Les Houches Lectures on Community Ecology: From Niche Theory to Statistical Mechanics

Wenping Cui [1], Robert Marsland III [2], and Pankaj Mehta [3⋆]

1 Kavli Institute for Theoretical Physics, University of California Santa Barbara
2 Pontifical University of the Holy Cross
3 Dept. of Physics and Faculty of Computing and Data Science, Boston University
⋆ pankajm@bu.edu

November 2, 2026

arXiv:2403.05497v2 [q-bio.PE] 30 Oct 2026

Abstract

Ecosystems are among the most interesting and well-studied examples of self-organized complex systems. Community ecology, the study of how species interact with each other and the environment, has a rich tradition. Over the last few years, there has been a growing theoretical and experimental interest in these problems from the physics and quantitative biology communities. Here, we give an overview of community ecology, highlighting the deep connections between ecology and statistical physics. We start by introducing the two classes of mathematical models that have served as the workhorses of community ecology: Consumer Resource Models (CRM) and the generalized Lotka-Volterra models (GLV). We place a special emphasis on graphical methods and general principles. We then review recent works showing a deep and surprising connection between ecological dynamics and constrained optimization. We then shift our focus by analyzing these same models in "high-dimensions" (i.e. in the limit where the number of species and resources in the ecosystem becomes large) and discuss how such complex ecosystems can be analyzed using methods from the statistical physics of disordered systems such as the cavity method and Random Matrix Theory.

Contents

1 Introduction **2**

2 Fundamental questions **5**

3 Ecological Processes **5**

4 An introduction to consumer resource models **7**
4.1 MacArthur's Consumer Resource Model 8
4.2 Externally supplied resources 10
4.3 Essential resources 11

5 Generalized Lotka-Volterra Theory **12**
5.1 Deriving the GLV Equations 12
5.2 Invasion analysis and Modern Coexistence Theory 13
5.3 MacArthur's Consumer Resource Model 14
5.4 Externally supplied resources 16

6 Niche theory and optimization **16**
6.1 MacArthur's Minimization Principle 17
6.2 MacArthur's Consumer Resource Model 18
6.3 Externally supplied resources 19
6.4 Essential resources 20

7 Ecology in high-dimensions **21**
7.1 Making the leap to high-dimensions 21
7.2 May's Stability Criteria and Random Matrix Theory 22

8 Cavity Method for Lotka–Volterra model **23**
8.1 The Cavity Solution 25
8.2 Beyond Replica Symmetry 29

9 Cavity Method for MacArthur Consumer-Resource Models **31**
9.1 Self-consistency equations for species 33
9.2 Self-consistency equations for resource 35

10 Conclusions **37**

11 Acknowledgments **38**

A Consumer resource models **38**
A.1 Substitutable Resources 39
A.2 Nonsubstitutable Resources 39
A.3 Mass action 39
A.4 Marginal benefit foraging 39
A.5 Self-renewing resources 40
A.6 Externally supplied resources 40
A.7 Explicit equations for all models 40
A.7.1 Substitutable, Mass action, Self-renewing 40
A.7.2 Substitutable, Mass action, Externally supplied 40
A.7.3 Nonsubstitutable, Mass action, Self-renewing 40
A.7.4 Nonsubstitutable, Mass action, Externally supplied 41
A.7.5 Substitutable, Marginal benefit, Self-renewing 41
A.7.6 Substitutable, Marginal benefit, Externally supplied 41
A.7.7 Nonsubstitutable, Marginal benefit, Self-renewing 41
A.7.8 Nonsubstitutable, Marginal benefit, Externally supplied 41

B Self-limitation and stability **42**

References **43**

1 Introduction

Life is everywhere on our planet. No matter where one looks, from the boiling hot springs of Yellowstone to the glaciers of Antartica, one can find thriving ecosystems. Despite the enormous variation in temperature, energy supplies, and physical resources, in all these settings

organisms interact with each other and their environments to self-organize and survive. Understanding this process is one of the fundamental goals of ecology.

Every ecosystem is different and shaped by the physical environments in which it functions. The organisms that can be found in the desert, where water is a scarce and precious resource, are very different from the kind of organisms that thrive in the rain forest where water is abundant. At a completely different scale, one finds microbial ecosystems – complex assemblages of microbes that metabolize and produce small molecules. Furthermore, far from being passive creatures in a fixed environment, organisms almost always transform the environments in which they live. The most dramatic example of such a transformation is the great oxidation event two billion years ago by the photosynthetic *cynobacteria*.

Understanding how ecosystems can assemble and function in light of all this complexity and difference has been a central strand of ecological research for the last century. These efforts have birthed an extremely rich theoretical and experimental tradition aimed at identifying the principles underlying community ecology. A central result of this work is the realization that despite all the innumerable differences between ecosystems – very different physical environments, the great heterogeneity in organism physiologies and traits – there exist a common set of core principles that fundamentally shape community assembly and function. In the language of physics, this is the statement ecosystems are shaped by *universal* processes.

These lectures introduce and build upon one of the most successful and rich theoretical paradigms for understanding ecosystems: niche-theory. As discussed below, niche-theory emphasizes the central role played by ecological competition and selection in shaping communities. Many of the ideas of niche theory have their origins in seminal papers from the 1960's by Richard Levins and Robert MacArthur that introduced many of the core ideas and mathematical models underlying niche theory [1–4]. Since then, niche theory has been further developed and expanded upon by many ecologists giving rise to an extremely sophisticated and coherent framework for understanding community ecology. These ideas have been summarized in a number of wonderful books [5–7]. For this reason, we are neither capable of, nor have we attempted to, give a comprehensive introduction to the ecological literature.

Instead, the focus of these lectures is on reformulating these ideas with the purpose of presenting them in a language and style more accessible to physicists. This process has often resulted in alternative derivations/formulations of key ideas. Throughout the review, we have tried our best to retain what we consider the best aspects of the theoretical ecology tradition - an emphasis on the use of graphical methods to produce robust theorems using realistic and generalizable models (see Levins' wonderful essay "The role of Model Building in Population Biology" [8]) – and combine this with the best of the physics pedagogical tradition.

The timing of the lectures is inspired by the exciting and fruitful developments over the last decade in the physics of living systems. This period has seen a renewed interest in niche-theory and community ecology from the viewpoint of statistical physics. These works have sought to extend and integrate many of the ideas of niche theory with techniques and tools from statistical physics, leading to rapid progress and a plethora of new insights. However, this literature is often scattered across journals in many different fields (ecology, physics, microbial ecology, biology) making it difficult to formulate a coherent picture. For this reason we have made a concerted attempt to provide a concise yet coherent introduction to these ideas.

This renewed interest in ecology among physicists has been driven by big experimental advances. This has been especially true of in the microbial realm, where the adoption of quantitative experimental designs centered around the rapid progress in DNA sequencing has allowed for the ability to catalogue the species present in a microbial ecosystem on a scale and resolution that was unimaginable just twenty years ago. This wealth of data has revealed that microbial communities are extremely diverse with tens to thousands of different strains/species present in most ecosystems. This has raised a number of new and fundamen-

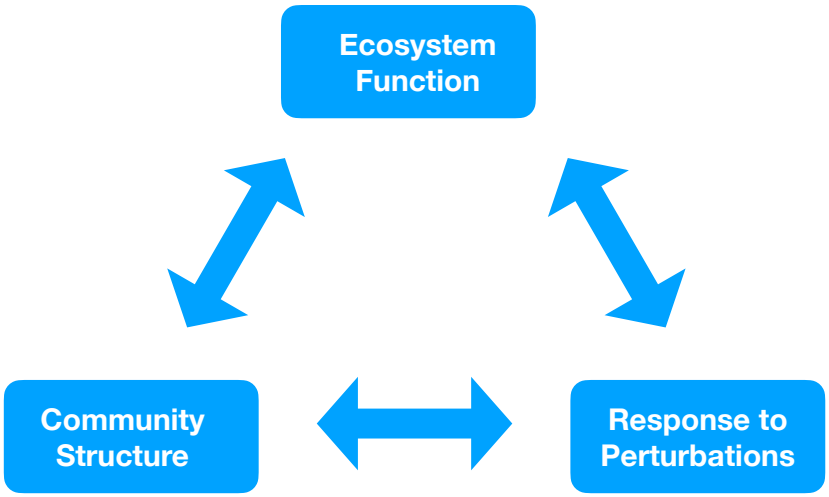


Figure 1: The three core problems and their interrelations.

tal questions about the diversity, function, and structure of microbial ecosystems. To answers these questions, physicists and ecologists have also started to design quantitative experiments that probe how the physical and chemical environment shapes the structure and function of microbial communities [9–18]. One of the fundamental insights that has resulted from these experiments are that microbes fundamentally transform the environments in which they live – suggesting that any theory of microbial ecosystems must account for environmental engineering and the strong coupling between organism and environment [19, 20].

The immense diversity seen in real ecosystems combined with the success of alternative frameworks for understanding diversity seen in ecosystems such as neutral theory (see [21] and [22] for review directed at statistical physicists) has also resulted in a renewed interest understanding how niche theory can be adapted to diverse ecosystems with many species and resources. This has led to a rich body of recent works exploring niche-theory in the "high-dimensional" limit. Much of this work draws inspiration from the rich literature in the statistical physics of disordered systems and random matrix theory.

Unfortunately, there are many interesting recent research directions that we have been unable to cover here. These include new graphical methods that build upon those presented here [23, 24] , the very impressive new work on dynamics and chaos in high-dimensional ecosystems (including dynamical mean-field theories) [25–31], the extension of consumer resource models to the microbial ecosystems by including metabolism and metabolic cross-feeding [12, 18, 32–35], and macroecological laws [36]. It is our hope that these lectures can serve as a starting point for readers interested in exploring this literature.

Finally, we note in line with the tradition of the Les Houches lectures we have kept citations to a minimum. We apologize for the many works we have inevitably failed to cite. We wish to emphasize that these lectures reflect the *personal perspective* of the authors (Cui, Marsland, Mehta), developed over the last few years working in ecology. In no way should these lectures be taken to be a comprehensive or complete overview of the field. In line with this, many of the derivations of classic results presented in the low-dimensional chapter are new to the literature. However, we have found these derivations to be more intuitive and less mathematical and for this reason have included these in the lectures.

2 Fundamental questions

Many works in community ecology are concerned with three inter-related problems. The first of these is to understand and predict *community structure*. Community structure is a very broad concept that refers to any property of the communities that reside in an ecosystem [37–40]. These include the *richness* of an ecosystem –the number of different species in a community , the *diversity* of a community – measures of species biodiversity that also accounts for differences in species abundances (e.g. Shannon entropy of species abundances in an ecosystem), and even the *predicability* of community assembly - will you end up with the same community in similar environments?

The second major problem in community ecology is to understand *ecosystem function* [41–44]. Ecosystem function is usually defined as the physiochemical and biological processes that occur in an ecosystem that are necessary for maintaining life [45]. Examples including processes that create biomass and energy, material transformations necessary for ecosystems to thrive (e.g nutrient recycling, biomass production), and often is used to to denote just about any property of an ecosystem that can help sustain human life. This vagueness in definition often leads to numerous debates, controversies, and disagreements. Chief among these is whether the notion of function is something imposed externally by humans or is it an internal property of the ecosystem itself. In our experience, in the literature both of these meanings of ecosystem function are commonly employed.

The third major reoccurring problem addressed in community ecology is understanding how ecosystem properties (including community structure and ecological function) respond to perturbations [46, 47]. Examples of the kind of perturbations that are of interest include changes to the physical environment (e.g. warming of temperatures), the response of the ecosystem to immigration/invasion by new organisms, or the effect temporal disturbances. These questions are often motivated by important practical questions related to the preservation and management of environments or human health (e.g. the effects of antibiotics on the gut *microbiome* - the collection of microbes that live inside the human gut).

A major goal of modern ecological research is to relate these three concepts [48]. For example, a large line of recent research has been concerned with asking how biodiversity affects ecological function [45, 49]. Another interesting line of research seeks to understand how the structure of food webs shapes community structure and stability in response to perturbations [50].

Finally, we note that while many of these questions and ideas were originally formulated for macroscopic ecosystems such as forests or coral reefs, the last few years have seen an increased interest in using these ideas to understand microbial ecosystems [51–54]. Microbes offer a particularly nice and tractable setting for exploring these problems since it is possible to directly measure community structure using DNA sequencing and the small size of these ecosystems makes it possible to tightly control the physical and biochemical environments.

3 Ecological Processes

In order to situate niche theory in the broader context of community ecology, it is worth briefly considering a very general model of community ecology that incorporates the four major ecological processes thought to be important for ecosystems: (i) *ecological selection* - the idea that organisms that are more suited to an environment will survive in an ecosystem; (ii) *stochasticity* - processes such as demographic noise that introduce randomness; (iii) *dispersal* - species live in local communities and can immigrate between these communities; and finally (iv) *speciation* - the process through which new species can emerge from evolution. In what follows, we

will always focus on timescales over which speciation is negligible (although it can be included as a special kind of "dispersal" from an imaginary pool of possible phenotypes).

A very general set of stylized equations that incorporates the first three of these processes is illustrated in Fig. 2*(a)* and consists of a set of local patches, $a, b, c, \ldots$, a set of species labeled by latin letters i whose population in patch a is given by N_{ia}, and local environmental variables labeled by greek letters α whose value in patch a is given by $E_{\alpha a}$. Within each patch, the dynamics are described by the following set of differential equations:

$$\frac{dN_{ia}}{dt} = N_{ia} g_i(\mathbf{N}_a, \mathbf{E}_a) + \sum_b \lambda^i_{ab} N_{ib} + \sqrt{D_i N_{ia}}\xi_{ia}(t) \tag{1}$$

$$\frac{dE_{\alpha a}}{dt} = h_{\alpha a}(\mathbf{E}_a, t) + \sum_j N_{ja} q_{j\alpha}(\mathbf{N}_a, \mathbf{E}_a) \tag{2}$$

where λ^i_{ab} is the migration rate of species i from patch b to patch a, and $\xi_{ia}(t)$ is Gaussian white noise.

The reader should not dwell too much on the details of Equation (2). This equation is too general to be amenable to mathematical analysis. Instead, the goal of writing Eq. (2) is to get a sense of how one can represent the ecological processes discussed above mathematically. Notice that the g_i contain the information about ecological *selection* since these functions encode how well a species will grow. The dependence of g_i on $\mathbf{N}$ and $\mathbf{E}$ is the mathematical manifestation of the fact that ecological selection is inseparable from interactions with other species and the shared environment. The λ^i_{ab} term represents *dispersal*. This terms controls how species immigrate between patches. The noise term $\sqrt{D_i N_i}\xi_i(t)$ models the effects of demographic *stochasticity*. The vector $h_{\alpha a}$ represents the intrinsic behavior of the environment in the absence of other organisms, and is known as the *supply vector* in contemporary niche theory, where it usually refers to an external supply of consumable resources. The impact of the organisms on the environment is captured by the second term, where the per-capita impact of each species j on the environment is measured by an *impact vector* $q_{j\alpha}$. This impact vector is the mathematical manifestation of the fact that species fundamentally modify the environments in the which they live.

The major branches of ecological theory shown in Fig. 2 each emphasize different aspects of this framework. *Consumer Resource Models* (CRMs) - the main model underlying contemporary niche theory – and generalized *Lotka-Volterra* (GLV) models – the main models underlying modern coexistence theory – focus on interactions with the environment and with other species, respectively. *Neutral theory* considers a different limit, where selective differences among species are absent, and species distributions are driven by drift and dispersal. Despite the differences of focus, these models have some overlap: interactions with the environment in CRMs produce effective interactions between species and thus can be implicitly included in a GLV model. Both of these models reduce to neutral theory in the limit of vanishing selective differences ($g_i = g$ for all i). Recent work from statistical physics literature also suggests that niche and neutral theory are actually better thought of as theories for two different "ecological phases" – akin to a ferromagnet and paramagnet in spin systems – rather than complete theories themselves [55, 56].

In the following sections, we give an overview of niche theory. Niche theory neglect drifts and dispersal and focus on the effects of selection. To build ecological intuitions, we start by considering "low-dimensional" ecosystems with 2-3 species and 2 environmental factors. We focus on the two most common types of mathematical models: generalized consumer resource models (CRMs) and generalized Lotka-Volterra models (GLVs). Whereas CRMs explicitly model the environment, GLVs model the environment implicitly through its effect on species-species interactions. Having build our ecological intuition in this low-dimensional set-

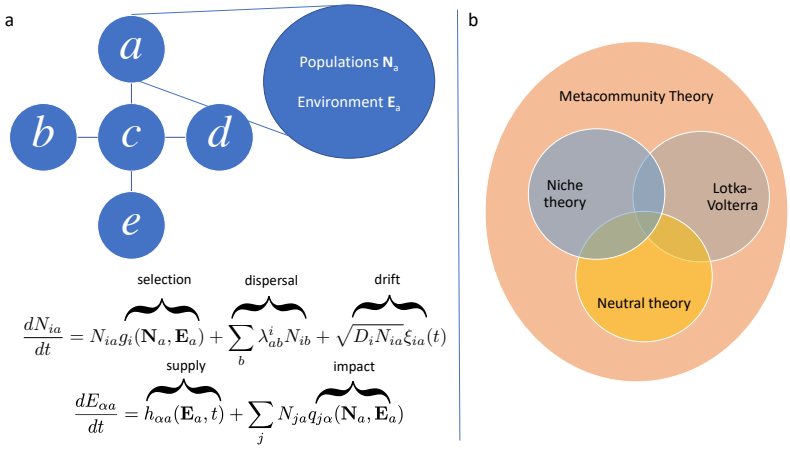


Figure 2: **The metacommunity framework.** *(a)* The metacommunity framework incorporates the three fundamental ecological processes of selection, dispersal and drift within a unified mathematical framework. *(b)* Within this framework, three areas of intense study have been niche theory, generalized Lotka-Volterra theory and neutral theory. The domains of applicability for these theories partially overlap.

ting, we then consider "high-dimensional" ecology where both the number of species and resources are assumed to be large. In this setting, we show how ideas from the statistical physics of disordered systems can be adapted to ecology in order to understand complex ecosystems.

4 An introduction to consumer resource models

One of the most influential class of ecological models in theoretical ecology are consumer resource models (CRMs). CRMs assume that all interactions between species are mediated by the consumption of external resources. Mathematically, this means that the g_i and $q_{j\alpha}$ in Eq. (1-2) are functions of $\mathbf{E}$ alone, and do not depend directly on the population sizes $\mathbf{N}$. The only environmental factors considered in CRMs are consumable resources. For this reason, we use the more familiar notation R_α instead of E_α, where each R_α represents the abundance of one type of resource in the local patch. As noted above, CRMs neglect drift and dispersal focusing on selection. With these restrictions, we obtain the following set of differential equations for the dynamics within a single patch:

$$\frac{dN_i}{dt} = N_i g_i(\mathbf{R}) \tag{3}$$

$$\frac{dR_\alpha}{dt} = h_\alpha(\mathbf{R}) + \sum_j N_j q_{j\alpha}(\mathbf{R}). \tag{4}$$

The elimination of direct interactions between species gives this class of models a bipartite structure, which is amenable to graphical representation, especially for ecosystem with two resources. Fig. 3 shows how the growth rate, impact vectors and supply can be represented in the two-dimensional space of resource abundances. The set of environmental states $\mathbf{R}$ where $g_i(\mathbf{R}) = 0$ comprise the zero net growth isocline (ZNGI), a line that separates states of negative growth rate from states of positive growth rate for each species. The impact $q_{i\alpha}(\mathbf{R})$ for each species can be represented as a vector field in this space, as can the supply vector $h_\alpha(\mathbf{R})$.

This representation immediately leads to geometric criteria for identifying fixed points. In two dimensions, there is an established set of three geometric conditions that together guarantee stable coexistence of two species in a large class of models [6, 7, 57]:

1. The ZNGI's of both species must intersect.

2. The supply vector h_α (evaluated at the intersection point) must lie within the cone spanned by the negative impact vectors $-q_{i\alpha}$ (also evaluated at the intersection point).

3. The impact of each species must be biased relatively more towards the resource that has a relatively larger effect on its own growth rate.

The first condition immediately follows from requiring $dN_i/dt = 0$ for all surviving species, and is valid in any dimension. This is the source of the *Competitive Exclusion Principle*, which says that at most M consumer species can stably coexist. The competitive exclusion principle is simply the statement that, generically, at most M ZGNIs can intersect at a single point in an M-dimensional space in the absence of additional constraints or degeneracies in growth rates.

The second condition guarantees that a set of positive population sizes N_i can be found whose combined impact $\sum_j N_j q_{j\alpha}$ perfectly balances the supply h_α, so that $dR_\alpha/dt = 0$. Rewriting the second equation in Eq. 4 with the left hand side set to zero gives

$$h_\alpha(\mathbf{R}) = \sum_j N_j \left(-q_{j\alpha}(\mathbf{R})\right). \tag{5}$$

Since species abundances must be positive $N_j > 0$, the equation has the geometric interpretation that the the supply h_α must lie within the cone spanned by the negative impact vectors. This condition is also valid in any dimension, and for any number of species, not just in $M = S = 2$. For the most commonly used models the supply vector can be regarded as pointing directly towards a "supply point," as depicted in Fig. 3. The supply point represents the steady-state resource abundances in the absence of consumers. The graphical representation is further simplified in such cases, since the full supply vector field no longer needs to be directly considered, and the satisfaction of the coexistence condition depends simply on whether the supply point (which is often modeled as a directly accessible control parameter) falls inside the coexistence cone.

The third condition is the most subtle. The first two conditions establish that $dN_i/dt = dR_\alpha/dt = 0$, but imply nothing about the stability of this fixed point. In two dimensions, Tilman showed that the stability of the fixed point in a large class of two-dimensional models can be determined by comparing the relative orientation of the impact vectors to the relative orientation of the ZNGI's (see also Appendix B) [57].

In the rest of this section, we illustrate this graphical approach and the conclusions that can be drawn from it using three canonical examples. See Appendix A for a systematic taxonomy of the main kinds of classic niche theory models, with additional examples.

4.1 MacArthur's Consumer Resource Model

We begin with MacArthur's Consumer Resource Model (MCRM), which was the first model developed within this framework, and remains a fundamental reference point for ecological theory ([4, 58]). This model describes competition for "substitutable" resources using simple mass-action kinetics. Individuals of consumer species i encounter and consume units of resource α at a rate $c_{i\alpha}R_\alpha$, where the $c_{i\alpha}$ are constants expressing the preferences of species i for the different types. Resource α carries an energy density w_α. Some amount of energy per individual m_i is required to sustain the population at steady state, and excess energy leads to population growth, with an efficiency factor e_i converting from energy intake to growth rate. Finally, the resources themselves are self-renewing, such that in the absence of consumers, each resource type α follows its own independent logistic growth law with carrying capacity

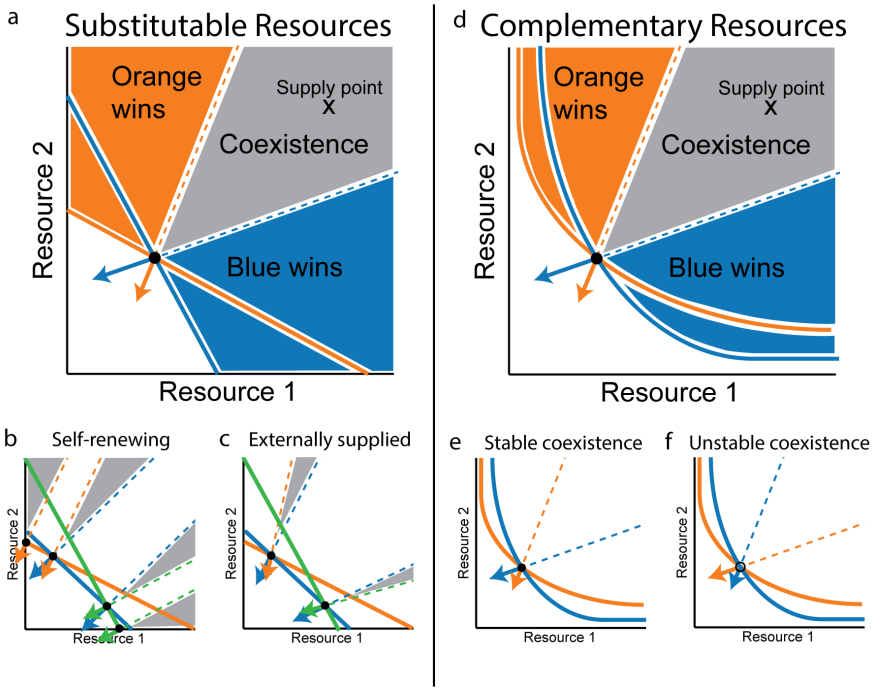


Figure 3: **Niche Theory Examples.** *(a,b,c)* Different resource supply choices with substitutable resources yield different coexistence cone geometries. *(d,e,f)* Stability depends on relative orientation of impact vectors.

K_α and low-density growth rate r_α. Putting this all together, we have:

$$\frac{dN_i}{dt} = N_i e_i \left[\sum_\beta w_\beta c_{i\beta} R_\beta - m_i \right] \tag{6}$$

$$\frac{dR_\alpha}{dt} = \frac{r_\alpha}{K_\alpha} R_\alpha (K_\alpha - R_\alpha) - \sum_j c_{j\alpha} N_j R_\alpha. \tag{7}$$

Graphical representations of this model for two or three species and two resources are shown in Fig. 3*(a,b)*. In sketching these figures, we have used the fact that the equilibrium condition $dR_\alpha/dt = 0$ for each resource can be arbitrarily rescaled without affecting the results of the analysis. The standard practice in the literature is to multiply both sides by w_α/R_α, yielding:

$$0 = w_\alpha r_\alpha K_\alpha (K_\alpha - R_\alpha) - \sum_j w_\alpha c_{j\alpha} N_j. \tag{8}$$

This can equivalently be viewed as the equilibrium condition for a model with supply vector

$$h_\alpha = w_\alpha r_\alpha K_\alpha (K_\alpha - R_\alpha) \tag{9}$$

which points directly at the supply point K_α, and impact vector

$$q_{i\alpha} = -w_\alpha c_{i\alpha} \tag{10}$$

which is minus the gradient $\partial g_i / \partial R_\alpha$ of the growth rate. This rescaling is extremely convenient for the graphical representation because the impact vectors are always perpendicular to the ZNGI's, and the coexistence condition reduces to the requirement that the supply point fall

inside the (negative) cone spanned by these vectors. If the supply point falls outside this cone but above at least one of the ZNGI's, then one of the species will persist, driving the other one extinct. As shown in Fig. 3*(a)*, this allows us to divide up the whole resource space into three regions based on the outcome of competition when the supply point is placed within the region: the "coexistence cone" where both species persist and two regions where one of the species competitively excludes the other. Note that the third criterion for stable coexistence above is always satisfied in this model, because the angle of the impact vector is tied to the angle of the ZNGI, and so the equilibrium point is always linearly stable.

Fig. 3*(b)* illustrates some additional features of this model that follow from the graphical representation. In the figure, we have drawn ZNGI's for three species, resulting in two possible coexistence cones. Notice as expected from the competitive exclusion principle, at most two of the species can coexist at anytime. The first thing to note is that the third intersection, between the orange and green species, is not a possible coexistence point, because the blue species has a positive growth rate there, and will always re-invade. The second feature is the presence of two cones at the intersection of the inner ZNGI's with the resource axes. The existence of these cones does not depend on the number of species, but they were omitted from panel *(a)* for simplicity, since they do not affect the outcome of competition. When the supply point lies inside one of these cones, one of the resources is driven to extinction, and the equilibrium state lies at the intersection of the ZNGI with the resource axis.

When these cones are included in the diagram, then the competitive exclusion limit is saturated whenever the supply point falls in any cone. Although only one species survives in the boundary cones, the number of species still equals the number of *persisting* resources. This leads us to the final interesting feature of the diagram, which is the fact that the inner boundaries of neighboring cones are always parallel. This follows immediately from the fact that the impact vectors are perpendicular to the ZNGI's. It implies that the cones cover the full 90-degree angle of the quadrant of positive resource abundances. As the magnitude of the supply grows (i.e., the distance of the supply point from the origin), it therefore requires increasingly precise fine-tuning of the supply point to avoid landing in one of the cones, with the result that the competitive exclusion bound is generically saturated in the limit of large supply.

4.2 Externally supplied resources

The logistic growth of the resources in the absence of consumers makes the MCRM a reasonable model of competition for substitutable biotic resources, such as for herbivores competing for plants. But if the resources are externally supplied nutrients, a different model is needed. For example, in microbial ecology, there can be competition for various organic molecules that can be used as carbon sources. In such cases, a better model for the supply vector is a simple linear law $h_\alpha = \kappa_\alpha - \omega_\alpha R_\alpha$ such as would arise from an ideal chemostat with incoming nutrient flux κ_α and dilution rate ω_α. With this supply, we obtain:

$$\frac{dN_i}{dt} = N_i e_i \left[\sum_\beta w_\beta c_{i\beta} R_\beta - m_i \right] \tag{11}$$

$$\frac{dR_\alpha}{dt} = \kappa_\alpha - \omega_\alpha R_\alpha - \sum_j c_{j\alpha} N_j R_\alpha. \tag{12}$$

In this case, the true supply vector already points directly towards the supply point $\kappa_\alpha/\omega_\alpha$. Notice that the impact vector $q_{i\alpha} = -c_{i\alpha} R_\alpha$ depends on the resource concentration and is no longer perpendicular to the ZNGI. The factor of R_α rotates it towards the most abundant resource, compressing all the supply cones towards the resource axes as shown in Fig. 3*(c)*.

The first consequence of this modified cone geometry is that the extinct resource cones disappear. When $R_\alpha = 0$ for one of the resources, the impact vector becomes parallel with the other resource axis, shrinking the angle spanned by this cone to zero. The fact that no equilibrium state can have $R_\alpha = 0$ for any α can also be seen by direct inspection of the Eq. (12), where $dR_\alpha/dt = \kappa_\alpha > 0$ at zero concentration of resource α. The second consequence is that adjacent coexistence cones no longer have parallel boundaries. This means that a finite fraction of possible supply points remain outside all the cones even in the limit of infinite total supply. In the high-dimensional setting, this prevents random high-dimensional ecosystems with linear resource dynamics from saturating the competitive exclusion bound [59].

4.3 Essential resources

Both of the previous examples assumed that resources were substitutable, and that the same growth rate could be obtained with any combination of resources with the same combined energy content. But many of the resources required by organisms play distinct functional roles, all of which must be filled for the organism to survive. Plants, for example, require sufficient supplies of carbon, nitrogen and phosphorous to be present in the soil. If any one of the three is too scarce, the plant will not survive, regardless of the available quantities of the other two. This phenomenon implies that the ZNGI's do not intersect with any of the axes, and must have the roughly hyperbolic shape sketched in Fig. 3*(d)*. Such ZNGI's can be obtained from a model of the form:

$$\frac{dN_i}{dt} = N_i \left\{ \left[\sum_\beta (c_{i\beta} R_\beta)^n \right]^{\frac{1}{n}} - m_i \right\} \tag{13}$$

$$\frac{dR_\alpha}{dt} = \kappa_\alpha - \omega_\alpha R_\alpha - \sum_j \nu_{j\alpha} N_j \left[\sum_\beta (c_{i\beta} R_\beta)^n \right]^{\frac{1}{n}} \tag{14}$$

with $n < -1$. In the limit $n \to -\infty$, this model reduces to Liebig's Law of the Minimum, where the growth rate of each species only depends on the concentration of the most limiting resource:

$$\frac{dN_i}{dt} = N_i \left[\min_\beta (c_{i\beta} R_\beta) - m_i \right] \tag{15}$$

$$\frac{dR_\alpha}{dt} = \kappa_\alpha - \omega_\alpha R_\alpha - \sum_j \nu_{j\alpha} N_j \min_\beta (c_{j\beta} R_\beta) \tag{16}$$

We have chosen the impact vector in the standard way, in which the nonlinear term $\left[\sum_\beta (c_{i\beta} R_\beta)^n\right]^{\frac{1}{n}}$ is treated as a total resource uptake rate, which is partitioned among the available resources according to the fixed biomass stoichiometry of the species $\nu_{i\alpha}$ [60].

In this model, the direction of the impact vector is set by the biomass stoichiometry $\nu_{i\alpha}$, regardless of the resource concentration and of the parameters $c_{i\alpha}$ and n that define the shape of the ZNGI. This means that the third criterion for stable coexistence is not necessarily satisfied, and fixed points can exist that are linearly unstable. Panels *(e)* and *(f)* of Fig. 3 represent two possible choices of biomass stoichiometry that lead to the same coexistence cone, with only the first one leading to stable coexistence.

The existence of different kinds of fixed points in this model opens the door to more complex attractors. For example, limit cycles, heteroclinic cycles, and chaos have all been observed in models of this basic type when $M, S \geq 3$ [61–63].

5 Generalized Lotka-Volterra Theory

In the introduction to these lectures, we noted that ecological theory considers two kinds of interactions that affect an organism's growth rate: interactions with the environment and interactions with other species. In the previous section, we saw how CRMs focus on the first kind of interaction. Historically, the earliest attempts at mathematical modeling of ecosystems focused on the second kind of interactions, reducing the role of the environment to a mediator of species-species interactions [64–66]. The most influential model for understanding species-species interactions is the Lotka-Volterra model. This model was first developed by Lotka and Volterra to describe predator-prey oscillations, and later extended to more general scenarios of interactions among larger collections of species [55, 64, 67].

In this section, we derive the multi-species (or "generalized") Lotka-Volterra equations (GLV) as a limit of the framework introduced in eq. (1-2). We then review how these equations have been used to analyze coexistence in two-species communities. Finally, we provide additional intuition for these results and point out potential complications by applying them to the effective interactions that emerge from the consumer resource models (CRMs) studied above. We will see that "higher-order" interactions generically arise from resource-explicit models, and that often, the strength of interaction between a pair of species becomes context-dependent.

5.1 Deriving the GLV Equations

From the physicist's perspective, the GLV equations are analogous to elasticity theory. We take a complex nonlinear system and describe it in terms of the phenomenological linear-response parameters that characterize its behavior near a stable fixed point. This can be straightforwardly carried out in the case of a fixed environment $\mathbf{E} = \text{const}$. Expanding the growth rate g_i to linear order in the perturbation $\mathbf{N} - \bar{\mathbf{N}}$ from a fixed point $\bar{\mathbf{N}}$, we obtain:

$$\frac{dN_i}{dt} = N_i \left[g_i + \sum_j \frac{\partial g_i}{\partial N_j}(N_j - \bar{N}_j) + \mathcal{O}((N_j - \bar{N}_j)^2) \right] \tag{17}$$

where g_i and its derivatives are evaluated at $\mathbf{N} = \bar{\mathbf{N}}$. If all species are present at nonzero abundance then we know $g_i(\bar{\mathbf{N}}) = 0$. But the g_i is included in this expression to allow for expansion about fixed points where $N_i = 0$ for some i. This will be required in the invasion analysis below.

If this were the exact model, with no higher-order terms, the intrinsic growth rate r_i of species i would be found by extrapolating to $N_j = 0$ for all j in the sum, yielding:

$$r_i = g_i(\bar{\mathbf{N}}) - \sum_j \frac{\partial g_i}{\partial N_j} \bar{N}_j. \tag{18}$$

The effect of other organisms on the growth rate is given by the "competition coefficients" α_{ij}, equal to

$$\alpha_{ij} = -\frac{\partial g_i}{\partial N_j}. \tag{19}$$

In terms of these parameters, the local dynamics are

$$\frac{dN_i}{dt} = N_i \left[r_i - \sum_j \alpha_{ij} N_j + \mathcal{O}((N_j - \bar{N}_j)^2) \right]. \tag{20}$$

This is the canonical GLV equation. Since the derivatives that define r_i and α_{ij} are evaluated at $\mathbf{N} = \bar{\mathbf{N}}$, they will in general take on different values at different fixed points, including when different assemblages of species are present.

The GLV framework is frequently employed to model environmentally mediated interactions as well as direct interactions. Doing this requires somehow writing the environmental state as a function of the population sizes $\mathbf{E}(\mathbf{N})$. Once this is achieved, the growth rate $g_i(\mathbf{N}, \mathbf{E}(\mathbf{N}))$ is again a function of the population sizes alone, and the GLV parameters are readily obtained using the chain rule:

$$r_i = g_i(\bar{\mathbf{N}}, \bar{\mathbf{E}}) - \sum_j \left(\frac{\partial g_i}{\partial N_j} + \sum_\beta \frac{\partial g_i}{\partial E_\beta} \frac{\partial E_\beta}{\partial N_j} \right) \bar{N}_j \tag{21}$$

$$\alpha_{ij} = -\frac{\partial g_i}{\partial N_j} - \sum_\beta \frac{\partial g_i}{\partial E_\beta} \frac{\partial E_\beta}{\partial N_j}. \tag{22}$$

If the environmental dynamics have a stable fixed point for arbitrary fixed $\mathbf{N}$, this fixed point defines a function $\mathbf{E}(\mathbf{N})$ that can be used in the above definitions. But the resulting GLV dynamics do not necessarily match the original dynamics, even in the vicinity of the fixed point [68].

5.2 Invasion analysis and Modern Coexistence Theory

One of the main uses to which the GLV equations have been put is the derivation of conditions for species coexistence. A categorization of the main factors affecting coexistence has been developed with the help of this framework, and is known as Modern Coexistence Theory (MCT) [37, 69, 70]. Instead of studying the properties of fixed points, as we did with niche theory above, MCT obtains sufficient conditions for coexistence in terms of the ability of each species to invade the rest of the community. While this approach has some limitations, which will emerge in the examples below, it has the distinct advantage of not depending on how well the GLV equations approximate the dynamics. The analysis is conducted entirely in terms of the sign of the growth rate g_i of the hypothetical invader at a fixed point of the resident community, for which the GLV equations provide an exact expression.

The central results of this theory concern the interaction of two species [37, 69]. We start by computing the equilibrium abundance of each species i when it is the only species present:

$$\bar{N}_{i/j} = \frac{r_{i/j}}{\alpha_{ii/j}} \tag{23}$$

where the $/j$ in the subscript indicates that these quantities are being evaluated at the fixed point where the other species j is absent. If we now imagine introducing species j at vanishingly small density, its initial per-capita growth rate is:

$$\begin{aligned} g_{j/j} &= r_{j/j} - \alpha_{ji/j} \bar{N}_{i/j} \\ &= r_{j/j} - \alpha_{ji/j} \frac{r_{i/j}}{\alpha_{ii/j}}. \end{aligned} \tag{24}$$

For species j to successfully invade, this quantity must be positive, which implies:

$$\frac{r_{j/j}}{r_{i/j}} \frac{\alpha_{ii/j}}{\alpha_{ji/j}} > 1. \tag{25}$$

The two species i and j are guaranteed to stably coexist when each one is capable of invading the other, that is, when:

$$\frac{r_{j/j}}{r_{i/j}} \frac{\alpha_{ii/j}}{\alpha_{ji/j}} > 1 > \frac{r_{j/i}}{r_{i/i}} \frac{\alpha_{ij/i}}{\alpha_{jj/i}}. \tag{26}$$

This is a sufficient condition for coexistence of two species in any model within our framework, as long as the relationship $\mathbf{E}(\mathbf{N})$ can be defined. It is not a necessary condition, because it is possible for a species that fails to invade at low density to successfully invade at high density. But it does facilitate a robust classification of factors that promote coexistence.

This classification is normally formulated by distinguishing two contributions to the expressions that appear in the inequality. This distinction implicitly assumes that the GLV parameters are independent of which species is resident, so that the $/j$ notation can be dropped from eq. (26). The quantities r_i, r_j can then be interpreted as estimates of the intrinsic "fitness" of the each species, and the same fitness ratio r_j/r_i appears on both sides of (26). The expression can therefore be rearranged to give:

$$\frac{\alpha_{ii}}{\alpha_{ji}} > \frac{r_i}{r_j} > \frac{\alpha_{ij}}{\alpha_{jj}}. \tag{27}$$

Factors that tend to reduce fitness differences (making r_i/r_j closer to 1) are known as "equalizing mechanisms" of coexistence. But the coexistence condition (27) shows that equalizing mechanisms are insufficient to achieve coexistence, because even as the ratio approaches 1, the ratios of competition coefficients still remain. We see that coexistence requires at least one of the species to compete with itself more than it competes with the other species (i.e., $\alpha_{ii} > \alpha_{ij}$). Factors that increase competition among members of the same species relative to competition with the other species are known as "stabilizing mechanisms" of coexistence.

In the following sections, we present the GLV approximations to the CRMs presented in Section 4 above. In the original MCRM (eq. 6-7), the approximation is exact, and stabilizing mechanisms have a mechanistic interpretation in terms of decreased niche overlap. When the resources are externally supplied (eq. 11-12), the GLV parameters become context-dependent, and the equalizing and stabilizing mechanisms can no longer be cleanly separated as in eq. (27).

5.3 MacArthur's Consumer Resource Model

Additional intuition about the operation of stabilizing mechanisms comes from analyzing the GLV approximation to the MCRM [4, 58]. We will consider a slightly generalized version of the model, allowing depletion rates $d_{i\alpha}$ to be independent of the growth law parameters $c_{i\alpha}$:

$$\frac{dN_i}{dt} = N_i\left[\sum_\beta w_\beta c_{i\beta} R_\beta - m_i\right] \tag{28}$$

$$\frac{dR_\alpha}{dt} = \frac{r_\alpha}{K_\alpha} R_\alpha(K_\alpha - R_\alpha) - \sum_j d_{j\alpha} N_j R_\alpha. \tag{29}$$

Following the procedure outlined above, we write the resource abundances as a function of the population sizes by setting $dR_\alpha/dt = 0$. This equation has two solutions, $R_\alpha = 0$, and

$$R_\alpha = K_\alpha\left(1 - \frac{1}{r_\alpha}\sum_j d_{j\alpha} N_j\right). \tag{30}$$

If the term in brackets is negative, then $dR_\alpha/dt < 0$ for all positive R_α, and so $R_\alpha = 0$ is the stable solution. Otherwise, $R_\alpha = 0$ is unstable, because a small addition of the resource will lead to a positive growth rate. The stable equilibrium state is therefore given by

$$R_\alpha(\mathbf{N}) = \max\left[0, K_\alpha\left(1 - \frac{1}{r_\alpha}\sum_j d_{j\alpha} N_j\right)\right]. \tag{31}$$

Inserting this into eqs. (21), we obtain an effective GLV of the form:

$$\frac{dN_i}{dt} = N_i \left[r_i - \sum_j \alpha_{ij} N_j + \mathcal{O}((N_j - N_j^*)^2) \right]. \tag{32}$$

with

$$r_i = \sum_{\beta \in \mathbf{M}^*} w_\beta c_{i\beta} K_\beta - m_i \tag{33}$$

and

$$\alpha_{ij} = \sum_{\beta \in \mathbf{M}^*} \frac{w_\beta K_\beta}{r_\beta} c_{i\beta} d_{j\beta}, \tag{34}$$

where $\mathbf{M}^*$ represents the set of resources with nonzero abundance. Eq. 33 has a simple interpretation. The intrinsic growth rate r_i is the difference between the maximum consumption capacity $\sum_{\beta \in \mathbf{M}^*} w_\beta c_{i\beta} K_\beta$ at the given resource supply levels and the maintenance cost m_i.

The formula for the competition matrix in eq. (34) carries an ecologically significant meaning, showing how the strength of competition is determined by the overlap between the requirements $c_{i\alpha}$ of one species and the impacts $d_{j\alpha}$ of the other. The overlap is weighted by a product of three factors: (a) the value w_α of each resource, with more energy-rich resources counting for more in the sum, (b) the low-density growth timescale $1/r_\alpha$ for each resource, with rapid renewal decreasing the significance of the resource as a locus of competition and (c) the resource carrying capacity K_α, giving resources with higher carrying capacity a greater contribution to the interaction.

Note that if no resources are driven to extinction, then both r_i and α_{ij} are independent of which fixed point is chosen (e.g., whether both species, only species i or only species j is present in the two-species case). We can therefore drop the extra index indicating which species is absent, and simply write r_i and α_{ij} as assumed in the simplified form of the coexistence condition (27).

These considerations suggest that the scaled quantity α_{ij}/α_{ii} is a natural definition of the "niche overlap" between species i and j. This definition is ambiguous, however, since it is not symmetric under exchange of i and j. If we symmetrize by taking the geometric mean, we can define the niche overlap ρ by

$$\rho = \sqrt{\frac{\alpha_{ij}\alpha_{ji}}{\alpha_{ii}\alpha_{jj}}}. \tag{35}$$

With this definition, and recalling that the values of the GLV parameters do not depend on which species is the resident and which is the invader (as long as no resources go extinct), we can simplify the coexistence condition (26) into the more interpretable form [37, 69]

$$\frac{1}{\rho} > \frac{r_1}{r_2} > \rho. \tag{36}$$

In this context, stabilizing mechanisms are features of the ecosystem that reduce the niche overlap, opening up more space for species with different maintenance costs m_i or consumption capacity $\sum_{\beta \in \mathbf{M}^*} w_\beta c_{i\beta} K_\beta$ (the two components of r_i) to stably coexist.

5.4 Externally supplied resources

As noted above, a special feature of the MCRM is that the "intrinsic" growth rate r_i and the competition coefficient α_{ij} between species i and species j (Eq. 34) can be fully determined by the properties of these two species and of the resources, regardless of which species or combination of species is currently resident. The one exception occurs when the addition of a third species drives one of the resources extinct, so that the set $\mathbf{M}^*$ of resources to be summed over changes. This caveat is often ignored in the literature.

In other models, the context-dependence of the GLV parameters is much more apparent. For the case of externally supplied substitutable resources (Eq. 11-12), solving for the resource equilibrium yields:

$$R_\alpha(\mathbf{N}) = \frac{\kappa_\alpha}{\omega_\alpha + \sum_j c_{j\alpha} N_j}. \tag{37}$$

Inserting this into the equation for the consumer dynamics (Eq. 11) gives:

$$\frac{dN_i}{dt} = N_i \left[\sum_\beta \frac{w_\beta c_{i\beta} \kappa_\beta}{\omega_\beta + \sum_j c_{j\beta} N_j} - m_i \right]. \tag{38}$$

This is no longer of the GLV form. But in the vicinity of the equilibrium state N_i^*, we can still use eq. (21-22) to locally obtain the GLV parameters

$$r_i = \sum_\beta w_\beta c_{i\beta} R_\beta^* \frac{2\kappa_\beta - \omega_\beta R_\beta^*}{\kappa_\beta} - m_i \tag{39}$$

$$\alpha_{ij} = \sum_\beta \frac{w_\beta (R_\beta^*)^2}{\kappa_\beta} c_{i\beta} c_{j\beta}. \tag{40}$$

The contribution of each resource to the sums in both equations now depends on its equilibrium abundance R_α^*. This complicates the above analysis of stabilizing versus equalizing mechanisms, since in general it is no longer true that $r_{i/j} \approx r_{i/i}$. In communities with more than two species, this also generates effective higher-order interactions of all orders, since the interaction between any pair of species is affected by all the other species with overlapping consumption preferences.

6 Niche theory and optimization

Ecological models where species interact reciprocally (i.e. how species i affects species j is identical to how species j affects species i) can be recast in the language of optimization. This optimization perspective provides a powerful set of analytical tools. In the papers where MacArthur first presents his Consumer Resource Model, he already started to investigate the conditions under which this ecological model can be represented in terms of an analogous optimization principle [66, 71]. Recent work has extended his original results and identified when such an optimization principle can be found [72]. One of the key lessons from these works is that it is much more natural to optimize over the state of the environment (i.e. the resource abundances) than over population sizes. When we do this, the objective function has a transparent ecological interpretation.

Below, we start with MacArthur's original result, where the optimization is performed over the space of population sizes in the GLV approximation to the MCRM. We then briefly describe

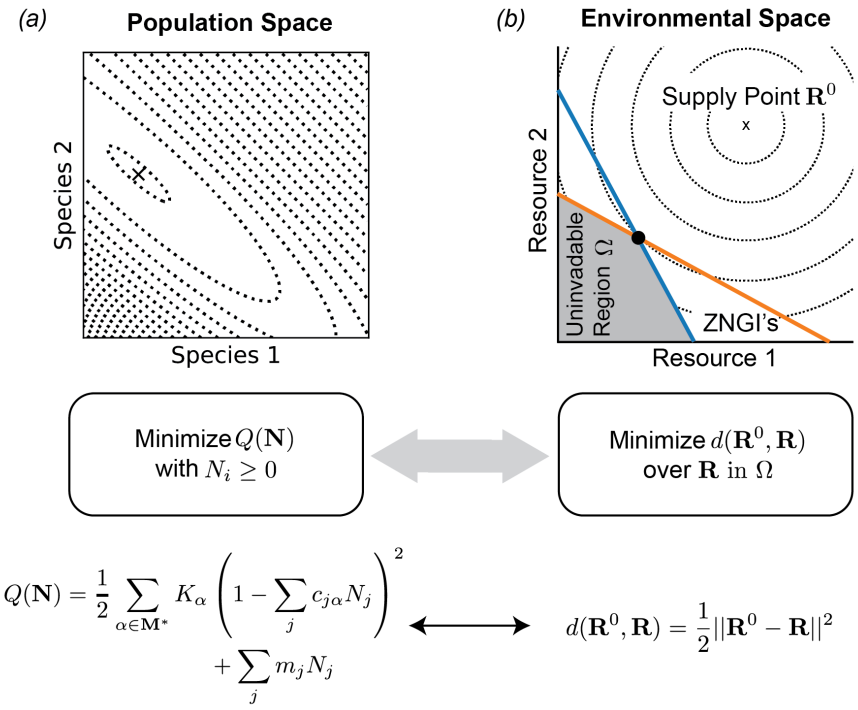


Figure 4: **Reinterpreting MacArthur's Minimization Principle.** *(a)* MacArthur's objective function $Q(\mathbf{N})$ (eqs. 41-42) plotted as a function of the two population sizes. The 'x' shows the steady state reached by numerically simulating the full model of eq. (6-7) using the parameters $r_\alpha = m_i = w_\alpha = 1$ ($\alpha = 1,2$, $i = 1,2$), $K_1 = 4.8$, $K_2 = 2.85$, $c_{1\alpha} = (0.5, 0.3)$ and $c_{2\alpha} = (0.4, 0.6)$. As MacArthur predicted, the simulation ends at the minimum of Q. The panel also gives the simpler form of Q obtained when $r_\alpha = w_\alpha = 1$. *(b)* The same ecosystem viewed in resource space. Contours show the environmental perturbation $d(\mathbf{R}^0, \mathbf{R})$, which measures how far the resource abundances are from the supply point $\mathbf{R}^0$. The colored lines are the zero net-growth isoclines (ZNGIs) of the two species, and the shaded region Ω is where no species can invade (i.e. $g_i(\mathbf{R}) \leq 0$ for all i). The steady state (black dot) is the point in Ω that minimizes d. Since $r_\alpha = w_\alpha = 1$, here d is just the Euclidean distance. Figure adapted from Ref. [72].

the optimization problems in resource space solved by the equilibrium states of the CRMs introduced in Section 4. We encourage the reader to consult [72] for further details and additional examples.

6.1 MacArthur's Minimization Principle

MacArthur noticed that the GLV representation of the MCRM given above in eq. (20) and (33-34) has a special structure: the per-capita growth rate of each species is proportional to the derivative of a single quadratic function of the population sizes N_i:

$$\frac{dN_i}{dt} = -e_i N_i \frac{\partial Q}{\partial N_i} \tag{41}$$

with

$$Q(\mathbf{N}) = \frac{1}{2}\sum_{\alpha\in\mathbf{M}^*} r_\alpha^{-1} K_\alpha w_\alpha \left(r_\alpha - \sum_j c_{j\alpha}N_j\right)^2 + \sum_j m_j N_j. \tag{42}$$

To check this, simply take the partial derivative of Q with respect to N_i and compare it to the GLV equations. Notice what eq. (41) tells us about the steady state. Any species that survives ($N_i > 0$) must sit at a point where $\partial Q/\partial N_i = 0$. Because of the minus sign in eq. (41), the dynamics always move downhill in Q, so this point is a minimum and not a maximum. Species that have gone extinct ($N_i = 0$) must not be able to re-invade the community. This requires that $\partial Q/\partial N_i > 0$ (i.e. increasing N_i would increase Q). In other words, Q is also minimized along these directions, this time at the boundary $N_i = 0$ set by the requirement that populations be non-negative [73, 74]. Fig. 4(a) shows $Q(\mathbf{N})$ for an ecosystem with two consumers. The 'x' marks where a direct numerical simulation of the full MCRM (eq. 6-7) ends up. As expected, it lies at the minimum of Q.

MacArthur's observation tells us a great deal about the steady states of this model. For example, since Q is convex, it has only one local minimum, and hence the model has a unique stable steady state. However, the objective function Q lacks an intuitive interpretation except in some special cases, and the essential assumptions required to obtain an optimization principle remained unclear. In the following sections, we show how the optimization perspective can be applied to a surprisingly large class of models, with a uniform physical interpretation in terms of how far the environment is pushed away from the supply point. The crucial assumption turns out to be the symmetry of the environmentally-mediated interactions among species [72]. In modern language, this is simply the statement that interactions between species are reciprocal.

It is also possible to generalize this optimization framework to models where interactions are not reciprocal. However, in this case, using it to actually find the steady state is more involved (see Section 6.4 below and [72] for a detailed discussion).

6.2 MacArthur's Consumer Resource Model

We begin with MacArthur's original model, presented in eqs. (6-7) above. At steady-state, we have that

$$0 = N_i g_i(\mathbf{R}) = N_i e_i \left[\sum_\beta w_\beta c_{i\beta} R_\beta - m_i \right] \tag{43}$$

$$0 = \frac{r_\alpha}{K_\alpha} R_\alpha (K_\alpha - R_\alpha) - \sum_j c_{j\alpha} N_j R_\alpha, \tag{44}$$

where in the first line we have defined the growth rate of species i:

$$g_i(\mathbf{R}) = e_i \left[\sum_\beta w_\beta c_{i\beta} R_\beta - m_i \right]. \tag{45}$$

To proceed, we rewrite the second of these equations in terms of the function

$$d(\mathbf{R}^0, \mathbf{R}) = \frac{1}{2} \sum_\alpha w_\alpha r_\alpha K_\alpha^{-1} (R_\alpha - R_\alpha^0)^2. \tag{46}$$

It turns out that $d(\mathbf{R}^0, \mathbf{R})$ is the natural measure of environmental perturbation and can be thought of as a (weighted) Euclidean distance between the resource abundances $\mathbf{R}$ and the supply point $\mathbf{R}^0$. For the MCRM, the supply point is just given by the carrying capacities of the resources, $R_\alpha^0 = K_\alpha$. A straightforward calculation shows that we can rewrite Eq. 44 as

$$0 = \frac{\partial d}{\partial R_\alpha} + \sum_j \frac{N_j}{e_j} \frac{\partial g_j}{\partial R_\alpha}. \tag{47}$$

If we in addition note that both species and resource abundances must be positive, the steady-state conditions can be summarized by the following four conditions:

$$\text{Steady populations} \quad 0 = \frac{dN_i}{dt} \propto \frac{N_i}{e_i} g_i(\mathbf{R}) \tag{48}$$

$$\text{Steady environment} \quad 0 = -\frac{\partial d}{\partial R_\alpha} - \sum_j \frac{N_j}{e_j} \frac{\partial g_j}{\partial R_\alpha} \tag{49}$$

$$\text{Noninvasibility} \quad 0 \geq g_i(\mathbf{R}) \tag{50}$$

$$\text{Feasible populations} \quad 0 \leq \frac{N_i}{e_i}. \tag{51}$$

Readers familiar with optimization theory will recognize these four equations. They are exactly the Karush-Kuhn-Tucker (KKT) conditions for minimizing the function $d(\mathbf{R}^0, \mathbf{R})$ subject to the inequality constraints $g_i(\mathbf{R}) \leq 0$ [75, 76]. The KKT conditions are the natural extension of the method of Lagrange multipliers to problems where the constraints are inequalities rather than equalities. In our ecological problem, the role of the Lagrange multipliers (called KKT multipliers in this context) is played by the rescaled species abundances N_i/e_i.

In particular, these four conditions are sufficient to show that the steady-state resource abundances $\mathbf{R}$ are solutions to the following constrained optimization problem

$$\begin{aligned} \min_R \quad & d(\mathbf{R}^0, \mathbf{R}) \\ \text{s.t.} \quad & g_j(\mathbf{R}) \leq 0 \\ & R_\alpha \geq 0, \end{aligned} \tag{52}$$

with $\mathbf{R}^0 = \mathbf{K}$. Since the resource dynamics in eq. (44) generally has two potential solutions: $R_\alpha = 0$ and $R_\alpha > 0$, we impose additional constraints $R_\alpha \geq 0$ in the optimization process to find the steady-state solution with the maximum number of surviving resources.

Let us pause to understand the ecological meaning of this objective function. Notice that $d(\mathbf{R}^0, \mathbf{R})$ does not treat all resources equally. Instead, each resource enters with its own weight $w_\alpha r_\alpha K_\alpha^{-1}$, which sets how much a given change in its abundance "costs". It is helpful to think about each of the three factors separately. The factor w_α is the nutritional value of resource α. If w_α is small, the resource does little to support consumer growth, and it does not matter much if its abundance shifts. The factor r_α is the rate at which the resource regenerates. A resource that grows back quickly is hard to deplete, so a shift of a given size represents a larger disturbance than the same shift in a resource that regenerates slowly. Finally, the factor K_α^{-1} tells us that what matters is the size of a perturbation relative to the carrying capacity. In other words, a change of fixed absolute size is less important for a resource with a large carrying capacity.

In the constrained optimization problem, the species abundances N_i are simply the generalized Lagrange multipliers (KKT multipliers) corresponding to the S inequality constraints $g_i(\mathbf{R}) \leq 0$. Notice that the steady-state species condition (known as the K.K.T. condition in the constrained optimization literature) is simply the statement that $\frac{N_i}{e_i} g_i(\mathbf{R}) = 0$. It states that species is either extinct (i.e. $g_i(\mathbf{R}) < 0$ meaning that $N_i = 0$) or that a species survives at steady-state (i.e. $g_i(\mathbf{R}) = 0$ meaning that $N_i > 0$). Furthermore, from the definition of Lagrange multipliers, we see that the species abundance N_i can also be thought of as how much the objective $d(\mathbf{R}, \mathbf{R}^0)$ changes if we infinitesimally change the growth rate $g_i(\mathbf{R})$ of species i.

6.3 Externally supplied resources

The environment-based perspective on optimization facilitates generalization to other models not considered by MacArthur, including the model of externally supplied susbstitutable resources discussed above in eq. (11-12). For this model, d is no longer a quadratic function.

Instead, it takes the form of a weighted Kullback-Leibler (KL) divergence (see [72] for detailed derivation):

$$d(\mathbf{R}^0, \mathbf{R}) = \omega \sum_\alpha w_\alpha \left[R^0_\alpha \ln \frac{R^0_\alpha}{R_\alpha} - (R^0_\alpha - R_\alpha) \right] \tag{53}$$

where the supply point is given by $R^0_\alpha = \kappa_\alpha/\omega$. The constrained minimization of this function at equilibrium can easily be verified by taking the derivative and substituting into the KKT conditions as given above in eq. (48-51). The appearance of the KL divergence here is quite natural. It is commonly used to measure how different two vectors with positive entries are from each other (e.g. two probability distributions or two sets of chemical concentrations) [77]. Just like in MacArthur's model, the nutritional value w_α sets how much each resource counts. There is, however, one important difference. Since $d(\mathbf{R}^0, \mathbf{R})$ blows up as $R_\alpha \to 0$, the minimum can never lie on the boundary $R_\alpha = 0$. In other words, for externally supplied resources we no longer need to impose the constraint $R_\alpha \geq 0$ by hand – it is automatically satisfied.

6.4 Essential resources

We now turn to the case of nonsubstitutable resources with fixed biomass stoichiometry, as represented by eq. (13-14) above with $n < -1$. The environmentally mediated interactions among consumers in these models are non-reciprocal because a species can deplete resources that do not limit its own growth. In other words, how a species changes the environment is no longer tied to how the environment affects its growth. In particular, the limit $n \to -\infty$ ("Liebig's Law of the Minimum") gives the simplest qualitative example of this since each species has a single limiting resource that controls the growth rate and is insensitive to the concentrations of all the other resources. It can be shown that in this limit, the steady-state resource abundances minimize the same objective functions as above (eq. (53)), but with the supply point $\mathbf{R}^0$ replaced by an effective supply point $\tilde{\mathbf{R}}^0$. The effective supply point $\tilde{\mathbf{R}}^0$ corrects the external supply vectors $\mathbf{R}^0$ by accounting for depletion of non-limiting resources at steady state as follows:

$$\tilde{R}^0_\alpha = R^0_\alpha - \omega^{-1} \sum_{i,\alpha \neq \beta_i} N^*_i \nu_{i\alpha} c_{i\alpha} R^*_\alpha, \tag{54}$$

where the sum is over all species i that are limited by some resource β_i other than α. The constrained minimization of $d(\mathbf{R}, \tilde{\mathbf{R}}^0)$ in equilibrium can be verified as before by simply substituting into the KKT conditions (48-51).

While the notion of the effective supply point defined in eq. (54) allows the equilibrium conditions to be formally rewritten as an optimization problem, it also makes the objective function depend explicitly on the solution to the problem. A similar issue is encountered in machine-learning models with latent variables, where the most likely value of the latent variable is itself a function of the best-fit model parameters [78]. In machine learning, this conundrum is solved through an iterative algorithm called Expectation Maximization (EM), where the model fitting is performed at each iteration with the current estimate of the most likely latent variable values, and then the latent variables are updated using the new model parameters. An analogous approach proves empirically successful on a range of asymmetric ecological models, including models of essential resources [72].

7 Ecology in high-dimensions

7.1 Making the leap to high-dimensions

Natural ecosystems often have an astounding degree of phylogenetic and physiological diversity. For example, microbial communities are extremely diverse, ranging from 500-1000 species in human guts [79] to over 10^3 species in marine ecosystems [80]. Recent advances in DNA sequencing technologies makes it possible to measure microbial communities abundances at high resolution and has opened a precision era in microbial ecology [81, 82]. Understanding and modeling such complex datasets challenges current theories and analytical approaches. This has motivated the statistical physics community to start developing a theory of high-dimensional ecosystems.

In low dimensional ecosystems, carefully measuring the traits of each individual in the ecosystems is tractable. However, in high dimensions, this approach becomes intractable for both practical and theoretical reasons. Practically, it is impossible to characterize the preferences and interactions of hundreds of species. Theoretically, even if we could carry out this task, it becomes very difficult to make senes of the resulting high-dimensional dynamics.

A similar problem is encountered in physics when modeling systems such as gases which are composed of a large ensemble of particles. One pertinent example from statistical mechanics is the example of *particles in a box*. At any time t, the microstate of a system of N particles is described by $6N$ numbers, namely the positions $\vec{x}_i(t)$ and the momenta $\vec{p}_i(t)$, of the $i = 1, \dots N$ particles. When N is small, the evolution of the microstate can be predicted precisely by solving Hamilton's equations for all particles. However, when $N \gg 1$ (e.g. for $N \sim 10^{23}$ as is the case for a gas), the complexity of predicting the microstate is too high to be feasible. Instead, one is forced to describe the system *statistically*. Making an analogy between particles in a box and species in an ecosystem, this suggests that analytical approaches and theoretical insights derived from small ecosystems with a few species and resources may not scale up to large, diverse ecosystems.

We have stressed the practical difficulties in high dimension. Actually, taking the thermodynamic limit $N \to \infty$ can also lead to a number of simplifications. As we all know, statistical mechanics provides a prescription for successfully dealing with systems with a very large number of degrees of freedom [83]. For the *particles in a box* example, instead of trying to predict the behavior of individual particles, we can make accurate statements about the macroscopic quantities such as temperature and pressure which reflect averages over millions of particles. As a result of this averaging, macroscopic quantities often have universal behaviors. For example, the pressure, temperature, volume and entropy are related by Maxwell's relations regardless of if the gas particles being analyzed are oxygen, nitrogen or mixture of the two. This reflects the fact that in the thermodynamics limit, the behaviors we will observe will be *typical*.

The analogy between species in ecosystems and particles in box yields some general lessons from statistical mechanics that we can use to study high-dimensional ecology. First, we should concentrate of finding relationship between macroscopic properties of the ecosystem rather than the behaviors of any particular species or resource. Second, we should focus on characterizing the typical behaviors we would expect to see in diverse ecosystems. Finally, on a technical level, we must exploit the central limit theorem by averaging over species and resource traits to make predictions about ensembles of ecosystems rather than any particular ecosystem.

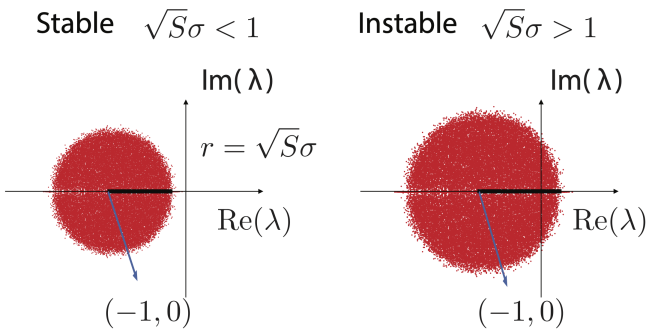


Figure 5: Schematic for May's stability criteria. The red scatter points are the eigenvalues on the complex plane.

7.2 May's Stability Criteria and Random Matrix Theory

One of the first examples of the use of ideas from statistical mechanics to ecology was the pioneering work by physicist-turned-ecologist Robert May [84]. May was inspired by Wigner's work on Uranium showing that one could recapitulate many aspects of experimental data by replacing the exact Uranium Hamiltonian by a large symmetric random matrix. The underlying reason for this was that the predicted properties of Uranium were *typical*. In other words, these properties did not depend on the detailed interactions of Uranium but instead were properties of any sufficiently complex system. Inspired by this, Wigner showed that one could model some aspects of complex Hamiltonians (i.e. the level statics) with large random matrices.

May transplanted these ideas to an ecological setting. He considered a general ecosystem with S species, with abundances N_i, whose dynamics take the form

$$\frac{dN_i}{dt} = g_i(N_1, N_2, \ldots, N_S) = g_i(\mathbf{N}) \tag{55}$$

He further assumed that the ecosystem had a fixed point given by $\bar{\mathbf{N}}$. He then asked about the nature of the dynamics around the fixed point. In this setting, the population dynamics can be described by Taylor expanding the dynamics around the equilibrium point $\bar{\mathbf{N}}$ to get

$$\frac{dN_i}{dt} = \sum_j J_{ij}(N_j - \bar{N}_j). \tag{56}$$

May wanted to know whether the fixed point $\bar{\mathbf{N}}$ is stable when the number of species is large ($S \gg 1$). Since eq. (56) is linear, the answer is determined by the eigenvalue $\lambda_{\max}$ of J_{ij} with the largest real part. If $\lambda_{\max}$ is negative, any small perturbation decays and the system returns to $\bar{\mathbf{N}}$. In contrast, if $\lambda_{\max}$ is positive, small perturbations grow and the ecosystem is driven away from the fixed point (i.e. the fixed point is unstable).

To answer this question, May made use of results from Random Matrix Theory. In the 1960's, Jean Ginibre calculated the eigenvalue distribution of large random matrices whose entries are independent Gaussian random variables [85]. This result was later generalized by Girko and is now known as the Circular Law. It says that if the entries J_{ij} are drawn independently from a distribution with zero mean and variance σ^2, then as $S \to \infty$ the eigenvalues uniformly fill a disk of radius $r = \sqrt{S}\sigma$ in the complex plane (see Fig. 5). In particular, the eigenvalue with the largest real part sits on the edge of this disk.

May then performed a simple thought experiment. He started with an ecosystem of S species that do not interact with each other, but where each species limits its own growth (i.e. $J_{ii} = -1$ and $J_{ij} = 0$ for $i \neq j$). All the eigenvalues of this matrix equal -1, so this ecosystem is stable no matter how many species it contains. He then asked what happens when we turn

on interactions between species by drawing the off-diagonal elements J_{ij} at random. The -1's on the diagonal simply shift the disk of eigenvalues to the left by one. Using the Circular Law, we then find that the eigenvalue with the largest real part is

$$\lambda_{\max} \sim -1 + \sqrt{S\sigma^2}. \tag{57}$$

In other words, the ecosystem becomes unstable once the typical combined effect of all the other species on a given species, $\sqrt{S\sigma^2}$, is as large as the self-regulation of that species. This leads to May's famous stability criterion for the largest number of species S^* that can stably coexist:

$$\sqrt{S^*\sigma^2} = 1. \tag{58}$$

Thus, according to May, the more strongly species interact (i.e. the larger σ), the fewer species an ecosystem can support before it becomes unstable.

As is clear from the argument above, a large number of assumptions were made in this derivation. Nonetheless, this result proved to be extremely influential. Before the 1970s, ecologists believed that diversity enhanced ecosystem stability. May's stability criteria challenged this idea and lead to what is now known as the "diversity-stability debate" [47]. Since May's original publications, theorists showed how this criteria could be violated by changing the May's original assumptions, including but not limited to, adding biologically realistic correlation structures [86], modular structures [87], accounting for the fact that the community matrix depends on population sizes [88], and considering high-order interactions Lotka–Volterra dynamics [89].

8 Cavity Method for Lotka–Volterra model

One major shortcoming of May's argument is that he considered purely linearized dynamics. Over the last few years, physicists have extended these ideas to more complex ecological models such including the Generalized Lotka-Volterra model (GLV) and the consumer resource models (CRMs). One crucial difference between May's linear dynamics and these latter two classes of models is that in both GLVs and CRMs, species can go extinct. This gives rise to additional self-organization due to species extinctions that fundamentally change the qualitative picture derived by May.

We begin by outlining the cavity solution to the GLV. Our derivation closely follows [90], though we deploy what we consider to be more transparent notation. We consider a GLV described by S equations of the form

$$\frac{dN_i}{dt} = g_i N_i (r_i - N_i - \sum_j A_{ij} N_j), \quad i = 1, 2, \ldots, S. \tag{59}$$

Here g_i are the intrinsic growth rates, r_i are the carrying capacities and A_{ij} encode inter-species interactions, with positive and negative values representing competition and mutualism interactions, respectively. We are interested in characterizing the steady-state solutions to these equations (i.e. $\frac{dN_i}{dt} = 0$) when the A_{ij} are sampled from a random distribution. In other words we would like to characterize the statistical properties of $\{N_j\}$ that satisfy the fixed point equations

$$0 = N_i (r_i - N_i - \sum_{j \neq i} A_{ij} N_j). \tag{60}$$

To do so, we will make use of the zero temperature cavity method [91]. The cavity method has been successfully used in a wide variety of settings, ranging from spin glasses to computer

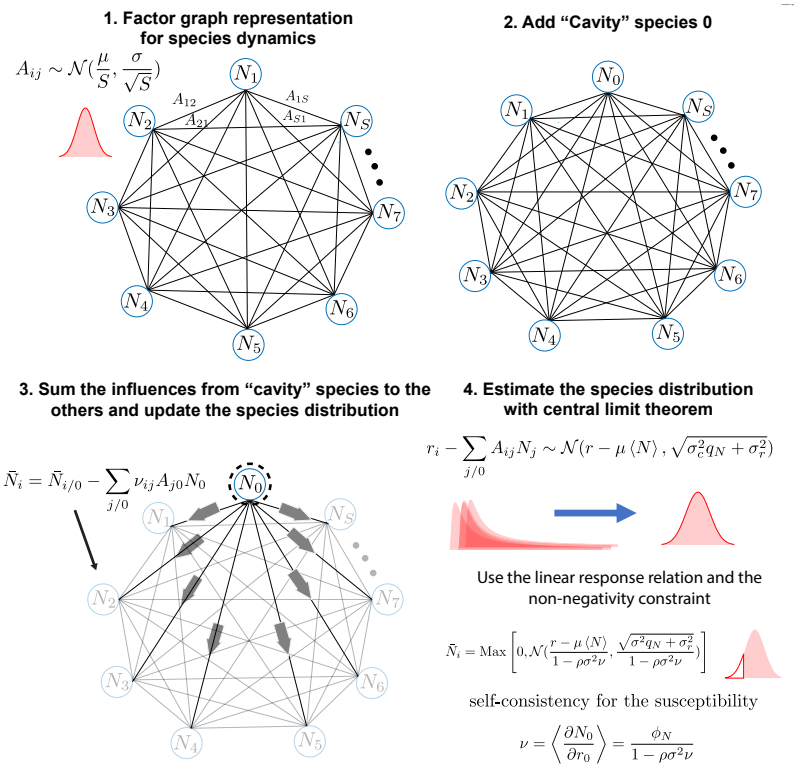


Figure 6: Schematic outlining steps in cavity solution for Lotka–Volterra model. **1.** The species dynamics in eq. (60) are expressed as a factor graph. The edges are bi-directional and sampled from a Gaussian distribution. **2.** Add the "Cavity" species 0 as the perturbation. **3.** Sum the resource abundance perturbations from the "Cavity" species 0 at steady state and update the species abundance distribution to reflect the new steady state. **4.** Employing the central limit theorem and the non-negativity constraint, the species distribution is expressed as a truncated normal distribution. The susceptibility appearing in the species distribution is the self-consistency relation.

science and combinatorial optimization. Over the last five or six years, there has been a concerted effort to develop and generalize the cavity method to solve ecological models with non-negative continuous variables (species and resource abundance) [34, 90, 92, 93].

The basic idea behind the cavity method is to derive self-consistency equations by relating an ecosystem with S species to another ecosystem with S+1 species where the interaction coefficients A_{ij} are drawn from the same random distribution (see Figure 6). In the thermodynamic limit $S \to \infty$, the macroscopic observables of these two systems will be indistinguishable. In other words, statistical observables such as the first and second moment of species abundances $\langle N \rangle$, $\langle N^2 \rangle$ in the two ecosystems, will be identical up to order $\sim 1/S$ corrections which become negligible in the thermodynamic limit.

Technically, we do this by adding a new "cavity" species 0 to the original ecosystem. The addition of this species can be viewed as a small perturbation to the original equilibrium state and hence can be calculated using perturbation theory. This allows us to derive self-consistency equations characterizing the observables of interest. In what follows, we assume that the A_{ij} are sampled from a Gaussian distribution with mean $\langle A_{ij} \rangle = \frac{\mu}{S}$ and variance $var(A_{ij}) = \frac{\sigma^2}{S}$. The Gaussian assumption is not essential. In fact, what is important is that the distribution is not long-tailed with finite first two moments give by the expression above.

In order to employ the cavity method, it is helpful to rewrite

$$A_{ij} = \frac{\mu}{S} + \sigma a_{ij} \tag{61}$$

with

$$\begin{aligned} \langle a_{ij} \rangle &= 0 \\ \langle a_{ij} a_{kl} \rangle &= \frac{1}{S}\delta_{ik}\delta_{jl} + \frac{\rho}{S}\delta_{il}\delta_{jk}, \end{aligned} \tag{62}$$

where $-1 \leq \rho \leq 1$ measures the amount of non-reciprocity between species-species interactions. In particular, $\rho = 1$ corresponds to the case of reciprocal interactions where species i affects species j the same way species j affects species i. The scaling in the mean and variance is chosen to keep the interaction term $\sum_j A_{ij}\tilde{N}_j$ in eq. (60) independent of system size.

The parameters r_i, the carrying capacities in the GLV, are also sampled from a Gaussian distribution with mean r and variance σ_r. Importantly, we assume that the r_i and A_{ij} are uncorrelated. To aid calculations, it is helpful to write

$$r_i = r + \delta r_i, \tag{63}$$

where

$$\begin{aligned} \langle \delta r_i \rangle &= 0 \\ \langle \delta r_i \delta r_j \rangle &= \sigma_r^2 \delta ij. \end{aligned} \tag{64}$$

8.1 The Cavity Solution

To perform the cavity calculation, we perturb the original ecosystem with a new "cavity" species N_0 with interaction coefficient A_{0j} and A_{j0} (see Figure 6). The equations for this new ecosystem with $S+1$ species take the form

$$\begin{aligned} 0 &= N_i(r_i - N_i - \sum_{j\neq i} A_{ij}N_j - A_{i0}N_0) \\ 0 &= N_0(r_0 - N_0 - \sum_{j=1}^{S} A_{0j}N_j), \end{aligned} \tag{65}$$

where for notational simplicity we have dropped the bar indicating steady-state.

The key observation is that since A_{i0} is of order $1/S$, the term $A_{i0}N_0$ is a small perturbation whose effects can be calculated perturbatively. In particular, from the viewpoint of a species i, the addition of species zero is the same as slightly changing r_i, namely $r_i \to r_i + \delta r_i$ with $\delta r_i = -A_{i0}N_0$. Motivated by this observation, we define the susceptibility function

$$\nu_{ij} = \frac{\partial N_i}{\partial r_j}. \tag{66}$$

If we denote the steady state abundances in the original S-species ecosystem by $N_{i/0}$ and the steady state abundances in the $(S+1)$-species ecosystem after the addition of species 0 by N_i, first order perturbation theory tell us that these two quantities are related by the expression

$$N_i \quad = \quad N_{i/0} - \sum_{j=1}^{S} \nu_{ij} A_{j0} N_0. \tag{67}$$

We can now substitute this relationship into the second equation in eq. (65) to get

$$0 = N_0(r_0 - N_0 - \sum_{j=1}^{S} A_{0j}(N_{j/0} - \sum_{i=1}^{S} \nu_{ji} A_{i0} N_0)). \tag{68}$$

Substituting eq. (61) and eq. (63) and solving for N_0 gives two potential solutions over the reals: $N_0 = 0$ and

$$N_0 = \frac{r + \delta r_0 - \sum_{j=1}^{S} \sigma a_{0j} N_{j/0} - \mu \langle N \rangle}{1 - \sigma^2 \sum_{j,i=1}^{S} \nu_{ij} a_{0j} a_{i0}}, \tag{69}$$

where we have defined $\langle N \rangle = \frac{1}{S}\sum_j N_j$ and we have ignored all terms of order $1/S$. Since species abundances must be *positive*, the latter solution may not exist if the resulting expression for N_0 is negative. If we also require that the solution be noninvasible (i.e. that if a positive solution exist we choose that solution), we can solve for N_0 to get

$$N_0 = \max\left(0, \frac{r + \delta r_0 - \sum_{j=1}^{S} \sigma a_{0j} N_{j/0} - \mu \langle N \rangle}{1 - \sigma^2 \sum_{j,i=1}^{S} \nu_{ij} a_{0j} a_{i0}}\right) \tag{70}$$

The max function simply takes the maximum value in the bracket. To leading order in $1/S$, we can replace the $a_{0j}a_{i0}$ in the double sum by its expectation values so that

$$\sum_{j,i=1}^{S} \nu_{ij} a_{0j} a_{i0} = \frac{\rho}{S} \sum_j \nu_{jj} = \rho\, \nu, \tag{71}$$

where we have defined the average trace of the susceptibility matrix

$$\nu = \frac{1}{S} \sum_j \nu_{jj}. \tag{72}$$

Substituting this into Eq. 70 , we get

$$N_0 = \max\left(0, \frac{r + \delta r_0 - \sum_{j=1}^{S} \sigma a_{0j} N_{j/0} - \mu \langle N \rangle}{1 - \sigma^2 \rho\, \nu}\right). \tag{73}$$

So far, we have not done anything out of the ordinary. However, at this point we will make use of the power of "self-averaging" and typicality. Instead, of characterizing the solutions

to Eq. 70 for a particular choice of δr_i and a_{ij}, we will instead ask about the property of solutions "averaged" over the different realizations of these variables. In other words, we imagine solving Eq. 70 for many different choices of a_{ij} and r_i, where these variables are all drawn from the same distribution, and ask about the distribution of solutions of N_0 for these different choices. Rather than focus on the full distribution, we will characterize the distribution by three quantities: the fraction of times that N_0 is non-zero, ϕ_{N_0}, the mean value of the species abundance $\langle N_0 \rangle$, and the second moment of the species distribution $\langle N_0^2 \rangle$. Here, the brackets denote averages over different realizations of the random interactions a_{ij} and δr_i.

Since we are interested in the behavior of solutions over different realizations of the random interactions, we can use the central limit theorem to further simplify Eq. 70. It is clear that if we make a plot of the the sum $\sum_{j=1}^{S} \sigma a_{0j} N_{j/0}$ in the numerator over different realizations of the a_{0j}, it will be have like a Gaussian random variable with mean

$$\langle \sum_{j=1}^{S} \sigma a_{0j} N_{j/0} \rangle = S\sigma \langle a_{0j} \rangle \langle N \rangle = 0, \tag{74}$$

and variance

$$\langle (\sum_{j=1}^{S} \sigma a_{0j} N_{i/0})^2 \rangle = \sigma^2 \sum_{jk} \langle a_{0j} a_{0k} \rangle N_{j/0} N_{k/0} = \sigma^2 \langle N^2 \rangle \tag{75}$$

where we have made use of Eq. 62 and defined the quantities

$$\begin{aligned} \langle N \rangle &= \frac{1}{S} \sum_i N_{i/0} \\ \langle N^2 \rangle &= \frac{1}{S} \sum_i N_{i/0}^2. \end{aligned} \tag{76}$$

Furthermore, by assumption δr_i is a random normal variable with variance σ_r^2 that is uncorrelated with a_{ij}. Hence, the quantity $\sum_{j=1}^{S} \sigma a_{0j} N_{i/0} + \delta r_i$ is also a random normal variable with mean 0 and variance given by the sum of the variances of the individual terms: $\sigma_r^2 + \sigma^2 \langle N^2 \rangle$.

We can combine these observations with eq. (73) to write an expression for the distribution of N_0 over different realizations of a_{ij} and δr_i. Namely, if we define z_N to be a standard normal random variable with mean 0 and standard deviation 1, then the distribution of solutions for N_0 takes the form

$$N_0 = \max\left(0, \frac{r - \mu \langle N \rangle + \sqrt{\sigma_r^2 + \sigma^2 \langle N^2 \rangle} z_N}{1 - \rho \sigma^2 \nu}\right). \tag{77}$$

Notice that the the right hand argument in the max

$$\frac{r - \mu \langle N \rangle + \sqrt{\sigma_r^2 + \sigma^2 \langle N^2 \rangle} z_N}{1 - \rho \sigma^2 \nu} \tag{78}$$

is just the formula for a Gaussian distribution with mean $(r - \mu \langle N \rangle)/(1 - \rho \sigma^2 \nu)$ and variance $(\sigma_r^2 + \sigma^2 \langle N^2 \rangle)/(1 - \rho \sigma^2 \nu)^2$. Thus, the full distribution of N_0 given in Eq. 77 is a truncated Gaussian. A truncated Gaussian can be completely characterized by three quantities: the probability that the new species survives ϕ_{N_0} (i.e. the probability that $N_0 > 0$) and its first two moments, $\langle N_0 \rangle$ and $\langle N_0^2 \rangle$.

Note that Eq. 77 still depends on three unknown quantities: the average trace of the susceptibility ν, as well as the first and second moments of the species distributions $\langle N \rangle$ and $\langle N^2 \rangle$.

We emphasize that $\langle N \rangle$ and $\langle N^2 \rangle$ are moments over all species in an ecosystem for a *single* realization of the random parameters a_{ij} and r_i. The same holds true for ν. In order to make progress, we now invoke the idea of self-averaging. Namely, the statement that as the number of species S goes to infinity, the statistical distribution characterizing N_i for single realization of the random parameters should be identical to the distribution characterizing N_0 over many different realizations of the random parameters. In the language of the statistical physics of disordered systems, we focus on the replica symmetric solution.

Self-averaging and replica symmetry follow from the assumptions that: (i) the new species we have introduced is statistically indistinguishable from the original S species and (ii) as S gets large, we can think of the entire ecosystem as being composed of smaller ecosystems, each with a different independent realization of the random parameters. As we will briefly discuss below, the replica symmetric solution is not always valid and more complicated things can happen, especially as one increases the variance of the underlying distribution from which the parameters are drawn. However, the replica-symmetric solution always serves as a good starting point for understanding the behavior of high-dimensional ecosystems.

Assuming replica symmetry and self-averaging, we can equate expectations of N_0 computed using the truncated Gaussian distribution in Eq. 77, with averages calculated over all species from a single realization. This allows us to derive self-consistency equations of the form:

$$\begin{aligned}
\phi_{N_0} = \phi_N &= \frac{1}{S}\sum_i^S \theta(N_{i/0}) \\
\langle N_0 \rangle = \langle N \rangle &= \frac{1}{S}\sum_i^S N_{i/0} \\
\langle N_0^2 \rangle = \langle N \rangle &= \frac{1}{S}\sum_i^S N_{i/0}^2 \\
\left\langle \frac{\partial N_0}{\partial r_0} \right\rangle &= \frac{1}{S}\sum_i \nu_{ii},
\end{aligned} \tag{79}$$

where in the last line we have used the definition of susceptibility Eqs. 66 and 72. Explicitly differentiating Eq. 77 with respect to r_0, this las self-consistency relation yields

$$\nu = \left\langle \frac{\partial N_0}{\partial r_0} \right\rangle = \frac{\phi_N}{1-\rho\sigma^2\nu}. \tag{80}$$

where $\phi_N = \frac{S^*}{S}$ is the fraction of nonzero species in the ecosystem. Note that ϕ_N results from the fact that if $N_0 = 0$, the corresponding derivative is zero.

In order to write the first three self-consistent equations more explicitly, it is helpful to define the following notation. Let $y = \max\left(0, \frac{c}{b}z + \frac{a}{b}\right)$, with z being a Gaussian random variable with zero mean and unit variance. Then its j-th moment is given by

$$\begin{aligned}
\left\langle y^j \right\rangle = & \quad \frac{1}{\sqrt{2\pi}} \int_{-\frac{a}{c}}^{\infty} e^{-\frac{x^2}{2}} \left(\frac{c}{b}x + \frac{a}{b}\right)^j dx \\
= & \quad \left(\frac{c}{b}\right)^j \frac{1}{\sqrt{2\pi}} \int_{-\frac{a}{c}}^{\infty} e^{-\frac{x^2}{2}} \left(x + \frac{a}{c}\right)^j dx \\
= & \quad \left(\frac{c}{b}\right)^j w_j(\frac{a}{c}),
\end{aligned} \tag{81}$$

where in the last line we have defined the functions

$$w_j(\frac{a}{c}) = \frac{1}{\sqrt{2\pi}} \int_{-\frac{a}{c}}^{\infty} e^{-\frac{z^2}{2}} \left(z + \frac{a}{c}\right)^j dz. \tag{82}$$

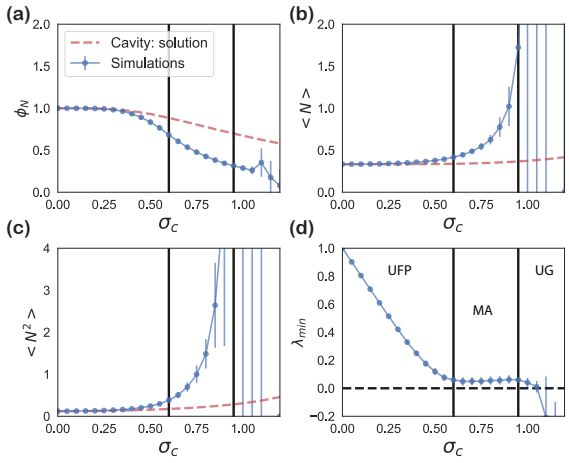


Figure 7: Comparison between the cavity solution (equation 83 - 85) and simulations for symmetric interactions ($\rho = 1$), (a) the fraction of surviving species $\phi_N = \frac{S^*}{S}$ and (b) the first moment $\langle N \rangle$ and (c) the second moment $\langle N^2 \rangle$ of the species distributions as a function of σ_c. (d) The minimum eigenvalue of the submatrix A^*_{ij} at at different σ_c. The error bar shows the standard deviation from 500 numerical simulations with $S = 200$, $\mu = 2.$, $r = 1.$, $\sigma_r = 0.1$ and $\rho = 1$.The black solid lines separate the results in three different regimes: unique fixed point, multiple attractors and unbound growth.The black dashed line in Panel (d) is the base line $\lambda = 0$ for comparison.

In terms of the w_j, the self-consistency equations take the form

$$\nu = \frac{\phi_N}{1-\rho\sigma^2\nu}, \quad \phi_N = w_0\left(\frac{r-\mu\langle N\rangle}{\sqrt{\sigma_r^2+\sigma^2\langle N^2\rangle}}\right), \tag{83}$$

$$\langle N\rangle = \frac{\sqrt{\sigma_r^2+\sigma^2\langle N^2\rangle}}{1-\rho\sigma^2\nu} w_1\left(\frac{r-\mu\langle N\rangle}{\sqrt{\sigma_r^2+\sigma^2\langle N^2\rangle}}\right), \tag{84}$$

$$\langle N^2\rangle = \frac{\sigma_r^2+\sigma^2\langle N^2\rangle}{(1-\rho\sigma^2\nu)^2} w_2\left(\frac{r-\mu\langle N\rangle}{\sqrt{\sigma_r^2+\sigma^2\langle N^2\rangle}}\right). \tag{85}$$

These self-consistency equations can be solved numerically in Mathematica. Fig. 7 and Fig. 8 show the comparison between the cavity solution and 500 independent numerical simulations for various ecosystem properties for symmetric ($\rho = 1$) and uncorrelated ($\rho = 0$) interactions A_{ij}. As can be seen in the figures, our analytic expressions agree remarkably well over a large range of σ_c. Notice, for the ecosystem with symmetric interactions (Fig. 7), the numerics and analytic expression start to disagree for $\sigma_c \sim 0.5$.

8.2 Beyond Replica Symmetry

The underlying reason for the disagreement between the analytic predictions and the numerical simulations for ecosystems with symmetric interactions is that the replica symmetric solution becomes unstable for $\sigma_c \sim 0.5$, indicating a phase transition to the "replica-symmetry broken" phase. The breaking of replica symmetry is actually closely related to May's original arguments about stability of ecosystems.

To see this, it is helpful to recall that one important criteria for thinking about whether a system is stable is to ask how sensitively it responds to perturbations in external parameters. We emphasize that this criteria is closely related to, but different from, asking about the stability of ecosystems to perturbing the species abundances. It is conceptually helpful to consider

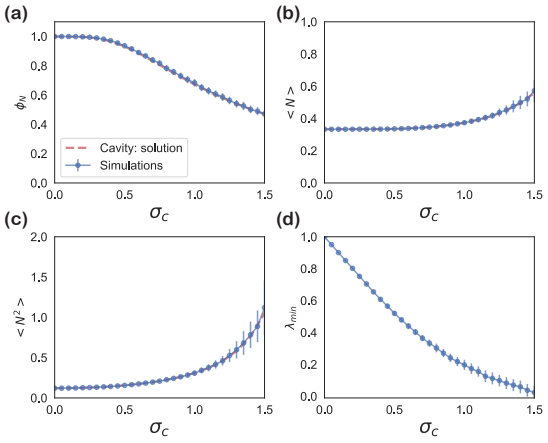


Figure 8: Comparison between the cavity solution (equation 83 - 85) and simulations for uncorrelated interactions, $\rho = 0$, (a) the fraction of surviving species $\phi_N = \frac{S^*}{S}$ and (b) the first moment $\langle N \rangle$ and (c) the second moment $\langle N^2 \rangle$ of the species distributions as a function of σ_c. (d) The minimum eigenvalue of the submatrix A^*_{ij} at at different σ_c. The error bar shows the standard deviation from 500 numerical simulations with $S = 200$, $\mu = 2.$, $r = 1.$, $\sigma_r = 0.1$ and $\rho = 0$.

this distinction in greater mathematical detail. To do so, we will focus only on the S^* surviving species, rather than all S original species in the regional species pool. From Eq. 60, we know at steady-state we must have (where i runs over the S^* surviving species)

$$0 = \frac{dN_i}{dt} = N_i(r_i - \sum_{k=1}^{S^*} A^*_{ik} N_k), \tag{86}$$

where A^*_{ij} denotes the $S^* \times S^*$ interaction sub-matrix of the full interaction matrix A_{ij} restricted to surviving species and the index i runs over the S^* surviving species.

Then, the stability to perturbations in species abundances (the kind of stability considered by May), can be calculated by linearizing this equation to get an equation for the deviations δN_i from the steady-state abundances N_i:

$$\frac{d\delta N_i}{dt} = -\sum_{k=1}^{S^*} N_i A^*_{ik} \delta N_k, \tag{87}$$

where δ_{ik} is the Kronecker Delta function (i.e. $S^* \times S^*$ identity matrix). This is the analogue of Eq. 56. Notice that this stability is set by the eigenvalues of the matrix $-N_i A^*_{ij}$ which unlike the case May considered depends on the steady-state abundances at the steady-state.

For this reason, it will be helpful to work with a slightly different quantity which ask about the sensitivity of steady-state abundances N_i to small changes in the external parameters r_j. This is precisely the susceptibility, Eq. 66, but now restricted to surviving species. We denote this restricted susceptibility by ν^*_{ij}. We can solve for it by dividing Eq. 86 by N_i and then differentiating by r_j to get

$$0 = \delta^*_{ij} - \sum_{j=1}^{S^*} A^*_{ik} \frac{dN_k}{dr_j}, \tag{88}$$

where the Kronecker-delta function δ^*_{ij} comes from the fact that $\frac{\partial r_i}{\partial r_j} = \delta_{ij}$. Inverting this equation gives

$$\nu^*_{ij} = \frac{dN_i}{dr_j} = (A^*_{ij})^{-1} \tag{89}$$

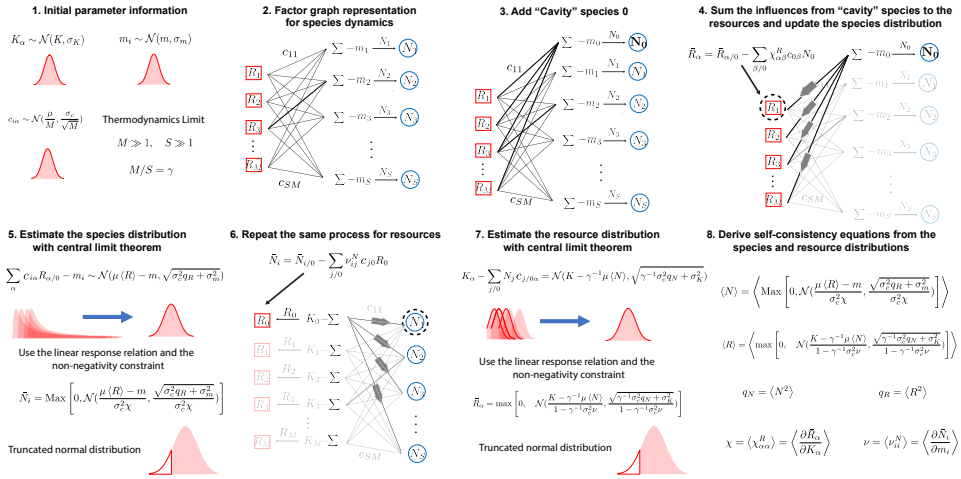


Figure 9: Schematic outlining steps in cavity solution. **1.** The initial parameter information consists of the probability distributions for the mechanistic parameters: K_α, m_i and $C_{i\alpha}$. We assume they can be described by their first and second moments. **2.** The species dynamics $N_i(\sum_\alpha C_{i\alpha}R_\alpha - m_i)$ in eqs. (90) are expressed as a factor graph. **3.** Add the "Cavity" species 0 as the perturbation. **4.** Sum the resource abundance perturbations from the "Cavity" species 0 at steady state and update the species abundance distribution to reflect the new steady state. **5.** Employing the central limit theorem and the non-negativity constraint, the species distribution is expressed as a truncated normal distribution. **6.** Repeat **Step 2-4** for the resources. **7.** The resource distribution is also expressed as a truncated normal distribution. **8.** The self-consistency equations are obtained from the species and resource distributions. Figure taken from Ref. [59], © 2020 American Physical Society.

It turns out that RS breaking corresponds to the case where A^*_{ij} develops a zero eigenvalue, indicating that the susceptibility diverges. In other words, the system becomes infinitely sensitive to external perturbations.

Fig. 7 (D) shows the minimum eigenvalue of A^*_{ij} as σ_c is varied. The minimum eigenvalue decrease monotonically with increasing σ_c until it is close to zero. And then the cavity solution fails. This correspond to the emergence of multiple attractors (MA phase). Further increasing σ_c results in unbounded growth. The full phase diagram and consequences of this observations were first worked out in [90]. To analyze these phases, one can recast the problem using the replica approach with replica symmetry breaking [94].

One of the most striking things about the GLV model is that the minimum eigenvalue of A^*_{ij} gets pinned exactly to zero over an extended region of the parameter space (region indicated by MA in Fig. 7 (D)). This corresponds to the phenomena of marginal stability. In other words, though the system is not unstable, it is infinitely sensitive to small changes in external parameters. This marginal stability is unique to the symmetric GLV model and is an indication that the ecosystem self-organizes through species extinction. Generically with non-reciprocal interactions, the marginal stable phase disappears and is replaced by a dynamic phase that can exhibit chaotic fluctuations [25–30].

9 Cavity Method for MacArthur Consumer-Resource Models

Just as we can use the cavity method to analyze the Lotka-Volterra model in high-dimensions, we can also adapt the cavity method to analyze consumer resource models (CRMs) in high

dimensions. We will be interested in developing a statistical understanding of CRMs in the limit where the number of resources M and the number of species S in the regional species is large. In these lectures, we limit our analysis to MacArthur's original model (eq. (6) and eq. (7)). For simplification, we set $w_\alpha = 1$ and $r_\alpha = K_\alpha$ and get the following dynamics:

$$\frac{dN_i}{dt} = N_i(\sum_\beta c_{i\beta}R_\beta - m_i), \tag{90}$$

$$\frac{dR_\alpha}{dt} = R_\alpha(K_\alpha - R_\alpha - \sum_j N_j c_{j\alpha}). \tag{91}$$

The basic idea behind the cavity method is to derive self-consistency equations by relating an ecosystem with S species and M resources to another ecosystem with $S+1$ species and $M+1$ species. Unlike in the GLV model were we just had to one new species to the ecosystem, here it is important that we add a new species N_0 and new resource R_0 to the ecosystem (see Figure 9). We once again will work in the thermodynamic limit, $S, M \to \infty$, but with the ratio $S/M = \gamma^{-1}$ fixed. As for the GLV model, we further assume that the system is self-averaging and can be described using a "replica symmetric" ansatz. With these assumptions, just as in the GLV model, we will use perturbation theory to relate solutions in the (S, M)-ecosystem to the $(S+1, M+1)$-ecosystem. The logic behind the cavity solution of the MacArthur CRM is very similar to the derivations for the GLV model and for this reason this section is more condensed than the rest of these lectures.

The basic steps needed to derive the cavity solution are shown in Fig. 9 and outlined below. **Step 1**: the consumer preferences, $c_{i\alpha}$, are drawn from a Gaussian distribution with mean μ/M and variance σ_c^2/M. They can be deposed into a deterministic and fluctuating component:

$$c_{i\alpha} = \mu/M + \sigma_c d_{i\alpha}, \tag{92}$$

where the fluctuating part $d_{i\alpha}$ obeys the statistics

$$\langle d_{i\alpha} \rangle = 0 \tag{93}$$

$$\left\langle d_{i\alpha} d_{j\beta} \right\rangle = \frac{\delta_{ij}\delta_{\alpha\beta}}{M}. \tag{94}$$

We also assume that both the carrying capacity K_α and the minimum maintenance cost m_i are independent Gaussian random variables with mean and covariance given by

$$\langle K_\alpha \rangle = K \tag{95}$$

$$\text{Cov}(K_\alpha, K_\beta) = \delta_{\alpha\beta}\sigma_K^2 \tag{96}$$

$$\langle m_i \rangle = m \tag{97}$$

$$\text{Cov}(m_i, m_j) = \delta_{ij}\sigma_m^2. \tag{98}$$

We also define $\langle R \rangle = \frac{1}{M}\sum_\beta R_\beta$ and $\langle N \rangle = \frac{1}{S}\sum_j N_j$ to be the average resource and average species abundance, respectively. In terms of these quantities, we can re-write eq. (90) and eq. (91) as

$$\frac{dN_i}{dt} = N_i(\mu \langle R \rangle - m + \sum_\beta \sigma_c d_{i\beta}R_\beta - \delta m_i) \tag{99}$$

$$\frac{dR_\alpha}{dt} = R_\alpha(K + \delta K_\alpha - R_\alpha - \gamma^{-1}\mu \langle N \rangle - \sum_j \sigma_c d_{j\alpha}N_j) \tag{100}$$

where $\delta K_\alpha = K_\alpha - K, \delta m_i = m_i - m$ and $\gamma^{-1} = S/M$.

In **Step 2**, we express eq. (100) as a bipartite factor graph model for visualization. At **step 3**, we add a "cavity" species N_0 and a "cavity" resource R_0 into the ecosystem,

$$\frac{dN_0}{dt} = N_0(\mu \langle R\rangle - m + \sum_\beta \sigma_c d_{0\beta} R_\beta - \delta m_0) \tag{101}$$

$$\frac{dR_0}{dt} = R_0(K + \delta K_0 - R_0 - \gamma^{-1}\mu \langle N\rangle - \sum_j \sigma_c d_{j0} N_j). \tag{102}$$

Adding new species and resource will perturb the original steady state. To characterize the perturbations, we introduce the following susceptibility matrices:

$$\chi^R_{\alpha\beta} = \frac{\partial R_\alpha}{\partial K_\beta}, \qquad \chi^N_{i\alpha} = \frac{\partial N_i}{\partial K_\alpha}, \tag{103}$$

$$\nu^R_{\alpha i} = \frac{\partial R_\alpha}{\partial m_i}, \qquad \nu^N_{ij} = \frac{\partial N_i}{\partial m_j}. \tag{104}$$

We can express the steady-state species and resource abundances in the $(S+1, M+1)$ system with a first-order Taylor expansion around the (S, M) values. To do so, notice that in the presence of the new species

$$\begin{aligned} 0 &= N_i(\sum_\beta c_{i\beta} R_\beta - (m_i - c_{i0} R_0)), \\ 0 &= R_\alpha((K_\alpha - N_0 c_{0\alpha}) - R_\alpha - \sum_j N_j c_{j\alpha}) \end{aligned} \tag{105}$$

Since in eq. (92), the contribution of the mean, $\frac{\mu}{M} \sim \mathcal{O}(\frac{1}{M})$, is much smaller than the fluctuation term, $\sigma_c d_{i0} \sim \mathcal{O}(\frac{1}{\sqrt{M}})$, to leading order the perturbations to m_i, and K_α in eqs. (100) and (99) are given by $\sigma_c d_{i0} R_0$ and $\sigma_c d_{0\alpha} N_0$, respectively. Namely, in the presence of the new species (to leading order in $1/M$ or $1/S$) one has

$$N_i = N_{i/0} - \sigma_c \sum_{\beta/0} \chi^N_{i\beta} d_{0\beta} N_0 - \sigma_c \sum_{j/0} \nu^N_{ij} d_{j0} R_0 \tag{106}$$

$$R_\alpha = R_{\alpha/0} - \sigma_c \sum_{\beta/0} \chi^R_{\alpha\beta} d_{0\beta} N_0 - \sigma_c \sum_{j/0} \nu^R_{\alpha j} d_{j0} R_0 \tag{107}$$

Note $\sum_{j/0}$ and $\sum_{\beta/0}$ mean the sums exclude the new species 0 and the new resource 0. These are just the analogue of Eqs. 67 for the GLV model The next step is to plug eq. (106) and eq. (107) into eq. (101) and eq. (102) and solve for the steady-state value of N_0 and R_0.

9.1 Self-consistency equations for species

For the new "cavity" species, the steady equation takes the form

$$\begin{aligned} 0 \quad &= N_0(\mu \langle R\rangle - m - \delta m_0 - \sigma_c^2 N_0 \sum_{\alpha/0,\beta/0} \chi^R_{\alpha\beta} d_{0\alpha} d_{0\beta} \\ &\quad - \sigma_c^2 R_0 \sum_{\beta/0, j/0} \nu^R_{\beta j} d_{0\beta} d_{j0} + \sum_{\beta/0} \sigma_c d_{0\beta} R_{\beta/0} + \sigma_c d_{00} R_0). \end{aligned} \tag{108}$$

Notice that each of the sums in this equation is the sum over a large number of uncorrelated random variables, and can therefore be well approximated by Gaussian random variables for large enough M and S.

As in the cavity solution of the GLV, we can replace the double sums over $d_{j0}d_{0\beta}$ and $d_{0\alpha}d_{0\beta}$ by their expectation values using eq. (93) and eq. (94):

$$\sum_{\beta/0,j/0} \nu^R_{\beta j} d_{0\beta} d_{j0} = \frac{1}{M} \sum_{\beta/0,j/0} \nu^R_{\beta j} \delta_{j0}\delta_{0\beta} = 0 \tag{109}$$

$$\sum_{\alpha/0,\beta/0} \chi^R_{\alpha\beta} d_{0\alpha} d_{0\beta} = \frac{1}{M} \sum_{\alpha/0,\beta/0} \chi^R_{\alpha\beta} \delta_{\alpha\beta} = \chi \tag{110}$$

where $\chi = \frac{1}{M}\sum_{\alpha/0,\beta/0} \chi^R_{\alpha\beta}\delta_{\alpha\beta} = \frac{1}{M}\mathrm{Tr}(\chi^R_{\alpha\beta})$ is the average susceptibility. Substituting these expressions into eq. (109) , we obtain

$$\begin{aligned} 0 &= N_0(\mu\langle R\rangle - m - \sigma_c^2 \chi N_0 \\ &+ \sum_{\beta/0} \sigma_c d_{0\beta} R_{\beta/0} - \delta m_0) + \mathcal{O}(M^{-1/2}). \end{aligned} \tag{111}$$

Instead of looking for a solution of N_0 for one realization of the random parameters, we can ask about the statistical distribution of the solution space for many different realizations of the random parameters. In this case, we know that $\sum_{\beta/0}\sigma_c d_{0\beta}R_{\beta/0}$ can be well described by a Gaussian with mean

$$\sigma_c \langle \sum_{\beta/0} d_{0\beta} R_{\beta/0}\rangle = \sum_{\beta/0} \langle d_{0\beta}\rangle R_{\beta/0} = 0 \tag{112}$$

and variance

$$\langle \left(\sum_{\beta/0} \sigma_c d_{0\beta} R_{\beta/0} \right)^2 \rangle = \sigma_c^2 q_R, \tag{113}$$

where q_R is the second moment of the resource distribution,

$$q_R = \langle R^2 \rangle = \frac{1}{M}\sum_{\beta} R^2_{\beta/0}.$$

Introducing an auxiliary random normal variable z_N with zero mean and unit variance, this allows us to replace the sum below with a random variable as follows:

$$\sum_{\beta/0} \sigma_c d_{0\beta} R_{\beta/0} - \delta m_0 = z_N \sqrt{\sigma_c^2 q_R + \sigma_m^2},$$

where we have used the fact that by assumption δm_0 is uncorrelated with $d_{0\beta}$ and $R_{\beta/0}$. We can solve eq. (111) in terms of the quantities just defined to get

$$\mu\langle R\rangle - m - \sigma_c^2 \chi N_0 + \sqrt{\sigma_c^2 q_R + \sigma_m^2} z_N = 0. \tag{114}$$

Rearranging this to solve for N_0 and noting that $N_0 \geq 0$, one gets

$$N_0 = \max\left(0, \quad \frac{\mu\langle R\rangle - m + \sqrt{\sigma_c^2 q_R + \sigma_m^2} z_N}{\sigma_c^2 \chi}\right) \tag{115}$$

which is the formula for a truncated Gaussian. In writing this solution, we have enforced the condition that, if possible, N_0 must be *positive*. Ecologically, this is the statement that we are interested in the statistical description of uninvasible fixed points.

Combining eq. (115) and eq. (81), by invoking self-averaging and replica symmetry we can easily write down the self-consistency equations for the fraction of non-zero species as well as the moments of the species abundances at the steady state:

$$\phi_N = \frac{S^*}{S} = w_0\left(\frac{\mu\langle R\rangle - m}{\sqrt{\sigma_c^2 q_R + \sigma_m^2}}\right) \tag{116}$$

$$\langle N\rangle = \left(\frac{\sqrt{\sigma_c^2 q_R + \sigma_m^2}}{\sigma_c^2\chi}\right) w_1\left(\frac{\mu\langle R\rangle - m}{\sqrt{\sigma_c^2 q_R + \sigma_m^2}}\right) \tag{117}$$

$$q_N = \left(\frac{\sqrt{\sigma_c^2 q_R + \sigma_m^2}}{\sigma_c^2\chi}\right)^2 w_2\left(\frac{\mu\langle R\rangle - m}{\sqrt{\sigma_c^2 q_R + \sigma_m^2}}\right) \tag{118}$$

9.2 Self-consistency equations for resource

We now derive the equations for the steady-state of the resource dynamics. Inserting eq. (107) into eq. (102) gives:

$$0 = R_0(K - R_0 - \gamma^{-1}\mu\langle N\rangle + \sigma_c^2 N_0 \sum_{\beta/0,j/0} \chi^N_{j\beta} d_{j0} d_{0\beta} \tag{119}$$
$$+\sigma_c^2 R_0 \sum_{i/0,j/0} \nu^N_{ij} d_{i0} d_{j0} - \sum_{j/0} \sigma_c d_{j0} N_{j/0} - \sigma_c d_{00} N_0 + \delta K_0)$$

where $\gamma^{-1} = \frac{S}{M}$.

To proceed, we note in the double sums above we can replace the sums over the $d_{i0}d_{j0}$ and $d_{j0}d_{0\beta}$ by their expectation values using eq. (93) and eq. (94) to get

$$\sum_{i/0,j/0} \nu^N_{ij} d_{0i} d_{0j} = \frac{S}{M} \times \frac{1}{S}\mathrm{Tr}(\nu^N_{ij}) = \gamma^{-1}\nu \tag{120}$$

and

$$\sum_{\beta/0,j/0} \chi^N_{j\beta} d_{j0} d_{0\beta} = 0. \tag{121}$$

Furthermore, notice that $\sigma_c d_{00} N_0$ is of order $1/\sqrt{M}$ and can be ignored in the thermodynamic limit. Finally, since we are interested in the statistical properties when we average over random parameters we can replace the sum over $N_{j/0}$ by Gaussian random variable of the form

$$\sum_{j/0} \sigma_c d_{j0} N_{j/0} + \delta K_0 = z_R \sqrt{\sigma_c^2 \gamma^{-1} q_N + \sigma_K^2}, \tag{122}$$

where z_R is unit normal random variable and $q_N = \langle N^2\rangle = \frac{1}{S}\sum_j N_j^2$. In writing this, we have once again used the fact that δK_0 is uncorrelated with d_{j0} and $N_{j/0}$.

We can substitute these equations into eq. (120) and solve for R_0. If we further require that the resource are uninvasible (i.e. we always choose the positive solution for R_0 if it exists), then the distribution of R_0 is described by a truncated Gaussian of the form

$$R_0 = \max\left(0, \frac{K - \gamma^{-1}\mu\langle N\rangle + \sqrt{\sigma_c^2\gamma^{-1} q_N + \sigma_K^2}\, z_R}{1 - \gamma^{-1}\sigma_c^2 \nu}\right). \tag{123}$$

This allows us to write down the self-consistency equations for the fraction of non-zero resources as well as the moments of their abundances at the steady state:

$$\phi_R = \frac{M^*}{M} = w_0\left(\frac{K - \gamma^{-1}\mu\langle N\rangle}{\sqrt{\sigma_c^2\gamma^{-1} q_N + \sigma_K^2}}\right) \tag{124}$$

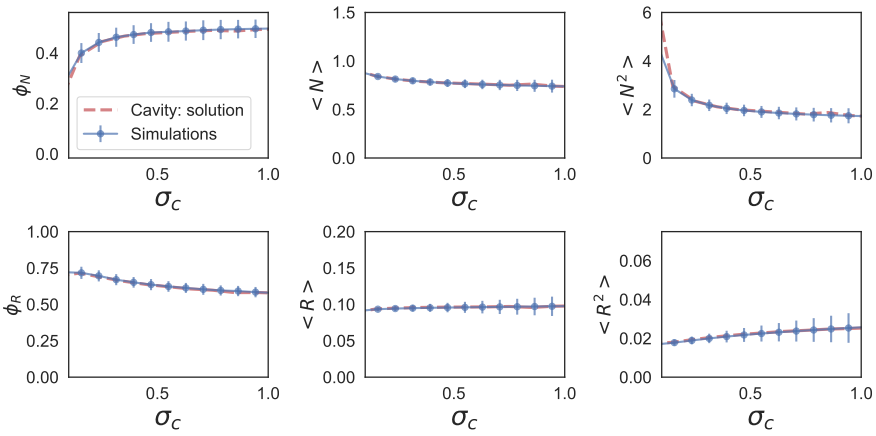


Figure 10: Cavity predictions (obtained by solving the self-consistency equations derived in the text) compared to numerical simulations of MacArthur's consumer resource model with Gaussian consumer preferences. Panels show, as a function of σ_c, the fraction of species that survive, $\phi_N = \frac{S^*}{S}$, the fraction of resources that survive, $\phi_R = \frac{M^*}{M}$, and the first and second moments of the species and resource abundances. The error bar shows the standard deviation from 100 numerical simulations with $M = S = 100$ $\mu = 1.$, $K = 1.$, $\sigma_K = 0.1$, $m = 0.1$, σ_m=0.01. Simulations were run using the CVXPY package [95].

$$\langle R\rangle = \left(\frac{\sqrt{\sigma_c^2\gamma^{-1}q_N + \sigma_K^2}}{1-\gamma^{-1}\sigma_c^2\nu}\right) w_1\left(\frac{K-\gamma^{-1}\mu\langle N\rangle}{\sqrt{\sigma_c^2\gamma^{-1}q_N+\sigma_K^2}}\right) \tag{125}$$

$$q_R = \left(\frac{\sqrt{\sigma_c^2\gamma^{-1}q_N + \sigma_K^2}}{1-\gamma^{-1}\sigma_c^2\nu}\right)^2 w_2\left(\frac{K-\gamma^{-1}\mu\langle N\rangle}{\sqrt{\sigma_c^2\gamma^{-1}q_N+\sigma_K^2}}\right). \tag{126}$$

One can also use these equations to solve for the trace of the susceptibilities. The susceptibilities are given by averaging ν_{ii}^N and $\chi_{\alpha\alpha}^R$,

$$\chi = \left\langle \frac{\partial R_\alpha}{\partial K_\alpha}\right\rangle = \frac{\partial\langle R\rangle}{\partial K} = \frac{\phi_R}{1-\gamma^{-1}\sigma_c^2\nu}, \tag{127}$$

$$\nu = \left\langle \frac{\partial N_i}{\partial m_i}\right\rangle = \frac{\partial\langle N\rangle}{\partial m} = -\frac{\phi_N}{\sigma_c^2\chi} \tag{128}$$

Solving above two equations yields

$$\chi = \phi_R - \gamma^{-1}\phi_N, \quad \nu = -\frac{1}{\sigma_c^2}\frac{\phi_N}{\phi_R-\gamma^{-1}\phi_N}. \tag{129}$$

Collectively Eqns 116 -118, 124- 126, and 129 define 8 self-consistency equations for the 8 quantities ϕ_N, $\langle N\rangle$, $\langle N^2\rangle$, ϕ_R, $\langle R\rangle$, $\langle R^2\rangle$, χ, and ν that can be solved numerically.

Fig. 10 shows a comparison between the cavity solution and 1000 independent numerical simulations for various ecosystem properties such as the fraction of surviving species S^*/S, the fraction of surviving resources M^*/M, and the first and second moment of the species and resource distributions. As can be seen in the figure, our analytic expressions agree remarkably well over a large range of σ_c.

As a further check on our analytic solution, we ran simulations where the $c_{i\alpha}$ were drawn from different distributions. One pathology of choosing $c_{i\alpha}$ from a Gaussian distribution is

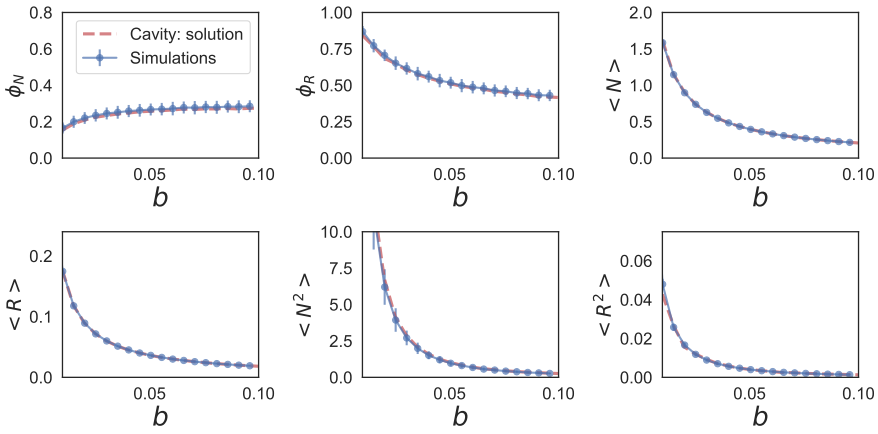


Figure 11: Comparison between cavity solutions and simulations for MacArthur's consumer resource model. The consumer matrix is sampled from an uniform distribution $\mathcal{U}(0, b)$. The error bar shows the standard deviation from 500 numerical simulations with $M = S = 100$ $\mu = 1.$, $K = 1.$, $\sigma_K = 0.1$, $m = 0.1$, $\sigma_m = 0.01$. Simulations were run using the CVXPY package [95].

that when σ_c is large, many of consumption coefficients are negative. Another peculiarity of a Gaussian is that every species interacts with every other species To test whether our cavity solution still describes ecosystems when $c_{i\alpha}$ are strictly positive or sparse, we compare our cavity solution to simulations where the $c_{i\alpha}$ are drawn from either a Bernoulli or uniform distribution. The results are shown in Figs. 11 and 12. As before, there is remarkable agreement between theoretical predictions and numerical simulations. The fact that our analytics also describe simulations from these alternative distributions highlights the incredible power of the central limit theorem and the cavity method. In fact, we expect that many properties diverse ecosystems can be modeled using random ecosystems even when the consumer preferences have considerable structure, an idea explored using the cavity method and random matrix theory here [96].

10 Conclusions

The goal of these lectures was to familiarize the reader with recent work at the intersection of statistical physics and community ecology. The bulk of the lectures focused on the two main models of community ecology: the generalized Lotka-Volterra model (GLV) and MacArthur's Consumer Resource Models (MCRM). We first presented these models in low-dimensions, where graphical methods could be used to develop intuition and understanding. We then discussed the deep connection between ecosystems with reciprocal interactions and constrained optimization before proceeding to analyze these same models in "high-dimensions" using methods from the statistical physics.

We hope that these lectures can serve as a compact point of reference for those interested in community ecology. While many of the results presented here are scattered in the literature, it has been our experience there is a dearth of comprehensive pedagogical resources for beginners. It is our sincere hope that these lectures can fill this void. As emphasized in the introduction, in no way are these lectures comprehensive. Instead, they should be viewed as a good starting point for delving into this rich and exciting research area.

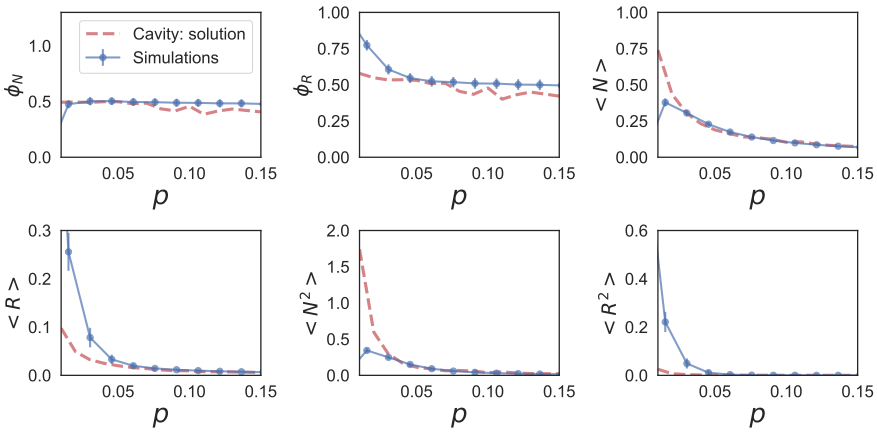


Figure 12: Comparison between cavity solutions and simulations for MacArthur's consumer resource model. The consumer matrix is sampled from a Bernoulli distribution $Bernoulli(p)$. The error bar shows the standard deviation from 500 numerical simulations with $M = S = 100$ $\mu = 1.$, $K = 1.$, $\sigma_K = 0.1$, $m = 0.1$, $\sigma_m = 0.01$. Simulations were run using the CVXPY package [95].

Function	Choice 1	Choice 2
g_i	Substitutable	Non-substitutable
$q_{i\alpha}$	Mass action	Marginal benefit
f_α	Self-renewing	Externally supplied

Table 1: Table of modeling choices

11 Acknowledgments

We would like to thank the Mehta group (especially Zhijie "Sarah" Feng and Emmy Blumenthal) and our collaborators for countless conversations and insights. We are extremely grateful to Arvind Murugan for his deep engagement with the work presented here, including critically reading an early version of these lectures and his many suggestions for improving the presentation. We would also like to thank Akshit Goyal for helpful comments on the presentation. This work was supported by NIH NIGMS R35GM119461 and a CZI Theory grant to PM.

A Consumer resource models

While all consumer resource models are broadly based on one idea - consumer growth depends on resource availability and consumer growth impacts resource availability, the mathematical forms can vary significantly based on detailed assumptions made about the nature of consumer growth, resource consumption by consumers and resource renewal. Although in principle g_i, $q_{i\alpha}$ and f_α are chosen from an infinite variety of possible functions, in practice there are only a few forms that are commonly used. We will consider two possibilities for each, generating eight possible models, tabulated in Table 1.

$n \to -\infty$	Essential
$n < 0$	Interactively essential
$0 < n < 1$	Complementary
$n = 1$	Substitutable
$n > 1$	Antagonistic
$n \to +\infty$	Switching

Table 2: Taxonomy of resource interactions

A.1 Substitutable Resources

For substitutable resources, the growth rate takes the form

$$g_i(\mathbf{R}) = \sum_\alpha w_\alpha c_{i\alpha} R_\alpha - m_i \tag{130}$$

where $c_{i\alpha}$ is a set of foraging preferences, w_α represents the value of each resource in the "common currency" (e.g., energy), and m_i is a minimal amount of consumption required to sustain growth.

A.2 Nonsubstitutable Resources

A general form for the growth rate with nonsubstitutable resources was given by Schreiber and Tobiason (2003):

$$g_i(\mathbf{R}) = \left[\sum_\alpha w_\alpha (c_{i\alpha} R_\alpha)^n\right]^{\frac{1}{n}} - m_i \tag{131}$$

where the exponent n can be any real number. Different values of n correspond to different ecological categories of resources, as shown in Table 2 (cf. Koffel 2016). The weights w_α are typically set to 1, but are included so that this model reduces exactly to the above substitutable model when $n = 1$.

A.3 Mass action

The simplest and most common way of modeling the impact vectors is to assume mass action kinetics, where the depletion rate of resource α by consumer species i is proportional to the rate of encounters between that resource and that species. This gives:

$$q_{i\alpha}(\mathbf{R}) = -\nu_{i\alpha} R_\alpha \tag{132}$$

for some set of constants $\nu_{i\alpha}$ with units of $[\text{consumer density}]^{-1} \times [\text{time}]^{-1}$.

A.4 Marginal benefit foraging

(see [69], p. 171)

The resource depletion rate in the mass action model is not related in any definite way to the growth rates. But in real systems, these two quantities are clearly related, because the growth is sustained by uptake of resources. One way of modeling this relationship is with marginal benefit foraging, where the relative foraging effort invested in each resource type is proportional to the partial derivative of the growth rate with respect to the (value-weighted) abundance of that resource. Instead of arbitrary constants $\nu_{i\alpha}$ determining the encounter rates

between consumers and resources, the rates are now tied to the growth law by the following formula:

$$q_{i\alpha}(\mathbf{R}) = -\frac{a_i}{w_\alpha}\frac{\partial g_i}{\partial R_\alpha}R_\alpha \tag{133}$$

where a_i is a constant that sets the overall encounter rate, which we will always take to be equal to 1.

A.5 Self-renewing resources

In MacArthur's original consumer resource model, he considered biotic resources such as plants serving as food for herbivores, and modeled the intrinsic resource dynamics with a logistic law:

$$f_\alpha(\mathbf{R}) = \frac{r_\alpha}{K_\alpha}R_\alpha(K_\alpha - R_\alpha) \tag{134}$$

where K_α is the carrying capacity and r_α is the low-density growth rate.

A.6 Externally supplied resources

When the resources are not alive, it is usually more physically reasonable to model them as diffusing in from some external source at some rate κ_α, and decaying or diluting at a fixed rate ω_α, leading to the linear law

$$f_\alpha(\mathbf{R}) = \kappa_\alpha - \omega_\alpha R_\alpha. \tag{135}$$

A.7 Explicit equations for all models

A.7.1 Substitutable, Mass action, Self-renewing

$$\frac{dN_i}{dt} = N_i\left[\sum_\beta w_\beta c_{i\beta}R_\beta - m_i\right] \tag{136}$$

$$\frac{dR_\alpha}{dt} = \frac{r_\alpha}{K_\alpha}R_\alpha(K_\alpha - R_\alpha) - \sum_j v_{j\alpha}N_jR_\alpha. \tag{137}$$

A.7.2 Substitutable, Mass action, Externally supplied

$$\frac{dN_i}{dt} = N_i\left[\sum_\beta w_\beta c_{i\beta}R_\beta - m_i\right] \tag{138}$$

$$\frac{dR_\alpha}{dt} = \kappa_\alpha - \omega_\alpha R_\alpha - \sum_j v_{j\alpha}N_jR_\alpha. \tag{139}$$

A.7.3 Nonsubstitutable, Mass action, Self-renewing

$$\frac{dN_i}{dt} = N_i\left\{\left[\sum_\beta w_\beta (c_{i\beta}R_\beta)^n\right]^{\frac{1}{n}} - m_i\right\} \tag{140}$$

$$\frac{dR_\alpha}{dt} = \frac{r_\alpha}{K_\alpha}R_\alpha(K_\alpha - R_\alpha) - \sum_j v_{j\alpha}N_jR_\alpha \tag{141}$$

A.7.4 Nonsubstitutable, Mass action, Externally supplied

$$\frac{dN_i}{dt} = N_i \left\{ \left[\sum_\beta w_\beta (c_{i\beta} R_\beta)^n \right]^{\frac{1}{n}} - m_i \right\} \tag{142}$$

$$\frac{dR_\alpha}{dt} = \kappa_\alpha - \omega_\alpha R_\alpha - \sum_j \nu_{j\alpha} N_j R_\alpha \tag{143}$$

A.7.5 Substitutable, Marginal benefit, Self-renewing

$$\frac{dN_i}{dt} = N_i \left[\sum_\beta w_\beta c_{i\beta} R_\beta - m_i \right] \tag{144}$$

$$\frac{dR_\alpha}{dt} = \frac{r_\alpha}{K_\alpha} R_\alpha (K_\alpha - R_\alpha) - \sum_j c_{j\alpha} N_j R_\alpha. \tag{145}$$

A.7.6 Substitutable, Marginal benefit, Externally supplied

$$\frac{dN_i}{dt} = N_i \left[\sum_\beta w_\beta c_{i\beta} R_\beta - m_i \right] \tag{146}$$

$$\frac{dR_\alpha}{dt} = \kappa_\alpha - \omega_\alpha R_\alpha - \sum_j c_{j\alpha} N_j R_\alpha. \tag{147}$$

A.7.7 Nonsubstitutable, Marginal benefit, Self-renewing

$$\frac{dN_i}{dt} = N_i \left\{ \left[\sum_\beta w_\beta (c_{i\beta} R_\beta)^n \right]^{\frac{1}{n}} - m_i \right\} \tag{148}$$

$$\frac{dR_\alpha}{dt} = \frac{r_\alpha}{K_\alpha} R_\alpha (K_\alpha - R_\alpha) - \sum_j \left[\sum_\alpha w_\alpha (c_{j\alpha} R_\alpha)^n \right]^{\frac{1-n}{n}} c_{j\alpha}^n R_\alpha^{n-1} N_j R_\alpha. \tag{149}$$

A.7.8 Nonsubstitutable, Marginal benefit, Externally supplied

$$\frac{dN_i}{dt} = N_i \left\{ \left[\sum_\beta w_\beta (c_{i\beta} R_\beta)^n \right]^{\frac{1}{n}} - m_i \right\} \tag{150}$$

$$\frac{dR_\alpha}{dt} = \kappa_\alpha - \omega_\alpha R_\alpha - \sum_j \left[\sum_\alpha w_\alpha (c_{j\alpha} R_\alpha)^n \right]^{\frac{1-n}{n}} c_{j\alpha}^n R_\alpha^{n-1} N_j R_\alpha. \tag{151}$$

B Self-limitation and stability

We can derive conditions for stability in the range of models described above. While the detailed analysis differs from case to case, qualitatively, the stability criterion is that each species must limit itself more than it limits others. The typical formulation of this in the niche theory framework is that each species consumes most of the resource that is most limiting to it. Here we generalize this statement to arbitrary numbers of species and resources.

Here, we first consider substitutable resources but different resource replenishment dynamics and impact vectors $q_{i\alpha}$. We show that non-substitutable resources can be handled by locally expanding the contours

We fill focus on the basic example of MCRM with two resources and two species, but allowing an arbitrary impact vector $v_{i\alpha}R_\alpha$ that need not be related to $c_{i\alpha}$:

$$\frac{dN_i}{dt} = N_i\left[\sum_\alpha c_{i\alpha}R_\alpha - m_i\right] \tag{152}$$

$$\frac{dR_\alpha}{dt} = R_\alpha\left[\frac{r_\alpha}{K_\alpha}(K_\alpha - R_\alpha) - \sum_j N_j v_{j\alpha}\right]. \tag{153}$$

The easiest thing to do is move to the "fast resource limit" (see below) and check the stability of the competition matrix. Following the derivation from the main text, and assuming no resources go extinct, we obtain:

$$\alpha_{ij} = \sum_\beta \frac{K_\beta}{r_\beta} c_{i\beta} v_{j\beta}. \tag{154}$$

For two species and two resources, we can explicitly compute the eigenvalues, finding:

$$\lambda = \frac{1}{2}\Bigg(\langle c_1|v_1\rangle + \langle c_2|v_2\rangle \pm \sqrt{(\langle c_1|v_1\rangle + \langle c_2|v_2\rangle)^2 - 4\frac{K_1K_2}{r_1r_2}|c||v|}\Bigg) \tag{155}$$

where we have defined the inner product

$$\langle x|y\rangle \equiv \sum_\beta \frac{K_\beta}{r_\beta} x_\beta y_\beta, \tag{156}$$

and $|c|$ is the determinant of the matrix $c_{i\alpha}$. We see that all the eigenvalues are positive, implying stability, only when $|c|$ and $|v|$ have the same sign:

$$|c||v| > 0. \tag{157}$$

Writing out the determinants explicitly, we have:

$$|c| = c_{11}c_{21}\left(\frac{c_{22}}{c_{21}} - \frac{c_{12}}{c_{11}}\right) = c_{11}c_{21}(\tan\theta_1^c - \tan\theta_2^c) \tag{158}$$

$$|v| = v_{11}v_{21}\left(\frac{v_{22}}{v_{21}} - \frac{v_{12}}{v_{11}}\right) = v_{11}v_{21}(\tan\theta_1^v - \tan\theta_2^v) \tag{159}$$

where the θ's denote the angle that the consumption vector and depletion vector of each species make with the resource 1 axis. The stability criterion thus implies that the species

whose depletion vector $\nu_{i\alpha}$ makes a bigger angle with the resource 1 axis must also have the biggest angle for its consumption vector. If we say that the species with a larger θ_i^c is limited by resource 2, while the one with a smaller θ_i^c is limited by resource 1, then the equilibrium is stable whenever the depletion from each species is more biased towards its own limiting resource than the depletion from the other species is. But both depletion vectors could very well be extremely biased towards the same resource – it's just the order that matters. And the magnitude of the impact doesn't matter at all, just the relative angle.

References

[1] R. Levins, *Theory of fitness in a heterogeneous environment. i. the fitness set and adaptive function*, The American Naturalist **96**(891), 361 (1962).

[2] R. Levins, *Theory of fitness in a heterogeneous environment. ii. developmental flexibility and niche selection*, The American Naturalist **97**(893), 75 (1963).

[3] R. MacArthur and R. Levins, *The Limiting Similarity, Convergence, and Divergence of Coexisting Species*, The American Naturalist **101**, 377 (1967).

[4] R. MacArthur, *Species Packing and Competitive Equilibrium for Many Species*, Theoretical Population Biology **1**, 1 (1970).

[5] R. Levins, *Evolution in changing environments: some theoretical explorations*, Princeton University Press (1968).

[6] D. Tilman, *Resource Competition and Community Structure.*, Princeton University Press, Princeton, NJ (1982).

[7] J. M. Chase and M. A. Leibold, *Ecological niches: linking classical and contemporary approaches*, University of Chicago Press, Chicago, IL (2003).

[8] R. Levins, *The strategy of model building in population biology*, American scientist **54**(4), 421 (1966).

[9] Z. Frentz, S. Kuehn and S. Leibler, *Strongly deterministic population dynamics in closed microbial communities*, Physical Review X **5**(4), 041014 (2015).

[10] B. Momeni, L. Xie and W. Shou, *Lotka-volterra pairwise modeling fails to capture diverse pairwise microbial interactions*, Elife **6**, e25051 (2017).

[11] C. Ratzke and J. Gore, *Modifying and reacting to the environmental ph can drive bacterial interactions*, PLoS biology **16**(3), e2004248 (2018).

[12] J. E. Goldford, N. Lu, D. Bajić, S. Estrela, M. Tikhonov, A. Sanchez-Gorostiaga, D. Segrè, P. Mehta and A. Sanchez, *Emergent simplicity in microbial community assembly*, Science **361**, 469 (2018).

[13] H. Mickalide and S. Kuehn, *Higher-order interaction between species inhibits bacterial invasion of a phototroph-predator microbial community*, Cell systems **9**(6), 521 (2019).

[14] S. Estrela, J. C. Vila, N. Lu, D. Bajić, M. Rebolleda-Gómez, C.-Y. Chang, J. E. Goldford, A. Sanchez-Gorostiaga and Á. Sánchez, *Functional attractors in microbial community assembly*, Cell Systems **13**(1), 29 (2022).

[15] M. Dal Bello, H. Lee, A. Goyal and J. Gore, *Resource–diversity relationships in bacterial communities reflect the network structure of microbial metabolism*, Nature Ecology & Evolution **5**(10), 1424 (2021).

[16] J. Diaz-Colunga, N. Lu, A. Sanchez-Gorostiaga, C.-Y. Chang, H. S. Cai, J. E. Goldford, M. Tikhonov and Á. Sánchez, *Top-down and bottom-up cohesiveness in microbial community coalescence*, Proceedings of the National Academy of Sciences **119**(6), e2111261119 (2022).

[17] J. Hu, D. R. Amor, M. Barbier, G. Bunin and J. Gore, *Emergent phases of ecological diversity and dynamics mapped in microcosms*, Science **378**(6615), 85 (2022).

[18] K. Gowda, D. Ping, M. Mani and S. Kuehn, *Genomic structure predicts metabolite dynamics in microbial communities*, Cell **185**(3), 530 (2022).

[19] R. Levins and R. C. Lewontin, *The dialectical biologist*, Harvard University Press (1985).

[20] R. Lewontin and R. Levins, *Organism and environment* (1997).

[21] J. Rosindell, S. P. Hubbell and R. S. Etienne, *The unified neutral theory of biodiversity and biogeography at age ten*, Trends in ecology & evolution **26**(7), 340 (2011).

[22] S. Azaele, S. Suweis, J. Grilli, I. Volkov, J. R. Banavar and A. Maritan, *Statistical mechanics of ecological systems: Neutral theory and beyond*, Reviews of Modern Physics **88**(3), 035003 (2016).

[23] Z. Wang, A. Goyal, V. Dubinkina, A. B. George, T. Wang, Y. Fridman and S. Maslov, *Complementary resource preferences spontaneously emerge in diauxic microbial communities*, Nature communications **12**(1), 6661 (2021).

[24] E. Blumenthal and P. Mehta, *Geometry of ecological coexistence and niche differentiation*, Phys. Rev. E **108**, 044409 (2023), doi:10.1103/PhysRevE.108.044409.

[25] 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 **117**(25), 14572 (2020), doi:10.1073/pnas.1915313117, https://www.pnas.org/doi/pdf/10.1073/pnas.1915313117.

[26] A. Altieri, F. Roy, C. Cammarota and G. Biroli, *Properties of equilibria and glassy phases of the random lotka-volterra model with demographic noise*, Phys. Rev. Lett. **126**, 258301 (2021), doi:10.1103/PhysRevLett.126.258301.

[27] F. Roy, M. Barbier, G. Biroli and G. Bunin, *Complex interactions can create persistent fluctuations in high-diversity ecosystems*, PLOS Computational Biology **16**(5), 1 (2020), doi:10.1371/journal.pcbi.1007827.

[28] V. Ros, F. Roy, G. Biroli, G. Bunin and A. M. Turner, *Generalized lotka-volterra equations with random, nonreciprocal interactions: The typical number of equilibria*, Phys. Rev. Lett. **130**, 257401 (2023), doi:10.1103/PhysRevLett.130.257401.

[29] V. Ros, F. Roy, G. Biroli and G. Bunin, *Quenched complexity of equilibria for asymmetric generalized lotka–volterra equations*, Journal of Physics A: Mathematical and Theoretical **56**(30), 305003 (2023), doi:10.1088/1751-8121/ace00f.

[30] T. Arnoulx de Pirey and G. Bunin, *Aging by near-extinctions in many-variable interacting populations*, Phys. Rev. Lett. **130**, 098401 (2023), doi:10.1103/PhysRevLett.130.098401.

[31] A. Mahadevan, M. T. Pearce and D. S. Fisher, *Spatiotemporal ecological chaos enables gradual evolutionary diversification without niches or tradeoffs*, Elife **12**, e82734 (2023).

[32] R. Marsland III, W. Cui, J. Goldford, A. Sanchez, K. Korolev and P. Mehta, *Available energy fluxes drive a transition in the diversity, stability, and functional structure of microbial communities*, PLOS Computational Biology **15**(2), e1006793 (2019).

[33] R. Marsland III, W. Cui and P. Mehta, *A minimal model for microbial biodiversity can reproduce experimentally observed ecological patterns*, Scientific reports **10**(1), 3308 (2020).

[34] P. Mehta and R. Marsland III, *Cross-feeding shapes both competition and cooperation in microbial ecosystems*, arXiv preprint arXiv:2110.04965 (2021).

[35] P.-Y. Ho, B. H. Good and K. C. Huang, *Competition for fluctuating resources reproduces statistics of species abundance over time across wide-ranging microbiotas*, Elife **11**, e75168 (2022).

[36] J. Grilli, *Macroecological laws describe variation and diversity in microbial communities*, Nature communications **11**(1), 4743 (2020).

[37] P. Chesson, *Mechanisms of maintenance of species diversity*, Annual review of Ecology and Systematics **31**, 343 (2000).

[38] R. E. Ricklefs, *A comprehensive framework for global patterns in biodiversity*, Ecology letters **7**(1), 1 (2004).

[39] M. Vellend, *Conceptual synthesis in community ecology*, The Quarterly review of biology **85**(2), 183 (2010).

[40] J. HilleRisLambers, P. Adler, W. Harpole, J. Levine and M. Mayfield, *Rethinking community assembly through the lens of coexistence theory*, Annual Review of Ecology, Evolution, and Systematics **43** (2012).

[41] D. Tilman, *The ecological consequences of changes in biodiversity: a search for general principles*, Ecology **80**(5), 1455 (1999).

[42] A. L. Downing and M. A. Leibold, *Ecosystem consequences of species richness and composition in pond food webs*, Nature **416**(6883), 837 (2002).

[43] S. I. Dodson, S. E. Arnott and K. L. Cottingham, *The relationship in lake communities between primary productivity and species richness*, Ecology **81**(10), 2662 (2000).

[44] D. U. Hooper, F. Chapin Iii, J. Ewel, A. Hector, P. Inchausti, S. Lavorel, J. H. Lawton, D. Lodge, M. Loreau, S. Naeem *et al.*, *Effects of biodiversity on ecosystem functioning: a consensus of current knowledge*, Ecological monographs **75**(1), 3 (2005).

[45] D. S. Srivastava and M. Vellend, *Biodiversity-ecosystem function research: is it relevant to conservation?*, Annu. Rev. Ecol. Evol. Syst. **36**, 267 (2005).

[46] A. R. Ives and S. R. Carpenter, *Stability and diversity of ecosystems*, science **317**(5834), 58 (2007).

[47] K. S. McCann, *The diversity–stability debate*, Nature **405**(6783), 228 (2000).

[48] M. Loreau, *From populations to ecosystems: Theoretical foundations for a new ecological synthesis (MPB-46)*, vol. 46, Princeton University Press (2010).

[49] M. Vellend, *The biodiversity conservation paradox*, American Scientist **105**(2), 94 (2017).

[50] N. Rooney and K. S. McCann, *Integrating food web diversity, structure and stability*, Trends in ecology & evolution **27**(1), 40 (2012).

[51] A. Shade, H. Peter, S. D. Allison, D. Baho, M. Berga, H. Bürgmann, D. H. Huber, S. Langenheder, J. T. Lennon, J. B. Martiny *et al.*, *Fundamentals of microbial community resistance and resilience*, Frontiers in microbiology **3**, 417 (2012).

[52] F. D. Hulot, G. Lacroix, F. Lescher-Moutoué and M. Loreau, *Functional diversity governs ecosystem response to nutrient enrichment*, Nature **405**(6784), 340 (2000).

[53] P. B. Adler, J. HilleRisLambers and J. M. Levine, *A niche for neutrality*, Ecology letters **10**, 95 (2007).

[54] J. M. Chase, B. J. McGill, P. L. Thompson, L. H. Antão, A. E. Bates, S. A. Blowes, M. Dornelas, A. Gonzalez, A. E. Magurran, S. R. Supp *et al.*, *Species richness change across spatial scales*, Oikos **128**(8), 1079 (2019).

[55] C. K. Fisher and P. Mehta, *The transition between the niche and neutral regimes in ecology*, Proceedings of the National Academy of Sciences **111**(36), 13111 (2014).

[56] D. A. Kessler and N. M. Shnerb, *Generalized model of island biodiversity*, Physical Review E **91**, 042705 (2015).

[57] D. Tilman, *Resources: a graphical-mechanistic approach to competition and predation*, The American Naturalist **116**(3), 362 (1980).

[58] P. Chesson, *MacArthur's consumer-resource model*, Theoretical Population Biology **37**, 26 (1990).

[59] W. Cui, R. Marsland III and P. Mehta, *Effect of resource dynamics on species packing in diverse ecosystems*, Physical Review Letters **125**(4), 048101 (2020).

[60] D. Tilman, *Resource competition and community structure*, vol. 17, Princeton University Press (1982).

[61] J. Huisman and F. J. Weissing, *Biodiversity of plankton by species oscillations and chaos*, Nature **402**, 407 (1999).

[62] J. Huisman and F. J. Weissing, *Biological conditions for oscillations and chaos generated by multispecies competition*, Ecology **82**, 2682 (2001).

[63] J. Huisman and F. J. Weissing, *Fundamental unpredictability in multispecies competition*, The American Naturalist **157**, 488 (2001).

[64] V. Volterra, *Variation and fluctuations of the number of individuals of animal species living together*, In *Animal Ecology*. McGraw-Hill (1926).

[65] R. MacArthur and R. Levins, *The limiting similarity, convergence, and divergence of coexisting species*, The american naturalist **101**(921), 377 (1967).

[66] R. MacArthur, *Species packing and competitive equilibrium for many species*, Theoretical population biology **1**(1), 1 (1970).

[67] A. J. Lotka, *Contribution to the theory of periodic reactions*, The Journal of Physical Chemistry **14**(3), 271 (1909).

[68] J. P. O'Dwyer, *Whence lotka-volterra?*, Theoretical Ecology **11**(4), 441 (2018).

[69] A. D. Letten, P.-J. Ke and T. Fukami, *Linking modern coexistence theory and contemporary niche theory*, Ecological Monographs (2017).

[70] G. Barabás, R. D'Andrea and S. M. Stump, *Chesson's coexistence theory*, Ecological Monographs **88**(3), 277 (2018).

[71] R. MacArthur, *Species packing, and what competition minimizes*, Proceedings of the National Academy of Sciences **64**, 1369 (1969).

[72] R. Marsland III, W. Cui and P. Mehta, *The minimum environmental perturbation principle: A new perspective on niche theory*, The American Naturalist **196**(3), 291 (2020).

[73] M. Gatto, *Comments on" macarthur's minimization principle: A footnote"*, The American Naturalist **119**, 140 (1982).

[74] M. Gatto, *A general minimum principle for competing populations: some ecological and evolutionary consequences*, Theoretical Population Biology **37**, 369 (1990).

[75] S. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge University Press, Cambridge, UK (2004).

[76] D. P. Bertsekas, *Nonlinear programming*, Athena Scientific, Belmont, MA (1999).

[77] R. Rao and M. Esposito, *Nonequilibrium thermodynamics of chemical reaction networks: Wisdom from stochastic thermodynamics*, Physical Review X **6**, 041064 (2016).

[78] P. Mehta, M. Bukov, C.-H. Wang, A. G. Day, C. Richardson, C. K. Fisher and D. J. Schwab, *A high-bias, low-variance introduction to machine learning for physicists*, Physics Reports **(in press)** (2019).

[79] C. Huttenhower, D. Gevers, R. Knight, S. Abubucker, J. H. Badger, A. T. Chinwalla, H. H. Creasy, A. M. Earl, M. G. FitzGerald, R. S. Fulton *et al.*, *Structure, function and diversity of the healthy human microbiome*, Nature **486**(7402), 207 (2012).

[80] S. Sunagawa, L. P. Coelho, S. Chaffron, J. R. Kultima, K. Labadie, G. Salazar, B. Djahanschiri, G. Zeller, D. R. Mende, A. Alberti *et al.*, *Structure and function of the global ocean microbiome*, Science **348**(6237), 1261359 (2015).

[81] J. G. Caporaso, C. L. Lauber, W. A. Walters, D. Berg-Lyons, C. A. Lozupone, P. J. Turnbaugh, N. Fierer and R. Knight, *Global patterns of 16s rrna diversity at a depth of millions of sequences per sample*, Proceedings of the national academy of sciences **108**(Supplement 1), 4516 (2011).

[82] A.-E. Saliba, A. J. Westermann, S. A. Gorski and J. Vogel, *Single-cell rna-seq: advances and future challenges*, Nucleic acids research **42**(14), 8845 (2014).

[83] S.-K. Ma, *Modern theory of critical phenomena*, Routledge (2018).

[84] R. M. May, *Will a large complex system be stable?*, Nature **238**(5364), 413 (1972).

[85] J. Ginibre, *Statistical ensembles of complex, quaternion, and real matrices*, Journal of Mathematical Physics **6**(3), 440 (1965).

[86] S. Allesina and S. Tang, *Stability criteria for complex ecosystems*, Nature **483**(7388), 205 (2012).

[87] J. Grilli, T. Rogers and S. Allesina, *Modularity and stability in ecological communities*, Nature communications **7**(1), 1 (2016).

[88] T. Gibbs, J. Grilli, T. Rogers and S. Allesina, *Effect of population abundances on the stability of large random ecosystems*, Physical Review E **98**(2), 022410 (2018).

[89] J. Grilli, G. Barabás, M. J. Michalska-Smith and S. Allesina, *Higher-order interactions stabilize dynamics in competitive network models*, Nature **548**(7666), 210 (2017).

[90] G. Bunin, *Ecological communities with lotka-volterra dynamics*, Physical Review E **95**(4), 042414 (2017).

[91] M. Mézard and G. Parisi, *The cavity method at zero temperature*, Journal of Statistical Physics **111**, 1 (2003).

[92] M. Advani, G. Bunin and P. Mehta, *Statistical physics of community ecology: a cavity solution to MacArthur's consumer resource model*, Journal of Statistical Mechanics **2018**, 033406 (2018).

[93] M. Barbier and J.-F. Arnoldi, *The cavity method for community ecology*, bioRxiv p. 147728 (2017).

[94] G. Biroli, G. Bunin and C. Cammarota, *Marginally stable equilibria in critical ecosystems*, New Journal of Physics **20**, 083051 (2018).

[95] A. Agrawal, R. Verschueren, S. Diamond and S. Boyd, *A rewriting system for convex optimization problems*, Journal of Control and Decision **5**(1), 42 (2018).

[96] W. Cui, R. Marsland III and P. Mehta, *Diverse communities behave like typical random ecosystems*, Physical Review E **104**(3), 034416 (2021).