Statistical physics of community ecology: a cavity solution to MacArthur's consumer resource model

Madhu Advani,

Harvard University,Cambridge, MA 02142

Guy Bunin, and

Technion-Israel Institute of Technology, Haifa, 3200003, Israel

Pankaj Mehta[*]

Dept. of Physics, Boston University, Boston, MA 02215

Abstract

A central question in ecology is to understand the ecological processes that shape community structure. Niche-based theories have emphasized the important role played by competition for maintaining species diversity. Many of these insights have been derived using MacArthur's consumer resource model (MCRM) or its generalizations. Most theoretical work on the MCRM has focused on small ecosystems with a few species and resources. However theoretical insights derived from small ecosystems many not scale up large ecosystems with many resources and species because large systems with many interacting components often display new emergent behaviors that cannot be understood or deduced from analyzing smaller systems. To address these shortcomings, we develop a statistical physics inspired cavity method to analyze MCRM when both the number of species and the number of resources is large. Unlike previous work in this limit, our theory addresses resource dynamics and resource depletion and demonstrates that species generically and consistently perturb their environments and significantly modify available ecological niches. We show how our cavity approach naturally generalizes niche theory to large ecosystems by accounting for the effect of collective phenomena on species invasion and ecological stability. Our theory suggests that such phenomena are a generic feature of large, natural ecosystems and must be taken into account when analyzing and interpreting community structure. It also highlights the important role that statistical-physics inspired approaches can play in furthering our understanding of ecology.

I. INTRODUCTION

One of the most stunning aspects of the natural world is the diversity of species present in most ecosystems. The community structure of ecosystems are shaped through a complex interplay of the externally supplied resources available in an ecosystem, competition for these resources, as well as stochasticity [1–4]. A fundamental problem in community ecology is to understand how these processes give rise to observed pattern of species abundances. A rich theoretical framework has been developed to address this problem.

---

[*] pankajm@bu.edu.

Niche-based theories have emphasized the role of competition for resources [2, 5–10], while neutral theory has highlighted the role of stochastic effects [4, 11–13], and several works have investigated the interplay between stochasticity and competition [14–18].

Many of these theoretical insights have been synthesized in what is commonly referred to as contemporary niche theory. Contemporary niche theory highlights the role played by equalizing mechanisms, processes that decrease fitness differences between organisms, and stabilizing mechanisms, processes that decrease competition for resources. These basic organizational schema have been successfully applied to understand community structure in a wide range of settings [1–3].

One of the simplest and most influential mathematical models for niche theory is MacArthur's consumer resource model (MCRM) [2, 7, 8, 10]. Most analysis of MCRM – including those that inform contemporary niche theory and modern coexistence theory – have focused on small ecosystems with a few species and and few resources [2, 7, 8, 10]. However, it is unclear to what extent the theoretical insights derived from ecosystems with just a few species can be scaled up to diverse, natural ecosystems. One of the defining features of large complex systems is that they often display new "emergent behaviors" that cannot be understood or deduced from analyzing small systems with just a few parts [19–22]. For this reason, it is essential to directly analyze large ecosystems with many resources and species and ask how they differ from the few-species ecosystems that have been analyzed previously. Recently, several works suggest that large ecosystems can exhibit unexpected behaviors such as phase transitions, emergent community-level cohesion, and the analogues of critical points [15, 23–27]. This highlights the need for new theoretical frameworks for directly analyzing large, heterogeneous ecosystems.

Perhaps the most successful and ubiquitous approaches for analyzing large systems in statistical physics is mean field theory. We emphasize that what is meant by a mean field theories in statistical physics is distinct from the way it is commonly understood in ecology [28, 29]. Unlike most usages in ecology, mean field theories in physics account for not only the means of various quantities but also fluctuations around the mean. In this paper, whenever we use the term mean field theory, we will mean it in this broader statistical physics definition rather than the narrow usage common in ecology. Mean field models have long history in statistical physics and have played a central role in the study of phase transitions and collective emergent behaviors in physical systems [30, 31]. Most mean field theories in physics focus on homogenous systems with identical components and couplings. However, more sophisticated variants such as the cavity method can be used to analyze heterogeneous "disordered systems" [32]. Here, we develop a statistical physics inspired mean field theory, based on a generalization of the cavity method, and use it to analyze diverse ecosystems. In this paper, we will refer to this as the cavity method (CM).

Our methods are inspired by and build upon recent work showing the connection between community ecology the physics of disordered systems [15, 23, 24, 27, 33–38]. It is also closely related to the statistical mechanics of interacting socio-economic agents [39]. However, unlike these previous works our analysis explicitly incorporates resource dynamics, including resource heterogeneity and depletion. This allows us to naturally

connect our results to contemporary niche theory and modern coexistence theory. One of the most striking aspects of our analysis is that we nd that modification of fitness and carrying capacity due to collective phenomena is a generic feature of all diverse ecosystems [40]. In diverse ecosystems, organisms can and do significantly reshape their environments by changing resource abundances and, importantly, depleting resources. Moreover, we show that many of the central theoretical quantities in our novel CM have natural ecological interpretations that generalize many classical quantities and results of niche theory to large ecosystems and quantify the effect of collective phenomena in shaping community structure.

II. MACARTHUR CONSUMER RESOURCE MODEL

In this work, we will analyze one of the canonical and most influential models in community ecology: MacArthur's Consumer Resource Model (MCRM) [7, 8]. MCRM consists of S species or consumers with abundances N_i ($i = 1 \ldots S$) that can consume one of M substitutable resources with abundances R_a ($a = 1 \ldots M$). The consumer preferences of species i for resource a are encoded by a $S \times M$ matrix, c_{ia}.

In the MCRM, the growth rate $g_i(\mathbf{R})$ of a species depends of the concentration of all the resources. To model the growth rate, following MacArthur, we assume that a species i have some minimum maintenance cost, m_i, that they must meet. The growth rate, $gi(\mathbf{R})$, is proportional to amount of resources consumed, weighted by a quality factor w_α, minus this maintenance cost

$$g_i(\mathrm{R}) = \sum_\alpha C_{i\alpha} w_\alpha R_\alpha - m_i . \quad (1)$$

If $g_i > 0$, then this is also the growth rate of species i.

The resources can be considered as "biotic", i.e. organisms which are themselves being consumed, this is the scenario the original MCRM was designed to describe. These resources have their own internal dynamics which, following MacArthur, we assume can be modeled using logistic growth. Furthermore, when a resource is consumed, it's abundance is reduced. This ecological dynamics is captured by the following coupled, nonlinear differential equations

$$\frac{dN_i}{dt} = N_i g_i(\mathrm{R})$$
$$\frac{dR_\alpha}{dt} = F_\alpha(R_\alpha) - \sum_i N_i c_{i\alpha} R_\alpha, \quad (2)$$

where $F_a(R_a) = R_a(K_a - R_a)$ describes the resource dynamics in the absences of consumption and K_a is the carrying capacity of each resource a. Such dynamics are standard when modelling biotic resources. Note also that both the species and resource

abundances N_i and R_α must be strictly non-negative. For our analysis of the MCRM, it will be useful to define an "effective resource capacity"

$$K_\alpha^{eff}(\mathbf{N}) = K_\alpha - \sum_i \mathbf{N}_i c_{i\alpha} \quad (3)$$

that accounts for depletion of resources by consumers [23]. The MCRM can be rewritten in terms of $K_\alpha^{eff}(\mathbf{N})$ as

$$
\begin{aligned}
\frac{dN_i}{dt} &= N_i g_i(\mathbf{R}), \\
\frac{dR_\alpha}{dt} &= R_\alpha \big(K_\alpha^{eff}(\mathbf{N}) - R_\alpha \big).
\end{aligned}
\quad (4)
$$

A crucial property of these equations is that resources can be completely depleted from the environment. This will play an important role in what follows. Finally, we emphasize that these equations are identical those analyzed by MacArthur, Chesson, and others in deriving modern niche theory.

III.   STATISTICAL MECHANICS APPROACH TO MACARTHUR'S CONSUMER RESOURCE MODEL

Previous approaches to analyzing the MCRM have largely been confined to small ecosystems with a few species and resources. Here, we consider the opposite limit of large, diverse ecosystems where both the number of species and number of resources is large, S, $M \gg 1$. In this limit, the number of parameters needed to define the ecosystem dynamics becomes extremely large. To overcome this problem, we follow a long tradition in theoretical ecology pioneered by Robert May of looking at the case where the parameters are drawn from a random distribution [41]. This allows us to ask questions about the behavior of a generic, diverse ecosystem.

We consider the case where all the consumption coefficients $c_{i\alpha}$, resource carrying capacities K_a, and maintenance costs m_i are drawn from a random distribution. Our analytic calculations depend only on the mean and variances of the probability distributions. Denoting the expectation value of a parameters x over a distribution by $\langle x \rangle$, then we denote the mean and variances of our parameters by: $\langle c_{i\alpha} \rangle = \mu_c/S$, $\left\langle (c_{i\alpha} - \langle c_{i\alpha} \rangle)^2 \right\rangle = \sigma_c^2/S$, $\langle m_i \rangle = m$, $\left\langle (m_i - \langle m \rangle)^2 \right\rangle = \sigma_m^2$, $\langle K_\alpha \rangle = K$, and $\left\langle (K_\alpha - \langle K_\alpha \rangle)^2 \right\rangle = \sigma_K^2$. We can also define a parameter $\gamma = M/S$ that measures the ratio of resources to species. This description is consistent with all species being generalists, that is, feeding from many resources. An examples is a predator which can feed on many different species with each prey making up a small part of its diet. In such a scenario, it is reasonable that the flux from each resources decreases with the number of resources M, through the scaling of $c_{i\alpha}$.

A. Invasion, ecological stability, and self-consistency

One of the cornerstones of community ecology is the idea of invasion [6, 42, 43]. In our analysis, we will ask under what circumstances a new species can invade an ecosystem. Denote the growth rate of species i when it tries to invade the ecosystem g_i^{inv}. We will call this the invasion growth rate. Since we are interested in statistical properties, we will be primarily concerned with the mean and variances of the invasion growth rate averaged over all species i in the regional species pool: $\left\langle g_i^{inv} \right\rangle = g$ and $\left\langle \left(g_i^{inv} \right)^2 \right\rangle - \left\langle g_i^{inv} \right\rangle^2 = \sigma_g^2$.

The key idea that we will exploit in our analysis is the observation that as S and M get large, both the invasion growth rates, g_i^{inv}, and the effective carrying capacities, K_α^{eff} are the sum of a large number of small terms. Each individual resource makes only a small contribution (of order $1/M$) to the growth of any consumer, and every consumer makes an order $1/S$ contribution to the effective resource capacity. Thus, from the central limit theorem, the distribution of growth rates g_i^{inv} and the distribution of effective resources K_α^{eff} in the ecosystem can be well-approximated by a normal distribution. In the language of the cavity method of statistical physics, this corresponds to the replica symmetric solution. For future reference, denote the means and variance of the effective carrying capacity by $\left\langle K_\alpha^{eff} \right\rangle = K_\alpha^{eff}$ and $\left\langle \left(K_\alpha^{eff} \right)^2 \right\rangle - \left\langle K_\alpha^{eff} \right\rangle^2 = \sigma_{K^{eff}}^2$ (see Figure 1).

This suggests the following intuition for thinking about our ecosystem. Each species, i, has a invasion growth rate drawn from a normal distribution. In other words, we can think of $g_i^{inv} \approx g + \sigma_g z_i$, where z_i is a standard normal variable. Similarly, each resource has an effective carrying capacity that is also drawn from a normal distribution, with $K_\alpha^{eff} \approx K^{eff} + \sigma_K \tilde{z}_\alpha$, with $\tilde{z}_\alpha$ standard, normal variable. In general, the means and variances $\left(g, K^{eff}, \sigma_g^2, \sigma_{K^{eff}}^2 \right)$ depend on the abundances of all other species and resources. Our statistical mechanics inspired mean field approach exploits this observation to self-consistently solve for the means and variances of the invasion growth rate and effective resource carrying capacity. In the physics literature, these is known as cavity method (CM). In general, this is a very subtle calculation but can be done using a generalized cavity equation (see below and in appendix).

In order to derive the CM self-consistency equations, we consider a system with S species and M resources and ask what happens when we add an additional species and resource to the system. We denote the abundances of the additional species and resource by N_0 and R_0 respectively. This two-step cavity where both a resource and species is removed is similar to the procedure employed to analyze the Hopfield model and compressed sensing [44, 45] and is necessary to correctly capture subtle correlations between resource and species dynamics. This approach is intimately related to classic works by MacArthur and Levins that analyzed ecological dynamics by asking if a new species could invade an ecosystem [6]. Whereas their analysis was applicable to small ecosystems with a few species, our analysis is valid for large, diverse ecosystems.

Since the number of species and resources in the original ecosystem is large (S, $M \gg 1$), the addition of the new resource and consumer represent a small perturbation of the original system. For this reason, it is useful to define two susceptibilities, χ and ν that measure the sensitivity of an ecosystem to small perturbations. The resource susceptibility, χ, measures the average change in the mean resource abundance at steady-state if we slightly increase the supply of all the externally supplied resources. Denoting the steady-state value of a quantity

X by $\bar{X}$, we can mathematically define χ, as

$$\chi = \frac{1}{M} \sum_{\alpha} \frac{\partial \bar{R}_{\alpha}}{\partial K_{\alpha}}. \quad (5)$$

The average species-cost susceptibility, ν, measures the change in mean species abundances if we slightly decrease the minimum fitness cost (or equivalently increase the growth rate),

$$\nu = \frac{1}{S} \sum_{i} \frac{\partial \bar{N}_i}{\partial g_i} = -\frac{1}{S} \sum_{i} \frac{\partial \bar{N}_i}{\partial m_i}. \quad (6)$$

These susceptibilities characterize the sensitivity of an ecosystem to perturbations and are derivatives which can be directly measured in experiments. To estimate ν, one would alter the environment to put stress on the consumers, for instance in the context of microbes, one might force microbes to express a useless protein such as GFP as done in [?] to reduce m_i. To estimate χ, one would alter the environment to vary the carrying capacity and measure how the resource abundance fixed point changes, this could be done in many ways for one example of an experiment varying the environment to impact resource abundances see [?].

In terms of these quantities, one can derive a simple expression for the steady-state abundances of a newly added consumer and resource (see Appendix):

$$\bar{N}_0 = \frac{\max\left[0, g + \sigma_g z_0\right]}{\gamma \sigma_c^2 \chi},$$

$$\bar{R}_0 = \frac{\max\left[0, K^{eff} + \sigma_{K^{eff}}^2 \tilde{z}_0\right]}{1 - \sigma_c^2 \nu},$$

where as above z_0 and $\tilde{z}_0$ are independent, unit normal variables. These equations have a beautiful and straightforward interpretation. A new species added to the system will have an invasion growth rate $g_0^{inv} = g + \sigma_g z_0$, which is normally distributed. If the growth rate is negative, it will not be able to invade the system and go extinct. If its growth rate is positive when introduced in the ecosystem, then it survives with an abundance proportional to its invasion growth rate. We emphasize that this proportionality constant can differ significantly from what would be expected in a single-species ecosystem and depends on all the other

resources and species present in the ecosystem through the susceptibility χ and the variance of the consumption coeffcients σ_c^2. For this reason, the invasion growth rate of a species when it invades an ecosystem is positively correlated with its abundance. Similarly, the new resource is depleted if its effective carrying capacity is negative. Otherwise the steady-state abundance of the new resource is proportional to its effective carrying capacity. These equations are similar to the arguments of MacArthur and Levins on the necessary conditions for invasibility to large ecosystems [6]. They also generalize results for species abundances derived in [27] using the Lotka-Volterra equation and the results in [24, 38, 39] which ignored resource depletion and resource fluctuations.

B. Comparison with numerics

Unlike small ecosystems, we cannot analytically solve for the all the resource and species abundances. However, we can take a statistical approach that allows us to calculate statistical properties of species and resource abundances at steady-state. We also restrict our analysis to uninvadable steady-states, defined as a steady-state which cannot be invaded by any species. This, both simplifies the mathematics, and allows us to more directly relate our calculations to ecology.

Using (7) is it possible to derive self-consistency equations for the fraction of species in the regional species pool that survive, ϕ_N, the mean abundance of the species $\langle N \rangle = 1/S \sum_i N_i$, and variance and second moment of surviving species abundances, $\langle (\delta N)^2 \rangle$ and $q_N = \langle (\delta N)^2 \rangle + \langle N \rangle^2 = 1/S \sum_i N_i^2$ respectively. We can also calculate the analogous equations for resources: the fraction of resources with non-zero abundance, ϕ_R, the mean abundance of resources $\langle R \rangle = 1/M \sum_\alpha R_\alpha$, and variance and second moment of the resource abundances, $\langle (\delta R)^2 \rangle$ and $q_R = \langle (\delta R)^2 \rangle + \langle R \rangle^2 = 1/M \sum_\alpha R_\alpha^2$. The equations are derived in Appendix C and can be solved numerically.

The validity of our derivation is dependent on the MCRM having replica symmetry in the correspondence between the closely related cavity and replica methods (see e.g. [?]), and while we do not rigorously prove replica symmetry is not broken in this setting, we expect this assumption to be correct because there is a convex Lyapunov function (see e.g. [8]) and therefore just one fixed-point in the MCRM. Broken replica symmetry occurs when the landscape is complex, see [?]. The other assumption is a large system size $S, M \gg 1$. The theory is expected to be exact for infinite systems, but as we will now discuss, our simulations demonstrate that moderately large ecosystems are well predicted by the theory in the setting of Gaussian distributed consumption coefficients for which it was derived.

To check the accuracy of our CM, we compared our analytic predictions to numerical simulations (see Figure 2). We simulated (2) for two different choices of distributions for the c_{ia}. In the first set of simulation, the c_{ia} were binary random variables with $c_{ia} = 1$ with probability p and $c_{ia} = 0$ with probability $1 - p$. The probability p can be viewed as the level of generalism in the regional species pool. As $p \to 0$, all organisms in the community are specialist and consume a handful of resources. When $p \to 1$, the community consists of

generalists who can consume almost all resources. In the second set of simulations, we drew the consumption coeffcients from a Gaussian distribution with the same mean and variance as the corresponding Bernoulli distribution with probability p.

As shown in Fig. 2, our analytic results agree remarkably well with numerical simulations. The agreement between theory and numerics is nearly exact when c_{ia} are drawn from a Gaussian and shows qualitative agreement even when the consumption coeffcients c_{ia} are binary random variables. This is a result of the Gaussianity assumptions used to derive the cavity equations (see Appendix). The discrepancy between the binary case and Gaussian case stems from the fact that the for large S and M the c_{ia} are strictly positive for the binary case but generically contain some negative elements for Gaussian distributions. A negative c_{ia} implies that species i produces resource a at a fitness cost to itself. Thus, all simulations with Gaussian include a small fraction of public good producers that are accounted for in our theoretical calculations but are absent in the simulations with binary variables.

Despite these differences, for both choice of distributions the fraction of surviving species declines with increasing p. This is consistent with the basic idea of nichetheory that as p increases, there is increased competition resulting in greater competitive exclusion. In contrast, the mean abundances of surviving species and resources shows a non-monotonic behavior as a function of p in both numerical simulations and analytics (see appendix and Fig. 5 for additional simulation results).

IV.   GENERALIZING NICHE THEORY TO LARGE ECOSYSTEMS

The MacArthur consumer resource model has played a central role in the development of niche-based theories of community assembly [2, 7–10]. However, most of these analyses have focused on small ecosystems with just a few species and resources. Here, we discuss the ecological implications of our analysis for understanding community assembly in large ecosystems with many species and resources.

A.   Relating MCRM parameters to ecology

We begin by relating the parameters of the MCRM to more ecologically meaningful quantities such as the niche overlap, fitness, zero net- growth isoclines (ZNGI), and impact vectors. In ecology, the niche overlap, ρ, measures how much two species compete for the same resources. The larger the niche overlap, the more species compete. For small ecosystems, the niche overlap is bounded between 0 and 1, with a niche overlap of zero meaning the species do not compete for resources and a niche overlap of one indicating the species have identical consumption profiles. In the context of the two species MacArthur resource model, the niche-overlap between species can be thought of as the percentage of variance explained if one performs a regression of the first consumer's consumption vector against the consumption vector of the second species[1, 7, 8]. Using this observation, we can naturally extend the idea of niche overlap to entire ecosystems by defining an ecosystem-level niche overlap ρ in terms of the mean and variances of the consumption coeffcients $c_{i\alpha}$:

$$\rho = \frac{\mu_c^2}{\mu_c^2 + \sigma_c^2}. \quad (8)$$

One useful way of thinking about ρ is that it measures the niche-overlap between two species randomly drawn from the regional species pool. It is easy to see that when $\sigma_c^2 \ll \mu_c$, all species have nearly identical consumption preferences and $\rho \rightarrow 1$. In contrast when $\sigma_c^2 \gg \mu_c$, species will have very distinct consumer preferences and $\rho \rightarrow 0$.

Another fundamental quantity in contemporary niche theory is the ecological fitness of an organism, $f_i = \sum_\alpha c_{i\alpha} K_\alpha - m_i$ [1,8]This fitness is the initial growth rate of organism i in the *absence* of other species. In general, the actual growth rate of a species will differ significantly from the fitness if the resource abundances differ significantly from the resource carrying capacities K_α. For this reason, we will refer to this as the "naive" fitness.

We show in the appendix that it is also possible to relate our parameters directly to ZGNIs and generalized impact vectors.

B. Niche overlap and coexistence

One of the fundamental results of niche-based theories is that as the niche-overlap between species increases, coexistence become more and more diffcult [1]. The underlying reason for this is species that have similar consumer preference are more likely to compete with each other, resulting in competitive exclusion. Thus, increasing the niche-overlap in the community should decrease the fraction of species ϕ_N that can co-exist in a community. On the other hand, stabilizing mechanisms that decrease the tness differences between species should increase coexistence. We can parameterize the fitness differences in the community by the dimensionless quantity σ_m/m equal to the standard deviation over the mean of the maintenance costs m_i over all species in the regional species pool. This choice of parameterization is in line with contemporary niche theory where fitness differences are dened as the difference in growth rates when species have identical consumption preferences [1]. Figure 3 shows ϕ_N as a function of the niche overlap ρ and σ_m^2/m. This choice of niche-overlap corresponds to varying the probability p for having a non-zero $c_{i\alpha}$ from 0:1 to 0:9 (see Fig. 2). As predicted by niche theories, increasing ρ leads to increased competition and a smaller ϕ_N. In constrast decreasing σ_m/m_i at a fixed ρ, leads to a larger fraction of species surviving. Thus, in this regard large ecosystems behave quite similarly to predictions made by analyzing smaller models.

C. Resource depletion, effective fitness, and carrying capacity

One important feature of our analysis, which we are able to analyze statistically with our CM approach, is the large-scale depletion of resources. As shown in Fig. 2, species can signiffcantly change the resource profile and deplete a large fraction of resources initially present in the environment. This collective behavior can change which species survive and

thrive in an environment. One way to measure this effect and the reshaping of the resource profile is to measure the correlation between the naive fitness of an organism, $f_i = \sum_\alpha c_{i\alpha} \max\{K_\alpha, 0\} - m_i$, and its steady-state abundance in the ecosystem $\bar{N}_i$. The fitness f_i measures the growth rate of organism i if it is introduced into an environment in the absence of other species. For this reason, we expect f_i to be highly predictive of $\bar{N}_i$. when resource abundance profiles are not significantly perturbed by consumption. On the other hand, it is possible for correlation between f_i and $\bar{N}_i$ to decrease significantly.

Fig. 4 shows f_i versus $\bar{N}_i$ for numerical simulations where the $c_{i\alpha}$ drawn from a binomial distribution with $p = 0{:}1$ and $\sigma_m/m = 0.1$, as well as the case where parameters are Gaussian random variables with mean an variance matching the binomial setting. From the figure, it is clear there is a significant correlation between f_i and $\bar{N}_i$. Organisms with higher fitness disproportionately survive in the ecosystem. However, a significant number of organisms that have a high naive fitness f_i can still go extinct in the ecosystem (black points). The difference between plots (A,B) and (C,D) is that in the former K_a and m_i are kept positive by ensuring they are drawn from a gamma distribution and the consumer preferences $c_{i\alpha}$ are always positive since they are binary (1 with probability p or 0 otherwise). In (C,D), each of these parameters is drawn from a Gaussian distribution, but with the same mean and variance as in (A,B). This allows K_a, m_i, and $c_{i\alpha}$ to be negative. We will be agnostic about the physical meaning of negative $c_{i\alpha}$: it could for instance be related to a common good production by a species at an energetic cost to itself or simply that species i is preyed on by resource a. A more careful breakdown of relevant behaviors which could lead to the ecological engineering phenomena we will discuss such as poisons and common good production, and this is a direction of future work, but not considered in the present manuscript.

We find that allowing these negative consumer preference, does not dramatically alter the model predictions in terms of number/abundance or species/resources: we demonstrate a relatively good agreement between models which allow (Gaussian) and don't allow (binomial) negative consumer preferences in Fig. 2. However, we also find that the two cases lead to different ecological engineering phenomena when one observes individual species in the environment as in Fig. 4. Here, in panel D the red points correspond to resources a with a negative carrying capacity ($K_a < 0$) which end up in the environment due to "help" from a set of fit species (i_1, ...) because of the negative consumer preference $\left(c_{i_1\alpha}, ...\right)$.

In Fig. 4 C, red points indicate species with negative fitness that can stably exist in the community $\left(f_i < 0, \bar{N}_i > 0\right)$ due to the effect of negative consumer preferences, which can boost resources above their carrying capacity. Thus, we see species that cannot survive in the environment in the absence of other species can fixate (red points). Importantly, this emergent phenomenon is a collective property of the whole ecosystem and results from a complex interplay between organisms and environment and is an interesting direction for future study.

Additionally, Fig. 4 shows predictions from our CM for the correlation between f_i and $\bar{N}_i$.

Within our replicasymmetric ansatz, these correlations are described by normal distributions whose variances and covariances can be calculated using our self-consistent equations. The contour lines represent half a standard deviation spread of our normal distribution. Our theory qualitatively captures the shape of the correlation between f_i and $\bar{N}_i$. We give explicit

expression for these correlations as well as the mutual information between species abundances and naive fitness in Appendix E.

V DISCUSSION

Niche-based theories have played a fundamental role in shaping our understanding of community assembly and community ecology. In this work, we use ideas and methods from statistical physics to analyze a canonical model in community ecology, MacArthur's Consumer Resource Model (MCRM). Unlike previous works, our statistical physics inspired approach allows us to analyze large ecosystems with many species and resources. Our results suggest that organisms can significantly perturb their environments. The abundance of resources can be significantly altered and resource can even be completely depleted. We nd that such niche-construction is a generic feature of MCRM. This suggests that in complex ecosystems, organisms actively construct their environment. To quote Levins and Lewontin, "they are not the passive objects of external forces, but creators and modulators of these forces" [22]. These effects are even more dramatic when consumers can produce public goods at a fitness cost to themselves. In this case, species and resources that could not survive in isolation can fixate in the ecosystem.

To carry out our analysis, we developed a sophisticated theory based on the cavity method. One of the most striking things about our analysis is that many physical quantities that appear in the "cavity equations" have natural ecological interpretations in terms of invasion growth rates and effective carrying capacities. The underlying reason for this is that the cavity methods is based on asking how ecosystems are perturbed when a new species and a new resource are introduced into the ecosystem. Conceptually, this is very similar in spirit to many classical arguments in community ecology pioneered by Levins and MacArthur that ask whether a new species can invade [2, 6]. This naturally allows us to generalize many of the results from niche-based theories to large, diverse ecosystems. However, the price we pay for using our cavity approach is that we are limited to making statistical predictions.

An important question for future investigation is to ask how our results change if we make the model more realistic. In the MCRM, all species are assumed to have a linear, Type I functional response. It will be interesting to generalize our model to non-linear functional responses. We have also neglected the effects of environmental and demographic stochasticity. Stochasticity can induce phases transitions in ecosystems from a nichelike phase where competitive effects dominate community assembly to an ecologically neutral-like phase where stochasticity is the primary determinant of community structure [15, 23]. It will be interesting to see if the techniques developed here can be generalized to this more complicated setting. Finally, we have assumed that our population can be modeled as a well-

mixed community. However, spatial effects can qualitatively change the behavior of cellular populations [46, 47] and are likely to play an important role in community assembly.

VI. ACKNOWLEDGEMENTS

Appendix A:: Basic Setup

We brie y summarize MacArthur's classical consumer resource model (MCRM). Species $i = 1 \ldots S$ grows at a rate proportional to its utilization of resources, $R_\alpha, \alpha = 1...M$, in the environment. This is described by the equation:

$$\frac{1}{N_i}\frac{dN_i}{dt} = \sum_\alpha c_{i\alpha}R_\alpha - m_i + h_i, \quad (A1)$$

where w_α is the value of one unit of resource to a species (e.g. ATPs that can be extracted); $c_{i\alpha}$ is the rate at which species i consumes resource α and converts that into a "growth rate", m_i; m_i is the minimum amount of resources that must be consumed in order to have a positive growth rate. We have also added a small perturbation h_i to the system that will do a linear expansion in. The original MCRM corresponds to the choice $h_i = 0$.

We define the growth rate to be

$$g_i(R) = \sum_\alpha c_{i\alpha}W_\alpha R_\alpha - m_i. \quad (A2)$$

In consumer resource model, resources satisfy their own dynamical equations:

$$\frac{dR_\alpha}{dt} = R_\alpha\big(K_\alpha + b_\alpha - R_\alpha\big) - \sum_j c_{j\alpha}N_j R_\alpha, \quad (A3)$$

where the first term (with $b_\alpha = 0$) describes the resource dynamics in the absence of any species and the second term models the consumption of resource by species in the environment, and b_α is small perturbation. The original MCRM corresponds to the choose $b_\alpha = 0$. Furthermore, define the effective carrying capacity

$$K_\alpha^{eff} = K_\alpha - \sum_j c_{j\alpha}N_j. \quad (A4)$$

We will consider the case when the consumer preferences $c_{i\alpha}$ are random variables that can be characterized by their means and variances. In particular,

$$\langle c_{i\alpha} \rangle = \frac{\mu_c}{S}, \quad (A5)$$

and

$$\langle c_{i\alpha} c_{j\beta} \rangle = \frac{\sigma_c^2}{S} \delta_{ij} \delta_{\alpha\beta} + \frac{\mu_c^2}{S^2} \approx \frac{\sigma_c^2}{S}. \quad (A6)$$

To perform the cavity equations, it is useful to define several other quantities. Let us a define the fluctuating part of the consumer preferences $d_{i\alpha}$ as

$$c_{i\alpha} \equiv \frac{\mu_c}{S} + \sigma_c d_{i\alpha}. \quad (A7)$$

Then, we have that

$$\langle d_{i\alpha} \rangle = 0, \quad (A8)$$

and

$$\langle d_{i\alpha} d_{j\beta} \rangle = \frac{\delta_{ij} \delta_{\alpha\beta}}{S}. \quad (A9)$$

We will also assume that the carrying capacities are drawn from a Gaussian distribution with

$$\langle K_\alpha \rangle = K, \quad (A10)$$

and

$$\langle \delta K_\alpha \delta K_\beta \rangle = \langle (K_\alpha - K)(K_\beta - K) \rangle = \delta_{\alpha\beta} \sigma_K^2. \quad (A11)$$

Finally, we assume that the minimum survival coeffcients are also drawn from Gaussian distribution with

$$\langle m_i \rangle = m, \quad (A12)$$

and

$$\left\langle \delta m_i \delta m_j \right\rangle = \left\langle \left(m_i - m\right)\left(m_j - m\right) \right\rangle = \delta_{ij} \sigma_m^2. \quad \text{(A13)}$$

For future reference, it will also be helpful to define the ratio

$$\gamma = \frac{M}{S}, \quad \text{(A14)}$$

the average resource abundance,

$$\langle R \rangle = \frac{1}{M} \sum_\alpha R_\alpha, \quad \text{(A15)}$$

and the average species abundance

$$\langle N \rangle = \frac{1}{S} \sum_j N_j. \quad \text{(A16)}$$

With these definitions, notice that we can rewrite (A1) as

$$\frac{1}{N_i} \frac{dN_i}{dt} = \mu \gamma \langle R \rangle - m + \sigma \sum_\alpha d_{i\alpha} R_\alpha - \delta m_i, \quad \text{(A17)}$$

And rewrite (A3) as

$$\frac{1}{R_\alpha} \frac{dR_\alpha}{dt} = \left[K - \mu_c \langle N \rangle\right] - R_\alpha - \sigma_c \sum_j d_{j\alpha} N_j + \delta K_\alpha. \quad \text{(A18)}$$

We can de ne the mean growth rate of the population

$$g = \mu_c \gamma \langle R \rangle - m \quad \text{(A19)}$$

and the mean e ective capacity of resources in the ecosystem to be

$$K^{eff} = K - \mu_c \langle N \rangle \quad \text{(A20)}$$

as in the main text. In terms of these quantities, we can rewrite these equations as

$$\frac{1}{N_i}\frac{dN_i}{dt} = g + \sigma \sum_\alpha d_{i\alpha}R_\alpha - \delta m_i,$$

$$\frac{1}{R_\alpha}\frac{dR_\alpha}{dt} = K^{eff} - R_\alpha - \sigma_c \sum_j d_{j\alpha}N_j + \delta K_\alpha. \tag{A21}$$

The terms on the right hand sides of the equation above have a natural interpretation as the "fluctuating parts" of the growth rate and effective carrying capacity. In particular, we have rewritten the growth rate for species i as the sum of the mean growth rate g and a uctuating component δg_i defined as

$$\delta g_i = \sigma \sum_\alpha d_{i\alpha}R_\alpha - \delta m_i. \tag{A22}$$

We have split the effective carrying capacity of resource α is divided into a mean K^{eff} and fluctuating component δK_α^{eff} defined as

$$\delta K_\alpha^{eff} = -\sum_j d_{j\alpha}N_j + \delta K_\alpha. \tag{A23}$$

Appendix B:: Deriving the Species and Resource Distributions

To derive the cavity equations, we will relate a system with S species and M resources to a new system where we add an additional resource R_0 and and additional species N_0. Thus, the cavity equations relate a ecosystem with $S + 1$ and $M + 1$ resources to a ecosystem with S and M resources.

Then we can write equations for this new ecosystem (to leading order in S):

$$\frac{1}{N_i}\frac{dN_i}{dt} = g + \sigma_c \sum_\alpha d_{i\alpha}R_\alpha - \delta m_i + \sigma_c d_{i0}R_0, \tag{B1}$$

and

$$\frac{1}{R_\alpha}\frac{dR_\alpha}{dt} = K^{eff} - R_\alpha - \sigma_c \sum_j d_{j\alpha}N_j + \delta K_\alpha - \sigma_c d_{0\alpha}N_0. \tag{B2}$$

We can also write down the corresponding equations for the new resource and species:

$$\frac{1}{N_0}\frac{dN_0}{dt} = g + \sigma_c \sum_\alpha d_{0\alpha} R_\alpha - \delta m_0 + \sigma_c d_{00} R_0, \quad \text{(B3)}$$

and

$$\frac{1}{R_0}\frac{dR_0}{dt} = K^{eff} - R_0 - \sigma_c \sum_j d_{j0} N_j + \delta K_0 - \sigma_c d_{0\alpha} N_0. \quad \text{(B4)}$$

We now focus on steady-state. Let us denote the steady-state value of a quantity X by $\bar{X}$. Then, we can define some susceptibilities that are extremely useful for what follow:

$$\chi_{i\beta}^{(N)} = \frac{\partial \bar{N}_i}{\partial K_\beta} \quad \text{(B5)}$$

$$\chi_{\alpha\beta}^{(R)} = \frac{\partial \bar{R}_\alpha}{\partial K_\beta}$$

and

$$v_{ij}^{(N)} = \frac{\partial \bar{N}_i}{\partial m_j}$$

$$v_{\alpha j}^{(R)} = \frac{\partial \bar{R}_\alpha}{\partial m_j}. \quad \text{(B6)}$$

Now we are in a position to perform the cavity calculation. Let us denote the steady-state value of a quantity X in the absence of the new resource and species as $\bar{X}_{/0}$. Then, since the addition of a resource and species represents a small perturbation (order 1/S), we can write:

$$\bar{N}_i = \bar{N}_{i/0} - \sigma_c \sum_\beta \chi_{i\beta}^{(N)} d_{0\beta} \bar{N}_0 - \sigma_c \sum_j v_{ij}^{(N)} d_{j0} \bar{R}_0, \quad \text{(B7)}$$

and

$$\bar{R}_\alpha = \bar{R}_{\alpha/0} - \sigma_c \sum_\beta \chi_{\alpha\beta}^{(R)} d_{0\beta} \bar{N}_0 - \sigma_c \sum_j v_{\alpha j}^{(R)} d_{j0} \bar{R}_0. \quad \text{(B8)}$$

We can now plug in these expressions into the steady-state equations for N_0 and R_0. This gives:

$$0 = \bar{N}_0 \left[g + \sigma d_{00} R_0 + \delta m_0 + \sigma_c \sum_\alpha d_{0\alpha} R_{\alpha/0} - \sigma_c^2 \sum_{\alpha\beta} \chi_{\alpha\beta}^{(R)} d_{0\alpha} d_{0\beta} N_0 - \sigma_c^2 \sum_{\alpha, j} v_{\alpha j}^{(R)} d_{0\alpha} d_{j0} R_0 \right].$$

(B9)

If we now take leading order contributions to S in this expression, and take expectation value over expressions this reduces to

$$0 = \bar{N}_0 \left[g + \sigma d_{00} R_0 - \gamma \sigma^2 \frac{1}{M} \sum_\alpha \chi_{\alpha\alpha}^{(R)} N_0 + \delta m_0 + \sigma \sum_\alpha d_{0\alpha} R_{\alpha/0} \right]. \quad \text{(B10)}$$

Notice that, to leading order in S, we can model the term $\delta m_0 + \sigma \sum_\alpha d_{0\alpha} R_{\alpha/0}$, which is just the invasion growth rate minus the mean growth rate $g_0^{inv} - g$, as a Gaussian random eld with mean 0 and variance:

$$\sigma_g^2 = \sigma_c^2 \gamma \frac{1}{M} \sum_\alpha \bar{R}_{\alpha/0}^2 + \sigma_m^2 = \sigma_c^2 \gamma q_R + \sigma_m^2, \quad \text{(B11)}$$

and

$$q_R = \frac{1}{M} \sum_\alpha R_{\alpha/0}^2. \quad \text{(B12)}$$

If we let z_N be random field with mean zero and unit variance, and define the average suspectibility

$$\chi = \frac{1}{M} \sum_\alpha \chi_{\alpha\alpha}^{(R)} \quad \text{(B13)}$$

then we can write the equation for N_0 as

$$0 = \bar{N}_0 \left[\mu \gamma \langle R \rangle - m - \gamma \sigma^2 \chi \bar{N}_0 + \sigma_g z_N \right]. \quad \text{(B14)}$$

We can also derive a similar equation for R_0. This is given by

$$0 = R_0 \left[K^{eff} - R_0 - \sigma_c \sum_j d_{j0} N_{j/0} - \sigma_c^2 \sum_{j\beta} \chi_{i\beta}^{(N)} d_{j0} d_{0\beta} \bar{N}_0 \right. \quad \text{(B15)}$$
$$\left. - \sigma_c^2 \sum_{j,k} v_{jk}^{(N)} d_{j0} d_{k0} \bar{R}_0 + \delta K_0 - \sigma_c d_{00} N_0 \right].$$

Using the same logic as above, to leading order we have

$$0 = R_0 \left[K^{eff} - R_0 + \sigma_c^2 v \bar{R}_0 + \sigma_{K^{eff}} z_R \right], \quad \text{(B16)}$$

where

$$v = \frac{1}{S} \sum_j v_{jj}^{(N)}, \quad \text{(B17)}$$

and

$$\sigma_{K^{eff}}^2 = \sigma_K^2 + \sigma_c^2 q_N, \quad \text{(B18)}$$

with

$$q_N = \frac{1}{S} \sum_j N_{j/0}^2. \quad \text{(B19)}$$

We can solve these equations and get

$$N_0 = \frac{\max\left[0, g + \sigma_g z_N\right]}{\gamma \sigma^2 \chi} \quad \text{(B20)}$$

$$R_0 = \frac{\max\left[0, K^{eff} + \sigma_{K^{eff}} z_R\right]}{1 - \sigma^2 v}. \quad \text{(B21)}$$

Thus, the distributions for N and R are given by truncated Gaussians.

Appendix C:: Self Consistency Equations

Let us now write some self-consistency equations in the replica symmetric phase. Let us define the number of non-zero species and resources as S^* and M^* respectively. Furthermore, de ne

$$\phi_S = \frac{S^*}{S}, \quad \text{(C1)}$$

$$\phi_M = \frac{M^*}{M}. \quad \text{(C2)}$$

Our goal is, given some parameters $\{K, \sigma_K, m, \sigma_m, \mu, \sigma, S, M\}$, to find the values for $\{\phi_S, \phi_N, \langle N \rangle, \langle R \rangle, q_R, q_N, \chi, v\}$. Since there are eight unknowns we will need eight equations. It will also be useful to define:

$$\Delta_g = \frac{\mu \gamma \langle R \rangle - m}{\sigma_g}, \quad \text{(C3)}$$

$$\Delta_{K^{eff}} = \frac{K - \mu \langle N \rangle}{\sigma_{K^{eff}}}, \quad \text{(C4)}$$

and the function

$$w_j(\Delta) = \int_{-\Delta}^{\infty} \frac{dz}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} (z + \Delta)^j. \quad \text{(C5)}$$

First, let us write self-consistency equations for the susceptibilities. Taking derivatives with respect to m and K of (B21) and noting that the fraction of non-zero species and non-zero resources is $_N$ and $_R$ respectively gives

$$v = -\frac{\phi_N}{\gamma \sigma_c^2 \chi}, \quad \text{(C6)}$$

$$\chi = \frac{\phi_R}{1 - \sigma_c^2 v}. \quad (C7)$$

Notice now that if we define $y = \max\left[0, \frac{a}{b} + \frac{c}{b}z\right]$ with z a gaussian random variable we have that:

$$\langle y^j \rangle = \left(\frac{b}{c}\right)^j \int_{-\frac{b}{a}}^{\infty} dz \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} \left(z + \frac{b}{a}\right)^j. \quad (C8)$$

We can now use the fact that (B21) implies that the species distribution and resource distribution is given by a truncated Gaussian to write self consistency equations for the fraction of nonzero resources and species as well as the the moments of their abundances:

$$\phi_N = w_0(\Delta_g), \quad (C9)$$

$$\phi_R = w_0\left(\Delta_{K^{eff}}\right), \quad (C10)$$

$$\langle N \rangle = \left(\frac{\sigma_g}{\gamma \sigma_c^2 \chi}\right) w_1(\Delta_g), \quad (C11)$$

$$\langle R \rangle = \left(\frac{\sigma_{K^{eff}}}{1 - \sigma_c^2 v}\right) w_1\left(\Delta_{K^{eff}}\right), \quad (C12)$$

$$q_N = \langle N^2 \rangle = \left(\frac{\sigma_g}{\gamma \sigma_c^2 \chi}\right)^2 w_2(\Delta_g), \quad (C13)$$

$$q_R = \langle R^2 \rangle = \left(\frac{\sigma_{K^{eff}}}{1 - \sigma_c^2 v}\right)^2 w_2\left(\Delta_{K^{eff}}\right). \quad (C14)$$

Together (C4), (C7), and (C14) define the 10 selfconsistency equations we need, along with the de nitions (B11) and (B18). We solve these mean field equations numerically using the sum of squared differences between the left and right sides of equations (C9-C14) as an energy function which we minimize using the basinhopping optimization algorithm from the scipy.optimize. The algorithm uses random perturbations, local minimization, and an accept or reject criterion to attempt to minimize function which may be non-convex. The parameters we used were a temperature of 1, a step size of 0:5, and 5 iterations or initializations. Note that the equations can also be solved iteratively, but we found these solutions were stable for a smaller set of parameter values using this approach.

Appendix D:: Zero net-growth nullclines and generalized impact vectors

We can also easily relate our mean-field quantities to ZGNIs. Recall, that ZGNI's delineate range of resource conditions in which a species maintains a positive growth rate [9, 10]. Each species i defines a ZGNI in the resource space R_i^{ZGNI} defined by the equation $g_i\left(R_i^{ZGNI}\right) = 0$. Geometrically, we can view R_i^{ZGNI} as a hyperplane in resource space whose dot product with the consumption coeffcients c_{ia} of species i equals m_i (see Eq. 1). If $m_i \gg 1$, the ZGNI is well-approximated by the plane perpendicular to c_{ia}. We can calculate some statistical properties of these ZGNI. Notice that the mean value of each component is just

$$\left\langle R_{i\alpha}^{ZGNI}\right\rangle = \frac{m}{\gamma\mu_c}, \quad \text{(D1)}$$

and the expected value of the square is just

$$\left\langle \left(R_{i\alpha}^{ZGNI}\right)^2\right\rangle = \frac{\sigma_m^2 + m^2}{\gamma\sigma_c^2}. \quad \text{(D2)}$$

In Modern Niche Theory, another important quantity is the impact vector of a species i. The impact vector describes how resources are depleted by the addition of another individual. Here, we introduce the idea of generalized impact vectors that measure how the steady state concentration of a resource α changes due to the introduction of a species j Alternatively, we can consider a system without resource β and then ask how it's addi-tion changes the species abundance of a species i. These de ne the generalized impact vectors (GIVs).

These are of course the leading order contribution (in S) to the cavity equations under the replica symmetric assumption, namely (B8) and (B7). Thus, the components of the two "generalized' impact vectors are given by:

$$\bar{R}_\alpha - \bar{R}_{\alpha/j} = -\sigma_c \sum_\beta \chi_{\alpha\beta}^{(R)} d_{j\beta}\bar{N}_j, \quad \text{(D3)}$$

$$\overline{N}_i - \overline{N}_{i/\beta} = -\sigma_c \sum_j v_{ij}^{(N)} d_{j\beta} \overline{R}_\beta, \quad \text{(D4)}$$

Appendix E:: Comparison of individual fitness to true growth rate and steady state abundance

We want to quantify how much the naive fitness (the growth rate of an organism without other the growth rate of an organism without other growth rate g_i^{inv}, which in turn is closely related to the steady state abundance of each of the species in the community. From the definition of the invasion growth rate:

$$g_i^{inv} = \sum_\alpha c_{i\alpha} R_{\alpha/i} - m_i, \quad \text{(E1)}$$

and

$$f_i = \sum_\alpha c_{i\alpha} K_\alpha - m_i, \quad \text{(E2)}$$

we can compute the level of correlation between the two using the CM. To begin, we compute the means of each of these distributions:

$$\langle f_i \rangle = \sum_\alpha \langle c_{i\alpha} \rangle \langle K_\alpha \rangle - \langle m_i \rangle = \gamma \mu_c K - m, \quad \text{(E3)}$$

and

$$\langle g_i^{inv} \rangle = \sum_\alpha \langle c_{i\alpha} \rangle \langle R_{\alpha/i} \rangle - \langle m_i \rangle = \gamma \mu_c \langle R \rangle - m. \quad \text{(E4)}$$

The final equality in the preceding equation holds in the asymptotic limit of large S and M. Note that the con-sumer preferences $c_{i\alpha}$ of an individual species are independent of the resources steady state levels when that species i is not included in the community as in the previous equation. We can also compute the correlation between fluctuations from the mean naive fitness and the mean invasion growth rate: i.e. if a species has a higher or lower individual fitness, if a species has a higher or lower individual in the community? To understand this correlation, we define $\delta f_i = f_i - \langle f_i \rangle$ and $\delta g_i^{inv} = g_i^{inv} - \langle g_i^{inv} \rangle$ and compute:

$$\langle \delta f_i \delta g_i^{inv} \rangle = \frac{\left\langle \left(\sum_\alpha \left[\sigma_c d_{i\alpha} K_\alpha + \frac{\mu_c}{S} \left(K_\alpha - \langle K_\alpha \rangle \right) \right] + \delta m_i \right)\right.}{\left. \cdot \left(\sum_\alpha \left[\sigma_c d_{i\alpha} R_{\alpha/i} + \frac{\mu_c}{S} \left(R_{\alpha/i} - \langle R_{\alpha/i} \rangle \right) \right] + \delta m_i \right) \right\rangle}$$

In the large S limit, the important terms remaining in the average above are:

$$\langle \delta f_i \delta g_i^{inv} \rangle = \gamma \frac{\sigma_c^2}{M} \sum_\alpha R_{\alpha/i} K_\alpha + \sigma_m^2, \quad \text{(E5)}$$

thus in the asymptotic limit we can write the correlation between the two forms of fitness as:

$$\langle \delta f_i \delta g_i^{inv} \rangle = \gamma \sigma_c^2 \langle R_{\alpha/i} K_\alpha \rangle + \sigma_m^2. \quad \text{(E6)}$$

To compute the correlation between carrying capacity and resource level we modify (B21) from our capacity calculation, which yields such a relationship:

$$R_\alpha(K_\alpha) = \frac{\max\left[0, K_\alpha - \mu_c \langle N \rangle + \sigma_c \sqrt{q_N} z_N \right]}{1 - \sigma_c^2 v}. \quad \text{(E7)}$$

Using this relationship and letting k be drawn from the same distribution as K_α, where

$$R(k) = \frac{\max\left[0, k - \mu_c \langle N \rangle + \sigma_c \sqrt{q_N} z_N \right]}{1 - \sigma_c^2 v}, \quad \text{(E8)}$$

we compute

$$\langle k R(k) \rangle_{k, z_N}. \quad \text{(E9)}$$

The full form of this integral is thus:

$$\left\langle \int dk\, k\, R(k) e^{-\frac{(k - K)^2}{2\sigma_K^2}} \right\rangle_{z_N}. \quad \text{(E10)}$$

By rewriting this as

$$
\left(\int dk (k-K) R(k) e^{-\frac{(k-K)^2}{2\sigma_K^2}} + \int dk K R(k) e^{-\frac{(k-K)^2}{2\sigma_K^2}} \right)_{z_N}, \quad \text{(E11)}
$$

it may be simplified via integration by parts on the first term, yielding:

$$
\langle k R(k) \rangle_{k, z_N} = \sigma_K^2 \langle R'(k) \rangle + K \langle R(k) \rangle
$$

$$
= \frac{1}{1-\sigma_v^2} \Big(\sigma_K^2 w_0 \big(\Delta_{K^{eff}} \big) + K w_1 \big(\Delta_{K^{eff}} \big) \Big), \quad \text{where } \Delta_{K^{eff}} = \frac{K - \langle N \rangle}{\sigma_{K^{eff}}}, \text{ and}
$$

$$
w_j(\Delta) = \int_{-\Delta}^{\infty} \frac{dz}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} (z+\Delta)^j. \quad \text{(E12)}
$$

Note, we can also compute Pearson's correlation coeffcient for these two tness metrics:

$$
c_\rho = \frac{\langle \delta f \delta g^{inv} \rangle}{\sqrt{\langle (\delta f)^2 \rangle \langle (\delta g^{inv})^2 \rangle}} \quad \text{(E13)}
$$

$$
= \frac{\gamma \sigma_c^2 \langle k R(k) \rangle_{k, z_N} + \sigma_m^2}{\sqrt{\big(\sigma_m^2 + \gamma \sigma_c^2 (\sigma_K^2 + K^2) \big) \big(\gamma \sigma_c^2 \langle R^2 \rangle + \sigma_m^2 \big)}}
$$

$$
= \frac{\frac{\gamma \sigma_c^2}{1-\sigma_c^2 v} \Big(\sigma_K^2 w_0 \big(\Delta_{K^{eff}} \big) + K w_1 \big(\Delta_{K^{eff}} \big) \Big) + \sigma_m^2}{\sqrt{\big(\sigma_m^2 + \gamma \sigma_c^2 (\sigma_K^2 + K^2) \big) \big(\gamma \sigma_c^2 \langle R^2 \rangle + \sigma_m^2 \big)}}. \quad \text{(E14)}
$$

Using (C10) and (C12) we can write the preceding expression as:

$$
\frac{\frac{\gamma \sigma_c^2}{1-\sigma_c^2 v} \left(\sigma_K^2 \phi_R + K \left(\frac{1-\sigma_c^2 v}{\sigma_{K^{eff}}} \right) \langle R \rangle \right) + \sigma_m^2}{\sqrt{\big(\sigma_m^2 + \gamma \sigma_c^2 (\sigma_K^2 + K^2) \big) \big(\gamma \sigma_c^2 \langle R^2 \rangle + \sigma_m^2 \big)}}. \quad \text{(E15)}
$$

1. Abundance vs naive tness

Note that given the relationship that N is a scaled version of g^{inv} where all negative values are truncated to zero (B21), it follows that we can compute the correlation between $\widetilde{N}_i$ and f_i

where $\widetilde{N}_i = \dfrac{g_i^{inv}}{\gamma \sigma_c^2 \chi}$. Where we let $N_i = \widetilde{N}_i$ if $\widetilde{N}_i > 0$ and $N_i = 0$ otherwise. To better understand

that the correlation between the abundance and fitness of a species, we compute the correlation between $\widetilde{N}$ and f:

$$\langle \delta \widetilde{N} \delta f \rangle = \frac{1}{\gamma \sigma_c^2 \chi} \left(\frac{\gamma \sigma_c^2}{1 - \sigma_c^2 v} \left(\sigma_K^2 \phi_R + K \left(\frac{1 - \sigma_c^2 v}{\sigma_{K^{eff}}} \right) \langle R \rangle \right) + \sigma_m^2 \right), \quad (E16)$$

$$\langle (\delta \widetilde{N})^2 \rangle = \frac{\left(\gamma \sigma_c^2 \langle R^2 \rangle + \sigma_m^2 \right)}{\left(\gamma \sigma_c^2 \chi \right)^2}, \quad (E17)$$

and

$$\langle (\delta f)^2 \rangle = \sigma_m^2 + \gamma \sigma_c^2 \left(\sigma_K^2 + K^2 \right). \quad (E18)$$

Using these covariances, along with the means:

$$\langle f \rangle = \gamma \mu_c K - m, \quad (E19)$$

$$\langle \widetilde{N} \rangle = \frac{\gamma \mu_c \langle R \rangle - m}{\gamma \sigma_c^2 \chi}, \quad (E20)$$

we are able to generate theoretical predictions for the distribution of f, N. See Fig 4 where the theoretical plot of f, $\widetilde{N}$ is compared with values of fittness f and abundance N for all species in a network over many realizations.

2. Resource capacity versus resource abundance

In the same Fig 4, we additionally plot a theoretical prediction overlayed with numerics of how the resource capacities K_a are related to the resource abundances R_a.

If we define $\tilde{R}_\alpha = \dfrac{K_\alpha^{eff}}{1 - \sigma_c^2 v}$, , which is equal to the prediction of R_α when the resources

abundance is positive, and use the fact

$$\delta K_\alpha^{eff} = -\sum_j d_{j\alpha} N_j + \delta K_\alpha, \quad \text{(E21)}$$

then we may combine these two relations to yield:

$$\delta \tilde{R}_\alpha = \frac{1}{1 - \sigma_c^2 v} \left(-\sum_j d_{j\alpha} N_{j/\alpha} + \delta K_\alpha \right). \quad \text{(E22)}$$

We can thus compute the correlations:

$$\frac{1}{M} \sum_\alpha \delta K_\alpha \delta K_\alpha \to \sigma_K^2, \quad \text{(E23)}$$

$$\frac{1}{M} \sum_\alpha \delta K_\alpha \delta \tilde{R}_\alpha \to \frac{\sigma_K^2}{1 - \sigma_c^2 v}, \quad \text{(E24)}$$

$$\frac{1}{M} \sum_\alpha \delta \tilde{R}_\alpha \delta \tilde{R}_\alpha \to \frac{\sigma_c^2 q_N + \sigma_K^2}{\left(1 - \sigma_c^2 v\right)^2}. \quad \text{(E25)}$$

Also the means are easy to compute:

$$\langle K_\alpha \rangle \to K, \quad \text{(E26)}$$

$$\langle \tilde{R}_\alpha \rangle \to \frac{K - \mu \langle N \rangle}{1 - \sigma_c^2 v}. \quad \text{(E27)}$$

This allows us to make theoretical predictions for how the resource abundances are correlated to resource capacities.

3. Mutual information between individual fitness and true growth rate

The mutual information between two Gaussian variables x and z is simply (note the means of these random variables do not contribute so we will assume them are zero mean random variables):

$$I(x,z) = \frac{1}{2}\ln\left(\frac{\sigma_x^2\sigma_z^2}{\sigma_x^2\sigma_z^2 - \sigma_{xz}^2}\right). \quad \text{(E28)}$$

Thus,

$$I\left(f, g^{inv}\right) = -\frac{1}{2}\ln\left(1 - c_\rho^2\right). \quad \text{(E29)}$$

This gives us a theoretical prediction using the predicted form for the correlation coeficient (E15).

Appendix F:: Additional simulations and notes

We discussed how the theoretical curves were generated in Appendix C. The numerical simulations were performed by solving the corresponding ODEs (4) and integrating numerically until time 50, 000 with 1000 steps. Although it is not always needed, we improved the accuracy by additionally including a small amount of migration noise which we lowered linearly to a negligible roundoff error over the course of the integration to help ensure that a species that was favored to survive would not go extinct.

We also ran simulations in other regimes, such as the one shown in Fig 5 where we consider fixing $\mu_c = 1$ while varying σ_c to study the setting when we are less interested in comparing specialists to generalists and more interested in the effect of niche overlap and how a high overlap in the generating distribution can reduce the number of surviving species.

References

[1]. Chesson P, Annual review of Ecology and Systematics , 343 (2000).

[2]. Tilman D, Resource Competition and Community Structure.(MPB-17), 17 (Princeton University Press, 1982).

[3]. Vellend M, The Quarterly review of biology 85, 183 (2010). [PubMed: 20565040]

[4]. Hubbell SP, The Unified Neutral Theory of Biodiversity and Biogeography (MPB-32) (Princeton University Press, 2001).

[5]. Hardin G, Science 131, 1292 (1960). [PubMed: 14399717]

[6]. Macarthur R and Levins R, The American Naturalist 101, 377 (1967).

[7]. MacArthur R, Theoretical population biology 1, 1 (1970). [PubMed: 5527624]

[8]. Chesson P, Theoretical Population Biology 37, 26 (1990).

[9]. Chase JM and Leibold MA, Ecological Niches: Linking Classical and Contemporary Approaches (University of Chicago Press, 2003).

[10]. Letten AD, Ke P-J, and Fukami T, Ecological Monographs (2017).

[11]. Volkov I, Banavar JR, Hubbell SP, and Maritan A, Nature 424, 1035 (2003). [PubMed: 12944964]

[12]. Rosindell J, Hubbell SP, and Etienne RS, Trends in Ecology & Evolution 26, 340 (2011). [PubMed: 21561679]

[13]. Rosindell J, Hubbell SP, He F, Harmon LJ, and Etienne RS, Trends in Ecology & Evolution 27, 203 (2012). [PubMed: 22341498]

[14]. Chisholm RA and Pacala SW, Theoretical Ecology 4, 195 (2011).

[15]. Fisher CK and Mehta P, Proceedings of the National Academy of Sciences 111, 13111 (2014).

[16]. Jeraldo P, Sipos M, Chia N, Brulc JM, Dhillon AS, Konkel ME, Larson CL, Nelson KE, Qu A, Schook LB, et al., Proceedings of the National Academy of Sciences 109, 9692 (2012).

[17]. Tilman D, Proceedings of the National Academy of Sciences of the United States of America 101, 10854 (2004), PMID: 15243158. [PubMed: 15243158]

[18]. Gravel D, Canham CD, Beaudet M, and Messier C, Ecology Letters 9, 399409 (2006).

[19]. Anderson PW et al., Science 177, 393 (1972). [PubMed: 17796623]

[20]. Levin SA, Ecology 73, 1943 (1992).

[21]. Levins R, American scientist 54, 421 (1966).

[22]. Levins R and Lewontin RC, The dialectical biologist (Harvard University Press, 1985).

[23]. Dickens B, Fisher CK, and Mehta P, Physical Review E 94, 022423 (2016). [PubMed: 27627348]

[24]. Tikhonov M and Monasson R, Physical Review Letters 118, 048103 (2017). [PubMed: 28186794]

[25]. Kessler DA and Shnerb NM, Physical Review E 91, 042705 (2015).

[26]. Tikhonov M, Elife 5, e15747 (2016). [PubMed: 27310530]

[27]. Bunin G, arXiv preprint arXiv:1607.04734 (2016).

[28]. Kadanoff LP, Statistical physics: statics, dynamics and renormalization (World Scienti c Publishing Company, 2000).

[29]. Violle C, Enquist BJ, McGill BJ, Jiang L, Albert CH, Hulshof C, Jung V, and Messier J, Trends in *ecology & evolution* 27, 244 (2012). [PubMed: 22244797]

[30]. Kadanoff LP, Statistical physics: statics, dynamics and renormalization (World Scienti c Publishing Co Inc, 2000).

[31]. Landau LD, Lifshitz E, and Pitaevskii L, "Statistical physics, part i," (1980).

[32]. Opper M and Saad D, Advanced mean field methods: Theory and practice (MIT press, 2001).

[33]. Diederich S and Opper M, Physical Review A 39, 4333 (1989).

[34]. Eissfeller H and Opper M, Physical review letters 68 2094 (1992). [PubMed: 10045302]

[35]. Rieger H, Journal of Physics A: Mathematical and General 22, 3447 (1989).

[36]. Tokita K, Physical review letters 93, 178102 (2004). [PubMed: 15525129]

[37]. Barbier M, Arnoldi J-F, Bunin G, and Loreau M, bioRxiv , 145862 (2017).

[38]. Yoshino Y, Galla T, and Tokita K, Journal of Statistical Mechanics: Theory and Experiment 2007, P09003 (2007).

[39]. De Martino A and Marsili M, Journal of Physics A: Mathematical and General 39, R465 (2006).

[40]. Erwin DH, Trends in Ecology & Evolution 23, 304 (2008). [PubMed: 18457902]

[41]. May RM, Nature 238, 413 (1972). [PubMed: 4559589]

[42]. Shea K and Chesson P, Trends in Ecology & Evolution 17, 170 (2002).

[43]. Tilman D, Resource competition and community structure, Vol. 17 (Princeton University Press, 1982).

[44]. Shamir M and Sompolinsky H, Physical Review E 61, 1839 (2000).

[45]. Ramezanali M, Mitra PP, and Sengupta AM, arXiv preprint arXiv:1509.08995 (2015).

[46]. Korolev KS, Avlund M, Hallatschek O, and Nelson DR, Reviews of modern physics 82, 1691 (2010). [PubMed: 21072144]

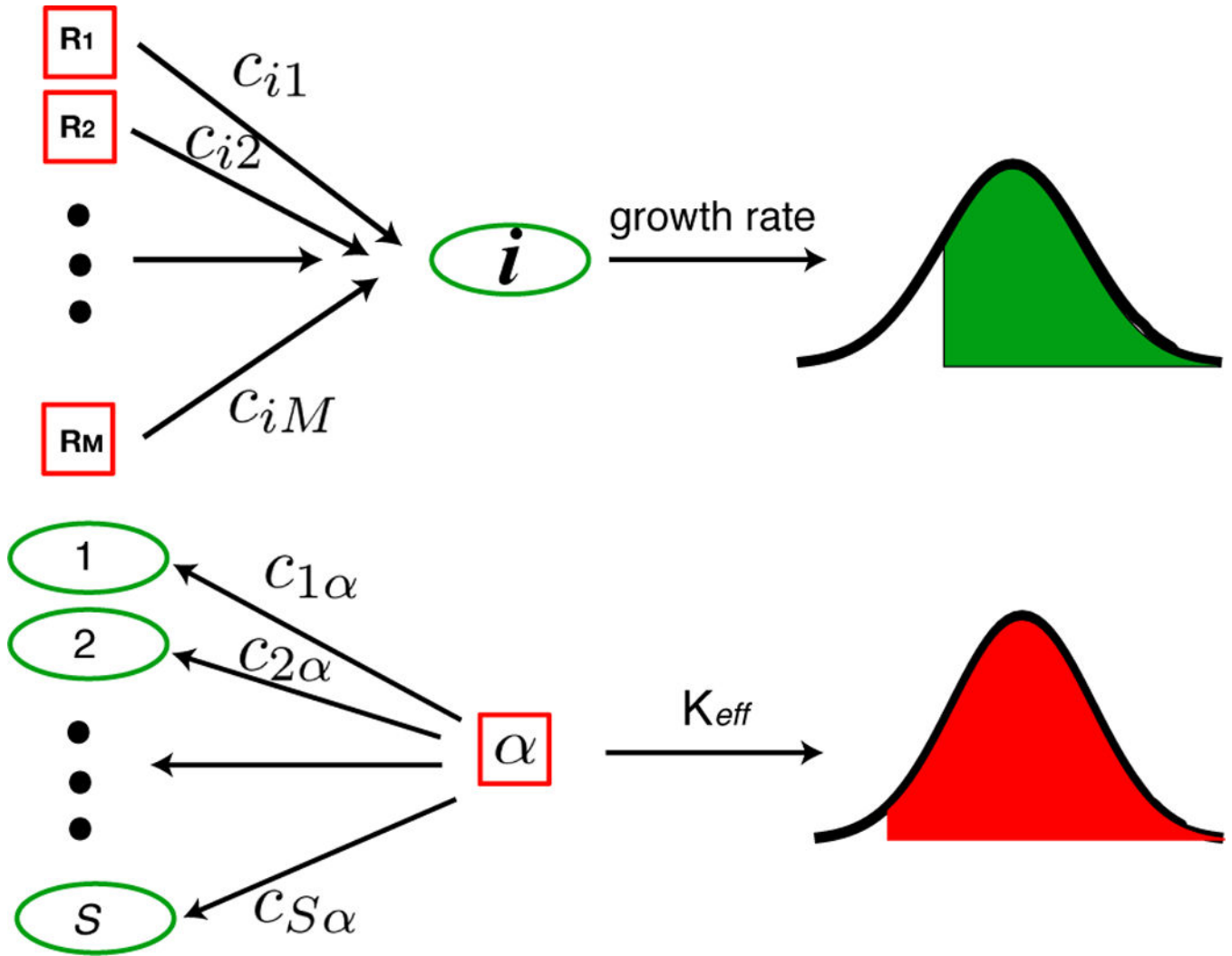[47]. Loreau M, Proceedings of the National Academy of Sciences 95, 5632 (1998).

FIG. 1. Analyzing MacArthur Consumer's Resource Model for large, diverse ecosystems.
(Top) The growth rate for a species i is a sum of terms resulting from consuming $M \gg 1$ resources. For this reason, from the central limit theorem, it can be well modeled by a (truncated) normally distributed variable. (Bottom) Each resource α is consumed by $S \gg 1$ consumers. From the central limit theorem, the effective carrying capacity of the resource (i.e. the resource abundance at steady-state) can also be modeled using a (truncated) normal distribution. The truncation is due to the fact that neither species nor resource abundances can become negative. To derive our analytic cavity equations, we require selfconsistency for the means and variance of these distributions.
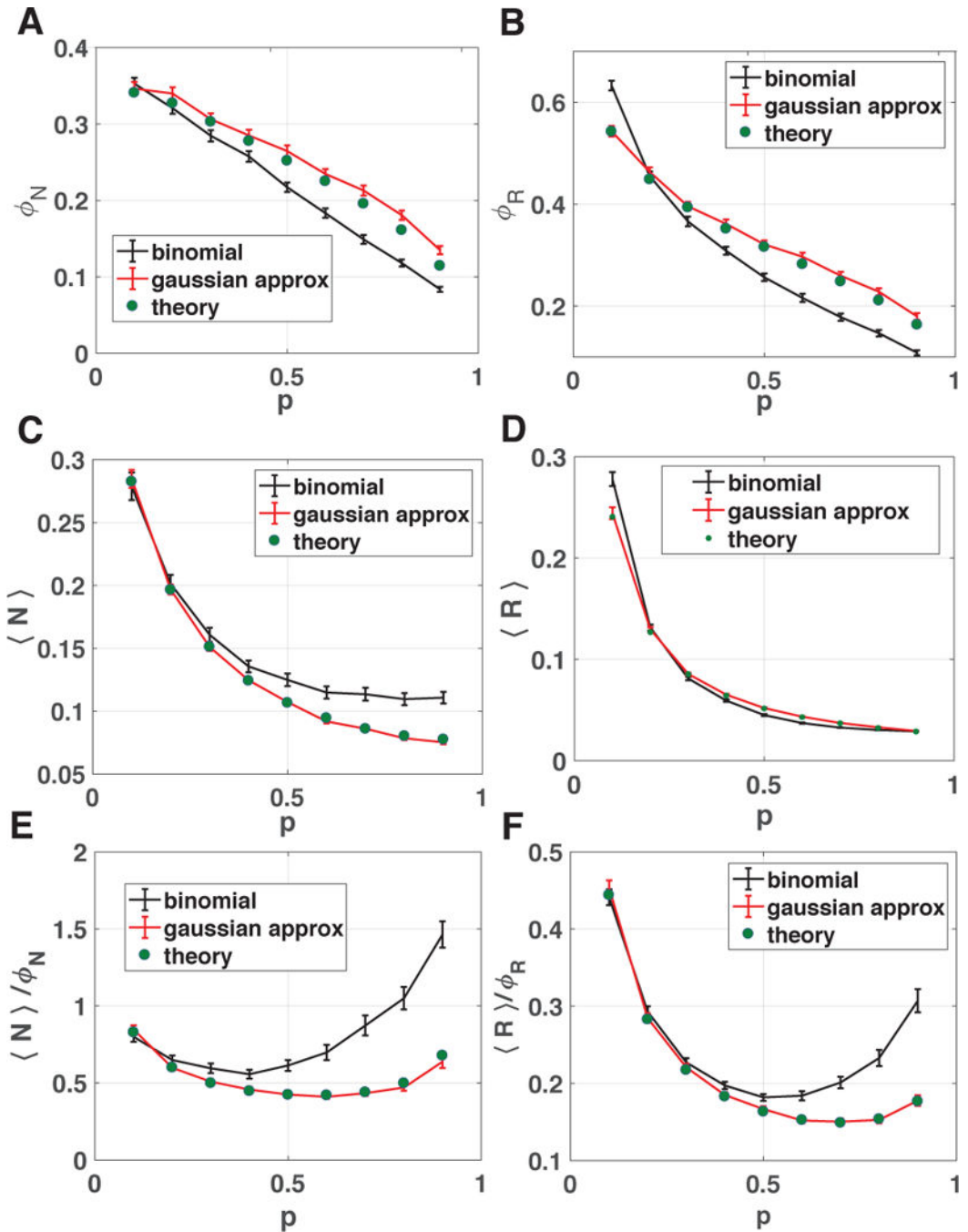
FIG. 2. Comparison of numerical simulations with theory.

Ecosystems were simulated with the consumption coefficients $c_{ia} = 0, 1$ drawn from a Bernoulli distribution with probability p (black lines) or a Gaussian distribution with same mean and variance as the binomial distribution (red line). Here $S = M = 30$, $K = 1$, $\sigma_k = 1$, $m = 1$, $\sigma_m = .1$, and 250 trials were used in these simulations: the error bars denote ± 2 standard deviations. The K_a and m_i were drawn iid from a gamma distribution in the binomial plot to ensure nonnegativity of the parameters, and a Gaussian distribution in the Gaussian approximation plot. The numerical results were compared to theoretical predictions from the

self-consistent cavity equations for ecologically stable steady-states (green circles). (A) Fraction of species that survive ϕ_N, (B) fraction of resources that are not depleted ϕ_R, (C) average abundance of all species $\langle N \rangle$, (D) average abundance of all resources $\langle R \rangle$ as a function of the probability p. (E) Mean-abundance of surviving species $\langle N \rangle / \phi_N$ and (F) mean-abundance of surviving resources $\langle R \rangle / _R$.
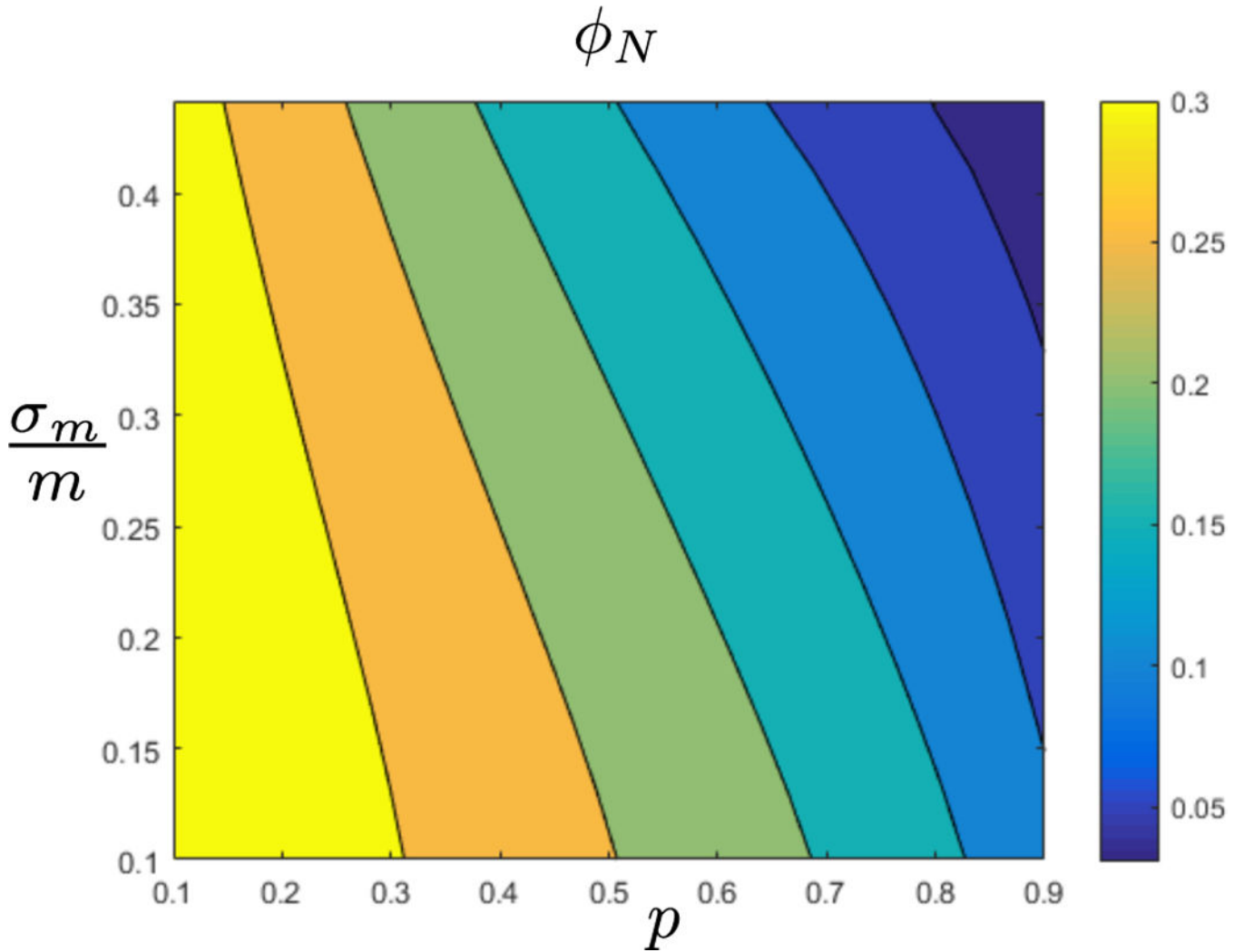
FIG. 3. Co-existence and diversity.
Here we use the same parameters as in the previous plot, apart from allowing σ_m to vary, and show the cavity prediction of the fraction of surviving species ϕ_N as a function of p and the standard deviation over mean σ_m/m of the maintenance costs m_i of species. In this regime increasing p leads to more similar species by increasing the niche overlap ρ as defined in (8) from $\rho \approx 77$ to $\rho \approx 99$ in the range shown above. As predicted by niche theories, increasing ρ leads to increased competition and a smaller ϕ_N while decreasing σ_m/m_i leads to larger fraction of species surviving at a fixed p.
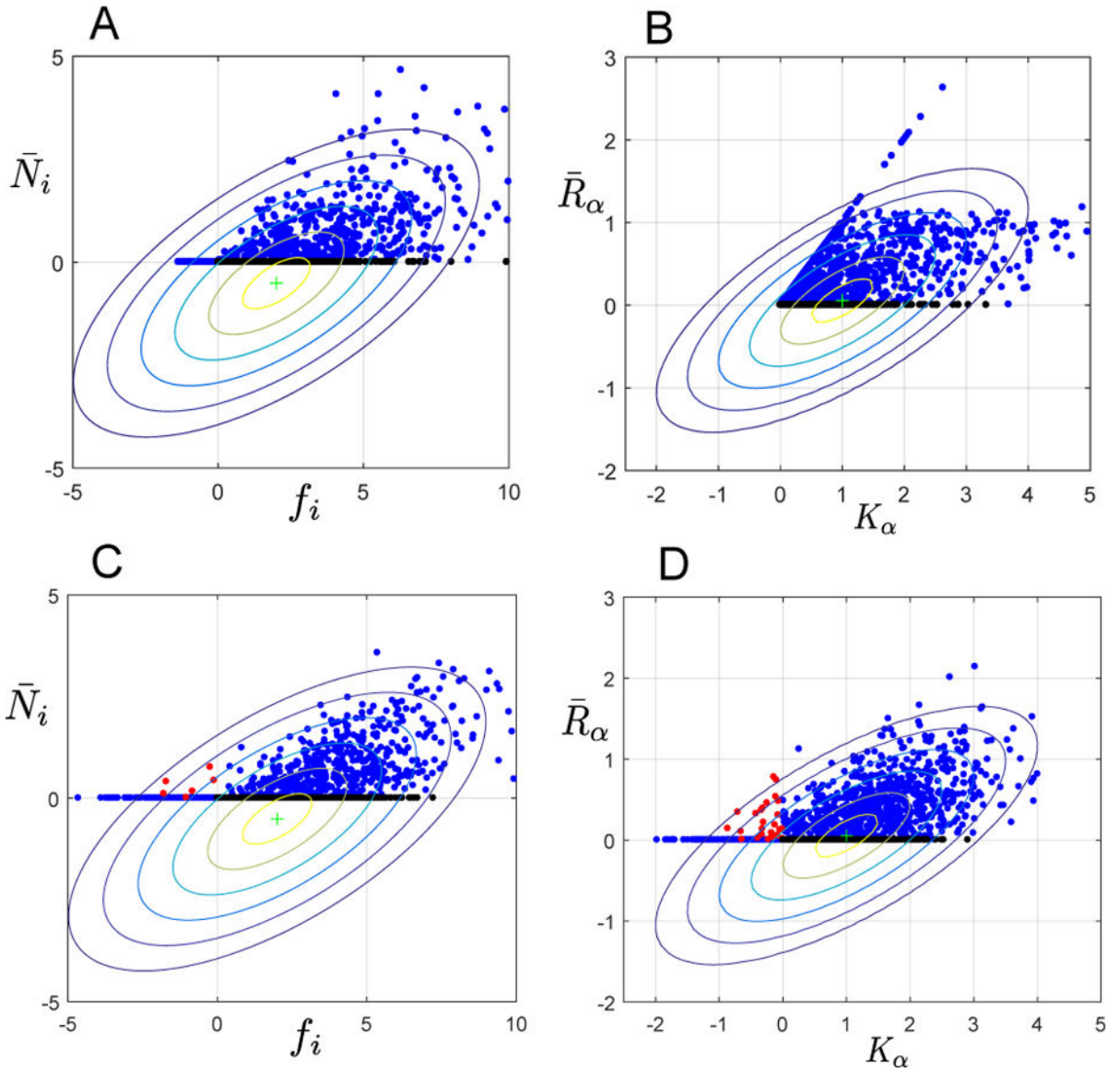
FIG. 4. Predicting the effective fttness and effective carrying capacity

(A) Steady-state abundance $\bar{N}_i$ versus the fitness $f_i = \sum_\alpha c_{i\alpha} \max\{K_\alpha, 0\} - m_i$ for each species i. Fitness is defined as the initial growth rate of species i in the environment in the absence of all other species. Points colored black are species with positive fitness that go extinct in the community $(f_i > 0, \bar{N}_i = 0)$. (B) Comparison of the steady state resource levels $\bar{R}_\alpha$ with their capacity K_α. The filled circles are generated from simulations with M = S = 30 resources and species, the data is generated from 50 separate trials. Parameters for simulations as in Figure 2 with $p =$ and $\sigma m/m = 0.1$. Black points indicate resources which have a positive capacity but go extinct in the community. The difference between plots for (A,B) and (C,D) is that in the former K_α and m_i are always positive because they are drawn from a gamma distribution and Bernoulli distribution respectively. In (C,D), each of these parameters is drawn from a Gaussian distribution with the same mean and variance as in (A,B). This means that a small fraction of the K_α, m_i, and $c_{i\alpha}$ are negative. Negative $c_{i\alpha}$

corresponds to production of resource a by species i at a fitness cost to itself (i.e. public good production). In (C,D), red points indicate species with negative fitness that can stably exist in the community $\left(f_i < 0, \overline{N}_i > 0 \right)$ or resources with negative K_a that are produced by the ecosystem. Contours show theoretical predictions of our CM for correlation between $\overline{N}_i$ and f_i as well as $\overline{R}_\alpha$ and K_a (see Appendix E for details). Each contour represents half a standard deviation of our theory.
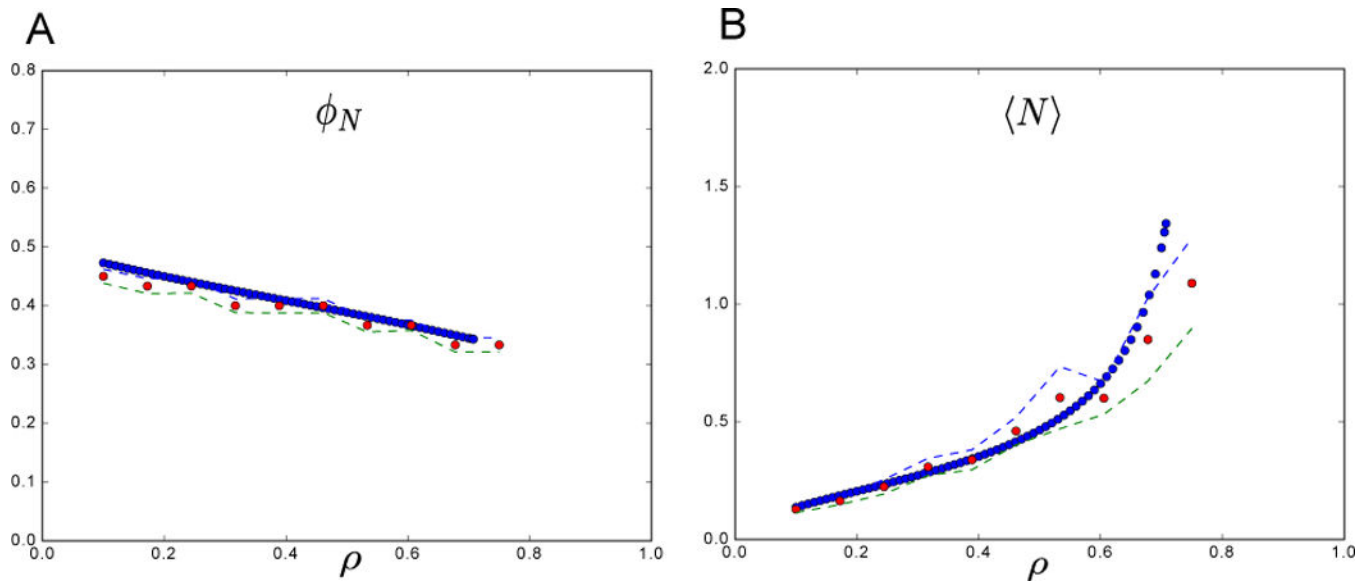
FIG. 5. Comparing CRM theory vs simulation - another setting.

All parameters defining the species were drawn from a Gaussian distribution as in main text but with $\sigma_m = 1$ (a larger maintenance cost variance). Interestingly, we see a different behavior than in Fig 2 in that the average abundance $\langle N \rangle$ of all species is increasing with increasing niche overlap. Also, the number of surviving species is reduced by high niche overlap as it is in Fig 2, which makes sense since when species are very similar they will compete more directly for the same resources. The parameters used are: $\sigma_m = 1$; $m = 1$; $\sigma_k = 1$; $K = 1$, and simulations were run assuming an equal number of species and resources $S = M = 30$ ($\gamma = 1$). Additionally, $\mu_c = 1$ was fixed and σ_c varied (note that the niche overlap ρ is determined by these two variables). The theoretical predictions from the cavity solution (blue), match well with the numerical solutions (red) for the CRM model averaged over 100 trials with and plotted with ±2 standard deviation error bars (dashed lines).