bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

The cavity method for community ecology

Matthieu Barbier[1] and Jean-François Arnoldi[1]

June 8, 2017

Abstract

This article is addressed to researchers and students in theoretical ecology, as an introduction to "disordered systems" approaches from statistical physics, and how they can help understand large ecological communities. We discuss the relevance of these approaches, and how they fit within the broader landscape of models in community ecology. We focus on a remarkably simple technique, the cavity method, which allows to derive the equilibrium properties of Lotka-Volterra systems. We present its predictions, the new intuitions it suggests, and its technical underpinnings. We also discuss a number of new results concerning possible extensions, including different functional responses and community structures.

Contents

1 Context **2**
1.1 Disordered systems . . . 2
1.2 Models of ecological dynamics . . . 4
1.2.1 The disordered systems framework . . . 4
1.2.2 Example: History dependence . . . 6

2 General framework **7**
2.1 The assembly process . . . 7
2.2 The assembled state . . . 7

3 What can we predict? **9**
3.1 Functioning . . . 10
3.2 Diversity and coexistence . . . 10
3.3 Stability . . . 11
3.4 Selected features . . . 13
3.5 Discussion . . . 13

[1]Centre for Biodiversity Theory and Modelling, Theoretical and Experimental Ecology Station, CNRS and Paul Sabatier University, 09200 Moulis, France.

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

4 Mathematical setting **14**
4.1 Scaling of interactions: when are large disordered communities not neutral? . . . 15
4.1.1 Inter-species versus intra-species effects . . . 15
4.1.2 Comparing communities of different sizes and connectivities 16
4.1.3 Tradeoffs and global stability in large systems . . . 17
4.2 Couplings rather than interactions . . . 18
4.2.1 Allowing heterogeneous self-interactions . . . 18
4.2.2 Reinterpreting γ . . . 20
4.3 Computing the assembled state . . . 20

5 Detailed calculations **23**
5.1 The cavity trick . . . 23
5.2 Universality of cavity results . . . 25
5.3 Abundance distribution . . . 26
5.4 Carrying capacities . . . 28
5.5 Productivity . . . 30
5.6 Response to perturbations . . . 30
5.6.1 Multistability order parameter . . . 31
5.6.2 Secondary extinctions . . . 32
5.6.3 Variability . . . 32

6 Sensitivity testing **32**
6.1 Distributions . . . 32
6.2 Interaction strength . . . 32

7 Extending the model **34**
7.1 Functional response . . . 34
7.2 Heterogeneous means . . . 36
7.3 Species groups . . . 38
7.4 Ordered species . . . 39
7.4.1 Linear order: competition-colonization tradeoff . . . 39
7.4.2 Exponential order: predation . . . 40

8 Appendices **42**

1 Context

1.1 Disordered systems

Many foundational insights of ecological theory were derived from the careful study of simplified patterns of interaction between a few distinct entities (individuals or species). Examples include trophic chains [18], competition and exclusion over a few resources [19, 34], mutualism [12] and parasite-host interactions [23]. These archetypal structures can be used as modules [14, 5] to understand more complex systems, when this complexity does not interfere with

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

their order and structure. For instance, our intuitions on multitrophic communities combine simple trophic mechanisms to understand inter-guild relations, and simple competitive mechanisms for intra-guild dynamics. Many theories also distinguish between ecosystems shaped by top-down control from predators, and those shaped by bottom-up control from resources. In all cases, these intuitions can only work if these interaction structures remain distinct even in a complex community.

In parallel, neutral models [15] have shown that many quantitative predictions can be made with minimal assumptions of structure, or even of distinctiveness of the entities. In most such works, the only accounting for an ecosystem's diversity is its number of otherwise identical species. These theories belong to the *mean-field* class, meaning that intrinsic differences between entities are erased statistically, owing to large numbers and some form of stochasticity. More precisely, such differences lose their importance in the collective dynamics – and therefore also in individual species' dynamics, provided that those are mostly shaped by collective factors (e.g. aggregate competition from many other species).

Disordered systems stand between these extremes, in that they exhibit heterogeneity without order: they represent situations where one cannot neglect the existence of differences between species, but no clear-cut structural patterns can be identified. As in neutral theories, it is either impossible or unnecessary to assign clear "roles" to individual species. But instead of being modelled as identical, all species appear to be completely heterogeneous in their biological features and interactions, without regular motifs. What we can know is the range and frequency of these traits in the community as a whole, and we investigate how much can be predicted about an ecosystem from that limited knowledge. While the notion of disordered system originates in physics[1], perhaps the simplest analogy is found in models of neural networks [24]. Each neuron has a distinct set of interactions, but it is difficult to identify the role of any given neuron, or group them into functional classes; mainly, we can access summary statistics of "how different" neurons and their interactions are in a given network.

This approach shares a conceptual starting point with Random Matrix Theory, which has long become a staple of theoretical ecology [22, 1]. In both cases, one considers a "simple sort of complexity": in large disordered communities, most details cease to matter and a few general mechanisms emerge, allowing us to unify many superficially different systems and models. However, Random Matrix Theory has developed with a specific focus on stability, whereas disordered systems techniques give access to properties – functioning, coexistence, abundance distributions, trait selection – which are usually predicted using neutral or structured models. In fact, many ecological models already fall into this category, especially numerical models that contain many species with uncertain attributes (lacking empirical or mechanistic constraints) such as large random

[1] It became prominent with the study of disordered materials such as glasses [26], before being applied to algorithmics and information theory [25], chemistry [27], economics [29] and systems biology [30].

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Lotka-Volterra communities. What has perhaps not been sufficiently realized is the availability of mathematical techniques for making broader and more robust statements about such systems, thus helping guide, validate or even bypass heavy simulation work. It is only recently that these techniques and ideas have been brought to bear on ecological problems [11, 16, 6, 32, 33].

Of particular interest is the application of this framework to *ecosystem assembly*, i.e. how dynamics and interactions shape the realized community from a pool of potential invaders. Indeed, it is easier to imagine that the pool itself is disordered: it may consist of species immigrating from a number of different environments, with independent evolutionary histories. By contrast, the assembled state that emerges from that pool will be less random. By following the process of assembly from that pool, it is possible to see which properties are selected by the ecological dynamics, i.e. how traits in the assembled community differ from those in the pool. Multiple works have thus explored what shapes species coexistence and invadability, in the context of assembly from a fixed regional pool [6], from immigration and stochastic extinction in an island biogeography perspective [16], or from two consumer communities "coalescing" – starting to compete on the same resources [32]. These methods have proven apt at generalizing to many species what was previously understood for modules of a few heterogeneous species, e.g. the conditions of competitive coexistence [33].

1.2 Models of ecological dynamics

1.2.1 The disordered systems framework

Practically speaking, the notion of disordered systems, and the fact that it allows for a simple and generic analytical approach, comes with significant methodological potential: it connects a class of commonly used models in ecology, synthesizing their insights and explaining why and when they may give similar or different predictions. We illustrate this in another work [3] with a "model meta-analysis" demonstrating that a variety of community structures, including mutualistic networks and resource competition, can still fall within the realm of disordered systems and create assembly patterns that we can predict with this framework.

This also means that if there were reasons to outright reject our results as a valid representation of ecological reality, then an entire class of models would become disqualified at once, as well as many qualitative arguments based on certain notions of complexity. That would be a useful result for ecological theory, limiting the field of investigation in models and simulations.

More concretely, to help situate this framework among preexisting work in theoretical ecology, we show in Fig. 1 a bird's eye view of the main types of dynamical models we identified in the field. They roughly fall between three extremes. First, fully independent populations, where we use physiological and behavioral traits to predict existence and functioning within a given environment. Second, fully dependent (i.e. very strongly interacting) populations where, for instance, the coexistence of competitors requires precise tradeoffs between their

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

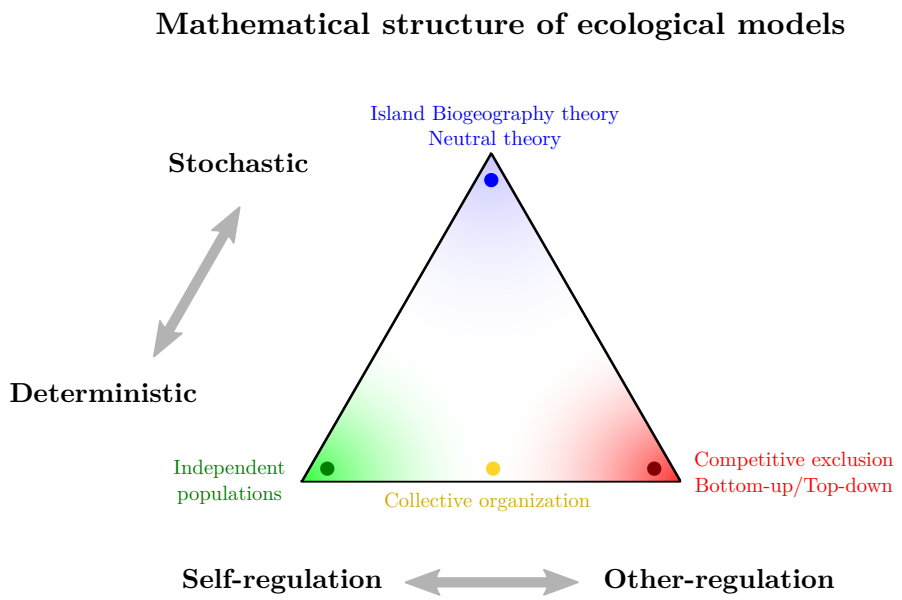

Figure 1: An abstract representation of the main classes of dynamical models in theoretical ecology, and the regime of collective organization where the notions and methods developed here are meant to apply.

traits. Third, fully stochastic models, where the characteristics of species and their interactions are all blurred and dominated by noise coming from internal (demographic) or external (environmental) sources. On this graph, our position is represented by the orange dot, a regime of collective organization where interactions are individually weak but collectively strong.

The two main axes of opposition on this graph are the following. First, deterministic versus stochastic: in deterministic approaches such as ours, we generally expect that real ecosystem properties can be understood from the equilibrium or near-equilibrium properties of a given dynamical model, controlled by some combination of intrinsic species fitness and by their interactions. The opposite pole includes theories that are dominated by stochastic and transient dynamics, where "noise" – in the demographics, the environment, or the immigration process – prevents the system from ever being close to some deterministic assembled equilibrium[2]. Those include neutral theories [15] and island biogeography and its variants [20, 10, 16]. These two viewpoints are complementary; only careful confrontation to empirical data can decide whether an ecosystem's dynamics are

[2]The term of "equilibrium" is ambiguous as it has been used for the Theory Island Biogeography among others. A fully stochastic system can indeed be characterized by probability distributions that attain a stationary state – in physics, this is called a Non-Equilibrium Stationary State, or driven stationary state, to distinguish it from the concept of equilibrium in deterministic dynamics and from thermodynamic equilibrium.

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

governed by transient, far-from equilibrium dynamics such as neutral turnover, or by the sort of near-equilibrium features described here.

The second axis of opposition confronts models whose dynamics are dominated by species characteristics, and those that are determined by interactions. Independent populations illustrate the former, competitive exclusion the latter. In these two extremes, results are often very sensitive to model assumptions: for instance, the colonization-competition tradeoff model ensures global coexistence by requiring a very precise ordering of competitive and dispersive ability. It may be surprising to see that the middle-ground – collective organization, where interactions are individually weak but collectively strong (because distributed among many species) – gives rise to very generic and predictable patterns.

This is the class of systems that we can describe with this framework. Note however that it could, fairly easily, be extended to include noise; by contrast, systems where species are strongly influenced by other individual species tend to have very dramatic dynamics (e.g. predator-prey cycles) which typically preclude the existence of any non-trivial equilibrium. Such dynamics are not entirely beyond the reach of disordered system approaches, but will almost certainly require more involved techniques than the simple one presented here.

1.2.2 Example: History dependence

Importantly, each class of models makes different predictions for the mechanisms of various ecological phenomena. We can illustrate it in the case of history-dependence (e.g. founder effects). In a deterministic equilibrium viewpoint, the long-term state of a system is dependent on its history only if its dynamics allow for multiple attractors. Of course, even with a single global attractor (e.g. a single final, stable and uninvadable community), if the time to reach it is very long, the system may retain some memory of its previous states at visible timescales. But this memory should become fainter as time passes, and any shock should only reset the dynamics to some earlier stage along a similar trajectory. This perspective is compatible with the intuitive notion of maturity[3] as found e.g. in Margalef [21].

By contrast, in a stochastic perspective, any composition is to some degree an accident, though it may hold for a long time if there are positive feedbacks (e.g. self-reinforcing density dependence) that transform an early advantage into a long-lasting one. Nevertheless, the composition may well undergo a dramatic shift, due to a shock or even out of steady noise, without it signalling any change in intrinsic species fitness. This is closer to the usual intuition behind the founder effect.

[3]Even with many local attractors, this picture may remain valid if the dynamics exhibit *aging*, i.e. the sytem is prone to falling into deeper and deeper attractors as time passes.

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

2 General framework

2.1 The assembly process

Here is the central conceit of this theory of assembly: we define a pool of species that can invade the community, and whose dynamical traits, such as growth rates and interactions, are known only statistically. We then allow sequences of invasions until the community reaches an uninvadable state, and investigate the properties of this "assembled" state – among others, species diversity, abundance, and how traits selected by the dynamics differ from the pool baseline.

This can represent a variety of physical settings: the pool could be a entire region and the community a local patch [28], they could be a mainland and an island connected by migration fluxes [20, 10, 16], or even oceanic water and the microbiome of an individual sponge [31]. More generally, this suggests a stable reservoir of biodiversity with some known aggregate properties, and an environment that is being populated from that reservoir, even if the reservoir is not larger (e.g. a forest serving as a biodiversity shelter for surrounding plains), provided that it can serve as a stable supply of potential invaders for the timescales considered.

But this can also represent, more abstractly, a "separation of concerns" between, for example, physiology and ecology. Properties of the pool reflect the traits of species "in potential"; by contrast, the assembled community retains only the subset of traits that are dynamically selected to survive together. Thus, in an abstract model, the pool could represent the range of species that are allowed to exist by physiological constraints, while assembly would single out combinations of such species that can coexist in the same ecosystem. Even in the spatial interpretation suggested before, there is such a separation of concerns: species-species interactions are properties of the pool, but species-environment interactions (e.g. carrying capacities) could be specific to a location, leading to different assembled states in different locales.

This means that the size of the pool, S, admits very different interpretations depending on the ecological scenario represented: S could be regional diversity or mainland diversity, if one is interested in invasion through immigration, but it could also represent the mutant strains that try to invade a microbial community over a given timescale (the pool being disordered would then reflect an assumption of random mutations). Carefully teasing apart these various interpretations and the hypotheses they hide is one of our purposes in this article.

2.2 The assembled state

Within this general setting, we decide to look only at the final assembled state, i.e. a feasible, stable and uninvadable community. As discussed in Sec. 1.2.1, this stands in stark contrast with theories that are dominated by stochastic and transient dynamics.

The mathematical techniques of disordered systems can be applied to many different dynamics. For convenience, we investigate the very commonly used

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Lotka-Volterra dynamics[4]:

$$\frac{d}{dt}N_i = N_i \left(R_i - D_i N_i - \sum_{j \neq i}^{S} A_{ij} N_j \right) \tag{1}$$

Here A_{ij} is the interaction between species i and j, and R_i is the growth rate of species i, that combines growth on external resources and mortality (and could thus be negative). D_i is intra-species competition, hereafter *self-interaction*. We can define

$$D_i = R_i / K_i \tag{2}$$

where K_i is properly the carrying capacity of species i in the case of consumers with $R_i > 0$. If $R_i < 0$, then K_i is negative and cannot be interpreted as a carrying capacity, but remains an important mathematical quantity (a measure of self-interaction or density-dependent mortality).

We will concern ourselves only with the equilibrium solutions, meaning $dN_i/dt = 0$ and

$$0 = R_i - D_i N_i - \sum_{j \neq i}^{S} A_{ij} N_j. \tag{3}$$

Only the species for which a positive solution $N_i > 0$ exists will then survive together in an assembled state. This state is obtained after long immigration sequence, during which all species that could invade did. As we explain below in Sec. 3.3, in a certain region of parameters, the precise sequence of invasions is irrelevant to the final state, which is a global attractor of the dynamics.

Important: To simplify the exposition, we assume starting now that all the self-interaction terms are equal $D_i = \langle D \rangle = 1$, so that $K_i = R_i$. This assumption is relaxed and the consequences of heterogeneous D_i are discussed starting in Sec. 4.

We will proceed to show that, to predict many properties of the assembled state, it is sufficient to know four simple statistical features of the pool of potential invaders. First, we need a measure of intrinsic fitness differences between species; for this we take ζ the standard deviation of the growth rates R_i. Second, we need to characterize interactions at a collective level. We thus ask how much a species is affected by the *sum* of all its interactions: how adverse this total interaction is on average, represented by a parameter μ ($\mu > 0$ means that biomass is on average lost through interactions, $\mu < 0$ means it is gained), and how much that varies from species to species, represented by a parameter σ. Hence, for any species, the total effect of interactions can be understood to fall roughly in the range $\mu \pm \sigma$. Finally, to understand feedbacks between species, we need to know how similar the effect of species i and j is to the reciprocal effect of j on i, and we encapsulate this in a parameter γ.

[4]Note however that in Sec. 7.1, we extend these results to a saturating functional response.

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Formally, these parameters are defined as follows:

$$\zeta^2 = \left\langle R_i^2 \right\rangle - \left\langle R_i \right\rangle^2 , \qquad \mu = s \left\langle A_{ij} \right\rangle ,$$
$$\sigma^2 = s(\left\langle A_{ij}^2 \right\rangle - \left\langle A_{ij} \right\rangle^2), \qquad \gamma = \frac{s}{\sigma^2} \left(\left\langle A_{ij} A_{ji} \right\rangle - \left\langle A_{ij} \right\rangle^2 \right) . \tag{4}$$

where $s = Sc$ is the average number of interaction partners for each species (c being the connectivity of the interaction network). This factor s means that μ and σ^2 indeed represent the *total* interaction strength and variance faced by one species from all its partners. Given our definition of γ, it takes values in the range $[-1, 1]$: $\gamma = 1$ means that two partners affect each other identically, while $\gamma = -1$ means that reciprocal effects are maximally different, see Sec. 4.2.2. Examples from competitive and trophic communities are given in the next section.

These parameters suggest how disordered systems differ from both mean-field and structured systems. The various species traits retain a non-negligible heterogeneity, encapsulated by ζ and σ. Yet, little more information is needed than those simple metrics, which give a basic estimate of "how different" species can be. The precise interpretation of these abstract parameters will become clearer, first from seeing what role they play in model results, and later in more in-depth discussion.

All higher-order correlations, that capture more precise structural information about the pool, are neglected. If more order is needed (e.g. for trophic systems, given that trophic roles are an essential feature that is not represented here), one could consistently push the formalism further, incorporating more and more structural properties in the form of correlations[5].

3 What can we predict?

While the following sections discuss the methods that we will use to solve the equations above, this one gives an overview of some of the main system properties that can be computed in this framework. These results fall roughly under four headings: functioning, diversity, stability, and selected features.

We present these results in the abstract parameter space defined in (4). The interpretation of the parameters comes in Sec. 4. Importantly, the results that we show are exact for the region of parameters where there exists a single, globally stable, uninvadable final assembled state. However, most results hold approximately past the boundary of this region, and we later discuss in Sec. 4.1 reasons why many systems that fit our basic assumptions are likely to be found in this region. The complex region that lays beyond is still being investigated.

We first give some details about the properties we compute, then discuss general trends in the parameter space.

[5]The reader may recognize this idea as similar to that of moment equations, see e.g. [7].

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

3.1 Functioning

For functioning, we use the total biomass

$$N_{tot} = \sum_i N_i \tag{5}$$

and total productivity or growth from external resources

$$G_{tot} = \sum_i N_i R_i \Theta(R_i) \tag{6}$$

with the Heaviside function $\Theta(x) = 1$ if $x > 0$, $\Theta(x) = 0$ otherwise. Throughout, we make no difference between biomass and abundance as we are not considering individual mass as a particular species trait; however the framework could easily be extended to take it into account. Our choice is equivalent to assuming that biomass has no systematic correlation with either interactions or carrying capacities in the pool of potential invaders, so that converting abundances to biomass only requires multiplying by the average individual's biomass in the community, which we can arbitrarily set to 1.

3.2 Diversity and coexistence

For diversity we use two main metrics. The first is the number of species surviving in the assembled state $S^* = S\phi$, see Fig. 2. As one would expect, coexistence decreases with increasing σ and ζ, since more variance in growth rates and interactions entails that some species are more likely to be overall stronger competitors than some others. Interestingly, as we discuss below, decreasing γ tends to increase coexistence.

The second is Simpson's index, which gives a measure of biomass concentration; i.e. its inverse tells us how effectively diverse the community is, in terms of the biomass being equitably distributed among all species. The index is computed as

$$I = \sum_i \left(\frac{N_i}{N_{tot}} \right)^2 \tag{7}$$

and is therefore a property of the abundance distribution. In fact, the full distribution can be computed within this framework, with some caveats discussed in Sec. 5.

3.3 Stability

We discuss three aspects of stability: the existence of multiple equilibria, variability around an equilibrium under stochastic perturbations, and the consequences of permanent community shifts – be they due to environmental changes or species extinction. All three can be shown to be related via simple expressions to pool parameters and to the fraction of surviving species ϕ.

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

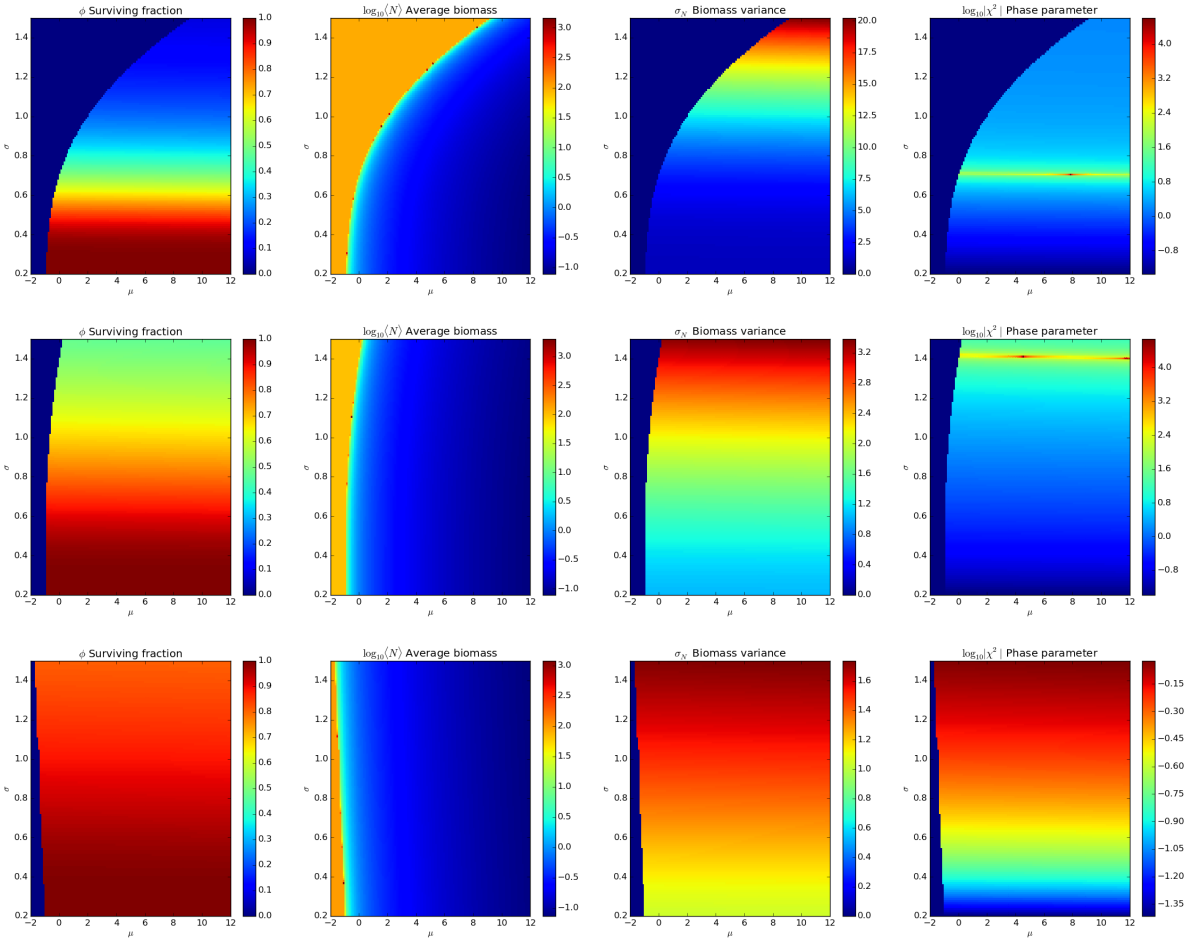

Figure 2: Coexistence, abundance and stability properties in the space (μ, σ) for $\gamma = 1, 0, -1$ (top to bottom) and $\zeta = 0$. From left to right: coexistence ϕ, average species biomass $\langle N \rangle$ and biomass variance σ_N. The uniform area in the left of each graph signals the parameter region where some interactions are mutualistic enough that abundances diverge, see Fig. 4 below. As is apparent here, when $\zeta = 0$, the average strength of couplings between species μ only affects the average biomass $\langle N \rangle$, while all other properties are controlled by heterogeneity σ and reciprocity γ. Results for $\zeta = 0.3$ are shown below in Fig 5, showing that μ becomes more significant.

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

First, we can find an expression for *invariability* as defined in [2], i.e. the inverse of the response of the system to demographic noise.

$$\mathcal{I} = 1 - \frac{\mu}{s} - \frac{\sigma^2}{2}(\gamma + 1)\frac{\phi}{1 - \mu/s} \tag{8}$$

It can formally be related to the response to other types of noise, and is also the basis for deriving a prediction for Taylor's law.

Second, we can derive the number of secondary extinctions resulting from the extinction of one species in the system. In a disordered system, secondary extinctions are more or less numerous depending on parameters, but never make up a significant fraction of the community, due to species being generally regulated by the community as a whole rather than by individual partners.

The final important result is χ the order parameter for multistability. Let us imagine that we perturb all the carrying capacities

$$K_i \to K_i + \xi_i \tag{9}$$

with a perturbation that is on average $\langle \xi_i \rangle = 0$, representing a small change in the environment that leads to different species being favored, without making the environment overall better or worse for the community. Then we can ask how much the abundances of the species are changed in response,

$$N_i \to N_i + x_i \tag{10}$$

and this change (with zero mean) is characterized by its variance, which we compute in Sec. 5.6 below:

$$\chi^2 = \left\langle \frac{x_i^2}{\xi_i^2} \right\rangle = \frac{1}{\frac{1}{4}\left(\sqrt{u^2 - 4\gamma\sigma^2\phi} + u\right)^2 - \sigma^2\phi}. \tag{11}$$

where $u = 1 - \mu/s$. This expression reflects the deep relationship between coexistence ϕ and sensitivity to the environment χ.

The divergence of χ^2 when the denominator of (11) vanishes, i.e. when

$$\sigma^2 \to \frac{1}{\phi}\left(\frac{1 - \mu/s}{1 + \gamma}\right)^2 \tag{12}$$

signals the boundary of the region where multiple equilibria exist: precisely at the boundary, an infinitesimal change in intrinsic fitness can lead to different species being present in the final assembled state. Of course, while the quantity formally diverges, the community's response is not truly infinitely steep – it only signals that the approximations we use to compute the single globally attractive equilibrium break down. Indeed, as one gets close to the divergence, χ stops predicting the simulated response to a press perturbation, which plateaus as shown in Fig. 3.

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

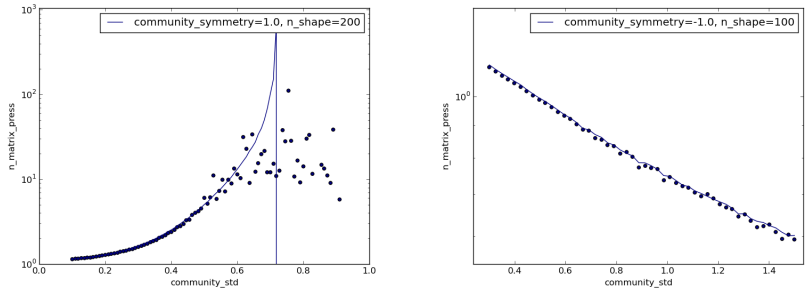

Figure 3: Average effect of a random perturbation of carrying capacities, calculated as χ (solid line) or directly from the assembled community in simulations (dots), depending on coupling heterogeneity σ. Left: $\gamma = 1$, with divergence of χ at finite σ, revealing the loss of stability of the single global attractor and the start of the multistable phase. Right: $\gamma = -1$, no divergence (no multistable phase).

Hence, it is possible to construct the *phase diagram* shown in Fig. 4, which indicates where in parameter space one will find a single, multiple or no equilibria. As we discuss more below, lowering γ has a stabilizing effect, as it extends the single equilibrium phase – in fact, it is apparent from (12) above that, in the limit $\gamma = -1$, the threshold to the multiple equilibria phase becomes infinite, and there can only be one attractor which is globally stable.

3.4 Selected features

Previous work [6] has shown that it is possible to use this framework to infer not only the distributions of abundances, growth rates and interactions in the assembled state, but also correlations between those, including the frequency of various small network motifs [14]. How these distributions differ from the pool baseline tells us how non-random properties have been selected from a random pool. This gives us worthwhile information on which factors are likely to enable a species to survive into the assembled state, or conversely, how ecosystems may adapt to ensure coexistence. We demonstrate this idea in this article with the selection of carrying capacities or growth rates, but see the Appendices in [6] for details on the selection of interaction patterns.

3.5 Discussion

Most striking is the role of γ the reciprocity of interactions: as γ becomes negative, sensitivity to perturbations decreases, the single attractor region expands and coexistence increases. This suggests that asymmetrical interactions – including commensalism, amensalism and trophic interactions – lead to easier

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

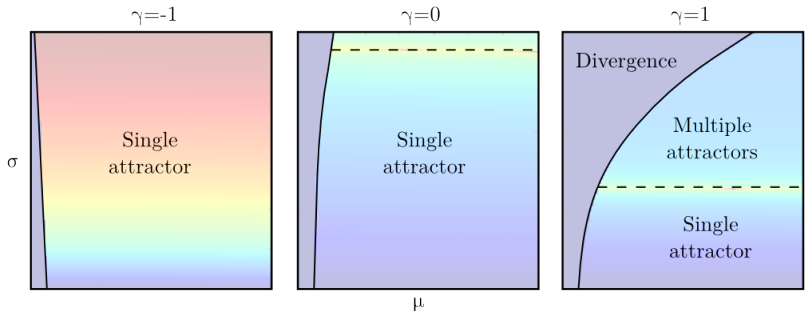

Figure 4: Phase diagram in the (μ, σ) plane for various γ and $\zeta = 0$, based on phase parameter χ^2 from Sec. 3.3. There are three regions depending on parameters. In the first region, there is a single global attractor, meaning that – after a long enough sequence of invasions and extinctions – a given community will reach a single uninvadable state, no matter what sequence it went through. In the second region however (up of the dashed line, signalling a divergence of χ^2), there are multiple such attractors, and the final assembled state depends on the history of the system. As $\gamma \to -1$, the transition to the second region disappears. Finally, the third region (left of the solid line, signaling a Rdivergence of $\langle N \rangle$) has no attractor, as some species have mutualistic interactions that are stronger than any negative feedback and their abundances grow to infinity. This signals a breakdown of the Lotka-Volterra approximation for more realistic dynamics.

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

assembly than symmetrical ones, echoing the conclusions of other authors [1] who noted "remarkable differences between predator-prey interactions, which increase stability, and mutualistic and competitive, which are destabilizing". Yet, we can also show that $\gamma < 0$ may very well be achieved by competitive interactions. It is in fact possible to construct trophic and competitive communities so that the latter have more negative γ than the former. Hence, this statement should not be taken as generally valid, and the meaning of γ should be discussed much more carefully, as we do in Sec. 4.2.2.

In the single attractor region, many different ecological scenarios become indistinguishable as long as one is only interested in the eventual assembled state. Whether invaders from the pool come from migrations or mutations, frequent or rare, correlated or not with the species dominating the community at any given time, interspersed with extinctions and recoveries or not – these distinctions become irrelevant in the long run, as any dynamical trajectory that does not involve irreversible changes in the pool of potential invaders (e.g. a permanent extinction) will end at the same point.

If we look at a pool that lies on the boundary between the two regions, multiple attractors first emerge as small deviations from the single attractor that existed below that boundary - most species remain, and only a few rare ones distinguish the multiple attractors by their absence or presence. Therefore, even though the results presented in this work focus on the single attractor region, where our analytical results hold exactly, we will demonstrate multiple times that our framework still provides good approximations some distance into the multiple attractor region. When our framework ceases to be predictive altogether, it is likely that the various attractors are different enough that perturbations may engender significant regime shifts.

If $\zeta \neq 0$, the model predictions seen on Fig. 5 are still quite similar to those shown in Fig. 2 for $\zeta = 0$, except for the fact that average coupling μ plays a much more important role: instead of controlling only the average biomass $\langle N \rangle$, it affects all the metrics, favoring the existence of a global attractor while reducing coexistence.

4 Mathematical setting

This section is meant to be an qualitative explanation of the mathematical concepts that are crucial for a better understanding of the methods and results shown in this paper. Sec. 5 below gives a detailed explanation of the derivation, of interest to the more mathematically-inclined reader.

We recall that the central equation we investigate is (3) which defines the equilibrium for Lotka-Volterra dynamics:

$$0 = R_i - D_i N_i - \sum_{j \neq i}^{S} A_{ij} N_j.$$

Before solving this equation, we show that it is possible to better isolate the role of the various parameters, thanks to two simple tricks of rescaling. First,

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

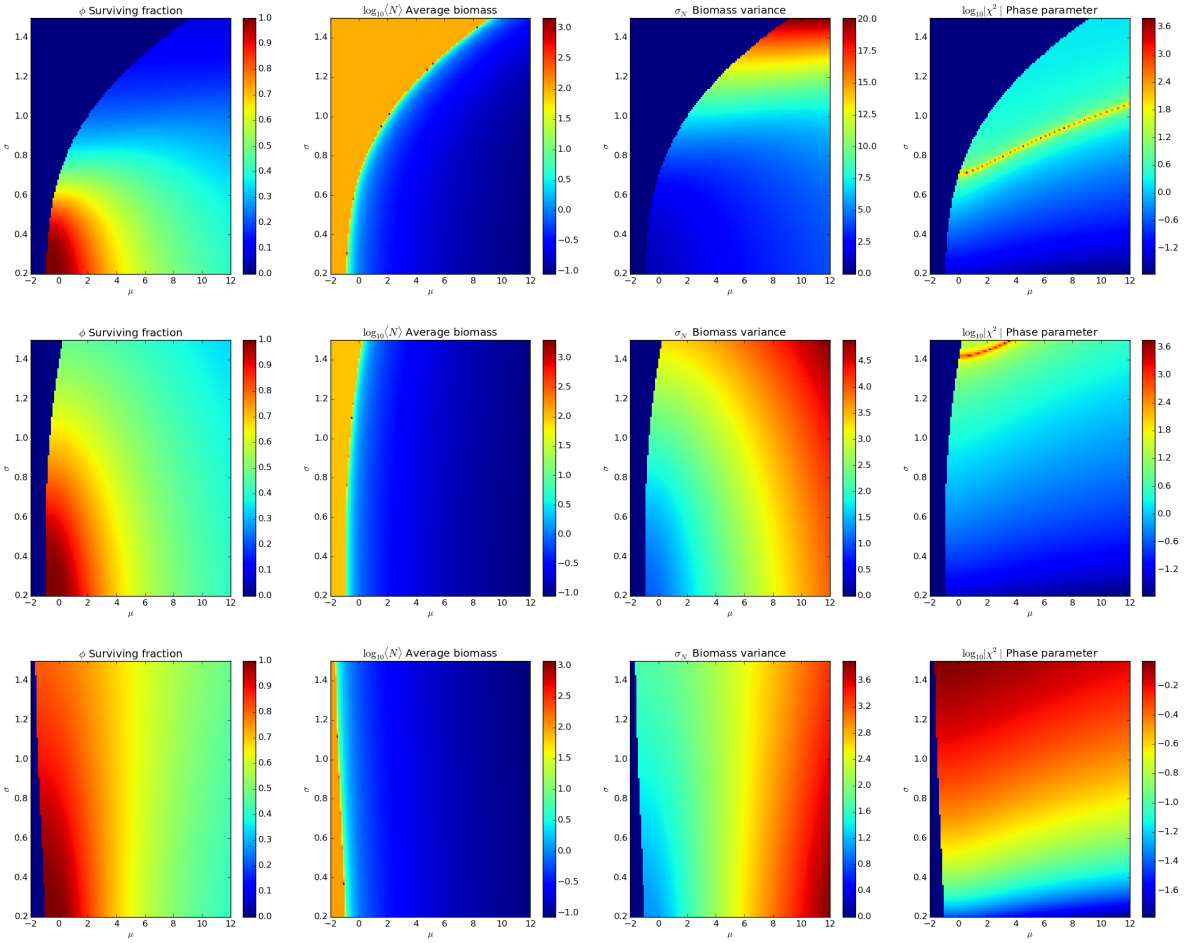

Figure 5: Coexistence, abundance and stability properties in the space (μ, σ) for $\gamma = 1, 0, -1$ (top to bottom) and $\zeta = 0.3$. From left to right: coexistence ϕ, average species biomass $\langle N \rangle$ and biomass variance σ_N. See Fig. 2 for results at $\zeta = 0$.

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

it is easily seen in (1) that R_i, D_i and A_{ij} could all be multiplied by the same number, without changing the dynamics (only making them globally faster or slower). Hence we can arbitrarily choose the time scale so that, for instance, growth rates have unit mean:

$$\langle R_i \rangle = 1. \tag{13}$$

Now, we can look more specifically at equilibria, i.e. equation (3). Clearly, we could multiply all D_i and A_{ij} by some number, and divide all N_i and N_j by the same number, without changing the equilibrium: the average intensity of interactions, inter- and intra-species, simply changes the average abundance at equilibrium. Hence we can also arbitrarily take

$$\langle D_i \rangle = 1 \tag{14}$$

so that now, interactions A_{ij} are measured in units of the average self-interaction, and abundances are measured in units of "carrying capacity" – indeed, if it were alone with positive growth rate, the average equilibrium abundance of the species would be around 1 (or equal to 1 if $D_i = 1$). Neither of these choices subtract anything from the generality of the equations. An important consequence is that any change of the absolute value of growth rates or interactions, for instance via nutrient enrichment, will affect only the total amount of biomass and no other property of the assembled state.

The second trick is using the fact that interactions can always be rewritten as

$$A_{ij} = A + \sigma_A a_{ij} \tag{15}$$

with two parameters: the average interaction $A = \langle A_{ij} \rangle$ and the variance of interactions $\sigma_A^2 = \langle A_{ij}^2 \rangle - A^2$, so that the rescaled interactions a_{ij} have zero mean and unit variance

$$\langle a_{ij} \rangle = 0, \qquad \langle a_{ij}^2 \rangle = 1 \tag{16}$$

For now, we make no further assumption on what the distribution of the A_{ij} and a_{ij} may be, beyond the fact that it has a well-defined mean and variance. We simply suggest that is important to know both how intense interactions are on average, i.e. A, and how much they differ between pairs of species, σ_A.

Hence, we transform (3) into the following equation

$$0 = R_i - D_i N_i - A \sum_j N_j - \sigma_A \sum_j a_{ij} N_j \tag{17}$$

and using

$$\sum_{j \neq i}^{S} N_j = S \langle N \rangle - N_i. \tag{18}$$

we finally get

$$0 = (R_i - SA \langle N \rangle) - N_i (D_i - A) - \sigma_A \sum_j a_{ij} N_j \tag{19}$$

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

The first term is an effective growth rate including the overall biomass loss due to competition ($A > 0$) or gain due to mutualism ($A < 0$). The second term contains the effective intra-species competition, and the third the relative gain or loss of fitness of species i due to the *heterogeneity* of inter-species interactions. Multiple insights already arise from this simple reformulation, without even solving the equation, and we discuss them now.

4.1 Scaling of interactions: when are large disordered communities not neutral?

4.1.1 Inter-species versus intra-species effects

It is well-known that intra-species competition plays a large role in stabilizing the dynamics: interactions with other species become important only if they are not negligible compared to this self-interaction. Strongly interacting communities can exhibit complex dynamics with multiple equilibria and limit cycles. If interactions with others are negligible, each species (with positive growth rate) simply follows its own logistic equation, and the system is sure to evolve toward a globally attractive equilibrium.

The equation (19) above shows that the effective importance of interactions, compared to self-interaction (and beyond a simple change in the mean biomass), is controlled by the ratio

$$\frac{\sigma_A}{D_i - A} \tag{20}$$

This gives a clear intuition of why species heterogeneity, as encapsulated by σ_A, can be important in the dynamics. If species are too similar, $\sigma_A \to 0$, then the abundances N_i decouple from each other: species are only coupled collectively through the average value $\langle N \rangle$ in the effective growth rate, except in the very special case of the Hubbell point $D_i = A$ where all inter- and intra-specific interactions are equal [15, 16]. Away from that singular point, we see that what truly destabilizes the system is not only how *intense* interactions are, i.e. A at the denominator, but also how *diverse* they are, i.e. σ_A at the numerator.

4.1.2 Comparing communities of different sizes and connectivities

A second central insight comes from asking: what does it mean to say that two pools with different sizes S have the *same amount* of heterogeneity? Let us assume that each species is connected to s_i other species in the interaction matrix, on average

$$\langle s_i \rangle = s = cS \tag{21}$$

with $c \in [0, 1]$ the connectivity of the network. If s is the same in the two pools[6], then species from either pool have the same average number of interactions. In that case, it makes sense to directly compare the properties of these interactions,

[6]Note that keeping the same s for different pool sizes S means that connectivity $c = s/S$ must decrease with pool size.

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

i.e. the values of A and σ_A,s to decide which pool has stronger or more diverse interactions.

However, let us assume that c is the same for both pools, for instance $c = 1$. In that case, the number of interactions, and hence the strength of the interaction between one species and all the others, will increase with pool size S proportionally to the quantity

$$F_i = \sum_{j \neq i}^{s} A_{ij} = sA + \sigma_A \sum_{j \neq i}^{s} a_{ij} \tag{22}$$

(this gives a general trend; the real interaction strength will depend on species abundances). This quantity is on average

$$\langle F_i \rangle = sA \tag{23}$$

which increases with the size of the community. This term is seen in (19) to intervene in the effective growth rate. As for how much F_i varies between species, that comes from the sum

$$\sum_{j \neq i}^{s} a_{ij} \sim \pm\sqrt{s} \tag{24}$$

By assumption, this sum is an unbiased random walk of mean 0 and typical step size 1; hence, after s steps (with s large enough), no matter the distribution of a_{ij}, the random walk will typically have gone a distance of order $\sqrt{s}$ toward either positive or negative values. There are two important assumptions here: one of large numbers, which can only be checked numerically (in fact, we can often find good agreement even for $s \sim 10$), and one of randomness which is discussed more carefully in Sec. 5. If one accepts these assumptions for now, the final interaction strength perceived by any species falls roughly in the range

$$F_i \sim sA \pm \sqrt{s}\,\sigma_A \tag{25}$$

This is a crucial point: if we consider different systems with the same values of A and σ_A, but different values of S or c, then we can see from (25) that, the larger and more connected the pool, the smaller the relative variation in F_i compared to its mean. In fact, for very large communities, there will be practically no difference in the interaction strength perceived by different species, making them *mean-field* (all species are effectively identical from the point of view of interactions) rather than disordered.

Instead, two pools of different size S have "equal" interaction strength and heterogeneity if their pools exhibit the same values of

$$\mu = sA = s\,\langle A_{ij} \rangle\,, \qquad \sigma^2 = s\,\sigma_A^2 = s(\langle A_{ij}^2 \rangle - A^2) \tag{26}$$

rather than the same values of A and σ_A. To understand this, we can see that, no matter the pool size or connectivity, a species' total interaction with others is given by

$$F_i \sim \mu \pm \sigma. \tag{27}$$

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

We can indeed show that μ and σ (in addition to a few other parameters) are what controls the properties of the assembled state – systems with different S but the same μ and σ will obey the same trends in terms of diversity, stability, abundance distribution, and so on. This rescaling means that systems with more connections between species (i.e. larger S or larger c) must have proportionally more varied interactions to retain the same σ, so that heterogeneity still matters and is not "averaged out".

4.1.3 Tradeoffs and global stability in large systems

We now further investigate this inverse perspective: instead of asking how *community parameters* μ and σ change with diversity S if we keep the same distribution of *individual parameters* A_{ij}, we ask how these A_{ij} should change with S to achieve a given type of community μ and σ. This simple question actually contains an important insight on the properties of large assembled communities. It is easy to see that if one wants to fix connectivity c as well as μ and σ, but increase S, then necessarily interactions should vary like

$$\langle A_{ij} \rangle = \frac{\mu}{s} \propto \frac{1}{S}, \qquad \sigma_A = \frac{\sigma}{\sqrt{s}} \propto \frac{1}{\sqrt{S}}. \tag{28}$$

For large enough systems $S \gg 1$, if we maintain the same values of μ and σ, we thus expect $\sigma_A \gg \langle A_{ij} \rangle$. Having a standard deviation much larger than the mean is not achievable with usual distributions[7], unless we allow the interactions to take any sign (the mean can thus go to zero).

Hence, due to simple mathematical constraints, large systems must have at least one of these four properties: large μ, small σ, small connectivity c, or mixed interactions allowing for all signs. For instance, a large competitive community where all species can potentially interact ($c = 1$) must have either large μ, meaning very strong competition and low biomass, or small σ, meaning that heterogeneities in species interactions cease to matter and the system is effectively mean-field. This "biomass–heterogeneity" tradeoff in large connected competitive communities, which we emphasize comes from simple algebra independent of any model, is indeed very apparent in our analysis of resource consumer models in [4].

This tradeoff has a very important consequence: as shown in recent work [35], small σ is effectively the condition under which a complex ecosystem can be reduced to a single dimension, leading to predictions of "universal" behavior [8]. Furthermore, small σ and small ζ taken together constitute the basic hypothesis of neutral theory – that we can take all species as identical in large communities, despite their very real heterogeneities. Hence, the fact that simple algebraic constraints can limit σ in large competitive communities (unless interactions are

[7] It can be achieved with a power-law tail whose exponent is arbitrarily close to 2, meaning that the variance of the distribution is close to diverging. This is somewhat equivalent to effectively having low connectivity, since there will be a few very strong interactions and a massive number of almost-zero ties that do not even sum up to much.

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

extremely strong) provides formal grounding for the intuitions behind "universal" and neutral dynamics.

4.2 Couplings rather than interactions

4.2.1 Allowing heterogeneous self-interactions

At multiple points in the discussion above, we have suggested that interactions are not, by themselves, the quantities that control the dynamics: instead, that role is taken by *couplings*, i.e. the ratios of interactions to self-interaction, which represents how much the influence of another species could cause one to deviate from a simple logistic growth.

In Sec. 2, we presented results assuming that all self-interactions were equal, $D_i = 1$. But in fact, we can easily see that we can transfer our conclusions to the Lotka-Volterra equilibrium equation (3) *divided by* D_i[8]:

$$\begin{aligned} 0 &= \frac{R_i}{D_i} - N_i - \sum_{j \neq i}^{S} \frac{A_{ij}}{D_i} N_j \\ &= K_i - N_i - \sum_{j \neq i}^{S} \alpha_{ij} N_j. \end{aligned} \tag{29}$$

Here, we recall $K_i = R_i/D_i$ the carrying capacity of species i, and $\alpha_{ij} = A_{ij}/D_i$ the coupling of species i to species j. Thus, all the results we presented in Sec. 2 still hold if, instead of using the variance of growth rates, we use that of carrying capacities, and instead of using the statistics of A_{ij}, we use those of α_{ij}. In other words, we formally redefine the parameters of (4) as follows:

$$\begin{aligned} \zeta^2 &= \left\langle K_i^2 \right\rangle - \left\langle K_i \right\rangle^2, & \mu &= s \left\langle \alpha_{ij} \right\rangle, \\ \sigma^2 &= s(\left\langle \alpha_{ij}^2 \right\rangle - \left\langle \alpha_{ij} \right\rangle^2), & \gamma &= \frac{s}{\sigma^2} \left(\left\langle \alpha_{ij} \alpha_{ji} \right\rangle - \left\langle \alpha_{ij} \right\rangle^2 \right). \end{aligned} \tag{30}$$

We can relabel the parameters and retain the same mathematical structure, and their interpretation does not change too dramatically – except in the important case of γ, as we soon discuss. Previously in this section, we used $A = \langle A_{ij} \rangle$ and σ_A extensively to discuss simple and general properties of the dynamics – the distribution of the interactions A_{ij} is otherwise entirely unspecified, and could take arbitrary forms (normal, exponential, uniform, gamma, multimodal...) except for the pathological. We saw in examples that almost all other properties of this distribution $P(A_{ij})$ were irrelevant to assembly for large enough communities, one of the very counter-intuitive and attractive traits of disordered systems. In fact, we see now that, if self-interactions differ, we should use the α_{ij} instead; but thankfully, results remain insensitive to the details of

[8] In fact, this is a useful parametrization, since the Lotka-Volterra approximation to more realistic dynamics may not hold if couplings are too large $|\alpha ij| > 1$ (e.g. mutualistic interactions would then cause abundances to diverge). Hence couplings also give an idea of the range of validity of the model.

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

$P(\alpha_{ij})$. The subtlety is found in the fact that σ now includes not only the heterogeneity of interactions, but also that of self-interaction – in other words, species with identical interactions but very different growth rates and carrying capacities will have large σ, with all the consequences shown in Sec. 2 in terms of coexistence and multistability.

This of course requires that $D_i \neq 0$ – and in fact that $A < D_i$, as the case of very strong interactions $A > D_i$ tends to lead to very unstable dynamics that we wish to avoid entirely in this approach[9], as discussed in 1.2.1.

4.2.2 Reinterpreting γ

However, there is one crucial change in interpretation. The parameter $\gamma \in [-1, 1]$ represents the average degree of reciprocity between the dynamical coupling of species i to j, and that of j to i. The extreme case $\gamma = 1$ indicates fully symmetrical couplings $\alpha_{ij} = \alpha_{ji}$, while $\gamma = -1$ means anticorrelated couplings. Unfortunately, while this parameter is the one that truly encapsulates how asymmetry between interaction partners affects the assembly process, its interpretation is very subtle and counter-intuitive.

Let us first discuss γ in the previous case where all self-interactions are equal, $D_i = 1$, as even that simpler situation has some difficulties. If $A = 0$, then $\gamma < 0$ suggests trophic interactions, since A_{ij} and A_{ji} will typically be of opposite signs. However, if $A \gg \sigma_A$, then even with $\gamma = -1$, $A_{ij} \sim A + \sigma_A$ and $A_{ji} \sim A - \sigma_A$ will have the same sign. They will simply be maximally anticorrelated: for every pair of species, one will affect the other much more than it is affected by it. This can for instance give rise to "rock-paper-scissors" intransitive competition.

If we now return to varied self-interactions $D_i \neq 1$ so that couplings are different from bare interactions, even that refined notion of reciprocity becomes more complex. Consider for example the classic resource competition models [19, 9, 34]. By construction, interactions come from two species harvesting the same resource and are therefore perfectly symmetrical, $A_{ij} = A_{ji}$. However, couplings are *not* – in fact, depending on parameters, the model can give any γ between -1 and 1, and most plausible parameter values correspond to $\gamma < 0$. In any pair of competitors, each reduces the other's growth equally, yet this will on average have a much larger impact on the dynamical trajectory of one species than the other.

Hence, the proper way to understand γ is in relative terms: no matter whether the interaction is positive or negative or even symmetrical, if $\gamma = -1$, one species will respond much more to the interaction (in the sense that it will be more affected compared to the trajectory it would have if alone), while for $\gamma = 1$ they will respond equally. Given that assembly is a process of selecting species that are most compatible among a larger pool, it can be understood that what really matters is not the net outcome of each interaction, but how

[9] A classic case is Levins and Culvers's colonization-competition tradeoff [17] where only a very special and asymmetrical structure ensures the survival of more than a single species. We discuss it further in [4].

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

differently it affects the partners' dynamics. But despite its predictive power, any intuitive interpretation of this parameter can only be made in the context of a more concrete model.

4.3 Computing the assembled state

At the start of Sec. 4, we showed that the equilibrium of the Lotka-Volterra equations (3), which we recall is given by

$$0 = R_i - D_i N_i - \sum_{j \neq i}^{S} A_{ij} N_j,$$

can be rewritten in all generality as

$$0 = (R_i - SA \langle N \rangle) - N_i (D_i - A) - \sigma_A \sum_j a_{ij} N_j. \tag{31}$$

given the mean $A = \langle A_{ij} \rangle$ and the variance of interactions $\sigma_A^2 = \langle A_{ij}^2 \rangle - A^2$, with the rescaled interactions $a_{ij} = (A_{ij} - A)/\sigma_A$.

As we argued in Sec. 4.2.1, the quantities that truly control the dynamics are not "bare" interactions A_{ij}, but couplings

$$\alpha_{ij} = \frac{A_{ij}}{D_i}$$

which represent how much species j can cause species i to deviate from its baseline (logistic) trajectory controlled only by self-interaction D_i. We now rewrite

$$\alpha_{ij} = \frac{\mu}{s} + \frac{\sigma}{\sqrt{s}} a_{ij} \tag{32}$$

with

$$\langle a_{ij} \rangle = 0, \qquad \langle a_{ij}^2 \rangle = 1. \tag{33}$$

Then, we have

$$0 = K_i - N_i - \sum_{j \neq i}^{S} \alpha_{ij} N_j.$$

and instead of (31) we get

$$0 = k_i - u N_i - \frac{\sigma}{\sqrt{s}} \sum_j a_{ij} N_j \tag{34}$$

where we have defined the effective self-interaction

$$u = 1 - \frac{\mu}{s} \tag{35}$$

and the effective carrying capacities

$$k_i = K_i - \mu \langle N \rangle \tag{36}$$

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Following previous work [6], we use the cavity method from physics [26] to find the properties of the equilibria of equation (34). Technical details are given in Sec. 5. To understand the results and their limitations, it is enough to grasp the following simplified discussion, introducing some important concepts.

Let us take an assembled state created from a pool with $S-1$ species. When the final species i invades, if S is large, the properties of the assembled state should not change significantly[10]. Hence, we imagine that we can move from $S-1$ to S species holding everything constant, except for the fact that the other species have their final abundance N_j shifted a small amount, in the direction suggested by their couplings with i,

$$N_j \to N_j - v a_{ji} N_i \tag{37}$$

with v a positive "response coefficient" that we can compute. Inserting this shifted abundance in (34), a feedback is created on species N_i. This is where the parameter γ enters the picture: if $\gamma > 0$ the feedback is positive (e.g. a species reduces its competitors' abundance, which in turn allows it to grow more), while if $\gamma < 0$ the feedback is negative (e.g. a predator reduces its preys' abundance, which stifles its growth). This explains how negative γ can play a stabilizing role in the dynamics, as we have seen above starting in Sec. 3.3.

The finite trick is to see that this equation expresses N_i as a random variable obeying:

$$N_i = \frac{k_i - \sigma a n}{1 - \frac{\mu}{S} - \sigma^2 \gamma v} \tag{38}$$

where a is a Gaussian random variable of mean 0, derived from the random walk (24) on the a_{ij}'s, and n is a random variable drawn from the distribution $P(N)$ of surviving species. This simplification can be made technically precise, and relies on a crucial property of disordered systems: there is no correlation between the a_{ij}, which represent direct couplings, and the N_j *before* the shift (37), which are shaped by all the other indirect couplings in the system. These two could not be taken as independent parameters if each species depended strongly on a few others, but the disordered nature of interactions means that species abundances are shaped mostly by collective dynamics involving the whole community. A failure of this approach to correctly predict the behavior of some model could signal that local ordered dynamics prevail over collective disordered dynamics.

The second central assumption that is made here is the existence of a single global attractor, meaning that no matter which initial condition is taken, the same assembled state is reached provided all species are given a chance to invade when they can. This premise is hidden in the fact that we assume that equilibrium species abundances do not shift drastically with the addition of a new

[10] This assumption is correct if there is only one globally stable assembled state; the method should later be extended to settings with alternative stable states, where the addition of a new species could cause a sudden shift between different equilibria.

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

species. As we have mentioned in Sec. 3.3, we can compute a quantity that becomes infinite precisely at parameter values for which there ceases to be a single attractor, thus signalling the appearance of potential tipping points. Methods in physics have been derived for the express purpose of investigating such multistable regimes, and quantifying the number and stability of attractors. In the case of Lotka-Volterra dynamics, however, this task is ongoing.

The assembled state must be feasible and uninvadable, and therefore, studying it reveals what controls species coexistence. But it also contains information on the ecosystem's sensitivity to external change: the quantities that arise naturally when computing the assembled state are long-term responses to a press perturbation, i.e. how the whole community's equilibrium shifts in response to some species being pushed toward an abnormal abundance. This can come from durable environmental changes that affect the fitness of some species, but also from the introduction of a new invader into the system.

5 Detailed calculations

5.1 The cavity trick

The assembly process shapes the community by changing the abundances of species until an equilibrium is reached. From the imposed properties of the pool, assembly uses the abundances as its tool to select and reweigh species features, crafting a state that is feasible, stable and uninvadable by any other species in the pool. Hence, all the properties of the solution can be derived from the distribution of abundances and their correlations with species features (such as carrying capacities and interactions).

The logic of the cavity method is remarkably simple: 1) we compute what happens after a single invasion in an already large community, and 2) we then require that the equilibrium features before and after the invasion be unchanged. The second part means that we are computing the fixed point of the invasion process, i.e. an equilibrium which cannot be invaded nor destabilized, rather than a single-step invasion as is often done in population genetics (testing whether a given mutant species can invade).

As shown in Fig. 6, the calculation of the invasion is made simpler by assuming that, even though the invader itself retains its nonlinear dynamics, the effect of the invader on each existing species is small enough to be approximated by a linear change of abundance. This assumption may seem intuitive if the final abundance of each species is the result of many interactions, so adding one more should not cause a dramatic change. A more formal justification is given in Appendix.

Let us introduce species 0 into a community of $S-1$ species. Its equilibrium abundance N_0 changes other N_i a little, as they now obey:

$$0 = k_i - uN_i - \sigma_A \sum_j a_{ij} N_j - \sigma_A a_{i0} N_0 \tag{39}$$

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

(where we recall that $\sigma_A = \sigma/\sqrt{s}$) This is equivalent to changing their carrying capacities k_i a little

$$0 = (k_i - \sigma_A a_{i0} N_0) - uN_i - \sigma_A \sum_j a_{ij} N_j \tag{40}$$

We now assume that N_i responds linearly to this small perturbation: using a Taylor expansion around the previous equilibrium abundance, without species 0, we can write

$$N_i \approx N_i^* - \frac{dN_i}{dk_i} \sigma_A a_{i0} N_0 \tag{41}$$

In fact, all other species in the community perceive the effect of adding species 0, so the full linear response can be generically expressed as

$$N_i = N_i^* - \sigma_A \sum_j v_{ij} a_{j0} N_0 \tag{42}$$

where we define the response coefficient

$$v_{ij} = \frac{dN_i}{dk_j} \tag{43}$$

Now, this shift feeds back into the equation for N_0, which is

$$0 = k_0 - uN_0 - \sigma_A \sum_i a_{0i} N_i \tag{44}$$

$$= k_0 - \left(u - \frac{\sigma^2}{s} \sum_{i,j} a_{0i} v_{ij} a_{j0} \right) N_0 - \sigma_A \sum_i a_{0i} N_i^* \tag{45}$$

As we discuss in Appendix, the complicated-looking sum on i and j is in fact a self-averaging quantity, meaning that it tends to become a single well-defined number for all species in the community.

$$\frac{\sigma^2}{s} \sum_{i,j} a_{0i} v_{ij} a_{j0} \approx \frac{\sigma^2}{s} \sum_{i=1}^{s^*} a_{0i} v_{ii} a_{i0} \approx \frac{\sigma^2}{s} s\phi\gamma v \tag{46}$$

with

$$v = \langle v_{ii} \rangle \,. \tag{47}$$

and $s^* = s\phi$ is the number of species that interact with species 0 and are still alive in the assembled state (we recall that ϕ is the overall fraction of surviving species). To interpret v, let us remark that, if species were independent, then

$$N_0 = k_0, \qquad v_{00} = \frac{N_0}{k_0} = 1 \quad \text{(independent species)}. \tag{48}$$

Hence, $v < 1$ indicates that species abundances are not defined by their intrinsic fitness, but by their effective fitness including interactions – it is possible to

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

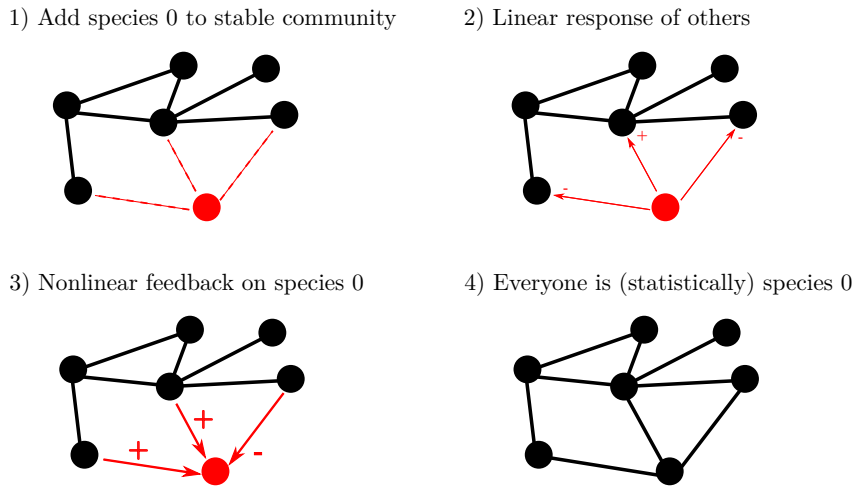

Figure 6: Cartoon of the cavity method.

change their carrying capacity without affecting their abundance proportionally. The lower v, the more different this effective fitness.

Finally, we obtain the equation for N_0 which will be solved below.

$$\left(u - \phi\sigma^2\gamma v\right) N_0 = k_0 - \frac{\sigma}{\sqrt{s}} \sum_i a_{0i} N_i^* \tag{49}$$

5.2 Universality of cavity results

Isolating N_0 on the left-hand side,

$$N_0 = \frac{k_0 - \frac{\sigma}{\sqrt{s}} \sum_i a_{0i} N_i^*}{\hat{u}} = \frac{k_0}{\hat{u}} - \frac{\Sigma_0}{\hat{u}} \tag{50}$$

where

$$\hat{u} = u - \phi\sigma^2\gamma v \tag{51}$$

and we have defined

$$\Sigma_0 = \frac{\sigma}{\sqrt{s}} \sum_i a_{0i} N_i^* \tag{52}$$

We can see in N_0 two main contributions: the effective carrying capacity k_0, and a fitness shift Σ_0 arising only from to the *heterogeneity* of interspecies interactions. The former is obviously model-dependent (we have no reason to constrain the distribution of carrying capacities), but the latter becomes universal in the limit of many interactions. This relies on the sum $\sum_i a_{0i} N_i^*$ being effectively a random walk.

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Here, we are going to assume that interactions between species are not correlated *in the pool* (we discuss a weaker assumption that is still tractable in Sec 7.4). This lack of correlation has two consequences. First, note that a_{0i} is uncorrelated with N_i^*: the abundance of species i *before species* 0 *was introduced* depends on a_{ji} for $j \neq 0$, but bears no relationship to a_{0i}. More importantly, a_{0i} and a_{0j} are uncorrelated, which means that all terms in the sum above are uncorrelated variables, and if they have finite variance, then the sum must tend to a normal law by the Central Limit Theorem. This normal law has mean 0 and variance $s\phi \left\langle N^2 \right\rangle^*$, where the mean is taken here over $P(N^*)$ the equilibrium distribution of values for N_i^*.

Hence, for any model where interactions have finite variance, the final abundance distribution will be composed from the model-specific distribution of carrying capacities, and from a generic Gaussian term emerging from interactions[11]. And even if $P(N^*)$ has a model-specific shape due to $P(K)$, the interaction term, being Gaussian, is always entirely parametrized by the first two moments of $P(N^*)$, meaning that one never needs a lot of information on $P(K)$ to correctly evaluate the effect of interactions.

This entire reasoning however weakens if $P(K)$ and therefore $P(N)$ have fat tails; any other structure (e.g. multimodal $P(K)$) will somewhat increase the necessary number of species and interactions to converge toward the Gaussian distribution of σ_0, but fat tails may prevent this convergence altogether (if variance is formally infinite) or make it dramatically slow. This possibility would call for an extreme value calculation instead: what would matter in Σ_0 would not be its variance anymore, but its extreme values, which obey a different but still universal distribution.

5.3 Abundance distribution

Equation (50) allows for $N_0 < 0$, hence, N_0 is a "virtual abundance" that is equal to the real abundance of some species if $N_0 > 0$, or signals that the species goes extinct if $N_0 \leq 0$. We can compute the equilibrium abundance distribution by finding out $P_0(N_0)$ the distribution of N_0 as a function of that of N_i^*, then require that the two distributions are equal for positive values.

We can now compute the first moments of the distribution of N_0 as a function of those of N_i^*. Taking the mean of both sides in (50), we find

$$\langle k_0 \rangle = \langle K \rangle - \phi\mu \langle N \rangle^*, \qquad \langle \Sigma_0 \rangle = 0 \tag{53}$$

(where we used $\langle N \rangle = \phi \langle N \rangle^*$ since when we introduced $\langle N \rangle$ earlier, we took the mean over all S species, including those that are dead) hence

$$\langle N_0 \rangle = \frac{\langle k_0 \rangle}{\hat{u}}, \tag{54}$$

while taking the variance, we find

$$\mathrm{Var}(k_0) = \zeta^2, \qquad \mathrm{Var}(\Sigma_0) = \phi\sigma^2 \left\langle N^2 \right\rangle^* \tag{55}$$

[11] Further clarifications on how randomness intervenes in interactions can be found in Appendix "The role of disorder".

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

hence

$$\sigma_0^2 = \mathrm{Var}(N_0) = \frac{\zeta^2 + \sigma^2 \phi \left\langle N^2 \right\rangle^*}{\hat{u}^2} \tag{56}$$

Then, we approximate $P_0(N_0)$ by a normal law with the mean and variance given above. Hence, we can express $\langle N \rangle^*$, $\left\langle N^2 \right\rangle^*$ and ϕ by restricting $P_0(N_0)$ to its positive values

$$\phi = \int_0^\infty dN_0 \; P_0(N_0) \tag{57}$$

$$\langle N \rangle^* = \frac{1}{\phi} \int_0^\infty dN_0 \; P_0(N_0) \; N_0 \tag{58}$$

$$\left\langle N^2 \right\rangle^* = \frac{1}{\phi} \int_0^\infty dN_0 \; P_0(N_0) \; N_0^2 \tag{59}$$

$$\tag{60}$$

where the $1/\phi$ factor ensures normalization. On the other hand, v is given by taking the derivative (43) in (50)

$$v = \langle v_{00} \rangle = \left\langle \frac{dN_0}{dk_0} \right\rangle = \frac{1}{u - \phi \sigma^2 \gamma v} \tag{61}$$

We now have a closed set of four equations for four unknowns: v, ϕ, $\langle N \rangle^*$, $\left\langle N^2 \right\rangle^*$ which appear on the left-hand side of each equation and in $P_0(N_0)$. These equations can be solved together numerically. It is then immediate to compute quantities such as total biomass

$$N_{tot} = S \phi \langle N \rangle^* \tag{62}$$

and Simpson index

$$I = \sum_i \left(\frac{N_i}{N_{tot}} \right)^2 = \frac{\left\langle N^2 \right\rangle^*}{S \phi (\langle N \rangle^*)^2} \tag{63}$$

More generally,

$$P_0(N_0) = \int P(k_0, a_{0i}, N_i) \; \delta \left(N_0 - \frac{1}{u - \gamma v} \left(k_0 - \frac{\sigma}{\sqrt{s}} \sum_{i=1}^{s} a_{0i} N_i \right) \right) \tag{64}$$

The equilibrium solution is obtained by solving this equation under the constraint that $P(N^*) \propto P_0(N_0)$. If $P(K_i)$ is normal, we get that $P_0(N_0)$ is also exactly normal in the limit of many interactions (large s)

$$P_0(N_0) = \frac{e^{-(N_0 - \langle N_0 \rangle)^2 / \sigma_0^2}}{\sqrt{2\pi} \sigma_0}, \tag{65}$$

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

for any reasonable distribution $P(a_{0i})$. Then, we restrict it to positive values to get the real abundance distribution at equilibrium

$$P(N^*) = \frac{P_0(N_0)}{\phi}, \qquad N^* > 0 \tag{66}$$

$$P(N^*) = 0, \qquad N^* \leq 0 \tag{67}$$

If $P(K_i)$ is not normal but has thinner tails (e.g. a uniform distribution), $P(N^*)$ will still be very close to normal, as its tails will come from the interaction term. If however $P(K_i)$ is exponential for instance, $P(N^*)$ inherits its longer tail. This may result in the moment method giving a poorer approximation to all computed quantities, as seen below in Sec. 6 – in that case, we instead need to integrate (64) numerically, which can easily be done in Fourier transform.

The previous calculation assumed K_i and A_{ij} to be independent variables. If there is a correlation between them, then

$$\sigma_0^2 = \frac{\zeta^2 + \sigma_N^2 - 2\phi \langle k_0 a_{0i} \rangle}{(u - \sigma\gamma v)^2} \tag{68}$$

where

$$\langle k_0 a_{0i} \rangle = \langle K_0 A_{0i} \rangle - \frac{\mu}{S}. \tag{69}$$

Hence, the correlation can be accounted for by a shift in the effective value of ζ.

The fact that the abundance distribution found above is a truncated Gaussian contradicts much evidence on empirical abundance distributions being fat-tailed – possibly log-normal or even power-law. No fault lies with the mathematical analysis, as it is in perfect quantitative agreement with simulations, see [4]. Hence the issue will occur for any system with Lotka-Volterra dynamics (1) with random interactions and without noise.

Still, this discrepancy is not a definite mark against random Lotka-Volterra models. There are two avenues to addressing it: first, the addition of noise, which is a central ingredient in neutral models that successfully predict abundance distributions. Second, the role of spatial aggregation in the data, as it is well known that aggregating over Gaussian distributions with different variances can create the appearance of more complex distributions and even fat tails - indeed, this is seen in an extension of this work to metacommunities (to be published).

5.4 Carrying capacities

We have seen above that equilibrium abundances are made up of two independent contributions: one coming from the effective carrying capacity, the other a shift in abundance due to the heterogeneity of interactions

$$N_0 = \frac{k_0}{\hat{u}} - \frac{\Sigma_0}{\hat{u}}$$

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

We also recall that

$$k_0 = K_0 - \mu \langle N \rangle$$

If we want to know the equilibrium distribution of carrying capacities K_i^*, we start from the joint distribution of abundance N_0 and effective carrying capacity k_0, which we will then restrict only to surviving species $N_0 > 0$.

$$P(N_0, k_0)\, dN_0\, dk_0 = P(\Sigma_0, k_0)\, d\Sigma_0\, dk_0 \tag{70}$$

hence

$$P(N_0, k_0) = \frac{d\Sigma_0}{dN_0} P(\Sigma_0, k_0) = \hat{u} P(\Sigma_0, k_0) \tag{71}$$

Now, since Σ_0 and k_0 are independent, $P(\Sigma_0, k_0) = P(\Sigma_0)P(k_0)$, so

$$P(N_0, k_0) = \hat{u} P(\Sigma_0 = k_0 - \hat{u} N_0) P(k_0) \tag{72}$$

We argued above that $P(\Sigma_0)$ is Gaussian with mean 0 and variance σ_Σ^2

$$P(\Sigma_0) = \mathcal{N}(\Sigma_0, 0, \sigma_\Sigma^2) \tag{73}$$

with

$$\sigma_\Sigma^2 = \sigma^2 \phi \left\langle N^2 \right\rangle^* \tag{74}$$

Now, we want to isolate $P_0(N_0)$ within $P(N_0, k_0)$, so that the integral over $N_0 > 0$ causes the moments we have already computed, ϕ, $\langle N \rangle$ and $\left\langle N^2 \right\rangle$, to appear. We recall

$$P_0(N_0) = \frac{e^{-(\hat{u} N_0 - \langle k_0 \rangle)^2 / (\zeta^2 + \sigma_\Sigma^2)}}{\sqrt{2\pi}\sigma_0}, \qquad \sigma_0^2 = \frac{\zeta^2 + \sigma^2 \phi \left\langle N^2 \right\rangle^*}{\hat{u}^2} = \frac{\zeta^2 + \sigma_\Sigma^2}{\hat{u}^2}$$

With some manipulations (assuming a normal distribution for K_i), we can see that

$$P(N_0, k_0) = P_0(N_0) \mathcal{N}\left(k_0, \frac{\zeta^2 \hat{u} N_0 + \sigma_\Sigma^2 \langle k_0 \rangle}{\zeta^2 + \sigma_\Sigma^2}, \frac{\zeta^2 \sigma_\Sigma^2}{\zeta^2 + \sigma_\Sigma^2} \right). \tag{75}$$

In particular, we find the averages over surviving species

$$\begin{aligned} \langle K_i \rangle^* &= \int_0^\infty dN_0\, P_0(N_0) \int_{-\infty}^\infty dk_0\, \frac{P(N_0, k_0)}{P_0(N_0)} \left(k_0 + \mu \langle N \rangle \right) \\ &= \langle K \rangle + \frac{(\hat{u} + \mu) \langle N \rangle + \langle K \rangle}{1 + \sigma^2 \langle N^2 \rangle / \zeta^2} \end{aligned} \tag{76}$$

and

$$\begin{aligned} \langle K_i N_i \rangle^* &= \int_0^\infty dN_0\, P_0(N_0)\, N_0 \int_{-\infty}^\infty dk_0\, \frac{P(N_0, k_0)}{P_0(N_0)} \left(k_0 + \mu \langle N \rangle \right) \\ &= \frac{\hat{u} \left\langle N^2 \right\rangle + \mu \langle N \rangle^2 + \frac{\sigma^2}{\zeta^2} \left\langle N^2 \right\rangle \langle K \rangle \langle N \rangle}{1 + \frac{\sigma^2}{\zeta^2} \langle N^2 \rangle} \end{aligned} \tag{77}$$

where we recall $\langle N \rangle = \phi \langle N \rangle^*$, $\left\langle N^2 \right\rangle = \phi \left\langle N^2 \right\rangle^*$.

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

5.5 Productivity

Given the abundance distribution $P(N_i)$, it is possible to compute the productivity distribution – or at least the distribution of growth on external resources,

$$g_i = R_i N_i \tag{78}$$

which acts as a proxy for productivity. If the quantities that are drawn independently are K_i, α_{ij} and D_i, then straightforwardly

$$P(g_i) = \int dN_i \; P(g_i, N_i) \tag{79}$$

$$= \int_0^\infty dN_i dR_i \; P(R_i, N_i) \; \delta(R_i - g_i/N_i) \tag{80}$$

$$= \int_0^\infty dN_i dK_i dD_i \; P(K_i, N_i) \; P(D_i) \; \delta(K_i D_i - g_i/N_i) \tag{81}$$

and we get $P(K_i, N_i)$ from the previous section. In particular, taking the average over surviving species, we find (if all $K_i > 0$)

$$\langle g_i \rangle^* = \int dg_i \; P(g_i) \; g_i = \langle K_i N_i \rangle^* \tag{82}$$

where we used $\langle D_i \rangle = 1$. If however the independent quantities are R_i, A_{ij} and D_i, the calculation is less straightforward: we must now start from

$$N_0 = \frac{1}{D_0 \hat{u}} \left(R_0 - S^* A \langle N \rangle^* - \hat{\Sigma}_0 \right)$$

where

$$\hat{\Sigma}_0 = \sum_i (A_{0i} - A) N_i^*, \qquad \left\langle \hat{\Sigma}_0 \right\rangle = 0, \qquad \mathrm{Var}\left(\hat{\Sigma}_0\right) = S^* \sigma_A^2 \left\langle N^2 \right\rangle^* \tag{83}$$

and we now have $P(R_0, \hat{\Sigma}_0, D_0) = P(R_0) \; P(\hat{\Sigma}_0) \; P(D_0)$, hence

$$P(N_0|R_0) = \int d\hat{\Sigma}_0 dD_0 \; P(\hat{\Sigma}_0) \; P(D_0) \; \delta \left(N_0 - \frac{1}{D_0 \hat{u}} \left(R_0 - S^* A \langle N \rangle^* - \hat{\Sigma}_0 \right) \right) \tag{84}$$

and we get the total productivity

$$G_{tot} = S \int_0^\infty dN_0 \int_0^\infty dR_0 \; R_0 N_0 \; P(N_0|R_0) \; P(R_0) \tag{85}$$

5.6 Response to perturbations

Let us now assume that some small permanent (press) perturbation ξ_i is added to the system at equilibrium (acting only on the surviving species)

$$0 = k_i + \xi_i - u N_i - \frac{\sigma}{\sqrt{s}} \sum_j a_{ij} N_j \tag{86}$$

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

We can write $N_i = N_i^* + x_i$, with x_i the displacement around equilibrium, which obeys

$$0 = \xi_i - \mu \langle x \rangle - u x_i - \frac{\sigma}{\sqrt{s}} \sum_j a_{ij} x_j \tag{87}$$

a very similar equation to that obeyed by equilibrium abundances, except for the fact that the carrying capacity K_i has been replaced by the perturbation ξ_i, and of course that x_i is not constrained to be positive. Following the same logic as before, we isolate the response of species 0

$$(u - \phi\sigma^2\gamma v) x_0 = \xi_0 - \mu \langle x \rangle - \frac{\sigma}{\sqrt{s}} \sum_j a_{ij} x_j^* \tag{88}$$

If there are no constraints on x_0 (see Sec. 5.6.2 below), we can simply assume

$$\langle x_0 \rangle = \langle x \rangle = \langle x_i^* \rangle \tag{89}$$

and thus get the average displacement:

$$(u - \phi\sigma^2\gamma v + \mu) \langle x \rangle = \langle \xi \rangle \tag{90}$$

5.6.1 Multistability order parameter

To test whether the equilibrium we are obtaining as the fixed point of the invasion process is globally stable, or might be exited through some small perturbation in the appropriate direction, we must consider the response to a random perturbation with zero mean,

$$\langle \xi \rangle = 0 \tag{91}$$

representing random shifts in relative fitness that advantage some species over others, without changing the mean species fitness. Since the mean is zero, the response to the perturbation is characterized by the variance,

$$\chi^2 = \left\langle \frac{x_i^2}{\xi^2} \right\rangle \tag{92}$$

Now the equation above, using $\langle x \rangle = 0$, can be taken to the square to get:

$$(u - \phi\sigma^2\gamma v)^2 \left\langle x_0^2 \right\rangle = \xi_0^2 + \phi\sigma^2 \left\langle x_j^* \right\rangle^2 \tag{93}$$

hence

$$\chi^2 = \frac{1}{(u - \phi\sigma^2\gamma v)^2 - \phi\sigma^2} \tag{94}$$

which is seen to diverge when $\sigma = \sigma_c$ such that

$$\phi\sigma_c = u - \phi\sigma_c^2\gamma v \tag{95}$$

and by using $(u - \phi\sigma^2\gamma v)v = 1$ we find as above in (12)

$$\sigma_c = \frac{1}{\phi} \left(\frac{u}{1+\gamma} \right)^2 \tag{96}$$

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

5.6.2 Secondary extinctions

In the case of secondary extinctions following the extinction of species 0, the perturbation is

$$\xi_i = \frac{\sigma}{\sqrt{s}} a_{i0} N_0^* \tag{97}$$

Now however, the perturbation is potentially large enough that we may need to impose $x_i > -N_i^*$, meaning that the response cannot cause a species' abundance to become negative. Hence, in a way similar to our previous calculation on abundances, the average change in abundance due to the extinction is

$$\langle x \rangle = \int dN_i P(N_i) \int_{-N_i}^{\infty} dx_0 P_0(x_0) x_0 \tag{98}$$

while the fraction of secondary extinctions is given by

$$\Delta\phi = \int dN_i P(N_i) \int_{-\infty}^{-N_i} dx_0 P_0(x_0) \tag{99}$$

5.6.3 Variability

Calculations for the variability are more technical and left in Appendix.

6 Sensitivity testing

6.1 Distributions

In this section, we test how sensitive the results above are to the choice of distributions for growth rates and interactions. Throughout, we have assumed that these distributions had finite mean and variance, which is not the case for, e.g., a power-law tailed distribution $P(x) \sim x^{-a}$ with exponent $a \leq 3$.

For reasons explained in Sec. 5.2 above, we expect results to be more sensitive to the distribution of growth rates than to that of interactions. Indeed, Fig. 7 shows that, while normal or uniform distributions lead to very good agreement with predictions, distributions with fatter tails (e.g. exponential tails) – and a fortiori multimodal distributions – lead to significant deviations from our baseline predictions. It is however not difficult to adapt the theory to fit any distribution $P(R_i)$ as long as it is known.

6.2 Interaction strength

We now test how small the interaction strength must be compared to self-interaction for cavity predictions to work. Fig. 8 shows that when average coupling strength is significantly larger than 0.1, the method stops being as accurate, in particular with fewer species surviving than expected.

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

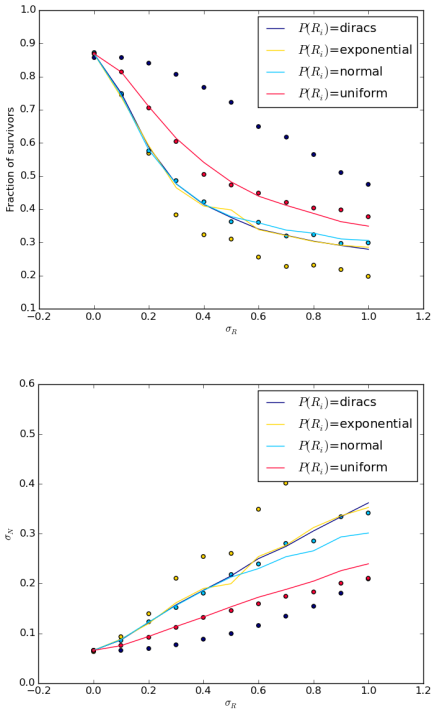

Figure 7: Sensitivity of results to distribution of growth rates $P(R_i)$.

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

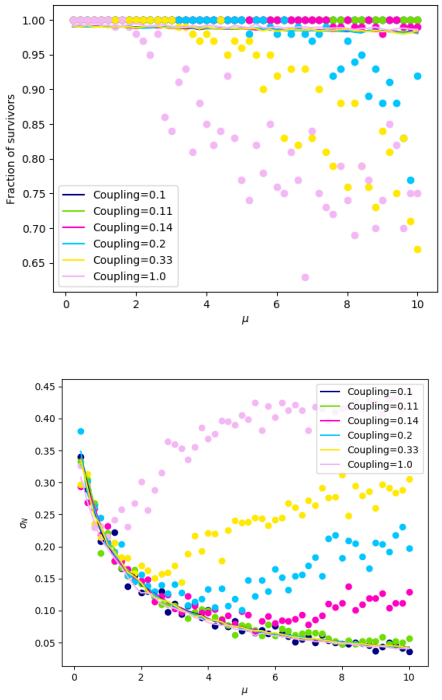

Figure 8: Sensitivity of results to interaction strength. By varying the connectivity from 0.1 to 1, we change the average interaction strength needed to achieve the same value of μ. Each curve is labelled here by the average coupling strength needed to obtain $\mu = 10$.

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

7 Extending the model

There are two main types of limitations to the model so far. The first is dynamical: we started from Lotka-Volterra dynamics, which might work as an approximation in many limits, but do not always enable us to find the equilibria reached by more complex dynamics. Since the most usual variant to Lotka-Volterra dynamics is having a nonlinear functional response, especially a saturating one, we treat that case below to show that it is still largely amenable (at least approximately or qualitatively) to our approach.

The second limitation is structural. Even though species are heterogeneous, we have treated them as if they all belong to the same functional group: their traits are all drawn from the same distributions. These distributions may be multimodal, but that does not convey the fact that different positions in the community may be truly different in many respects at once. There are two simple ways to address this concern, one discrete and one continuous.

The continuous approximation, which we call "ordered species", supposes that we can characterize the network by a simple axis of possible niches or roles, knowing only how various traits (e.g. carrying capacity, interaction mean and variance) correlate to linear order with position on that axis.

The discrete approximation, which we call "species groups", instead divides the network into distinct groups, each disordered inside, but with deterministic patterns within and between groups. In essence, this is extending simple reductionist descriptions (e.g. one predator and one prey) by replacing each species by a group of heterogeneous but functionally similar species, treated like a disordered community.

7.1 Functional response

The coefficients A_{ij} come from two distinct factors: first, how much an *individual* of species i can affect an individual of species j, which may be a property of their behavior and biology, and second, how often these individuals may meet and interact. For simplicity, let us say that the individual-level effects are fixed. Then, the frequency of pair encounters is what controls interaction strength. The interaction between species i and j is independent from any other interaction in relatively "sparse" environments where any pair can meet, and adding new individuals or new species always increases the number of such meetings. On the other hand, interactions may saturate if there is a limiting factor such as plants competing only with their neighbors, or a predator taking time to handle prey, preventing it from consuming more than a given number of prey among those available in a given timespan.

The same reasoning led Holling [13] to postulate a saturating functional response in the interaction between two species: instead of being a number, A_{ij} becomes a function of abundances such as

$$\tilde{A}_{ij}(N_1, ...N_S)\, N_i N_j = \frac{A_{ij} N_i N_j}{1 + \frac{1}{\langle A \rangle N_c} \sum_k A_{ik} N_k} \tag{100}$$

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

with constants A_{ij}, their average $\langle A \rangle$ and the population threshold N_c. This means that the interaction between species i and j cannot keep growing indefinitely with $N_i N_j$ the product of their abundances, as in Lotka-Volterra equations. Hence, the sum is bounded

$$N_i \sum_j \tilde{A}_{ij} N_j = N_i \frac{\sum_k A_{ik} N_k}{1 + \frac{1}{AN_c} \sum_k A_{ik} N_k} \tag{101}$$

$$\approx A N_i N_c \qquad \text{if } \left| \sum_k A_{ik} N_k \right| \gg |AN_c| \tag{102}$$

Analytically, it is possible to deal with the saturating functional response by a simple approximation. For each species, we replace it with a piecewise linear function:

$$\sum_j \tilde{A}_{ij} N_j \approx F\left(\sum_k A_{ik} N_k \right), \qquad F(x) = \begin{cases} x & |x| < |AN_c| \\ AN_c & |x| \geq |AN_c| \end{cases} \tag{103}$$

then at the community level, we use the real saturating function to evaluate the expected fraction f of species whose interactions are saturated.

For simplicity, we will "cheat" and instead apply the piecewise functional response to the *couplings* (which differ from interactions if species have heterogeneous self-interaction D_i), i.e.

$$0 = K_0 - N_0 - \tilde{F}\left(\sum_i \alpha_{0i} N_i \right), \qquad \tilde{F}(x) = \begin{cases} x & |x| < |\mu N_c / s| \\ \mu N_c / s & |x| \geq |\mu N_c / s| \end{cases} \tag{104}$$

Avoiding this simplification is possible but requires keeping track of more parameters, and does not make much of a qualitative change.

Going through the same steps as before, then for each species

$$N_0 = \frac{1}{u - (1-f)\phi\sigma^2\gamma v} \begin{cases} K_0 - \mu\phi \langle N \rangle^* - \frac{\sigma}{\sqrt{s}} \sum_i a_{0i} N_i^* & \text{with probability } 1-f \\ K_0 - \mu N_c / s & \text{with probability } f \end{cases} \tag{105}$$

Notice the $1-f$ factor in front of the feedback term in the denominator: only species whose interactions have not saturated will contribute to the feedback of the invader on itself through its interaction partners. The mean is now

$$\langle N_0 \rangle = \frac{\langle K \rangle - \mu(f N_c / s + (1-f)\phi \langle N \rangle^*)}{u - (1-f)\phi\sigma^2\gamma v} \tag{106}$$

while taking the variance, we find

$$\sigma_0^2 = \frac{\zeta^2 + (1-f)^2 \sigma^2 \phi \langle N^2 \rangle^*}{(u - (1-f)\phi\sigma^2\gamma v)^2} \tag{107}$$

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

The rest of the calculations follow as in Sec. 5.3, except we additionally need to know the fraction f of species with saturated interactions. This is where the real saturation function reappears (slightly rearranged):

$$f = \int_{-\infty}^{\infty} dx P\left(x = \sum_k \alpha_{0k} N_k\right) \frac{1}{1 + \left|\frac{\mu N_c}{sx}\right|} \tag{108}$$

where, as usual, we can approximate $P(x)$ by a normal distribution with

$$\langle x \rangle = |\mu| \phi \tag{109}$$

$$\mathrm{Var}(x) = \sigma^2 \phi \left\langle N^2 \right\rangle^* \tag{110}$$

Note that if we want to make an equivalent model without saturation, we can apply the equivalence

$$\tilde{\mu} = \mu \left(f \frac{N_c}{s \langle N \rangle^*} + (1-f)\phi \right), \qquad \tilde{\sigma} = (1-f)\sigma, \qquad \tilde{\gamma} = \frac{\gamma}{1-f} \tag{111}$$

which is possible only if $\tilde{\gamma} \in [-1, -1]$. Then, a Lotka-Volterra system with parameters $\tilde{\mu}$, $\tilde{\sigma}$ and $\tilde{\gamma}$ will have identical equilibrium properties to the system with a saturating functional response studied here. Of course, this equivalence can only be computed once we know f and $\langle N \rangle^*$, i.e. after having solved the calculation above for this system.

In Fig. 9, we show that predictions from our calculations match simulation results quite well, despite many simplifications which tend to create a more abrupt transition between the saturated and unsaturated regimes. By the same reasoning, we can easily extend the approach to any functional response, including an approximation of Type-III sigmoid response, with two thresholds instead of a single one.

7.2 Heterogeneous means

Let us come back to the cavity method:

$$0 = K_0 - N_0 - \sum_i^{S^*} A_{0i} N_i \tag{112}$$

Now let us assume that

$$A_{0i} = \langle A \rangle_0 + \frac{\sigma_0}{\sqrt{S}} a_{0i} \tag{113}$$

where the mean and the variance both depend on species 0.

$$0 = K_0 - S^* \langle A \rangle_0 \langle N \rangle^* - N_0 - \frac{\sigma_0}{\sqrt{S}} \sum_i^{S^*} a_{0i} N_i \tag{114}$$

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

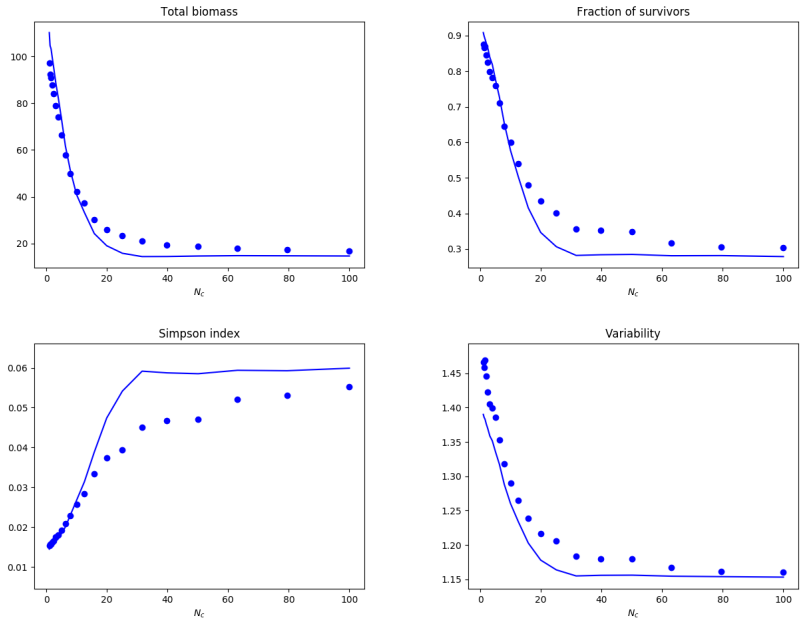

Figure 9: Effect of saturating functional response. Symbols are simulations with a Holling Type II functional response, while solid lines are analytical predictions using the approximate linear-saturating functional response (104).

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

where we still assume that $\langle N \rangle$ over partners of 0 does not differ from the community average. Then,

$$N_i \approx N_i^* - \frac{dN_i}{dK_i}\frac{\sigma_i}{\sqrt{S}} a_{i0} N_0 \tag{115}$$

hence

$$\left(1 - \frac{1}{S}\sum_i \sigma_i \sigma_0 a_{i0} a_{0i} \frac{dN_i}{dK_i}\right) N_0 = K_0 - S^* \langle A \rangle_0 \langle N \rangle^* - \frac{\sigma_0}{\sqrt{S}} \sum_i^{S^*} a_{0i} N_i^* \tag{116}$$

$$\approx (1 - \phi \sigma_0 \sigma \gamma v) N_0 \tag{117}$$

where

$$v = \left\langle \frac{1}{1 - \phi \sigma_0 \sigma \gamma v} \right\rangle \approx \frac{1}{1 - \phi \sigma^2 \gamma v} \tag{118}$$

and finally

$$\langle N_0 \rangle = v(\langle K \rangle - \phi \mu \langle N \rangle^*) \tag{119}$$

$$\langle N_0^2 \rangle = \langle N_0 \rangle^2 + v^2(\sigma_K^2 + S^{*2}(\sigma_{\langle A \rangle}^2) \langle N \rangle^{*2} - 2S^*(\langle K_0 A_0 \rangle - \langle K \rangle \langle A \rangle) + \sigma^2 \phi \langle N^2 \rangle^*) \tag{120}$$

7.3 Species groups

The equations can be extended to any structure comprised of discrete groups, with disordered interactions within and between groups, but different statistics for each set of interactions. Coming back to the equilibrium equation, we can write for species i in group x (which contains S^x species) as

$$0 = K_i^x - N_i^x - \sum_{j,y} \alpha_{ij}^{xy} N_j^y \tag{121}$$

Thus, we now have vector $\boldsymbol{\zeta}$ and matrices $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$, $\boldsymbol{\gamma}$, defined by

$$\alpha_{ij}^{xy} = \frac{\mu^{xy}}{S} + \frac{\sigma^{xy}}{\sqrt{S}} a_{ij}^{xy} \tag{122}$$

with

$$\langle a \rangle = 0, \qquad \langle a^2 \rangle = 1, \qquad \langle a_{ij}^{xy} a_{ji}^{yx} \rangle = \gamma^{xy}. \tag{123}$$

The equations to solve are the same as above, except there are now four equations per group, all coupled: for each group we solve (dropping the * for concision)

$$\phi^x = \int_0^\infty dN_0 \; P_0^x(N_0) \tag{124}$$

$$\langle N \rangle^x = \frac{1}{\phi^x} \int_0^\infty dN_0 \; P_0^x(N_0) \; N_0 \tag{125}$$

$$\langle N^2 \rangle^x = \frac{1}{\phi} \int_0^\infty dN_0 \; P_0^x(N_0) \; N_0^2 \tag{126}$$

$$v^x = \frac{1}{\tilde{u}^x} \tag{127}$$

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

with

$$\langle N_0 \rangle^x = \frac{\langle K \rangle^x - \sum_y \mu^{xy} \varphi^y \langle N \rangle^y}{u^x}, \tag{128}$$

$$(\sigma_0^2)^x = \frac{\zeta^2 + \sum_y (\sigma^{xy})^2 \varphi \langle N^2 \rangle^y}{(u^x)^2}, \tag{129}$$

$$\tilde{u}^x = 1 - \sum_y \varphi^y v^y \gamma^{xy} \sigma^{xy} \sigma^{yx}. \tag{130}$$

where

$$\varphi^x = \phi^x \frac{S^x}{S}. \tag{131}$$

7.4 Ordered species

7.4.1 Linear order: competition-colonization tradeoff

We can take the continuous limit of the group approach. Let us assume that interactions depend only on rank difference:

$$\mu(x, y) = \mu + \tilde{c}_\mu(x - y) \tag{132}$$

$$\sigma^2(x, y) = \sigma^2 + \tilde{c}_\sigma(x - y) \tag{133}$$

$$\gamma(x, y) = \gamma + \tilde{c}_\gamma(x - y) \tag{134}$$

$$\langle K \rangle (x) = K + \tilde{c}_K(x) \tag{135}$$

If furthermore we choose linear functions $\tilde{c}_*(z) = c_* z$ we have

$$\langle N_0 \rangle (x) = \frac{K + c_K x - (\mu + c_\mu x) \langle \varphi N \rangle + c_\mu \langle \varphi N x \rangle}{\tilde{u}(x)}, \tag{136}$$

$$\sigma_0^2(x) = \frac{\zeta^2 + (\sigma + c_\sigma x) \langle \varphi N^2 \rangle - c_\sigma \langle \varphi N^2 x \rangle}{\tilde{u}^2(x)}, \tag{137}$$

$$\tilde{u}(x) = 1 - ((\gamma + c_\gamma x) \langle \varphi v \rangle - c_\gamma \langle \varphi v x \rangle)\sigma^2. \tag{138}$$

Hence there is an additional set of unknowns: for $\Psi = N, N^2, v$ we must compute

$$\langle \varphi \Psi \rangle = \int dx P(x) \int_0^\infty dN P_0(N, x) \Psi \tag{139}$$

and

$$\langle \varphi \Psi x \rangle = \int dx P(x) x \int_0^\infty dN P_0(N, x) \Psi \tag{140}$$

where $P(x)$ is the probability distribution of x.

In the competition-colonization example, we have

$$\mu(x, y) = \mu \frac{y - x + 1}{2} \tag{141}$$

$$\langle K \rangle (x) = 2K(1 - x) \tag{142}$$

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

7.4.2 Exponential order: predation

If on the other hand we choose exponential functions $\tilde{c}_*(z) = c_* e^z$ we have

$$\langle N_0 \rangle (x) = \frac{K + c_K x - \mu \langle \varphi N \rangle + c_\mu e^x \langle \varphi N e^{-x} \rangle}{\tilde{u}(x)}, \tag{143}$$

$$\sigma_0^2(x) = \frac{\zeta^2 + \sigma \langle \varphi N^2 \rangle - c_\sigma e^x \langle \varphi N^2 e^{-x} \rangle}{\tilde{u}^2(x)}, \tag{144}$$

$$\tilde{u}(x) = 1 - (\gamma \langle \varphi v \rangle - c_\gamma \langle \varphi v e^x \rangle)\sigma^2. \tag{145}$$

References

[1] Stefano Allesina and Si Tang. Stability criteria for complex ecosystems. *Nature*, 483(7388):205–208, 2012.

[2] Jean-François Arnoldi, Michel Loreau, and Bart Haegeman. Resilience, reactivity and variability: A mathematical comparison of ecological stability measures. *Journal of theoretical biology*, 389:47–59, 2016.

[3] Matthieu Barbier, Jeff Arnoldi, and Guy Bunin. Generic assembly patterns in large ecological communities. *bioRxiv preprint 145862*, 2017.

[4] Matthieu Barbier, Jeff Arnoldi, Michel Loreau, and Guy Bunin. Generic assembly patterns in large ecological communities. 2017.

[5] Jordi Bascompte and Daniel B Stouffer. The assembly and disassembly of ecological networks. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 364(1524):1781–1787, 2009.

[6] Guy Bunin. Interaction patterns and diversity in assembled ecological communities. *arXiv preprint arXiv:1607.04734*, 2016.

[7] Amar Gandhi, Simon Levin, and Steven Orszag. Moment expansions in spatial ecological models and moment closure through gaussian approximation. *Bulletin of mathematical biology*, 62(4):595–632, 2000.

[8] Jianxi Gao, Baruch Barzel, and Albert-László Barabási. Universal resilience patterns in complex networks. *Nature*, 530(7590):307–312, 2016.

[9] M Gatto. A general minimum principle for competing populations: Some ecological and evolutionary consequences. *Theoretical Population Biology*, 37(3):369–388, 1990.

[10] Dominique Gravel, François Massol, Elsa Canard, David Mouillot, and Nicolas Mouquet. Trophic theory of island biogeography. *Ecology letters*, 14(10):1010–1016, 2011.

[11] Jacopo Grilli, Tim Rogers, and Stefano Allesina. Modularity and stability in ecological communities. *Nature communications*, 7, 2016.

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

[12] J Nathaniel Holland, Donald L DeAngelis, and Judith L Bronstein. Population dynamics and mutualism: functional responses of benefits and costs. *The American Naturalist*, 159(3):231–244, 2002.

[13] Crawford Stanley Holling. The functional response of predators to prey density and its role in mimicry and population regulation. *Memoirs of the Entomological Society of Canada*, 97(S45):5–60, 1965.

[14] Robert D Holt. Community modules. In *Multitrophic interactions in terrestrial ecosystems, 36th Symposium of the British Ecological Society*, pages 333–349. Blackwell Science Oxford, 1997.

[15] Stephen P Hubbell. *The unified neutral theory of species abundance and diversity*. Princeton University Press, Princeton, 2001.

[16] David A Kessler and Nadav M Shnerb. Generalized model of island biodiversity. *Physical Review E*, 91(4):042705, 2015.

[17] Richard Levins and David Culver. Regional coexistence of species and competition between rare species. *Proceedings of the National Academy of Sciences*, 68(6):1246–1248, 1971.

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

[19] Robert Mac Arthur. Species packing, and what competition minimizes. *Proceedings of the National Academy of Sciences*, 64(4):1369–1371, 1969.

[20] Robert H MacArthur and Edward O Wilson. *Theory of Island Biogeography*. Princeton University Press, 2015.

[21] Ramon Margalef. Diversity, stability and maturity in natural ecosystems. In *Unifying concepts in ecology*, pages 151–160. Springer, 1975.

[22] Robert M May. Will a large complex system be stable? *Nature*, 238:413–414, 1972.

[23] Robert M May and Roy M Anderson. Parasite—host coevolution. *Parasitology*, 100(S1):S89–S101, 1990.

[24] M Mézard. The space of interactions in neural networks: Gardner's computation with the cavity method. *Journal of Physics A: Mathematical and General*, 22(12):2181, 1989.

[25] Marc Mézard and Andrea Montanari. *Information, physics, and computation*. Oxford University Press, 2009.

[26] Marc Mézard, Giorgio Parisi, and Miguel Virasoro. *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications*, volume 9. World Scientific Publishing Co Inc, 1987.

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

[27] Andrea Montanari, M Müller, and Marc Mézard. Phase diagram of random heteropolymers. *Physical review letters*, 92(18):185509, 2004.

[28] Nicolas Mouquet and Michel Loreau. Community patterns in source-sink metacommunities. *The american naturalist*, 162(5):544–557, 2003.

[29] Pierre Paga and Reimer Kühn. Contagion in an interacting economy. *Journal of Statistical Mechanics: Theory and Experiment*, 2015(3):P03008, 2015.

[30] Alessandro Seganti, Andrea De Martino, and Federico Ricci-Tersenghi. Boolean constraint satisfaction problems for reaction networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2013(09):P09009, 2013.

[31] Torsten Thomas, Lucas Moitinho-Silva, Miguel Lurgi, Johannes R Björk, Cole Easson, Carmen Astudillo-García, Julie B Olson, Patrick M Erwin, Susanna López-Legentil, Heidi Luter, et al. Diversity, structure and convergent evolution of the global sponge microbiome. *Nature Communications*, 7, 2016.

[32] Mikhail Tikhonov. Community-level cohesion without cooperation. *Elife*, 5:e15747, 2016.

[33] Mikhail Tikhonov and Remi Monasson. A collective phase in resource competition in a highly diverse ecosystem. *arXiv preprint arXiv:1609.01270*, 2016.

[34] David Tilman. *Resource competition and community structure*. Princeton university press, 1982.

[35] Chengyi Tu, Jacopo Grilli, and Samir Suweis. Universality of resilience patterns in generalized lotka volterra dynamics and beyond. *arXiv preprint arXiv:1606.09630*, 2016.

8 Appendices

A. The role of disorder

Computing the equilibrium state in effect requires inverting the matrix a_{ij} in the equation:

$$0 = k_i - uN_i - \sigma_A \sum_j a_{ij} N_j \tag{146}$$

In principle we could expand this equation, substituting

$$N_j = \frac{k_j - \sigma_A \sum_j a_{jk} N_k}{u} \tag{147}$$

into the sum, and so on recursively, causing all paths $a_{ij}a_{jk}...$ $(i \to j \to k...)$ in the community matrix to appear in the first equation above. The crucial

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

idea of disordered systems methods is that, for any reasonably complex system, these paths soon lose coherence and start to interfere with each other. These interferences are so complex as to produce seemingly-random results.

One way to see it is that each term in the sum is a (possibly very long) product of coefficients a_{kl} and is weakly correlated with other such terms – unless the same coefficients appear again and again in the same combinations in many terms. That can happen in a very structured community, or one with only small loops (the extreme is a collection of independent few-species motifs: in each motif, there are not enough paths for them to interfere and create a seemingly random result). This enables the cavity trick, which retains only the first-order feedback of a species on itself, and for the rest needs only the mean and variance of interactions, as the effect of interactions is essentially random.

Let us see a precise example of these ideas. In the main text, we noted that the feedback of a species on itself through the rest of the network is a self-averaging quantity. Consider the following sum:

$$\sum_{i,j}^{s^*} a_{0i} v_{ij} a_{j0} \tag{148}$$

which occurs in (45) (we recall $s^* = \phi s$ is the average number of surviving interaction partners for species 0). We separate the sum into two parts:

$$\sum_{i}^{s^*} a_{0i} v_{ii} + \sum_{i \neq j}^{s^*} a_{0i} v_{ij} a_{j0} \tag{149}$$

Recall that $\langle a_{0i} a_{i0} \rangle = \gamma$ while $\langle a_{0i} \rangle = 0$. Hence, the second sum is a random walk with $s^*(s^* - 1)$ terms, and its total scales like s. Now, we should note that

$$v_{ij} = \frac{dN_i}{dk_j}$$

for $i \neq j$ is created by the interaction between species i and j, which is at best proportional to $\sigma/\sqrt{s}$ (if the species are linked directly) or even weaker.

Finally, we get the respective scalings:

$$\sum_{i}^{s^*} a_{0i} a_{i0} v_{ii} \propto s^* \gamma v \tag{150}$$

$$\sum_{i \neq j}^{s^*} a_{0i} v_{ij} a_{j0} \propto s^* \frac{\sigma}{\sqrt{s}} v \tag{151}$$

In other words, even though the sum over $i \neq j$ has many more terms, it ends up being of order $1/\sqrt{s}$ compared to the first sum. That sum itself has variations of order $1/\sqrt{s}$ compared to its mean, depending on which species is isolated as species 0. As s becomes larger, the feedback thus tends to become a single number, no matter which species we consider.

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

B. Linear response

Typical cavity method (Press)

Notation: $a_i^j x_j = \sum_j a_{ij} x_j$. $\langle y_\cdot \rangle = \frac{1}{S}\sum_i y_i$. Let ξ be a random press-perturbation. Its long term effect on the equilibrium associated to the Jacobian B is

$$x = -B^{-1}\xi$$
$$ux_i = \xi_i - \mu \langle x_\cdot \rangle - \sigma a_i^j x_j - \sigma a_{i0} x_0;\ u = 1 - \mu/S$$

Where we isolated the effect of a generic species 0 (this is the starting point of the cavity method). Assuming a small effect of that species we may write

$$x_j = x_{j/0} - \sigma x_0 \mathcal{V}_j^k a_{k0};\ \mathcal{V}_j^k := \frac{\partial x_j}{\partial \xi_k}$$

Under the replica symmetry, species 0 is generic and its addition or removal does not change any statistical properties of community, in particular $\langle x_{\cdot/0} \rangle = \langle x_\cdot \rangle$. Then

$$ux_0 = \xi_0 - \mu \langle x_\cdot \rangle - \sigma a_0^j x_{j/0} + \sigma^2 x_0 \underbrace{\mathcal{V}_j^k a_0^j a_{k0}}_{\approx \gamma \left\langle \mathcal{V}_{\cdot}^{\cdot} \right\rangle}$$

$$\left(u - \sigma^2 \gamma \mathcal{V}\right) x_0 \approx \xi_0 - \mu \langle x_\cdot \rangle - \sigma a_0^j x_{j/0};\ \mathcal{V} = \left\langle \mathcal{V}_{\cdot}^{\cdot} \right\rangle$$

Taking the square

$$\left(u - \sigma^2 \gamma \mathcal{V}\right)^2 x_0^2 \approx \xi_0^2 + \underbrace{\sigma^2 a_0^j a_0^i x_{j/0} x_{i/0}}_{\approx \sigma^2 \langle x_\cdot^2 \rangle} + \mu^2 \langle x_\cdot^2 \rangle + \mu\sigma \langle x_\cdot^2 \rangle a_0^j x_{j/0} + \mathcal{O}(\xi_0)$$

taking the mean over ξ_0, knowing that $\mathbb{E}\xi_i = 0$; $\mathbb{E}\xi_i\xi_j = \delta_{ij}$, and by linearity, $\mathbb{E}_\xi x_i = 0$, gives

$$\left(u - \sigma^2 \gamma \mathcal{V}\right)^2 \mathbb{E} x_0^2 \approx 1 + \sigma^2 \langle \mathbb{E} x_\cdot^2 \rangle$$

Taking the mean over all species gives

$$\chi^2 := \langle \mathbb{E} x_\cdot^2 \rangle \approx \frac{1}{\left(u - \sigma^2 \gamma \mathcal{V}\right)^2 - \sigma^2}$$

To conclude we need $\mathcal{V}$. For that we write

$$\left(u - \sigma^2 \gamma \mathcal{V}\right) \mathcal{V} \approx 1 - \sigma \underbrace{\left\langle a_\cdot^j \mathcal{V}_j^j \right\rangle}_{\approx 0}$$

Hence it is solution to (knowing that at $\sigma = 0$, $\mathcal{V} = u$)

$$\sigma^2 \gamma \mathcal{V}^2 - u\mathcal{V} + 1 = 0$$

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

we thus have

$$\chi^2 \approx \frac{1}{u^2\left(1-v\right)^2-\sigma^2};\ v=\frac{\sigma^2\gamma\mathcal{V}}{u}=\frac{1-\sqrt{1-\frac{4\sigma^2\gamma}{u^2}}}{2}$$

In the Lotka-Volterra assembly perspective, B is the interaction matrix and only a fraction ϕ of species survive. We still have $u=1-\mu/S$ but

$$\chi^2 \approx \frac{1}{u^2\left(1-v\right)^2-\phi\sigma^2}$$

Externally applied press

In the assembly perspective, the Jacobian matrix reads

$$J_{ij}=-\frac{r_i}{k_i}N_i^*B_{ij}$$

where r_i is the growth rate of species i and k_i its carrying capacity. A press acting proportionally to $N_i^{*\nu}$ will induce a displacement

$$x_i=\frac{k_i}{r_i}N_i^{*\nu-1}\zeta_i-a_i^jx_j$$

so we can simply adapt the above calculations by replacing ξ_i by $\frac{k_i}{r_i}N_i^{*\nu-1}\zeta_i$. This gives us

$$\left\langle \mathbb{E}_\zeta x_\cdot^2\right\rangle \approx \chi^2\left\langle (\frac{k}{r})^2N^{*2\nu-2}\right\rangle$$

To compute (mean case) variability we must solve the Lyapunov equation

$$BC+CB^\top=-\mathbb{I}\Leftrightarrow 2C+AC+\left(AC\right)^\top=\mathbb{I}$$

Off diagonal terms, if $\varphi=\frac{S-1}{S}$, read

$$2C_{kj}=-\frac{1-u}{u}\left(C_{kk}+C_{jj}+\varphi\left\langle C_{\cdot\neq k}\right\rangle+\varphi\left\langle C_{\cdot\neq j}\right\rangle\right)-\frac{\sigma}{u}\left(a_k^lC_{lj}+a_j^lC_{lk}\right)$$

taking the mean, the self-averaging of the terms proportional to σ makes the problem reduce to the mean-field case, which gives us the relationship

$$\left\langle C_{\neq}\right\rangle \approx -\frac{S-\mu}{S+\mu(\varphi S-1)}\left\langle C_{=}\right\rangle$$

The interaction diversity (σ) enters on the diagonal (variances):

$$2C_{jj}=1-\sigma a_j^k2C_{kj}-2\mu\varphi\left\langle C_{\cdot\neq j}\right\rangle$$

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

in replacing C_{kj} in $\sigma a_j^k 2C_{kj}$ by its expression written above, we only need to track terms proportional to σ^2, the other will self average. So:

$$2C_{jj} \approx 1 + \frac{\sigma^2}{u} \underbrace{a_j^k a_k^l C_{lj}}_{\approx \gamma C_{jj}} + \frac{\sigma^2}{u} \underbrace{a_j^k a_j^l C_{lk}}_{\approx \langle C_= \rangle} - 2\mu\varphi \langle C_{\cdot \neq j} \rangle$$

$$2\left(1 - \frac{\sigma^2 (\gamma+1)}{2u}\right) \langle C_= \rangle \approx 1 - 2\mu\varphi \langle C_\neq \rangle$$

so that

$$\langle C_= \rangle \approx \mathcal{V} = \frac{1}{2\mathcal{I}}; \mathcal{I} = \mathcal{I}_{\mathrm{MF}} - \frac{\sigma^2}{2}(\gamma+1)\frac{S}{S-\mu};\ \mathcal{I}_{\mathrm{MF}} = \frac{1+\mu\varphi}{1+\mu\varphi/u} = 1 - \mu/S + \mathcal{O}\left(S^{-1}\right)$$

when $\mu = 0$, $u = 1$, $\mathcal{I}_{\mathrm{MF}} = 1$. When $\sigma = 0$, the expression for the mean variance is exact (mean-field case).

Demographic noise in the assembly perspective for identical carrying capacities and growth rates

B must be replaced with the Jacobian matrix

$$J = -D_* B = -D_* - D_* A;\ D_* = \mathrm{diag}(N^*)$$

Where the N_i^* are the equilibrium abundances of surviving species. The Lyapunov equation becomes

$$D_* BC + CB^\top D_* = -D_* \Leftrightarrow D_* C + CD_* + D_* AC + (AC)^\top D_* = D_*$$

Off diagonal terms if $\varphi = \frac{S^*-1}{S} = \phi - 1/S \approx \phi$ read

$$u\left(N_k^* + N_j^*\right) C_{kj} + N_k^* \left(\mu\varphi \langle C_{\cdot \neq j} \rangle + \frac{\mu}{S} C_{jj} + \sigma a_k^l C_{lj}\right) + N_j^* \left(\mu\varphi \langle C_{\cdot \neq k} \rangle + \frac{\mu}{S} C_{kk} + \sigma a_j^l C_{lk}\right) = 0$$

let us write

$$\mathcal{N}_k^{(j)} = \frac{2N_k^*}{N_k^* + N_j^*};\ \ \mathcal{N}_k^{(k)} = 1 \text{ and } \left\langle \mathcal{N}_\cdot^{(\cdot)} \right\rangle = 1$$

then

$$2C_{kj} = -\frac{\mu/S}{u}\left(C_{kk}\mathcal{N}_j^{(k)} + \mathcal{N}_k^{(j)} C_{jj} + \varphi \langle C_{\cdot \neq k} \rangle \mathcal{N}_j^{(k)} + \varphi \mathcal{N}_k^{(j)} \langle C_{\cdot \neq j} \rangle\right) - \frac{\sigma}{u}\left(\mathcal{N}_k^{(j)} a_k^l C_{lj} + \mathcal{N}_j^{(k)} a_j^l C_{lk}\right)$$

taking the mean over $k \neq j$ gives the same approximate relationship between $\langle C_\neq \rangle$ and $\langle C_= \rangle$ as for the mean field case (and is exact when $\sigma = 0$). On the diagonal:

$$2C_{jj} = 1 - \sigma a_j^k 2C_{kj} - 2\mu\varphi \langle C_{\cdot \neq j} \rangle$$

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

The abundances only enter via the expression of C_{kj} but always as a ratio $\mathcal{N}_k^{(j)}$ or $\mathcal{N}_j^{(k)}$ whose average is 1 . We get

$$2C_{jj} \approx 1 + \frac{\sigma^2}{u} \underbrace{\mathcal{N}_j^{(k)} a_j^k a_k^l C_{lj}}_{\approx \phi\gamma \left\langle \mathcal{N}_j^{(\cdot)} \right\rangle C_{jj}} + \frac{\sigma^2}{u} \underbrace{\mathcal{N}_k^{(j)} a_j^k a_j^l C_{lk}}_{\approx \phi\gamma \left\langle \mathcal{N}_\cdot^{(j)} \right\rangle \langle C_= \rangle} - 2\mu\varphi \langle C_{\cdot \neq j} \rangle$$

taking the mean over j gives

$$2 \langle C_= \rangle \approx 1 + \frac{\sigma^2}{u} \phi (\gamma + 1) \langle C_= \rangle - 2\mu\varphi \langle C_\neq \rangle$$

so that

$$\langle C_= \rangle \approx \mathcal{V} = \frac{1}{2} \frac{1}{\mathcal{I}_{\mathrm{MF}} - \frac{\sigma^2 \phi}{2u} (\gamma + 1)}$$

when $\mu = 0$, $u = 1$, $\mathcal{I}_{\mathrm{MF}} = 1$. All in all the approximation for (adimensional) invariability is

$$\mathcal{I} \approx \mathcal{I}_{\mathrm{MF}} - \frac{\sigma^2}{2} (\gamma + 1) \frac{\phi S}{S - \mu};\ \mathcal{I}_{\mathrm{MF}} = \frac{1 + \mu\phi}{1 + \mu\phi/u} = 1 - \mu/S + \mathcal{O}(S^{-1})$$

More general case in the assembly perspective

In general B must be replaced with the Jacobian matrix

$$J = -DB = -D - DA;\ D = \mathrm{diag}\,(N^* r/k)$$

Under stochastic perturbations scaling as $N_i^{*\nu/2}$ ($\nu = 1$ demographic noise, $\nu = 2$ environmental noise) the Lyapunov equation becomes

$$DBC + CB^\top D = -D_*^\nu \Leftrightarrow DC + CD + DAC + (AC)^\top D = D_*^\nu$$

For off-diagonal terms we may reproduce the previous calculations by replacing N_i^* by $\mathcal{K}_i = N_i^* r_i / k_i$ and

$$\mathcal{K}_k^{(j)} = \frac{2\mathcal{K}_k}{\mathcal{K}_k + \mathcal{K}_j};\ \ \mathcal{K}_k^{(k)} = 1 \text{ and } \left\langle \mathcal{K}_\cdot^{(\cdot)} \right\rangle = 1$$

On the diagonal however we now have:

$$2C_{jj} + A_j^l C_{lj} + A_j^l C_{lj} = \frac{k_j}{r_j} N_j^{*\nu - 1}$$

which, as above, leads us to

$$\left(1 - \frac{\sigma^2}{2u} \phi (\gamma + 1)\right) \langle C_= \rangle \approx \left\langle \frac{k}{2r} N^{*\nu - 1} \right\rangle - \mu\varphi \langle C_\neq \rangle$$

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

so that

$$\langle C_= \rangle \approx \mathcal{V}_\nu = \left\langle \frac{k}{r} N^{*\nu-1} \right\rangle \mathcal{V}$$

In particular, for $\nu = 0, 1, 2$ (immigration, demographic and environmental noise, respectively):

$$\mathcal{V}_0 = \left\langle \frac{k/r}{N^*} \right\rangle \mathcal{V};\ \mathcal{V}_1 = \left\langle \frac{k}{r} \right\rangle \mathcal{V};\ \mathcal{V}_2 = \left\langle \frac{k}{r} N^* \right\rangle \mathcal{V}$$

Taylor's law

relates species mean abundance and their variance. In our context, under some stochastic noise acting independently on species (mean case scenario) we have that

$$C_{jj} \approx \frac{1}{2} \frac{k_j}{r_j} N_j^{*\nu-1} + \frac{\sigma^2 \gamma}{2} \frac{\phi S}{S-\mu} \left\{ \left\langle \mathcal{K}_j^{(\cdot)} \right\rangle C_{jj} + \left\langle \mathcal{K}_\cdot^{(j)} \right\rangle \langle C_= \rangle \right\} - \mu\phi \langle C_{\cdot \neq j} \rangle$$

We roughly approximate $\left\langle \mathcal{K}_j^{(\cdot)} \right\rangle$ and $\left\langle \mathcal{K}_\cdot^{(j)} \right\rangle$ by 1. Using that

$$2 \langle C_{\cdot \neq j} \rangle \approx -\frac{1-u}{u} \left(\langle C_= \rangle + C_{jj} + \phi \langle C_\neq \rangle + \phi \langle C_{\cdot \neq j} \rangle \right)$$

so

$$-\mu\phi \langle C_{\cdot \neq j} \rangle \approx \frac{\mu\phi\mu}{2(S-\mu) + \phi\mu} \left(\langle C_= \rangle + C_{jj} + \phi \langle C_\neq \rangle \right)$$

If we define

$$\frac{\mu\phi\mu}{2\,(S-\mu) + \phi\mu} = f_S(\mu, \phi)$$

This leads us to

$$\left(1 - \frac{\sigma^2\gamma}{2} \frac{\phi S}{S-\mu} - f_S(\mu,\phi) \right) C_{jj} \approx \frac{1}{2} \frac{k_j}{r_j} N_j^{*\nu-1} + \left(\frac{\sigma^2\gamma}{2} \frac{\phi S}{S-\mu} + f_S(\mu,\phi) \right) \mathcal{V}_\nu$$

if μ is of order zero in $1/S$ then f_S is of order one and for S large enough

$$\left(1 - \frac{\sigma^2\phi\gamma}{2} \right) C_{jj} \approx \frac{1}{2} \frac{k_j}{r_j} N_j^{*\nu-1} + \frac{\sigma^2\phi\gamma}{2} \mathcal{V}_\nu$$

We thus expect an exact power law $\mathrm{Var}(N_i) = C_{jj} \sim N_i^{\nu-1}$ when $\gamma = 0$ and σ^2 large enough (so that N_j and k_j become uncorrelated).

bioRxiv preprint doi: https://doi.org/10.1101/147728; this version posted June 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

C. Other calculations

Generating A with correlation to K

Here is how to generate A_{ij} with prescribed correlation to K_i:

$$A_{ij} = \frac{\mu}{S} + f(K_i) + \sigma a_{ij} \tag{152}$$

$$\langle A_{ij} \rangle_j = \frac{\mu}{S} + f(K_i) \tag{153}$$

$$\langle A_{ij} \rangle = \frac{\mu}{S} + \langle f(K_i) \rangle \tag{154}$$

Let's set $\langle K_i \rangle = 1$, then:

$$\langle K_i A_{ij} \rangle = \frac{\mu}{S} + \langle f(K_i) K_i \rangle \tag{155}$$

$$\langle K_i A_{ij} \rangle - \langle K_i \rangle \langle A_{ij} \rangle = \langle f(K_i) K_i \rangle - \langle f(K_i) \rangle \tag{156}$$

$$\left\langle A_{ij}^2 \right\rangle = \frac{\mu^2}{S^2} + 2 \langle f(K_i) \rangle \frac{\mu}{S} + \left\langle f(K_i)^2 \right\rangle + \frac{\sigma^2}{S} \tag{157}$$

$$\langle A_{ij} A_{ji} \rangle = \frac{\mu^2}{S^2} + 2 \langle f(K_i) \rangle \frac{\mu}{S} + \langle f(K_i) \rangle^2 + \frac{\sigma^2 \gamma}{S} \tag{158}$$

$$\left\langle A_{ij}^2 \right\rangle - \langle A_{ij} \rangle^2 = \left\langle f(K_i)^2 \right\rangle - \langle f(K_i) \rangle^2 + \frac{\sigma^2}{S} \tag{159}$$

$$\langle A_{ij} A_{ji} \rangle - \langle A_{ij} \rangle^2 = \frac{\sigma^2 \gamma}{S} \tag{160}$$

Now, let's set $f(K_i) = c(K_i - 1)$ so that

$$\langle f(K_i) \rangle = 0 \tag{161}$$

$$\left\langle f(K_i)^2 \right\rangle = c^2 (\left\langle K_i^2 \right\rangle - 2 \langle K_i \rangle + 1) \tag{162}$$

$$= c^2 \sigma_K^2 \tag{163}$$

Then

$$\langle A_{ij} \rangle = \frac{\mu}{S} \tag{164}$$

$$\langle K_i A_{ij} \rangle - \langle K_i \rangle \langle A_{ij} \rangle = c \left\langle K_i^2 \right\rangle - c \langle K_i \rangle = c \sigma_K^2 \tag{165}$$

$$\left\langle A_{ij}^2 \right\rangle - \langle A_{ij} \rangle^2 = c^2 \sigma_K^2 + \frac{\sigma^2}{S} \tag{166}$$

$$\frac{\langle A_{ij} A_{ji} \rangle - \langle A_{ij} \rangle^2}{\left\langle A_{ij}^2 \right\rangle - \langle A_{ij} \rangle^2} = \frac{\gamma}{1 + Sc^2 \frac{\sigma_K^2}{\sigma^2}} \tag{167}$$

Hence, achieving a prescribed correlation $C = c\sigma_K^2$ requires to generate a matrix a_{ij} with parameters

$$\tilde{\sigma}^2 = \sigma^2 - Sc^2 \sigma_K^2, \qquad \tilde{\gamma} = \gamma (1 + sc^2 \sigma_K^2 / \sigma^2) \tag{168}$$

before inserting it into $A_{ij} = \frac{\mu}{S} + c(K_i - 1) + \sigma a_{ij}$.