DISORDERED SYSTEMS IN COMMUNITY ECOLOGY: A TUTORIAL

Matthieu Barbier[1]

---

[1] Correspondance: contact@mrcbarbier.org
Affiliation: CIRAD, UMR PHIM, 34090 Montpellier, France
& PHIM Plant Health Institute, University of Montpellier, CIRAD, INRAE, Institut Agro, IRD, 34090 Montpellier, France
& Institut Natura e Teoria en Pirenèus, Surba, France.
ORCID: 0000-0002-0669-8927.

Introduction

A basic fact of community ecology is that most ecosystems are comprised of many species and processes. Mathematical models trying to capture these ingredients and explain empirical phenomena usually face the problem of having vast numbers of unknown parameters and uncertain assumptions. It is not obvious when such mathematical models can truly be quantitatively predictive, or even conceptually helpful – to this day, this branch of ecological theory carries a heavy burden of proof.

This document is an informal introduction to disordered systems in community ecology, an approach inspired by statistical physics, whose starting point is trying to capture a few central ecological processes or mechanisms and replacing many other unknown details by some form of randomness.

The challenge comes in two parts: synthesizing observable phenomena into general(-ish) concepts that can be given mathematical counterparts of this kind, and understanding said mathematical objects enough to make them usable (e.g. predictive or explanatory). The first part requires conceptual effort from theoretical ecologists, a lot of which happens either before writing any equation, or going back to reality after all the math is done. Physics-inspired thinking and tools help with the second part, guiding modelling choices and revealing their consequences more robustly than standard explorations through numerical simulations.

My hope with this manual is to help people on each side better understand the broader picture and how to concretely work on this challenge, at least in the context of the specific approach I have been using, which I think has a lot of potential for connecting and synthesizing a range of existing work and questions in theoretical community ecology.

Disclaimers:

- **This is a work in progress, major changes and corrections are likely to happen.**
- This guide is informally written, and meant as a frank and practical compendium of all the ideas, issues, assumptions and tricks I wish I had been told as I started working on this topic. I will not hesitate to voice opinions. I will strive to be inclusive of ideas, but not exhaustive in citations, and I will allow myself biases in choices of references as long as I feel all necessary information is conveyed. There exist more traditionally written overviews or related efforts, see e.g. [1] for a mathematical audience, or [2] and Barbier & Leibold (in prep), though none cover all the material I would like to discuss here.
- It contains a lot of speculation, shortcuts and simplifications. I will strive to signal those whenever possible, be explicit about the level of certainty I assign to various statements (through "epistemic notices" at the start of setions and mid-text when relevant) and, of course, cite works going in depth into the arguments or the mathematics that are simply broached upon here.

Structure and entry points:

This guide is meant to be useful to researchers coming from a variety of backgrounds, from empirical ecologists wondering what the fuss is about, to theoretical physicists trying to figure out how they can contribute to ecological understanding. As a consequence, the guide is modular, with several entry points to choose from:

- 1 **Modelling:** how do we go from reality to a certain class of mathematical objects (dynamical equations with random parameters), what are the explicit or hidden assumptions in this process
- 2 **Theory:** what can we say about these mathematical objects, their behaviors and predictions

- 3 **Epistemology:** what is the meaning of randomness and what are conceptual justifications for using it in community ecology
- 4 **Empirics:** what are the most salient general phenomena we are trying to explain in many-species ecosystems, and what are their possible explanations in our framework

Readers with an ecology background are likely to be happier starting with either of the latter two chapters, whereas those with a complex systems (physics, applied math...) background will probably be more comfortable starting with the former two, as well as the brief introduction below.

Main questions tackled here

Figure 1 illustrates the three broad questions I find most salient in community ecology:

- **Biodiversity:** why do we encounter so many species, why are some common and some rare?
- **Dynamics:** how does a community change over time, what makes it sensitive or resilient to sources of change?
- **Functioning:** what does a community (or a species) use and produce, what is the role of each component species or individual in that process?

Of the three axes shown in the figure, two axes, biodiversity and dynamics, will be extensively discussed in this guide. Questions of functioning will receive some attention but notably less, for reasons that I present briefly below and in Chapter 1.

For all these questions, there is a tension in the literature between two broad types of explanations (this line of thinking is pursued further in Chapter 3):

- **the classical ecology approach:** deterministic explanations based on a limited number of hopefully well-understood factors (**"low-dimensional"** explanations), usually particular to a given system

  e.g. "increasing nitrogen supply in lakes leads to a predominance of blue-green algae"

- **the classical complex systems approach:** explanations based on a large number of often poorly-understood factors (**"high-dimensional"** explanations), usually probabilistic and generic i.e. meant to apply to many different systems

  e.g. "communities with a high diversity of species and interactions are more likely to be unstable and fluctuate in composition"

Of course this distinction is not absolute, and there are approaches that mix aspects of both types[2].

Clearly, our central approach here is the second one. But I believe it is fair to say that the most important existing ecological insights into functioning are of the first kind. This does not mean that the complex systems approach has nothing to say there – there is ample literature on the topic, although often oriented in focus toward resource consumption, see e.g. [4] – but I feel it has a harder time addressing how empirical ecologists, conservationists, ecosystem managers... think about ecosystem functioning.

Even for other questions, I believe it especially important for researchers from complex systems to keep in mind the suspicion that, perhaps, whichever ecological phenomenon they are interested in modelling may be "low-dimensional" in some significant way, and I will provide both theoretical and empirical arguments for this. Conversely, I invite ecologists to consider the value of exploring the "high-dimensional" mindset and the kind of questions and answers it may provide.

Ultimately, I believe we need a mix of both, though it is unclear at this stage whether this should arise as different models for different concerns, or as models combining a crucial few deterministic factors with a probabilistic and generic outlook on all the other aspects of each system.

---

[2] For instance, using body mass and metabolic rates as the main features characterizing each species in order to predict ecological patterns, see e.g. [3], is a "low-dimensional", deterministic, yet generic explanation (applying across all forms of life).

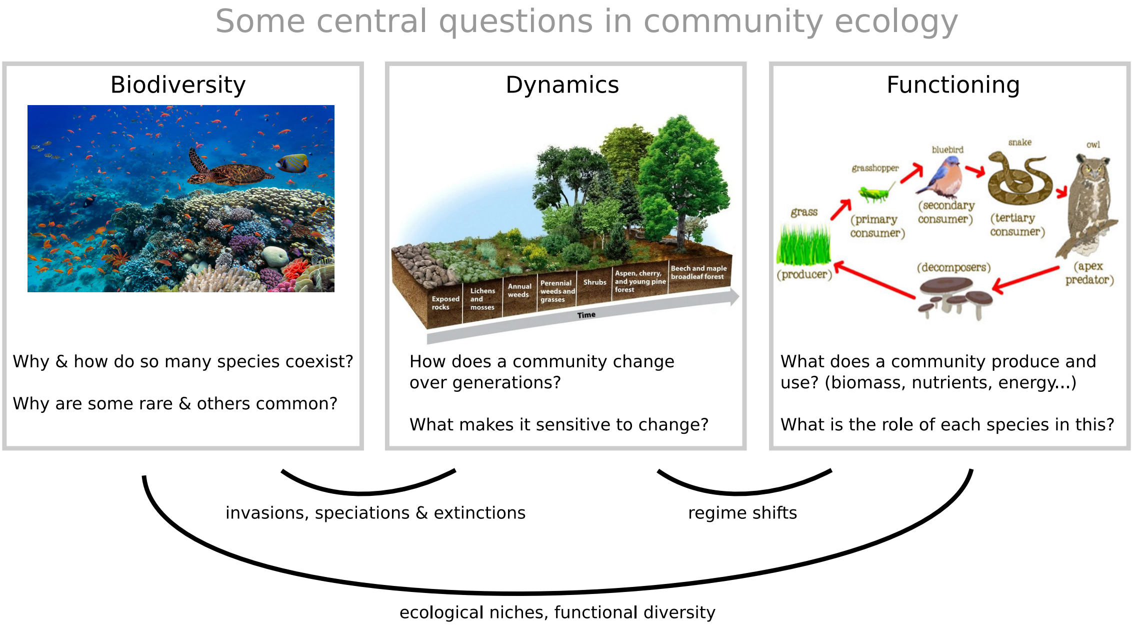

Figure 1: Some central questions of community ecology. There are at least three axes of investigation, all connected but arguably shaped by the influence of different disciplines: an axis of biodiversity originating in naturalist studies, an axis of dynamics notably tied to mathematical modelling and forecasting, and an axis of functioning influenced by systems thinking and physiology, seeing ecological communities as influx-outflux relationships or organisms.

Chapter 1

Modelling

Summary

This chapter is meant as an entry point to our approach for the mathematical modeller, regardless of prior knowledge in community ecology. It will focus on giving a rapid overview of what kinds of models may be appropriate to deal with the kind of data that we have, rather than explaining the history of those models or deep insights into their workings or their ecological interpretations. Other researchers might prefer to start elsewhere while still possibly benefitting in some aspects of this chapter. I argue that we land on Lotka-Volterra models, not anything simpler or more complex, because they are qualitatively adequate for the most universal patterns observed in empirical abundance distributions, and list a few reasons to go further, e.g. functioning.

A classical way of building up to mathematical models in population and community ecology is what I would call the **mechanism-based** approach: we start from simple settings such as a single individual or population, list processes and mechanisms such as reproduction, resource consumption, competition, movement in space... and show how they could be expressed in mathematical terms. I direct the interested reader toward existing textbooks that do this, e.g. [5].

Here I tend to mainly follow a different, **outcome-based** approach: introducing model ingredients based on how they can explain some empirical observations in many-species communities, regardless of whether these ingredients provide a plausible or realistic description of underlying mechanisms. If the distinction between the two approaches or the justification for the latter is unclear, I suggest starting by reading Appendix A at the end of this chapter.

I think the most basic question of community ecology (see Fig. 1) can be stated as such:

How can we explain why various taxa (e.g. species) are found, in conjunction or not, at certain points in space and time?

As noted in Chapter 4, the most widely available type of data relevant to this question are tables of presence/absence or abundance of various species across a variety of samples, sites, timepoints, environments, etc. and our specific approach to this question leads us to ignore explanations that are very specific to a taxon or environment, and instead focus on more generalizable aggregate properties such as community-wide statistics and distributions.

1.1 Sampling models

Epistemic status: Not my work, not checked very thoroughly.

A natural first step is to try to find simple regularities in the data itself: can we find a probabilistic model that generates data with the same distributions and statistics as the empirical observations?

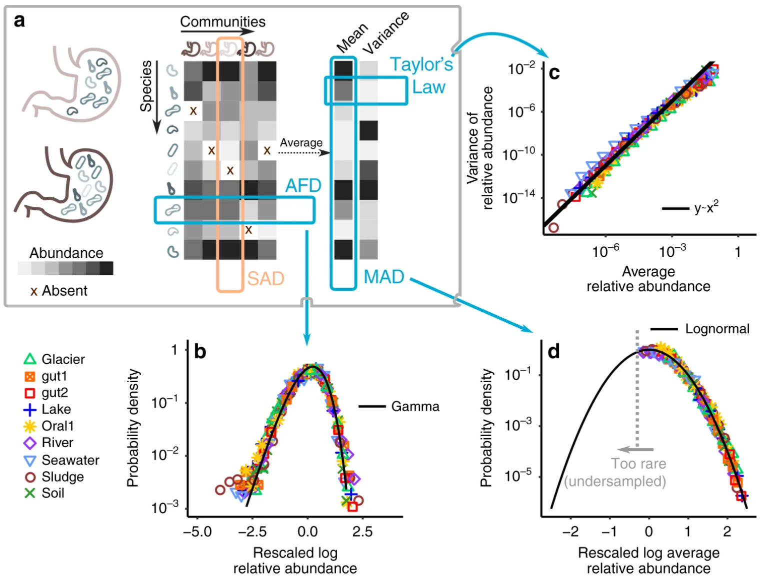

Figure 1.1: A simple model summarizing statistical regularities for several microbial datasets of (relative) abundance of various species (rows) across various samples or time points (columns), reproduced from Grilli [6]. The various patterns seen here are discussed in more detail in Chapter 4.

While such phenomenological approaches (random sampling [6], Maximum Entropy [7], stochastic geometry [8]) postulate almost no ingredients beyond the observed patterns, those very few ingredients often take the form of hidden or seemingly benign assumptions, which often turn out to play a deviously profound role in predicted outcomes.

As a central recent example, I briefly recall a series of work by Grilli and colleagues on several microbial datasets of abundance of many species across many samples, suggesting that a number of distinct patterns in that data (expounded on in Chapter 4) can be summarized by drawing species abundances from a set of simple rules:

- fat-tailed distributions for abundances and their fluctuations: a lognormal distribution of "typical abundances" or "carrying capacity" K_i characterizing each species i in a given dataset, and a Gamma distribution of fluctuations of each species' abundance $N_i(x,t)$ around that typical value between samples x or times t, see Fig. 1.1 and [6]

- occasional jumps in carrying capacity for at least part of the species [9]

- correlated carrying capacities for the same species across environments [10]

- correlated carrying capacities between species based on phylogenetic proximity [11]

Practical note: Trying to reproduce these results in the same or different datasets is a good way to encounter many subtleties of ecological data analysis. Issues with sampling (not seeing rare species) can dramatically bias a naive analysis of abundance distributions, for instance giving a rather convincing appearance of a power-law shape for the mean abundance distribution even in fake data generated using a lognormal for the very same distribution.

In this model (or suite of models), species are noninteracting, and their abundances are driven independently by "environmental factors" varying in space and time. Their empirical success does not necessarily entail that this is actually the microscopic truth, and indeed we suggest in chapters 2 and 4 that similar results could arise with fluctuations entirely driven by species interactions instead.

1.2 Stock and flux models

Going beyond models of noninteracting species, we may want to explicitly incorporate interactions and causality via a dynamical model, in order for instance to better predict the outcomes of experimental modifications of the system.

The simplest option is to think of each species (or other variable) as a stock, and model fluxes between these stocks, most simply through linear relationships[1]

$$\frac{dN_i}{dt} = s_i + \sum_j F_{ij} N_j \tag{1.1}$$

with sources s_i (or drains $s_i < 0$) and F_{ij} coefficients for influxes or outfluxes of i controlled by j. Its equilibrium

$$0 = s_i + \sum_j F_{ij} N_j$$

is a simple mass balance equation saying that all the fluxes coming in and out of a compartment must sum to zero.

Such representations are found throughout ecology, associated with

- models of chemical (e.g. nutrient) concentrations
- internal population structure, i.e. splitting a species into multiple categories (different ages, sizes, genetic types) with linear fluxes between them (e.g. aging, somatic growth, mutations ... transferring individuals between categories)
- empirical/management ecosystem models like Ecopath/Ecosim [12] where compartments are indeed often species, and fluxes represent transfers of biomass or nutrients e.g. due to predation

Furthermore, while this point might be lost on readers at this stage and I will come back to it later, many models somehow imposing "limits" on how much interaction can happen between species, e.g. saturated or ratio-dependent predation [13, 14], will in fact recover exactly or approximately something like the linear model (1.1) and many arguments in favor of these limiting mechanisms may simply be arguments in favor of being linear or close enough.

1.3 Lotka-Volterra as an archetype

1.3.1 Why Lotka-Volterra, why stop there

Not quite my reason why: As a simple matter of principle, the above models fail to capture a property that seems essential when we talk about biological populations, i.e. their ability for exponential growth, one of ecology's major concerns since Malthus and Darwin. Exponential growth would suggest a dynamical model that looks like $dN/dt = RN$, and then, since that growth cannot go on forever, we would replace the constant growth rate R by a diminishing function such as $R(N) = r(1 - N/K)$, giving the logistic model with an equilibrium ("carrying capacity") $N = K$. We could then make that equilibrium depend on other species as well, giving the classic Lotka-Volterra model

$$\frac{dN_i}{dt} = r_i N_i (1 - \sum_j A_{ij} N_j)$$

One notable property is that the equilibrium of this model is still a mass balance equation like the equilibrium of (1.1), so we may try to simply connect the two classes of models through their equlibrium properties.

However, it is not at all obvious that the model's ability to exhibit exponential growth is truly essential to understanding the patterns that we see in data – we could imagine such extreme growth to be relevant only in edge cases (some epidemics, a Petri dish, an anthropogenic invasion) and largely irrelevant in the

---

[1] It is worth noting that linear models are both an interesting starting point and quite singular in many ways.

majority of natural communities. It could be that, fundamentally, most observed abundances are not driven by these exponential tendencies, but tightly controlled by other causal factors (e.g. nutrient fluxes, oceanic currents...), in which case a Lotka-Volterra model might entirely miss the correct causality and fail to predict any realistic dynamics.

My reason why: Rather than just a matter of principle, I propose an empirical argument for the relevance of using Lotka-Volterra-like models when dealing with abundance data.

In almost all settings, especially natural ones, species abundances are seen to fluctuate quite dramatically in time and space (see for instance databases of population time series [15]), and these fluctuations very often seem to be multiplicative in nature – meaning that populations go up and down by some fraction of their current abundance, rather than by roughly constant increments – as evidenced by various widespread patterns [16, 6]. Such multiplicative fluctuations are very likely associated with something like exponential growth, with a changing growth rate $R(t)$, perhaps due to external factors in the environment or due to interactions with other species that are themselves fluctuating.

As I will argue below in Sec. 1.5, I don't believe that going "further" than Lotka-Volterra can introduce similar deep qualitative changes that appear as universally mandated by available data. This is not to say that I necessarily think Lotka-Volterra is the right way to model ecosystems in general, let alone any specific ecosystem, but I feel that the most valid reasons to discard Lotka-Volterra are tied to:

- either considerations particular to certain ecosystems or interactions
- or principled (ecological or mathematical) considerations, such as wanting to account for other aspects of an ecosystem than just species abundances, rather than broad empirical patterns within said abundances.

As noted in Chapter 4, however, it is not entirely implausible that exponential growth and LV might be the wrong model for those large fluctuations; to quote from that chapter:

> "[If indeed] the patterns we are seeing are best explained by exponential tendencies modulated by changing biotic contexts, then we may posit that the central dynamical process to take into consideration is reproduction (the engine of exponential growth), which is what unifies species – or at least local populations – and demarcates them strictly from each other. If however environmental fluctuations are the main driver, it is not obvious that species are the right grain – there is usually significant variation in environmental preferences within a species, and similar preferences across species, therefore it could become more relevant to use a model of ecotypes."

1.3.2 The focal model

Our focal model throughout this document (as it can recover, at least qualitatively and often quantitatively, many of the properties of the other options above) is a spatialized Lotka-Volterra model with growth rates r_i, interactions A_{ij} (the carrying capacity or abundance in the absence of other species is given by $K_i = 1/A_{ii}$), perturbations ξ_i, and dispersal (movement) between adjacent patches x and y with rates D_i

$$\frac{dN_i(x,t)}{dt} = r_i N_i(x,t)\left(1 - \sum_j A_{ij} N_j(x,t)\right) + \xi_i(x,t) + \sum_y D_i\left(N_i(y,t) - N_i(x,t)\right) \tag{1.2}$$

What is N? Prima facie, stock, abundance, biomass etc. but it is actually better to think in terms of *dynamical importance*: everything else being equal, species that end up with large N will typically have more impact on others[2]. The idea that our variables effectively represent dynamical importance generalizes better than the straightforward interpretation of N as abundances, for reasons detailed in Sec. 1.4.2 and in Chapter 4.

---

[2] This is especially true as we later draw parameters such as A_{ij} independently for all species.

What is r? The combination of all sources of growth and mortality that do not depend on the abundance of any species (so-called "density-independent" factors), and may be a function of species traits as well as temporally constant environmental features.

What is the environment? All the things we are not modelling that **are not** part of significant dynamical feedbacks – this may include non-biological phenomena such as the weather, but also other species such as prey or predators that are impacting the focal species we are modelling, but are not significantly impacted by them, or on very different time scales.

What is ξ? Any contribution to abundance variations due to "perturbations" broadly defined, i.e. variable factors that we are not modelling explicitly – possibly happening at larger scales (like weather conditions) or at smaller scales (like individual births and deaths at uncertain times) or in adjacent systems (e.g. a sudden migratory influx from a neighboring community). We can approximate changes in growth rate r by choosing perturbations proportional to N, i.e. $\xi(N,t) = N\hat{\xi}(t)$.

What is A_{ij}? Interactions between all the variables that we are modelling because they might be part of significant feedbacks.

As discussed below, when using a Lotka-Volterra model, we assume that these various forces are (more-or-less) **additively combined** in the **instantaneous** dynamics, but can become entangled in complex ways in their long-term consequences.

1.3.3 Important features

The main ecological features that this model strives to capture are the following, roughly in decreasing order of generality.

- species can be extinct ($N_i = 0$) or present, and the composition of surviving species, i.e. the set Ω of species i such that $N_i > 0$, is a major characteristic of an ecosystem state[3].
- populations tend to grow or decay exponentially when rare, but this exponential growth is impacted by various factors. Here, we further assume that we can separate instantaneous growth into contributions from external and internal factors: on the one hand, interactions (A_{ij}) between the modelled variables, and on the other hand r_i and ξ_i which represent fixed and fluctuating impacts of an environment (both abiotic and biotic) that is not itself impacted by the modelled variables. We note that these contributions of environment and interactions, while they are separated and simply additive here in short-term growth, become unavoidably entangled in their long-term effects [17] (for instance, it is not possible to decompose equilibrium abundances into a simple sum of contributions of the environment and contributions of species interactions or dispersal).
- spatial fluxes can be approximated by simple "diffusion-like" dispersal between adjacent patches, which is enough to qualitatively capture many of the consequences of accounting for space and movement. We could further assume that growth rates r_i or interactions A_{ij} depend on the spatial location to introduce spatial heterogeneity. We also note that, here, interactions only happen within one locality, thus the community (or site) scale can be defined as the spatial scale over which species interactions can be modelled as well-mixed – we do not consider here systems where, for instance, different species might interact over vastly different scales, see e.g. [18].

---

[3] In fact, it is a bit of a peculiarity of the most basic version of the model ($\xi_i = D_i = 0$) that it allows only one ecosystem state per composition: for a given set Ω, assuming they can coexist at equilibrium, their abundances must verify:

$$0 = 1 - \sum_{j \in \text{survivors}} A_{ij} N_j^* \qquad \text{for all } i. \tag{1.3}$$

Such a linear system can only have one solution (per set of survivors); however, that equilibrium need not be stable (and reachable), as we discuss in Chapter 2.

As noted above, any fundamental dynamical properties of LV are due to the exponential tendency of population growth, i.e. dN_i/dt is (mainly) proportional to N_i in (1.2). This tendency may be lost if we have direct fluxes $F_{ij}N_j$ as seen above in (1.1) rather than bilinear interactions of the form $A_{ij}N_iN_j$. This has been proposed to change the dynamics possibly quite dramatically [19], notably by introducing "bottlenecks" – some outcomes of inter-species dynamics can be slowed down due to the intra-specific processes. Dispersal in (1.2) can play a similar role, but it is also special as it involves a conserved quantity (see below in Sec. 1.5.2). The most obvious consequences are that these fluxes can rescue extinct species, and they induce much less extreme feedbacks and fluctuations[4]

1.4 Disorder in Lotka-Volterra-like models

Simply put, disorder is drawing all the parameters in our dynamical model at random, and trying to find the "hyperparameters" or macroscopic features (e.g. maybe which distributions to use, maybe what mean and variance and correlations they should have), that most successfully reflect empirical phenomena. Conceptual and epistemological reasons for why this may (or may not) be a good idea are discussed in Chapter 3.

The most important ingredient of disordered systems theory, compared to previous probabilistic approaches such as sampling theories (a la Grilli [6]) or neutral theories, is assuming that interactions A_{ij} can be modelled as random values – while assuming randomness in other parameters (e.g. r_i, D_i in (1.2)) typically does not lead to altogether different ways of thinking[5], assuming random interactions has surprisingly far-reaching implications for coexistence and dynamics, as we strive to show through theoretical predictions in Chapter 2.

We also note a reason why we may simply stick with Lotka-Volterra, ignoring futher refinements such as nonlinearities in interactions (functional responses), dispersal, etc. While these details may matter a lot in few-species models, in the disordered limit, we speculate that they often fall within "irrelevant" details, in the sense that they can impact quantitative outcomes but they do not really alter the space of qualitative behaviors of the focus model, or only in a very straightforward way. See for instance nonlinearities in [20, 21, 22] and neural network models: in most of these cases, a lot of the phenomenology is similar to that of the random LV model, and differences are easy to guess from basic intuitions.

1.4.1 Low-dimension versus high-dimension

As discussed in the Introduction of the overall document and in Chapter 3, disordered systems approaches are "high-dimensional" explanations, relying on the existence of many processes and details that we need not know exhaustively to obtain a predictive model – yet, there is plenty of ecological phenomena that may point at "low-dimensional" structures being quite important as well.

A very important candidate for low-dimensional phenomenology is the existence of "public goods" that all the species in an ecosystem interact through (impacting it and being impacted by it) – for instance, water availability in an arid environment [23], or pH in a microbial community [24]. It is thus worth keeping in mind that, every time we propose a disorder-based model, it can be useful to confront it to one or several low-dimensional alternatives and see if we have more evidence for one or the other (e.g. as we did in [25]).

1.4.2 Applicability of disorder

A simple objection to the use of randomness is that species typically seem to be much more different from each other than simple random draws from a distribution. To clarify this point, we can think about *degrees* of heterogeneity:

- "variance-like" heterogeneity: quantitative variation within a group of species that can be adequately quantified by statistical moments (e.g. a mean and a variance) of their parameters, recognizing that we will expect to occasionally see some species that lay at the extremes of the distribution.

---

[4] These are consequences that go beyond just linear fluxes, and extend to a lot of "sub-exponential behavior".

[5] Still, some candidates for novel behaviors: desynchronization of oscillatory dynamics, Anderson localization due to dispersal coefficients fluctuating randomly in space, maybe a potential for Turing patterns when pairs of species with asymmetric interactions have sufficiently different dispersal.

- "systematic" heterogeneity: irreducible qualitative differences among such groups (e.g. autotrophs versus heterotrophs).

Disorder is primarily applicable to "variance-like" heterogeneity [6] Differences between classes (for instance, the fact that a plant usually cannot consume or parasitize an animal) typically must be added "by hand" into our model, because they are highly impactful and structured.

. Even so, an important lesson from experience is that differences that appear irreducible and fundamental can **sometimes** be ignored, and entities that seem like they belong to different classes can **sometimes** be bunched together for practical purposes. For example, even species that are very different on a *per capita* basis, such as bacteria & whales, may be more comparable in terms of biomass (as demonstrated e.g. by allometric relations holding over many classes of organisms [3] at least for the purpose of understanding certain macroecological patterns.

Consequently, whenever we think about using randomness in a model, we must seek the representation of the system that makes species as equivalent as possible. To illustrate, we start with a simple theoretical example, using the non-spatial, unperturbed version of the equations (1.2) i.e.

$$\frac{dN_i}{dt} = r_i N_i \left(1 - A_{ii} N_i - \sum_{j \neq i}^{S} A_{ij} N_j \right)$$

We see that $A_{ii} = 1/K_i$ defines the carrying capacity K_i, i.e. the equilibrium abundance in the absence of other species. Let us assume for now that we are looking at a competitive system where all species have a carrying capacity $K_i > 0$ (though everything that follows can be extended to e.g. trophic systems where predators cannot survive without prey).

We could then change variables to $\eta_i = N_i/K_i$, usually called the relative yield [25], and still have a Lotka-Volterra model:

$$\frac{d\eta_i}{dt} = r_i \eta_i \left(1 - \eta_i - \sum_{j \neq i}^{S} \alpha_{ij} \eta_j \right) \tag{1.4}$$

with different interaction coefficients ($\alpha_{ij} = A_{ij} K_j / K_i$). If the A_{ij} are independent random parameters, then the α_{ij} are not independent (and conversely): they are correlated by row and by column due to the factors K_i. This is ecologically meaningful, as it can, for instance, largely decide whether many species coexist or go extinct[7].

If we thus first choose variables and write our model such that species are as equivalent as possible – which translates here to choosing parameters (α, or K and A, or others) to be as narrowly distributed and uncorrelated as possible.

We note that this is not necessarily the most intuitive representation. For instance, in a resource competition setting where individuals of one species consume much more than individuals of another, the best variable to express the importance of a species might not be its abundance, but rather its total consumption (see e.g. [26]).

Hereafter we will use terminology suggesting that our variables N_i represent abundances, but it is crucial to recognize that this will not always be the right choice; instead they should be understood as a general proxy for a species' importance in community dynamics, perhaps expresses in biomass units or consumption rates etc. When we see apparent inequalities between species, we should first ask how much is really due to some species being more or less important, and how much is due to the fact that we are not measuring their importance in the right units?

---

[6] Initially, it was proposed to model physical systems with many particles of roughly the same nature, not ecosystems where organisms differ utterly in scale and physiology, like viruses from trees.

[7] Let us make the empirically-plausible assumption that the carrying capacities K_i are widely distributed (e.g. as a lognormal). Then, if the A_{ij} are independent and narrowly distributed, individuals from all species have comparable per capita competitive impacts on others, and thus, a few species with very large K will likely win the competition and exclude most others.

By contrast, if the α_{ij} are independent and narrowly distributed, each species impacts others through the fraction of carrying capacity it has N_i/K_i, no matter how small that capacity is (as appears in some experiments discussed in [25]). This is sensible if, for instance, some species consume much more resources per capita: then their K is smaller, but a few individuals from that species can still impact other species a lot. In this case there is no telling whether species with large K will win or lose the competition.

1.5 Broader landscape of disordered models

Here we briefly discuss model ingredients that are not explored in this guide, but did or could receive a disordered treatment.

1.5.1 Functions and niches

Common wisdom seems to indicate that, while species composition tends to fluctuate dramatically in complex communities, there is a much more stable notion of "functional composition" or niches – in other words, different species may occupy the same role.

Ecological function or niche – what a species does, its role in an ecosystem – is poorly defined, but is often taken to encompass resources consumed and services provided (e.g. metabolites and other chemical products, pollination, decomposition...). It is usually not as measurable and as universal as species abundances:

- the functions occupied by organisms cannot be observed directly, and we often use low-dimensional proxies: a limited number of measurable traits that are thought to be determining for their roles ("functional"), with body size being the most universal and often used as a single axis of variation structuring ecosystems [27],
- at the other extreme, phylogenetic relatedness is sometimes taken as a proxy for functional relatedness, which could be seen as an infinite-dimensional notion of function (it would take an infinite number of relevant traits to avoid any convergent evolution and ensure that phylogenetically distant species are always functionally different)
- most other examples tend to be more domain specific e.g.
  - the leaf economics spectrum [28].
  - the main situation where one has many explicitly measured traits is molecules consumed and produced, directly and extensively observable through metabolomics, but this is largely restricted to microbial communities and may not exhaust the function/niche of the species (i.e. they may also differ in many ways not captured by their metabolic needs and products)

Therefore, it is much harder to reliably evaluate a community's functional compositon than its taxonomic (species) composition.

Functions/niches could be represented implicitly e.g. in the parameters of a Lotka-Volterra model. Most simply, we could assume that growth rates and interactions are derived from underlying functions or niches, such as species competing if they occupy the same niche, and indeed we can for instance have a Lotka-Volterra approximation of resource competition that captures a lot of the important phenomenology [29, 30].

But we could also want to tend toward a hybrid model between mass balance-like stability (for the functions) and Lotka-Volterra-like dynamics (for the species) This suggests a model where both species and functions are explicitly represented. This would allow us to go beyond explaining patterns of abundances only, and toward patterns related to function (e.g. questions such as "are all niches saturated or are there still open niches that some invader could come and occupy").

1.5.2 Other potentially important model ingredients

Here are some more mathematical ingredients which I speculate could qualitatively impact the aggregate many-species patterns of interest to us:

- Conservation laws: models in physics and chemistry often behave in very particular ways because of strong laws such as the conservation of energy and matter. Such laws are typically absent in ecological models, but having invariants (symmetries, constants of motion) could drastically change the space of possible dynamics, see e.g. in chemical networks [31].
  - The most obvious example here is spatial fluxes, which conserve abundance (no individual is created or killed by dispersal itself, though they may die in places along the way), and this is a non-negligible reason why I am inclined to consider space explicitly in (1.2), rather than just for purposes of being "realistic".

  - Another salient example is nutrient conservation and cycling. Lotka-Volterra models do not account for the fact, perhaps made salient by agronomic advances such as Liebig's, that specific nutrients (e.g. nitrogen) are found to be strong limiting factors on the growth and reproduction of many species. As a consequence, many models explicitly incorporate stoichiometry of a few basic nutrients, especially when dealing with plants or aquatic life (as other lifeforms tend to have less flexibility in their composition). While nutrients are not strictly conserved (there are influxes and leaks), some or many ecosystems may be in a regime where a large fraction of the nutrients involved in population growth is cycling locally, and this could possibly lead to very different dynamics.

- Fluctuating interactions: another way to account for hidden complexity in interactions is to make them fluctuate over time, see e.g. [**?**]. If these fluctuations are strong enough to dwarf the time-averaged differences between species, we expect to go back to a kind of neutral model, since no species will maintain a particular role or advantage over time through its interactions, except through the action of chance (which may well become entrenched due to rich-get-richer mecanisms). It is not entirely clear how much of a qualitative difference fluctuating interactions may make – in some implementations they can amplify inequalities between species to the point of allowing "condensation" where a single species becomes an outlier holding a non-negligible fraction of the whole system's biomass. These kind of outliers exist in economic data e.g. a single city comprising a third of a country's population, and can be modelled with highly fluctuating interactions [32]. For ecological data, it is not obvious to me that we truly encounter such outliers, versus simply large values within a well-defined probability distribution.

- Really non-additive (e.g. multiplicative) interactions: our model (1.2) assumes that the impacts of other species are somewhat substitutable, since their combined effect is simply the sum of their individual effects. One alternative is multiplicative interactions, where effects combine as a product: in the example of multiple essential resources, if a single resource is missing, consumer growth is zero. Long multiplicative chains of interactions would be a natural way of obtaining wide distributions (e.g. abundances). Many models of higher-order interactions combine additive and multiplicative aspects: they often use products like $A_{ijk...}N_iN_jN_k...$, but they usually sum over many such terms, and this additive aspect tends to outweigh the multiplicative aspect and lead to broadly similar phenomenology to that of our focal model, see e.g. [33].

Other ingredients not mentioned here (e.g. higher-order interactions [34], more nonlinear growth or interaction terms [13], discrete individuals and mechanistic derivations for growth and interactions [35]...) fall within the class of those that I cannot yet see produce an equally deep qualitative difference, though they may play a role in quantitative fits or particular patterns.

Appendix A: Justifying the outcome-based approach

The most common way of building a model in ecology is finding ingredients that one assumes are important (e.g. species, nutrients, interactions...), putting them together in a reasonable way, and trying to understand and interpret their outcomes.

There is also a long tradition of "null models" attempting the opposite: start from known behaviors or patterns (e.g. species diversity, abundance distribution), and try to find the simplest set of model ingredients that give us these outcomes, ideally assuming as little as possible beyond that[8].

In the latter perspective, many differences in ingredients, even striking ones such as a model being discrete or continuous, deterministic or probabilistic, spatial or non-spatial, can be treated as irrelevant if they do not impact the target outcomes. One way of articulating this idea is: all details matter, but they do not necessarily matter for what we care about.

---

[8] On the spectrum between these two traditions, one extreme is represented by the purest mechanistic reductionism, anchoring ecology in physics and chemistry where ingredients are truly knowable as exactly as possible, e.g. [36]. The other extreme includes the phenomenological approaches discussed below. Most approaches have a little bit of both, and there is no real sharp divide except those imposed by habit.

Predictions regarding different aspects of the same system may require different degrees of detail. Fidelity to microscopic ingredients is often unnecessary to make good aggregate predictions (atoms are not bouncy balls at all, but that description is enough to understand an ideal gas). Furthermore, trying to increase microscopic fidelity does not necessarily improve accuracy when predicting large-scale outcomes, and can even detract from it if poorly estimated parameters are added (for instance, it might be better to model individuals in a crowd as mindless atoms than using a very wrong model of their psychology).

Hence, we want to think of models not in terms of how they fit our microscopic intuition, but:

- what kind of macroscopic behaviors and patterns they can encompass at all (e.g. a strictly static model cannot be used to predict dynamical patterns),
- whether we can adjust model parameters so that predicted patterns fit their empirical counterparts quantitatively,
- whether parameters fitted using some patterns also correctly predict other patterns – in which case the model is a workable description of macroscopic reality, even if its details turn out to be untrue to microscopic reality.

Chapter 2

Theory

Summary

This chapter attempts to summarize the state-of-the-art of our understanding of our main mathematical objects, without concerning itself too much with empirical relevance (except inasmuch as what qualifies as "main" results here is strongly skewed by conceptual relevance to comunity ecology). The main question being answered here is: what can random interactions between species and random (spatio-temporal) environments entail in terms of biodiversity or dynamics? I will also much more briefly consider the impact of spatial fluxes on the answers to that question, since knowledge there is more limited and speculative.

Here we attempt to give a bird's eye view of the main theoretical behaviors of the Random Lotka-Volterra model (chosen on grounds detailed in Chapter 1)

$$\frac{dN_i(x,t)}{dt} = r_i N_i(x,t)\left(1-\sum_j A_{ij}N_j(x,t)\right) + \xi_i(x,t) + \sum_y D_i\left(N_i(y,t)-N_i(x,t)\right) \tag{2.1}$$

that is, the phenomena, dynamics, regimes etc. that this model can explore.

NB: What do I mean by behaviors? Simpler models such as sampling models (see Chap 1) tend to have only one possible qualitative behavior, e.g. they predict that species abundances follow a lognormal distribution, with only quantitative aspects changing as we vary parameter values. But more complex models have the potential for qualitative changes of behavior: a model of competing species may display either coexistence or exclusion; a predator-prey model may reach a fixed equilibrium or a cyclical trajectory. Different behaviors can avail us of different patterns to compare to data, as we cannot, for instance, study temporal correlations in a fixed equilibrium. Often, the existence of multiple qualitative behaviors in a model reflects the fact that several processes are at play in the model, and their relative importance may change depending on parameter values, leading to very different outcomes.

One might expect that any dynamical behavior is possible in the Random Lotka-Volterra model, since drawing coefficients at random could potentially lead to any structure. Yet, perhaps surprisingly, this model only has a limited set of behaviors of interest, as some occur predictably and all the other possibilities become vanishingly unlikely when species numbers are large.

This section summarizes the regimes in the Random Lotka-Volterra model that were exposed across a wide range of previous works, without going through precise mathematical results, instead emphasizing the basic intuitions that characterize each regime. We list a brief description of every regime in Tables 1 and 2, which are all detailed in the rest of this section.

2.1 Quick summary

In brief, Random-Lotka Volterra (and similar) models typically will display one of three main behaviors[1] [37]:

- **simple equilibrium** where species abundances tend toward a well-defined reference value tracking how “fit” they are in their abiotic and biotic environment;
- **fluctuating dynamics** where either external noise or chaos created by complex interactions leads to permanently fluctuating abundances without a stable ordering of which species is more or less fit [38];
- **multiple equilibria** where the observed composition depends on past history through priority effects, and where some form of long-term dynamics (which could be interpreted as ecological succession) can happen through jumps between many alternate stable states [39].

These chaotic and successional dynamics in many-species Random LV are very different from preexisting ecological intuitions of few-species chaos [40] or deterministic (Clementsian) succession (see discussion in [41]). They mainly rely on the exploration of many different combinations of species, either undirected in time (chaos) or directed on average (succession) even though individual jumps between compositions are largely unpredictable. These three main behaviors can be further detailed in terms of how many species survive and thrive at any point in time, depending largely on the strength of interactions, as we now explain.

2.2 Regimes driven by interactions (no noise, no space)

The simplest version of a random Lotka-Volterra model already provides some important insights about how interaction statistics affect model regimes.

We can start with the most basic Lotka-Volterra equation with relative yields (1.4) for many species, $S \gg 1$,

$$\frac{d\eta_i}{dt} = r_i \eta_i \left(1 - \eta_i - \sum_{j \neq i}^{S} \alpha_{ij} \eta_j \right)$$

The dynamics will tend to bring the expression in parentheses close to zero, even if it doesn’t cancel out exactly (it will if there is a stable equilibrium, but that is not always the case).

Rough ’order of magnitude’ arguments already outline distinct parameter regimes within this model. The mean and variance of the sum $\sum_j \alpha_{ij}\eta_j$ give a simple estimate of how much the species’ abundances will differ from their carrying capacities (we recall that $\eta_i = N_i/K_i$). Interactions do not matter if $\sum_j \alpha_{ij}\eta_j \ll 1$, and so we can call this the “weak interaction” limit, where every species mainly follows its own intra-specific dynamics.

Otherwise, we see that the sum contains $S-1$ terms of the form $\alpha_{ij}\eta_j$. When S is large, the expression in parentheses is thus unlikely to sum to zero if all the α_{ij} and η_j are numbers $O(1)$, meaning “of order 1” – practically speaking, often some value like 0.1 or 10.

Although it is possible that interactions with a particular structure may still allows for perfect net cancellation of interaction effects, this is extremely unlikely if interactions are random. For random interactions, either interactions are small, $|\alpha_{ij}| \ll 1$, or very few of them are nonzero $P(\alpha_{ij} \neq 0) \ll 1$, or abundances are all small (compared to carrying capacity), $\eta_j \ll 1$ or few are nonzero. Thus we can already foresee that the model has at least three main parameter regimes delineated by interaction strength.

---

[1] We remark that when the species number S is not very large (as a rule of thumb, less than 20 in simple simulations), the boundaries between the regimes discussed here are not so sharp, and communities whose parameters are drawn from the same distribution may fall in one regime or another depending on luck of the draw; however for many species, most regimes become clearly distinguishable.

Regimes	**Coexistence**	**Dynamics**	**Comments**
(a) Diffuse interactions: pairwise interactions are small, but the combined effect of all interaction partners is significant			
1. Feasible	All species	Stable point equilibrium	Shrinks as $1/\sqrt{\log S}$ when S increases holding μ, σ constant [42]
2. Unique Fixed Point	Many species	Stable (uninvadable) point equilibrium	[43, 44]
3. Chaotic turnover	All species may persist$^{(1)}$, turnover of large set of thriving species	Large uncorrelated abundance fluctuations	Fluctuation strength increases with species diversity[45, 46]
(b) Sparse interactions: each species can have a significant impact on the abundance of a few others			
4. Intransitive loops	All species persist	Abundance fluctuations	Rather found with asymmetric interactions (Marcus & Bunin in prep)
5. Multi-stable cliques	Potentially large number of species per equilibrum	Many stable equilibria, history dependence	Rather found with symmetric interactions [47], can lead to directional "succession" [39]
(c) Strong interactions: each species can have a significant impact on the abundance of many others (poorly understood regime)			
6. Strong interaction chaos	All species may persist$^{(1)}$, turnover of small set of thriving species	Large correlated fluctuations	[48]
7. Multi-attractor regime	Few species per attractor	Many (?) attractors, sometimes chaos within each attractor	[48]

Table 2.1: Regimes arising from species interactions (see also Fig. 2.1). Coexistence: how many species can persist in the system in the sense of being sometimes able to grow from rare, and how many "thrive" i.e. are found with large abundances (not very small compared to their carrying capacity). $^{(1)}$ Abundance fluctuations can lead to species extinctions in finite populations, but every species can regrow from rare; space can prevent true extinctions, see Table 2.2. Dynamics: presence and multiplicity of equilibria, or persistent fluctuations. More detailed predictions, e.g. species abundance distributions, are discussed in Chapter 4.

	Interaction regime	**Consequences**
Weak noise	Any	Increased abundance fluctuations
Strong noise	Any	**New regime:** Noise-dominated (”neutral-like”) dynamics
Spatial heterogeneity in parameters	1	Species track environmental preferences
	all others	Same modulated by interactions
Multiple sites (or continuous space) with dispersal	1, 2, 4	No qualitative change
	3, 6	Allows persistence of fluctuations even if species go locally extinct [38]
	5, maybe 7	Spatial domains [41], directional “succession” [39]
	5 & 7, maybe 3 & 6	**New regime:** Interaction-driven patch dynamics [49]
Strong noise + space	Any	**New regime:** Noise-driven patch dynamics

Table 2.2: Effects of adding noise (in the sense of stochasticity in the species dynamics) and space to the model depending on the species interaction-driven regimes described in Table 2.1. Some of these effects are speculative as this has been far less studied; we note them.

Diffuse interactions:

Interactions are individually small, but their sum is potentially not negligible (as a collective impact of all interaction partners together on each target species), i.e. if we define aggregate parameters μ and σ as

$$\mu \equiv \left\langle \sum_j^S \alpha_{ij} \right\rangle = S \left\langle \alpha_{ij} \right\rangle , \qquad \sigma^2 \equiv \mathrm{var}\left(\sum_j^S \alpha_{ij} \right) = S \, \mathrm{var}(\alpha_{ij}) \tag{2.2}$$

then this regime corresponds to

$$\mu, \sigma^2 = O(1) \tag{2.3}$$

In that case, many species (a macroscopic fraction, i.e. a number that can be arbitrarily large and keeps increasing proportionally if we increase S while holding μ, σ constant) can coexist with $\eta = O(1)$ at any given time.

Strong interactions:

A single species can play a significant role in the abundance of another

$$\left\langle \alpha_{ij} \right\rangle , \mathrm{var}(\alpha_{ij}) = O(1) \tag{2.4}$$

Most likely, few species can coexist at any given time and place (a fixed number or one that increases slower than S, e.g. $\log S$). There is also, as a more fragile case, the possibility that many species coexist with very small abundances when they are all “similar” enough (e.g. a neutral model with $\alpha_{ij} = 1$ for every interaction and $\eta_i = 1/S$ for every species)

(Very) sparse interactions:

Most interactions are entirely negligible, only O(1) interactions per species are nonzero (or not very small). This results in a system where many species can coexist but have abundances and occurences that are strongly

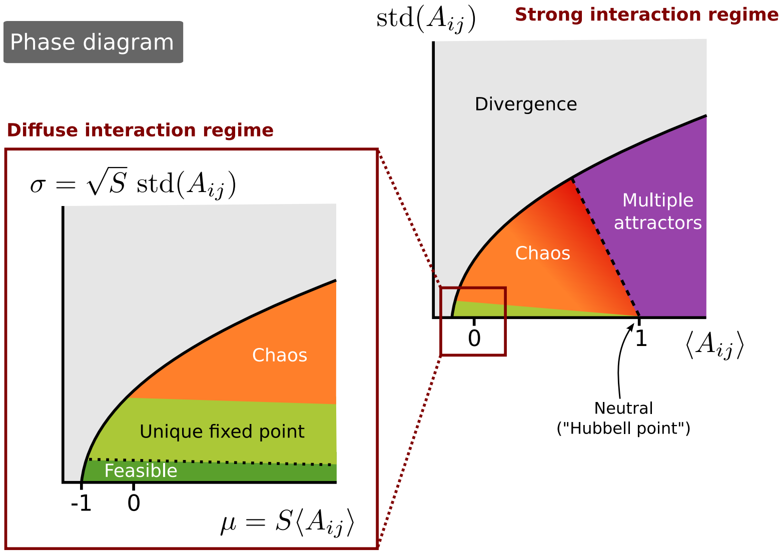

Figure 2.1: Diagram of dynamical regimes for the Random Lotka Volterra model without noise or space (and excluding very sparse interactions). Most existing theoretical studies based on disordered systems are limited to the diffuse interaction regime shown in inset, where species interactions are important in aggregate but very weak individually (i.e. μ and σ remain moderate even for very large species number S), as this is the regime where theoretical tools are most applicable and the most robust predictions can be made. The strong interaction regime is less understood, mainly through simulations, except in some particular limits. The grey region indicates where abundances become infinite due to an excess of mutualistic interactions, i.e. a breakdown of the LV model, which can be solved in various ways. In the inset, the feasible region does not have the same scaling as others: it shrinks to 0 as $S \to \infty$ while keeping μ, σ constant, as it becomes infinitely unlikely for all species to coexist, however any finite μ, σ allow a significant fraction of species to coexist.

affected by a small number of other species. In this case, the resulting regimes may be more complex that the three we listed above. This case has a tendency to behave more like few-species models, including more of a menagerie of different dynamics and behaviors than the other two cases, see e.g. [50].

This non-spatial RLV model further shows some subdivisions within these three regimes that further allows us to make more refined prediction that depend only a few aggregate parameters, i.e. statistics of the individual species parameters: mainly the sum-of-interactions mean and variance μ and σ^2 defined above in (2.2), heterogeneity in carrying capacities var(K) and additional correlations between parameters, such as interaction symmetry $\gamma = \mathrm{corr}(A_{ij}, A_{ji})$ (see [21] for intuitions regarding these parameters and their meaning). It has been shown theoretically [43] that many details (e.g. higher order statistics such as the third or fourth moments of interactions, or which distribution they are drawn from – normal, uniform, ect.) do not affect the outcomes that we are considering here in large (S¿20) biotas.

2.2.1 Diffuse (weak to moderate) interactions

Three dynamical phases are predicted to exist when interactions are weak to moderate (as defined above), meaning that each pairwise interaction has limited impact on a species, but the total impact of all its interaction partners is significant. Previous studies, e.g. [51], have found suggestive empirical evidence for all three phases predicted in that regime, though that does not necessarily mean that this is the empirically correct regime.

Feasible phase:

All species coexist and reach a stable point equilibrium together. This requires very low-variance interactions. Theoretical conditions are given in [42]; they basically require $\sigma \sim 1/\log S$, because the more species there are, the more likely it is that at least one of them will go extinct due to an unlucky draw of its interactions.

Unique fixed point phase:

A fraction of species go extinct, but the remainder coexist stably in an equilibrium that cannot be invaded by any of the extinct species (if we reintroduce them, they go extinct again). This requires moderate interactions whose variance is not too large ($\sigma = S\text{var}(\alpha)$ is below a threshold value σ_c of O(1) given in [43]) which we call the **stability bound**.

Chaotic turnover phase:

At any given time, the dynamics tend to approach an equilibrium where a significant fraction of species thrive at significant abundances, but all such equilibria are unstable and can be invaded by other species that were previously going extinct. Thus, a constant turnover occurs through a kind of "pinball" dynamics, the system bouncing between unstable equilibria with different compositions[2].

This phase arises because, when interactions vary enough between species pairs, $\sigma > \sigma_c$, equilibria lose their stability. This was first predicted by May [53] in a simpler linear setting, and more specifically demonstrated for Lotka-Volterra and similar models in [54, 55, 43, 44].

Here are several key facts about this phase:

- all species get to thrive at some point, no matter how disadvantaged they are by the luck of the draw in the original species pool. We must therefore distinguish what we mean by species being present in the system: all S are observable in the long run, a smaller (but not insignificant) fraction are temporarily advantaged in the sense that they would grow in abundance if added to the system at a given time, and a smaller fraction still have large abundances at a given time.

- the typical trend (though it is very noisy in each specific instance) for each species is a relatively quick jump from very small to significant abundances, followed by a slower decay toward insignificance, and these jumps appear uncorrelated between species and randomly distributed over time

- species that are rare at a given time, and even rare for most of the time, may nevertheless play a crucial role in the dynamics.

- the number of species coexisting with significant abundances at any given times lies strictly below the stability bound (see above), due to the perturbing effect of incoming (currently rare but going up) species on the abundant ones.

- when S is not very large, we can get limit cycles (periodic behavior) instead of chaos. But as we increase S, boundaries become sharp, and cycles in the "chaotic" phase are decreasingly likely and increasingly complex until only chaos occurs in the limit $S \to \infty$.

- the high-dimensional type of chaos we describe here is very different from the low-dimensional chaos that has been described in population models with one or few species (e.g. the logistic map, dynamics with time lags [40] Schaffer, etc). First, it is driven by diversity, and the intensity of fluctuations should thus increase if more species are added. Second, it is tied to species going toward extinction and being able to regrow later to significant abundances.

- it can be difficult to distinguish these fluctuations from others that would be driven externally, e.g. by environmental stochasticity, as both translate to a growth rate that changes over time (furthermore, the temporal asymmetry with sudden jumps then decay had been proposed as a marker of externally-driven fluctuations in other fields such as finance [56], and is here a basic feature of internally=driven dynamics)

---

[2] This can be related to the notion of a heteroclinic network, see e.g. [52]

- this many-species chaos can extinguish itself: if species can go extinct due to finite population size, then the large chaotic fluctuations in abundances found here lead to populations going to ever lower numbers, until some species go extinct. Unless species can then reinvade, this prcoess continues until the remaining $S*$ species are such that $\sigma^* = \sqrt{S^* \mathrm{var}(\alpha)}$ falls below the critical value σ_c. Chaos can be maintained by having immigration from an external pool (or mutation) that allows species to reinvade, and we explain below in Sec. 2.4 that it can also persist in a metacommunity [38].

2.2.2 Strong interactions

When interactions are strong, i.e. individually $O(1)$ (e.g. $\alpha_{ij} = 0.5$), very few (also $O(1)$) species can coexist at any given time. This means that many of the theoretical benefits of looking at a many-species system, such as having well-defined dynamical phases, may be lost. While this parameter regime is only starting to be understoo [48], we can roughly distinguish two dynamical regimes, chaos and multistability described below.

The transition between the two regimes occurs around $\langle \alpha \rangle = 1$, but it is possible for systems to be in a state that mixes chaos and multistabilty, i.e. a dynamical landscape with multiple basins of attraction but at least some attractors are chaotic rather than fixed points [48]. As noted above, we do not expect well-separated phases, since few species survive at significant abundance at any given point in time and space, and thus, it is not obvious that we can benefit from the tendency of dynamics to tend toward a few well-defined archetypes when there are many interacting variables.

Strong interaction chaos:

This resembles the chaotic phase described above, still with noticeably intermittent dynamics (the typical trajectory is that species jump to large abundances in a relatively short time, and then decay), but with

- few ($O(1)$ rather than a significant fraction of S) species thriving at large abundances at any given time
- strong correlations between species, to quote [48]: "As one moves deeper into the chaotic phase, the collective correlation increases continuously, as the effective community size drops, suggesting a seamless transition from a weak-interaction, chaotic regime amenable to full dynamical mean-field theory treatment to the strongly correlated regime that we have analyzed by simulations and the approximate focal-species model."

Multiple attractors:

There exist many different attractors that each allow a relatively small subset of coexisting species. As noted in the summary above (Sec 2.1), this picture contrasts with some classic ecological intuitions on multistability in multiple ways. This model will *not* exhibit regime shifts between just two or a few very distinct states, like savannas and forests or lake eutrophication [57]; a different picture of the world is needed for those.

Furthermore, with at least some small noise, the system can undergo slow dynamics of jumping from one attractor ot another [39] but they differ from the classical ecological concept of succession as proposed by Clements for plant associations: the dynamics do not go determinitically from one stage to the next, each characterized by a single set of species; instead we see much more stochastic trajectories through many possible combinations. Instead, this occurs wiith many species forming many different low diversity associations, with adjacent "spots" retaining distinct compositions despite spatial fluxes. Coexisting subsets are "cliques" of species that do not compete too strongly, and are stable against all other species invading in small numbers. But a large enough invasion (or an invasion coupled with an environmental perturbation) can cause species in the clique to go extinct and the system to jump to a different stable equilibrium corresponding to another clique.

2.2.3 Sparse interactions

Very sparse interactions, where a substantial majority of pairwise coefficients A_{ij} are negligable, have barely been studied [47]. One reason is biological: many well-studied ecological networks are relatively sparse, but those usually describe interactions *between* different functional groups or trophic positions, e.g. food webs or plant-pollinator networks (sparsity then corresponds to the prevalence of specialized predation or pollination links). Instead, interactions that can most plausibly be treated as fully disordered involve competitive or multifactorial interactions within a guild of similar species (see Sec 1.4.2), e.g. competition between grasses in a field, or a mixture of positive and negative effects between soil bacteria, and there is no reason to expect that such interactions should be sparse.

It seems that due to the fact that a few partners have a significant impact on each species, details about the interactions tend to matter more than in the diffuse interaction regime. For instance, these models often have a large number of moderately-important transitions between phases that differ due to some details rather than in response to overall changes in parameter (. [47]). Work done with sparse interactions are nevertheless interesting in that they can be strong and still allow many species to coexist. Thus, despite their *a priori* unrealistic character for competitive or multifactorial interactions, we could imagine that they may represent some scenarios encountered in many-species communities.

This work also highlights the distinction between **intransitive loops**, i.e. a regime of abundance fluctuations driven by mechanisms somewhat akin to rock-paper-scissors interactions (Marcus & Bunin in prep), and **multistable cliques** [58, 47], i.e. a regime where species compete strongly enough to exclude their interaction partners, and thus coexist only in "cliques" with species that they do not interact with. The former picture requires asymmetric interactions (one species impacting another but not being impacted in return), while the latter tends to be associated with symmetric interactions.

The equilibria observed in the multistable phase tend to be very numerous and to “look relatively alike”, i.e. many will tend to have comparable total biomass, diversity, etc. Nevertheless, we can see a kind of stochastic succession dynamics: Bunin has shown that equilibria can be assigned a metric of “maturity” that correlates roughly with species diversity, such that perturbations will on average cause more jumps toward more mature states [39] (see also [59] in a related model).

2.3 Noise-dominated regimes

The analysis of the simple case we described above serves as a useful baseline model to which we can start to add other elements. We can start by asking if our assumption of complete determinism can be expanded to allow for noise (real stochasticity). To do this we use the Lotka-Volterra equations we had before (equation 1 above) and focus on $\xi_i(t)$. Here we ignore the terms related to dispersal and come to those in the following section.

$$\frac{dN_i(t)}{dt} = r_i N_i(t)\left(1 - \sum_j A_{ij} N_j(t)\right) + \xi_i(t)$$

. Heree $\xi_i(t)$ is a stochastic noise, i.e. a number that varies over time like a random walk (for now let us assume it is white noise, i.e. uncorrelated across time), that can represent within-species processes like individual birth and death, or external factors like weather.

If interactions create less variation over time than the perturbations[3] the dynamics of species i will be driven by this noise $\xi_i(t)$ much more than by ecological features. Otherwise, the effects will be much smaller. If noise is important it will increasingly lead to models that behave like the 'neutral' theory of Hubbell (2001) or to related 'stochastic logistic equation (SLE) models (Grilli). In the case where there is no noise, this behavior only occurs at one point in the deterministic model (see Figure 4) but actual stochasticity expands the area in parameter space (refs?) as we discuss below.

One important consideration is that the resulting behavior depends on whether and how ξ_i might depend on N_i.

Demographic noise

Fluctuations from discrete birth-death events happening at random times can be shown [60] to induce noise that increases like the square root of total population abundance, i.e. it can be approximated here as

$$\xi_i(t) = \sqrt(N_i)\hat{\xi}_i(t) \tag{2.5}$$

---

[3] i.e. $\mathrm{var}\left(\sum_j A_{ij} N_i(t) N_j(t)\right) \ll \mathrm{var}(\xi_i(t))$

As shown in [61] the phenomenology is then very close to the original neutral theory of ecology [62], or models of fixation of neutral mutations[4].

Environmental noise

(i.e. changes in conditions that randomly impact growth rates r_i) gives

$$\xi_i(t) = N_i \, \hat{\xi}_i(t) \tag{2.6}$$

The model then behaves like the Stochastic Logistic Equation studied by [6] or the environmentally-perturbed neutral model studied by [61]. An important difference with the original neutral theory is that fluctuations are much larger and more rapid, so species can go extinct on much shorter time scales.

The noise-free ($\xi = 0$) chaotic phase discussed above is expected to look very similar to these dynamics dominated by environmental noise. To see this, we can first rewrite the dynamical equation as:

$$\frac{dN_i}{dt} = r_i N_i(t) \left(1 - \sum_j A_{ij} N_j(t) + \frac{\hat{\xi}_i(t)}{r_i} \right)$$

Chaos occurs when the term $\sum_j A_{ij} N_j(t)$ acts like a noise perturbing the growth rate, as would environmental stochasticity, thus the consequences are similar to having nonzero $\hat{\xi}$. The main difference is that fluctuations in the chaotic regime are created by interactions with other species, and will depend on the diversity and abundances of species in the community. In particular, these endogenous fluctuations drop to zero when diversity falls below a threshold, see Sec. 2.2.1), whereas $\hat{\xi}_i$ is in principle completely independent of other species. An important consequence of this is that external noise may have to be tuned to some level to preserve species diversity, neither too weak nor too strong depending on the system, whereas endogenous fluctuations (chaos) are “self-tuning” [38].

To conclude, our focal model somewhat subsumes neutral theory and Stochastic Logistic Equation (SLE) as particular noise-dominated limiting cases of the RLV model and highlights how noise enhances the parameter space for behavior that resembles these cases. The framework we describe here also highlights how discriminating among cases where deterministic RLV dynamics drive dynamics into the ’chaotic regime’ from simpler models that are driven by noise. One solution to this may be to focus on how species diversity itself is stabilized (rather than just fluctuations in composition and abundances).

2.4 Spatial (meta-community) behaviors

The introduction of space, bringing us to the full equations (2.1), poses two additional questions:

- How spatial heterogeneity in parameters impacts dynamics and shapes measurable patterns
- How species dispersal between patches can change the dynamics, for instance by rescuing locally extinct species, even in an homogeneous environment where all parameters (carrying capacities, interactions...) are constant across space

The analysis of this full model naturally produces more complex outcomes and regimes and the results of these analyses are poorly studied. Extending ’disordered models’ to address these questions increases the number of critical aggregate parameters quite significantly since we now need at the very least: the variance of dispersal rates D_i (and whether they correlate with other species properties), and the spatial variance and autocorrelation of growth rates or carrying capacities, which represent the spatial features of the environment.

Random Lotka-Volterra metacommunities have been less studied than single communities, so we have a limited array of established results to draw from, and in particular few known behaviors that are truly novel

---

[4] We remark that neutral models originally assume a zero-sum game (the total number of individuals is fixed and divided among species), which can be recovered here (almost exactly) by having identical A_{ij} for all i, j (including A_{ii}). We could also assume non-interacting species $A_{ij} = 0$ for $i \neq j$, in which case the total number of individuals obviously increases with species number, and is much more free to fluctuate over time. Such models, and everything between the two limits, have also been studied under the name of neutral models, see e.g. [63]. Many predictions will be similar across this whole range of models, e.g. abundance distributions [64, 63], while others are obviously different e.g. change of total biomass with diversity.

compared to well-established ecological intuitions. **A significant fraction of the explanations below are therefore speculative.**

2.4.1 Heterogeneous environments

First, without considering dispersal as an additional process, we can simply ask what happens if we let parameters such as growth rates and interactions (or just carrying capacities) vary across the landscape. This question has barely been explored in our focal model, although ongoing work is homing in on this problem.

Here, we have little grounds so far to expect novel behaviors compared to what is already well understood in metacommunity theory – for instance, more heterogeneous environments will allow more different species to coexist at the regional level, etc. At the broad level, we thus expect **no new qualitative regime** from this addition, though of course it could induce many quantitative variations.

2.4.2 Dispersal in a homogeneous environment

As for simple metapopulations, we can roughly separate three regimes of dispersal [65], starting with two extremes. On one hand, there are isolated communities, where dispersal is weak enough that patches are effectively isolated from each other. At the other extreme, a well-mixed system, where dispersal is strong enough that all patches are equalized and synchronized in dynamics, effectively going to a single large well-mixed patch. In-between, there may exist a regime where nontrivial spatial dynamics can occur. The observed regime depends on both dispersal rate and the landscape scale. As discussed in [65], for many models and questions, interesting behaviors happen at the boundaries between these three regions.

Here we focus on how spatial extension and dispersal change dynamics even if the environment is otherwise homogeneous, in the sense that dynamical parameters do not vary between localities. We note that an environment can be considered homogeneous even if has variations, provided they occur on small scales: when the correlation length of heterogeneity is below the typical spatial scale of dispersal, which increases with dispersal rate and can extend over many patches even if dispersal occurs only between neighboring patches, see e.g. [18].

In the diffuse interaction regime, we do not see truly novel spatio-temporal dynamics, but a relatively straightforward extension of previous equilibrium or chaotic regimes. In the equilibrium regimes, species' abundances now track their environmental optimum in the feasible phase, or some nontrivial combination of environment and interactions in the Unique Fixed Point phase ([18], Barbier & Leibold in prep). The main novelty is that adding space makes it more likely to observe the dynamical regimes characterized by abundance fluctuations. As discussed in Sec. 2.2.1, chaotic dynamics tend to extinguish themselves in a well-mixed system[5]. However they can be rescued by having multiple patches, with dispersal strong enough to rescue extinct species but too low to synchronize dynamics; asynchronous dynamics emerge easily, requiring infinitesimally small heterogeneities in the landscape if there are many patches [38]. There is a striking effect at play here that can already be understood from a single metapopulation: a population that has on average negative growth rates in every patch can keep rescuing itself forever when patches are coupled, see e.g. [30] and references therein.

With strong interactions however, we can see new patterns in space: for instance, ecotones separating clearly distinct domains [41], or directional dynamics with succession, community range expansion etc [43]. Thus we potentially enter the region of **patch dynamics** in metacommunity theory [49], with local extinctions and invasions driving strong impacts, see also Sec. 4.2.3.

---

[5] There are very subtle caveats there that we don't understand well: for high extinction threshold, it's mainly transient behaviors (typically leading to much wider fluctuations than those found later in a stationary chaotic behavior when that is observed in metacommunities [38]) leading to all these extinctions; if one lowers the extinction cutoff a lot, it seems that chaotic dynamics are mainly extinguished by deterministic loss of diversity (i.e. species cannot survive in the long term in a single patch unless they could do so at a fixed point). So it is not necessarily chaos killing itself through "activated events" (i.e. large fluctuations touching the extinction threshold), though this may still happen in principle.

Chapter 3

Epistemology

Summary

This (currently short) chapter attempts to demystify the use of randomness and simple approaches to "complex systems", explaining when and why we can expect them to be useful or not, what kind of information they made provide or not, by explaining what is even meant or assumed by randomness and its inclusion in our models.

Ecosystems with many species and processes can seem overwhelmingly complex, but the accumulation of complicated ingredients does not necessarily entail that the whole will behave in complicated ways[1]. We can imagine at least three plausible scenarios:

1. Among all these details, a few matter much more than the rest and we can reduce our focus to those. For instance, body size might be the foremost factor distinguishing aquatic species and structuring their ecological interactions [27], and then this structure very likely (though not necessarily) has to be incorporated into a model for it to successfully predict any phenomenon. Ee can call this "classical simplicity", since identifying such key mechanisms is a traditional goal of ecology.

2. Many or all details matter roughly equally and influence outcomes in different, weakly correlated, directions. Their combined impact will often be simpler than expected (they average out, or create random-like variation), and only certain aggregate properties persist as important causes and consequences – we can call this "emergent simplicity".

3. Many or all details matter in a deeply correlated and interconnected way. We intuit that such a situation is fragile by definition: displacing a single detail can impact all the others in a major way, like a computer program where removing one character will cause the whole to stop working. We can call this "true complexity", and many biological examples exist (e.g. the genetic program in an organism). But we feel it is legitimate to wonder whether ecosystems are sufficiently closed and co-evolved to allow the emergence and persistence of such fragile structures (and whether these structures are the leading forces shaping the behavior of the entire ecosystem).

Our main ideas here are that i) classical and emergent simplicity (options 1 and 2), in the right mix, might suffice to capture many aspects of complex-looking ecosystems, and ii) emergent simplicity (option 2) has not been explored enough compared to its classical counterpart, being an equally interesting starting point to improve our understanding of ecological phenomena.

One aim of this guide is to demystify models of emergent simplicity: show when they are useful, when they teach us something new, but also when they do not, and be honest with any conceptual difficulty in applying them. Such models, including disordered systems, can reveal phenomena that *only* happen when there are many species and details, but also tell us when the behavior of a large heterogeneous community is a straightforward extension of what could be understood from only two or a few species. Emergence is

---

[1] Conversely, simple ingredients do not necessarily lead to a simple system.

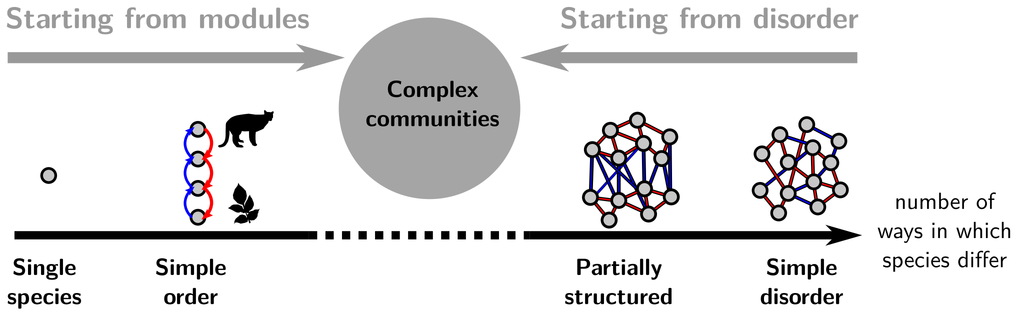

Figure 3.1: Illustrating the argument from high dimensionality for disorder in Sec. 3, i.e. the existence of many factors shaping interactions between species. When species differ importantly in just a few ways, e.g. trophic height, we can organize them in simple modules like food chains. When they differ in a vast number of ways, their interactions become largely unpredictable and, in effect, random-like. Real communities probably lie somewhere in-between, and can be approached from both sides, e.g. modules or partially structured models.

only sometimes magical, and studies of complex-looking models contain many red herrings. We believe a honest and pragmatic presentation will show that only a few new intuitions suffice to capture the main ideas of emergent simplicity and add it to one's intellectual palette.

We now try to explain why this outlook naturally leads to the use of randomness in our models. We propose three explanations (all in some sense plausible and grounded in mathematical arguments developped in other works), in the hope that at least one of them will resonate and provide a working intuition for when it is a good idea to turn to randomness.

3.1 Three ways of thinking about randomness

3.1.1 Objective motivation: a confluence of many factors

As noted above, ecological theory has often focused on situations where a few key parameters matter more than everything else.

On the other hand, when there are many independent sources of variation, their combined effects often behave like random variables. The canonical example is a dice throw where many deterministic microscopic factors (spin, speed, orientation, angle to substrate, etc) actually determine the outcome but interfere so much with each other that none of them dominates, and the result is effectively random.

In other words, randomness can be seen on objective grounds as an "attractor" in the space of all possible phenomena, in the sense that more and more complicated systems will often tend to act more and more like randomness generators.

We may treat pairwise species interactions as random if they are sufficiently *independent*: if the impact of one species on another is not predictable from the interactions of these two species with other ones. In a strict competitive hierarchy, or in a food chain, we can rather easily extrapolate how species A & C will interact if we know how A & B and how B & C interact. If however, say, A & B interact mainly through light limitation and B & C mainly through pathogens, or some complex and context-dependent combinations of many such factors, then we cannot truly predict the outcome of any interaction from the knowledge of others.

Being random-like simply requires that the many interactions in the network are sufficiently uncorrelated, which is a likely outcome if these interactions arise from many different traits and factors.

A subtler take: Even if each individual interaction is in a sense deterministic and predictable, their longer-term outcomes in a complex community might still be as if they were random. If there are many pathways and modes of interaction, a species can have a negative direct impact on another (e.g. a predator on its prey), but a positive indirect impact through other variables (e.g. a predator by recycling nutrients). Many positive and negative indirect effects running along various paths through other species can end up interfering with each other, and acting as random noise.

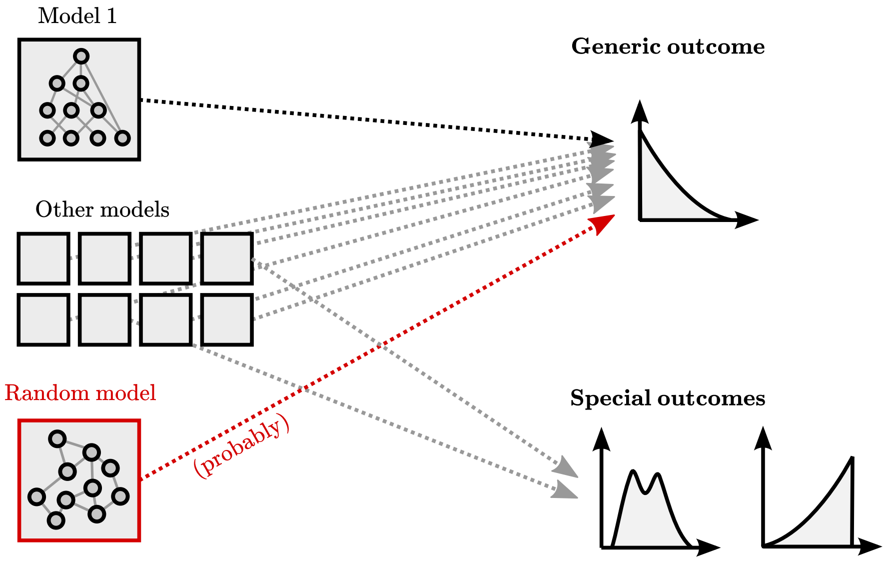

Figure 3.2: Illustrating the argument from robustness in Sec. 3, i.e. the idea that we may use random models because they are a simple way to access "generic" outcomes that we expect to arise in many nonrandom models [21]. Drawing parameters at random is thus different from 'exploring all the possibilities' – on the contrary, for large numbers of parameters, it is unlikely to stumble upon any non-generic possibility.

If however species are organized in sufficiently distinct groups, roles, axes, etc., these direct and indirect paths all sum up into a coherent impact and we cannot use this randomness assumption, e.g. as trophic levels, all paths between a top predator and a plant tend to be correlated in their impacts: the top predator may eat multiple carnivores, each of which eat multiple herbivores, but it all results in a cosnistently negative impact on the plant.

3.1.2 Heuristic motivation: baseline models for robust phenomena

Even if we do not accept randomness as being "objectively" a good representation of the dynamics we are modelling, we can still adopt it on heuristic grounds. We admit that we are dealing with intricate complex systems, possibly devoid of any true randomness, but we only focus on behaviors and patterns that are robust to these details because we believe these predictions would also hold even if we were to make changes in the details of the studied system (e.g. shuffle carying capacities, growth rates, or interaction coefficients). It is then plausible that the same phenomena or patterns will be very frequently be observed if these details are drawn at random and deviations would require very specific changes that would be very unlikely (Fig. 3.2). Choosing a random model is useful simply because it is easier to manipulate than most non-random ones with the same behaviors, and it thus serves as a "baseline" model for this aspect of the dynamics.

3.1.3 Subjective motivation: aggregation and statistical equivalence

A third option is a subjective Bayesian viewpoint: 'Randomness' simply reflects the degree to which we are uncertain or uncaring about certain details.

Aggregating implies that the variables we are aggregating over (e.g. the abundances of various species) are, in some sense, equivalent or exchangeable – that we do not mind adding apples and oranges if what we care about is total fruit availability. This does not mean that these variables are identical in every way, but only that none is "special" in how it contributes to the aggregate.

This clearly depends on which aggregate pattern is our focus: it may seem strange to add together the biomass of predators and prey into a metric of total biomass given that they have very different ecological implications, but some predictable patterns (e.g. how much fish biomass can we extract from an ocean, regardless of their trophic level) assume an equivalence between them even though others (e.g. how efficiently is sunlight converted into fish biomass) do not.

Random models (and neutral models as a subcase) appear like a natural choice when we believe that we are indeed justified in treating species as statistically equivalent for a given pattern. In a random interaction network, species are not identical, but no species occupies a very unique role – even a well-connected “hub” species may not be a remarkable outlier because it occurs within the expected distribution of means and variances, but can be modeled as a representative sample of the overall community’s distribution of connectedness. We give a more concrete discussion of this notion of statistical equivalence in Sec. 1.4.2.

Chapter 4

Empirics

Summary

This chapter has two aims: it attempts to synthesize the state of the art on the question of "how well our models do in the world", and then offers a larger shopping list of observables one could measure empirically and confront to theoretical predictions. Many empirical observations seem compatible with a model of noninteracting species driven by environmental noise *or* interacting species in a chaotic regime – it is not easy however to discriminate between these two options (or some combination of the two). Experimental response to perturbations might be the most approriate and informative way to probe community structure, but this probing is more subtle than in other disciplines because we rarely know how much within-species processes are impacted by perturbations. I list various other caveats and issues that come with any attempt at confronting theory and data, to explain why there have been few, if any, smoking guns so far for disorder and similar complex systems approaches in community ecology.

Empirical community ecology often focuses on particular species/taxa and factors (e.g. cyanobacteria and nitrogen). By taking a disordered systems approach, we choose to lose the labels and focus only on observables that are somehow aggregated over individual species, samples and times – typically distributions (e.g. the distribution of species abundances in a community), statistics of these distributions (e.g. the variance or unevenness of species abundances), and relations between such quantities or statistics (e.g. Taylor's law relating mean and variance of abundances).

Furthermore, since we are employing LV-type models (1.2), all the patterns that can be confronted to theory must be constructed entirely from:

- model variables, typically abundances $N_i(x,t)$ (or perhaps "dynamical importance", see below),
- model parameters (number of species S, growth rates r_i, interactions A_{ij}...)
- perturbations $\xi_i(x,t)$
- observation parameters: number of samples, time range, spatial scale...

As we noted in various places, this means that we cannot very satisfactorily address patterns of ecosystem functioning beyond rather simple ones.

4.1 Stylized facts

Epistemic status: Rough approximations & speculation, but most of the facts are truly salient in natural observations (i.e. I would be surprised by any many-species observational dataset not conforming to them). I discuss differences with experimental settings whenever relevant.

Here we summarize some qualitative facts that often emerge from empirical observations across a variety of ecosystems, and how they may be explained in the disordered LV framework. This section refers to the theoretical regimes and predictions detailed in Chapter 2.

Many-species coexistence

Epistemic status: Peculiarly not easy in experiments with simple environments e.g. a Petri dish with a simple culture medium.

A rather massive literature addresses the question of coexistence mechanisms. But very often, the proposed mechanisms allow a few species to coexist, or can multiply the number of species that can coexist by a moderate factor (uncoutable examples including [34]). Far fewer mechanisms are actually credible for going from a few to a truly large number (hundreds, thousands...) of coexisting species.

It's broadly understood that there are two main options: either **fitness equalization**, meaning that no species has a significant advantage (from single-species dynamics and multi-species interactions) over any other across *all* space and time, or **self-regulation**, meaning that the competitive impact of a species on others is relatively weak compared to its impact on itself, and thus, no matter its advantage, it will eventually limit itself enough to not overtake others.

In slightly more detail, we can find four regimes within our model that could sustain large diversity:

- No overall fitness advantage: species have comparable growth advantages (at least in some sites or part of the time), either because they are truly similar in traits and interactions, or because they are sufficiently close to a trade-off surface (for instance, species may all fall on a spectrum from faster grower to stronger competitor, or between various resource consumption strategies, with the advantages of each being well-balanced so that none of them takes over)
- Diffuse interactions (possibly as an effective result of spatio-temporal averaging, i.e. interactions between individuals could be very strong, but the community-level impacts could be much weaker, e.g. due to species being segregated into many microhabitats)
- Sparse interactions (likewise possibly as an effective result, e.g. interactions concentrate among few species pairs that are very often present together)
- And an idea (at this stage less well-defined for me) of coexistence fundamentally driven by spatial or temporal heterogeneity (e.g. out-of-equilibrium fluxes) despite strong interactions and large fitness differences. It is not entirely clear to me that this truly won't be morally equivalent to one of the former three options when properly aggregated over time and space, despite a significant corpus on nonequilibrium coexistence [66, 67].

The main alternative to these four explanations is strong structure, but it is not obvious to identify structures that lead to many-species coexistence without effectively boiling down to one of these options.

Wide abundance distributions

As discussed in Sec. 4.2.1, realistic abundance distributions are usually very wide (power-law or log-normal).

Among possible explanations within our framework:

- Chaotic or noise-dominated regimes produce the right kind of distributions: no species is truly advantaged forever, but the rich-get-richer mechanism of exponential growth allows some lucky ones to blow up at a given time
- Otherwise regimes where differences between species arise mainly from interactions would give narrower distributions of the dynamical variables, but it may be that measured abundances are very bad proxies for these variables (Sec. 4.2.1)
- Otherwise, regimes where differences between species arise mainly from the environment can be arbitrarily tuned to produce wide abundance distributions (i.e. we can postulate a wide distribution of

carrying capacities K by fiat, since we are not modelling where those K come from). Species interactions in that case can only be very small corrections on what is imposed by the environment – if they were significant, then the environmentally-advantaged species should wipe out all the others[1]

Few dominant species

Many experiments and observations across biomes and organism types suggest that, despite a large diversity, a large fraction of abundance or biomass is dominated by very few species. This can fit with the strong interaction regime in general.

Large abundance fluctuations

All the chaotic or noise-dominated regimes (see [16] for possible evidence of a mix of demographic and environmental noise at different levels depending on species and population sizes)

Wide distribution of ranges

There is significant evidence that a few species are dominant over a large fraction of the landscape (or region, or set of samples...). This is not necessarily what would be obtained by simply coupling many patches displaying chaotic fluctuations, where there is no guarantee for the same species to prevail in many places at once (depending on the strength of coupling of course).

Hence

- a chance explanation: dominance is truly by chance and spatial coupling is strong enough to synchronize it or entrench the winner (this would however tend to help extinguish chaotic fluctuations [38])
- a deterministic explanation: the carrying capacities or interactions deciding who comes up on top are consistently biased in favor of the same species in many different places. While we can of course draw the average carrying capacity for each species as a random number across the whole landscape [6], there is no clear explanation for doing so without positing

Simplifying the system leads to more dominance, less diversity

Epistemic status: This is not as crisp an empirical fact as many of the other ones, which simply require looking at bare data.

Many systems seem to lend themselves to observations along the following lines: adding much of a particular resource often tends to dramatically favor one species over others, usually leading to a reduction of diversity together with an increase in total biomass. On the contrary, adding a source of mortality tends to lower total biomass but increase diversity.

This might, if it is truly a broad observation, lend some credence to the notion of the environment's "complexity" being a key factor in diversity, and that species may often be near some sort of trade-off unless some external factor suddenly advantages one. This kind of response to perturbations is not something that is readily obtained in a model with purely random parameters, but it may require a particular structure underlying carrying capacities, interactions and perturbations, thus perhaps pushing us away from a completely disordered Lotka-Volterra model and toward one structured by (poorly-defined) "resources" or "functions", see Chapter 1.

Putting it all together: what is the most plausible regime within our theory?

Disclaimer: there is no obvious reason why a single regime of our model should be the most representative of everything that is found in nature, even when some stylized facts seem to recur.

---

[1] Unless interactions are somehow rigged to correlate negatively with environmental advantage, which typically would mean that abundances are not the correct variables to reflect dynamical importance (very abundant species would need to have a much weaker per capita impact, suggesting that per capita is not the best way to compare interactions).

In terms of accounting for diversity and fluctuation patterns, the most obvious candidates are either the regime dominated by environmental noise (where species interactions do not matter much or only for some marginal patterns) or the diffuse or strong chaotic regimes (where species interactions are at the heart of the dynamics).

The first type of explanation is effectively "passing the buck" of explaining why species abundances and dynamics display certain features, since it now relies entirely on a correct description of the environment's own statistics and variations, rather than on the interplay between the variables in our (LV-like) dynamical model. This viewpoint, embodied e.g. in sampling models [6] that do away with interactions entirely, can certainly account for some of the observed phenomenology, and looking more into precise dynamical features (e.g. perturbations and responses) is likely necessary to discriminate reliably between the two options.

One example feature that could help identify the chaotic turnover regime is that we expect decreasing temporal correlations between species with increasing diversity, as seen in [68] (who give a different explanation: dilution effects). Another might be observing that species invasions lead to change in the composition of the existing community [69].

Among the less obvious and more speculative consequences of the dilemma between the two regimes listed above is the question of whether a description in terms of species is most appropriate: if the patterns we are seeing are best explained by exponential tendencies modulated by changing biotic contexts, then we may posit that the central dynamical process to take into consideration is reproduction (the engine of exponential growth), which is what unifies species – or at least local populations – and demarcates them strictly from each other. If however environmental fluctuations are the main driver, it is not obvious that species are the right grain – there is usually significant variation in environmental preferences within a species, and similar preferences across species, therefore it could become more relevant to use a model of ecotypes.

4.2 List of patterns and predictions

4.2.1 Abundances only

Many sources of data in community ecology take the form of a distribution of species abundance (or at least presence/absence) for various species in various samples across space or time. Several patterns or distributions can be constructed from that kind of dataset [6].

We present the theoretical predictions for these patterns, but we must also reemphasize the major caveat that the variables whose distribution we are describing are not necessarily truly the species abundances, but rather their dynamical importance, which could be instead tied to e.g. biomass or relative yield or resource consumption, see Sec. 1.4.2.

Abundance distributions

We show in Table 4.1 the basic predictions of the RLV model in terms of distributions of the variables N_i. Most empirical data supports a Gamma or otherwise wide distribution, rather than a narrow distribution (e.g. Gaussian or truncated Gaussian)

This may seem to rule out certain regimes of the RLV, which produce Gaussian distributiosn. Yet it has sometimes been suggested that the "correct" variables (see Sec. 1.4.2) such as relative yields N_i/K_i, may indeed follow a Gaussian distribution [25], with the K_i themselves being widely distributed, for reasons that require further explanation. Since we often lack knowledge of the K_i, the information of the N_i alone is thus not a surefire way to rule out the corresponding dynamical regimes.

Additionally, a lot of work around neutral theory has consisted in trying to validate and understand the claim that SADs are sometimes bimodal, with a cluster of abundant species and a long distribution of low-abundance ones. This is somewhat similar to what can be observed in RLV in the chaotic turnover regime, or in the spatial case with dispersal, though evidence is spotty and complicated.

We will not spend much time discussing the various diversity and evenness metrics that can be extracted from the abundance distribution – e.g. mean and variance of abundances, Hill numbers, ... as they represent less complete information (but they might be more robust to model details)

Regime	**SAD****
Feasible	Narrow (Gaussian) [43]
Unique Fixed Point	Narrow (truncated Gaussian) [43]
Chaotic turnover	Wide (truncated power-law, Gamma) [48]
Noise dominated	Wide (Gamma) [70]
Multistable phase	Unknown, likely narrow [37, 48]

Table 4.1: Predicted abundance patterns (* Species Abundance Distribution, or distribution of whichever variable actually matters in the dynamics, see main text.

Abundance co-distributions across space/time

McGill [8] has proposed that different mechanisms will generate the same commonly-studied spatial patterns if they lead to the same "stochastic geometry", i.e. an effectively random distribution of individuals in space, where only a small number of statistical properties matter.

A particularly simple type of stochastic geometry might be one that (i) specifies the degree of clustering of individuals of the same species, (ii) specifies the (e.g. lognormal) distribution of species abundances, and (iii) otherwise assumes species to be independently distributed in space. This idea has been picked up and discussed and tested empirically in recent articles by various groups [71, 6, 68].

We can discuss when a given type of stochastic geometry is expected to arise from a Random Lotka-Volterra metacommunity. For instance, the above three assumptions are actually quite likely to hold in the chaotic regime with diffuse interactions, since temporal fluctuations are highly uncorrelated between species – despite the fact that these fluctuations only exist due to the very interactions between all these species. But the parameter regime that best maintains chaos is one where dispersal does not strongly couple local populations of the same species, so intraspecific spatial correlations should be rather limited. It remains to be seen whether empirical values of dispersal and spatiotemporal correlations will prove compatible with this chaotic regime.

4.2.2 Abundances and parameters

If we have independent ways of measuring the parameters in our equations, i.e. carrying capacities K_i, interactions A_{ij} etc., there are more patterns we can measure: for instance, theoretical predictions on how parameters for coexisting species differ from parameters for non-coexisting ones [72].

With a few exceptions (like the experiment analyzed in [25]), this is however rarely possible in terms of empirical data. In most cases, we cannot directly measure all the parameters in our model, as that would require at the very least $O(S^2)$ experiments: growing all species in isolation to get the K_i, all species pairs to get the A_{ij}, and some more combinations to test the model.

Since disordered systems predictions do not require full knowledge of the parameters, we may still hope to obtain enough partial knowledge to test some theoretical results, e.g. sampling only some of the pairwise interactions to estimate their mean and variance, and. To our knowledge, no test of RLV following this principle has succeeded quantitatively, but qualitative agreement between theory and experiments was found in [51], where a few of the regimes in Chapter 2 were observed.

4.2.3 Abundances and perturbations (stability)

In Chapter 2, we classified dynamics in the RLV model by their **qualitative stability**: in the absence of perturbations ($\xi = 0$), would the dynamics tend to a single stable equilibrium, to persistent fluctuations, or could it fall in various attractors depending on past history?

But we may also ask about the **quantitative stability** of a community: how much does it respond to perturbations $\xi_i(t)$, how evenly between species, etc. Many quantitative metrics of stability have been proposed to characterize how stable or unstable a community is, accounting for its response to various such perturbations [73] e.g.

- Press perturbations, i.e. ξ_i constant over time (e.g. a new permanent source of growth or mortality is added, like humans harvesting some species at a fixed rate)

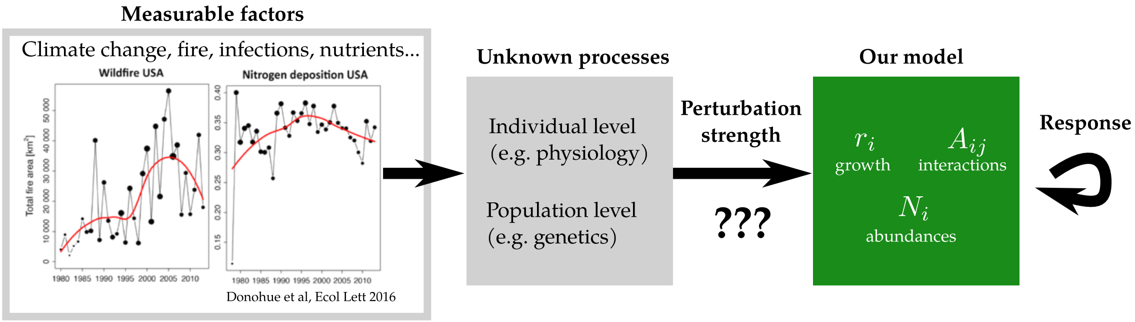

Figure 4.1: The problem with trying to figure out how much an ecological community's dynamics is driven by external factors versus species interactions, or using perturbations to probe its structure: while many aspects of a perturbation are measurable, we almost never know the actual impact on our model variables or parameters (with the exception of perturbations destroying a known percentage of a population, e.g. forest fires, and even then there are subtleties).

- Pulse perturbations, i.e. instantaneous impacts $\xi_i(t) = \delta(t-t')\xi_i$ (e.g. a sudden fire or resource bloom)
- Noise $\xi_i(t) = W(t)\xi_i$ with W a random walk (e.g. weather fluctuations)
- Periodic perturbations $\xi_i(t) = \sin(\omega t + \phi)\xi_i$ (e.g. night-day or seasonal cycles)

as well as species invasions or extinctions. These perturbations represent impacts on the population dynamics, and we are interested in how demography then plays out (e.g. how the community recovers from, or even amplifies, the perturbation).

A systematic way of thinking about stability and comparing it across communities is still lacking but see [74, 75] for intuitions, a few of which are summarized below and in Appendix.

The problem of measuring stability

A major problem arises when trying to compare ecosystems in terms of stability. We often take time series $N(t)$ from various ecosystems, look at their temporal variability or total displacement or some such metric, and use this to conclude than an ecosystem is more or less stable than another.

But it almost never occurs that we really know how strong the perturbation ξ was that induced the observed variation in abundances N. As illustrated in Fig. 4.1, when we incorporate a disturbance as $\xi_i(t)$ in our equations, it is almost always the phenomenological result of physiology, resistance, demographic phenomena within each species, etc. for which we do not usually have good quantitative understanding.

Meaningful metrics of stability, i.e. those we can use to estimate interaction strengths or other processes of interest, are usually not so much about "how much an ecosystem moves in total" and more about "how much an ecosystem moves relative to how much you perturb it", hence expressions like $dN/d\xi$ in Table 4.3.

Predictions that can be tested are more qualitative, such as the existence of relations between various stability properties.

Perturbations as probes

The response of a community to perturbations is an essential prism through which we can understand what is happening inside. For instance,

- when a perturbation ξ_i selectively targets some species i (e.g. a disease, an influx of a specific resource...), we can study whether its impact on another species j is weaker or stronger than their impact on the targeted species, to have a measure of how strongly interacting these species are [76]. The "amplification" of the response, e.g. the ratio of $dN_j/d\xi_i$ to $dN_i/d\xi_i$ in the case of press pertubations, is somewhat robust to the fact that we do not know the magnitude of ξ_i (in the linear limit it is independent of that magnitude).

  One complication, however, is that the two species may have different sensitivities or "biomass scales" (e.g. the abundance N_j of plants may be much larger than the abundance N_i of herbivores, then we might expect dN_j to always be larger than dN_i). Thus, we may only be able to reliably probe more robust quantities, such as the product

$$\lambda = \frac{dN_i/d\xi_j \; dN_j/d\xi_i}{dN_i/d\xi_i \; dN_j/d\xi_j}$$

which represents how strong the feedbacks through other species are compared to feedbacks of a species on itself (e.g. in a two-species LV, it would be $\lambda = A_{ij}A_{jj}/A_{ii}A_{jj}$, but in a more complex model it is a bit more subtle)

- this notion can be generalized into several metrics of how "collective" the dynamics are within the network of interacting species: e.g. how far the perturbations "propagate" away from the original speices, and how predictable the long-term responses are from short-term ones [77]
- studying "invasion" events when a species is added to the community is also a way of probing community stability [78], and allows us to define whether a community is "saturated", i.e. behaves as if it can sustain a maximum diversity [79]

Relationships between stability properties

In an arbitrarily complex community, the sensitivity to perturbations of diverse types, or applied to distinct species, could be very different (e.g. a perturbation on a top predator or basal species may have widely divergent impacts). In a disordered community however, we expect stability properties to be connected in predictable ways. We give some examples of these relations:

- the species' responses to press perturbations and invasions are tightly connected [78] and tell us about qualitative stability as well: the transition to chaos is seen in a dramatic increase in the sensitivity to a zero-mean press on carrying capacities [43]

Stability & ordering by abundances

In disordered systems, species contribute differentially to stability properties in a way that is largely determineed by the species' abundances, or rather by relative yields $\eta_i = N_i/K_i$ [75]. More precisely, some stability indicators, despite being computed at the community level, will mostly reflect the stability of the rarer species, and others that of the more abundant species.

For instance, it can be shown (see Appendix) that the commonly used stability indicator known as *asymptotic resilience*, i.e. the long-term rate at which a community returns to equilibrium, is very likely to simply be the return rate of the rarest species. As such, this metric does not clearly represent a "community property": the rest of the community could be much more stable than suggested by this metric [80, 74].

The fact that rarity or commonness are such strong predictors of a species' role in community stability is because a disordered community does not assign special positions to species. In a structured community, contributions to stability may have more to do with the particular role of a species, e.g. a top predator species may have disproportionate importance despite being rare, because of where it sits in the structure. A random network offers no notion of "where" a species sits, except a statistical one that can be summarized as: is the species overall favored or disfavored by all the other interactions in the network, i.e. is it abundant or rare. Being lucky or unlucky is a collective, context-dependent property tied to the whole network: if one species is more abundant than another in one community, the reverse is as likely to hold in another community, even if the environmental conditions are all identical.

This prediction has not been tested empirically yet, to the best of our knowledge, but it could be used to rethink attempts at connecting various stability metrics [73], at least inasmuch as we expect diverse metrics to be correlated only when they are probing the same (rare or abundant) subset of species.

Invasions and extinctions

In our many-species disordered models, it is natural to ask what happens to the community on average when any one species is added or removed. This can be closely connected to the community's response to press perturbations, as adding or removing a species is like a permanent change in the growth or mortality of all the species it interacted with.

In the diffuse interactions regime (Sec 2.2.1), the individual impact of any species on any other is rather weak. Thus, if an invasion or extinction causes secondary extinctions in the community, the latter will only concern rare species[2], simply because any perturbation can push them over the edge, not due to particular interactions between them and the added or removed species.

For strong interactions, we can understand under which conditions invaders are likely to cause more dramatic transformations to the community, see [78]. This happens when the invader experiences a strong positive feedback, e.g. when it can grow in abundance, kill some competitors which leads to a more favorable environment with even faster growth, and so on. This, and the emergence of "keystone" species whose extinction has dramatic impacts, can be understood by looking at the distribution of species' susceptibilities to press perturbations, and checking when they diverge or even become negative, as discussed in [78] and follow-ups (in prep).

4.2.4 Relations between properties across communities

Many commonly studied patterns are relationships between some of the observables listed above, i.e. how they co-vary across communities that differ in some basic ways, as illustrated in Table 4.2.

Pattern	Definition	Example studies
Species Area Relationship	S^* vs area	[7]
Diversity-variability	$\mathrm{var}(N)$ vs S^*	[81]
Diversity-invadability	invasion success vs S^*	[82]
Dissimilarity-Overlap	$\mathrm{corr}\left(\vec{N}(x), \vec{N}(y)\right)$ vs $\mathrm{corr}(\Omega(x), \Omega(y))$	[16]

Table 4.2: Some commonly studied composite patterns: relationship between pairs of observables (defined in Table 4.3) across some axis of variation between systems. (Note: $\vec{N}(x)$ is used to designate the vector of all species abundances in a community x, while $\Omega(x)$ designates the set of present species).

Inference from patterns

Using relationships between diversity and other properties has been one of the main ways of testing the disordered model and inferring its parameters.

By and large, one of the most easily and robustly inferred properties is average interaction strength, as it can be deduced from the relationship between species number S and total biomass [83].

One important test that would lend strong credence to the Random-Lotka Volterra model is that, if ecological fluctuations stem from species interactions as in the chaotic regimes, we would expect these fluctuations to become stronger and less correlated between species in communities with strong interactions or larger diversity. This has been found in at least one experiment [51] but field evidence has not been accrued yet[3].

We can ask two questions:

- Why are some patterns (e.g. SAD, SAR) far more studied than others?
- Is it useful to test many patterns at once, can we improve our ability to infer model parameters?

Appendix: Warnings and caveats for theory-data comparisons

An important direction to be explored is whether certain basic or composite patterns are more robust to various types of error, noise, uncertainty:

---

[2] Typically species whose abundances are of order μ/S or $\sigma/\sqrt{S}$.

[3] See notably evidence in [68], although they give a different explanation based on dilution effects.

- Measurement errors and biases
- Wrong variables, see Sec 1.4.2

While we can always try to assume the error distribution and infer the real pattern as latent variables, it is probably much safer to focus on patterns that are intrinsically less affected by these errors.

As an example, relative properties, such as relative change of biomass in response to a perturbation, are potentially more robust to errors of units/variables, but more prone to errors due to taking a ratio (which is bad when the error bars of the denominator can touch zero, for instance). This also connects to the problems of normalization discussed below.

Log or no log

Since abundances often vary over orders of magnitude, we may wonder whether it is meaningful to measure the statistical moments (mean, variance, etc.) of abundances, or whether we should look at the moments of their logarithms. Indeed, when random variables have a wide, fat-tailed distribution such as a lognormal or power-law, the mean and variance may be largely controlled by one or a few very large values, and thus not be good indicators of the whole abundance distribution.

This is not as simple it may seem, because the usual (arithmetic) mean and variance do matter for the dynamics – they determine the amount of interaction experienced by other species, for instance, or the flux of migrants going to other sites through dispersal[4].

It may be necessary to have a theoretical understanding in order to decide whether to take the log or not in specific cases: for instance, we see in [46] that, in the diffuse-interaction chaotic regime, the "correct" variables to employ to understand the dynamics are the bare abundances for the most abundant species, and the log abundances for the other ones.

Normalizing observables

Many properties must be normalized to allow for "fair" comparison between ecosystems. For instance, if we compare biomass production across ecosystems, we may want to normalize by something like system size, to have a notion of how efficient each ecosystem is relative to its size. Such a normalized metric may then be "scale-invariant" in the sense that we can meaningfully compare it across measurements done at different scales.

The choice of how to normalize is self-evident only when considering linear/additive metrics, such as biomass (the biomass of a system is always the sum of biomasses of subsystems). In that case, normalizing is always equivalent to taking an average.

There is no correct way to normalize a property like variability. Consider for instance: if variability is due to demographic noise, then we expect $\mathrm{var}(N(t))$ to be proportional to N(t) (see 4.2.3), so the correct normalization is to divide the variance by the total biomass. If variability is due to environmental noise, then var(N(t)) is proportional to $N(t)^2$, so we should normalize by the square of total biomass. Choosing the wrong normalization here will automatically mean that variance decreases or increases with total biomass (and thus with scale) regardless of anything else.

What this means is that "how to normalize" is not a property of the metric, but a property of the process: some processes will give scale-invariant metrics once normalized by $\sum N$, others by $(\sum N)^2$, others by something else or nothing at all. Only linear metrics escape this because linearity automatically entails that one should normalize by $\sum N$.

Appendix: Table of observables

Involved variables	**Example observables**

---

[4] Indeed, it's a recurrent theme in these models that counter-intuitive aspects arise from a subtle interplay between arithmetic and geometric means, see e.g. [30, 84].

Abundances $P(N)$, composition Ω	$N_{tot} = \sum_i N_i$ $S^* = \sum_i (N_i > 0)$
• Snapshot / Equilibrium	Fluctuations: $\mathrm{E}_i(C_{ii})$
• Nonequilibrium stationary $C_{ij} = \mathrm{cov}_{x,t}(N_i, N_j)$ $R(t,t') = \mathrm{cov}_{x,i}(N_i(t), N_i(t'))$	Autocorrelation time $\mathrm{E}_{t,\Delta t}(\log \lvert R(t+\Delta t)\rvert / \Delta t)$ Avalanche distributions [85] Spatial patterns [86]
+ **Parameters** $P(N,\theta)$	Relative yield $\mathrm{E}(N_i/K_i)$ Properties of survivors [72]: - $\mathrm{E}_i(K_i \vert N_i > 0) - \mathrm{E}_i(K_i)$ - $\mathrm{E}_{i,j}(A_{ij} \vert N_i, N_j)$[25]
+ **Perturbations**	
• Press ξ_i Susceptibility $\chi_{ij} = dN_i^*/d\xi_j$	$\lvert\Delta N\rvert = \mathrm{E}_i\big((N_i(\infty) - N_i(0))^2\big)^{1/2}$
• Extinction $N_i^* \to 0$ $(\xi_j = -A_{ji}N_i^*)$	Initial, median, final return rate to eq.
• Invasion $0 \to \epsilon$	Variability $\mathrm{E}_i(\mathrm{var}(N_i(t)))$
• Pulse $\xi_i(t) = \delta(t-t')\xi_i$	Taylor's law $\mathrm{var}(N_i)$ vs $\langle N_i \rangle$ and other stability-abundance relations (maybe involving N_i/K_i instead) [75]
• Noise $\xi_i(t) = W(t)\xi_i$, $\mathrm{cov}(\xi_i(t), \xi_j(t)) = \Sigma_{ij}$	
• Periodic $\xi_i(t) = \sin(\omega t + \phi)\xi_i$	

Table 4.3: Main observables: indicators (scalar numbers) aggregated over individual species and locations.

A few take-aways

I conclude with a small selection of take-away messages from all chapters:

- When dealing with many-species, high-dimensional systems, the assumption of disorder or emergent simplicity entails that many model details are of limited importance, so we can make robust predictions (qualitatively, sometimes even quantitatively).
- An important consequence is that a model like Random Lotka-Volterra has very few fitting parameters: theory predicts (and simulations confirm) that the details of interactions matter only through simple statistics such as mean and variance [21]. Where we might otherwise have assumed simple (e.g. Gaussian) interaction distributions for practicality, here we are guaranteed that there is truly no need to go to a more complicated one (under known conditions).
- The Random Lotka-Volterra model is a possible baseline to study a broad class of models that all share a few premises: population growth (and especially its exponential tendency) is the most central driver of the phenomena we are interested in, and all other processes (e.g. species interactions, environment, chance...) contribute in relatively simple ways to that driver. Within this class of models, RLV potentially encompasses, qualitatively or even quantitatively, the predictions of a number of other approaches: most evidently the Stochastic Logistic Equation [6], but also neutral theory [62] (except for speciation and phylogeny-related questions), etc.
- On the other hand, this model is not the be-all and end-all of disordered approaches in community ecology. We could imagine that different variables and processes would be a stronger driving or limiting force for community patterns, and modelling species abundances and their exponential tendencies would not provide much explanatory power. Each is a different way to "flatten" the complexity of an ecosystem, and many others e.g. focusing on environmental spatial structure or on metabolic networks, could still lead to disordered models.
- In a disordered interaction network, species do not have *a priori* roles like top predator. Instead, the most salient emergent outcome is that some species are lucky or unlucky, and thus abundant or rare (in some sense, usually relative to their carrying capacities). Therefore, many properties, such as how a species contributes to dynamics and stability, can be predicted by its rarity. Being lucky or unlucky is a collective, context-dependent property tied to the whole network: if one species is more abundant than another in one community, the reverse is as likely to hold in another community, even if the environmental conditions are all identical.
- Disordered LV-like dynamics fall into a few broad regimes or archetypes that dominate in distinct parameter regimes, including equilibrium-like, chaotic, multistable and noise-dominated ones. Chaotic and directional dynamics in many-species Random LV are very different from preexisting ecological intuitions of few-species chaos or deterministic (Clementsian) succession. They are strongly tied to species diversity: they rely on the exploration of many different combinations of species, either undirected in time (chaos) or directed on average by a metric of maturity (succession) even though individual jumps between compositions are largely unpredictable.
- Among empirical facts, there is a lot of focus on patterns of inequality between species (abundances, ranges, etc.). These admit multiple competing explanatiosn within the Random-Lotka Volterra model: strong inequalities could arise from broadly-distributed carrying capacities or environmental "fitness"

(as in [6, 10]), from noise like neutral theory [62], or from many-species chaotic dynamics. Several other stylized facts are compatible with both explanations focusing on the environment and explanations focusing on interactions, and finding reliable ways of detecting where real ecosystems may fall on that spectrum may be the most obvious next-step challenge for this approach.

- The other obvious challenge is going beyond pure disorder, i.e. representing many-species communities by a combination of a few salient deterministic features (such as each species' position along the food chain, in the case of trophic networks) and many other complex features that can be treated as effectively random. I currently believe that the right mix of disorder and simple structure could allow us to understand a large variety of phenomena observed in real ecosystems [21], but a general method for applying this idea to empirical communities and data is still lacking.

Bibliography

[1] Imane Akjouj, Matthieu Barbier, Maxime Clenet, Walid Hachem, Mylène Maïda, François Massol, Jamal Najim, and Viet Chi Tran. Complex systems in ecology: A guided tour with large lotka-volterra models and random matrices. *arXiv preprint arXiv:2212.06136*, 2022.

[2] Matthieu Barbier and Jean-François Arnoldi. The cavity method for community ecology. *bioRxiv*, page 147728, 2017.

[3] Ian A Hatton, Ryan F Heneghan, Yinon M Bar-On, and Eric D Galbraith. The global ocean size spectrum from bacteria to whales. *Science advances*, 7(46):eabh3732, 2021.

[4] Joy Bergelson, Martin Kreitman, Dmitri A Petrov, Alvaro Sanchez, and Mikhail Tikhonov. Functional biology in its natural context: A search for emergent simplicity. *Elife*, 10:e67646, 2021.

[5] Clarence Lehman, Shelby Loberg, and Adam Clark. *Quantitative Ecology: A New Unified Approach.* University of Minnesota Libraries Publishing, 2019.

[6] Jacopo Grilli. Macroecological laws describe variation and diversity in microbial communities. *Nature communications*, 11(1):1–11, 2020.

[7] John Harte. *Maximum entropy and ecology: a theory of abundance, distribution, and energetics.* OUP Oxford, 2011.

[8] Brian J McGill. Towards a unification of unified theories of biodiversity. *Ecology letters*, 13(5):627–642, 2010.

[9] Silvia Zaoli and Jacopo Grilli. A macroecological description of alternative stable states reproduces intra-and inter-host variability of gut microbiome. *Science Advances*, 7(43):eabj2882, 2021.

[10] Silvia Zaoli and Jacopo Grilli. The stochastic logistic model with correlated carrying capacities reproduces beta-diversity metrics of microbial communities. *PLoS computational biology*, 18(4):e1010043, 2022.

[11] Matteo Sireci, Miguel A Muñoz, and Jacopo Grilli. Environmental fluctuations explain the universal decay of species-abundance correlations with phylogenetic distance. *bioRxiv*, pages 2022–07, 2022.

[12] Villy Christensen and Carl J Walters. Ecopath with ecosim: methods, capabilities and limitations. *Ecological modelling*, 172(2-4):109–139, 2004.

[13] H Resit Akcakaya, Roger Arditi, and Lev R Ginzburg. Ratio-dependent predation: an abstraction that works. *Ecology*, 76(3):995–1004, 1995.

[14] Matthieu Barbier, Laurie Wojcik, and Michel Loreau. A macro-ecological approach to predation density-dependence. *Oikos*, 130(4):553–570, 2021.

[15] Maria Dornelas, Laura H Antao, Faye Moyes, Amanda E Bates, Anne E Magurran, Dušan Adam, Asem A Akhmetzhanova, Ward Appeltans, Jose Manuel Arcos, Haley Arnold, et al. Biotime: A database of biodiversity time series for the anthropocene. *Global Ecology and Biogeography*, 27(7):760–786, 2018.

[16] Michael Kalyuzhny and Nadav M Shnerb. Dissimilarity-overlap analysis of community dynamics: Opportunities and pitfalls. *Methods in Ecology and Evolution*, 8(12):1764–1773, 2017.

[17] Matthieu Barbier, Guy Bunin, and Mathew A. Leibold. Getting more by asking for less: Linking species interactions to species co-distributions in metacommunities. *in prep*, 2022.

[18] Yuval R Zelnik, Matthieu Barbier, David W Shanafelt, Michel Loreau, and Rachel M Germain. Linking intrinsic scales of ecological processes to characteristic scales of biodiversity and functioning patterns. *bioRxiv*, 2021.

[19] André M de Roos. Dynamic population stage structure due to juvenile–adult asymmetry stabilizes complex ecological communities. *Proceedings of the National Academy of Sciences*, 118(21):e2023709118, 2021.

[20] H Rieger. Solvable model of a complex ecosystem with randomly interacting species. *Journal of Physics A: Mathematical and General*, 22(17):3447, 1989.

[21] Matthieu Barbier, Jean-François Arnoldi, Guy Bunin, and Michel Loreau. Generic assembly patterns in complex ecological communities. *Proceedings of the National Academy of Sciences*, 115(9):2156–2161, 2018.

[22] Laura Sidhom and Tobias Galla. Ecological communities from random generalized lotka-volterra dynamics with nonlinear feedback. *Physical Review E*, 101(3):032101, 2020.

[23] Yuval R Zelnik, Ehud Meron, and Golan Bel. Gradual regime shifts in fairy circles. *Proceedings of the National Academy of Sciences*, 112(40):12327–12331, 2015.

[24] Christoph Ratzke, Jonas Denk, and Jeff Gore. Ecological suicide in microbes. *Nature ecology & evolution*, 2(5):867–872, 2018.

[25] Matthieu Barbier, Claire De Mazancourt, Michel Loreau, and Guy Bunin. Fingerprints of high-dimensional coexistence in complex ecosystems. *Physical Review X*, 11(1):011009, 2021.

[26] Jurg W Spaak and Frederik De Laender. Intuitive and broadly applicable definitions of niche and fitness differences. *Ecology letters*, 23(7):1117–1128, 2020.

[27] and others. Characteristic sizes of life in the oceans, from bacteria to whales. *Annual review of marine science*, 8:217–241, 2016.

[28] Ian J Wright, Peter B Reich, Mark Westoby, David D Ackerly, Zdravko Baruch, Frans Bongers, Jeannine Cavender-Bares, Terry Chapin, Johannes HC Cornelissen, Matthias Diemer, et al. The worldwide leaf economics spectrum. *Nature*, 428(6985):821–827, 2004.

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

[30] Tal Agranov and Guy Bunin. Extinctions of coupled populations, and rare event dynamics under non-gaussian noise. *Physical Review E*, 104(2):024106, 2021.

[31] Martin Feinberg. Chemical reaction network structure and the stability of complex isothermal reactors—i. the deficiency zero and deficiency one theorems. *Chemical engineering science*, 42(10):2229–2268, 1987.

[32] Matthieu Barbier and DS Lee. Urn model for products' shares in international trade. *Journal of Statistical Mechanics: Theory and Experiment*, 2017(12):123403, 2017.

[33] Tobias Galla. Random replicators with asymmetric couplings. *Journal of Physics A: Mathematical and General*, 39(15):3853, 2006.

[34] Theo L Gibbs, Gabriel Gellner, Simon A Levin, Kevin S McCann, Alan Hastings, and Jonathan M Levine. Can higher-order interactions resolve the species coexistence paradox? *bioRxiv*, pages 2023–06, 2023.

[35] RM Nisbet, EB Muller, K Lika, and SALM Kooijman. From molecules to ecosystems through dynamic energy budget models. *Journal of animal ecology*, pages 913–926, 2000.

[36] NS Jacobsen, C Lindemann, EA Martens, AB Neuheimer, K Olsson, A Palacz, AW Visser, N Wadhwa, and T Kiørboe. Characteristic sizes of life in the oceans, from bacteria to whales. *Annu. Rev. Mar. Sci*, 8:3–1, 2016.

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

[38] Felix Roy, Matthieu Barbier, Giulio Biroli, and Guy Bunin. Complex interactions can create persistent fluctuations in high-diversity ecosystems. *PLoS computational biology*, 16(5):e1007827, 2020.

[39] Guy Bunin. Directionality and community-level selection. *Oikos*, 130(4):489–500, 2021.

[40] Robert M May. Biological populations obeying difference equations: stable points, stable cycles, and chaos. *Journal of Theoretical Biology*, 51(2):511–524, 1975.

[41] Kevin Liautaud, Egbert H van Nes, Matthieu Barbier, Marten Scheffer, and Michel Loreau. Superorganisms or loose collections of species? a unifying theory of community patterns along environmental gradients. *Ecology letters*, 22(8):1243–1252, 2019.

[42] Pierre Bizeul and Jamal Najim. Positive solutions for large random linear systems. *Proceedings of the American Mathematical Society*, 149(6):2333–2348, 2021.

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

[44] Tobias Galla. Dynamically evolved community size and stability of random lotka-volterra ecosystems (a). *EPL (Europhysics Letters)*, 123(4):48004, 2018.

[45] Thibaut Arnoulx de Pirey and Guy Bunin. Aging by near-extinctions in many-variable interacting populations. *Physical Review Letters*, 130(9):098401, 2023.

[46] Thibaut Arnoulx de Pirey and Guy Bunin. Many-species ecological fluctuations as a jump process from the brink of extinction. *arXiv preprint arXiv:2306.13634*, 2023.

[47] Stav Marcus, Ari M Turner, and Guy Bunin. Local and collective transitions in sparsely-interacting ecological communities. *PLoS computational biology*, 18(7):e1010274, 2022.

[48] Emil Mallmin, Arne Traulsen, and Silvia De Monte. Chaotic turnover of rare and abundant species in a strongly interacting model community. *arXiv preprint arXiv:2306.11031*, 2023.

[49] Mathew A Leibold, Marcel Holyoak, Nicolas Mouquet, Priyanga Amarasekare, Jonathan M Chase, Martha F Hoopes, Robert D Holt, Jonathan B Shurin, Richard Law, David Tilman, et al. The metacommunity concept: a framework for multi-scale community ecology. *Ecology letters*, 7(7):601–613, 2004.

[50] Stav Marcus, Ari M Turner, and Guy Bunin. Local and collective transitions in sparsely-interacting ecological communities. *arXiv preprint arXiv:2110.13603*, 2021.

[51] Jiliang Hu, Daniel R Amor, Matthieu Barbier, Guy Bunin, and Jeff Gore. Emergent phases of ecological diversity and dynamics mapped in microcosms. *Science*, 378(6615):85–89, 2022.

[52] Sebastian J Schreiber. Generalist and specialist predators that mediate permanence in ecological communities. *Journal of Mathematical Biology*, 36(2):133–148, 1997.

[53] Robert M May. Will a large complex system be stable? *Nature*, 238(5364):413–414, 1972.

[54] Manfred Opper and Sigurd Diederich. Phase transition and 1/f noise in a game dynamical model. *Physical review letters*, 69(10):1616, 1992.

[55] Yoshimi Yoshino, Tobias Galla, and Kei Tokita. Statistical mechanics and stability of a model eco-system. *Journal of Statistical Mechanics: Theory and Experiment*, 2007(09):P09003, 2007.

[56] Didier Sornette and Agnès Helmstetter. Endogenous versus exogenous shocks in systems with memory. *Physica A: Statistical Mechanics and its Applications*, 318(3-4):577–591, 2003.

[57] Marten Scheffer, S Harry Hosper, Marie Louise Meijer, Brian Moss, and Erik Jeppesen. Alternative equilibria in shallow lakes. *Trends in ecology & evolution*, 8(8):275–279, 1993.

[58] Yael Fried, David A Kessler, and Nadav M Shnerb. Communities as cliques. *Scientific reports*, 6(1):1–8, 2016.

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

[60] Bart Haegeman and Michel Loreau. A mathematical synthesis of niche and neutral theories in community ecology. *Journal of theoretical biology*, 269(1):150–165, 2011.

[61] Michael Kalyuzhny, Ronen Kadmon, and Nadav M Shnerb. A neutral theory with environmental stochasticity explains static and dynamic properties of ecological communities. *Ecology letters*, 18(6):572–580, 2015.

[62] Stephen P Hubbell. The unified neutral theory of biodiversity and biogeography. In *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press, 2011.

[63] Bart Haegeman and Rampal S Etienne. Relaxing the zero-sum assumption in neutral biodiversity theory. *Journal of Theoretical Biology*, 252(2):288–294, 2008.

[64] Salvador Pueyo, Fangliang He, and Tommaso Zillio. The maximum entropy formalism and the idiosyncratic theory of biodiversity. *Ecology letters*, 10(11):1017–1028, 2007.

[65] Yuval R Zelnik, Jean-François Arnoldi, and Michel Loreau. The three regimes of spatial recovery. *Ecology*, 100(2):e02586, 2019.

[66] Peter Chesson. Mechanisms of maintenance of species diversity. *Annual review of Ecology and Systematics*, 31(1):343–366, 2000.

[67] György Barabás, Rafael D'Andrea, and Simon Maccracken Stump. Chesson's coexistence theory. *Ecological monographs*, 88(3):277–303, 2018.

[68] Thorsten Wiegand, Andreas Huth, Stephan Getzin, Xugao Wang, Zhanqing Hao, CV Savitri Gunatilleke, and IAU Nimal Gunatilleke. Testing the independent species' arrangement assertion made by theories of stochastic geometry of biodiversity. *Proceedings of the Royal Society B: Biological Sciences*, 279(1741):3312–3320, 2012.

[69] Jacob D O'Sullivan, Robert J Knell, and Axel G Rossberg. Metacommunity-scale biodiversity regulation and the self-organised emergence of macroecological patterns. *Ecology letters*, 22(9):1428–1438, 2019.

[70] Lana Descheemaeker, Jacopo Grilli, and Sophie de Buyl. Heavy-tailed abundance distributions from stochastic lotka-volterra models. *Physical Review E*, 104(3):034404, 2021.

[71] Julia Chacón-Labella, Marcelino de la Cruz, and Adrián Escudero. Evidence for a stochastic geometry of biodiversity: the effects of species abundance, richness and intraspecific clustering. *Journal of Ecology*, 105(2):382–390, 2017.

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

[73] Sonia Kéfi, Virginia Domínguez-García, Ian Donohue, Colin Fontaine, Elisa Thébault, and Vasilis Dakos. Advancing our understanding of ecological stability. *Ecology Letters*, 22(9):1349–1356, 2019.

[74] J-F Arnoldi, Azenor Bideault, Michel Loreau, and Bart Haegeman. How ecosystems recover from pulse perturbations: A theory of short-to long-term responses. *Journal of theoretical biology*, 436:79–92, 2018.

[75] Jean-François Arnoldi, Michel Loreau, and Bart Haegeman. The inherent multidimensionality of temporal variability: how common and rare species shape stability patterns. *Ecology Letters*, 22(10):1557–1567, 2019.

[76] Núria Galiana, Jean-François Arnoldi, Matthieu Barbier, Amandine Acloque, Claire de Mazancourt, and Michel Loreau. Can biomass distribution across trophic levels predict trophic cascades? *Ecology Letters*, 24(3):464–476, 2021.

[77] Yuval Zelnik, Nuria Galiana, Matthieu Barbier, Michel Loreau, Eric Galbraith, and Jean-Francois Arnoldi. How collectively integrated are ecological communities? *bioRxiv*, pages 2022–12, 2022.

[78] Jean-François Arnoldi, Matthieu Barbier, Ruth Kelly, György Barabás, and Andrew L Jackson. Invasions of ecological communities: Hints of impacts in the invader's growth rate. *Methods in Ecology and Evolution*, 13(1):167–182, 2022.

[79] Jacob D O'Sullivan, J Christopher D Terry, and Axel G Rossberg. Intrinsic ecological dynamics drive biodiversity turnover in model metacommunities. *Nature Communications*, 12(1):3627, 2021.

[80] Bart Haegeman, Jean-François Arnoldi, Shaopeng Wang, Claire de Mazancourt, José M Montoya, and Michel Loreau. Resilience, invariability, and ecological stability across levels of organization. *bioRxiv*, page 085852, 2016.

[81] Shaopeng Wang, Michel Loreau, Jean-Francois Arnoldi, Jingyun Fang, K Abd Rahman, Shengli Tao, and Claire de Mazancourt. An invariability-area relationship sheds new light on the spatial scaling of ecological stability. *Nature Communications*, 8(1):1–8, 2017.

[82] Jonathan M Levine, Peter B Adler, and Stephanie G Yelenik. A meta-analysis of biotic resistance to exotic plant invasions. *Ecology letters*, 7(10):975–989, 2004.

[83] Hugo Fort. Quantitative predictions from competition theory with an incomplete knowledge of model parameters tested against experiments across diverse taxa. *Ecological Modelling*, 368:104–110, 2018.

[84] Lorenzo Fant, Onofrio Mazzarisi, Emanuele Panizon, and Jacopo Grilli. Stable cooperation emerges in stochastic multiplicative growth. *arXiv preprint arXiv:2202.02787*, 2022.

[85] Samuel R Bray and Bo Wang. Forecasting unprecedented ecological fluctuations. *PLoS computational biology*, 16(6):e1008021, 2020.

[86] Felix May, Thorsten Wiegand, Sebastian Lehmann, and Andreas Huth. Do abundance distributions and species aggregation correctly predict macroecological biodiversity patterns in tropical forests? *Global Ecology and Biogeography*, 25(5):575–585, 2016.