### Les Houches lectures on Theoretical Ecology: high-dimensional models and extreme events

Ada Altieri<sup>1\*</sup>

1 Laboratoire Matière et Systèmes Complexes (MSC), Université Paris Cité, CNRS, 75013 Paris, France \* ada.altieri@u-paris.fr

#### Abstract

The study of ecological systems is gaining momentum in modern scientific research, driven by an abundance of empirical data and advancements in bioengineering techniques. However, a full understanding of their dynamical and thermodynamical properties, also in light of the ongoing biodiversity crisis, remains a formidable endeavor. From a theoretical standpoint, modeling the interactions within these complex systems – such as bacteria in microbial communities, plant-pollinator networks in forests, or starling murmurations – presents a significant challenge. Given the characteristic high dimensionality of the datasets, alternative elegant approaches employ random matrix formalism and techniques from disordered systems. In these lectures, we will explore two cornerstone models in theoretical ecology: the MacArthur/Resource-Consumer model, and the Generalized Lotka-Volterra model, with a special focus on systems composed of a large number of interacting species. In the second part, we will highlight timely directions, particularly to bridge the gap with empirical observations and detect macroecological patterns.

Copyright attribution to authors. This work is a submission to SciPost Physics Lecture Notes. License information to appear upon publication. Publication information to appear upon publication.	Received Date Accepted Date Published Date
--	--

#### <sup>2</sup> Contents

1

3	1	Intro	oduction	2			
4	2	Mac	Arthur's consumer-resource (CR) model:				
5	a simple instance for niche theory						
6		2.1	Low-dimensional criterion for feasibility and stability vs ZNGIs	4			
7		2.2	Rephrasing the CR model in high dimensions: statistical physics lens	5			
8		2.3	Steady-state solutions	5			
9			2.3.1 Interpretation of the CR model as a Constraint Satisfaction Problem	7			
10		2.4	How to recover the Generalized Lotka-Volterra (GLV) model from a CR model?	8			
11		2.5	Cavity method for the GLV model: zero-noise picture	9			
12	3	3 High-dimensional random GLV equations with stochasticity					
13		3.1	Itō's versus Stratonovich prescription	12			
14		3.2	Disordered free energy: replica method	13			
15		3.3	Saddle-point evaluation of the free energy	14			

16		3.4	The re	plica symmetric solution	14		
17		3.5	Nume	rical interpretation of the order parameters	16		
18		3.6	Stabili	ty analysis: Single equilibrium versus multiple equilibria	17		
19	4	Deri	rivation of the phase diagram at finite noise				
20		4.1	Zero-t	emperature limit of the RS solution	19		
21		4.2	Conclusions and Perspectives				
22			4.2.1	Dynamical Mean-Field Theory approach: asymmetric interactions a	and		
23				spatial dependence	21		
24			4.2.2	Heavy-tailed distributions	21		
25			4.2.3	Data-driven approaches	22		
26	26 <b>References</b>						
27							

#### Introduction 1 20

28

In recent years, the study of ecological communities has attracted growing interest from both 30 theoretical physics and quantitative biology. Despite significant advances, several foundational 31 questions remain unresolved. Among these is the long-standing *diversity-stability debate* [1,2], 32 which seeks to clarify the role of ecological diversity in maintaining ecosystem resilience, once 33 appropriate diversity metrics are defined. Another critical challenge lies in understanding how 34 ecosystems respond to external perturbations, whether from sudden shocks or gradual envi-35 ronmental changes. Ecosystems exhibiting abrupt responses to gradual environmental changes 36 are particularly concerning. Prominent examples include desertification in arid regions, coral 37 reef bleaching, and the collapse of tropical forests into treeless landscapes. These shifts are 38 frequently linked to the emergence of multiple stable states under specific conditions [3–5]. 39 Minor perturbations can induce transitions between these states, triggering severe declines 40 in biodiversity and ecosystem functionality. This raises a central theoretical question: Is the 41 number of feasible equilibria in such ecosystems exponential or sub-exponential relative to 42 system size?<sup>1</sup> Local multistability refers to the coexistence of numerous stable states, some of 43 which may require an extended time to equilibrate. Whether the number of such stable and 44 uninvadable states grows exponentially with the number of species – a behavior that could 45 be observed in simulations or tailored laboratory experiments – or on a much smaller scale 46 remains an open question. 47

Moreover, identifying key drivers of ecosystem fragmentation and segregation, as well 48 as developing quantitative precursor indicators, could be essential for predicting ecological 49 collapses. For instance, dispersal in spatially structured communities can rescue fragmented 50 ecosystems from extinction. Depending on demographic fluctuations, the transition from a 51 self-sustained state to ecosystem collapse may be continuous or abrupt, with the latter often 52 associated with tipping points, metastability, and hysteresis – concepts widely recognized not 53 only in ecology but also in environmental science, and condensed matter physics. 54 In these lecture notes, we aim at bridging the gap between statistical physics approaches

55

<sup>&</sup>lt;sup>1</sup>In disordered systems literature, two main reference classes are commonly distinguished: i) spin-glass landscapes, characterized by a sub-exponential number of free-energy minima in the system size, with free-energy barriers expected to be sub-extensive; ii) simple structural-glass landscapes, where the number of free-energy minima grows exponentially with the system size, and the barriers separating these minima are extensive. This kind of question might also be rephrased and adapted to ecological systems.

and the empirical analysis of large, heterogeneous ecosystems. Statistical physics offers effi cient tools for investigating complex dynamics of species-rich ecosystems and eventually ad dressing still open questions in this booming field. Notably, we shall discuss three reference
 models: neutral models, the Generalized Lokta-Volterra model, and the Consumer-Resource
 (or MacArthur) model.

• Neutral Models. Neutral models, exemplified by Hubbell's unified neutral theory of bio-61 diversity and biogeography [6], assume that all individuals are ecologically equivalent, 62 regardless of species identity. They share identical per-capita birth and death rates, as 63 well as equal migration and speciation probabilities. While this assumption implies that 64 species are statistically indistinguishable, demographic stochasticity creates variability: 65 species with larger populations are more likely to persist, while smaller populations 66 face a higher risk of extinction due to random fluctuations. These dynamics are often 67 modeled using stochastic birth-death processes, and in some cases, their stationary dis-68 tributions can be exactly obtained. As such, neutral models capture fluctuations that are 69 driven purely by demographic stochasticity. Neutral theories have since been extended 70 to incorporate additional factors and have successfully reproduced several macroecologi-71 cal patterns [7–9]. However, the assumption of *functional equivalence* among species has 72 been widely criticized for oversimplifying biological systems. In fact, species often ex-73 hibit significant differences in their environmental adaptations, and their interactions 74 actively shape the environment they inhabit. 75

Niche models. Conversely, models that explicitly account for species differences are 76 collectively referred to as niche models. Each niche is occupied by a single species ac-77 cording to the competitive exclusion principle. In other words, two species competing 78 for the same limited resource cannot coexist: the fittest one will dominate in the long 79 term, thus leading either to extinction or an evolutionary shift towards another niche. 80 Among this class of models, Lotka-Volterra equations deserve a special place [10-12]. 81 They were introduced in the early 20th century – in terms of a pair of first-order nonlin-82 ear differential equations – postulating that the demographic rates of a species depend 83 not only on its abundance but also on the abundance of other species in the commu-84 nity. This departure from neutrality enables niche models to explore how inter-specific 85 interactions drive coexistence, competition, and overall stability of a given ecological 86 community. 87

However, one of the main shortcomings of the niche framework is its limited predictive 88 ability. While it provides a way to explain the coexistence of two species by compar-89 ing their ecological requirements and preferences, it offers little insight into determin-90 ing the number of niches in a given environment without directly counting the species 91 that occupy them. MacArthur attempted to address this issue by proposing a famous 92 consumer-resource model. Accordingly, a stable equilibrium can only be achieved if the 93 number of competing species does not exceed the total available resources. However, 94 this upper bound is frequently violated in natural ecosystems, particularly in planktonic 95 communities, giving rise to the so-called *paradox of the plankton* [13]. 96

By focusing on two of these benchmark models, we aim at achieving a better understanding of the intricate dynamics of large ecosystems, thus offering analytical insights into their resilience to external perturbations, species coexistence, and biodiversity maintenance.

#### SciPost Physics Lecture Notes

# MacArthur's consumer-resource (CR) model: a simple instance for niche theory

The foundational concepts of competition and selection shaping communities can be traced
 back to the pioneering works of R. Levine and R. MacArthur in the sixties [14, 15].

Let us consider *S* species, characterized by their abundances  $N_i$  (i = 1, 2, ...S) and competing for *M* resources with abundances  $R_{\alpha}$  ( $\alpha = 1, 2..., M$ ). Their dynamics are defined by:

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

$$\frac{dR_\alpha}{dt} = F_\alpha(\mathbf{R}) - \sum_i N_i c_{i\alpha} R_\alpha = F_\alpha(\mathbf{R}) + \sum_i N_i q_{i\alpha}(\mathbf{R}) ,$$
(1)

where  $c_{i\alpha}$  is a  $S \times M$  matrix accounting for the metabolic strategies or consumer preferences 106 in terms of the  $\alpha$  resource, whereas  $g_i(\mathbf{R}) = \sum_{\alpha} c_{i\alpha} w_{\alpha} R_{\alpha} - \chi_i$  is the growth rate with qual-107 ity factor  $w_{\alpha}$ . In the spirit of MacArthur's seminal work, each species has its own minimum 108 requirement  $\chi_i$ . The function  $F_{\alpha}(\mathbf{R})$  is also called resource supply, which typically satisfies a lo-109 gistic form with  $F_{\alpha}(R_{\alpha}) = R_{\alpha}(K_{\alpha} - R_{\alpha})$ ,  $K_{\alpha}$  being the carrying capacity associated with resource 110  $\alpha$  . The formulation of such Consumer-Resource (CR) equations relies on the assumption that 111 interactions between species are mediated by the consumption of external resources. In other 112 words, the growth function and the impact vector for each species  $q_{ia}(\mathbf{R})$  do not depend on 113 the population size. 114

Given this system of  $S \times M$  coupled equations, one can define three different conditions:

- steady state population, by imposing  $\frac{dN_i}{dt} = 0$ ;
- non-invasibility if  $g_i(\mathbf{R}) \leq 0$ ;
- feasible equilibria provided  $N_i \ge 0$ .

#### 119 2.1 Low-dimensional criterion for feasibility and stability vs ZNGIs

Pioneering works in theoretical ecology have relied on models composed of a few species and few resources. In these cases, stability properties could be deduced by simple geometrical arguments [16, 17]. By denoting with  $c_{i\alpha}R_{\alpha} = -q_{i\alpha}(\mathbf{R})$  the *impact vector* of species *i*, one can deduce fixed-point properties. For the sake of simplicity, let us consider a 2-dimensional model. There exist three geometric conditions that allow us to determine stable coexistence of two competing species:

• The Zero Net Growth Isoclines (ZNGI), defined by the conditions  $g_i(\mathbf{R}) = 0$ , must intersect. This criterion is valid in any dimension and leverages the *Competitive Exclusion Principle* according to which at most *M* consumer species can safely coexist.

• The supply vector  $F_{\alpha}(\mathbf{R})$  evaluated at the fixed point must lie within the cone spanned by the negative impact vector. This second condition guarantees the existence of a set of positive population sizes, with  $N_i > 0$ . It corresponds to imposing  $\frac{dR_{\alpha}}{dt} = 0$ , which leads to  $F_{\alpha}(R_{\alpha}) = \sum_i N_i (-q_{i\alpha}(\mathbf{R}))$ , still holding regardless of the dimensionality of the system.

The impact of each species is biased toward the resource that significantly influences its growth rate, which is particularly challenging to identify in higher dimensions. Based on Tilman's geometrical framework, the orientation of the impact vector should be compared to that of the ZNGIS.

#### 137 2.2 Rephrasing the CR model in high dimensions: statistical physics lens

The aforementioned geometrical criteria offer an intuitive starting point for ecological inter-138 pretation, especially in the context of low-dimensional systems. However, in recent years, 139 there has been a growing focus on systems composed of an enormous number of interacting 140 components, i.e. by considering  $S, M \to \infty$  at fixed ratio  $\gamma = M/S$  [18–20]. Such a pic-141 ture leads to the neglect of individual details and zooms out on the typical behaviors of large 142 communities. In other words, since measuring how single-species parameters scale with the 143 system size is particularly hard, one can therefore resort to a coarse-grained description by 144 averaging over a large number of (functionally equivalent) microscopic degrees of freedom. 145 *Effective models* thus retain only a few fundamental features of the real system, using them to 146 capture general properties, such as macroscopic observables or collective dynamics. 147

The difficulty of inferring the interactions in diverse ecosystems, such as for microbial communities, stems from several factors: the high dimensionality of microbiome datasets, the shortage of long time-series experiments, the time-dependent nature of microbial interactions, and environmental filtering effects due to abiotic factors such as resources, temperature, and pH. Another challenge involves disentangling noisy fluctuations from true ecological signals in empirical data.

An alternative and elegant way out was originally pioneered by Robert May leveraging Random Matrix Theory to model species interaction networks: each entry of the adjacency matrix is drawn from a given probability distribution. This approach drastically reduces the number of model parameters to be estimated, from  $S^2$  to just a few [21]. Within a statistical physics scenario, one typically splits the consumer preference matrix into a deterministic part and a fluctuating contribution, that is:

$$\langle c_{i\alpha} \rangle = \frac{\mu_c}{M} , \quad \langle (c_{i\alpha} - \langle c_{i\alpha} \rangle)^2 \rangle = \frac{\sigma_c^2}{M} .$$
 (2)

160

$$\langle \chi_i \rangle = \chi , \quad \langle (\chi_i - \langle \chi_i \rangle)^2 \rangle = \sigma_{\chi}^2$$
(3)

$$\langle K_{\alpha} \rangle = K$$
,  $\langle (K_{\alpha} - \langle K_{\alpha} \rangle)^2 \rangle = \sigma_K^2$ . (4)

This description is consistent with a *generalist* picture, meaning that all species tend to grab a large number of resources. This is also consistent with the scaling of the metabolic strategy coefficients  $c_{i\alpha}$ : the outflow associated with each resource decreases with the amount of resources itself. This precise rescaling of the random coefficients will appear much clearer from a formal point of view in the following derivations. In other words, the scaling with *M* of the first two moments guarantees a proper thermodynamic limit.

#### 168 2.3 Steady-state solutions

To derive the mean-field self-consistent equations, we ask what happens if we add an additional species and resource to the pre-existing pool of S species and M resources. Intuitively, this is equivalent to exploring whether a new species can invade the ecosystem.

Formally, by denoting with  $N_0$  the new species and  $R_0$  the new resource, we aim to study correlations between resource and species entangled dynamics. However, because we always leverage a high-dimensional formalism in which  $S, M \gg 1$ , considering a system with S + 1species and M + 1 resources does not sensitively affect the resulting thermodynamic picture. This addition will result in a slight perturbation of the original system, to be studied in terms of two local susceptibilities:

$$\chi = \frac{1}{M} \sum_{\alpha} \frac{\partial R_{\alpha}^{*}}{\partial K_{\alpha}}, \qquad \nu = \frac{1}{S} \sum_{i} \frac{\partial N_{i}^{*}}{\partial g_{i}} = -\frac{1}{S} \sum_{i} \frac{\partial N_{i}^{*}}{\partial \chi_{i}}.$$
(5)

The first expression (left) encodes the change of the mean resource abundance if we slightly perturb the supply of all external resources, by summing over  $\alpha$ . Conversely, on the right, the other susceptibility quantifies the change of the mean species abundance upon decreasing their fitness cost (or increasing the growth rate). The so-called susceptibilities  $\chi$  and  $\nu$  can be advantageously measured in experiments, for instance investigating growth and community structure by the addition of a given molecule, i.e. glucose.

In the following, in Sec. 2.5, we will see how to explicitly derive the steady-state abundances for a simpler model, by using the so-called *cavity method*. For the sake of compactness, we show here the resulting expressions for  $N_0^*$  and  $R_0^*$ , to be obtained by means of the Gaussian nature of the  $c_{i\alpha}$ , i.e.  $c_{i\alpha} = \frac{\mu}{M} + \sigma_c \eta_{i\alpha}$  where  $\langle \eta_{i\alpha} \rangle = 0$ ,  $\langle \eta_{i\alpha} \eta_{j\beta} \rangle = \frac{\delta_{ij} \delta_{\alpha\beta}}{M}$ :

$$N_{0}^{*} = \max\left[0, \frac{g + \sigma_{g} z_{0}}{\gamma \sigma_{c}^{2} \chi}\right]$$

$$R_{0}^{*} = \max\left[0, \frac{K_{\text{eff}} + \sigma_{K \text{eff}} \tilde{z}_{0}}{1 - \gamma^{-1} \sigma_{c}^{2} \nu}\right]$$
(6)

where we have denoted as  $z_0, \tilde{z}_0$  two Gaussian variables with zero mean and unit variance. Eq. (6) will thus represent two truncated Gaussian distributions for the typical species and resource variables.  $K_{\text{eff}}(\mathbf{N}) = K_{\alpha} - \sum_{i} N_i c_{i\alpha}$  denotes instead an *effective resource capacity*.

The above expressions give rise to a straightforward interpretation: by integrating out the quenched disorder formally encoded in the matrix  $c_{i\alpha}$  and  $K_{\alpha}$ , we obtain an effective description where both the growth rates and the carrying capacities are normally distributed. Depending on the sign of the invasion growth rate,  $g + \sigma_g z_0$ , two options are possible:

If such a quantity is strictly positive, the species will survive with an abundance proportional to the invasion growth rate;

• otherwise, it will not be able to invade, thus going extinct in a finite time.

An analogous reasoning can be applied to  $K_{\text{eff}}$ : if the overall quantity  $K_{\text{eff}} + \sigma_{K_{\text{eff}}} \tilde{z}_0$  turns out to be negative, the new resource will be depleted; otherwise, it will be proportional to the carrying capacity itself.

In the previous expressions, we have assumed that the depletion and growth rates share the same matrix structure. However, we can also consider a slightly different variant:

$$\frac{dN_i}{dt} = N_i \left[ \sum_{\delta} c_{i\delta} w_{\delta} R_{\delta} - \chi_i \right]$$

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

203

Imposing  $\frac{dR_{\alpha}}{dt} = 0$  leads to a stability criterion, i.e.  $R_{\alpha} = K_{\alpha} - \frac{1}{\rho_{\alpha}} \sum_{j} d_{j\alpha} N_{j}$ , where  $\rho_{\alpha}$  denotes the ratio between the growth rate and the carrying capacity. The condition on  $R_{\alpha}$  is 204 205 associated with two distinct scenarios: i) if the term in parenthesis is negative,  $\frac{dR_{\alpha}}{dt} < 0$  for all positive  $R_{\alpha}$ , and therefore  $R_{\alpha} = 0$  will be the stable solution. Conversely, if the term in 206 207 parenthesis is positive,  $R_{\alpha} = 0$  becomes unstable and even a small perturbation will result 208 in a positive growth rate. Despite its simplicity, the consumer-resource model can display a 209 plethora of appealing dynamical behaviors. Provided a feasible fixed point, it turns out to be 210 stable if the two matrices are equal. In other words, by looking at the species abundance as a 211 function of time, it will converge to a unique fixed point. More complex scenarios, like limit 212 cycles and persistent fluctuations, can, however, emerge when the growth and depletion rates 213 are not identical, as qualitatively depicted in Figure 1. 214



Figure 1: Distinct consumption and growth probabilities can lead to instabilities and subsequently richer dynamics. Complex and out-of-equilibrium behavior, including both limit cycles and chaotic dynamics, can occur upon changing the correlation between consumption and growth rates. Figure adapted from [22] (courtesy of Y. Liu).

#### 215 2.3.1 Interpretation of the CR model as a Constraint Satisfaction Problem

In the limit of a large number of species and resources  $S, M \to \infty$ , the MacArthur model 216 has recently been reformulated as a constraint satisfaction problem, displaying the typical fea-217 tures of a perceptron model [18, 20]. Originally introduced in the theory of neural networks 218 and machine learning as a binary classifier, the perceptron makes predictions by applying a 219 weighted sum to the input features and using a threshold to assign a class label. A key quan-220 tity is represented by the fractional volume of the solution space occupied by the interacting 221 patterns, which corresponds to the maximum storage capacity of the network. In its convex 222 regime – meaning that the configurational landscape is characterized by a unique fixed point 223 - the critical transition occurs at zero interaction heterogeneity and fixed ratio between the 224 number of patterns and the total degrees of freedom (i.e., the number of bits), equal to two. 225 This transition separates a SAT region, where all constraints are simultaneously satisfied, from 226 an UNSAT region, where no solution can fulfill all clauses. 227

Rephrased in an ecological context, the former of Eqs. (7) can be recast as a function of a resource surplus,  $\Delta_i$ :

$$\frac{dN_i}{dt} \propto N_i \Delta_i , \qquad (8)$$

where again i = 1, ..., S. This novel framework, introduced by Tikhonov and Monasson [18], focuses exclusively on equilibrium configurations, leading to two possible scenarios: (i)  $N_i > 0$ and  $\Delta_i = 0$ , or (ii)  $N_i = 0$  and  $\Delta_i < 0$  (extinction). Conversely, the opposite case, for  $\Delta_i > 0$ , is prohibited by the model definition. Then, the total demand  $T_{\delta} = \sum_i N_i c_{i\delta}$  – where the index  $\delta = 1, ..., M$  denotes the number of available resources – is a function of the species abundances themselves, hence resulting in a feedback loop mechanism.

This model was initially devised to provide a clear geometric interpretation: metabolic 236 strategies are represented as S vectors in the M-dimensional space of resource availability. 237 The equality  $\sum_{\delta} c_{i\delta} w_{\delta} R_{\delta} = \chi_i$  defines a hyperplane, which divides the configuration space of 238 abundances into two regions: an unsustainable region, where the scalar product is smaller than 239 the effective requirement  $\chi_i$ , and a sustainable region, above this hyperplane, where a positive 240 resource surplus enables species to self-sustain and multiply. The single species requirement 241 can also be recast in terms of a random cost, such that  $\chi_i = \sum_{\delta} c_{i\delta} + \epsilon x_i$ ,  $\epsilon$  denoting an 242 infinitesimally small parameter and  $x_i$  a unit-variance Gaussian variable. Accordingly, one 243 can establish the full phase diagram of the model and investigate its stability as a function 244

of a restricted number of control parameters: the ratio between the species in the pool and the number of available resources  $\alpha = \frac{S}{M}$ ; the width of the cost distribution  $\epsilon$ ; the average and variance of the metabolic strategy distribution; and, if an additional layer of complexity is introduced, the variance of the resource supply. To characterize the system's dynamics, a Lyapunov function can be defined for each trajectory:

$$\mathcal{F}(\{N_i\}) = \sum_{\alpha} R_{\alpha} \log\left(\sum_i N_i c_{i\alpha}\right) - \sum_i N_i \chi_i .$$
(9)

One key advantage of defining a convex and bounded Lyapunov function is that it allows for the analysis of system stability in each phase by considering the harmonic fluctuations of such a function with respect to the species-abundance order parameter. In this system, it has been shown that a globally stable equilibrium state always exists, which, under certain conditions, may approach a marginally stable configuration [20].

The asymptotic behavior of the system can then be evaluated in the large- $\beta$  limit, which 255 corresponds to the zero-temperature regime of a statistical physics problem, where  $\beta = 1/T$ 256 is the inverse equilibrium temperature. As  $\epsilon \to 0$  limit, the model exhibits a phase transition 257 between two distinct regimes [18]: a shielded phase, where a collective and self-sustained 258 behavior arises, basically unaffected by external conditions; and a vulnerable phase, where 259 species cannot sustain themselves and turn out to be highly sensitive to environmental changes 260 and improvements. This kind of transition reminds the aforementioned SAT/UNSAT transition 261 occurring in the perceptron, interpreted as a linear signal classifier. 262

Moreover, in the low-temperature regime, the logarithmic term in the effective free energy 263 becomes the dominant contribution. This behavior is associated with vanishing  $\Delta_i$  for a subset 264 of selected species. Then, by analyzing the harmonic fluctuations of the Lyapunov function 265 around its minimum, one can extract the spectral density in both phases, revealing a critical 266 Marchenko-Pastur distribution [23]. This signals the onset of a marginal stability condition, 267 reminiscent of the glassy phenomenology observed in critical jammed systems, such as infinite-268 dimensional hard spheres and other continuous constraint satisfaction problems belonging to 269 the same universality class [20, 24]. 270

#### 271 2.4 How to recover the Generalized Lotka-Volterra (GLV) model from a CR model?

Different dynamical frameworks can be applied depending on the type of interaction being modeled—whether with the environment or among species. In the CR model, interactions primarily occur with the environment, whereas in the LV model, the environment serves as a mediator for inter-species interactions, akin to a thermal bath in physics. For the sake of simplicity, let us consider the scenario in which both depletion and growth rates are encoded in the same matrix,  $c_{i\alpha}$ .

$$\frac{N_i}{N_i} = \sum_{\delta} c_{i\delta} R_{\delta} - \chi_i$$

$$\frac{\dot{R}_{\alpha}}{R_{\alpha}} = \frac{r_{\alpha}}{K_{\alpha}} (K_{\alpha} - R_{\alpha}) - \sum_j c_{\alpha j}^T N_j$$
(10)

Each row of the matrix is an *M*-dimensional vector, the uptake profile of the species *i*. If we allow resource dynamics to be much faster than those of the competing species, we can apply an adiabatic approximation and set the latter equations to zero:

$$m_{\alpha}(K_{\alpha}-R_{\alpha})-\sum_{j}c_{\alpha j}^{T}N_{j}=0, \qquad (11)$$

where  $m_{\alpha}$  denotes the ratio between the growth rate and the carrying capacity for each  $\alpha$ . Equation (11) implies a closed-form expression for the resources as a function of species' abundances

$$R_{\alpha} = K_{\alpha} - \frac{1}{m_{\alpha}} \sum_{j} c_{j\alpha} N_j .$$
<sup>(12)</sup>

As a consequence, the former dynamic turns out to be:

$$\frac{\dot{N}_i}{N_i} = \sum_{\delta} \left[ c_{i\delta} \left( K_{\delta} - \frac{1}{m_{\delta}} \sum_j c_{j\delta} N_j \right) \right] - \chi_i = \left( \sum_{\delta} c_{i\delta} K_{\delta} - \chi_i \right) - \sum_{\delta} \left( \frac{1}{m_{\delta}} \right) \sum_j c_{i\delta} c_{\delta j}^T N_j .$$
(13)

<sup>285</sup> It can be rephrased in a more compact form:

$$\frac{N_i}{N_i} = \mathcal{U}_i - A_{ii}N_i - \sum_{j \neq i} A_{ij}N_j \tag{14}$$

<sup>286</sup> by slightly shaping the interaction matrix

$$A_{ij} = \sum_{\delta} c_{i,\delta} c_{\delta,j}^T \,. \tag{15}$$

Note that we have made use of two assumptions: i) we have assumed  $m_{\delta} = m = 1$ , namely species-independent parameters and constant over time; ii) we have split the sum over *j* in Eq. (13) into a diagonal and off-diagonal contribution, and eventually swapped the summation over resources and species. In this rephrasing, the interactions are written as a scalar product between uptake profiles, which allows us to recover the Generalized Lotka-Volterra equations with symmetric interactions  $A_{ij} = A_{ji}$ . A detailed derivation can be found in the SI of [25].

#### 293 2.5 Cavity method for the GLV model: zero-noise picture

One possibility for investigating how a system reacts to external perturbations is to apply linear stability analysis, linearizing the dynamical equations around their fixed point. This approach reflects May's original intuition.

However, to account for more complex scenarios, especially those featuring multiple equilibria or persistent fluctuations, one can use the *cavity method*, a powerful technique in the analysis of disordered systems. Therefore, we consider the multi-species dynamics for the species abundances  $N_i$ , without any demographic or environmental noise:

$$\frac{dN_i}{dt} = r_i N_i \left( K_i - N_i - \sum_{j \neq i} A_{ij} N_j \right).$$
(16)

As in previous frameworks, the index i = 1, 2, ..., S runs over the total number of species in the pool. Positive interaction coefficients  $A_{ij} > 0$  denote competition, whereas negative coefficients stand for mutualism. For simplicity, we shall set the growth rates  $r_i = r = 1$ and the carrying capacities  $K_i = K$  (upon rescaling the other parameters). The analysis can nevertheless be extended in the case of random carrying capacities [26], providing only a quantitative change in terms of the resulting phase diagram.

Our goal is to characterize the steady-state solution, defined by  $\frac{dN_i}{dt} = 0$ , where the coefficients  $A_{ij}$  are extracted from a suitably defined random matrix. Let us consider then:

$$N_i \left( K - N_i - \sum_{j \neq i} A_{ij} N_j \right) = 0 \tag{17}$$

where, similarly to the MacArthur model,  $\langle A_{ij} \rangle = \mu/S$  and  $\langle A_{ij}^2 \rangle_c = \sigma^2/S$ . Such scaling is essential to ensure a well-defined thermodynamic limit, that is the quantity  $\sum_j A_{ij} N_j^*$  must be independent of the system size. Therefore, we assume that the elements scale as

$$A_{ij} = \frac{\mu}{S} + \sigma a_{ij} \tag{18}$$

<sup>312</sup> where the first and the second moments of the Gaussian variable  $a_{ij}$  read

$$\langle a_{ij} \rangle = 0$$
,  $\langle a_{ij} a_{kl} \rangle = \frac{1}{S} \delta_{ik} \delta_{jl} + \frac{\rho}{S} \delta_{il} \delta_{jk}$ . (19)

However, writing down a closed-form set of equations does not require the interactions to be Gaussian. The only condition is that the first two moments are finite, and the distribution does not exhibit long tails.

In the cavity formalism, a new species is added to the ecosystem with mutual interactions  $A_{0j}$  and  $A_{j0}$  with the rest of the community. By construction, the different trajectories  $\{N_i(t)\}$  evolving according to the dynamics (21) are independent of such interaction coefficients, which allow us to use central-limit arguments. The dynamics of the former *S* species can be written as

$$N_i \left( K - N_i - \sum_j A_{ij} N_j - A_{i0} N_0 \right) = 0 , \qquad (20)$$

<sup>321</sup> whereas for the new one, we have:

$$N_0 \left( K - N_0 - \sum_j A_{0j} N_j \right) = 0 .$$
 (21)

For compactness, we have dropped the asterisk in the steady-state solution. As we have seen for the consumer-resource model, adding a new species is equivalent to slightly perturbing the carrying capacity by an amount  $\delta K_i$ . Let us consider the local susceptibilities  $\chi_{ij} = \frac{\partial N_i}{\partial K_j}$ , and apply the linear response theory to rewrite the stationary solution in perturbation theory:

$$N_i = N_{i\setminus 0} - \sum_j \chi_{ij} A_{j0} N_0 .$$
<sup>(22)</sup>

The term  $N_{i\setminus 0}$  represents the abundance in the absence of the 0-th species, which gives rise to the *cavity* designation. Since  $A_{0i} \sim O\left(\frac{1}{S}\right)$ , the term  $A_{i0}N_0$  can be treated as a small perturbation. Plugging the expression above into Eq. (21) for the 0-th species, we end up with

$$N_{0}\left[K-N_{0}-\sum_{j}A_{0j}\left(N_{j\setminus 0}-\sum_{i}\chi_{ji}A_{i0}N_{0}\right)\right]=0.$$
 (23)

Solving the equation for  $N_0$ , and excluding the trivial vanishing solution, we obtain to the leading order:

$$N_{0} = \frac{K - \sum_{j} \sigma a_{0j} N_{j \setminus 0} - \mu \langle N \rangle}{1 - \sigma^{2} \sum_{ij} \chi_{ij} a_{0j} a_{i0}} .$$
(24)

According to the central limit theorem, one can claim that the contribution  $\sum_{i} a_{0i} \chi_{ii} a_{i0}$ will converge to its expectation value. We can therefore rewrite the term at the denominator as  $\rho \frac{1}{S} \sum_{i} \chi_{jj} = \rho \chi$  yielding

$$N_0 = \max\left[0, \frac{K - \sum_j \sigma a_{0j} N_{j\setminus 0} - \mu \langle N \rangle}{1 - \sigma^2 \rho \,\chi}\right]$$
(25)

which is required to be non-negative, otherwise the solution does not exist. Then, one need to evaluate the scaling of the typical fluctuations: the non-diagonal term is associated to corrections  $O(S^{-1/2})$ , which become subleading in the thermodynamic limit.

As a next fundamental hypothesis to close the equations on  $N_0$  we need to ask for **selfaveraging** property in the thermodynamic limit, that implies:

$$\left\langle \sum_{j=1}^{S} \sigma a_{0j} N_{j\setminus 0} \right\rangle = S \sigma \langle a_{0j} \rangle \langle N \rangle = 0 ,$$

$$\left\langle \left( \sum_{j=1}^{S} \sigma a_{0j} N_{j\setminus 0} \right)^{2} \right\rangle = \sigma^{2} \sum_{j,k} \langle a_{0j} a_{0k} \rangle N_{j\setminus 0} N_{k\setminus 0} = \sigma^{2} \langle N^{2} \rangle$$
(26)

where we have introduced  $\frac{1}{S}\sum_{j} N_{j\setminus 0} = \langle N_0 \rangle = \langle N \rangle$ ,  $\frac{1}{S}\sum_{j} N_{j\setminus 0}^2 = \langle N_0^2 \rangle = \langle N^2 \rangle$ . By gathering all ingredients together and introducing a standard normal variable *z*, we obtain the distribution of the solutions for  $N_0$ :

$$N_0 = \max\left[0, \frac{K - \mu \langle N \rangle + \sqrt{\sigma^2 \langle N^2 \rangle} z}{1 - \sigma^2 \rho \chi}\right]$$
(27)

which reflects the typical shape of a Gaussian distribution with mean  $(K - \mu \langle N \rangle)/(1 - \sigma^2 \rho \chi)$ and variance  $\sigma^2 \langle N^2 \rangle/(1 - \sigma^2 \rho \chi)^2$ . To the leading order in 1/*S*, the resulting distribution of the typical species abundance is a truncated (positive-definite) Gaussian law. It inherently implies the resolution of the following set of self-consistent equations:

$$\langle N \rangle = \frac{1}{S} \sum_{i} N_{i \setminus 0} , \qquad \langle N^2 \rangle = \frac{1}{S} \sum_{i} N_{i \setminus 0}^2 , \qquad (28)$$

346

$$\chi = \left\langle \frac{\partial N_0}{\partial \xi_0} \right\rangle = \frac{\phi}{1 - \sigma^2 \rho \chi} , \qquad \phi = \frac{1}{S} \sum_i \theta(N_0) . \tag{29}$$

where the susceptibility essentially accounts for the single-species response function to a slight perturbation, for instance in the carrying capacity or the growth rate. It can be recast in terms of the fraction of surviving species  $\phi$  and  $\chi$  itself. The symbol  $\theta(\cdot)$  denotes the Heaviside step function. See also [17] for an alternative derivation in a few benchmark models.

#### <sup>351</sup> 3 High-dimensional random GLV equations with stochasticity

We consider the following dynamical equations for the evolution of the species abundance  $N_i$ at time *t*, where the index *i* runs over the total number of species i = 1, ..., S:

$$\frac{dN_i}{dt} = N_i \left[ (K_i - N_i) - \sum_{j, (j \neq i)} A_{ij} N_j \right] + \eta_i(t) + \lambda_i =$$

$$= N_i \left[ -\nabla_{N_i} V_i(N_i) - \sum_{j, (j \neq i)} A_{ij} N_j \right] + \eta_i(t) + \lambda_i ,$$
(30)

 $N_i(t)$  is interpreted as a relative species abundance of species *i* at time *t*, meaning that the population is actually normalized by the total number of individuals populating the ecosystem in the absence of interactions (deterministic scenario, i.e.  $A_{ii} = 0$ ).

To account for demographic fluctuations, we introduce the variable  $\eta_i(t)$ , a white Gaussian noise with zero mean and variance  $\langle \eta_i(t)\eta_i(t')\rangle = 2TN_i(t)\delta_{ij}\delta(t-t')$ , where *T* represents the

noise amplitude. The amplitude T turns out to be inversely proportional to the total number 359 of individuals in the pool: hence, the larger the system size, the smaller the demographic noise 360 strength. We specifically follow Ito's convention for the stochastic part: the multiplicative na-361 ture of the equations ensures that for zero immigration ( $\lambda_i = 0$ ) – if the absorbing state  $N_i = 0$ 362 is reached – the population goes extinct and remains there at all later times. Alternatively, 363 to properly study the interplay with an intrinsic source of fluctuations, one can introduce a 364 species-independent immigration parameter  $\lambda$ , therefore preventing overall extinctions in a 365 finite time. 366

#### 367 3.1 Itō's versus Stratonovich prescription

Equation (30) is meaningless without selecting a proper discretization for the noise. If one adopts Itō's prescription, the time derivative of a generic observable O can be expressed as

$$\frac{d}{dt}\langle \mathcal{O}(\{N_j\})\rangle = \langle \sum_i \frac{\partial \mathcal{O}}{\partial N_i} \frac{dN_i}{dt} \rangle + T \langle \sum_i \frac{\partial^2 \mathcal{O}}{\partial N_i^2} N_i \rangle , \qquad (31)$$

where  $\langle \cdot \rangle$  stands for the average over the probability distribution  $P(\{N_j\}, t)$ , i.e. over the thermal noise.

Then, in the symmetric interaction case (for  $A_{ij} = A_{ji}$ ), one can show that the multi-species dynamics admit an invariant equilibrium-like probability distribution [26,27]. By plugging the dynamical equation (30) in the above expression, we obtain:

$$\frac{d}{dt}\langle \mathcal{O}(\{N_j\})\rangle = \left\langle \sum_i \frac{\partial \mathcal{O}}{\partial N_i} \left[ -N_i \nabla_{N_i} V_i(N_i) - N_i \sum_j A_{ij} N_j + \lambda \right] \right\rangle + T \left\langle \sum_i \frac{\partial^2 \mathcal{O}}{\partial N_i^2} N_i \right\rangle.$$
(32)

The three terms in parenthesis can be collected together and denoted as  $\tilde{F}(\{N_j\})$  allowing us to rewrite the time evolution of the average operator as

$$\frac{d}{dt}\langle \mathcal{O}(\{N_j\})\rangle = \int \prod_i dN_i P(\{N_j\}, t) \left[\sum_i \frac{\partial \mathcal{O}}{\partial N_i} \tilde{F}(\{N_j\}) + T \sum_i \frac{\partial^2 \mathcal{O}}{\partial N_i^2} N_i\right], \quad (33)$$

from which the resulting equation for  $\langle O(\{N_j\}) \rangle$  can easily be obtained by integration by parts. In the same spirit, the dynamical equation for the probability distribution reads:

$$\frac{\partial P(\{N_j\},t)}{\partial t} = \sum_{i} \left[ -\frac{\partial}{\partial N_i} \left[ \tilde{F}(\{N_j\}) P(\{N_j\},t) \right] + T \frac{\partial^2}{\partial N_i^2} \left[ P(\{N_j\},t) N_i \right] \right]$$
(34)

from which, by imposing the R.H.S. to be zero, we obtain the stationary probability distribution. Therefore, we ask the invariant probability distribution to scale as  $P \propto \exp(-\beta H)$ , with *H* an effective energy function, and inverse temperature  $\beta = 1/T$ . This implies  $\frac{1}{P} \frac{\partial P}{\partial N_i} = -\frac{1}{T} \frac{\partial H}{\partial N_i}$ from which, by integrating over  $N_i$ , we can deduce the associated energy function valid in the symmetric interaction case:

$$H = \sum_{i} V_{i}(N_{i}) + \sum_{i < j} A_{ij} N_{i} N_{j} + (T - \lambda) \sum_{i} \ln N_{i} .$$
(35)

A complete derivation can be found for other instances of conservative dynamics, for instance in the case of a non-logistic self-regulation contribution (see [28]).

In other words, the original dynamical process in Eq. (30) describes the time evolution of an interacting ecosystem whose thermodynamics is defined by Eq. (35). Note that – at variance with the simplest scenario corresponding to the analysis in the limit  $T \rightarrow 0$  and  $\lambda \rightarrow 0^+$  – introducing a small but finite immigration results in a qualitative behavioral change. The immigration parameter ensures that no species will go extinct thus replacing the concept of indefinite extinction with the concept *everything is everywhere* [29].

Conversely, Stratonovich discretization has the advantage of preserving invariance under time reversal. However, due to the multiplicative nature of the noise, the two discretization schemes are not formally equivalent and necessitate the introduction of an additional drift term:

$$\eta\sqrt{N} \to \eta\sqrt{N} - \frac{1}{2}\frac{\sqrt{2T}}{2\sqrt{N}}\sqrt{2TN} = \eta\sqrt{N} - \frac{T}{2} .$$
(36)

#### 396 3.2 Disordered free energy: replica method

The disorder average of the free energy can be computed through the replica method [30], a powerful technique originally introduced for dealing with spin-glass problems. It relies on the identity

$$-\beta F = \overline{\ln Z} = \lim_{n \to 0} \frac{\ln \overline{Z^n}}{n} .$$
(37)

where one first computes the replicated partition function for integer values of n, and only at the final stage considers the analytical continuation  $n \rightarrow 0$ . Despite being thermodynamically independent, replicas turn out to be correlated because they are taken over the same realization of the disorder. Therefore, integrating over the disorder introduces an effective coupling between them.

Let us thus proceed with the computation of  $\overline{Z^n}$ :

$$\overline{Z^n} = \int \prod_{i(38)$$

<sup>406</sup> By performing a Gaussian integration over the random variables  $A_{ij}$  with finite mean and <sup>407</sup> variance, we obtain:

$$\overline{Z^{n}} = \int \prod_{a=1}^{n} \prod_{i} dN_{i}^{a} \exp\left\{\sum_{i < j} \frac{\sigma^{2}}{2S} \left(\beta \sum_{a} N_{i}^{a} N_{j}^{a}\right)^{2} + \beta \sum_{a} \left[\sum_{i} \left[V_{i}(N_{i}) + (T - \lambda) \ln(N_{i})\right] + \frac{\mu}{S} \sum_{i < j} N_{i}^{a} N_{j}^{a}\right]\right\}$$
(39)

To decouple the first term, we thus introduce the following quantities corresponding respectively to the *overlap matrix* between two replicas *a*, *b* of the reference system and the *mean abundance*:

$$Q_{ab} = \frac{1}{S} \sum_{i=1}^{S} N_i^a N_i^b , \quad H_a = \frac{1}{S} \sum_{i=1}^{S} N_i^a$$
(40)

The replicated partition function can eventually be rewritten in terms of the new order parameters:

$$\overline{Z^n} = \int \prod_{a \le b} dQ_{ab} \prod_a dH_a \exp\left\{S\mathcal{A}\left(\{Q_{ab}, H_a\}\right)\right\}$$
(41)

413 where the free-energy action A reads:

$$\mathcal{A}(\{Q_{ab}, Q_{aa}, H_a\}) = -\frac{1}{2}\sigma^2\beta^2 \left(\sum_{a < b} Q_{ab}^2 + \frac{1}{2}\sum_a Q_{aa}^2\right) + \frac{\beta\mu}{2}\sum_a H_a^2 + \frac{1}{S}\sum_i \ln Z_i(\{Q_{ab}, H_a\}).$$
(42)

#### SciPost Physics Lecture Notes

414 Alternatively, one can recognize a Hubbard-Stratonovich transformation structure, which in 415 the case of a multivariate Gaussian integral reads:

$$\sqrt{\frac{(2\pi)^{N}}{\det\hat{A}}}\exp\left(-\frac{\mathbf{b}^{T}\hat{A}\mathbf{b}}{2}\right) = \int_{-\infty}^{\infty}\exp\left[-\frac{\mathbf{y}^{T}\hat{A}^{-1}\mathbf{y}}{2} + \mathbf{b}^{T}\cdot\mathbf{y}\right]d\mathbf{y},$$
(43)

which we apply to  $y = Q_{ab}$  (and  $H_a$ , in turn).

The last piece in Eq. (42), i.e. the effective partition function for the one-species abundances in different replicas, reads:

$$Z_i = \int \prod_a dN_i^a \exp\left[-\beta H_{\text{eff}}(\{N_i^a\}, \{Q_{ab}, H_a\})\right]$$
(44)

419 whose associated effective Hamiltonian is

$$H_{\rm eff} = -\beta \sigma^2 \left( \sum_{a < b} N_i^a N_i^b Q_{ab} + \frac{1}{2} \sum_a (N_i^a)^2 Q_{aa} \right) + \sum_a \left( \mu H_a N_i^a + V(N_i^a) + (T - \lambda) \ln N_i^a \right).$$
(45)

420

#### 421 3.3 Saddle-point evaluation of the free energy

In the  $S \to \infty$  limit only the values of  $Q_{ab}$  and  $H_a$  that extremize the action will contribute<sup>2</sup>:

$$\frac{\ln \overline{Z^n}}{S} = \mathcal{A}(\{Q_{ab}^*, H_a^*\}).$$
(46)

<sup>423</sup> Therefore, based on the stationarity condition, one obtains:

where the brackets indicate the thermal average over the resulting Hamiltonian  $H_{\text{eff}}$ . It depends on  $Q_{ab}$  and  $H_a$  requiring the explicit integration or, in a more straightforward way, the resolution of a set of self-consistency equations. As it is for spin-glass models,  $Q_{ab}$  represents the overlap matrix between two replicated configurations, in the presence of an external field, denoted by  $H_a$ .

#### 429 **3.4** The replica symmetric solution

Solving these high-dimensional equations without guessing a specific form for the overlap matrix  $Q_{ab}$  and the mean abundance  $H_a$  seems impractical. To address this challenging task, one can initiate the analysis by assuming a replica-symmetric (RS) ansatz. Since the action is symmetric under the exchange of replica indices, we can first claim that the solution respects this symmetry. Therefore, all fields must be equal and the overlap can only take two values:  $q_0$  for the overlap between different replicas, and  $q_d > q_0$  for the self-overlap (see Fig. 2).

$$Q_{ab} = q_0 \qquad a \neq b$$

$$Q_{aa} = q_d \qquad a = b \qquad (48)$$

$$H_a = h \qquad \forall a$$



Figure 2: Replica-Symmetric (RS) ansatz for the overlap matrix  $Q_{ab}$ , which is parametrized by two values only, the diagonal term (self-overlap) and the off-diagonal contribution (inter-state overlap).

The stability of the solution will be eventually checked by computing the Hessian of the free 436 energy, and notably its leading eigenvalue, the so-called replicon eigenvalue [30–32]. If there 437 were multiple equilibria, two distinct replicas would have a different overlap according to 438 whether they were in the same state or not. Since here distinct replicas always have overlap 439  $q_0$ , this ansatz corresponds to assuming that there is a unique equilibrium state. The "size" of 440 this state is characterized by the overlap between different replicas  $q_0$ : if it is large ( $q_0 \leq q_d$ ) 441 the configurations in the same state are very similar, so that the state is very localized in phase 442 space, if it is small  $(q_0 \ll q_d)$  the state is very wide. We can insert it in the action and in the 443 effective Hamiltonian: 444

$$\mathcal{A}(q_d, q_0, h) = -\frac{1}{2}\sigma^2\beta^2 \left(\frac{n(n-1)}{2}q_0^2 + \frac{n}{2}q_d^2\right) + \frac{\beta\mu n}{2}h^2 + \frac{1}{S}\sum_i \ln Z_i(q_d, q_0, h)$$
(49)

445

$$H_{eff} = -\frac{\beta \sigma^2}{2} \left( q_0 \left( \sum_a N_i^a \right)^2 + (q_d - q_0) \sum_a (N_i^a)^2 \right) + \sum_a \left( \mu h N_i^a + V(N_i^a) + (T - \lambda) \ln N_i^a \right)$$
(50)

To decouple the different replicas we introduce a Gaussian integration over an auxiliary variable  $z_i$ , thus obtaining:

$$Z_{i} = \int_{-\infty}^{+\infty} \frac{dz_{i}}{\sqrt{2\pi}} e^{-z_{i}^{2}} \left[ \int dN_{i} \exp\left\{-\beta H_{\text{eff}}^{RS}(N_{i};q_{d},q_{0},h;z)\right\} \right]^{n}$$
(51)

448

$$H_{\rm eff}^{RS} = -\beta \sigma^2 \frac{q_d - q_0}{2} N_i^2 + (\mu h - z \sqrt{q_0} \sigma) N_i + V(N_i) + (T - \lambda) \ln N_i$$
(52)

One of these terms is proportional to a fluctuating field z, which accounts for the randomness of the interactions. Considering the stationary condition, performing the analytic continuation for  $n \rightarrow 0$  and taking into account that after the integration over the fluctuating

 $<sup>{}^{2}</sup>H_{a}$  would actually need to be integrated over the imaginary axis; deforming the integration contour and using the method of steepest descent would then yield the stated result.

452 field z all species are equivalent, the self-consistent equations (47) become:

$$h = \int \mathcal{D}z \left( \frac{\int_{0}^{\infty} dN e^{-\beta H_{\text{eff}}^{RS}(q_{0},q_{d},h,z)} N}{\int_{0}^{\infty} dN e^{-\beta H_{\text{eff}}^{RS}(q_{0},q_{d},h,z)}} \right) = \overline{\langle N \rangle}$$

$$q_{d} = \int \mathcal{D}z \left( \frac{\int_{0}^{\infty} dN e^{-\beta H_{\text{eff}}^{RS}(q_{0},q_{d},h,z)} N^{2}}{\int_{0}^{\infty} dN e^{-\beta H_{\text{eff}}^{RS}(q_{0},q_{d},h,z)}} \right) = \overline{\langle N^{2} \rangle}$$

$$q_{0} = \int \mathcal{D}z \left( \frac{\int_{0}^{\infty} dN e^{-\beta H_{\text{eff}}^{RS}(q_{0},q_{d},h,z)} N}{\int_{0}^{\infty} dN e^{-\beta H_{\text{eff}}^{RS}(q_{0},q_{d},h,z)}} \right)^{2} = \overline{\langle N \rangle^{2}}$$
(53)

where we used the calligraphic notation for the Gaussian integral in z:  $\int Dz \equiv \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} e^{-z^2/2}$ . The brackets indicate the average over the Boltzmann distribution under the effective Hamiltonian, while the overbar denotes the average over the disorder, represented by the Gaussian unit-variance variable z. The two averages coincide with a single species abundance's thermal and disorder average, respectively. Given the resulting values of the above set of equations, one can eventually compute the disordered free energy as:

$$f = \frac{F}{S} = -\lim_{n \to 0} \frac{\ln \overline{Z^n}}{n\beta S} = -\lim_{n \to 0} \frac{\mathcal{A}(h, q_0, q_d)}{\beta n} .$$
(54)

Solving these equations analytically at finite  $\beta$  is impractical. It is thus necessary to employ a numerical scheme that, given an initial guess, allows us to solve them iteratively.

Algorithmic protocol

h

- Initialize the mean abundance *h*, and the two overlaps  $q_d$ ,  $q_0$  at the initial time;
- Solve Eqs. (53) iteratively upon progressively increasing  $\beta = 1/T$ . The introduction of a damping parameter for instance  $\alpha = 0.1$  can be beneficial to facilitate the convergence.

Therefore, by looking at the first equation, one should consider:

$${}^{t} \leftarrow \alpha \int \mathcal{D}z \frac{\int_{N_{c}}^{\infty} dN e^{-\beta H_{\mathrm{RS}}(q_{0}^{t-1}, q_{d}^{t-1}, h^{t-1}, z)} N}{\int_{N_{c}}^{\infty} dN e^{-\beta H_{\mathrm{RS}}(q_{0}^{t-1}, q_{d}^{t-1}, h^{t-1}, z)}} + (1-\alpha)h^{t-1}$$

 The algorithm is expected to converge if the error between the (t − 1)-value and t-value ≤ ε, a given arbitrary precision ε, for each of the three order parameters.

461

#### 462 3.5 Numerical interpretation of the order parameters

To validate the replica symmetric ansatz, one can simulate the dynamical system defined in Eq. (30). To generate a single ecosystem realization:

- First, one needs to sample the  $S \times S$  interaction matrix with rescaled parameters ( $\tilde{\mu}, \tilde{\sigma}^2$ );
- One sets up the initial conditions  $N_i(t = 0)$ , by assuming for instance a uniform distribution in [0, 1];
- Then, demographic fluctuations are sampled from a white-noise distribution with amplitude *T*.

The dynamical equations can then be integrated deterministically up to a given  $t_{\text{max}}$ , which denotes the maximal extent of the simulation. Also, to analyze variability, it is beneficial to perform the simulation for  $N_{\text{samples}}$  independent realizations, each obtained by varying the sources of randomness. This process produces a dataset  $\{N_i^p(t)\}_{t=0,...,t_{\text{max}};i=1,...,S;p=1,...,N_{\text{sample}}}$ . Eventually averaging over all these different contributions, one can have access to the dynamical correlator:

$$\mathbb{E}[N(t)N(t')] = \frac{1}{S} \frac{1}{N_{\text{sample}}} \sum_{i=1}^{S} \sum_{p=1}^{N_{\text{sample}}} N_i^p(t) N_i^p(t') , \qquad (55)$$

which, for sufficiently large samples ( $N \simeq 50-60$ ), shows a convergence in law. Moreover, at sufficiently large times, the correlator appears to satisfy the time translational invariance (TTI) property being a function only of the time difference:  $\mathbb{E}[N(t)N(t')] = C(t-t'), \forall t \ge t' > t_{wait}$ . The waiting time depends on the parameters' choice, notably ( $\sigma$ , T); however, in the single equilibrium phase, we are guaranteed that the dynamics converge to the TTI state and  $t_{wait}$ turns out to be typically  $10^2$ .



The (left) Figure shows the numerical comparison for two characteristic one-time observables: the self-overlap  $q_d$  and the mean abundance h in the single-equilibrium (RS) regime. The dashed line corresponds to the theoretical prediction; the full line results from the integration of Dynamical Mean-Field Theory equations. For more details see [27, 33].

481

482 We eventually find:

$$h = \mathbb{E}[N(t)], \quad q_d = C(0) = [N(t)^2], \quad q_0 = \lim_{\tau \to \infty} C(\tau) \simeq \mathbb{E}[N(t)N(t_{\max})], \quad (56)$$

where the L.H.S. is computed through the replica method and the R.H.S. provides the numer-ical comparison.

#### 485 3.6 Stability analysis: Single equilibrium versus multiple equilibria

In the symmetric interaction case, for  $A_{ij} = A_{ji}$ , the RS solution becomes unstable at  $\sigma_c = \frac{1}{\sqrt{2}}$ . In the asymmetric case, the instability bound is slightly modified to take the correlation between asymmetric coefficients into account [34]. Without correlation, the transition from the unique equilibrium to multiple attractors turns out to be independent of  $\mu$  and lies on the line  $\sigma = \sqrt{2}$ . For a generic asymmetry  $\rho$ , on the other hand,  $\sigma_c = \sqrt{2}/(1 + \rho)$ . Increasing the symmetry shifts the two transitions towards lower variance and stronger interactions. In this light, predation-prey relationships may contribute to stabilizing the community.

To investigate the stability of the different phases, we introduce the Hessian matrix of the free energy, which allows us to study the harmonic fluctuations in terms of the matrix  $\delta Q_{ab}$ . Thanks to the symmetry group properties of the replica space, the diagonalization of the Hessian matrix can be expressed as a function of three different contributions. Following [31, 35], we define three sectors: the *longitudinal*,  $\lambda_{\rm L}$ , the *anomalous*,  $\lambda_{\rm A}$ , and the *replicon*,  $\lambda_{\rm R}$ .

An instability leading to the breaking of replica symmetry is signaled by the Hessian matrix, projected on the so-called replicon sector, no longer being strictly positive definite. If its smallest eigenvalue reaches zero, this signals the onset of an instability, analogous to the de
 Almeida-Thouless line, as originally identified in the Sherrington-Kirkpatrick model for spin
 glasses [36].

Conversely, a zero longitudinal mode provides insights into spinodal points describing how a state opens up along an unstable direction. This phenomenon is particularly relevant for understanding the emergence of *tipping points* and catastrophic collapses, which typically occur through first-order phase transitions. To be more detailed, we consider the matrix of the second derivatives of the action, as in Eq. (42), differentiated with respect to the overlap matrix:

$$\mathcal{M}_{abcd} \equiv -\frac{\partial^2 \mathcal{A}}{\partial Q_{ab} \partial Q_{cd}} = \beta^2 \rho^2 \sigma^2 \Big[ \delta_{(ab),(cd)} - (\beta^2 \rho^2 \sigma^2) \overline{\langle N^a N^b, N^c N^d \rangle_c} \Big] .$$
(57)

<sup>509</sup> The subscript  $\langle \cdot \rangle_c$  denotes the connected part of the correlation function.

It should be noted that the limit  $n \rightarrow 0$  implies a negative multiplicity for *Q*s, thus entailing a change in the convexity of the free energy at the stationary point. Therefore, one should evaluate the corrections to the saddle point and select only those solutions with a positivedefined second derivative of the action.

<sup>514</sup> Within the replica formalism, Eq. (57) can eventually be recast as a function of three <sup>515</sup> different correlators, depending on the choice of the replica indices:

$$\mathcal{M}_{abcd} = M_{ab,ab} \left( \frac{\delta_{ac} \delta_{bc} + \delta_{ad} \delta_{bc}}{2} \right) + M_{ab,ac} \left( \frac{\delta_{ac} + \delta_{bd} + \delta_{ad} + \delta_{bc}}{4} \right) + M_{ab,cd} .$$
(58)

<sup>516</sup> In the specific case of the replicon sector, the associated eigenvalue reads:

$$\lambda_{\rm R} = (\beta \rho \sigma)^2 \left[ 1 - (\beta \rho \sigma)^2 \overline{\left( M_{ab,ab} - 2M_{ab,ac} + M_{ab,cd} \right)} \right]$$
(59)

where each of these contributions can be rewritten as a function of the species abundance correlators as follows

$$M_{ab,ab} - 2M_{ab,ac} + M_{ab,cd} = \left[ \langle (N^a)^2 (N^b)^2 \rangle - 2 \langle (N^a)^2 N^b N^c \rangle + \langle N^a N^b N^c N^d \rangle \right] .$$
(60)

As usual, the average  $\langle \cdot \rangle$  is performed over the effective Hamiltonian, while the disordered one is denoted by  $\overline{\cdot}$ . In the presence of a single equilibrium – namely in the replica-symmetric approximation – this leads to the following replicon eigenvalue:

$$\lambda_{\rm R} = (\beta \rho \sigma)^2 \left[ 1 - (\beta \rho \sigma)^2 \overline{(\langle N^2 \rangle - \langle N \rangle^2)^2} \right].$$
(61)

The averaged difference in the parenthesis accounts for fluctuations between the first and 522 second moment of the species abundances within one state, namely between the diagonal 523 value  $q_d$  and the off-diagonal contribution  $q_0$  of the overlap matrix. Such a difference can also 524 be interpreted as the response function of the single species to an infinitesimal perturbation. 525 The identification of a vanishing replicon eigenvalue is intricately linked to the emergence 526 of marginal states and diverging susceptibilities. This aspect is also pertinent in the context of 527 out-of-equilibrium aging dynamics. This stability analysis procedure can be repeated as many 528 times as needed, ranging from the 1RSB ansatz to the full-RSB solution. In the 1RSB scenario, 529 530 a key difference lies in the distinction between the intra-state average, computed within a single replica state, and the inter-state average, computed over a block of replicas of size m. 531 In a similar manner, the replicon eigenvalue reads: 532

$$\lambda_{\rm R}^{\rm 1rsb} = (\beta \rho \sigma)^2 \left[ 1 - (\beta \rho \sigma)^2 \overline{\langle \left( \langle N^2 \rangle_{\rm 1r} - \langle N \rangle_{\rm 1r}^2 \right)^2 \rangle_{m \cdot \rm r}} \right], \tag{62}$$

The point at which the 1RSB replicon mode vanishes marks a critical phase transition to a more structured phase.



Figure 3: Pictorial representation of a replica-symmetric landscape for the Generalized Lotka-Volterra model, depicted by a single fixed point (left); fractal landscape structure in the low-demographic noise phase, formally captured by a full-RSB solution (right).

#### <sup>535</sup> 4 Derivation of the phase diagram at finite noise

<sup>536</sup> Different regimes associated with increasing complexity can be pointed out by varying the <sup>537</sup> demographic noise amplitude, *T*, and the heterogeneity of the interactions,  $\sigma^2$  as pertinent <sup>538</sup> control parameters of the model. Below the blue line in Fig. (4), which reflects the vanishing <sup>539</sup> behavior of RS replicon eigenvalue, the landscape structure is no longer identified by a single <sup>540</sup> equilibrium. It would be depicted, instead, by a more complex, two-level structure, which <sup>541</sup> turns out to be associated with an additional order parameter as well.

Remarkably, upon further decreasing the demographic noise and increasing the variance 542 of the interaction matrix  $\sigma^2$ , the system undergoes a *Gardner transition* to a marginally stable, 543 amorphous phase, terminology borrowed from glassy physics. In this regime, each state – 544 also referred to as a basin – fragments into a fractal hierarchy of sub-basins (metabasins), as 545 illustrated in Fig. (3), right panel. As a result, the internal structure of the emerging sub-546 basins is formally described by the full replica symmetry breaking (full-RSB) solution of the 547 partition function (below the orange line, Fig. (4)). This regime is characterized by an infinite 548 hierarchy of broken symmetries in the order parameter [30]. In other words, the original 549 piecewise function parametrizing the overlap matrix will converge to a continuous function 550 q(x), with  $x \in [0, 1]$ . 551

Analytical calculations in high-dimensional hard-sphere systems and continuous constraint 552 satisfaction problems have allowed for the prediction of this type of transition, supporting the 553 emergence of a fractal hierarchy [37]. Hence, the infinite-dimensional theory corroborates the 554 hypothesis that an equilibrium glass state undergoes a structural change, transitioning from a 555 normal glass to a marginal one in the low-temperature regime. Remarkably, for the first time 556 in an ecological context, an analogy between low-temperature glassy systems and complex en-557 ergy landscapes in the Generalized Lotka-Volterra model has been highlighted in [27], building 558 on prior studies in this direction [20, 26, 34]. However, what a replica symmetry-breaking sce-559 nario precisely yields for biological and ecological communities – whether linked to a finite or 560 a hierarchical sequence of slower and slower timescales in the dynamics – represents still an 561 open and intriguing question. 562

#### 563 4.1 Zero-temperature limit of the RS solution

At  $T \to 0$ , the two order parameters tend to become degenerate, i.e.  $q_0 \to q_d$ . It is therefore useful to define an auxiliary quantity  $\Delta q = \beta \rho (q_d - q_0)$  that remains finite. In this limit, the integrals are exactly solvable given the vanishing trend of the logarithmic term in the Hamiltonian (35) and their consequent Gaussian nature.



Figure 4: Phase diagram of the GLV model in the presence of finite demographic fluctuations and random symmetric interactions, i.e.  $\alpha_{ij} = \alpha_{ji}$ . At sufficiently high demographic noise, the landscape is described by a single fixed point (above the blue line). At low-noise and high-heterogeneity values, the dynamics get trapped in different multiple-equilibria phases, which are captured either by a 1-RSB scenario or a full-RSB one. Figure adapted from [27].

The self-consistent equations for the order parameters turn out to be then:  

$$\begin{split} h &= \overline{N^*(z)} , \\ q_0 &= \overline{N^*(z)^2} , \\ \Delta q &= \frac{\overline{\theta(N^*(z)})}{H_{\rm RS}''(N^*(z))} \end{split} \tag{63}$$

568

where the  $\theta(\cdot)$  function in the last expression accounts only for the positive-definite species abundances. The saddle-point approximation eventually yields for the typical value:

$$N^{*}(z) = \max\left[0, \frac{K - \mu h + z\sigma\sqrt{q_{0}}}{1 - \Delta q\sigma^{2}}\right].$$
(64)

Equations (63) are explicitly written in the symmetric case, for  $\rho = 1$ , where the replica method safely applies. They immediately recall those obtained by the cavity method, also implying:

$$N^{*}(z) = \max\left[0, \frac{K - \mu \langle N \rangle + z\sigma \sqrt{q_{0}}}{1 - \rho \sigma^{2} \chi}\right], \qquad (65)$$

where  $\rho$  stands for the generic correlation between the interaction matrix couplings.

#### 575 4.2 Conclusions and Perspectives

<sup>576</sup> We shall now discuss a few exciting research directions that could pave the way not only <sup>577</sup> for a better theoretical understanding of large ecosystems – interpreted through the lens of disordered systems – but also for a more immediate detection of empirical macroecological
 patterns and quantitative contributions to biomedical research.

## 4.2.1 Dynamical Mean-Field Theory approach: asymmetric interactions and spatial de pendence

Non-reciprocal interactions break detailed balance. When the dynamics are no longer governed by conservative forces, defining a free energy and minimizing its Hessian on a proper subsector becomes unfeasible. However, asymmetric interactions can give rise to fascinating phenomena, including an exponential number of attractors, limit cycles, and chaotic behavior. We now revisit the dynamical equations for the Generalized Lotka-Volterra model with arbitrary correlation  $\rho$  between the interaction coefficients  $A_{ij}$ :

$$\frac{dN_i}{dt} = N_i \left[ (1 - N_i) - \sum_{j, (j \neq i)} A_{ij} N_j \right] + \eta_i(t) + \lambda_i .$$
(66)

In the large system size limit, for  $S \rightarrow \infty$ , this many-body problem can be conveniently rewritten in terms of a single-variable self-consistent stochastic process, also referred to as Dynamical Mean-Field Theory (DMFT):

$$\frac{dN(t)}{dt} = N(t) \left[ 1 - N(t) - \mu h(t) - \sigma \zeta(t) + \rho \sigma^2 \int_0^t dt' R(t, t') N(t) + H_{\text{ext}}(t) \right]$$
(67)

where  $h(t) = \mathbb{E}[N(t)]$  represents the mean abundance evolving in time,  $C(t, t') = \mathbb{E}[N(t)N(t')]$ 591 the correlation function, and  $R(t, t') = \mathbb{E}\left[\frac{\delta N(t)}{\delta H_{\text{ext}}(t')}\Big|_{H_{\text{ext}}=0}\right]$  the response function, thus com-592 puted w.r.t an external field  $H_{ext}$ . The notation  $\mathbb{E}[\cdot]$  denotes the average on all different sources 593 of randomness: random couplings, initial conditions, and demographic noise. The derivation 594 is similar to that to obtain the Langevin equation from Newtonian dynamics. However, an 595 additional layer of complexity arises here because the statistical properties of the effective 596 thermal bath – accounting for the interactions with the rest of the system – must be deter-597 mined self-consistently. 598

This DMFT formalism has been widely applied to a variety of interdisciplinary contexts, 599 starting from strongly correlated electrons, to the analysis of rheological properties of amor-600 phous systems under shear deformations, up to active matter. Moreover, this approach allows 601 us to go beyond a purely well-mixed approximation and model a multi-species metacommunity 602 in the presence of external noise and random interactions [38]. In the simultaneous limits of 603 an infinite number of species and spatial patches, the phase diagram can be determined an-604 alytically. For sufficiently large demographic fluctuations, the transition between an active 605 phase – where most species persist – and an inactive phase is continuous, falling within the 606 universality class of Directed Percolation. However, at lower values of demographic noise, the 607 scenario changes dramatically. In this regime, the transition becomes discontinuous, unveiling 608 novel features as well as the pivotal role of random interactions. 609

The effect of migration between distinct patches in the strong heterogeneity regime was also investigated in the presence of non-symmetric interactions and zero noise [39,40]. In the latter, chaotic behaviors associated with long-lived persistent fluctuations can take place.

#### 613 4.2.2 Heavy-tailed distributions

The RS solution discussed thus far predicts a truncated Gaussian distribution for the typical species abundances, which is known to be not the kind of distribution observed in Nature [41]. Nearly all the continuous statistical distributions (normal, exponential, gamma) have light
tails. However, when talking about ecological collapses and external perturbations, such as the
distribution of the largest fires within a limited region [42], the rate of population spread [43]
or even hurricanes [44], heavy-tailed distributions seem to be the right answer.

Moreover, one might wonder whether the phase transitions obtained in high-dimensional phase diagrams (multiple equilibria, Gardner, etc.) remain true beyond the mean-field approximation, although there is by now remarkable experimental evidence of multiple alternative stable states. In light of this, scaling and crossover theory might account for the long-range fluctuations that the mean-field approximation ignores.

Within the framework of disordered systems, accurately representing more realistic distri-625 butions requires incorporating two additional key ingredients: asymmetric interactions and 626 multiple fixed-point scenarios. Indeed, asymmetric interactions between species can result in 627 chaotic dynamics, in which a sharp timescale separation between many rare species – respon-628 sible for power-law trends – and a few abundant ones, associated with sudden turnovers can 629 be detected [45-47]. It is quite acknowledged that such heavy tails arise as an emergent prop-630 erties of large ecological communities, the result of the interplay between interactions and 631 stochasticity [48]. How a large number of species can coexist in those complex communities 632 and what is the mechanism according to which they are mostly dominated by rare species is 633 still not completely understood. Over the years several efforts have been done to recover such 634 distributions from simple statistical or mechanistic models: 635

- Heavy-tailed distributions can be obtained within an individual-based model (IBM) with 636 a maximum number of individuals [49], thus counterbalancing immigration of new 637 species with the interactions between species. In other words, the IBM are stochastic 638 cellular automatons in which immigration, growth rates, and extinction events can hap-639 pen at every time step with the rates. All individuals of a species are assumed to be 640 equivalent (same growth rates, same self-interaction strengths, same immigration pa-641 rameters). Then, species abundances evolve stochastically. [49, 50] showed separately 642 that depending on the immigration rate, both power law and lognormal abundance dis-643 tributions can be observed. By using experimental data of plankton and microbial com-644 munities, it has been found out that a log-normal distribution, rather than a power law, 645 fits best the distribution of abundances of the most abundant species. 646
- Alternatively, [48] proposes a stochastic GLV model with a global maximal capacity –
   meaning that communities are limited by finite resources and limited space taking
   both the immigration and the degree of connectivity into account as relevant control
   parameters. At high connectivity, the abundance distribution allows for the emergence
   of non-trivial power-law behaviors.

The abundance distribution of microbial communities is typically a lognormal. In IBMs, 652 broad distributions are the result of the system being self-organized at the edge of sta-653 bility [49]. Immigration increases the number of species, whereas interactions between 654 species cause individual extinction. Both the IBM and simplistic models with maximal 655 capacity give rise, in different ways, to heavy-tailed rank abundance distributions. No-656 tably, in the presence of logistic equations, a uniform distribution of the self-interaction 657 leads to a power-law abundance distribution. In contrast, lognormal distributions of the 658 growth rate and the self-interaction term give rise to lognormal trends. 659

#### 660 4.2.3 Data-driven approaches

The growing availability of observational datasets, particularly for microbial communities, allows increasingly detailed characterizations of local communities. Recent studies indicate that samples of the same community type, collected from various spatial or temporal locations,
hosts, etc. frequently exhibit shared statistical properties [29, 51, 52]. These findings support
the application of ensemble averages as representative markers of distinct community states,
a concept fundamental to statistical physics methodologies [53].

However, comparing model predictions with empirical abundance distributions as a crite-667 rion for model selection may lead to several challenges. Three notable approaches are usually 668 employed in this regard. First, longitudinal datasets reveal temporal changes in community 669 composition, enabling more precise comparisons with theoretical models. Although determin-670 istic and stochastic models may predict similar species abundance distributions, they differ in 671 the timescales of fluctuations and resulting abundance trajectories. A key challenge requires 672 distinguishing measurement errors, noise, and chaotic dynamics. New non-parametric time-673 series analysis methods are addressing these limitations [54]. Moreover, time-series proper-674 ties like critical slowdown can signal approaching tipping points, potentially leading to catas-675 trophic shifts in ecosystem states [55]. Detecting such transitions is crucial for ecological 676 forecasting and decision-making, where statistical physics methods can play a fundamental 677 role. 678

Second, incorporating ecological functions – such as productivity, metabolic activity, or trophic structures – into community descriptions can enhance the theoretical accuracy. The definition of *functional groups*, accounting for the biological knowledge of how species with a similar role contributes to ecosystem functions, often reduces variability [56, 57]. This multilayer approach enables better model selection and motivates new theoretical directions.

Third, some microbial communities exhibit ergodicity breaking, thus resulting in multiple attractor states [58–60]. In light of this, since multistability reflects distinct community properties [61,62], there is hope that the inferred information from data could improve predictions on species compositions and emerging patterns, also for classifying specific ecosystem services and distinct diseases in microbial dynamics [25,63].

#### **Acknowledgements**

The author thanks Abishek Dhar, Joachim Krug, Satya N. Majumdar, Alberto Rosso, and Grégory Schehr for the organization of Les Houches Summer School 2024, *Theory of Large Deviations and Applications*, where these lectures were delivered. The author also acknowledges Yizhou Liu, Nirbhay Patil, and Frédéric van Wijland for a careful reading of the manuscript. Explicit permission to reuse part of the material, specifically in Sections 3.2, 3.4, and 3.6,

<sup>695</sup> has been granted under APS license number RNP/25/JAN/087602.

Funding information A.A. acknowledges the support received from the Agence Nationale
de la Recherche (ANR) of the French government, under the grant ANR-23-CE30-0012- 01
("SIDECAR" project).

#### 699 References

[1] K. S. McCann, *The diversity-stability debate*, Nature **405**(6783), 228 (2000),
 doi:10.1038/35012234.

[2] I. A. Hatton, O. Mazzarisi, A. Altieri and M. Smerlak, *Diversity begets stability: Sublin- ear growth and competitive coexistence across ecosystems*, Science 383(6688), eadg8488

#### **SciPost Physics Lecture Notes**

- (2024), doi:10.1126/science.adg8488, Publisher: American Association for the Ad vancement of Science.
- [3] K. Van Meerbeek, T. Jucker and J.-C. Svenning, Unifying the concepts of stability and
   *resilience in ecology*, Journal of Ecology **109**(9), 3114 (2021).
- <sup>708</sup> [4] D. R. Amor, C. Ratzke and J. Gore, *Transient invaders can induce shifts between alternative* <sup>709</sup> stable states of microbial communities, Science advances **6**(8), eaay8676 (2020).
- [5] G. Aguadé-Gorgorió, J.-f. Arnoldi, M. Barbier and S. Kéfi, *A taxonomy of multiple stable* states in complex ecological communities, Ecology Letters 27(4), e14413 (2024).
- [6] S. P. Hubbell, *The Unified Neutral Theory of Biodiversity and Biogeography (MPB-32)*,
  Princeton University Press, ISBN 978-1-4008-3752-6, doi:10.1515/9781400837526,
  Publication Title: The Unified Neutral Theory of Biodiversity and Biogeography (MPB-32) (2011).
- [7] S. Azaele, S. Suweis, J. Grilli, I. Volkov, J. R. Banavar and A. Maritan, *Statistical mechanics* of ecological systems: Neutral theory and beyond, Rev. Mod. Phys. 88(3), 31 (2016).
- [8] J. Chave, Neutral theory and community ecology, Ecology Letters 7(3), 241 (2004),
   doi:10.1111/j.1461-0248.2003.00566.x.
- [9] J. Rosindell, S. P. Hubbell, F. He, L. J. Harmon and R. S. Etienne, *The case for ecological neutral theory*, Trends in Ecology & Evolution 27(4), 203 (2012), doi:10.1016/j.tree.2012.01.004.
- [10] A. Lotka, *Elements of physical biology*, Williams and Wilkins (1925).
- [11] P. J. Wangersky, Lotka-volterra population models, Annual Review of Ecology and Systematics 9, 189 (1978).
- [12] J. Hofbauer and K. Sigmund, *Evolutionary Games and Population Dynamics by Josef Hof- bauer*, doi:10.1017/CBO9781139173179 (1998).
- <sup>728</sup> [13] *The paradox of the plankton*, The American Naturalist **95**(882), 137 (1961).
- [14] R. MacArthur and R. Levins, *The limiting similarity, convergence, and divergence of coexisting species*, The american naturalist **101**(921), 377 (1967).
- [15] R. MacArthur, Species packing and competitive equilibrium for many species, Theoretical
   Population Biology 1(1), 1 (1970), doi:10.1016/0040-5809(70)90039-0.
- [16] D. Tilman, *Resources: A Graphical-Mechanistic Approach to Competition and Predation*,
   The American Naturalist **116**(3), 362 (1980), doi:10.1086/283633.
- [17] W. Cui, R. Marsland III and P. Mehta, Les Houches Lectures on Community Ecology: From
   Niche Theory to Statistical Mechanics, ArXiv:2403.05497 [cond-mat, q-bio] (2024).
- [18] M. Tikhonov and R. Monasson, *Collective phase in resource competition in a highly diverse ecosystem*, Physical review letters **118**(4), 048103 (2017).
- [19] A. Batista-Tomás, A. De Martino and R. Mulet, *Path-integral solution of macarthur's resource-competition model for large ecosystems with random species-resources couplings*, Chaos: An Interdisciplinary Journal of Nonlinear Science **31**(10) (2021).

- [20] A. Altieri and S. Franz, Constraint satisfaction mechanisms for marginal stability and criticality in large ecosystems, Physical Review E 99(1), 010401 (2019),
  doi:10.1103/PhysRevE.99.010401.
- [21] S. Allesina and S. Tang, *Stability criteria for complex ecosystems*, Nature 483(7388), 205
  (2012).
- [22] Y. Liu, J. Hu, H. Lee and J. Gore, *Complex ecosystems lose stability when resource con- sumption is out of niche*, Physical Review X 15(1), 011003 (2025).
- <sup>749</sup> [23] M. L. Mehta, *Random matrices*, Elsevier (2004).
- <sup>750</sup> [24] A. Altieri, *The jamming transition*, Springer (2019).
- [25] J. Pasqualini, A. Maritan, A. Rinaldo, S. Facchin, E. Savarino, A. Altieri and S. Suweis,
   *Microbiomes through the looking glass*, arXiv preprint arXiv:2406.07465 (2024).
- [26] G. Biroli, G. Bunin and C. Cammarota, *Marginally stable equilibria in critical ecosystems*,
   New Journal of Physics 20(8), 083051 (2018), doi:10.1088/1367-2630/aada58.

[27] A. Altieri, F. Roy, C. Cammarota and G. Biroli, Properties of Equilibria and Glassy Phases
 of the Random Lotka-Volterra Model with Demographic Noise, Physical Review Letters
 126(25), 258301 (2021), doi:10.1103/PhysRevLett.126.258301.

- [28] A. Altieri and G. Biroli, *Effects of intraspecific cooperative interactions in large ecosystems*,
   SciPost Physics 12(1), 013 (2022).
- [29] J. Grilli, Macroecological laws describe variation and diversity in microbial communities,
   Nature Communications 11(1), 4743 (2020), doi:10.1038/s41467-020-18529-y, Number: 1 Publisher: Nature Publishing Group.
- [30] M. Mézard, G. Parisi and M. A. Virasoro, *Spin glass theory and beyond: An Introduction* to the Replica Method and Its Applications, vol. 9, World Scientific Publishing Company
   (1987).
- [31] C. De Dominicis and I. Giardina, *Random fields and spin glasses: a field theory approach*,
   Cambridge University Press (2006).
- [32] A. Altieri and M. Baity-Jesi, *An introduction to the theory of spin glasses*, Encyclopedia of
   Condensed Matter Physics pp. 361–370 (2024).
- [33] F. Roy, G. Biroli, G. Bunin and C. Cammarota, Numerical implementation of dynamical mean field theory for disordered systems: application to the Lotka–Volterra model of ecosystems, Journal of Physics A: Mathematical and Theoretical 52(48), 484001 (2019), doi:10.1088/1751-8121/ab1f32.
- [34] G. Bunin, Interaction patterns and diversity in assembled ecological communities, arXiv
   preprint arXiv:1607.04734 (2016).
- <sup>776</sup> [35] A. J. Bray and M. A. Moore, *Evidence for massless modes in the'solvable model'of a spin* <sup>777</sup> glass, Journal of Physics C: Solid State Physics **12**(11), L441 (1979).
- [36] Stability of the sherrington-kirkpatrick solution of a spin glass model, Journal of Physics
  A: Mathematical and General 11(5), 983 (1978).
- [37] P. Charbonneau, J. Kurchan, G. Parisi, P. Urbani and F. Zamponi, *Fractal free energy landscapes in structural glasses*, Nature communications 5(1), 3725 (2014).

- [38] G. Garcia Lorenzana, A. Altieri and G. Biroli, *Interactions and Migration Rescuing Ecolog- ical Diversity*, PRX Life 2(1), 013014 (2024), doi:10.1103/PRXLife.2.013014.
- [39] F. Roy, M. Barbier, G. Biroli and G. Bunin, *Complex interactions can create persistent fluctuations in high-diversity ecosystems*, PLoS computational biology 16(5), e1007827 (2020).
- [40] 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).
- [41] C. T. Callaghan, L. Borda-de Água, R. van Klink, R. Rozzi and H. M. Pereira, Unveiling
   global species abundance distributions, Nature Ecology & Evolution pp. 1–10 (2023),
   doi:10.1038/s41559-023-02173-y.
- [42] M. A. Moritz, Analyzing extreme disturbance events: fire in los padres national forest,
   Ecological Applications 7(4), 1252 (1997).
- <sup>795</sup> [43] J. S. Clark, M. Lewis and L. Horvath, *Invasion by extremes: population spread with variation in dispersal and reproduction*, The American Naturalist **157**(5), 537 (2001).
- [44] R. W. Katz, Stochastic modeling of hurricane damage, Journal of Applied Meteorology
   41(7), 754 (2002).
- [45] D. A. Kessler and N. M. Shnerb, *Generalized model of island biodiversity*, Physical Review
   E 91(4), 042705 (2015).
- [46] Y. Xue, H. Chen, J. R. Yang, M. Liu, B. Huang and J. Yang, *Distinct patterns and processes* of abundant and rare eukaryotic plankton communities following a reservoir cyanobacterial
   bloom, The ISME Journal 12(9), 2263 (2018).
- [47] D. M. Perkins, A. Perna, R. Adrian, P. Cermeño, U. Gaedke, M. Huete-Ortega, E. P. White
   and G. Yvon-Durocher, *Energetic equivalence underpins the size structure of tree and phy- toplankton communities*, Nature communications 10(1), 255 (2019).
- <sup>807</sup> [48] L. Descheemaeker, J. Grilli and S. de Buyl, *Heavy-tailed abundance distributions from* <sup>808</sup> *stochastic lotka-volterra models*, Physical Review E **104**(3), 034404 (2021).
- [49] R. V. Solé, D. Alonso and A. McKane, *Self-organized instability in complex ecosystems*,
  Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences **357**(1421), 667 (2002).
- [50] Z. Heyvaert, *Towards a control of our microbiota* (2017).
- [51] E. Ser-Giacomi, L. Zinger, S. Malviya, C. De Vargas, E. Karsenti, C. Bowler and
  S. De Monte, Ubiquitous abundance distribution of non-dominant plankton across the
  global ocean, Nature Ecology & Evolution 2(8), 1243 (2018), doi:10.1038/s41559-0180587-2, Publisher: Nature Publishing Group.
- <sup>817</sup> [52] Y. Gao, A. Abdullah and M. Wu, *A unifying model of species abundance distribution*, <sup>818</sup> doi:10.1101/2024.06.14.599104, BioRxiv 2024.06.14.599104 (2024).
- [53] J. Pasqualini, A. Maritan, A. Rinaldo, S. Facchin, E. Savarino, A. Altieri and S. Suweis,
   *Microbiomes Through The Looking Glass*, doi:10.48550/ARXIV.2406.07465, Version Number: 1 (2024).

#### SciPost Physics Lecture Notes

- [54] T. L. Rogers, B. J. Johnson and S. B. Munch, *Chaos is not rare in natural ecosystems*, Nature
  Ecology & Evolution 6(8), 1105 (2022), doi:10.1038/s41559-022-01787-y, Number: 8
  Publisher: Nature Publishing Group.
- [55] V. Dakos and S. Kéfi, *Ecological resilience: what to measure and how*, Environmental
  Research Letters 17(4), 043003 (2022), doi:10.1088/1748-9326/ac5767.
- [56] S. Louca, M. F. Polz, F. Mazel, M. B. N. Albright, J. A. Huber, M. I. O'Connor, M. Ackermann, A. S. Hahn, D. S. Srivastava, S. A. Crowe, M. Doebeli and L. W. Parfrey, *Function and functional redundancy in microbial systems*, Nature Ecology & Evolution 2(6), 936 (2018), doi:10.1038/s41559-018-0519-1.
- [57] 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(6401), 469 (2018), doi:10.1126/science.aat1168, Publisher: American Association
  for the Advancement of Science Section: Research Article.
- <sup>835</sup> [58] J. Hong, W. Xue and T. Wang, *Emergence of alternative stable states in microbial commu-*<sup>836</sup> *nities undergoing horizontal gene transfer*, eLife **13** (2025).
- E. S. Wright, R. Gupta and K. H. Vetsigian, *Multi-stable bacterial communities exhibit extreme sensitivity to initial conditions*, FEMS Microbiology Ecology 97(6), fiab073 (2021).
- [60] W. Lopes, D. R. Amor and J. Gore, *Cooperative growth in microbial communities is a driver* of multistability, Nature Communications 15(1), 4709 (2024), doi:10.1038/s41467 024-48521-9, Publisher: Nature Publishing Group.
- E. S. Wright and K. H. Vetsigian, *Inhibitory interactions promote frequent bista- bility among competing bacteria*, Nature Communications 7(1), 11274 (2016),
  doi:10.1038/ncomms11274.
- [62] S. Estrela, J. C. C. Vila, N. Lu, D. Bajić, M. Rebolleda-Gómez, C.-Y. Chang, J. E. Goldford, A. Sanchez-Gorostiaga and Sánchez, *Functional attractors in microbial community assembly*, Cell Systems 13(1), 29 (2022), doi:10.1016/j.cels.2021.09.011.
- [63] D. R. Amor, Smooth functional landscapes in microcosms, Nature Ecology & Evolution
  7(11), 1754 (2023), doi:10.1038/s41559-023-02214-6, Number: 11 Publisher: Nature
  Publishing Group.