so we represent interactions by B_{ij}. If the abundance densities are replaced by absolute abundances x_i , then equation 2.9 transforms to:

$$\frac{dx_i}{dt} = r_i x_i (1 - \frac{x_i}{K_i}) + \frac{x_i}{A} \sum_{j \neq i} B_{ij} x_j \tag{3.1}$$

where we do not use the σ parameter which was effectively the proxy for the standard deviation of the interaction matrix. Since K_i and x_i have the same units, additional factors of A do not appear in that term. Next, we account for the fact that absolute carrying capacities should change with changing areas. We modify equation 3.1 as:

$$\frac{dx_i}{dt} = r_i x_i (1 - \frac{x_i}{K_i (\frac{A}{A_{init}})^{\gamma}}) + \frac{x_i}{A} \sum_{j \neq i} B_{ij} x_j \tag{3.2}$$

where A_{init} is the minimum area that supports all species from the pool in the assembled community – this can be fixed by running the model numerically. We fix $\gamma = 0.25$, but in general $\gamma < 0.5$ is consistent with the results that are reported. The non-trivial scaling of the carrying capacity corresponds to spatial clustering of conspecifics. Clustered conspecifics would already have saturated levels of negative density dependence that is expected to change less drastically with change in areas. If the area is held fixed, lower values of γ translate to high spatial aggregation and weak negative density dependence.

Island SARs are particularly amenable to analysis through the model described. For a given A, the dynamics results in the assembly of a community of species from the species pool. By progressively sweeping over different values of A, we can get assembled communities representing different islands where species richness varies monotonically with A.

Community assembly with immigration

Immigration is another key aspect of island systems that could influence diversity patterns through introduction of new species [6], or the introduction of individuals of the existing species (demographic immigration) [87]. Demographic immigration is incorporated into our model using an additional term that captures incoming individuals from the mainland source pool:

$$\frac{dx_i}{dt} = r_i x_i (1 - \frac{x_i}{K_i(\frac{A}{A_{init}})^{\gamma}}) + \frac{x_i}{A}\sum_{j \neq i} B_{ij}x_j + \lambda e^{-\beta/\sqrt{A}} \tag{3.3}$$

The immigration term – $\lambda e^{-\beta/\sqrt{A}}$ – depends on A such that smaller islands receive almost negligible contributions through immigration. The exponential function further contains the parameter β – we fix β=1000 – that appears analogously as a characteristic length scale in [88]. We set the extinction limit to 10^{-5} even though this framework never really drives abundances to zero. The effective immigration rate equals λ for large A, which means that λ imposes limits on the immigration rates possible for an island system. We can compare results for different levels of immigration using this parameter.

Structured communities and empirical comparisons

We introduced many parameters to reflect the ecological setting of the Island SARs but this exercise helps avert some restrictive assumptions. In any case, our model still builds on the framework of studying random ecosystems especially in terms of how species interactions are structured. We also study ecological communities with more realistic connectances and distribution of interactions. This allows us to compare community level patterns that are insensitive to addition of further structure. The differences prompt scrutiny of additional mechanisms that drive the spatial patterns.

The findings are discussed in the context of empirical studies that benefit from SAR data over a large range of island areas [89–93]. Of particular interest are [89, 90], where two different Island SAR forms are discussed in relation to a common mainland.

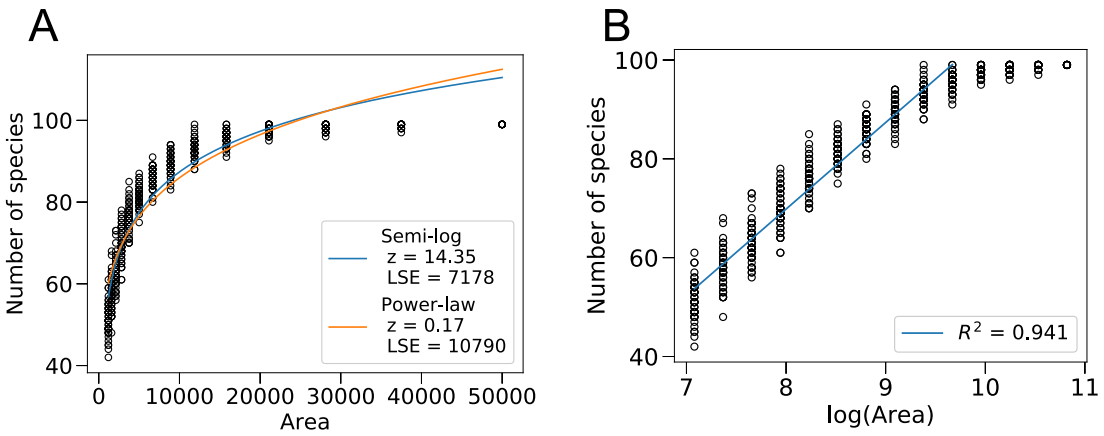

Figure 3.1: Species-area plots generated through 50 realizations of interaction matrix with mean = -1 and variance = 0.2. A_{init} = 50000. (A) The semi-log form shows a better fit. (B) The corresponding linear regression on a semi-log plot that shows an obvious upper asymptote.

Results

For no or low immigration rates, we get semi-log SARs (Figure 3.1). However, there exists an intermediate range of immigration rates that supports power-law SARs (Figure 3.2).

If interactions are structured such that the community contains many weak interactions and a fewer strong ones, then a higher skew towards weaker interactions shifts the SAR from a power-law to a semi-log relationship. Lower connectances – such as those found in real communities result in lower SAR slopes as in most natural communities.

Discussion

Our study identifies two major factors that differentiate semi-log and power-law SARs: immigration rates and skewness towards weak interactions. It is also evident that both functional forms show similar fits to the data except for very small and very large areas.

The case of remote archipelagos

A key result we report is that low immigration rates translate to semi-log SARs. As a crude test, empirical studies from remote archipelagos should

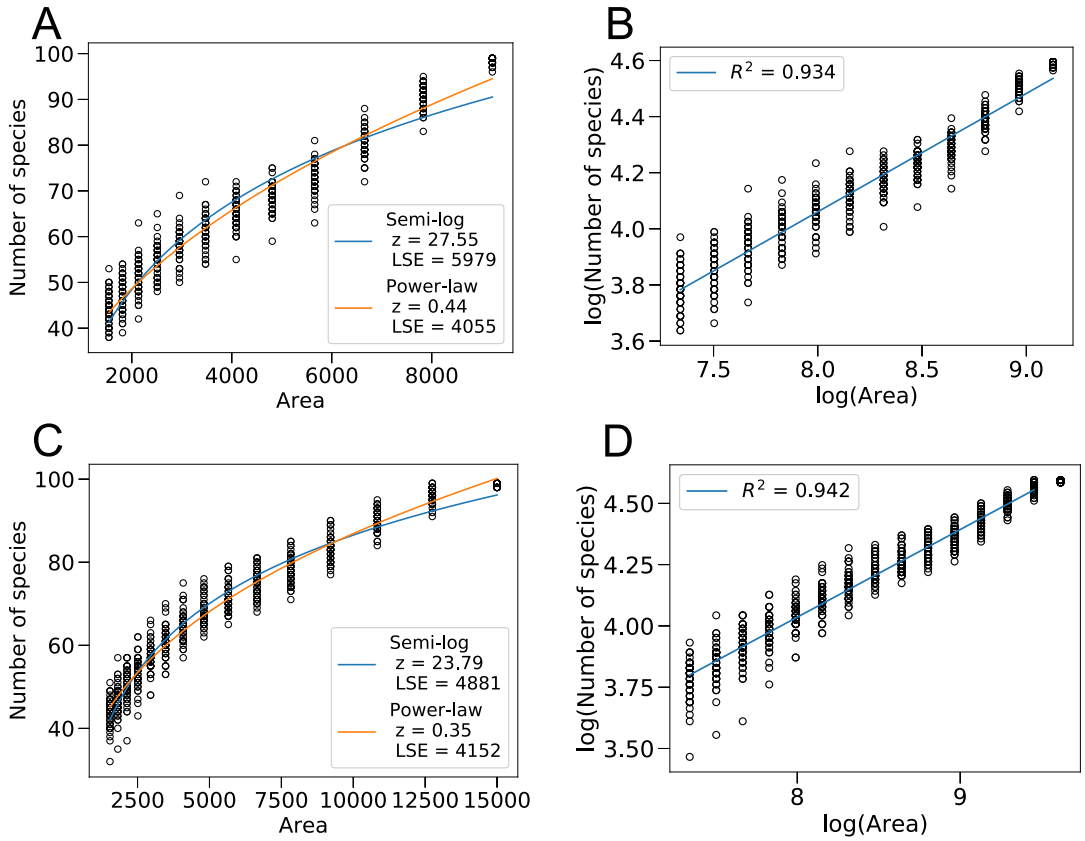

Figure 3.2: Species area plots demonstrating the better fit of power law SAR for intermediate values of immigration rates. Panels A and C show the fits for $\lambda = 0.1$ and $\lambda = 0.01$ respectively for 50 instances of the interaction matrix. Panels B and D correspond to the respective linear regressions on log-log plots. The interaction strength mean and variance are -1 and 0.2 respectively. $A_{init} = 15000$.

concur with this result. Two extensive studies from the Southwest Pacific are extremely well-suited to this end [89, 90]. The Solomon archipelago studied in [90] is more than 600 km away from the 'source island' of New Guinea. The authors find conclusive support for the semi-log form for islands that span over six orders of magnitude. The archipelago consists of three clusters of islands – each of which show very similar SAR slopes. In fact, the authors state that with increasing isolation of an archipelago, the SAR may shift in form from a power function to an exponential (semi-log). This observation is complemented by another study [89] (by the same authors), where they

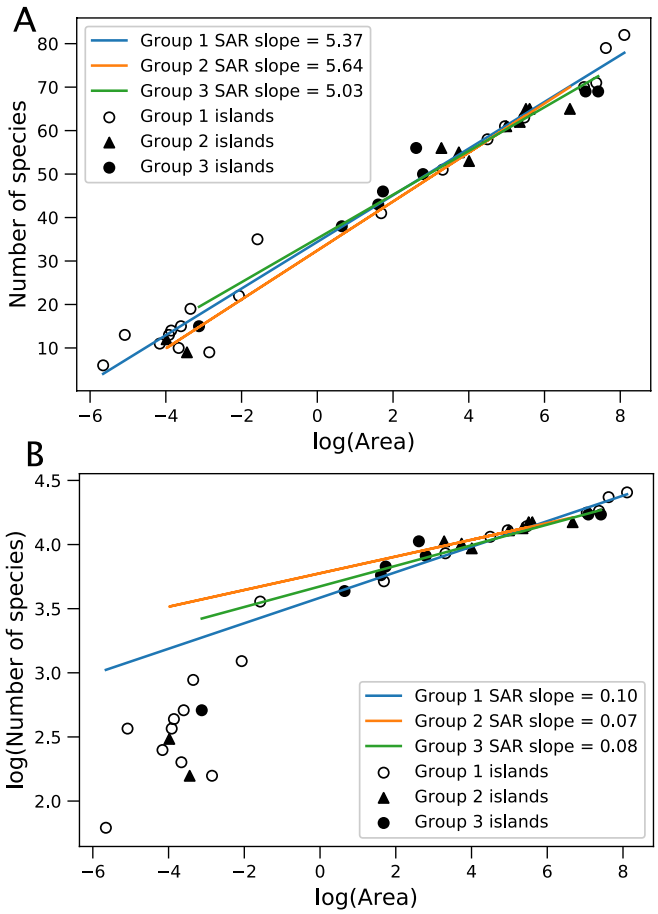

Figure 3.3: SAR plots for three groups of non-isolated islands within the Solomon Archipelago. These groups differ in how the islands within them were connected during the Pleistocene period. The islands in Group 3 did not have any history of connections. The semi-log relationship shows a good fit to data (A). The R^2 values for the regression lines are 0.978, 0.982 and 0.955 for Group 1, 2 and 3 respectively. The slopes for the different groups are very similar. Panel B shows a clear departure from a power-law relationship for smaller areas. The linear regression lines indicate a good fit for islands larger than one square mile. In particular, the R^2 value for such islands in group 1 is 0.976 from the power-law SAR. This demonstrates that a naive inference could support a power law, in spite of the islands spanning over four orders of magnitude in area ($>$ 1 square mile).

study islands that lie 5 to 500 km from New Guinea. That dataset shows a clear power-law for the SAR. Our finding on semi-log SARs is also supported

by data from distant archipelagos in the Atlantic [91] as well as the Indian Ocean [92, 93] (Supplementary Appendix S2 in Paper 1).

Conclusion

This study disentangles factors that might result in the two most widely observed SAR forms for islands. We conclude that the semi-log functional form is observed for islands systems with low immigration rates. On remote archipelagos, the difference between the two functional forms might only be evident on the smallest islands where the power-law form overestimates species richness. This calls for a systematic estimation of the Island SARs in field studies, given that islands systems have witnessed disproportionately large number of extinctions [94, 95].

3.2 Paper 2

Consider a large contiguous landscape characterized by a many species interacting over different spatial locations. What diversity patterns emerge if the landscape is understood as a collection of many habitat patches connected by high dispersal of species? Of particular interest is the limit of very high dispersal resulting in spatial coherence of species densities across patches. When studied for the theoretical extreme of total randomness and no spatial correlation in inter-species interaction strengths, spatial coherence simplifies the analytic investigation of complex ecological systems. We further extend the emerging simple picture to more structured landscapes so as to explore various aspects of habitat heterogeneity and its relationship to diversity patterns.

Methods

We use a spatially-extended version of the GLV equations where dispersal between well-mixed patches is captured through diffusion. The equations read as:

$$\begin{aligned}\frac{\partial \phi_i(\mathbf{x},t)}{\partial t} &= r_i \phi_i(\mathbf{x},t)\left(1 - \frac{\phi_i(\mathbf{x},t)}{K_i}\right) \\ &+ \phi_i(\mathbf{x},t)\sum_{j=1}^{N} A_{ij}(\mathbf{x})\phi_j(\mathbf{x},t) \\ &+ D\nabla^2\phi_i(\mathbf{x},t),\end{aligned} \tag{3.4}$$

where $\phi_i(\mathbf{x},t)$ are species abundance densities for species i at position $\mathbf{x}$ and at time t. As previously, r_i and K_i represent intrinsic growth rates and carrying capacities respectively. These equations resemble reaction-diffusion equations, but we discretize the spatial dimension such that a given point in space represents a well-mixed patch that can undergo local GLV dynamics in the absence of dispersal. The Laplacian operator $\nabla^2 = \sum_{\alpha=1}^{d} \partial^2/\partial x_\alpha^2$ (where d is the dimension of space) is discretized on a two-dimensional lattice as:

$$\nabla^2\phi_{i,\alpha\beta} = \left(\phi_{i,\alpha+1\beta} + \phi_{i,\alpha-1\beta} + \phi_{i,\alpha\beta+1} + \phi_{i,\alpha\beta-1} - 4\phi_{i,\alpha\beta}\right)/h^2. \tag{3.5}$$

which implies that the Laplacian operator acting on abundance density $\phi_{i,\alpha\beta}$ at spatial patch (α,β) transforms it based on the abundance densities at adjoining patches and the given patch. Beyond this cumbersome notation, the diffusion term has a simple function – to enable dispersal between patches such that the abundance densities are 'equalized' or homogenized. The equation 3.4 also contains interaction and growth terms that dictate the dynamics of the system in conjunction with the diffusion term.

One primary objective of this analysis is to understand how spatial heterogeneity in interactions affects diversity patterns. The interaction strengths $A_{ij}(\mathbf{x})$ between any pairs of species are allowed to vary with the patch location $\mathbf{x}$, although the local means and variances of the interaction matrices are kept constant. In effect, interaction heterogeneity is used as a placeholder for habitat heterogeneity.

High-dispersal and spatial coherence

In the high-dispersal limit, abundance densities become coherent across spatial patches. An observation greatly simplifies the analysis of these spatially-

extended systems – in the coherent limit, the system can be described by regular GLV equations with an effective interaction matrix $\bar{A}$ (cross ref fig). That is:

$$\frac{d\phi_i(t)}{dt} = r_i\phi_i(t)\left(1 - \frac{\phi_i(t)}{K_i}\right) + \phi_i(t)\sum_{j=1}^{N}\bar{A}_{ij}\phi_j(t), \tag{3.6}$$

Each entry within $\bar{A}_{ij}$ is an average of the corresponding pairwise interaction over all spatial patches. The standard deviation of $\bar{A}_{ij}$ could be used to study the stability and species richness of the spatially-extended ecosystem.

Spatial heterogeneity and species richness

It is unlikely that the interaction strength between two species is entirely uncorrelated across all spatial locations. A more reasonable setting is to consider some correlation between interaction strengths across space. We investigate how spatial heterogeneity affects the relationship between the number of habitats and species richness. This relationship has been actively studied in the context of the Area-Heterogeneity Trade-off (AHTO) hypothesis. The earliest proposals championed a unimodal relationship between the number of habitats (located in a fixed area) and species richness [96]. However, a recent experimental study reported a monotonic functional form [97].

The authors of [97] expected the positive relationship to be strengthened by increasing dispersal. The high dispersal coherent limit we described is conducive to theoretically understand these results. This is partly because the effect of environmental stochasticity is subdued when inter-species competition shows high variation [98]. High dispersal also increases the rate at which community assembly takes place, which means that stochastic effects are less dominant.

The simulation setup consists of nine patches, each of which could have a habitat characterized by an interaction matrix. The interaction strength for each species pair is correlated by a constant amount across habitats. We run our simulations for different spatial correlations ρ and plot the species richness resulting from an increasing number of habitats that could be randomly distributed over the nine patches.

It is unlikely that constant spatial correlations exist for very distant spatial patches. If a fixed correlation is assumed between adjacent spatial patches

only, then how do species richness patterns compare against those from uniformly correlated cases? We address this question by analysing the situation in one spatial dimension.

Results

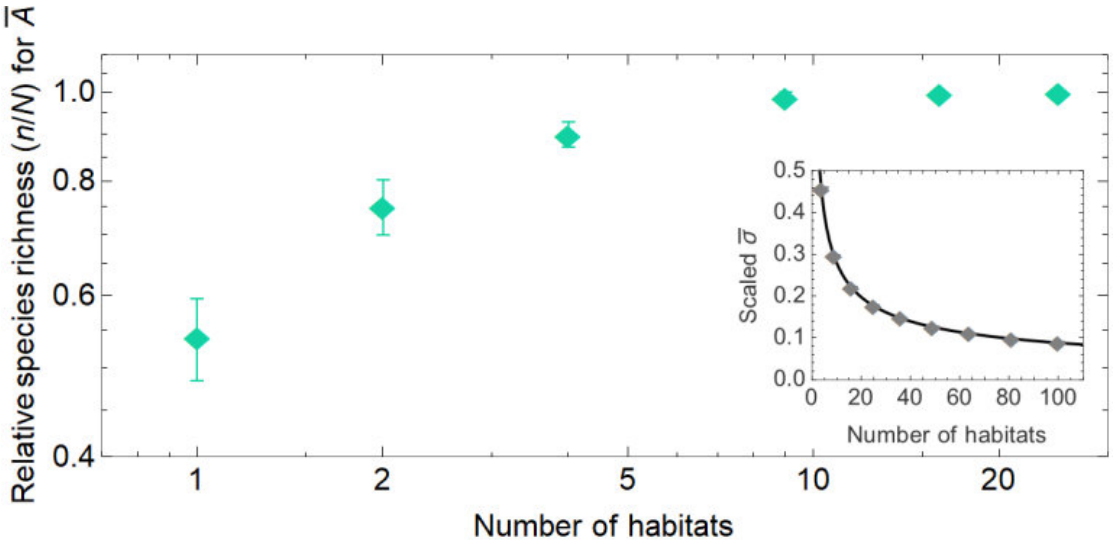

Figure 3.4: The plot shows how the relative species richness (n/N) increases with heterogeneity, in effect number of habitats. The data points are species richness averages from 20 runs of systems with $\mu_{A_{\alpha\beta}} = 0$ and $\sigma_{A_{\alpha\beta}} = 5/4\sqrt{cn}$ and one standard deviation errorbars. The species richness saturates at relative species richness 1, which corresponds to complete coexistence of all species in the original species pool. The inset shows the decrease of the standard deviation of the entries of the effective interaction matrix $\bar{\sigma}$. Since the effective system captures the dynamics and stability properties of the underlying metacommunity, this demonstrates (as in the framework of May) how the proxy for effective complexity of the metacommunity decreases, thereby allowing for a higher species richness.

If species interactions are totally uncorrelated in space, then the variance $\bar{\sigma}^2$ and mean $\bar{\mu}$ of the effective interaction matrix have exact analytical expressions. If each local interaction matrix has mean μ and standard deviation σ, then we find:

$$
\begin{aligned}
E[\bar{A}_{ij}] &= c\mu \\
&= \bar{\mu} \\
\\
Var[\bar{A}_{ij}] &= \frac{c}{G}(\sigma^2 + \mu^2(1-c)) \\
&= \bar{\sigma}^2,
\end{aligned}
\tag{3.7}
$$

where c is the connectance. Each non-zero entry corresponding to species i and j at spatial patch g is considered a stochastic variable X_{ijg} where $i = j = 1, 2, ..N$, $g = 1, 2...G$.

Equation 3.7 implies that $\bar{\sigma} \to 0$ in the limit of infinitely many habitats (limit $G \to \infty$). Equivalently, a fully-feasible and stable ecosystem is theoretically possible for indefinitely large local σ if there exist an infinite number of habitats connected by high dispersal.

Spatial correlation mediates heterogeneity-richness relationships

Low values of ρ result in a significantly positive relationship between species richness and the number of habitats, irrespective of whether the interaction matrix mean is zero or negative (Figure 3.5) . For the case of zero interaction mean and spatial correlation ρ, the analytical expression for the variance $\bar{\sigma}^2_\rho$ of the effective interaction matrix can be calculated. This reads as:

$$
\begin{aligned}
Var[\bar{A}_{ij\rho}] &= \frac{\sigma^2}{G}\left(1 + (G-1)\rho\right) \\
&= \bar{\sigma}^2_\rho.
\end{aligned}
\tag{3.8}
$$

where σ is the standard deviation of the local interaction matrices, as previously. In the limit of very large number of habitats G, we get

$$
Var[\bar{A}_{ij\rho}] \sim \rho\sigma^2 \tag{3.9}
$$

If one assumes spatial correlation ρ_{nn} between adjacent patches only, then the following result holds (for one-dimensional space):

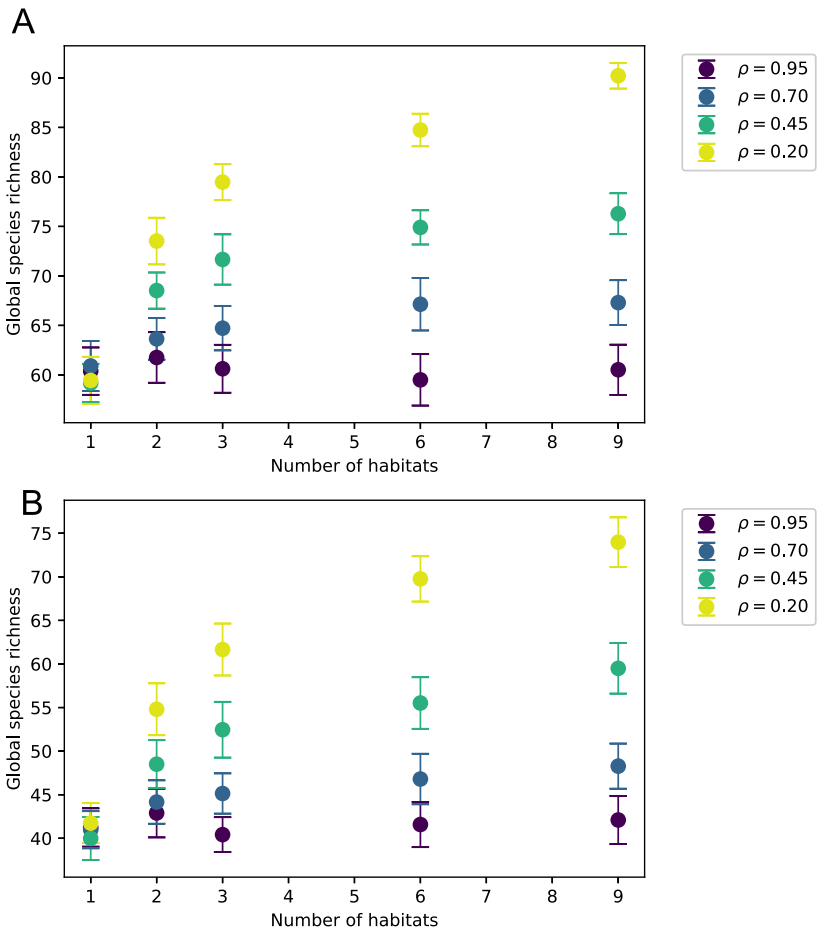

Figure 3.5: Relationship between habitat heterogeneity and species richness. For a given value of ρ, the plots show the mean species richness (with one standard deviation error bars) resulting from different number of habitats distributed over nine spatial patches. A given number of habitats corresponds to an equivalent number of correlated random interaction matrices such that the means are computed over 50 realizations of these matrices. The two panels correspond to (A) Mean interaction strength = 0. (B) Mean interaction strength = -0.5.

$$Var[\bar{A}_{ij\rho_{nn}}] = \frac{\sigma^2\left(G + 2\sum_{\eta=1}^{G-1}\rho_{nn}^{\eta}(G-\eta)\right)}{G^2} = \bar{\sigma}^2_{\rho_{nn}}. \tag{3.10}$$

If two habitat patches are separated by k patches between them, then the spatial correlation between them is equal to $(\rho_{nn})^{k+1}$. In other words, correlations fall off exponentially as the separation between patches increases. Given infinitely many spatial patches, $\bar{\sigma}_{\rho_{nn}} \to 0$. High species richness could still result from high spatial correlation between adjacent patches. The exponential decay of correlations with distance, paves way for large spatial variation in species interactions.

Discussion

Spatial coherence through high dispersal offers a unique perspective to understand diversity patterns in spatially-extended ecosystems. All the results we presented were obtained in this limit. We consistently found that spatial heterogeneity and high dispersal promote species richness for complex local communities.

The theoretical extreme of spatially uncorrelated interaction matrices provides simple yet rich insights. We show analytically that such a setting could allow for fully-feasible and stable metacommunities even if the local communities are extremely complex. Dispersal and heterogeneity have been shown to promote stability earlier as well [59].

The standard deviation of the effective interaction matrix could also be used to study diversity patterns in cases where species interactions are correlated uniformly everywhere or only between adjacent patches. We found that increasing the number of habitat patches in a fixed area results in significantly higher species richness, as opposed to the unimodal relationship advocated by the AHTO hypothesis. Since authors of [97] reported that increasing the number of habitats had a significantly positive effect on species richness, we surmise that this could be a consequence of high spatial variation in interspecies interactions.

A sufficiently large contiguous landscape would always have variation in species interaction strengths over space, which motivates the picture of many habitat patches connected to one another. The different cases we analysed consistently showed that spatial heterogeneity in interactions promotes high species richness. Factors that promote such heterogeneity could therefore explain how high diversity persists in large contiguous landscapes.

3.3 Paper 3

Biological invasions show up in a variety of contexts. Over the course of their history, ecosystems regularly witness arrival of new species through processes such as range expansion which could be a result of chance events or human intervention. Invasions could also come from within the community, such as when a new mutant strain evolves from an existing strain of bacteria within a microbial community. Invasion events are therefore the hallmark of ecological communities rather than mere aberrations. The study of invasions is thus intimately tied to understanding the assembly of ecosystems and their possible future states in the face of impending invasion events.

The outcomes of invasions are hard to predict, especially if the communities are large and the invading species are not rare. Large communities are characterized by species interactions that are hard to measure precisely. It is common to classify invasion outcomes into broad categories like coexistence, turnover, and invasion failure [99, 100]. This paper furthers the scope of prediction beyond these broad outcomes to identify the structure of ecological communities post invasion. The framework we develop, aims to predict the structural outcomes of invasion events, i.e., what set of species persist at the new equilibrium. It also allows much flexibility in initial invader abundance, structure of interactions and the functional response of the underlying model.

Methods

The prediction workflow is illustrated in figure 3.6. We divide the community's time evolution into three chronological states: pre-invasion, transient state and post-invasion state.

Let's consider population dynamics of the following form:

$$\frac{dX_i}{dt} = X_i R_i(\mathbf{X}) \tag{3.11}$$

where $R(\mathbf{X})$ is the effective growth rate function that also contains species interaction coefficients.

Our framework uses the following important result based on the concept of a strongly uninvadable state 2.44 – If the current state of the system is in the vicinity of a strongly uninvadable state, then the relative entropy between them decreases monotonically with time [101]. We use this result to prune

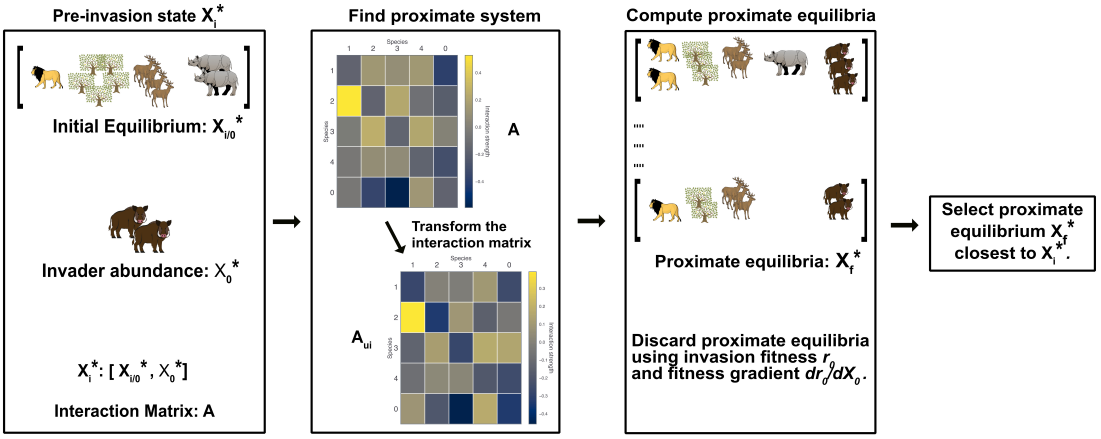

Figure 3.6: This illustration shows our prediction workflow in the most general case. The proximate equilibria here differ in terms of having different set of surviving species. As we discussed in the Methods section, a possible estimate for the invader's final abundance can also be used to discard certain outcomes for type I functional response.

the space of invasion outcomes by time evolving the current system state and computing its distance to possible uninvadable equilibria. If a system does not admit such equilibria, then we use the uninvadable equilibria of a sufficiently close system to compute distances from the current system state. We call such nearby systems as proximate systems and the corresponding equilibria as proximate equilibria.

Proximate systems can be constructed for models such as the gLV equations by making the self-interaction terms more negative. We describe another approach in the paper, but in general no unique procedure exists.

We further use the invasion fitness and fitness gradient to time evolve the invader's abundance. Time evolving the abundance state of the entire community is desirable but complicated, so the invader's abundance trajectory serves as a useful proxy. With regards to the equation 3.11, invasion fitness can be written as:

$$r_0 = R_0(X^*_{/0}) \tag{3.12}$$

where the subscript /0 denotes omission of the invader's abundance. Arnoldi et al. [100] defined fitness gradient by considering the addition of an invading species as a press perturbation on the growth rates of the resident community.

The fitness gradient can be expressed as:

$$\frac{dr_0}{dX_0} = \frac{\partial R_0}{\partial X_0} + \sum_{i \neq 0} \frac{\partial R_0}{\partial X_i} \frac{\partial X_i^*}{\partial X_0} \tag{3.13}$$

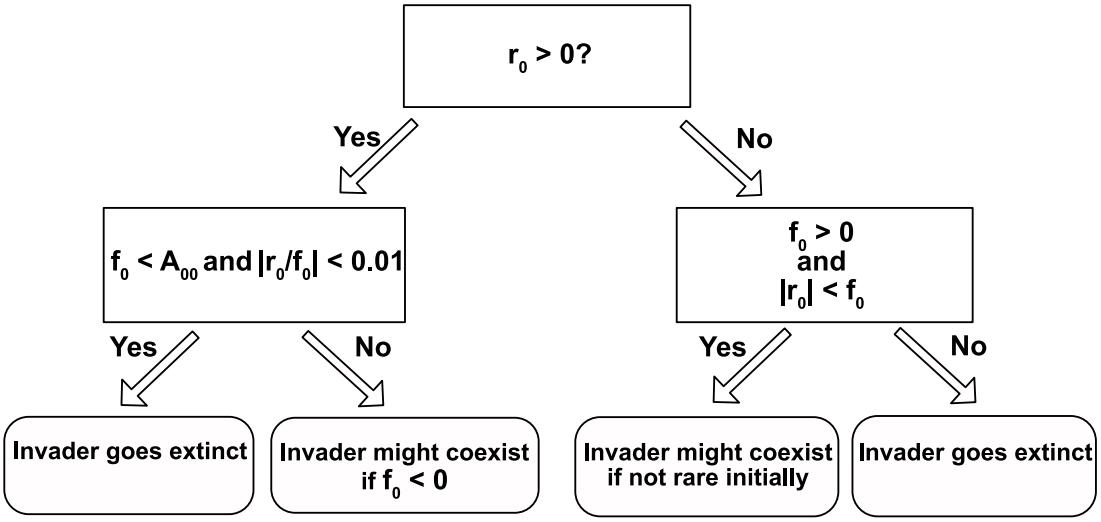

Figure 3.7: The decision tree determines the outcomes for the invader based on the values of the invasion fitness r_0 and f_0. The invader always goes extinct if its initial abundance is low and $r_0 < 0$.

The decision tree (Figure 3.7) uses the invasion fitness and the fitness gradient to determine the approximate trend of the invader's final abundance. In cases where the invader's extinction can't be determined, the final abundance is estimated as follows (Figure 3.7):

1. If $f_0 < 0$ and r_0 is not too large, the final invader abundance is approximately equal to $\frac{r_0}{|f_0|}$ ([100]). We use this approximation only if $|r_0| < 10 * |f_0|$.

2. Otherwise we set the invader's final abundance equal to the abundance from the corresponding proximate equilibrium.

After filtering out implausible proximate equilibria, we predict the structure of the invasion outcome using the proximate equilibrium that is closest to the pre-invasion state.

To quantify how well our framework performs, we define a predictive score that takes values between 0 and 1, where the score is 1 if the invasion outcomes structure is exactly predicted. For cases where a fraction of the extinct species are correctly identified, the prediction score is simply the ratio of the predicted number and the actual number of extinct species. In all other cases, the prediction score is 0.

We test our framework for many simulated settings including evolutionary contexts where new mutant strains invade an existing community. Recent experiments with microbes have shown that bottom-up assembly processes cannot build many larger complex communities [102]. We compare the results for invasions in communities which are assembled in two different ways:

1. For sequentially assembled ecological communities, we start with an initial feasible community of 10 species and introduce invaders one at a time. We track the predictive ability of our framework as the starting community is pruned by subsequent invasions. We expect that predicting invasion outcomes in this case should be easier than that of top-down assembled communities.

2. For the case of top-down assembled communities, we build a feasible community with some number of species thrown in together and then introduce an invader. We repeat this for a feasible resident community of the same size and with statistically similar parameters. The results are then compared for different sizes of resident communities that have between 10 and 100 species.

We also consider the case of type-II functional response with the following dynamical equations:

$$\frac{dX_i}{dt} = X_i\Big(r_i - \frac{r_i X_i}{K_i} + \epsilon \sum_j \frac{A_{ij}X_j}{1+\theta^{-1}\sum_k A_{ik}X_k} - \sum_j \frac{A_{ji}X_j}{1+\theta^{-1}\sum_k A_{jk}X_k}\Big) \tag{3.14}$$

A_{ij} is the positive per-capita effect on species i from j. We draw A_{ij} based on the body masses of different species. By picking a random niche value n_i between 0 and 1 to each species, the body masses w_i are assigned as $w_i = w_{\max}^{n_i} w_{\min}^{1-n_i}$ [103]. A species only feeds on prey that lies within its niche feeding range. We set $\epsilon = 0.2$ and $\theta = 1$.

We also analyze a food-web dataset from lakes in Switzerland which witnessed the introduction of European catfish a decade ago. We use the allometric dynamical food-web model described in [104]. Parameters such as metabolic and mortality rates are scaled with body sizes to infer different interactions in the food-web. The dynamics exhibits abundance fluctuations for long times, so we use our framework to forecast extinctions in the system. We first assemble a food-web with just the resident species, and then predict the outcomes for when the invading species is introduced.

Results

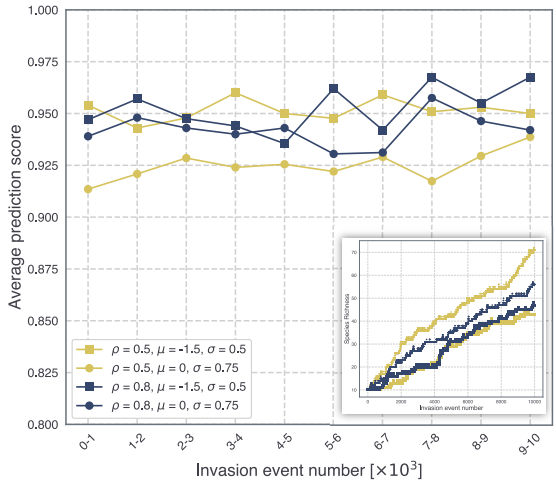

Figure 3.8: Starting with an initial feasible community of 10 species, the figure shows average prediction scores as invaders are added sequentially. The scores are averaged over 1000 consecutive invasion events. The initial resident community is generated using $\mu = 0$, $\sigma = 0.5$ and $\mu = -0.5$, $\sigma = 0.3$ for the 0 and negative mean cases respectively. The inset shows the progression of species richness for each of the cases. The two bottom-most trajectories in the inset correspond to the negative mean cases.

Figure 3.8 shows how the predictive score changes with successive invasion events when the invaders interaction profile is correlated to one resident species, and these two species strongly interact with each other. Under these conditions, invasion is much less successful than when such strong interactions do not exist (Figure 3.9). The plots with $\rho = 0.8$ are more in line with

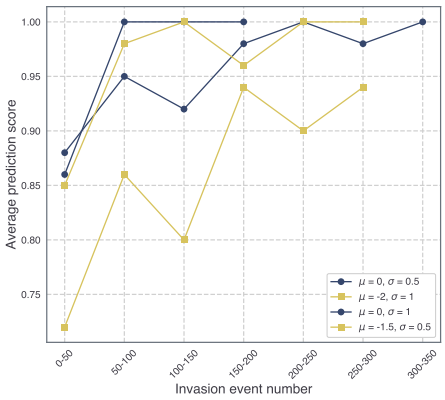

Figure 3.9: Plots for sequential invasion events when the invader's interaction profile is correlated to a randomly selected resident species, but no disproportionately strong interactions exist. All trajectories have $\rho = 0.5$. Most events result in coexistence, and hence no more species are added after a few hundred invasion events.

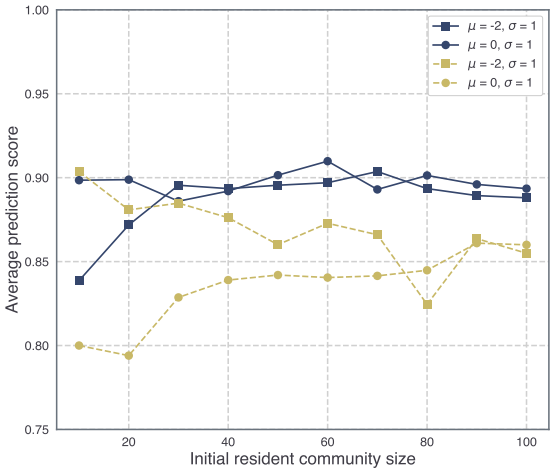

Figure 3.10: Plots for invasion of top-down assembled communities. For a given resident community size, the prediction scores are averaged over 1000 different communities with $\mu = 0$, $\sigma = 0.3$ and $\mu = -0.5$, $\sigma = 0.3$ for the 0 and negative mean cases respectively. The figure legend shows the parameters of the invader's interaction profile. $\rho = 0.5$ for all the cases shown. The dashed lines show results for initial invader abundance = 0.1.

invading mutants, where ρ is the correlation coefficient between the species interactions of the invader and one randomly selected resident species. μ and

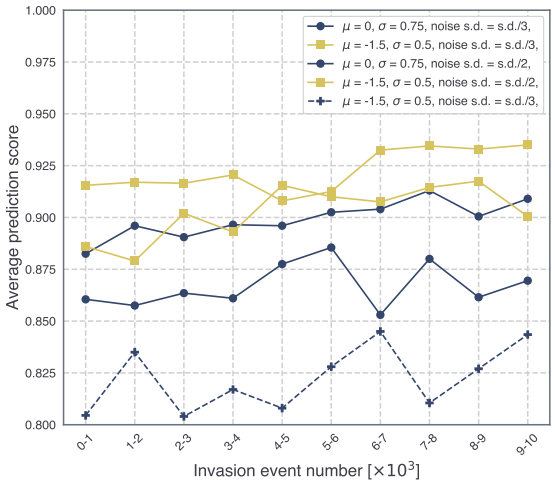

Figure 3.11: Plots for sequential assembly when noise is added to the interaction matrices. The s.d. here is equal to $\sigma/\sqrt{c * N}$ which corresponds to the distribution from which the invader's interactions are drawn. Noise is added to the full interaction matrix only for the dashed line case. Otherwise, we assume that the effects on the invader A_{0i} are known perfectly.

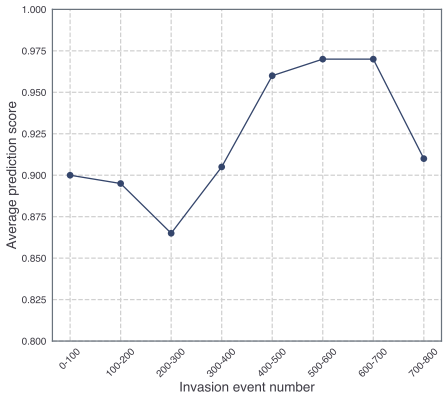

Figure 3.12: Plot for the case of type II functional response. The invader's interactions are drawn using a random niche value.

σ are the mean and standard deviation of the normal distribution from which each interaction term is drawn.

Figure 3.9 also corresponds to cases with sequential assembly, but the invaders interactions are drawn such that it does not have disproportionately strong interactions with any of the resident species. Most invasion events

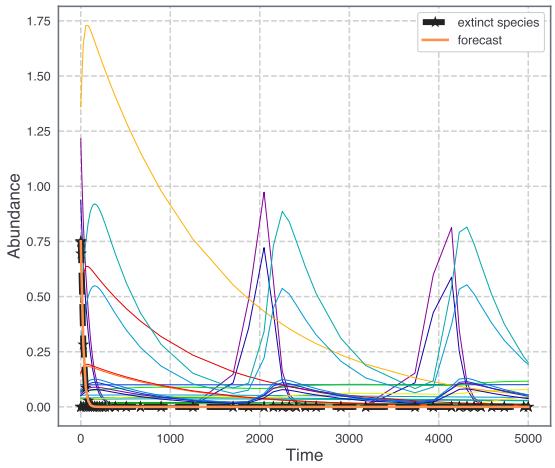

Figure 3.13: The dynamics of an aquatic food-web ([104]) after introduction of the invasive species. As in the dynamics, our framework forecasts the extinction of the Daphniidae species cluster (thick orange line) even though its initial average abundance is high.

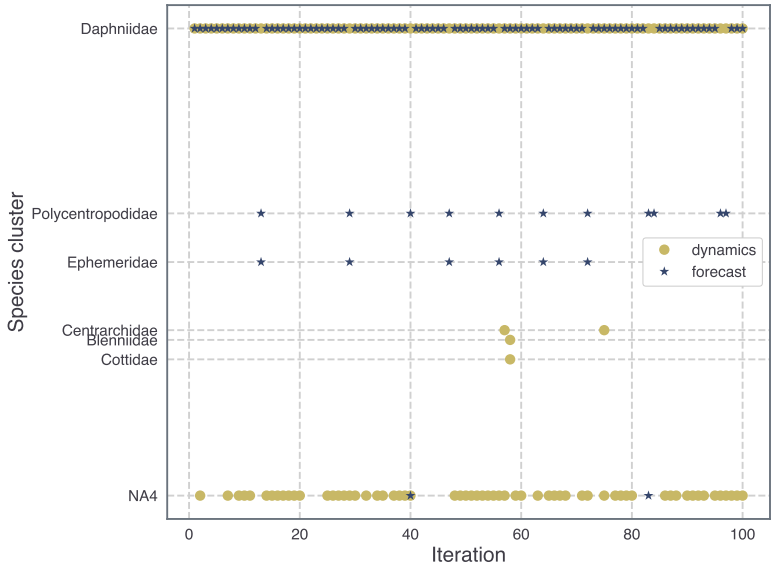

Figure 3.14: Comparisons of extinct species clusters from the simulated dynamics and our framework's forecasts. Different iterations correspond to adding white noise to the metabolic rates of species.

result in coexistence. We have more than 200 species with just around 300 invasion events and we therefore do not add any new invaders further.

Top-down assembly results in lower predictive scores, but the results are robust to changes in the pool size (Figure 3.10).

In figure 3.11, we show results for cases when interactions are known imperfectly. The prediction scores are relatively lower if we add noise to all the non-zero inter-species interactions, as shown by the dashed line. In all other cases, we omit noise from the interaction terms A_{0i} which capture the effects on the invader and also appear in the invasion fitness. Under these constraints, our approach captures a large fraction of the structural outcomes except those where the invader and resident species coexist. The scores are lower for cases with mean 0 because addition of noise could change the sign of the interactions. In real contexts, we expect that the sign of these interactions could be measured reliably in many settings, and adding noise while still preserving the signs would yield better results.

The case with type 2 functional response is shown in Figure 3.12. The predictive power is lower compared to the type 1 case.

With regards to the aquatic food-web [104], we see that one species cluster (Daphniidae) declines from high average abundance to extinction (Figure 3.13). This is seen over many different realizations, each corresponding to slight variations in metabolic rates (Figure 3.14). The forecasts from our framework consistently converge with this result from the simulations.

Discussion

The main strength of the proposed framework is its applicability to a broad range of settings even when the invaders are not initially rare.

Species interactions are difficult to estimate [105–108]. We therefore analysed the cases for which the interactions are known imperfectly. We found that the outcomes can be predicted better if the effects on the invader are known with higher certainty. This is in line with previous studies [100], and it also suggests that measurement effort should focus on such interactions.

Our framework is closer to the structural perspective in spirit [37, 54, 55]. While these studies work with fixed interaction matrices and explore the space of intrinsic growth rates resulting in feasibility, our work identifies variations in the interaction matrix that preserve the structural outcomes.

In our analyses, coexistence and single extinction outcomes were more predictable than extinctions of species pairs. Events with many extinctions could be missed by our framework. The possibility of even small structural changes

via extinctions should anyway provide useful early warnings for better management of such systems.

The extinction forecast from the aquatic food-web [104] we analyzed, consistently indicates extinction of the large zooplanktonic cluster Daphniidae. Given that Daphniidae are important food sources for fish in the lakes [109], invasion management could focus on measuring impacts of the catfish on these planktonic species.

It would be useful to come up with transformations that make more ecological sense since our current methods for finding proximate systems are motivated almost entirely by mathematical simplicity. Transformations that preserve the expected sign of interactions between species could be a good place to start.

3.4 Paper 4

May's influential paper [28] sparked the *complexity-stability debate* which has led to a flurry of papers – some addressing the framing of the problem [110], while others identifying factors that affect stability [103, 111–115]. The key result in May's paper drew upon Wigner's semi-circle law which states that the eigenvalue spectral density of large Hermitian random matrix ensembles has a semi-circular – more precisely, semi-elliptical – shape [38]. There had been related work on non-Hermitian matrices before May, which established the circular law stating that for random matrices of size $N \times N$, the eigenvalues are distributed with uniform density at each point within a disk when $N \to \infty$ [39, 40]. For ecosystems represented by large random community matrices, May showed that the radius of the disk grows with the connectance, standard deviation and the number of species. This helped put strong mathematical bounds on the stability of ecological systems, if one were to characterize stability in terms of asymptotic stability.

Since then there has been work on finding mathematical bounds on the eigenvalue spectrum of non-random matrices, but the efforts have been limited to only a few matrix classes such as those with elements having correlations across the diagonal [45, 116–119]. In this paper, we find mathematical bounds on stability which are ecologically interpretable and are applicable to any real square matrix. We further derive stability conditions which can be expressed in terms of row-sums of the interaction matrix, which in turn relate

to the *Energetic Equivalence Rule* [120]. This rule states that the total rate of consumption per population is constant across species.

Finding ecologically-interpretable stability bounds

We use two key results to derive the new bound. The first is *Gershgorin's circle theorem* which states that the eigenvalues of a matrix lie within the union of disks defined by each matrix row (or column) such that the disk radii R_i are equal to the sum of absolute values of the non-diagonal elements of the respective row (or column). Also the disk centres O_i are equal to the respective diagonal elements.

$$O_i = A_{ii} \qquad \text{and} \qquad R_i = \sum_{j \neq i} |A_{ij}| \tag{3.15}$$

We assume that all the diagonal entries are the same negative value, i.e., $A_{ii} = -s$ which is a common assumption in almost all past theoretical studies. Using this theorem, we can place bounds on the real and imaginary parts of the eigenvalues by using the disk with the largest radius:

$$Re[\lambda(A)] \leq Max[\sum_{j \neq i} |A_{ij}|] - s \tag{3.16}$$

$$Im[\lambda(A)] \leq Max[\sum_{j \neq i} |A_{ij}|] \tag{3.17}$$

where $Re[.]$ and $Im[.]$ denote the real and imaginary parts of any eigenvalue of matrix A, i.e., $\lambda(A)$. Also $Max[.]$ denotes the maximum value.

We can already use this theorem to find a preliminary bound on eigenvalues and compare it to May's results. Consider the case $s = 1$. For large random $n/timesn$ matrices with connected C, the law of large numbers implies that the sum $\sum_{j \neq 1} |A_{ij}|$ will be equal to nC times the mean of the distribution from which the terms $|A_{ij}|$ are drawn. In fact, these terms come from a folded normal distribution if A_{ij} are normally distributed. The mean of the folded distribution is equal to $\sqrt{\frac{2}{\pi}}\sigma$ if the mean of the underlying normal distribution is equal to 0 and the standard deviation is σ. In the most general case, the mean of the folded distribution is always than σ.

$$Re[\lambda(A)] < nC\sigma - 1 \tag{3.18}$$

Compared to May's bound which scales as $\sqrt{nC}$, this is a much weaker bound but fairly easy to prove.

The second theorem we use are the Bendixson-Bromwich inequalities [121, 122]. These rely on the fact that a matrix A can be expressed as the sum of a symmetric A_s and anti-symmetric matrix A_{as}, i.e. $A = A_s + A_{as} = (A + A^T)/2 + (A - A^T)/2$ where T denotes the transpose. The inequalities are:

$$Re[\lambda(A)] \leq Max[\lambda(A_s)] \qquad \text{and} \qquad Im[\lambda(A)] \leq Max[\lambda(A_{as})] \tag{3.19}$$

By combining equations 3.16 and 3.19, we get the most general version of the new bound:

$$Re[\lambda(A)] \leq Max[\lambda(A_s)] \leq Max[\frac{1}{2}\sum_{j\neq i}|A_{ij} + A_{ji}|] - s \tag{3.20}$$

This is a tighter bound than the bound 3.16 because $Max[\sum_{j\neq i}|A_{ij}|]$ requires that most of the $|A_{ij}|$ terms are also large. This implies that most terms $\frac{1}{2}\sum_{j\neq i}|A_{ij} + A_{ji}|$ are less than or equal to $|A_{ij}|$.

The bound contains a sum over absolute values of terms within a row of the interaction matrix or its symmetric part. Intuitively, the row corresponding to the species experiencing the strongest interaction effects should maximize this sum. In this sense, this bound already points us to interesting ecological directions. We apply it to specific ecologically-inspired cases to highlight the insights gained from this simple bound.

First consider that for a predator-prey pair, the interaction terms would have opposite signs for the two. Further, the interaction term for the predator will be a fraction ϵ of the prey's interaction term. ϵ is the conversion efficiency, whose value is typically around 0.1 [123, 124] but we consider it to lie between -1 and 1 to capture a broad range of interactions. The corresponding matrix is generally expressed as:

$$A_n(\epsilon) = \begin{bmatrix} -s & \dots & \epsilon A_{n1} \\ \vdots & \ddots & \vdots \\ -A_{n1} & \dots & -s \end{bmatrix} \tag{3.21}$$

where $A_{ij} > 0$. Instead of being a constant, ϵ can also be used as a correlation between elements lying symmetrically across the diagonal [45, 116]:

$$\langle A_{ij}, A_{ji} \rangle = -\epsilon \tag{3.22}$$

We also consider another ecological direction that relies on the fact that species are limited in terms of the total resources they can consume. Our bound is well-equipped to utilize this fact by virtue of containing the row sums that represent the overall consumption and loss terms for each species. We use the example of the gLV equations whose Jacobian is the following:

$$J_n = \begin{bmatrix} -\dot{S}_1 & \dots & \epsilon A_{n1} N_1^* \\ \vdots & \ddots & \vdots \\ -A_{n1} N_n^* & \dots & -\dot{S}_n \end{bmatrix} \tag{3.23}$$

where $\dot{S}_i = r_i N_i^* / K_i$, N_i^*, r_i and K_i are the equilibrium abundance, intrinsic growth rate and carrying capacity respectively for species i. ϵ is the conversion efficiency. We will find stability bounds corresponding to matrices 3.21 and 3.23 in the results section below.

Comparing bounds on gut microbes

We also compare the two versions of our bound 3.18 and 3.20 on data from stable microbial systems in the human gut [125]. The differences between using the standard Jacobian and just the interaction matrix are also of interest, partly because most bounds are based on the interaction matrix alone, thus ignoring the role of abundances [28]. Further, the abundances are central to relating our bounds to principles like the Energetic Equivalence Principle.

Results

For large matrices of the form 3.22, using our general bound 3.20 gives:

$$Re[\lambda(A(\epsilon))] \leq Max[\lambda(A_s(\epsilon))] \leq \Big(\frac{1}{2}\big|1-\epsilon\big| \times Max[\sum_{j\neq i} A_{ji}]\Big) - s \qquad (3.24)$$

for the reason that terms in the symmetric matrix A_s would have factors of $(1-\epsilon)$ or $(\epsilon-1)$ and the sum is over absolute values of each term yielding $|1-\epsilon|$.

For the case of a large random matrix with connectance C, as in the inequality 3.18, we get:

$$Re[\lambda(A(\epsilon)] \leq \big|1-\epsilon\big| nC\sigma - s \qquad (3.25)$$

In a similar way, one can find the bound along the imaginary axis by considering the anti-symmetric part of $A(\epsilon)$ which would give $\big|1+\epsilon\big|nC\sigma$. These bounds show a similar dependence on ϵ and σ as Girko's elliptical law [43] which also considers a similar correlation across the diagonal.

With regards to the matrix 3.23, we consider a simpler case with all $\dot{S}_i = s$. For each species, we divide their interactions into gain and loss terms depending on whether these correspond to consumption/gain or loss. Then our general bound for this matrix gives:

$$Re[\lambda(J_n)] \leq Max[\lambda(J_{nS})] \leq \frac{1}{2} Max[\sum_{j,\text{gain}} A_{ji}\big|-N_j^*+\epsilon N_i^*\big| \sum_{j,\text{loss}} A_{ji}\big|\epsilon N_j^* - N_i^*\big|] - s \qquad (3.26)$$

The gain and loss terms within $Max[.]$ can be further written as:

$$\sum_{j,\text{gain}} A_{ji} N_j^* \big|1-\epsilon\frac{N_i^*}{N_j^*}\big| \qquad \text{and} \qquad \sum_{j,\text{loss}} A_{ji} N_i^* \big|1-\epsilon\frac{N_j^*}{N_i^*}\big| \qquad (3.27)$$

$\big|\epsilon\frac{N_i^*}{N_j^*}\big|$ and $\big|\epsilon\frac{N_j^*}{N_i^*}\big|$ are $<< 1$ given that $\epsilon < 1$ and consumer/predator populations are typically smaller than resource/prey populations. Species i is a consumer in the gain terms while a resource in the latter. Hence,

$$Re[\lambda(J_n)] \leq \big(\frac{1}{2}Max[\sum_{j,\text{gain}} A_{ji}N_j^* + \sum_{j,\text{loss}} A_{ji}N_i^*] - s\big) = \big(\frac{1}{2}Max[\dot{D}_i + \dot{E}_i] - s\big)$$
$$= \frac{1}{2}(\dot{D} + \dot{E}) - s \tag{3.28}$$

where $\dot{D}$ and $\dot{E}$ are the maximum overall rates of consumption/gain rate and loss for any individual in the system across species. Crucially, this bound does not increase with the number of species, thus providing a resolution to May's paradox.

We now ask if there are species characteristics that correspond to maximum values of gain and loss rates. While the Energetic Equivalence principle is debated, it implies that the total consumption rate per individual should be inversely proportional to the equilibrium abundance N_i^* of that species. Combined with Damuth's rule, i.e. $N_i^* \propto M_i^{-3/4}$, we get that $\dot{D}_i \propto M_i^{3/4}$ where M is the mass of an average individual for that species. The bound would be maximized by species with the biggest-sized individuals by virtue of maximizing $\dot{D}_i$, more so if they are large predators which only have purely consumptive interactions. The effect is further accentuated by variation in the diagonal elements. The bound for the matrix with heterogeneous diagonal elements as in the Jacobian 3.23 is:

$$Re[\lambda(J_n)] \leq \big(\frac{1}{2}Max[\dot{D}_i + \dot{E}_i] - \dot{S}_i\big) \tag{3.29}$$

$\dot{S}_i$ depends on N_i^*. The biggest-sized species would minimize $\dot{S}_i$ via their abundance as a consequence of Damuth's law. By the combined effect of maximizing $\dot{D}$ (or the radius of the Gershgorin disk) and minimizing $\dot{S}_i$ (or bringing the centre of the disk closer to 0), the largest species are expected to play a major role in determining stability.

In figure 3.15, we see that bounds with the full Jacobian matrix are tighter than those using only the interaction matrix. This is expected because the full Jacobian contains abundances which are key to constraints on the net inter-species interactions 3.29. The general bound 3.20 also lies closer to the eigenvalues as compared to just the Gershgorin bound 3.18.

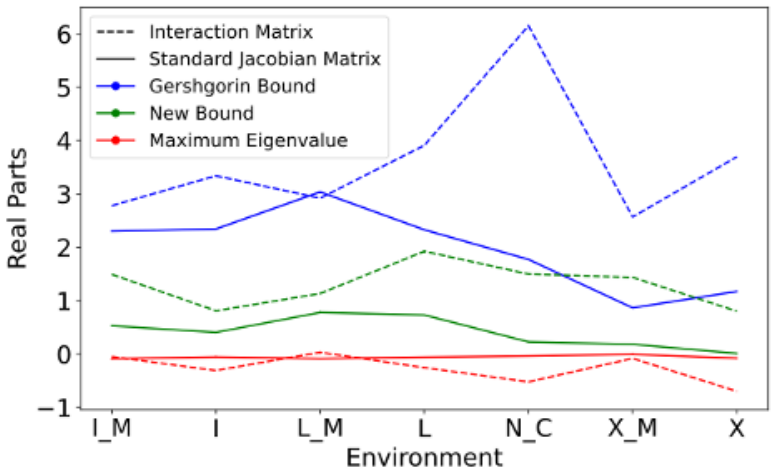

Figure 3.15: Calculated eigenvalues and geometric bounds based on gLV parameters measured from stable microbial systems in the human gut [125]. Notice that our new bound (green) always does a better job than the Gershgorin bound (blue) and places a fairly strong constraint on the actual calculated maximum eigenvalues (red). Moreover, bounds based on the standard Jacobian matrix (solid lines) are tighter than bounds based on the interaction matrix (dashed lines) used by May [28] and others. The labels at the bottom of the x-axis correspond to the particular environment and diet: I_M=Inulin + monosaccharide, I = Inulin, L=Laminarian, L_M=Laminarian + monosaccharide, N_C=No Carbohydrate (glycerol stock in ABB media), X=Xylan, and X_M=Xylan + monosaccharide.

Discussion and Conclusion

In this study, we found a mathematical bound on the stability of ecosystems whose form is ecologically interpretable and it is applicable to any type of interaction matrix. The bounds were derived using simple results from linear algebra Gershgorin circle theorem and the Bendixson-Bromwich inequalities. The resulting expression for the bound is also geometrically intuitive since one can interpret the terms as the radius and centre of the Gershgorin disks. A simple way of guaranteeing stability is to have small radii of the disks and making the centres more negative. These requirements directly translate to keeping the net interactions small and making the self-interactions/regulation more negative. From the bound 3.29, one can see that if the summed net interspecific interaction strength is smaller than the self-regulation term $\dot{S}_i$

for every species, then stability would certainly follow. This can be thought of as a condition for multi-species coexistence since the underlying Jacobian 3.23 assumes a feasible fixed-point.

We also relaxed a wide range of restrictive assumptions that are routinely invoked in studies on mathematical bounds on stability. This is captured by the matrix 3.23, which allows for dependence on species abundances as well as heterogeneity in the self-interaction terms. To our surprise, embracing this heterogeneity enabled simplifications giving the results 3.28 and 3.29 that are independent of the number of species while also being elegant and ecologically insightful.

These bounds imply that adding more number of species could still keep the system stable as long as the total consumption and loss rates are not very high. This could be achieved by having a fixed row-sum constraint that does not allow consumption to scale with the number of species. Overall our study provides a simple way to understand how real ecosystems could be both complex and stable.

3.5 Paper 5

This study uncovers the spatial patterns in genetic connectivity of the seagrass *Halodule uninervis* and compares it to the oceanographic dispersal barriers along the northwestern coast of Australia. One key objective is to understand the resilience of the contemporary genetic structure to catastrophic disturbances. To counter the limited knowledge of dispersal properties of this seagrass, the oceanographic dispersal barriers are investigated for a range of fragment dispersal durations and for many different timescales over which the genetic structure could persist. Seagrasses are important ecosystems that have a range of functions from being a food-source for threatened species such as dugongs [127] to sequestering carbon [128]. Interestingly, the study landscape encompasses a large area north of Shark Bay, which is a global stronghold of dugongs.

My contribution pertains to oceanographic dispersal barrier (ODB) analysis that is described below.

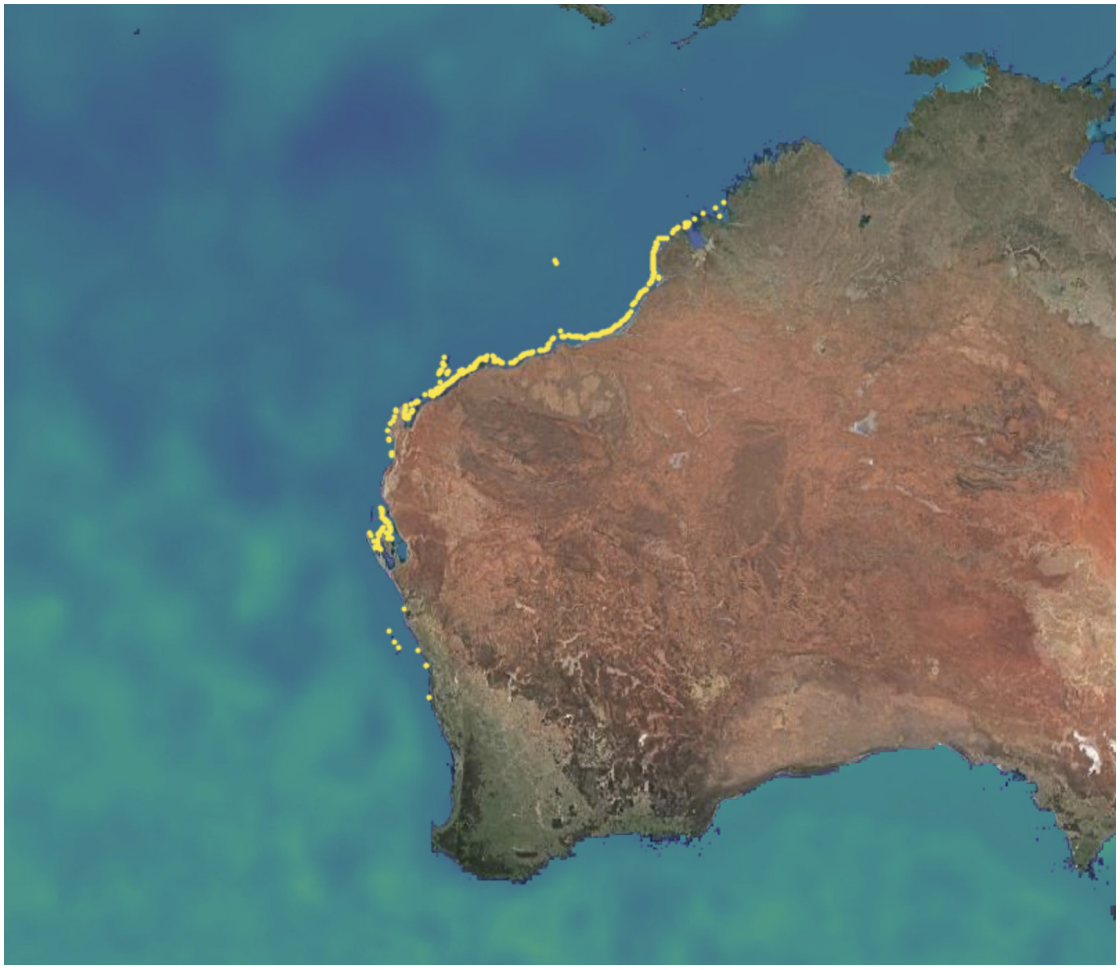

Figure 3.16: Seagrass population sites (yellow dots) along the north-western coast of Australia, which were analysed for dispersal barriers under different scenarios. A naive clustering of sites into metapopulations is possible based on geographical breaks in the distribution. However, this clustering is sub-optimal and might not capture the actual barriers to dispersal that is mediated by ocean currents and other biophysical factors. The shades of the ocean represent varying sea surface elevations that increase from lighter to darker shades. The map was generated using the HYCOM sea surface elevation data [126]

Methods

ODB analysis requires fragment dispersal matrices that are generated by simulating passive dispersal of vegetative fragments. Dispersal of biological species is sometimes modelled as a diffusion process but it is a much better approx-

imation of passive dispersal by media such as ocean currents. The generation process uses a biophysical oceanographic model [129] combining data on ocean currents [130], fragment viability and mortality rate [131]. Preliminary dispersal matrices are first obtained by calculating likelihood of dispersal from source to destination populations – i.e. between all pairs of sites – for fragment dispersal duration (FDD) of 2, 7, 14 and 28 days. These matrices were then projected forward in time (for 10, 50 and 100 years) using a recursive recipe that combines matrix operations incorporating birth, subsequent migration and growth to carrying capacity [132]. The resulting fragment dispersal matrices represent combinations of different FDDs and life-history based temporal mixing scenarios. Dispersal barriers that create almost disconnected subpopulations (i.e. clusters of sites) are identified using the algorithm defined in [29].

Oceanographic dispersal barrier (ODB) analysis

Identifying subpopulations broadly amounts to dividing sites into clusters such that sites across different clusters have much lower transition probabilities than sites within. We use a threshold parameter to determine whether two sites are sufficiently likely to allow dispersal between them and hence be included in the same subpopulation. As in [29], by sweeping over the threshold parameter and varying the penalty for clustering weakly connected sites together, many alternative sets of subpopulations can be found. These solutions are compared by calculating the overall 'leakage' between constituent subpopulations, to find the optimal aggregation.

Some additional heuristics are used to find optimal solutions for the study dataset. There is a critical transition probability corresponding to each solution, which equals the inverse of the threshold parameter. This is useful when two solutions have the same number of clusters and equal leakage – the better solution should have a higher critical transition probability, which amounts to stronger connections within the clusters. If the best dispersal scenario corresponds to zero leakage between clusters, then we try to identify barriers within the largest of those clusters by allowing 'minimal non-zero leakage' between potential sub-clusters. This places ODB analysis in line with the genetic methods since different subpopulations are never entirely disjunct genetically.

Results

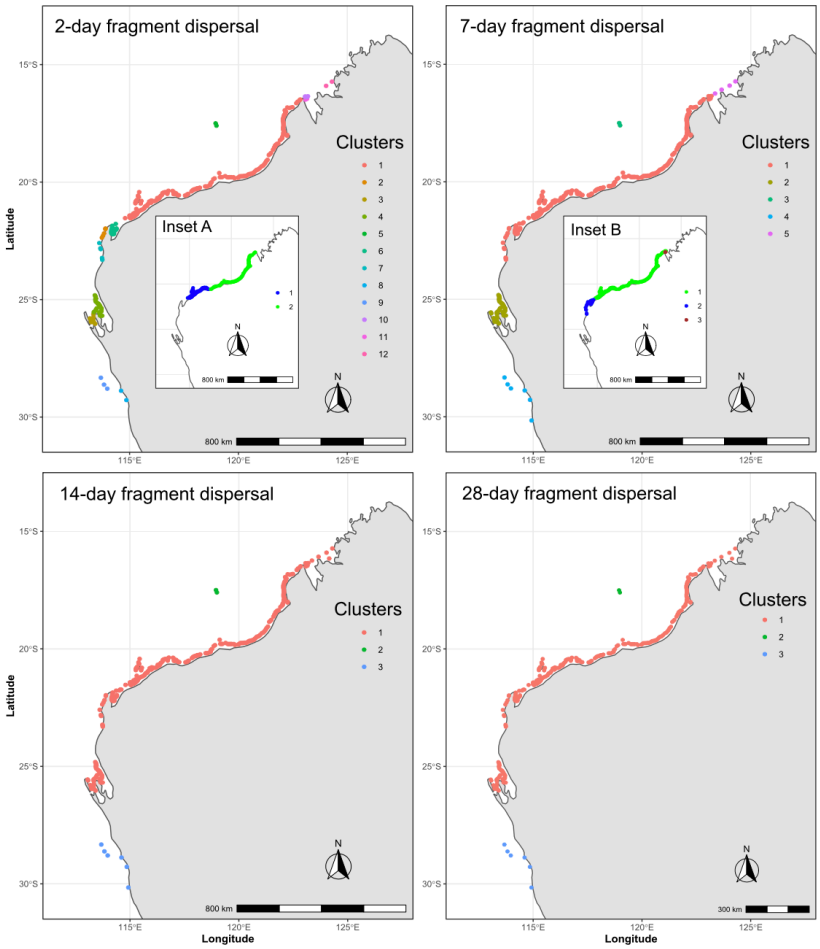

Figure 3.17: Results of the oceanographic barrier dispersal analysis for *H. uninervis* fragment dispersal based on basal modelling outputs from oceanographic modelling of 2-, 7-, 14- and 28-day fragment dispersal duration. Clusters indicate near zero leakage between identified groups. Inset A and B show clustering with minimal non-zero leakage for the base 2-day and the 7-day, respectively.

The genetic structure was best supported by the 2-day FDD model that identified 12 clusters with zero leakage that persisted up to the 100-year projection. The genetic results indicated an additional barrier (at Balla Balla) that was not captured by the ODB analysis. However, the largest cluster in the 2-day FDD split into two sub-clusters at this barrier when minimal non-

zero leakage was allowed. The 7-day FDD could also identify some important dispersal barriers but the number of predicted clusters declined greatly for the 14- and 28-day scenarios – 3 clusters each.

Discussion

The ODB analysis used a variation of the algorithm in [29], which is particularly suited for the case of zero leakage in the predicted clusterings. The genetic clustering best matched the results of the 2-day FDD especially with minimal non-zero leakage – but the 7-day model also managed to capture some key dispersal barriers. The likely fragment dispersal duration might lie somewhere between 2 to 7 days. Since dispersal barriers exist at smaller spatial scales at the edge of the distribution, conservation should be managed at these scales to prevent loss of genetic diversity. The results from two complementary approaches show good agreement despite utilizing very different types of data and analysis approaches. ODB analysis could therefore serve as a cheap alternative to collecting genetic samples, especially in large landscapes where extensive sampling requires considerable resources.

CHAPTER 4

Discussion, Outlook and Reflections

Thomas Kuhn in his article 'Objectivity, value judgment, and theory choice', described one of the characteristics of a good scientific theory as being broad scoped [133]. This means that a theory should have wider implications than the data it was initially designed to explain. This is a high standard that theoreticians revere even if they have a tunnel vision regarding problems or systems at the outset. An idea/theory that comes close to this ideal is the metabolic theory of ecology (MTE) which is fundamentally a model independent approach.

The rich tapestry of life unfolds at various levels of organization and across vastly different time scales. Using metabolic rate as a currency, MTE showed that individual level effects of temperature and body mass can be scaled up to study the analogous effects at the population or ecosystem level (section 2.7). In a similar fashion, cellular level metabolic rates and allometric relationships can be related to whole-organism metabolic rates. In fact these effects are not unidirectional in the levels of organization, and changes in whole-organism metabolic rates feedback on cellular size and metabolic rates [134] – findings which were much broader than the patterns and data that led to MTE. Models like gLV and consumer-resource models can also be used to study

sub-organismal processes. A very novel application is the modelling of self-organization of regulatory T-cells (tregs) that prevent autoimmune responses. Using a simple dynamical model that can be transformed to correspond with consumer-resource models, it was shown that tregs minimize niche overlap with each other, which is analogous to niche partitioning in ecological systems [135].

This thesis showed that ecological processes can generate a range of patterns at different levels of organization – species as well as the community level. With regards to dispersal, the constituent papers used this concept in very different ecological contexts.

Single species metapopulations have high dispersal probabilities for sites that lie within than across them. The coastal landscape we studied allowed for dispersal barriers that also indicate reduced gene-flow in an evolutionary sense. Paper 1 looked at complex ecological systems to understand the incidence of the island SAR. As discussed previously, a key ingredient that tipped the functional form of the relationship was dispersal of individuals from a mainland pool, i.e., demographic immigration. Mainland-island systems provided a simplified setting to study a spatial law (i.e. SAR) without explicitly invoking space. Paper 2 looked at the diametrically opposite case of contiguous landscapes by considering them as a patchwork of many habitats connected by high dispersal.

Interestingly, this drastic departure from an island-like configuration again facilitated a simplification with surprising results. High dispersal results in spatial coherence of species abundance densities, guaranteeing description of the system in terms of an effective interaction matrix. We used the properties of this matrix to get analytically exact results on stability and global species richness. The different cases converged on one result – spatial heterogeneity in interactions has a positive effect on species richness, and this effect becomes stronger as the number of habitats increase. While spatial coherence might be an idealized setting, it does provide a framework to study processes that could drive high diversity in large contiguous landscapes.

Examples of such landscapes include tropical lowland forests that support astoundingly many species of various types. Communities of species such as trees show very high local (i.e. α) diversity in these forests, which is not well understood [136, 137]. Evolutionary aspects might be key to solving these puzzles but the temporal sequence of dispersal and speciation is only recently

being uncovered.

A good example is the Amazon, where the absence of dispersal barriers means that dispersal and speciation are not trivially related. A recent phylogenetic study found evidence for widespread dispersal assembly of some diverse tree lineages [138]. Their proposal is to consider the entire Amazon basin as a metacommunity on evolutionary timescales.

Some interesting future avenues could be explored to understand these empirical findings. In particular, are speciation and widespread dispersal continuously in operation to generate high diversity? Or does speciation have a role only after local communities have been shaped by widespread dispersal? This does seem like an additional layer of complexity but stripping it away at the onset might deprive us of new patterns and simplifications that we constantly seek.

The tapestry of life is not static, but constantly evolving while still retaining traces of all that existed before. In this sense, its underlying fabric resembles a *palimpsest* – a manuscript whose earlier text has been washed off to accommodate new text while traces of the older writing still remain. One of the most striking examples of this is the Krakatau volcanic islands in the Sunda strait, Indonesia.

The island witnessed one of the strongest volcanic explosions ever recorded in 1883 which destroyed most of the island and the surrounding archipelago. What remained was just the southern part of the island called Rakata though most of the biodiversity was lost. Over the next century, many species dispersing from elsewhere recolonized the islands – not all that landed survived. This was of great interest to ecologists and evolutionary biologists since it provided a window to study ecosystem assembly from scratch [139]. By the mid 20th century, the numbers of some species groups stabilized. In fact, the islands contain almost the same number of bird species as islands of similar size in Java and Sumatra. While this is a tale of ecological succession and constant turnover of the species composition, the islands still exhibited macro scale patterns that were expected of an island of that size in the region. Many new species 'invaded' the island, but their fates were determined by the history of reassembly and other environmental factors. Such events raise many interesting questions about community assembly and invasions.

Are invasions more likely to fill certain niches or functional roles better than others? If species invasions homogenize a community [15], then how are

the environmental impacts on species traits and functional diversity altered? Large recolonization events are not exactly sequential because species keep arriving before any of the invasion events can converge to some equilibrium. There could still be a cohort of species that arrives and displaces an equilibrium of a recently established community. One can view invasions by many species as being sequential in species cohorts/groups. The failure of certain species to successfully invade could be more predictable than expected, given that only some kinds of species would first arrive in such ecosystems thus charting very specific assembly histories. It would be interesting to express stability bounds in terms of ecological constraints on invasions. Metrics like invasion fitness already represent simple constraints on effective growth rate of invaders arriving in low abundance.

While controlled experiments with microbes have provided great control on estimating parameters of those communities, there is much potential in using time series data from natural microbiomes such as the human gut to infer the underlying interaction strengths between species. Some very promising approaches in this regard have been recently proposed [86], so it would be interesting to compare the similarities between controlled and natural microbiomes and also understand how they differ.

Musings as a PhD researcher

My first reading as a PhD student was a book on complex networks that my supervisor handed to me [140]. Dabbling through the book and some related references, what grabbed my attention the most was that the social sciences have the longest tradition of quantitative research on real-world networks [141–144]. It was amusing to see the proliferation of ideas across disciplines, and how some ideas and representations lie dormant for decades before a breakout. Another very interesting example is that of ant-colony optimization algorithms. Marco Dorigo while working on his PhD in systems engineering was inspired by the ingenious work of Jean-Louis Deneoubourg and his colleagues on ant behaviour, specifically how ants use pheromone trails to find shortest paths to target locations [145]. The meta-heuristic algorithms that followed were directly inspired from the distributed intelligence of ants in solving search problems [146, 147].

I have gained much from being exposed to diverse academic traditions partly

because of my training, colleagues and interests. Statistical physics – although developed to study systems in the physical sciences – has developed many powerful frameworks and techniques that have been widely used in handling systems with many components. While none of these statistical physics models can capture the real-world complexity of large ecological systems, they can be employed to set baseline expectations by theoretically exploring various ecological mechanisms. This has helped explain some unexpected properties of systems such as hyper-diverse bacterial communities, which was earlier considered intractable because of the extremely high-dimensional parameter space [67, 148]. Not one size fits all, and therefore my work has found cases where these simple results cant be extrapolated to larger systems – many of which are not well-mixed but rather extended in space.

My PhD position is not tied to a specific grant or project, which has afforded me sufficient flexibility in framing and guiding my research. My research has moved from an extensive focus on simulations and modelling to shuttling between theory building and search for alternative representations of problems. Paper 5 was an ongoing project of my supervisor, and although I played a small role in the analysis and algorithm design, the results are expected to inform the planning of conservation reserves. My subsequent work has focused on abstract models to analyze the population dynamics of high-diversity ecological communities. At times, this entailed rethinking existing models to incorporate movement of species and the fact that species interactions could vary in space. These model extensions are inspired from ideas within statistical physics – where such spatial heterogeneity has been extensively studied – but not all ways of defining movement make biological sense. This ultimately made me realize that my findings apply better to certain types of species than the others – highlighting the context dependence of such complex adaptive systems.

My work in ecology and evolution draws from my interest in natural history. A major focus of natural history is detailed observations of individual organisms. Identification of individuals is of supreme importance, specifically placing their identity at not just the species but various other levels of organization. Classified by age and sex, individuals of the same species might look entirely different. In contrast, studies on macroecological patterns by definition focus on distributions and species agnostic metrics which have also been a characteristic of my work. Only recently I have reconciled the no-

tion of species identities and functional roles in my work on invasions in large ecological communities. Taking this view ties well to the question of what diversity is at a fundamental level, whether it is captured by numbers, forms or functions of species or their traits, and what representations can possibly meld these concepts and their inter-relations.

References

[1] F. Roy, M. Barbier, G. Biroli, and G. Bunin, "Complex interactions can create persistent fluctuations in high-diversity ecosystems," *PLoS computational biology*, vol. 16, no. 5, e1007827, 2020.

[2] M. T. Pearce, A. Agarwala, and D. S. Fisher, "Stabilization of extensive fine-scale diversity by ecologically driven spatiotemporal chaos," *Proceedings of the National Academy of Sciences*, vol. 117, no. 25, pp. 14 572–14 583, 2020.

[3] J. Hu, D. R. Amor, M. Barbier, G. Bunin, and J. Gore, "Emergent phases of ecological diversity and dynamics mapped in microcosms," *Science*, vol. 378, no. 6615, pp. 85–89, 2022.

[4] S. Drakare, J. J. Lennon, and H. Hillebrand, "The imprint of the geographical, evolutionary and ecological context on species–area relationships," *Ecology letters*, vol. 9, no. 2, pp. 215–227, 2006.

[5] F. W. Preston, "The canonical distribution of commonness and rarity: Part i," *Ecology*, vol. 43, no. 2, pp. 185–215, 1962.

[6] R. H. MacArthur and E. O. Wilson, "An equilibrium theory of insular zoogeography," *Evolution*, pp. 373–387, 1963.

[7] W. Dritschilo, H. Cornell, D. Nafus, and B. O'Connor, "Insular biogeography: Of mice and mites," *Science*, vol. 190, no. 4213, pp. 467–469, 1975.

[8] T. Bell *et al.*, "Larger islands house more bacterial taxa," *Science*, vol. 308, no. 5730, pp. 1884–1884, 2005.

[9] S. I. Glassman, K. C. Lubetkin, J. A. Chung, and T. D. Bruns, "The theory of island biogeography applies to ectomycorrhizal fungi in subalpine tree islands at a fine scale," *Ecosphere*, vol. 8, no. 2, e01677, 2017.

[10] J. P. O'Dwyer and R. Chisholm, "A mean field model for competition: From neutral ecology to the red queen," *Ecology letters*, vol. 17, no. 8, pp. 961–969, 2014.

[11] H. G. Andrewartha, L. C. Birch, *et al.*, *The distribution and abundance of animals*. University of Chicago press, 1954.

[12] R. Levins, "Some demographic and genetic consequences of environmental heterogeneity for biological control," *American Entomologist*, vol. 15, no. 3, pp. 237–240, 1969.

[13] M. A. Leibold *et al.*, "The metacommunity concept: A framework for multi-scale community ecology," *Ecology letters*, vol. 7, no. 7, pp. 601–613, 2004.

[14] M. Holyoak, M. A. Leibold, and R. D. Holt, *Spatial dynamics and ecological communities*, 2005.

[15] M. Dornelas *et al.*, "Assemblage time series reveal biodiversity change but not systematic loss," *Science*, vol. 344, no. 6181, pp. 296–299, 2014.

[16] J. B. Grace *et al.*, "Does species diversity limit productivity in natural grassland communities?" *Ecology letters*, vol. 10, no. 8, pp. 680–689, 2007.

[17] C. Roscher *et al.*, "Using plant functional traits to explain diversity–productivity relationships," *PloS one*, vol. 7, no. 5, e36760, 2012.

[18] B. J. Enquist *et al.*, "Scaling from traits to ecosystems: Developing a general trait driver theory via integrating trait-based and metabolic scaling theories," in *Advances in ecological research*, vol. 52, Elsevier, 2015, pp. 249–318.

[19] J. H. Brown, J. F. Gillooly, A. P. Allen, V. M. Savage, and G. B. West, "Toward a metabolic theory of ecology," *Ecology*, vol. 85, no. 7, pp. 1771–1789, 2004.

[20] M. Kleiber *et al.*, "Body size and metabolism," *Hilgardia*, vol. 6, no. 11, pp. 315–353, 1932.

[21] J. F. Gillooly, J. H. Brown, G. B. West, V. M. Savage, and E. L. Charnov, "Effects of size and temperature on metabolic rate," *science*, vol. 293, no. 5538, pp. 2248–2251, 2001.

[22] P. Yodzis and S. Innes, "Body size and consumer-resource dynamics," *The American Naturalist*, vol. 139, no. 6, pp. 1151–1175, 1992.

[23] U. Brose, R. J. Williams, and N. D. Martinez, "Allometric scaling enhances stability in complex food webs," *Ecology letters*, vol. 9, no. 11, pp. 1228–1236, 2006.

[24] R. Mac Arthur, "Species packing, and what competition minimizes," *Proceedings of the National Academy of Sciences*, vol. 64, no. 4, pp. 1369–1371, 1969.

[25] S. A. Levin, "Community equilibria and stability, and an extension of the competitive exclusion principle," *The American Naturalist*, vol. 104, no. 939, pp. 413–423, 1970.

[26] P. Chesson, "Quantifying and testing coexistence mechanisms arising from recruitment fluctuations," *Theoretical population biology*, vol. 64, no. 3, pp. 345–357, 2003.

[27] G. Barabás, R. D'Andrea, and S. M. Stump, "Chesson's coexistence theory," *Ecological Monographs*, vol. 88, no. 3, pp. 277–303, 2018.

[28] R. M. May, "Will a large complex system be stable?" *Nature*, vol. 238, no. 5364, pp. 413–414, 1972.

[29] M. N. Jacobi, C. André, K. Döös, and P. R. Jonsson, "Identification of subpopulations from connectivity matrices," *Ecography*, vol. 35, no. 11, pp. 1004–1016, 2012.

[30] M. Fiedler, "Algebraic connectivity of graphs," *Czechoslovak mathematical journal*, vol. 23, no. 2, pp. 298–305, 1973.

[31] M. E. Newman, "Finding community structure in networks using the eigenvectors of matrices," *Physical review E*, vol. 74, no. 3, p. 036 104, 2006.

[32] M. E. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical review E*, vol. 69, no. 2, p. 026 113, 2004.

[33] A. Clauset, M. E. Newman, and C. Moore, "Finding community structure in very large networks," *Physical review E*, vol. 70, no. 6, p. 066 111, 2004.

[34] M. N. Jacobi, "A robust spectral method for finding lumpings and meta stable states of non-reversible markov chains," *arXiv preprint arXiv:0810.1127*, 2008.

[35] S. P. Hubbell, "Neutral theory in community ecology and the hypothesis of functional equivalence," *Functional ecology*, vol. 19, no. 1, pp. 166–172, 2005.

[36] R. J. Williams and N. D. Martinez, "Simple rules yield complex food webs," *Nature*, vol. 404, no. 6774, pp. 180–183, 2000.

[37] R. P. Rohr, S. Saavedra, and J. Bascompte, "On the structural stability of mutualistic systems," *Science*, vol. 345, no. 6195, 2014.

[38] E. P. Wigner, "On the distribution of the roots of certain symmetric matrices," *Annals of Mathematics*, vol. 67, no. 2, pp. 325–327, 1958.

[39] J. Ginibre, "Statistical ensembles of complex, quaternion, and real matrices," *Journal of Mathematical Physics*, vol. 6, no. 3, pp. 440–449, 1965.

[40] M. Mehta, *Random matrices and the statistical theory of energy levels*, 1967.

[41] T. Tao and V. Vu, "Random matrices: The circular law," *Communications in Contemporary Mathematics*, vol. 10, no. 02, pp. 261–307, 2008.

[42] T. Tao, V. Vu, and M. Krishnapur, "Random matrices: Universality of esds and the circular law," 2010.

[43] V. Girko, "Elliptic law," *Theory of Probability & Its Applications*, vol. 30, no. 4, pp. 677–690, 1986.

[44] H. J. Sommers, A. Crisanti, H. Sompolinsky, and Y. Stein, "Spectrum of large random asymmetric matrices," *Physical review letters*, vol. 60, no. 19, p. 1895, 1988.

[45] S. Allesina and S. Tang, "Stability criteria for complex ecosystems," *Nature*, vol. 483, no. 7388, pp. 205–208, 2012.

[46] S. Tang and S. Allesina, "Reactivity and stability of large ecosystems," *Frontiers in Ecology and Evolution*, vol. 2, p. 21, 2014.

[47] B. S. Goh, "Global stability in many-species systems," *The American Naturalist*, vol. 111, no. 977, pp. 135–143, 1977.

[48] E. Kaszkurewicz and A. Bhaya, *Matrix diagonal stability in systems and computation*. Springer Science & Business Media, 2012.

[49] S. Persidskii, "Problem of absolute stability," *Automation and Remote Control*, vol. 12, pp. 1889–1895, 1969.

[50] B. Goh, "Nonvulnerability of ecosystems in unpredictable environments," *Theoretical Population Biology*, vol. 10, no. 1, pp. 83–95, 1976.

[51] A. A. Andronov, "Systems grossiers," in *Dokl. Acad. Nauk SSSR*, vol. 14, 1937, pp. 247–251.

[52] U. Bastolla, M. Lässig, S. C. Manrubia, and A. Valleriani, "Biodiversity in model ecosystems, i: Coexistence conditions for competing species," *Journal of theoretical biology*, vol. 235, no. 4, pp. 521–530, 2005.

[53] U. Bastolla, M. A. Fortuna, A. Pascual-García, A. Ferrera, B. Luque, and J. Bascompte, "The architecture of mutualistic networks minimizes competition and increases biodiversity," *Nature*, vol. 458, no. 7241, pp. 1018–1020, 2009.

[54] J. Grilli *et al.*, "Feasibility and coexistence of large ecological communities," *Nature communications*, vol. 8, no. 1, p. 14 389, 2017.

[55] S. Saavedra, R. P. Rohr, J. Bascompte, O. Godoy, N. J. Kraft, and J. M. Levine, "A structural approach for understanding multispecies coexistence," *Ecological Monographs*, vol. 87, no. 3, pp. 470–486, 2017.

[56] O. Arrhenius, "Species and area," *Journal of Ecology*, vol. 9, no. 1, pp. 95–99, 1921.

[57] H. A. Gleason, "On the relation between species and area," *Ecology*, vol. 3, no. 2, pp. 158–162, 1922.

[58] H. G. Martín and N. Goldenfeld, "On the origin and robustness of power-law species–area relationships in ecology," *Proceedings of the National Academy of Sciences*, vol. 103, no. 27, pp. 10 310–10 315, 2006.

[59] D. Gravel, F. Massol, and M. A. Leibold, "Stability and complexity in model meta-ecosystems," *Nature communications*, vol. 7, no. 1, pp. 1–8, 2016.

[60] B. Thomas, "On evolutionarily stable sets," *Journal of mathematical Biology*, vol. 22, pp. 105–115, 1985.

[61] J. Hofbauer, "On the occurrence of limit cycles in the volterra-lotka equation," *Nonlinear Analysis: Theory, Methods & Applications*, vol. 5, no. 9, pp. 1003–1007, 1981.

[62] S. Diederich and M. Opper, "Replicators with random interactions: A solvable model," *Physical Review A*, vol. 39, no. 8, p. 4333, 1989.

[63] M. Opper and S. Diederich, "Phase transition and 1/f noise in a game dynamical model," *Physical review letters*, vol. 69, no. 10, p. 1616, 1992.

[64] J. Hofbauer and K. Sigmund, *Evolutionary games and population dynamics*. Cambridge university press, 1998.

[65] K. Tokita, "Species abundance patterns in complex evolutionary dynamics," *Physical review letters*, vol. 93, no. 17, p. 178 102, 2004.

[66] Y. Yoshino, T. Galla, and K. Tokita, "Rank abundance relations in evolutionary dynamics of random replicators," *Physical Review E*, vol. 78, no. 3, p. 031 924, 2008.

[67] G. Bunin, "Ecological communities with lotka-volterra dynamics," *Physical Review E*, vol. 95, no. 4, p. 042 414, 2017.

[68] C. A. Serván, J. A. Capitán, J. Grilli, K. E. Morrison, and S. Allesina, "Coexistence of many species in random ecosystems," *Nature ecology & evolution*, vol. 2, no. 8, pp. 1237–1242, 2018.

[69] T. Wiegand *et al.*, "Consequences of spatial patterns for coexistence in species-rich plant communities," *Nature Ecology & Evolution*, vol. 5, no. 7, pp. 965–973, 2021.

[70] M. Uriarte *et al.*, "Trait similarity, shared ancestry and the structure of neighbourhood interactions in a subtropical wet forest: Implications for community assembly," *Ecology letters*, vol. 13, no. 12, pp. 1503–1514, 2010.

[71] G. Claireaux, C. Couturier, and A.-L. Groison, "Effect of temperature on maximum swimming speed and cost of transport in juvenile european sea bass (dicentrarchus labrax)," *Journal of Experimental Biology*, vol. 209, no. 17, pp. 3420–3428, 2006.

[72] C. R. White, L. A. Alton, T. S. Crispin, and L. G. Halsey, "Phylogenetic comparisons of pedestrian locomotion costs: Confirmations and new insights," *Ecology and evolution*, vol. 6, no. 18, pp. 6712–6720, 2016.

[73] W. Jetz, C. Carbone, J. Fulford, and J. H. Brown, "The scaling of animal space use," *Science*, vol. 306, no. 5694, pp. 266–268, 2004.

[74] A. M. Hemmingsen, "Energy metabolism as related to body size and respiratory surfaces, and its evolution," *Rep. Steno Mem. Hosp. Nord. Insulin Lab.*, vol. 8, pp. 1–110, 1960.

[75] T. Fenchel, "Intrinsic rate of natural increase: The relationship with body size," *Oecologia*, vol. 14, pp. 317–326, 1974.

[76] T. Southwood, R. May, M. Hassell, and G. Conway, "Ecological strategies and population parameters," *The American Naturalist*, vol. 108, no. 964, pp. 791–804, 1974.

[77] L. Blueweiss, H. Fox, V. Kudzma, D. Nakashima, R. Peters, and S. Sams, "Relationships between body size and some life history parameters," *Oecologia*, vol. 37, pp. 257–272, 1978.

[78] W. W. Hennemann, "Relationship among body mass, metabolic rate and the intrinsic rate of natural increase in mammals," *Oecologia*, vol. 56, pp. 104–108, 1983.

[79] V. M. Savage, J. F. Gillooly, J. H. Brown, G. B. West, and E. L. Charnov, "Effects of body size and temperature on population growth," *The American Naturalist*, vol. 163, no. 3, pp. 429–441, 2004.

[80] T. Norin and A. K. Gamperl, "Metabolic scaling of individuals vs. populations: Evidence for variation in scaling exponents at different hierarchical levels," *Functional ecology*, vol. 32, no. 2, pp. 379–388, 2018.

[81] C. L. Jerde, K. Kraskura, E. J. Eliason, S. R. Csik, A. C. Stier, and M. L. Taper, "Strong evidence for an intraspecific metabolic scaling coefficient near 0.89 in fish," *Frontiers in Physiology*, vol. 10, p. 456 930, 2019.

[82] C. R. White and R. S. Seymour, "Mammalian basal metabolic rate is proportional to body mass2/3," *Proceedings of the National Academy of Sciences*, vol. 100, no. 7, pp. 4046–4049, 2003.

[83] F. Bokma, "Evidence against universal metabolic allometry," *Functional Ecology*, pp. 184–187, 2004.

[84] D. S. Glazier, "A unifying explanation for diverse metabolic scaling in animals and plants," *Biological Reviews*, vol. 85, no. 1, pp. 111–138, 2010.

[85] J. Grilli, "Macroecological laws describe variation and diversity in microbial communities," *Nature communications*, vol. 11, no. 1, p. 4743, 2020.

[86] J. Camacho-Mateu, A. Lampo, M. Sireci, M. A. Muñoz, and J. A. Cuesta, "Sparse species interactions reproduce abundance correlation patterns in microbial communities," *Proceedings of the National Academy of Sciences*, vol. 121, no. 5, e2309575121, 2024.

[87] J. H. Brown and A. Kodric-Brown, "Turnover rates in insular biogeography: Effect of immigration on extinction," *Ecology*, vol. 58, no. 2, pp. 445–449, 1977.

[88] J. D. OSullivan, R. J. Knell, and A. G. Rossberg, "Metacommunity-scale biodiversity regulation and the self-organised emergence of macroecological patterns," *Ecology letters*, vol. 22, no. 9, pp. 1428–1438, 2019.

[89] J. M. Diamond, "Biogeographic kinetics: Estimation of relaxation times for avifaunas of southwest pacific islands," *Proceedings of the National Academy of Sciences*, vol. 69, no. 11, pp. 3199–3203, 1972.

[90] J. M. Diamond and E. Mayr, "Species-area relation for birds of the solomon archipelago," *Proceedings of the National Academy of Sciences*, vol. 73, no. 1, pp. 262–266, 1976.

[91] R. J. Whittaker *et al.*, "Functional biogeography of oceanic islands and the scaling of functional diversity in the azores," *Proceedings of the National Academy of Sciences*, vol. 111, no. 38, pp. 13 709–13 714, 2014.

[92] L. D. Gooriah, P. Davidar, and J. M. Chase, "Species–area relationships in the andaman and nicobar islands emerge because rarer species are disproportionately favored on larger islands," *Ecology and Evolution*, vol. 10, no. 14, pp. 7551–7559, 2020.

[93] L. D. Gooriah, P. Davidar, and J. M. Chase, *Data from: Species-area relationships in the andaman and nicobar islands emerge because rarer species are disproportionately favored on larger islands*, Dryad, Dataset, 2020.

[94] C. Loehle and W. Eschenbach, "Historical bird and terrestrial mammal extinction rates and causes," *Diversity and Distributions*, vol. 18, no. 1, pp. 84–91, 2012.

[95] D. R. Spatz *et al.*, "Globally threatened vertebrates on islands with invasive species," *Science advances*, vol. 3, no. 10, e1603080, 2017.

[96] O. Allouche, M. Kalyuzhny, G. Moreno-Rueda, M. Pizarro, and R. Kadmon, "Area–heterogeneity tradeoff and the diversity of ecological communities," *Proceedings of the National Academy of Sciences*, vol. 109, no. 43, pp. 17 495–17 500, 2012.

[97] E. Ben-Hur and R. Kadmon, "An experimental test of the area–heterogeneity tradeoff," *Proceedings of the National Academy of Sciences*, vol. 117, no. 9, pp. 4815–4822, 2020.

[98] A. M. Kramer and J. M. Drake, "Time to competitive exclusion," *Ecosphere*, vol. 5, no. 5, pp. 1–16, 2014.

[99] T. J. Case, "Invasion resistance arises in strongly interacting species-rich model competition communities.," *Proceedings of the National Academy of Sciences*, vol. 87, no. 24, pp. 9610–9614, 1990.

[100] J.-f. Arnoldi, M. Barbier, R. Kelly, G. Barabás, and A. L. Jackson, "Invasions of ecological communities: Hints of impacts in the invader's growth rate," *Methods in Ecology and Evolution*, vol. 13, no. 1, pp. 167–182, 2022.

[101] I. M. Bomze, "Cross entropy minimization in uninvadable states of complex populations," *Journal of Mathematical Biology*, vol. 30, pp. 73–87, 1991.

[102] C.-Y. Chang, D. Baji, J. C. Vila, S. Estrela, and A. Sanchez, "Emergent coexistence in multispecies microbial communities," *Science*, vol. 381, no. 6655, pp. 343–348, 2023.

[103] A. M. de Roos, "Dynamic population stage structure due to juvenile–adult asymmetry stabilizes complex ecological communities," *Proceedings of the National Academy of Sciences*, vol. 118, no. 21, e2023709118, 2021.

[104] C. Vagnon, R. P. Rohr, L.-F. Bersier, F. Cattanéo, J. Guillard, and V. Frossard, "Combining food web theory and population dynamics to assess the impact of invasive species," *Frontiers in Ecology and Evolution*, vol. 10, p. 913 954, 2022.

[105] P. Yodzis, "The indeterminacy of ecological interactions as perceived through perturbation experiments," *Ecology*, vol. 69, no. 2, pp. 508–515, 1988.

[106] P. A. Abrams, "Describing and quantifying interspecific interactions: A commentary on recent approaches," *Oikos*, vol. 94, no. 2, pp. 209–218, 2001.

[107] E. L. Berlow *et al.*, "Interaction strengths in food webs: Issues and opportunities," *Journal of animal ecology*, pp. 585–598, 2004.

[108] E. L. Berlow, J. A. Dunne, N. D. Martinez, P. B. Stark, R. J. Williams, and U. Brose, "Simple prediction of interaction strengths in complex food webs," *Proceedings of the National Academy of Sciences*, vol. 106, no. 1, pp. 187–191, 2009.

[109] F. Bourinet, O. Anneville, H. Drouineau, C. Goulon, J. Guillard, and A. Richard, "Synchrony in whitefish stock dynamics: Disentangling the effects of local drivers and climate.," *Journal of Limnology*, vol. 82, 2023.

[110] S. L. Pimm, "The complexity and stability of ecosystems," *Nature*, vol. 307, no. 5949, pp. 321–326, 1984.

[111] S. McNaughton, "Stability and diversity of ecological communities," *Nature*, vol. 274, no. 5668, pp. 251–253, 1978.

[112] P. Yodzis, "The stability of real ecosystems," *Nature*, vol. 289, no. 5799, pp. 674–676, 1981.

[113] D. Tilman, "Biodiversity: Population versus ecosystem stability," *Ecology*, vol. 77, no. 2, pp. 350–363, 1996.

[114] T. Gross, L. Rudolf, S. A. Levin, and U. Dieckmann, "Generalized models reveal stabilizing factors in food webs," *Science*, vol. 325, no. 5941, pp. 747–750, 2009.

[115] G. Barabás, M. J. Michalska-Smith, and S. Allesina, "Self-regulation and the stability of large ecological networks," *Nature ecology & evolution*, vol. 1, no. 12, pp. 1870–1875, 2017.

[116] S. Allesina and S. Tang, "The stability–complexity relationship at age 40: A random matrix perspective," *Population Ecology*, vol. 57, no. 1, pp. 63–75, 2015.

[117] S. Butler and J. P. ODwyer, "Stability criteria for complex microbial communities," *Nature communications*, vol. 9, no. 1, p. 2970, 2018.

[118] S. Butler and J. P. ODwyer, "Cooperation and stability for complex systems in resource-limited environments," *Theoretical ecology*, vol. 13, pp. 239–250, 2020.

[119] T. Gibbs, Y. Zhang, Z. R. Miller, and J. P. ODwyer, "Stability criteria for the consumption and exchange of essential resources," *PLoS computational biology*, vol. 18, no. 9, e1010521, 2022.

[120] J. Damuth, "Population density and body size in mammals," *Nature*, vol. 290, no. 5808, pp. 699–700, 1981.

[121] I. Bendixson, "Sur les racines d'une équation fondamentale," 1902.

[122] L. Mirsky, *An introduction to linear algebra*. Courier Corporation, 2012.

[123] W. A. Calder, *Size, function, and life history*. Courier Corporation, 1996.

[124] R. H. Peters, *The ecological implications of body size*. Cambridge university press, 1986, vol. 2.

[125] E. Ostrem Loss, J. Thompson, P. L. K. Cheung, Y. Qian, and O. S. Venturelli, "Carbohydrate complexity limits microbial growth and reduces the sensitivity of human gut communities to perturbations," *Nature ecology & evolution*, vol. 7, no. 1, pp. 127–142, 2023.

[126] J. A. Cummings and O. M. Smedstad, "Variational data assimilation for the global ocean," in *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications (Vol. II)*, Springer, 2013, pp. 303–343.

[127] H. Marsh, A. Grech, and K. McMahon, "Dugongs: Seagrass community specialists," in *Seagrasses of Australia*, Springer, 2018, pp. 629–661.

[128] J. W. Fourqurean *et al.*, "Seagrass ecosystems as a globally significant carbon stock," *Nature geoscience*, vol. 5, no. 7, pp. 505–509, 2012.

[129] J. T. Kool and S. L. Nichol, "Four-dimensional connectivity modelling with application to australia's north and northwest marine environments," *Environmental Modelling & Software*, vol. 65, pp. 67–78, 2015.

[130] E. P. Chassignet *et al.*, "The hycom (hybrid coordinate ocean model) data assimilative system," *Journal of Marine Systems*, vol. 65, no. 1-4, pp. 60–83, 2007.

[131] L. M. Hall, M. D. Hanisak, and R. W. Virnstein, "Fragments of the seagrasses halodule wrightii and halophila johnsonii as potential recruits in indian river lagoon, florida," *Marine Ecology Progress Series*, vol. 310, pp. 109–117, 2006.

[132] J. T. Kool, C. B. Paris, S. Andréfouët, and R. K. Cowen, "Complex migration and the development of genetic structure in subdivided populations: An example from caribbean coral reef ecosystems," *Ecography*, vol. 33, no. 3, pp. 597–606, 2010.

[133] T. S. Kuhn, "Objectivity, value judgment, and theory choice," in *Review of Thomas S. Kuhn The Essential Tension: Selected Studies in Scientific Tradition and Change*, D. Zaret, Ed., Duke University Press, 1981, pp. 320–39.

[134] V. M. Savage *et al.*, "Scaling of number, size, and metabolic rate of cells with body size in mammals," *Proceedings of the National Academy of Sciences*, vol. 104, no. 11, pp. 4718–4723, 2007.

[135] R. Marsland III, O. Howell, A. Mayer, and P. Mehta, "Tregs self-organize into a computing ecosystem and implement a sophisticated optimization algorithm for mediating immune response," *Proceedings of the National Academy of Sciences*, vol. 118, no. 1, e2011709118, 2021.

[136] R. Valencia, H. Balslev, and G. P. Y. Miño, "High tree alpha-diversity in amazonian ecuador," *Biodiversity & Conservation*, vol. 3, no. 1, pp. 21–28, 1994.

[137] E. G. Leigh Jr *et al.*, "Why do some tropical forests have so many species of trees?" *Biotropica*, vol. 36, no. 4, pp. 447–473, 2004.

[138] K. G. Dexter *et al.*, "Dispersal assembly of rain forest tree communities across the amazon basin," *Proceedings of the National Academy of Sciences*, vol. 114, no. 10, pp. 2645–2650, 2017.

[139] E. O. Wilson, *The diversity of life*. WW Norton & Company, 1999.

[140] M. Newman, A.-L. Barabási, and D. J. Watts, *The structure and dynamics of networks*. Princeton university press, 2011.

[141] J. L. Moreno, "Who shall survive? a new approach to the problem of human interrelations," 1934.

[142] A. Davis, B. B. Gardner, and M. R. Gardner, *Deep South: A social anthropological study of caste and class*. Univ of South Carolina Press, 2009.

[143] F. J. Roethlisberger and W. J. Dickson, *Management and the Worker*. Psychology press, 2003, vol. 5.

[144] M. E. Newman, "The structure and function of complex networks," *SIAM review*, vol. 45, no. 2, pp. 167–256, 2003.

[145] S. Goss, S. Aron, J.-L. Deneubourg, and J. M. Pasteels, "Self-organized shortcuts in the argentine ant," *Naturwissenschaften*, vol. 76, no. 12, pp. 579–581, 1989.

[146] M. Dorigo, "Positive feedback as a search strategy," *Technical report*, pp. 91–16, 1991.

[147] M. Dorigo, V. Maniezzo, and A. Colorni, "Ant system: Optimization by a colony of cooperating agents," *IEEE transactions on systems, man, and cybernetics, part b (cybernetics)*, vol. 26, no. 1, pp. 29–41, 1996.

[148] J. E. Goldford *et al.*, "Emergent simplicity in microbial community assembly," *Science*, vol. 361, no. 6401, pp. 469–474, 2018.