UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Dipartimento
di Fisica
e Astronomia
Galileo Galilei

MASTER DEGREE IN PHYSICS OF DATA

SAMUELE PIO LIPANI

COEXISTENCE AND DIVERSITY IN A CONSUMER RESOURCE MODEL WITH A SERIAL DILUTION SETTING

FINAL DISSERTATION

Supervisors:
Prof. J.Grilli
Prof. S.Suweis

ACADEMIC YEAR 2022/2023

To my dad
I kept the promise and I started to work
in your field of research,
Watch over me

Abstract

The coexistence of species and maintenance of diversity is the most important question in Ecology. Consumer resources models are one of the most interesting settings to model microbial ecosystems. The great majority of theoretical studies have been framed in the chemostat setting, where resources are constantly flowing through the system. On the other hand, both experiments and natural communities, like the gut microbiome, are better described as serial dilution processes, where resources are periodically replenished and consumed. Surprisingly, in this case theoretical results are scarce. The main goal of this thesis is to fill this gap. We study a general multispecies consumer resource model in serial dilution using both numerical tools and analytical techniques borrowed from disordered statistical physics. We find that under general conditions the number of coexisting species is lower than the chemostat setting. Furthermore, both the timescales to reach a stationary state and the shape of the species abundances distribution appear to depend dramatically on the connectivity properties of the consumption matrix: while for a fully connected model the system shows enormously long transients, by introducing a small sparsity such timescales reduce abruptly. We developed a novel algorithm to face this problem of long time scales for the convergence of the process.

Contents

1 Summary

Many experiments have been carried out in order to mimic natural environments like the gut microbiota and how it works. For the purpose of describing these natural communities one can use consumer-resource models. In this thesis we are going to analyze how we can describe these processes with two theoretical frameworks: the chemostat and the serial dilution. In section 2 we introduce the meaning of community Ecology and how mathematical models behave in portraying natural communities. In section 2.3 we show first examples of simple models of species interaction. Then, in section 3, we move to more complex frameworks like the Lotka-Volterra model and its general form. We introduce therefore disordered systems and how we can set a generalized consumer resource model by starting from a Lotka-Volterra. After a brief introduction to these models we present the main object of analysis of this work; the chemostat model is taken into account in section 4 where resources are constantly flowing through the system. While most of the studies have been framed in the chemostat setting, in section 5, we realize that a more realistic situation is represented by a serial dilution model. The latter, we will see, presents differences and analogies with the chemostat. In order to further examine the divergences of the two models, in section 6 we focus on a stochastic serial dilution process as pictured in B.Good 2022. To make it simpler, in section 6 and 7, we study and simulate it for one species and one resource and retrieve some considerations about the shape of the time series of species abundance in the end of each dilution step and its distribution. In section 8 we explore the model in a multispecies setting. On one hand we are adding complexity by adding species but we are also simplifying by hand the way resource are replenished. Instead of adding new resources randomly in each dilution step, as it was in section 7 with one species and one resource, we will use the same value $\overline{Y}$ to restore initial resources to their starting conditions. In section 8.2 we will describe some problems arising during the implementation of the algorithm. They refers to the ambiguous definition of an approximate steady state that allows us to stop the run within a feasible number of dilution steps. Finally in section 9 we will derive analytical results useful for diversity and coexistence predictions. We will be able to describe the evolution dynamics of the system in terms of the mean species abundance. At some point, in section 10 we point out the differences in convergence time scales of the algorithm when changing the sparsity of the consumption matrix; furthermore we develop a novel algorithm called "invasion analysis" that helps in detecting how many species are surviving in the environment even if it doesn't reach stationarity. Eventually in section 11 we carry out an analysis on how parameters of the model influence the number of surviving species with respect to the chemostat.

2 Introduction to community ecology

The basic definition of Ecology, the scientific study of the relationships between organisms and their environment, is rather vague and the word environment requires an explicit definition. An alternative definition of ecology that is more relevant to community ecology is the scientific study of the distribution and abundance of organisms (C.J.Krebs 1994; H.G.Andrewartha 1961). In population ecology, we are interested in understanding the variables that most likely govern the rates of growth, abundances, and distributions of biological populations. In this context, a population (also known as a biological population) refers to a collection of interbreeding organisms that are present in the same region or space at the same time (i.e., they are sympatric). These individuals are assumed to function as a functional unit because they interact with one another and interbreed with other members of the population. When a population is closed, neither immigration nor emigration of people from outside the population is anticipated. Populations aren't actually closed to immigration or emigration unless we're talking about a population on a far-off island, a mountaintop, or a secluded cave. We typically do not know which individuals in a population might be recent immigrants unless we have successfully marked every member of that population. These concepts are incorporated by P.Turchin 2003 in his definition of a population, which he defines as "a group of individuals of the same species that live together in an area large enough to permit normal dispersal and migration behavior, and in which population changes are primarily governed by birth and death processes." Since we are dealing with a group of people interacting in a specific place and time, a local population is different from a species or species population.

2.1 Fundamental principles and the use of mathematical models

What are the underlying principles that govern population growth? Population ecology is a quantitative field by necessity, and mathematically inclined ecologists have developed a range of predictive models to address questions about populations. The generation time of a population differs by several orders of magnitude, just as the mass of a single bacterium is many orders of magnitude smaller than the mass of an elephant. Therefore, no single population growth model is appropriate for all organisms or environments. As one moves from lower to higher levels of organization, it is a general rule of systems that properties that were absent at the lower levels are added. Consequently, a single organism is more than just a group of physiological systems. Similarly, a population has properties not evident from the study of individuals. Populations have growth rates distributions, age distributions, and spatial patterns.

2.2 The general laws of population ecology

According to P.Turchin 2003, population ecology is a robust, predicative science with a set of underlying principles that are very similar to Newton's laws. He has outlined these three essential ideas.:

- populations tend to grow exponentially
- populations show self-limitation (or bounded fluctuations)
- consumer–resource interaction tend to be oscillatory

In the first scenario, all populations exhibit a nonlinear, exponential growth pattern in the absence of density-dependent feedback from the environment. This is referred to as "the exponential law" by P.Turchin 2001, who sees a direct analogy to Newton's law of inertia. For more intricate mathematical explanations of population dynamics, the exponential law serves as a good starting point. The idea that per capita population growth declines with resource depletion forms the foundation of the second theorem or principle, self-limitation.

2.3 Simple models of species interaction

Life is fundamentally about consumption. In order to maintain themselves, grow, and reproduce, organisms must take resources from their environment. As the fundamental components of cells, all organisms require carbon and inorganic substances like nitrogen and phosphorus. We refer to these as limiting resources when one or more of them restrict growth, and we talk about resource limitation and resource competition when different species' growth is restricted as a result of the lack of available resources. When these resources are living things, we frequently refer to them as hosts, prey, or just victims. Predators, pathogens, parasites, herbivores, or even adversaries could be the consumers.

2.3.1 Predator-prey models

In general, we can conceptualize consumer-resource interactions as coupled equations:

$$\dot{X} = X f_X \left(X, Y\right) \tag{2.1}$$

$$\dot{Y} = Y f_Y \left(X, Y\right) \tag{2.2}$$

in which per-unit change in the resource is a function, f_Y, of the resource itself and the consumer. Likewise, per-unit change in the consumer is a function,f_X, of the consumer itself, and the resource. Let us start with a famous consumer-resource pairing, that of predator and prey, also known as enemy-victim interactions. Imagine that a population grows at rate

$$\dot{X} = bX - dX \tag{2.3}$$

This model describes how the growth of a population depends on a birth rate b and on a death rate d. Let us stick with this simplicity, and link a predator, Y, to the prey. A very simple way to do this is to include predators as another source of death of prey (aX), and prey as the source of life for predators (eaX),

$$\dot{X} = bX - dX - aX \tag{2.4}$$

$$\dot{Y} = eaX - mY \tag{2.5}$$

where $e < 1$ and is the relative efficiency with which predators turn captured prey into new predators. The units of these differential equations are individuals per unit time. Therefore, the units of each term must also be individuals per unit time. The units of b and d are units of prey per unit of prey per unit time.

$$\text{units}(b) = \frac{\text{prey}}{\text{prey}}\frac{[1]}{[t]} = [t^{-1}]$$

or simply per unit time, where a unit could be an individual or a gram. Units of attack rate or consumption rate, a, are also t^{-1}. Units of efficiency are simply new predators per unit prey,

$$\text{units}(e) = \frac{\text{predators}}{\text{prey}}$$

and units of m are predators per predator per unit time, or t^{-1}. When we combine all the units for all the terms for $\dot{Y}$ above, we get

$$\frac{\text{predators}}{[t]} = \left(\frac{\text{predators}}{\text{prey}} \cdot \frac{1}{[t]} \cdot \text{prey}\right) - \left(\frac{1}{[t]} \cdot \text{predators}\right)$$

The mere fact of efficiency, $e < 1$, shows us another general proposition about enemy-victim interactions; if we consider enemy and victim in the same units (e.g., grams), then the cost to the victim is always greater than the benefit to the enemy. This is simply the result of the second law of thermodynamics, and is a fundamental part of all predator-prey interactions (Holt 2011). By setting $\dot{Y} = 0$ we can make a statement that relates predator and prey abundances at steady state.

$$Y = \frac{ea}{m}X \quad ; \quad \frac{Y}{X} = \frac{ea}{m} \tag{2.6}$$

Thus, this simple theory predicts that at steady state, the relative abundances of predator and prey will be constant. Because the prey only grow (or die) exponentially, this theory can make no prediction about Y^*, save when $b = d$. However, it does predict that predator and prey will persist at a constant ratio, assuming the parameters are fixed.

3 Lotka-Volterra and consumer resource models

A more widely used model of predator-prey dynamics is that championed independently by Alfred Lotka and Vito Volterra. In this formulation, the per predator capture rate depends only on the prey abundance, not on the ratio of prey to predators. We refer to this as “mass effects”, where, as is often the case in chemical reactions, the rate of reaction depends merely on the concentration of the reactants. Molecules of each reactant do not interfere with each other. Organisms are not always so "well behaved". Nonetheless, it is a widely used, and often reasonable, approximation of enemy–victim interactions, especially at low densities of enemies.

When capture rate is independent of the number of other predators in the population, we describe it thus.

$$\dot{X} = bX - dX - aXY = rX - aXY \tag{3.1}$$

$$\dot{Y} = eaXY - mY \tag{3.2}$$

We find equilibria, or more correctly, zero net growth isoclines, by setting $\dot{X} = \dot{Y} = 0$. When we do this we get curious results.

$$\dot{Y} = 0 \longrightarrow X = \frac{m}{ea} \tag{3.3}$$

$$\dot{X} = 0 \longrightarrow Y = \frac{r}{a} \tag{3.4}$$

Each population’s growth rate is zero when the other population is at a fixed value. Looking at the phase space of the prey predator dynamics we see that the prey increase whenever the predator abundance is below a certain value (r/a). On the other hand the predator increase or decrease when the prey is above or below a particular value (m/ea). The only time these populations are at rest is when both of these are true.

3.1 From Lotka-Volterra to consumer-resource models

The equations of Lotka-Volterra can be naturally generalized to a system with S species:

$$\frac{\mathrm{d}X_i}{\mathrm{d}t} = \Big[a_i - \sum_{j=1}^{S} \alpha_{ij} X_j(t)\Big] X_i(t) \qquad i = 1, \ldots, S \tag{3.5}$$

where the coefficients α_{ij} encode the species interaction network. We can move to a disordered system then just by drawing the consumption matrix entries from probability distributions as prescribed in W.Cui 2021.

We can go a step further in generalization, introducing a non-linear species dependence:

$$\frac{\mathrm{d}X_i}{\mathrm{d}t} = \underbrace{G(X_1, X_2, \ldots, X_{i-1}, X_{i+1}, \ldots, X_S)}_{\text{interaction among species}} \underbrace{F(N_i)}_{\substack{\text{species}\\\text{growth}\\\text{function}}}$$

where it is important that G *does not* depend on N_i. If taking into account as an example the linear case, at stationarity, we have $\mathrm{d}X_i \,/\, \mathrm{d}t = 0 \ \forall i$. This can happen trivially if $X = 0$ (i.e. if the system is empty).

A more interesting equilibrium is given by the X^* that satisfies the solution of eq.(3.5):

$$a_i = \sum_{j=1}^{S} \alpha_{ij} X_j^*$$

In order to study the stability of the equilibrium we can compute the jacobian matrix:

$$J_{ij} = \left.\frac{\partial \dot{X}_i}{\partial X_j}\right|_{X^*} = X_i^* \alpha_{ij}$$

Since we are dealing with populations, the entries of X^* cannot be negative (feasibility condition):

$$X_i^* \geq 0 \qquad \forall i$$

If the adjacency matrix is antisymmetric, i.e. $\alpha_{ij} = -\alpha_{ji}$, then its eigenvalues will be purely imaginary. This holds true even if each row is multiplied by a positive number (e.g. X_i^* in our case), meaning that also the eigenvalues of J_{ij} are purely imaginary.

In this case, the equilibrium is stable, with the system oscillating wildly (neutrally stable). If, instead, the adjacency matrix is symmetric, i.e. $\alpha_{ij} = \alpha_{ji}$, with $\alpha_{ii} = 0$, we can study the global stability of the system. This is done by building a Lyapunov function $Q(N(t)) \equiv Q(t)$ that satisfies the following properties:

- $Q(t) > 0$ if $X(t) \neq X^*$
- $Q(t) = 0$ if and only if $X(t) = X^*$

If its derivative is non-positive, then X^* is a global equilibrium:

$$\frac{\mathrm{d}Q}{\mathrm{d}t} \leq 0 \Rightarrow X^* \text{ is a global equilibrium}$$

There is no general rule to build Lyapunov functions, just heuristics. Usually, good candidates are quadratic functions, such as the relative fluctuation among connected species:

$$Q(t) = \sum_{i,j=1}^{S} (X_i(t) - X_i^*)\alpha_{ij}(X_j(t) - X_j^*)$$

Differentiating:

$$\dot{X}_i = X_i\Big[a_i - \sum_j \alpha_{ij} X_j\Big]$$
$$a_i = \sum_j \alpha_{ij} X_j^*$$

And so:

$$\frac{\mathrm{d}Q}{\mathrm{d}t} = -2\sum_{i=1}^{S} X_i(t) J(t)^2 \qquad J(t) = \sum_{k=1}^{S} \alpha_{ik}(X_k(t) - X_k^*)$$

Since $X_i(t) \geq 0$ and $J^2 > 0$, we get:

$$\frac{\mathrm{d}Q}{\mathrm{d}t} \leq 0$$

meaning that X^* is a global equilibrium. This means that, if a system starts at $t = 0$ with S species (i.e. S non-zero components in $X(0)$), then the equilibrium will be such that:

$$X^* = (X_1^*, X_2^*, \dots, X_k^*, 0, \dots, 0)$$

with $S - k$ species going extinct. If we perturb this equilibrium reintroducing one species, it will go extinct again, since the equilibrium is global. Problems arise when we are asked to predict how many species are going to extinction in a general model like this one.

3.2 Consumer resource models

Before moving to pure consumer resource models it is useful to understand what competition really means between species that have overlaps in their predation niches. Which factors influence competition and how we can think of a simple model that takes into account competition?

Competition is most typically considered the interaction of individuals that die for a common resource that is in limited supply, but more generally can be defined as the direct or indirect interaction of organisms that leads to a change in fitness when the organisms share the same resource. The outcome usually has negative effects on the weaker competitors.

Consider a system of two competing species A and C, e.g. two different bacteria in a flask. Periodically, resources are restored, and the system is diluted, i.e. N individuals are picked at random and the others are removed. In general, if one of the two is better at consuming the resources than the other, it will thrive and grow faster, meaning that it will have more individuals. Thus, after each cycle of dilution, the ratio between the two populations will favor the fittest one, until the other disappears. However, if there are two different resources, and each species consumes only one of them, the two populations may coexist without issue. Thus, we would expect that in an environment with M resources, only $m \leq M$ species can coexist. This result goes under the name of competition exclusion principle. We can write now a model that exploits the definition of competition between species of an environment as follows. We consider S species and M resources:

$$\begin{cases} \dot{X}_i = X_i \left(\sum_{\alpha=1}^{M} v_\alpha a_{i\alpha} Y_\alpha - \delta_i \right) \\ \dot{Y}_\alpha = w_\alpha - \sum_{i=1}^{S} X_i a_{i\alpha} r_\alpha(Y_\alpha) \end{cases} \tag{3.6}$$

Where w_α is the supply rate, i.e the rate of growth of the α-th resource. $a_{i\alpha}$ is the metabolic energy of species i devoted to eat resource α, also called the metabolic strategy of i. In particular only if $a_{i\alpha} > 0$ species i can consume resource α. Then $r_\alpha(Y_\alpha)$ could be either the Monod function (for abiotic resources) or the logistic functional response (for biotic resources) or some simpler functional response, like a linear one in resources for their depletion because of the interaction with species. Eventually v_α is called efficiency of resource conversion into biomass of species i and δ_i the death rate of species i. In order to see competition at work we can simulate a simple system with ten species and five resources and figure out how the number of resources influences the

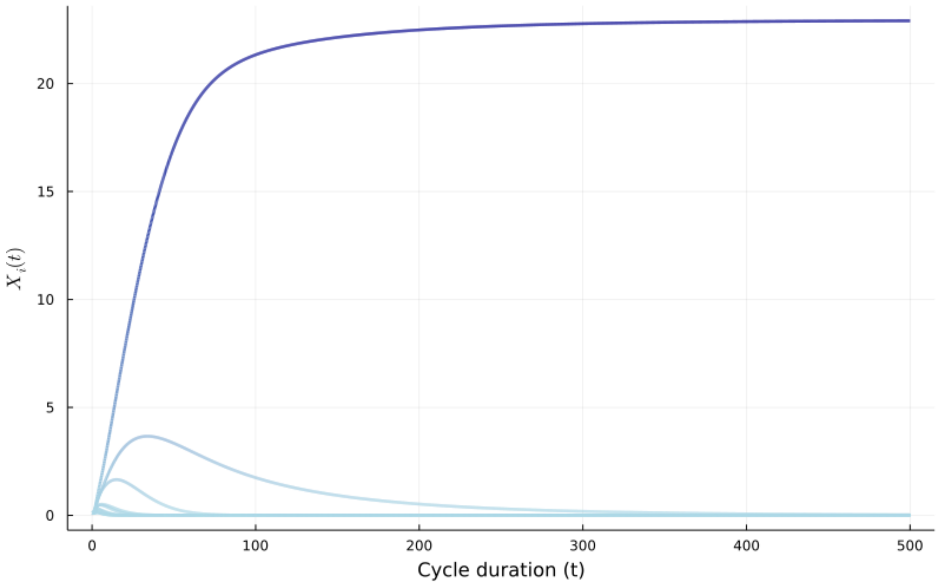


Figure 1: Simulations have been computed with all parameters drawn from a uniform distribution in $[0, 1]$.

final abundance of species. In Figure 1 one species out of ten survived at the end as prescribed by the competition exclusion principle.

3.2.1 Generalized consumer resource models

We are going to prove now how a general setting of a consumer resource model can be reduced into generalized Lotka-Volterra equations and viceversa. Indeed Generalized Lotka-Volterra equations can be easily separated into coupled sets of differential equations such that they resemble a consumer resource setting, that is simply having species and resources with separated dynamics into the model. It means in practice to distinguish two time scales for species and resources such that the latter are extinguished much faster than species reach saturation. We start from the consumer resource model for $i = 1, \ldots, S$ species of predators and $\alpha = 1, \ldots, M$ species of resources with resources growing logistically.

$$\begin{cases} \dot{X}_i = X_i \left(\sum_{\alpha=1}^{M} v_\alpha a_{i\alpha} Y_\alpha - \delta_i \right) \\ \dot{Y_\alpha} = Y_\alpha \left(w - \frac{w}{k} Y_\alpha - \sum_{i=1}^{S} X_i a_{i\alpha} \right) \end{cases}$$

By setting the resources variation to zero we are saying that resources are quenching faster than the extinction of the species of predators. For instance the characteristic time of dynamical evolution of the consumer species is much greater of the resources':

$$Y_\alpha = k - \frac{k}{w} \sum_{i=1}^{m} X_i a_{i\alpha} \tag{3.7}$$

If we substitute this into the first equation of the system we end up with m decoupled differential equations leading to the quasi stationary approximation of the consumer resource model:

$$\dot{X}_i = X_i \left(k \sum_{\alpha=1}^{M} v_\alpha a_{i\alpha} - \frac{k}{w} \sum_{i=1}^{S} \sum_{\alpha=1}^{M} X_i a_{i\alpha} v_\alpha a_{i\alpha} - \delta_i \right) \tag{3.8}$$

For simplicity we can set the product of the metabolic strategies equal to a kind of interaction matrix such that

$$\sum_{\alpha=1}^{M} a_{i\alpha} v_\alpha a^T_{\alpha j} = A_{ij} \tag{3.9}$$

We would have a simpler expression for the quasi stationary differential equations

$$\dot{X}_j = X_j \left(\sum_{\alpha=1}^{M} k v_\alpha a_{j\alpha} - \frac{k}{w} \sum_{j=1}^{S} A_{ij} n_j - \delta_j \right) \tag{3.10}$$

If we would find also the stationary steady states we can set the latter equations to zero and retrieve $\bar{n}_j$ from them:

$$\overline{X}_j = \sum_{j=1}^{S} A^{-1}_{ij} \left(\sum_{\alpha=1}^{M} w v_\alpha a_{j\alpha} - \frac{w}{k} \delta_j \right) \tag{3.11}$$

From this point on we will try to exploit the concepts regarding competition inside the consumer resource models to try to predict certain behaviors of experimental data taken from "serial dilution processes".

4 Chemostat prescription

A chemostat is defined as a steady-state bioprocess, where a microbial culture is continuously supplied with nutrients at a fixed rate and concomitantly harvested to keep the culture volume constant. We can easily write a model trying to mimic the chemostat behavior such that resources are externally supplied and species and resources are being diluted thanks to an intrinsic death rate.

$$\begin{aligned} \dot{X}_i &= X_i \left(\sum_\alpha a_{i\alpha} Y_\alpha - m_i \right) \\ \dot{Y}_\alpha &= K_\alpha - \omega_\alpha Y_\alpha - \sum_i X_i a_{i\alpha} Y_\alpha \end{aligned} \tag{4.1}$$

Species consumptions $a_{i\alpha}$ are random variables drawn from a Gaussian distribution with mean μ/M and variance σ_c^2/M. They can be deposed into $a_{i\alpha} = \mu/M + \sigma_c d_{i\alpha}$ where the fluctuation part obeys

$$\langle d_{i\alpha} \rangle = 0 \tag{4.2}$$

$$\langle d_{i\alpha} d_{j\beta} \rangle = \frac{\delta_{ij}\delta_{\alpha\beta}}{M} \tag{4.3}$$

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

$$\langle K_\alpha \rangle = K \tag{4.4}$$

$$\mathrm{Cov}(K_\alpha, K_\beta) = \delta_{\alpha\beta}\sigma_K^2 \tag{4.5}$$

$$\langle m_i \rangle = m \tag{4.6}$$

$$\mathrm{Cov}(m_i, m_j) = \delta_{ij}\sigma_m^2 \tag{4.7}$$

Let now $\langle Y \rangle = 1/M \sum_\alpha Y_\alpha$ and $\langle X \rangle = 1/S \sum_i X_i$ be the average resource and average species abundance. We can then rewrite eqs. (4.1) as

$$\dot{X}_i = X_i \left[\mu \langle Y \rangle - m + \sum_\alpha \sigma_c d_{i\alpha} Y_\alpha - \delta m_i \right] \tag{4.8}$$

$$\dot{Y}_\alpha = K + \delta K_\alpha - \left[\omega_\alpha + \gamma^{-1} \mu \langle X \rangle + \sum_j \sigma_c d_{j\alpha} X_j \right] Y_\alpha \tag{4.9}$$

Where $\delta K_\alpha = K_\alpha - K$, $\delta m_i = m_i - m$ and $\gamma = M/S$. In order to find the probability distribution, at stationarity, of final species and resources abundances we can use the cavity method.

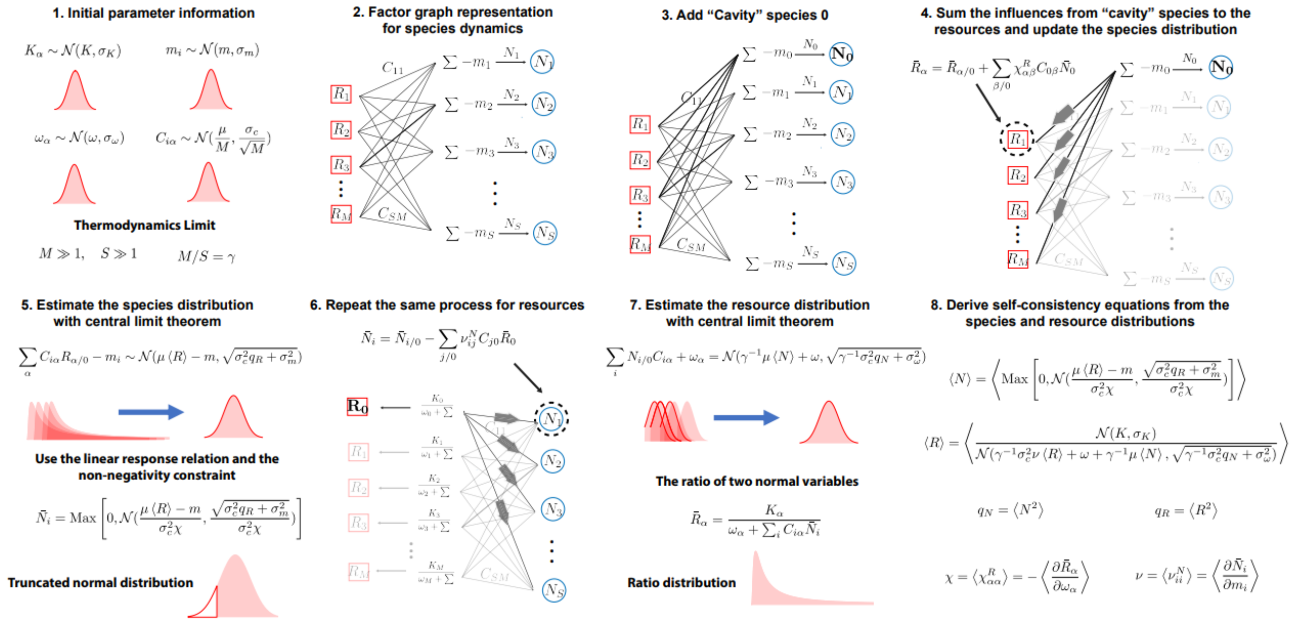


Figure 2: This figure has been taken from G. P.Mehta M. 2018.

4.1 Cavity method for SAD prediction

As we can see from Figure 2 and in R. P.Mehta W. 2020 the basic idea behind the cavity method is to derive self-consistency equations relating an ecosystem with M resources and S species to an ecosystem with $M + 1$ resources and $S + 1$ resources. This is done by adding a new "cavity" species 0 and a new "cavity" resource 0 to the original ecosystem. When $S, M \gg 1$, the effect of the new cavity species or resource is small and can be treated using perturbation theory. From eqs. (4.8) and (4.9) we can write the ecological model for the $(M+1,S+1)$ system where resource Y_0 and species X_0 have been introduced to the (S, M) system and solving everything for the new species and resource taking into account that according to neutral theory all species are equivalent in the thermodynamic limit, that is $S, M \gg 1$ and $M/S = \gamma$ fixed.

$$\dot{X}_0 = X_0 \left[\mu\langle Y \rangle - m + \sum_{\alpha} \sigma_c d_{0\alpha} Y_\alpha - \delta m_0 \right] \tag{4.10}$$

$$\dot{Y}_0 = K + \delta K_0 - \left[\omega_0 + \gamma^{-1}\mu\langle X \rangle + \sum_{i} \sigma_c d_{i0} X_i \right] Y_0 \tag{4.11}$$

4.1.1 Perturbations in cavity approach

Let us now define the following susceptibilities:

$$\begin{aligned}
\chi^Y_{\alpha\beta} &= -\frac{\partial Y^*_\alpha}{\partial \omega_\beta} && (4.12)\\
\chi^X_{i\beta} &= -\frac{\partial X^*_i}{\partial \omega_\beta} && (4.13)\\
\nu^Y_{\alpha j} &= \frac{\partial Y^*_\alpha}{\partial m_j} && (4.14)\\
\nu^X_{ij} &= \frac{\partial X^*_i}{\partial m_j} && (4.15)
\end{aligned}$$

Where we denote X^* and Y^* the steady-state value of species and resources. The main goal of this calculation is to derive a set of self-consistency equations that relates the ecological system characterized by $M+1$ resources (variables) and $S+1$ species (constraints) to that with the new species and the new resources removed: $(S+1, M+1) \longrightarrow (S, M)$. Since the introduction of a new species and resource represents only a small (order $1/M$) perturbation to the original ecological system, we can express the steady-state species and resource abundances in the $(S+1, M+1)$ system with a first order Taylor expansion around the (S, M) values.

$$\begin{aligned}
X^*_i &= X^*_{i/0} - \sigma_c \sum_{\beta/0} \chi^X_{i\beta} d_{0\beta} X^*_0 - \sigma_c \sum_{j/0} \nu^X_{ij} d_{j0} Y^*_0\\
Y^*_\alpha &= Y^*_{\alpha/0} - \sigma_c \sum_{\beta/0} \chi^Y_{\alpha\beta} d_{0\beta} X^*_0 - \sigma_c \sum_{j/0} \nu^Y_{\alpha j} d_{j0} Y^*_0
\end{aligned} \qquad (4.16)$$

Where $\sum_{\beta/0}$ and $\sum_{j/0}$ represent summations over all the species except the 0-th. The next step is to introduce these equations into the original ones and solve for the steady state of X_0 and Y_0. Employing the central limit theorem for the susceptibility quantities, R. P.Mehta W. 2020 found that

$$0 = X^*_0 \left[\mu\langle Y\rangle - m - \sigma_c^2 \chi X^*_0 + \sum_{\beta/0} \sigma_c d_{0\beta} Y^*_{\beta/0} + \sigma_c d_{00} Y^*_0 - \delta m_0 \right] + \mathcal{O}(M^{-1/2}) \qquad (4.17)$$

Where χ is the average susceptibility. After introducing an auxiliary Gaussian variable,z_X with zero mean and unit variance then

$$\sum_{\beta/0} \sigma_c d_{0\beta} Y^*_{\beta/0} + \sigma_c d_{0\beta} Y^*_0 - \delta m_0 = z_X \sqrt{\sigma_c^2 q_Y + \sigma_m^2} \qquad (4.18)$$

where $q_Y = \frac{1}{M}\sum_\beta R_\beta^2$ is the second moment of the resource distribution. It is possible now to solve eq.(4.17) in terms of quantities just defined

$$\mu\langle Y\rangle - m - \sigma_c^2\chi X_0^* + z_X\sqrt{\sigma_c^2 q_Y + \sigma_m^2} \leq 0 \tag{4.19}$$

That reduces to

$$X_0^* = \max\left[0, \frac{\mu\langle Y\rangle - m + z_X\sqrt{\sigma_c^2 q_Y + \sigma_m^2}}{\sigma_c^2\chi}\right] \tag{4.20}$$

Which is a truncated Gaussian. It means that those species that fall into the negative region of the distribution are predicted to extinct during the process. We have therefore a direct way to compute the diversity of the chemostat model.

5 Serial dilution prescription

To interrogate the dynamics of in vivo microbiotas, a common, top-down strategy is longitudinal sampling followed by 16S rRNA amplicon or metagenomic sequencing, as explained in (S.Gupta 2019), thereby generating a relative abundance time series. Longitudinal data analysis has revealed that in healthy humans, species abundances fluctuate around constant, host-specific values (J.G.Caporaso and R.Knight 2011; A.David 2013; J.Faith 2013). These time series have recently been found to have distinctive statistical signatures, sometimes referred to as macroecological dynamics, that can reflect the characteristics of the community and its environment (L.Descheemaeker 2020; J.Grilli 2020a; W.Shoemaker 2017). Time series modeling can shed light on the ecological mechanisms that underlie the observable patterns.

5.1 First approach to serial dilution

Directly measuring all parameters is inherently difficult because the network of resource consumption in a community will typically depend on thousands of underlying parameters. By using an indirect, coarse-grained approach, in which resources describe useful clusters of metabolites or niches and model parameters are randomly selected from a common statistical ensemble, we attempted to overcome this combinatorial complexity.

To determine the nature of time series statistics generated by resource competition, we considered a minimal CR model in which consumers compete for M resources via growth dynamics described by:

$$\begin{aligned} \dot{X}_i &= X_i \sum_{\alpha=1}^{M} a_{i\alpha} Y_j \\ \dot{Y}_\alpha &= -Y_\alpha \sum_{i=1}^{X} a_{i\alpha} X_i \end{aligned} \tag{5.1}$$

Here, X_i denotes the abundance of species i, Y_α the amount of resource α, and $a_{i\alpha}$ the consumption rate of resource α by consumer i. Individual resources in this model represent functional groups of metabolites or niches because the resources are defined at a coarse level.

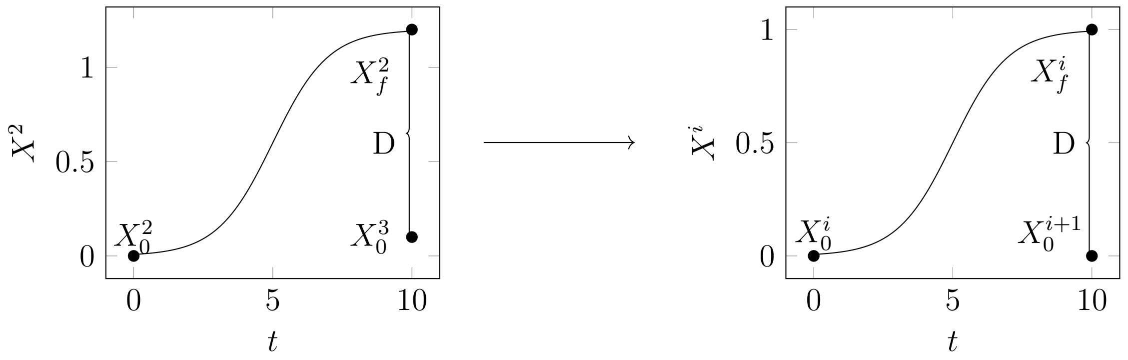
We made the assumption that the rates of resource consumption, $a_{i\alpha}$, were independent of the external environment and constant over time, allowing us to specify the community's inherent ecological characteristics with a set of $S \cdot M$ microscopic parameters. We hypothesized that typical profiles of resource consumption taken from a statistical ensemble might be able to capture the macroecological features of our experimental time series in order to reduce the complexity of this vast parameter space. This is a crucial simplification because, despite the fact that the values chosen at random will never exactly correspond to the rates at which a given microbiota consumes resources, prior research indicates that they can frequently recapitulate the overall behavior of sufficiently diverse communities (W.Cui 2021; question about the different models that can reproduce the consumption rates different from a simple random network matrix where each entry is randomly drawn from a uniform distribution with a certain probability p).

5.2 Illustrating serial dilution

Initial conditions in this process should not influence its stability and convergence. After we set initial conditions for both species and resources we let them evolve in one cycle such that all resources are deplenished $\mathrm{d}Y_\alpha/\mathrm{d}t = 0$ and $Y_\alpha(T) \approx 0$. After one evolution cycle we set as initial conditions of a further one:

$$\begin{aligned} X_i^{k+1}(0) &= \frac{X_i^k(T)}{D} \\ Y_\alpha^{k+1}(0) &= \frac{Y_\alpha^k(T)}{D} + \overline{Y}_\alpha^k \end{aligned} \tag{5.2}$$

Where D is the dilution factor and $\overline{Y}_\alpha^k$ are the type of resources that can be put into the environment to restore the consumed ones. We can visualize the simplest case with one species:



5.3 Biological meaning of the serial dilution setting

In order to simulate the punctuated turnover of gut microbiotas caused by repeated feedings and digestion of food between sampling times, we used a serial dilution scheme (A.Erez 2020). We know that experiments are designed in such a way that both bacteria and the remaining resources are taken through a small pipette and are diluted for the next step because we are considering serial dilution processes. Even if experiments have been carried out in a serial dilution setting many theoretical studies still use a chemostat model to describe what is happening in this type of natural communities. Let us go deeper in differences between the serial dilution setting and the chemostat

5.4 Chemostat and serial dilution

As we see in A.Erez 2020, when the dilution factor and the integration time in serial dilution are small we reduce to the chemostat limit; that is having the same small dilution that a chemostat provides during each time step and rescaling the dilution steps time scale such that dilution and evolution of the system are on the same time. In Figure 3 we notice that for small evolution times and small dilution factor D the species and resources abundance distributions are reduced to Gaussians. Furthermore species are not able to grow as much as their carrying capacity and resources are not totally deplenished. Now that we know what a chemostat is and that can be obtained as a limit of the serial dilution model we can dig deeper into more complicated prescriptions of the consumer-resource for serial dilution and explore its properties with a critical eye regarding what instead the chemostat predicts.

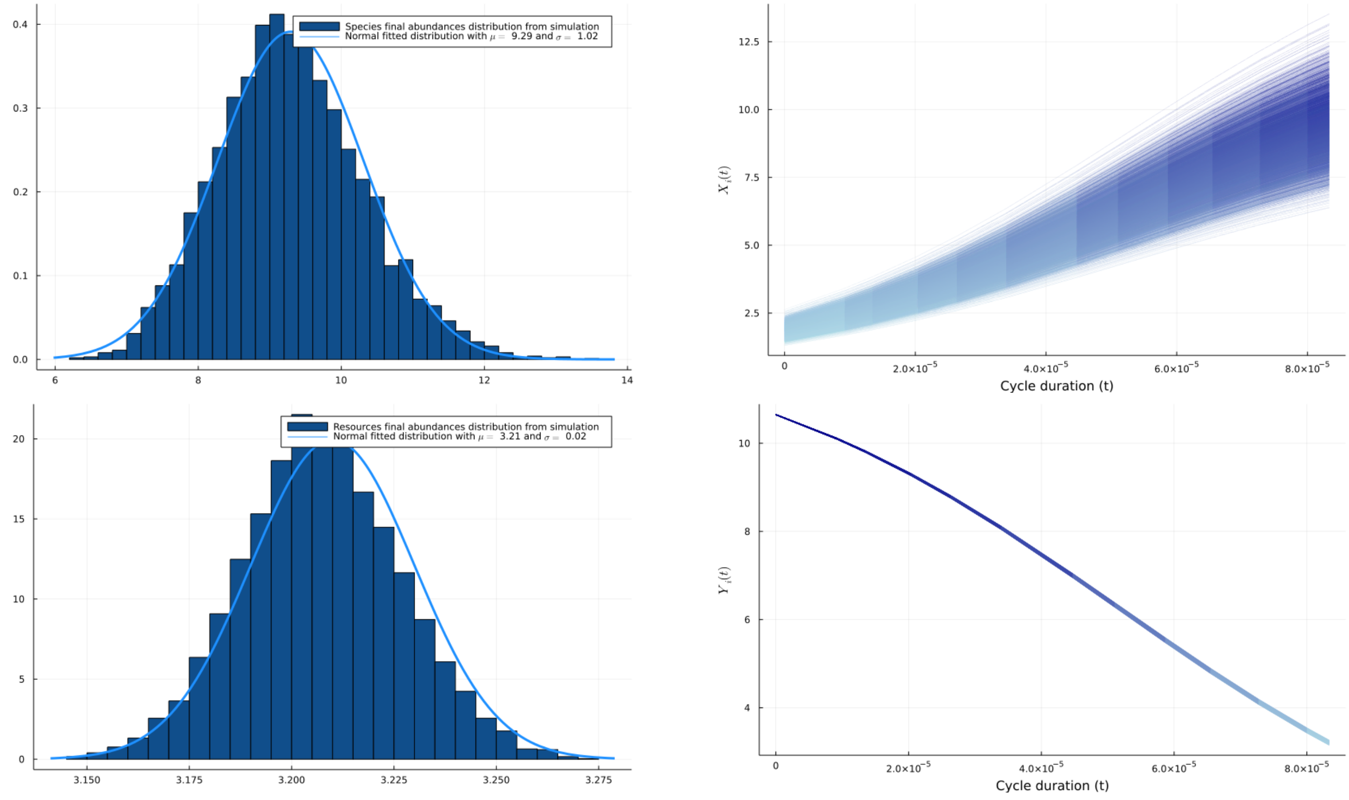


Figure 3: Simulation with $D = 5$, $t = 8.3 \cdot 10^{-5}$, abundance threshold $= 0.01$, $x = 0.01$ and $\mu = 3000$.

6 Serial dilution with stochastic resources replenishment after multiple dilutions

We can add some complications to the initial model for serial dilution presented in section 5. By following what B.Good 2022 stated in their article, during a sampling interval T, each dilution cycle was seeded with an initial amount of each resource, $Y_{\alpha,0}(T)$, and the dynamics was simulated until all resources were depleted ($\mathrm{d}Y_\alpha/\mathrm{d}t = 0$ for all α). The community was then diluted by a factor D and resources were replenished to their initial amounts $Y_{\alpha,0}(T)$. To mimic the effects of a reservoir of species that could potentially compete for the resources, we initialized the first dilution cycle of each sampling interval by assuming that S consumers were present at equal abundance. After that, more dilution cycles were run until an approximately ecological steady state was achieved. This roughly ecological steady state served as the basis for defining consumer abundances at sampling time T. According to the model's presumptions, dietary changes or other external fluctuations in the initial resource levels $Y_{\alpha,0}(T)$ could be the cause of any temporal variation in consumer abundances. We made the assumption that the initial resource levels experience a biased random walk around their average values Y_α in order to model these fluctuations:

$$Y_{\alpha,0}(T) = \left| Y_{\alpha,0}(T-1) - k\Big(Y_{\alpha,0}(T-1) - \overline{Y}_\alpha \Big) + \sigma \overline{Y}_\alpha \xi_\alpha(T) \right| \tag{6.1}$$

where ξ_α is a normally distributed random variable with zero mean and unit variance, σ determines the strength of how resources fluctuate, and k is the magnitude of a force ensuring the same resource environment on average over time. The absolute value require $Y_{\alpha,0}$ to be positive. If $k = 0$, $Y_{\alpha,0}(T)$ performs an unbiased random walk because there is no restoring force; if $k = 1$ $Y_{\alpha,0}(T)$ fluctuates about its set point $\overline{Y}_\alpha$ independent of its value at the previous sampling time. We used an ensemble approach to model the set points $\overline{Y_j}$, assuming that each $\overline{Y}_\alpha$ was independently drawn from a uniform distribution between 0 and Y_{max}. These assumptions yield a Markov chain of fluctuating resource amounts $Y_{\alpha,0}(T)$ and their corresponding consumer relative abundances $x_i(T) = X_i(T)/\sum_n X_n(T)$.

The total number of consumers in the reservoir S, the number of resources in the environment M, the sparsity S of the resource consumption matrix, and the resource fluctuation parameter σ and k are the five global parameters that have the greatest influence on the statistical properties of these time series. For our purposes, the absolute value of $Y_{\max}$ is not significant because it has no impact on the predictions of consumer relative abundances at ecological steady state.

6.1 Modeling consumption rates

We specifically looked at an ensemble where each $a_{i\alpha}$ was randomly chosen from a uniform distribution ranging from 0 to a_{max}. Each $a_{i\alpha}$ was set to zero with probability S to model the community's resource competition's sparsity. We can represent arbitrarily large communities using this ensemble approach and just two global parameters, S and a_{max}.

6.2 Implementing simulations

There are two different time scales that we need to consider: dilution and evolution. First of all we go back to the initial conditions of species and resources. Remember that each resource in this model represents a group of metabolites or a niche. This is because the model's definition of resources is coarse-grained. Initially, we have an equal number of consumers $X_i(0)$ $\forall i$ and resources $Y_{\alpha,0}(0)$. Consumers then compete until all resources have run out over a predetermined amount of time t. When this occurs, consumers are diluted by a factor D, introducing a new larger time scale T with respect to the consumer-resource dynamics, and resources are restored according to the Markov process of $Y_{\alpha,0}(T)$. Consumers do, in fact, participate in a cycle of resource depletion during time t, and they are subsequently diluted by a factor of D. However, on a larger scale, both Y_α and X_i behave like random variables.

An ecological steady state is reached when the dynamics in subsequent passages are identical, which is the case when all consumers are either extinct or have a growth ratio (the ratio between consumer's final and initial abundances within one passage, time scale t) equal to the dilution factor D. Due to the slow path of extinction of some consumers, reaching an exact ecological steady state is really time demanding, i.e we should implement a huge number of dilution cycles, even hundreds of passages. Thus, we assumed instead that between sampling times the system only approximately reaches an ecological steady state, defined as the growth ratios of all species changing by less than a threshold between subsequent passages that was defined as D plus a small fraction. We investigate now what it actually means to implement such an approximate steady state condition. Indeed we define a steady state when

$$\frac{X_f^i}{X_0^i} = D \tag{6.2}$$

It means that the fixed points for final abundances have been reached by all the species and their growth rate has been fixed equal to the dilution factor. If now we recall that

$X_f^i = DX_0^{i+1}$ we realize that the condition simplifies to:

$$\frac{X_0^{i+1}}{X_0^i} = 1 \tag{6.3}$$

Because of the numerical implementation and the finite computational precision we may want to set a threshold for the validity of the before mentioned condition. That is we require that the initial abundances of the subsequent step are different from the previous one more or less than $\varepsilon \cdot 100\%$.

$$\left| \frac{X_0^{i+1}}{X_0^i} - 1 \right| < \varepsilon \tag{6.4}$$

In B.Good 2022, they decide to first implement a bunch of dilution steps and then, by changing realization (sampling in a different time T), they vary noise added to $Y_{0,j}$ so that they make sure that each point they are taking into account for the time series is at stationary state. That is consumers reach after $6-8$ dilution steps their steady state, at least approximately. This process corresponds to a serial dilution in which noise for $Y_{0,j}$ is taken into account every n dilution steps (for example $n = 6$ if we want to recall the previous process setting)

6.3 One species and one resource case

If we consider one species and one resource we can easily derive some useful insights about the time series of both X_0^i and Y_0^i.

$$\begin{aligned} \dot{X} &= XRY \\ \dot{Y} &= -XRY \end{aligned} \tag{6.5}$$

That leads to

$$\begin{aligned} X(t) - X(0) &= Y(0) - Y(t) \\ X_\infty - X(0) &= Y(0) - \cancel{Y_\infty} \end{aligned} \tag{6.6}$$

We can write then the time series equations for both the species and the resource

$$X_0^i = DX_0^{i+1} + Y_0^i \tag{6.7}$$

Eventually we will have these two time series (forget about the zero pedix since it is not important for the derivation)

$$\begin{cases} Y^i = Y^{i-1} + k(\overline{Y} - Y^{i-1}) + \sigma \overline{Y} \xi^i \\ X^i = DX^{i-1} + Y^i \end{cases} \tag{6.8}$$

We can show that X^i is also a random variable following a Gaussian distribution in these way. We may derive the complete time series for X^i up to X^1 or just noticing that it is a sum of Gaussian random variables coming from Y^i. We can also make some calculations to theoretically compute the Taylor's law that mean species abundance and its variance are following if we are in the simple case of one species and one resource. Indeed:

$$X^{(i)} = D^{i+1}X^{(1)} + \sum_{j=1}^{i} D^j(\text{noise}) \tag{6.9}$$

Notice that the noise will come from the complete time series of Y^i and the exponent of D is a power not an index of the time series. Steps indices are indicated by parenthesis in these ambiguous cases. We see now the correlation between $\langle X^i\rangle$ and $\langle (X^i - \overline{X})^2\rangle$. If we multiply by X^i, X^{i-1}, Y^i we will end up with these system of equations recalling that $\overline{X} = \overline{Y}/(1-D)$

$$\begin{cases} \langle X^{i-1}X^i\rangle = D\langle X^2\rangle + \langle Y^iX^{i-1}\rangle \\ \langle X^2\rangle = D\langle X^{i-1}X^i\rangle + \langle Y^iX^i\rangle \\ \langle X^iY^i\rangle = D\langle X^{i-1}Y^i\rangle + \langle Y^2\rangle \end{cases} \tag{6.10}$$

From the third equation of the system we realize that X^i and X^{i-1} are, for different realizations of the process, the same random variable therefore

$$\langle X^{i-1}Y^i\rangle(1-D) = \langle Y^2\rangle \tag{6.11}$$

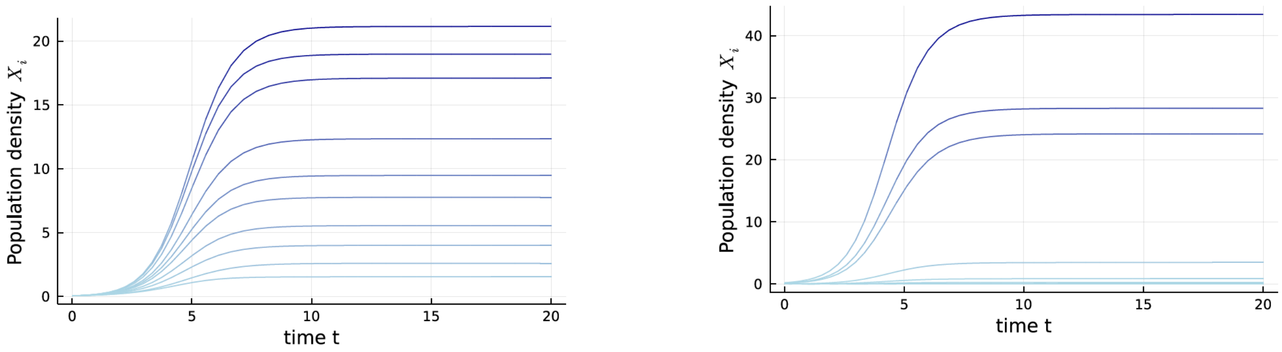
We end up with

$$\langle X^2\rangle = \frac{\langle Y^2\rangle}{(1-D)^2} \tag{6.12}$$

If now we compute $\langle (X^i - \overline{X})^2\rangle$ as $\langle X^2\rangle - \langle X\rangle^2$ and $\langle (Y^i - \overline{Y})^2\rangle$ we will have:

$$\begin{aligned} \langle (Y^i - \overline{Y})^2\rangle &= \frac{\overline{Y}^2\sigma^2}{(1-k)^2} \\ \langle (X^i - \overline{X})^2\rangle &= \frac{\overline{Y}^2\sigma^2}{(1-k)^2(1-D)^2} = \frac{\overline{X}^2\sigma^2}{(1-k)^2} \end{aligned} \tag{6.13}$$

Mean species abundance $\overline{X}$ and its variance $\langle (X^i - \overline{X})^2\rangle$ follow a power law with exponent two.



(a) First dilution cycle.

(b) Sixth dilution cycle reaching the approx s.s.

Figure 4: Species abundances in two steps of the dilution process; the first one and the approximate steady state.

6.4 Simulations of the model and first results

In order to see whether an approximate stationary state is reached during simulation we can plot the species initial conditions $X_{0,j}$ at each dilution step and see what is the maximum growth ratio that a species can reach after n steps. Simulations in Figure 4 have been computed with fifty species and thirty resources and a steady state threshold $x = 5\%$, that is the maximum growth rate we would expect is $1 + 0.05$. We see that in Figures 5 and 6 that within six dilution steps no species overcome that threshold.

6.4.1 Species and resources time series

Both species and resources are behaving as in a Markov chain since we are adding in every step a random resource to the system. Stochasticity is therefore brought into species time series as we have seen in eq.(6.8). We notice that Figures 7,8 and 9 show the random behavior of both species and resources abundances in each sampling time.

6.4.2 Taylor's law

Another important result has been obtained in the case of multiple species and multiple resources regarding the Taylor law calculated in eq.(6.13). The theoretical result however is valid only when we take into account one species and one resource; in this case we simulated the model for fifty species and thirty resources therefore the exponent of eq.(6.13) will be different than two as we can see in Figure 10

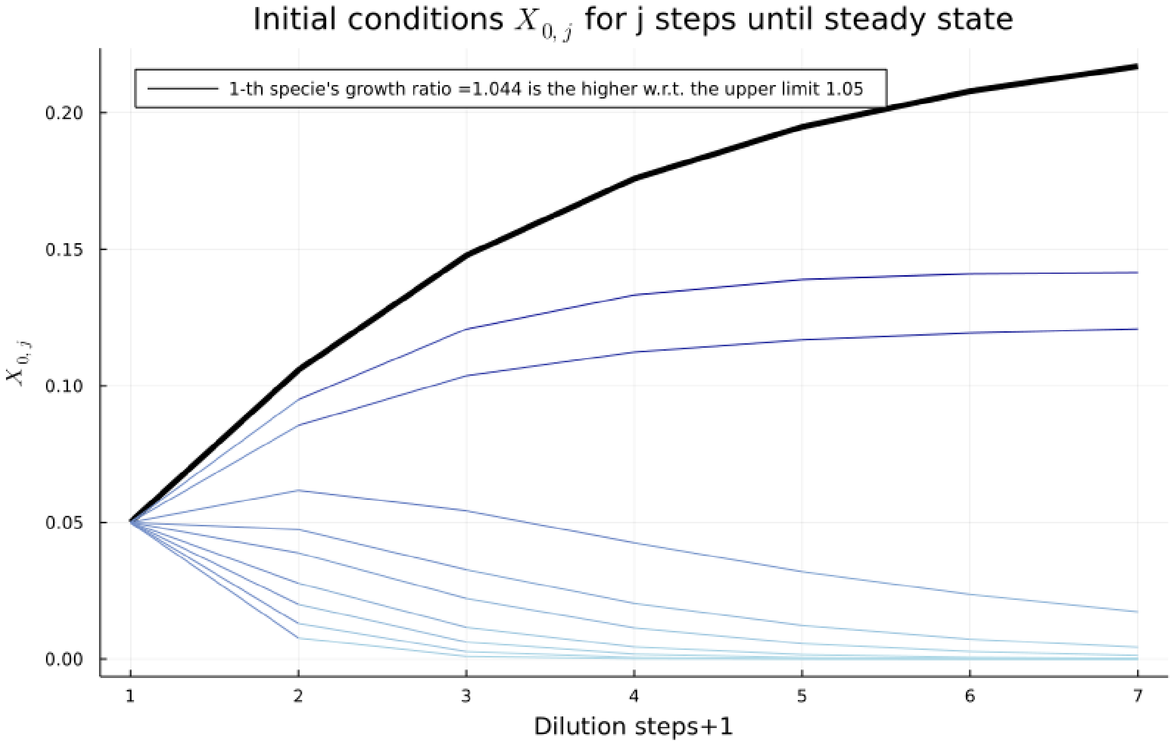


Figure 5: Species initial conditions $X_{0,j}$ versus the number of dilution steps. There is an additional dilution step since for the last one we are indirectly computing the initial condition of the next one. The curve in black represents the species that has the biggest growth rate.

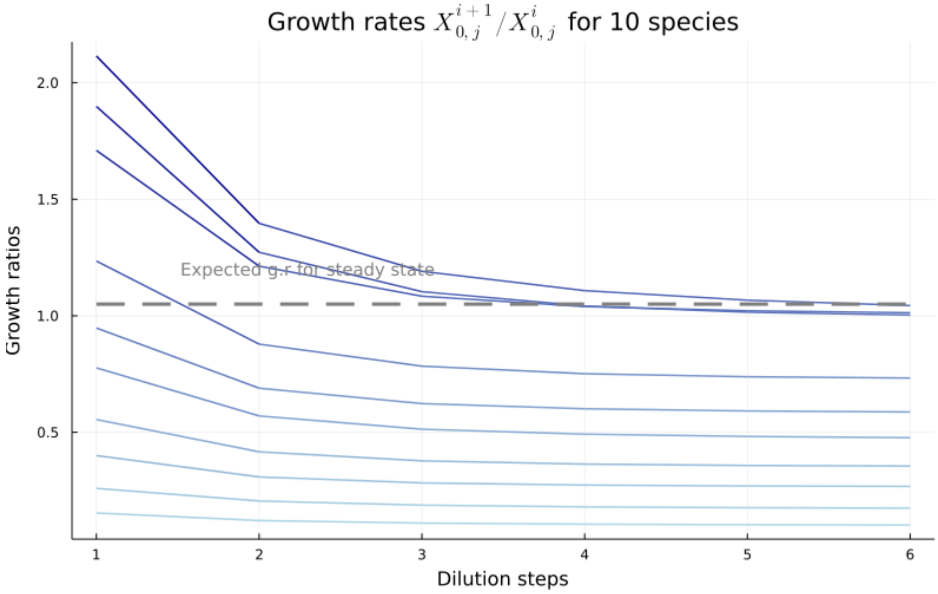


Figure 6: We are restricting the plot to 10 species and we are showing the growth rates of species versus the dilution steps until the stationary state is reached.

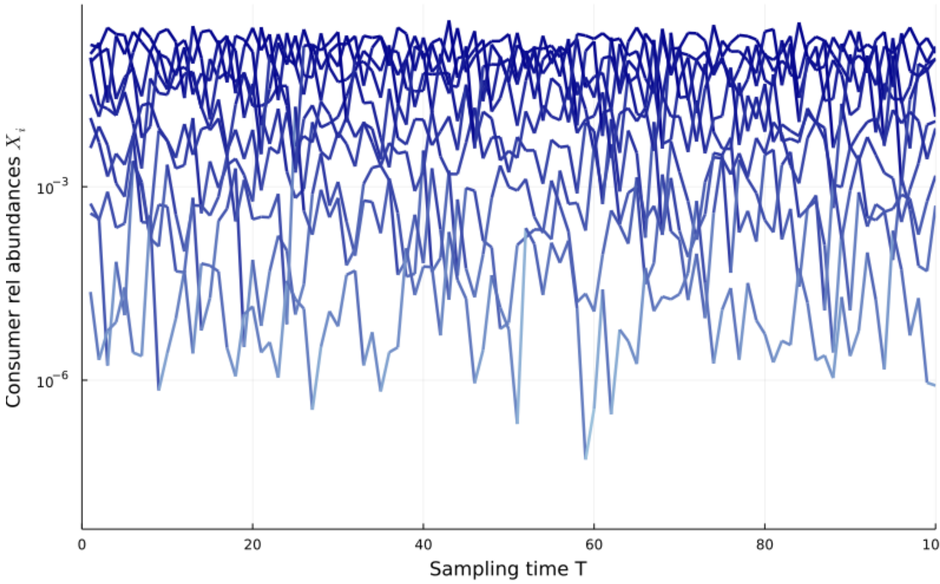


Figure 7: Species final abundances recorder after each dilution process realization reaches an approximate steady state.

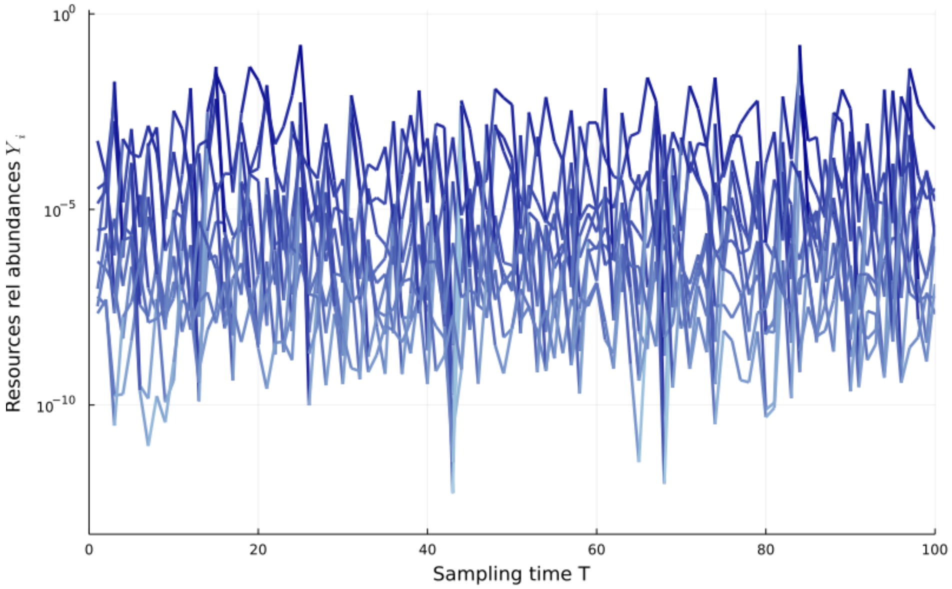


Figure 8: Resources final abundances recorder after each dilution process realization reaches an approximate steady state.

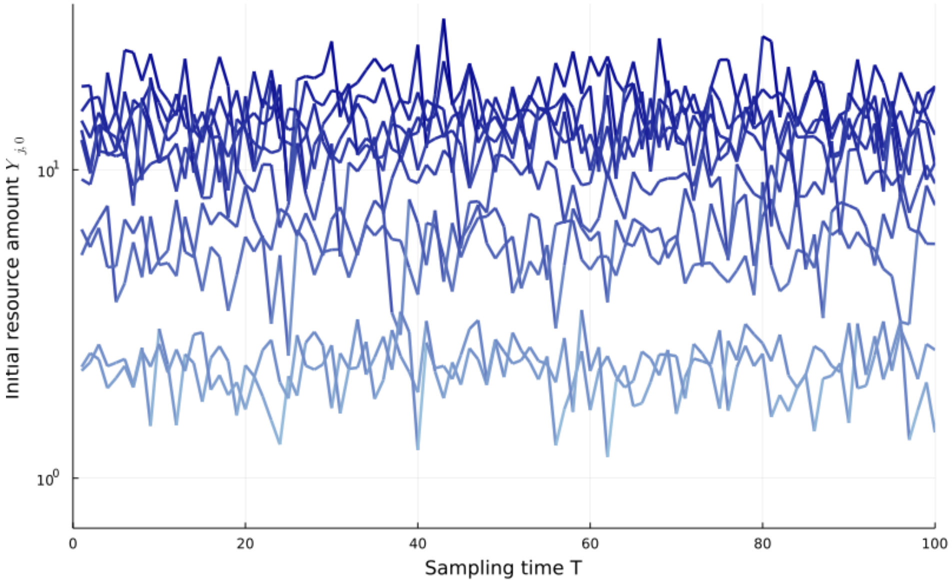


Figure 9: Initial amount of resources at each realization of a dilution process follow a Markov chain as we saw in eq.(6.8).

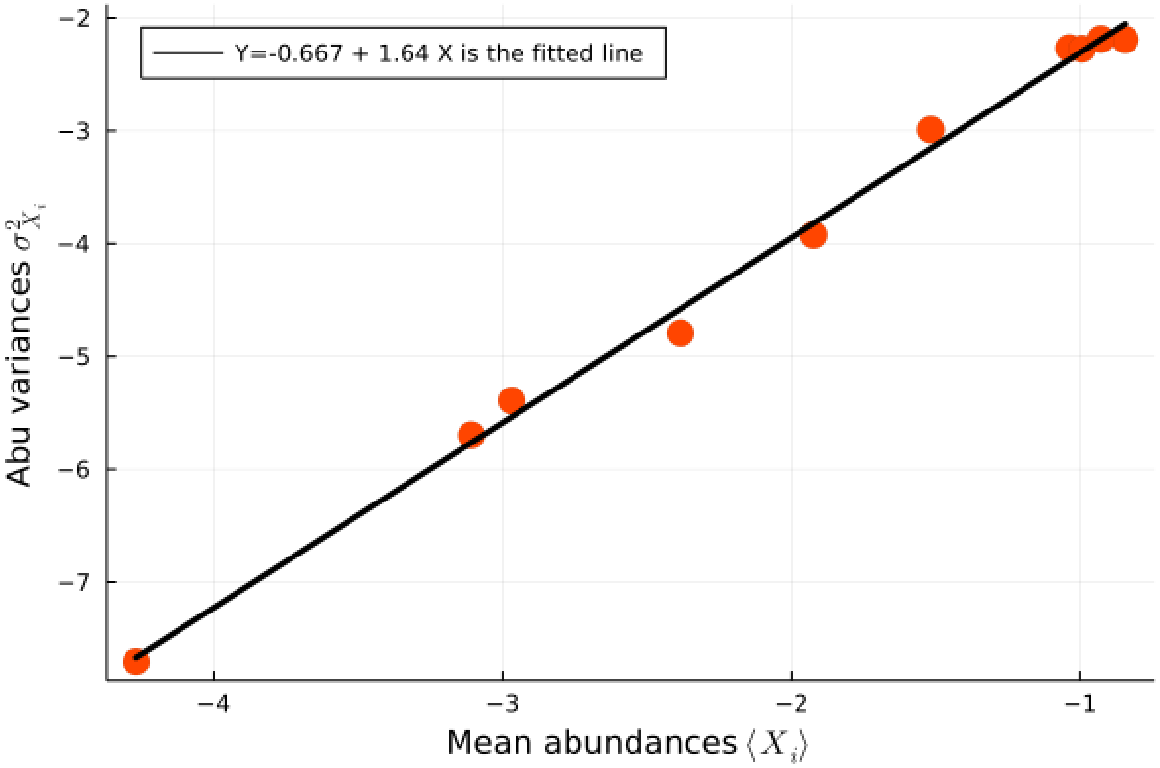


Figure 10: Species abundance variance $\sigma^2_{X_i}$ versus its mean $\langle X_i \rangle$. Each point represent one species while statistics is computed across each species samples recorded at the end of a realization, T, of a dilution process.

7 Serial dilution with stochastic resources replenishment after each dilution

As before we need to sample only those points of consumer abundances such that they reached an approximate steady state. In order to collect them we have to reach first that steady state. Instead of waiting the dilution process to reach stationarity with the same initial resources at each dilution step and only after that changing them according to the Markov chain for a new realization of the process, we will try something different. We want that resources are replenished at the beginning of a dilution step randomly following the Markov chain prescription. That is because it is more natural that animals are fed every day, for example, with different resources that maybe have memory of the day before (like diets indications). It means that we are supposed to change initial conditions for resources randomly at each dilution step instead of waiting the species abundances to reach an approximate steady state.

7.1 Time series abundances analysis, one species one resource

In order to analyze the process with a bottom top approach we choose to reduce the system to one species and one resource. Many interesting results can be obtained starting from this simple setting where we are surely neglecting inter-specific competition among species. The consumption matrix reduces to a random variable drawn from a uniform distribution $a \in [a, a_{max}]$.

7.1.1 Gaussian noise for resources, evolution $T \longrightarrow \infty$

Each dilution step resources are replenished following a Markov chain without memory. Noise is added at each dilution, randomly drawn from a Gaussian distribution with mean $\overline{Y}$ and σ standard deviation.

$$\begin{cases} \dot{X} = \mu XY \\ \dot{Y} = -\frac{\mu}{K} XY \end{cases} \tag{7.1}$$

If we write $\frac{\mathrm{d}X}{\mathrm{d}Y}$ we will find

$$\frac{\mathrm{d}X}{\mathrm{d}Y} = -\frac{1}{K} \tag{7.2}$$

Integrating between the initial and final state, if the final one corresponds to that such $Y(\infty)$ is equal to zero

$$X_t - X_0 = \frac{Y_0}{K} \longrightarrow X_t = X_0 + \frac{Y_0}{K} \tag{7.3}$$

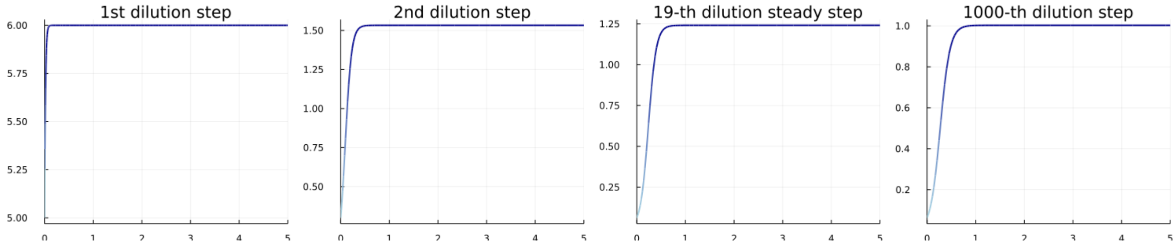


Figure 11: One species is growing in each dilution cycle until it eats all the resource. There are at most 1000 dilution steps and that species reaches the steady state in the 19-th step.

If $K = 1$ and by separating the time scales between the single evolution step and the whole dilution cycle with i steps we see that

$$X_t = X_f^{i+1}, \quad X_f^{i+1} = DX_0^{i+1}, \dots, D^i X_0^1 \qquad i = 1, \dots, T \tag{7.4}$$

therefore

$$X_0^{i+1} = \frac{X_0^i + Y_0^i}{D} \tag{7.5}$$

By taking into account the time evolution of Y_0^i

$$Y_0^i = \overline{Y} + \xi_x, \qquad \xi_x \sim \mathcal{X}(0, \sigma) \tag{7.6}$$

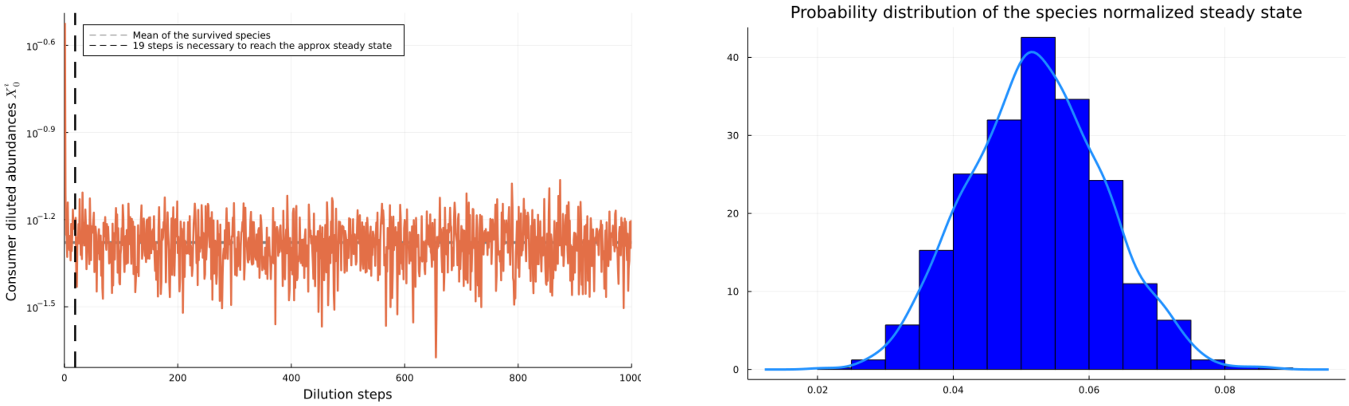
$$X_0^{(i+1)} = \frac{X_0^{(i)} + \overline{Y} + \xi_x}{D} \tag{7.7}$$

$$X_0^{(i+1)} = \sum_{n=1}^{i} \frac{\overline{Y}\xi_x}{D^n} + \frac{X_0^{(1)}}{D^i} \tag{7.8}$$

ξ is without the exponent since it is just a random variable while D^i means D to the power of i (remember that in case of ambiguity there are parenthesis to simbolize the steps). The sum of Gaussian random variables is Gaussian. In Figure 11 we see how that one species is evolving during each dilution step until it reaches the approximate steady state. In Figures 12a and 12b there are the simulated time series of species abundances at the end of each dilution step and the correspondent distribution.

7.1.2 Gaussian noise for resources, evolution $T \longrightarrow 0$

If we change the integration time of each dilution step we don't expect a Gaussian distribution anymore but a log normal since we would have the product of exponential



(a) Time series of the recorded abundance of one species at the end of each dilution process.

(b) Distribution of the time series points of species abundance for 1000 dilution steps.

Figure 12: Parameters of the simulations: S $=1$, M $=1$, Sparsity $=0.1$, $\sigma = 0.2$, $k = 1$, $a_{max} = 5$, $Y_{max} = 100$, $\overline{Y} = 50$, steps $= 1000$, tspan $= 5$, steady_threshold $= 0.01$.

terms containing noise. Such time span is not enough to let the abundances reach the steady state. There are also conditions to meet in terms of the dilution factor, the initial resource abundance, time and interaction strength in order to have non zero consumer abundances as we are going to see. If we start again from the relation between one species and one resource and their initial and final abundances:

$$Y_t = Y_0 + K(X_0 - X_t) \tag{7.9}$$

$$\frac{\mathrm{d}X}{\mathrm{d}t} = \mu Y_t X_t \tag{7.10}$$

By substituting Y_t into the dynamics of X_t we get

$$\frac{\mathrm{d}X}{\mathrm{d}t} = \mu X_t (Y_0 + K(X_0 - X_t)) \tag{7.11}$$

That is a logistic equation indeed

$$\frac{\mathrm{d}X}{\mathrm{d}t} = \mu (Y_0 + KX_0) X_t \left(1 - \frac{X_t}{\frac{Y_0 + KX_0}{K}}\right) \tag{7.12}$$

if $r = \mu(Y_0 + KX_0)$ and $\overline{K} = \frac{Y_0 + KX_0}{K}$ we get

$$X_t = \frac{X_0 \overline{K} \mathrm{e}^{rt}}{(\overline{K} - X_0) + X_0 \mathrm{e}^{rt}} \tag{7.13}$$

if $Y_0 \gg KX_0$ and for small t we can consider

$$X_t = \mathrm{e}^{\mu(Y_0+KX_0)t} X_0 \tag{7.14}$$

If we make explicit the dilution process condition and for X_0^i

$$X_0^{i+1} = \frac{X_0^i}{D} e^{\mu(Y_0+KX_0)t}, \quad \text{if } Y_0^i \gg KX_0^i \tag{7.15}$$

$$X_0^{i+1} \approx \frac{X_0^i}{D} e^{\mu Y_0^i t} \tag{7.16}$$

if Y_0^i noise is without memory again

$$X_0^{(i+1)} = \left(\frac{e^{\mu \bar{Y} t}}{D} \right)^i e^{\mu t (\sum_{n=1}^{i} \xi_x)} X_0^{(}1) \tag{7.17}$$

with the same considerations on D's exponent and ξ_x. We see that $Z = e^{\xi}$ if $\xi \sim \mathcal{X}(0, \sigma)$, that is Z is a log-normal. In both situations we can state what is the condition that ensures the survival of that species after the evolution time T. By starting from the resource evolution.

$$\begin{cases} \frac{\mathrm{d}X}{\mathrm{d}t} = \mu XY \\ \frac{\mathrm{d}Y}{\mathrm{d}t} = -\frac{\mu}{K} XY \end{cases} \tag{7.18}$$

if $K = 1$ then we can easily write that

$$X_t - X_0 = Y_0 - Y_t \longrightarrow Y_t = Y_0 + X_0 - X_t \tag{7.19}$$

then substituting that solution into the differential equation describing the evolution of species we have

$$\frac{\mathrm{d}X}{\mathrm{d}t} = \mu X (Y_0 + X_0 - X) \tag{7.20}$$

if we consider that both X_t and X_0 are much smaller than the initial condition of resources then: $X_0 \ll X_t \ll Y_0$ we end up with

$$\frac{\mathrm{d}X}{\mathrm{d}t} \approx \mu X Y_0 \longrightarrow X(t) = X(0) e^{(\mu Y_0)t} \tag{7.21}$$

And finally by rewriting $X(t) = X_f^{i+1} = D X_0^{i+1}$ we get

$$X_0^{i+1} = \frac{X_0^i}{D} e^{(\mu Y_0)t} = X_0^i \; e^{(\mu Y_0 t - \log D)} \tag{7.22}$$

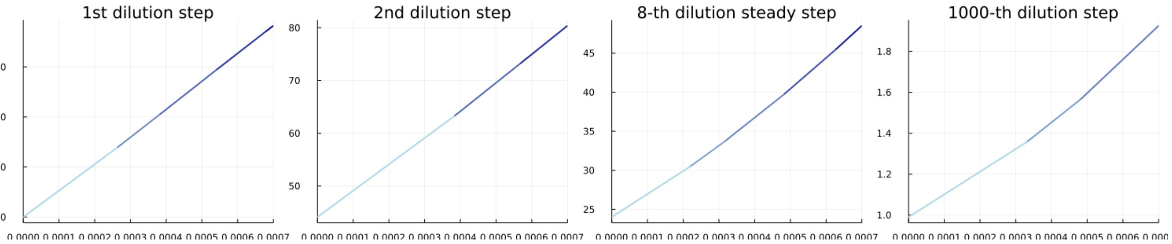
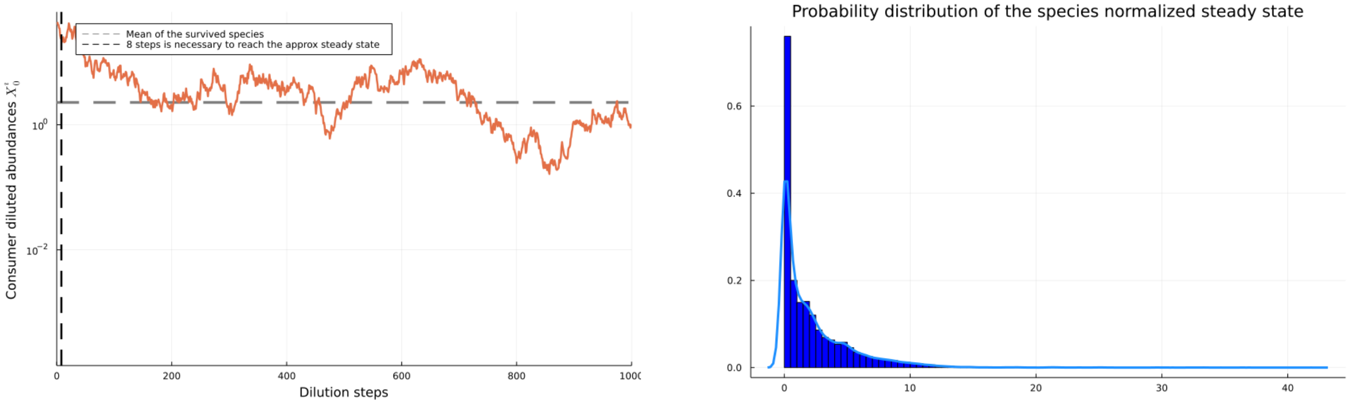


Figure 13: One species is growing in each dilution cycle until it eats all the resource. There are at most 1000 dilution steps and that species reaches the steady state in the 8-th step.



(a) Time series of the recorded abundance of one species at the end of each dilution process.

(b) Distribution of the time series points of species abundance for 1000 dilution steps.

Figure 14: Parameters of the simulations: S = 1, M = 1, Sparsity = 0.1, $\sigma = 0.2$, $k = 1$, $a_{max} = 5$, $Y_{max} = 100$, $\overline{Y} = 50$, steps = 1000, tspan = 0.007, steady_threshold = 0.01.

The only way X_0^{i+1} does not decrease is if the exponent of the exponential is greater or equal than zero, that is

$$\mu Y_0 t \geq \log D \tag{7.23}$$

In Figure 13 we see how that one species is evolving during each dilution step until it reaches the approximate steady state. In Figures 14a and 14b there are the simulated time series of species abundances at the end of each dilution step and the correspondent distribution.

8 Serial dilution with constant resources replenishment and multiple species

We consider now another scenario, a simpler one where resources are replenished at each time step with the same constant value $\overline{Y}$. The consumption matrix is chosen in such a way that its entries are drawn, as we will see, from a probability distribution, but they are set to zero following a sparsity parameter $S_p = 0.1$. It means that on average each row of the random matrix will have ten percent of its entries set to zero. As a direct physical consequence we are saying that species have a small overlap in their predation niches and a smaller fitness. Since there is no way that species decrease in abundance during a dilution cycle, there could be only two reasons for extinction:

- In the first cycles one or more species don't grow so much such that they reach the surviving threshold
- At the end of a dilution cycle, we dilute some species abundances so that they are too low and they end up in a smaller abundance than the above mentioned threshold

8.1 Model's properties

We can dig into some properties of the model:

$$\begin{aligned} \dot{X}_i &= X_i \sum_{\alpha=1}^{M} a_{i\alpha} Y_\alpha \\ \dot{Y}_\alpha &= -\frac{1}{Y} Y_\alpha \sum_{i}^{S} a_{i\alpha} X_i \end{aligned} \tag{8.1}$$

where S species and M resources are present in the system, X_i are the species abundances, Y_α the resources concentration, Y the consumption timescale and the consumer matrix is generated as:

$$a_{i\alpha} = \frac{\mu}{M} + d_{i\alpha}, \tag{8.2}$$

We see that μ is the deterministic mean and the quenched disorder d is generated as independent uniform random variable with zero mean and standard deviation $1/\sqrt{M}$. We consider a dynamics, with some initial conditions $X_i^0(0)$ and $Y_\alpha^0(0)$ where after initialization, the system evolves under eqs.(8.1) for a fixed time T, that we call a cycle. Then, after a cycle k, the new $k+1$ is initialized with diluted conditions:

$$\begin{aligned} X_i^{k+1}(0) &= \frac{X^k(T)}{D} && (8.3) \\ ,Y_\alpha^{k+1}(0) &= \frac{Y^k(T)}{D} + \overline{Y}, && (8.4) \end{aligned}$$

where the apex k indicate the cycle and D is the dilution factor. We are interested in studying the stationary state induced by this dynamics. First, consider some interesting aspects of the model. Consider the average species abundance and resources concentration:

$$\begin{aligned} \langle X \rangle_t &= \sum_i \frac{X_i}{S} && (8.5) \\ \langle Y \rangle_t &= \sum_\alpha \frac{Y_\alpha}{M} && (8.6) \end{aligned}$$

their time derivatives are related by:

$$\begin{aligned} \langle \dot{Y} \rangle &= -\frac{\gamma}{Y} \langle \dot{X} \rangle && (8.7) \\ & && (8.8) \end{aligned}$$

where $\gamma = S/M$ leading to:

$$\begin{aligned} \langle Y \rangle_t - \langle Y \rangle_0 &= \frac{\gamma}{Y} (\langle X \rangle_0 - \langle X \rangle_t) && (8.9) \\ \langle X \rangle_t &= \langle X \rangle_t + (Y/\gamma)(\langle Y \rangle_0 - \langle Y \rangle_t). && (8.10) \end{aligned}$$

As a first consequence, we have a conservation of the total biomass in a cycle:

$$B^k(t) = \langle X^k \rangle(t) + \frac{\gamma}{Y} \langle Y^k \rangle(t) = \langle X^k \rangle(0) + \frac{\gamma}{Y} \langle Y^k \rangle(0), \tag{8.11}$$

and that after dilution the total biomass is diluted and there is a influx of resources:

$$B^{k+1}(0) = \langle X^{k+1} \rangle(0) + \frac{\gamma}{Y} \langle Y^{k+1} \rangle(0) = \frac{B^k}{D} + \frac{\gamma \overline{Y}}{Y}. \tag{8.12}$$

Hence, it is easy to see that the biomass increases between cycles k and $k+1$ if the resource injection is big enough:

$$\begin{aligned} B^{k+1} &> B^k && (8.13) \\ \frac{\gamma \overline{Y}}{Y} &> B^k \left(\frac{D-1}{D} \right) && (8.14) \end{aligned}$$

and we can relate easily the k-th biomass with the initial one:

$$B^k = \frac{B^0}{D^k} + \frac{\gamma \overline{Y}}{Y} \sum_{l=0}^{k-1} D^{-l} \tag{8.15}$$

leading to:

$$\frac{D^{k+1}\gamma \overline{Y}}{Y(D-1)} \left(1 - (D-1)D^{-1}\frac{\gamma \overline{Y}}{Y} \right) > B^0 \tag{8.16}$$

As a second consequence, we can use the definition of the consumption matrix and eqs.(8.9) to make explicit the logistic growth of species:

$$\begin{aligned} \dot{X}_i &= X_i \sum_{\alpha=1}^{R} a_{i\alpha} Y_\alpha = X_i \left(\mu \langle Y \rangle_t + \sum_\alpha d_{i\alpha} Y_\alpha \right) \\ &= X_i \left[\mu \left(\langle Y \rangle_0 + \frac{\gamma \langle X \rangle_0}{Y} \right) - \frac{\gamma}{Y} \langle X \rangle_t + \sum_\alpha d_{i\alpha} Y_\alpha \right] && (8.17) \\ &= X_i \left[\frac{\mu}{Y} B_0 - \frac{\gamma}{Y} \langle X \rangle_t + \sum_\alpha d_{i\alpha} Y_\alpha \right]. && (8.18) \end{aligned}$$

We say the dynamics reaches a stationary state if, for all species and resources:

$$\begin{aligned} X_i^{k+1}(T) &= X_i^k(T), && (8.19) \\ Y_\alpha^{k+1}(T) &= Y_\alpha^k(T), && (8.20) \end{aligned}$$

using eqs.8.3 we obtain the condition including the dilution process:

$$\begin{aligned} X_i^k(T) &= D X_i^k(0) && (8.21) \\ Y_\alpha^k(T) &= D(Y_\alpha^k(0) - \overline{Y}). && (8.22) \end{aligned}$$

It is easy to calculate the stationary biomass. If $k \to \infty$ and $D > 1$,

$$\sum_{l=0}^{k-1} D^{-l} = \frac{D}{D-1} \tag{8.23}$$

and hence the total stationary biomass is independent of the initial conditions:

$$B^* = \frac{\gamma D \overline{Y}}{Y(D-1)} \tag{8.24}$$

In the particular case where at the end of the cycle resources are totally depleted this gives us the final stationary average species abundance. In the steady state, generally the number of surviving species would be $S^* < S$. Formally, we can integrate the model equations during a cycle and get:

$$\begin{aligned}
X_i(T) &= X_i(0)e^{\Lambda_i[T]} && (8.25)\\
\Lambda_i[T] &= \mu\int_0^T \langle Y\rangle_s ds + \sum_\alpha d_{i\alpha}\int_0^T Y_\alpha(s)ds && (8.26)\\
Y_\alpha(T) &= Y_\alpha(0)e^{-\Delta_\alpha[T]} && (8.27)\\
\Delta_\alpha[T] &= \mu\int_0^T \langle X\rangle_s ds + \sum_i d_{i\alpha}\int_0^T X_i(s)ds && (8.28)
\end{aligned}$$

Using the stationary conditions we get a restriction over the integrated growth/death rate:

$$\begin{aligned}
\Lambda_i^*[T] &= \log D && (8.29)\\
\Delta_\alpha^*[T] &= \log\left(\frac{Y_{\alpha(0)}}{D(Y_\alpha(0)-\bar{Y})}\right) = \log R_\alpha(T) && (8.30)
\end{aligned}$$

for all i, α and we have defined $R_\alpha(T) = Y_\alpha(0)/Y_\alpha(T)$. Consider first the species growth rate. Given that it is the same for all species, we have that:

$$\sum_\alpha d_{i\alpha}\int_0^T Y_\alpha(s)ds = \sum_\alpha d_{j\alpha}\int_0^T Y_\alpha(s)ds \qquad (8.31)$$

for all i, j. At the same time the average must be also the same value,

$$\langle\Lambda[T]\rangle = \log D = \mu\int_0^T \langle Y\rangle_s ds + \sum_\alpha(\sum_i \frac{d_{i\alpha}}{S})\int dsY_\alpha, \qquad (8.32)$$

if $S^* \to \infty$ we have that

$$\sum_i^{S^*} \frac{d_{i\alpha}}{M} \to 0 \qquad (8.33)$$

leading to:

$$\begin{aligned}
\mu\int_0^T \langle Y\rangle_s ds &= \log D && (8.34)\\
& && (8.35)
\end{aligned}$$

and hence for all i:

$$\sum_{\alpha} d_{i\alpha} \int_0^T Y_\alpha(s)ds = 0. \tag{8.36}$$

Consider now the resources death rate. By using eq.(8.9):

$$\Delta_\alpha[T] = \mu(\langle X\rangle_0 + Y\gamma^{-1}\langle Y\rangle_0) - Y\gamma^{-1}\mu \int_0^T \langle Y\rangle_s ds + \sum_i d_{i\alpha} \int_0^T X_i(s)ds \tag{8.37}$$

and in the stationary state by using eq.(8.34) we have:

$$\Delta^*_\alpha[T] = T\mu(\langle X\rangle_0 + Y\gamma^{-1}\langle Y\rangle_0) - Y\gamma^{-1}\log D + \sum_i d_{i\alpha} \int_0^T X_i(s)ds \tag{8.38}$$

Using the same argument as above and eq.(8.29), if $M \to \infty$ we have that:

$$\begin{aligned} \langle \Delta^*[T]\rangle &= \mu(\langle X\rangle_0 + Y\gamma^{-1}\langle Y\rangle_0) - Y\gamma^{-1}\log D & (8.39)\\ &= \langle \log Y^*_\alpha(T)\rangle = \Big\langle \log\left(\frac{Y_{\alpha(0)}}{D(Y_\alpha(0) - \bar{Y})}\right)\Big\rangle & (8.40)\end{aligned}$$

. Complementary, we can also study directly the statistical properties of $R_\alpha(T)$ by looking at:

$$\langle Y^*_\alpha(T)\rangle = \langle e^{\Delta^*_\alpha}\rangle = \exp\{\langle \Delta^*[T]\rangle + q_X(T)\} \tag{8.41}$$

where, using the properties of disorder:

$$q_X(T) = \langle (\sum_i d_{i\alpha} \int_0^T X_i(s)ds)^2\rangle = \int_0^T ds \int_0^T ds' \langle X(s)X(s')\rangle. \tag{8.42}$$

Putting everything together we get to:

$$\begin{aligned} \log\langle Y^*_\alpha(T)\rangle &= \langle \log Y^*_\alpha(T)\rangle + q_X(T) \\ &= T\mu(\langle X\rangle_0 + Y\gamma^{-1}\langle Y\rangle_0) - Y\gamma^{-1}\log D + q_X(T). & (8.43)\end{aligned}$$

If we try to simulate, for different realizations of the disorder (changing the consumption matrix entries randomly), the integral relations regarding

$$\left\{\int_0^T \mathrm{d}s Y_\alpha(s)\right\}_{\alpha=1,\dots,M} \qquad \left\{\int_0^T \mathrm{d}s X_i(s)\right\}_{i=1,\dots,X} \tag{8.44}$$

and then computing their averages, at stationarity, with respect to the disorder, we would find useful sanity checks for the model in terms of the integrated growth rate for species as we can see in eq.(8.34) and the vanishing term in eq.(8.33). Figure 15 shows indeed that the integral in eq.(8.34) is exactly *logD*, that is an integrated effective growth rate for species X_i.

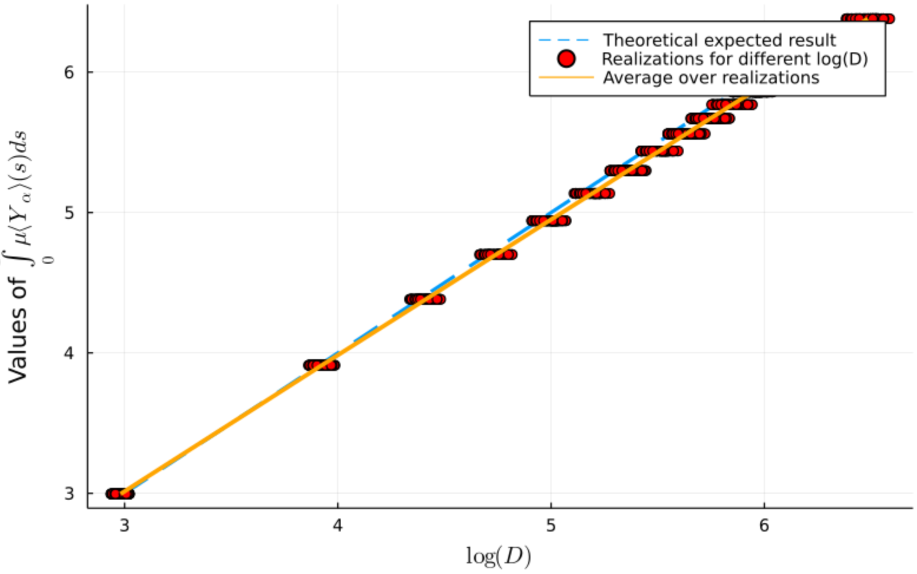


Figure 15: Each group of points is the integral of one of the resources over the integration time for a value of the dilution factor D. Averaging over the resources we have one point for each value of D that make the yellow line.

8.2 Adding a lower limit of growth rate of species

In doing the simulation we noticed that we were not imposing a lower limit to the growth rate of species abundances for each dilution step. There can arise some differences in the simulations, especially in the number of steps needed to accomplish the stationary state condition and the interpretation we are giving to that. From now on we are going to ask that on average species growth rates stay within a certain range determined by a fraction of the dilution factor:

$$\frac{1}{S}\sum_{j=1}^{S}\left|\frac{X_{j,0}^{i+1}}{X_{j,0}^{i}}-1\right|<\varepsilon \tag{8.45}$$

where $\varepsilon = 10^{-2}$ for example. Indeed from Figures 16,17 and 18 we see that survived species approach the approximate steady state without taking into account a lower bound for the decreasing growth rates but also in such a moment that there could be species decreasing more than we would expect; that is moving away from $\left|\frac{X_0^{i+1}}{X_0^i}-1\right| < \varepsilon \quad \forall i = 1,\ldots,$ dilution steps. The steady state however is reached more slowly with respect to the species that are actually surviving. It happens because we

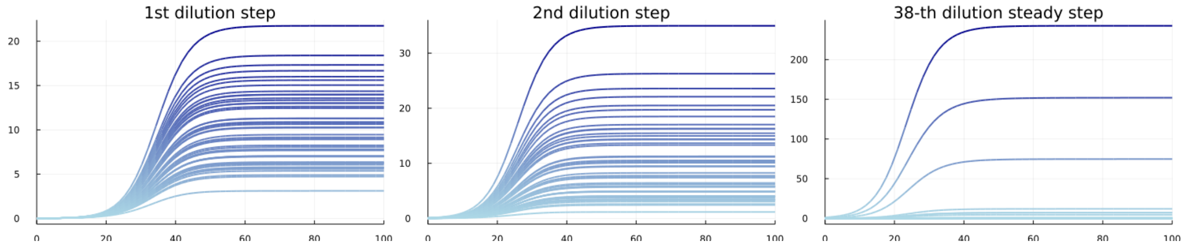


Figure 16: Species growing in each dilution step until they reach a *misleading* approximate steady state.

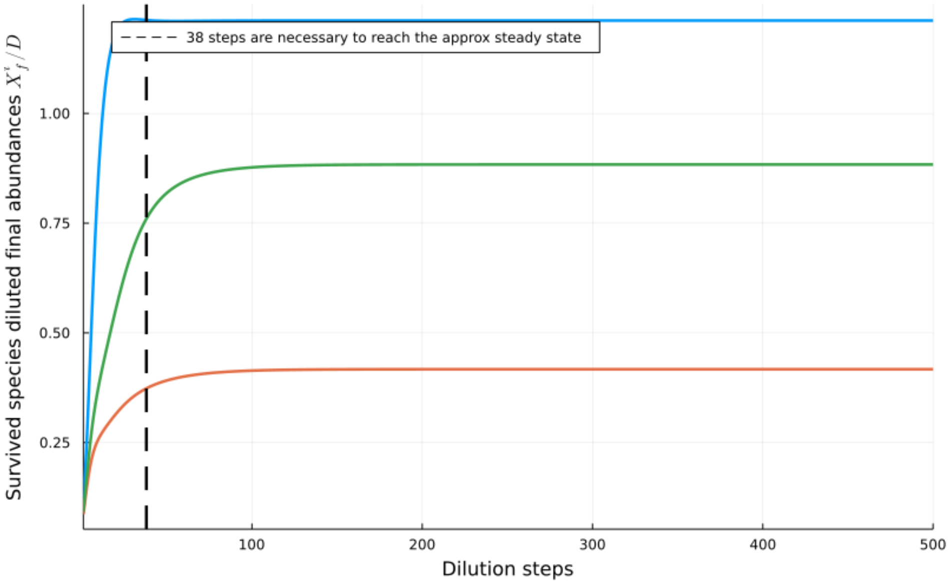


Figure 17: Final surviving species abundances versus the dilution steps.

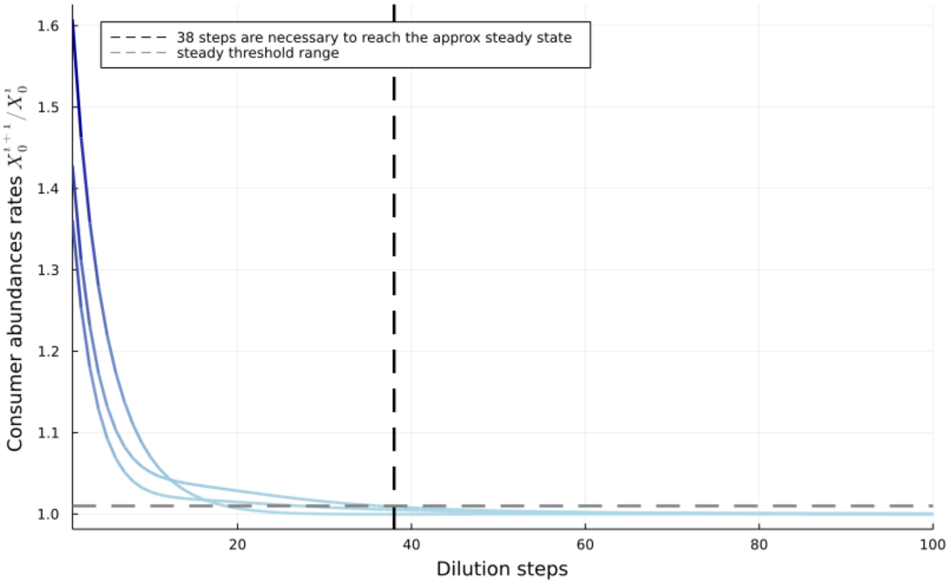


Figure 18: Species initial abundances rates across subsequent dilution steps.

are asking that also those that are going to be extinguished in the next iterations should meet the condition of reaching an initial abundance rate across steps within a given range. For this reason the extinguishing species are slowing down the cut off of the steady state of the system. Figure 19 shows that when adding a lower bound to the approximate steady state condition the number of dilution steps that are necessary to reach it increases. Furthermore in Figure 20 we see that if final species abundances are constant among different dilution steps it means that dilution D is the only factor that influences the process and we have reached the *approximate* steady state In Figure 21 we see that if we plot the rates of initial abundances between two subsequent dilution

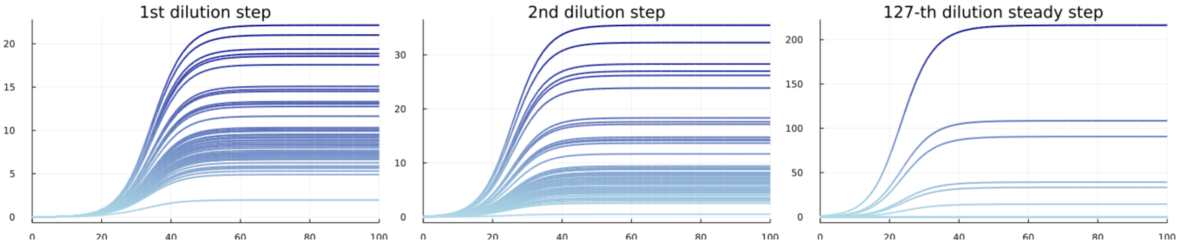


Figure 19: Species growing in each dilution step until they reach a refined approximate steady state.

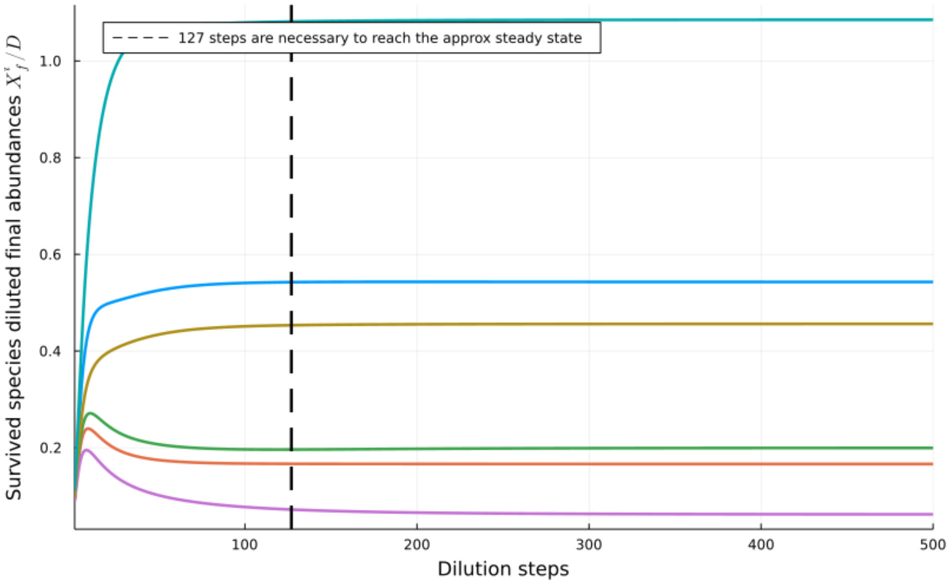


Figure 20: Final abundances of survived species versus the dilution steps.

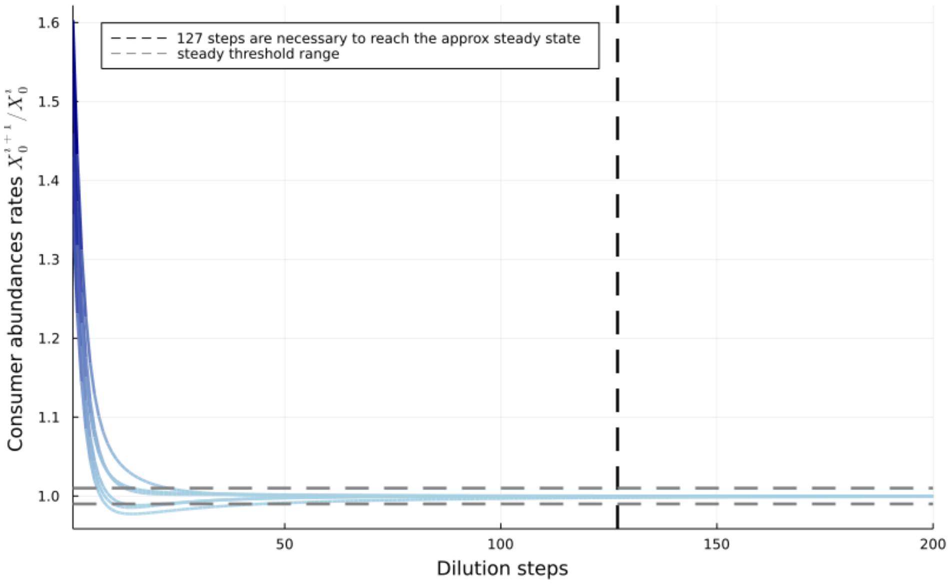


Figure 21: Initial species abundances ratio versus dilution steps.

steps we expect to see that the ratios reach the constant value of one since the total species abundances have reached stationarity; evolution in each cycle is therefore only influenced by the dilution. The new function that finds the steady step should return each dilution step the number of species that are surviving, that is those that succeeds to overcome the threshold that we chose beforehand. We are asking then that the ones dying are not exceeding an abundance of 10^{-2}. Furthermore in each dilution step we are changing the equations to be solved in order to contain only the number of species that are surviving so that remaining resources are shared among them. As a direct consequence the remaining species will have an absolute abundance in individuals that is larger (not that much since we are deleting species that are not able to survive and therefore that had a small final abundance) than before since now they can eat all the resources without competing with those that are now dead. We are not deleting resources since in every dilution step they are replenished with a constant value. Now one question arises: in each dilution step, in order to reach the approximate steady state and stop the simulation we are not taking into account those species that are dying in the subsequent step but still those that are going to be under the surviving threshold in more than one dilution step;that is we are not adding "memory" in the future to the steady state condition. Is this computation useful? Yes, it is because even if some species are going to die they are influencing, by competition, the growth rates of those that are surviving in the end. In Figure 22 we see the dilution factors D that are obtained calculating the ratio between the final and initial abundances of species in the last dilution step. Then we repeat the process by changing the consumption matrix and average the results over this disorder. We deduce then that the growth rates at stationarity, after the number of steady steps we found before, are within the desired limit, that is $D \pm Dx$ where x is fixed a priori and it is $\approx 10^{-2}$.

8.3 Normalization of the serial dilution setting

In order to simplify the evolution time scales of the simulations and make them coherent with respect to the change of parameters we can assign to species and resources initial conditions such values that we fix the dynamics of the system together with rescaled consumption matrix entries.

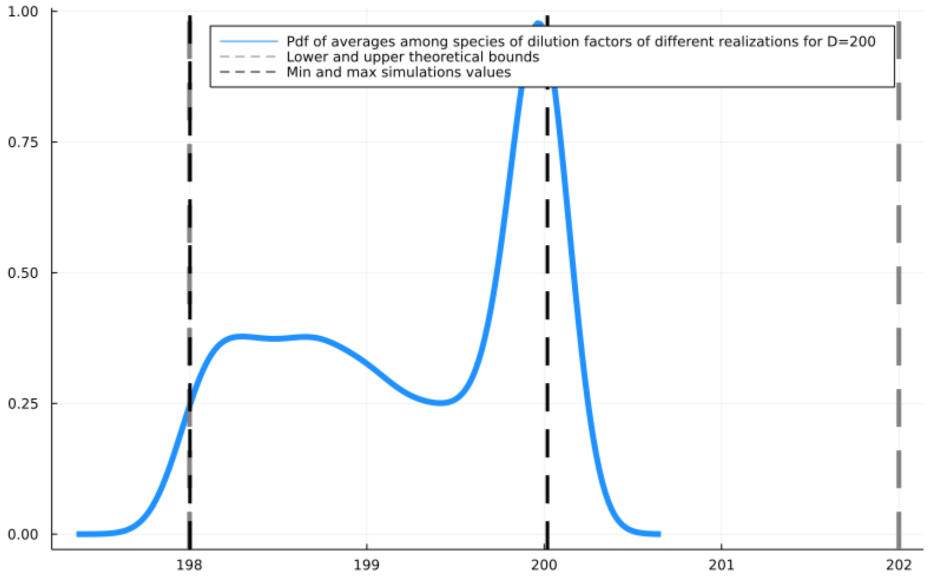


Figure 22: Dilution factors probability distribution averaged among different realizations of the process.

8.3.1 Species and resources initial conditions

First of all we can set the carrying capacity of the environment to 1 by rescaling the initial conditions of resources $Y_\alpha(0)$

$$\sum_{\alpha=1}^{M} Y_\alpha(0) = 1 \tag{8.46}$$

Recall that at each dilution step we are replenishing the amount of resources of a certain quantity $\overline{Y} = Y_\alpha(0),\ \forall\alpha$. As we can see from eq.(8.3), other than $\overline{Y}$ we are adding the diluted resources obtained as a result of the previous dilution step. However we realize that, since for each cycle resources have almost gone to extinction, that is they reach a total abundance less than 10^{-10}, we can say that the initial resource abundance in each cycle determines the carrying capacity of the system. Regarding species initial conditions we recall the eq.(8.24). Since at the steady state we expect that the total biomass of the system is only described by $X_i(T)$ we can easily derive the optimal initial condition of species that allows to reach the steady state. By exploiting the conserved quantity during a dilution step, that is the total biomass $B(t)$ $\forall t \in [0,T]$ we get:

$$B^*(T) \equiv \sum_i X_i^*(T) = B^*(0) \equiv \sum_i X_i(0) + \sum_\alpha Y_\alpha(0) \tag{8.47}$$

Since $\sum_i X_i^*(T) = D\sum_i X_i^*(0)$, if we define $Z(t) = \sum_\beta Z_\beta(t)$, where Z is valid for both species and resources, it follows that

$$\begin{aligned} X^*(0) &= \frac{Y^*(0)}{D-1} \\ X^*(T) &= \frac{Y^*(0)D}{D-1} \end{aligned} \tag{8.48}$$

It means that, as a sanity check for the system reaching stationarity we can compute the total final species abundance and compare it with $D/D-1$.

8.3.2 Normalized consumption matrix

On the other hand we can also fix the consumption dynamics of resources by changing the statistical moments of the consumption matrix.

$$a_{i\alpha} = \mu + d_{i\alpha}, \quad \text{where} \quad d_{i\alpha} \sim U(1-\sigma, 1+\sigma) \quad \text{and} \quad \mu = 1 \tag{8.49}$$

The main differences with respect to the previous choices concern the scalability of the mean and variance of these entries in terms of the number of resources. Furthermore the average $\langle a_{i\alpha}\rangle = 1$. It means that, for comparing results of different species we need to scale each time the values of σ and $\langle a\rangle$ proportionally to the number of resources M.

9 Further properties of the multispecies model

9.1 Stability of rescaled species abundances

In order to get a closed form for the probability distribution of the final abundances of the survived species let us try to simplify the model by looking at an important property that species abundances at stationarity may show. We write that $X_i(t) = n_i(t)\overline{X}(t)$ and plot the rescaled species abundances $n_i(t)$. In Figure 23 we observe at stationarity that n_i^* are constant during the dilution cycle. We would expect surely that each initial and final point of the rescaled species abundances falls on the same y position since at stationarity:

$$n_i^*(T) = \frac{X_i(T)}{\overline{X}(T)} = \frac{\cancel{D}X_i(0)}{\cancel{D}\overline{X}(0)} = n_i^*(0)$$

But on the other hand what we see is much stronger and can be analytically written as:

$$X_i(t) = n_i^*\langle X_i\rangle(t) \tag{9.1}$$

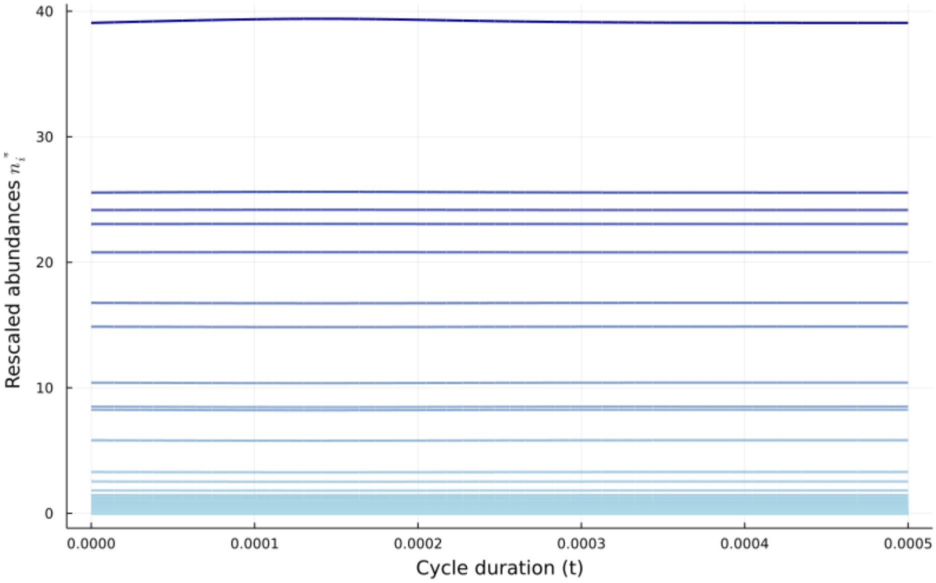


Figure 23: One realization of the serial dilution process with 5000 species and resources, $\mu = 1$, $D = 100$, $x = 0.01$ and abundance surviving threshold $= 0.01$.

and hence

$$\dot{X}_i(t) = n_i^* \langle \dot{X}_i \rangle (t) \tag{9.2}$$

If we call $\langle X_i \rangle = \overline{X}$, we can rewrite everything as:

$$\dot{X}_i(t) = n_i^* \dot{\overline{X}}(t) \tag{9.3}$$

We realize that, since during the stationary dilution cycle n_i^* are independent of time, their time derivative is zero, then:

$$\dot{n_i^*} = \frac{\dot{X}_i(t)}{\overline{X}(t)} - \frac{X_i(t)\dot{\overline{X}}(t)}{\overline{X}^2(t)} = 0 \tag{9.4}$$

Therefore

$$\frac{\dot{X}_i(t)}{X_i(t)} = \frac{\dot{\overline{X}}(t)}{\overline{X}(t)} \tag{9.5}$$

As a consequence it is easy to rewrite the dynamics of the species and resources in terms of their mean abundances over time.

$$\dot{\overline{X}}(t) = \overline{X}(t) \sum_{\alpha=1}^{M} a_{i\alpha} Y_\alpha(t) = \overline{X}(t) \left(\mu \overline{Y}(t) + \sum_{\alpha=1}^{M} d_{i\alpha} Y_\alpha(t) \right) \tag{9.6}$$

Since $\sum_{\alpha=1}^{M} d_{i\alpha} Y_\alpha$ is a linear combinations of independent random variables we can interpret them as a noise ξ_i and write the previous equation as

$$\dot{\overline{X}}(t) = \overline{X}(t)\left(\mu \overline{Y}(t) + \xi_i\right) \quad \forall i = 1, \dots S \tag{9.7}$$

Now we can make use of the conserved quantity in each dilution cycle that is the total biomass of species and resources $B(t) = \sum_i X_i(t) + \sum_\alpha Y_\alpha(t) \quad \forall t \in [0, T]$ to rewrite $\overline{Y}(t)$ in terms of $\overline{X}(T)$ and $\overline{X}(t)$.

$$B = \gamma^*(\cancel{\overline{Y}(T)} + \overline{X}(T)) = \gamma^*(\overline{Y}(t) + \overline{X}(t)) \longrightarrow \overline{Y}(t) = \overline{X}(T) - \overline{X}(t) \tag{9.8}$$

Since $\overline{Y}(T) \approx 0$ in the end of the steady state dilution step. We can eventually write the dynamical evolution of the mean species abundance $\overline{X}(t)$ as follows:

$$\dot{\overline{X}}(t) = \mu \overline{X}(t)\left(\overline{X}(T) - \overline{X}(t)\right) + \overline{X}(t)\xi \tag{9.9}$$

Where $\xi = 1/S^* \sum_i \xi_i$ because the previous $i = 1, \dots, S^*$ equations can be mediated over the number of survived species and they will be still valid. ξ then doesn't depend anymore on the species i.

9.2 Reduction to logistic

We see clearly that if the term that multiplies the noise in eq.(9.9) is subleading with respect to the rest the mean species abundance would behave like a logistic curve. In order to numerically prove that we can try to plot:

$$\log\left(\frac{\overline{X}(t)}{\overline{X}(T) - \overline{X}(t)}\right) - \log\left(\frac{\overline{X}(0)}{\overline{X}(T) - \overline{X}(0)}\right)$$

versus t and verify that the fitted line's slope is μ within a certain range of error. These quantities are obtained by integrating analytically eq.(9.9) without the noise

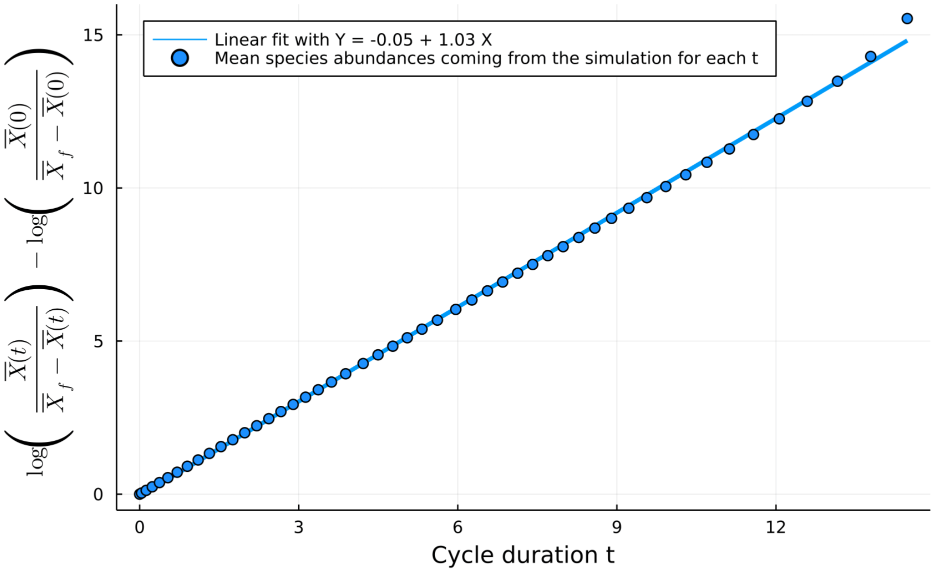


Figure 24: Logistic curve fit for the mean species abundance. Fitted line slope is 1.03 ± 0.2.

term. Indeed if we exponentiate both terms we end up with a logistic curve in terms of $\overline{X}(t)$:

$$\overline{X}(t) = \frac{\overline{X}(0)\mathrm{e}^{\mu t}}{1 - \frac{\overline{X}(0)}{\overline{X}(T)} + \left(1 - \frac{\overline{X}(0)}{\overline{X}(T)}\right)\mathrm{e}^{\mu t}} \tag{9.10}$$

Simulations give the result in Figure 24. First of all we proved also numerically that the mean species abundance can be rewritten as a logistic curve and that the noise term $\overline{X}(t)\xi$ is negligible since the line slope fall into the prescription of $\mu = 1$.
An initial sanity check to see whether the assumption regarding what we have seen in Figure 23 could be related to the form of the equation for resources evolution in time:

$$\dot{Y}_\alpha(t) = -Y_\alpha(t)\sum_{i=1}^{S} X_i(t)a_{i\alpha} \tag{9.11}$$

By substituting the assumption eq. (9.3) we end up with

$$\sum_{i=1}^{S} X_i(t)a_{i\alpha} = \overline{X}(t)\sum_{i=1}^{S} n_i^* a_{i\alpha}$$

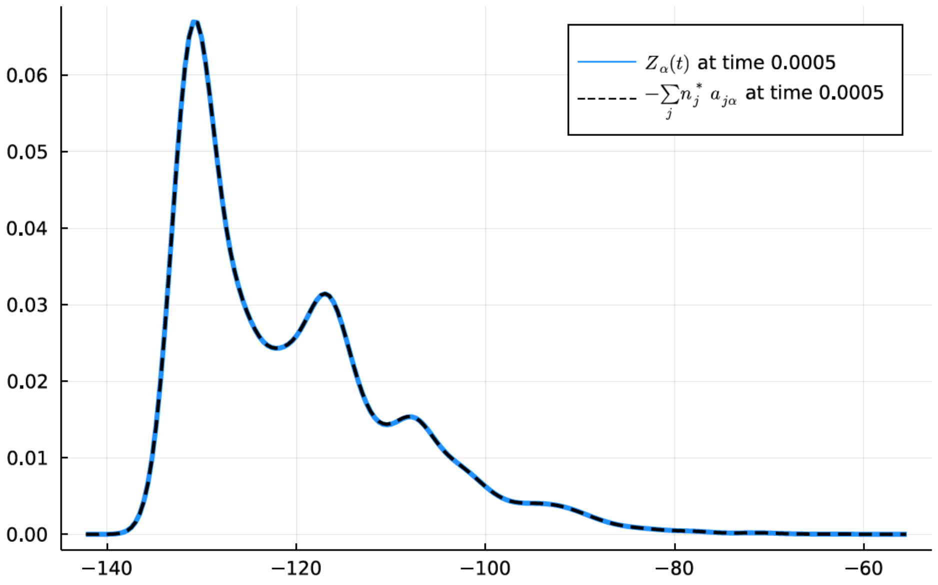


Figure 25: Resources abundances distribution and $-\sum_i n_i^* a_{i\alpha}$ distribution calculated at time $t = 15$ of the simulation

Integrating both hand sides we get

$$Z_\alpha(t) \equiv \frac{\log Y_\alpha(t) - \log Y_\alpha(0)}{\int_0^t \mathrm{d}s \overline{X}(s)} = -\sum_i n_i^* a_{i\alpha} \tag{9.12}$$

We could verify that actually the right hand side does not depend on time as we are assuming. In order to do that we are plotting in Figure 25 distributions of $Z_\alpha(t)$ and $-\sum_j n_j^* a_{j\alpha}$ at a random time t of the trajectories of n_i^*. We clearly see that the two sides of eq.(9.12) are equal and furthermore if we considered other time steps t the result would have been the same.

9.3 Dynamical Mean field theory for the serial dilution model

We will to use the dynamical cavity method to try solve the system in the stationary state. First, introduce a species death rate δ_i and a resource exit rate ω_α to be able to write the susceptibility, and then put them to zero:

$$\begin{aligned} \dot{X}_i &= X_i \left(\mu \langle Y \rangle_t + \sum_\alpha d_{i\alpha} Y_\alpha - \delta_i \right) \qquad (9.13) \\ \dot{Y}_\alpha &= -Y_\alpha \left(\mu \langle X \rangle_t + \sum_i d_{i\alpha} X_i + \omega_\alpha \right) \qquad (9.14) \end{aligned}$$

If we imagine to perturb the system by adding a new species and a new resources, with pedix 0, in small quantities we can think them as a perturbation of these death and exit rates:

$$\begin{aligned} \dot{X}_i &= X_i \left(\mu \langle Y \rangle_t + \sum_{\alpha/0} d_{i\alpha} Y_\alpha - (\delta_i - d_{i0} Y_0) \right) \qquad (9.15) \\ \dot{Y}_\alpha &= -Y_\alpha \left(\mu \langle X \rangle_t + \sum_{i/0} d_{i\alpha} X_i + (\omega_\alpha + d_{0\alpha X_0}) \right). \qquad (9.16) \end{aligned}$$

Now, we can think of expanding linearly the time values of the variables in such perturbations $\tilde{\delta}(t)_i = d_{i0} Y_0, \tilde{\omega}_\alpha(t) = d_{0\alpha} X_0(t)$, by defying the susceptibilities as coefficients of the Taylor series

$$\begin{aligned} \chi^Y_{\alpha\beta}(t,t') &= -\frac{\partial Y_\alpha(t)}{\partial \omega_\beta(t')} \qquad (9.17) \\ \chi^X_{i\beta}(t,t') &= -\frac{\partial X_i(t)}{\partial \omega_\beta(t')} \qquad (9.18) \\ \nu^Y_{\alpha j}(t,t') &= -\frac{\partial Y_\alpha(t)}{\partial \delta_j(t')} \qquad (9.19) \end{aligned}$$

$$\begin{aligned} & X_i(t) = X_{i/0}(t) - \sum_\alpha \int_0^t dt' \chi^X_{i\alpha}(t,t') d_{0\alpha} X_0(t') - \\ & - \sum_j \int_0^t dt' \nu^X_{ij}(t,t') d_{j0} R_0(t') \end{aligned} \qquad (9.20)$$

$$\begin{aligned} & Y_\alpha(t) = Y_{\alpha/0}(t) - \sum_\beta \int_0^t dt' \chi^Y_{\alpha\beta}(t,t') d_{0\beta} X_0(t') - \\ & - \sum_j \int_0^t dt' \nu^Y_{\alpha j}(t,t') d_{j0} R_0(t'). \end{aligned} \qquad (9.21)$$

Now, we can insert them back and obtain the following eq. (doing the same approx of G. P.Mehta M. 2018):

$$\dot{X}_0(t) = X_0 \left(\mu \langle Y \rangle (t) + \xi - \sigma^2 \int_0^t dt' \chi(t,t') X_0(t') \right) = X_0 \lambda_0(t). \tag{9.22}$$

At the stationary state, at the end of the cycle for the surviving species we have that:

$$\Lambda_0(T,0) = \int_0^T dt \left(\mu \langle Y \rangle (t) + \xi - \sigma^2 \int_0^t dt' \chi(t,t') X_0(t') \right) = \log D \tag{9.23}$$

By analysis above we know already that:

$$\int_0^T dt \mu \langle Y \rangle (t) \sim \log D \tag{9.24}$$

and hence we can conclude that:

$$\int_0^T dt \xi(t) = \sigma^2 \int_0^T dt \int_0^t dt' \chi(t,t') X_0(t') \tag{9.25}$$

The equation for the abundances now reads:

$$\begin{aligned} X_0^*(T) = \frac{D}{D-1} \Bigg(\mu \int_0^T X_0(s) \langle Y \rangle (s) ds + \\ + \int_0^T X_0(s) \xi_0(s) ds - \sigma^2 \int_0^T dt \int_0^t ds \chi(t,s) X_0(s) X_0(t) \Bigg) \end{aligned} \tag{9.26}$$

Up to now we can not solve the equation derived from the dynamical mean field in a self-consistent way; therefore we decide to proceed in studying numerically the properties and statistics of the species abundance distribution of the final survived species in the serial dilution process.

10 Analysis of convergence time scales

We stated before, as we can see in eq.(6.4), that a stationary steady state is reached when the initial abundances of two subsequent dilution cycles (j and $j+1$) $X_{i,j+1}(0)$ and $X_{i,j}(0)$ differ within a certain range of error ε. Even though this condition seems reasonable we want to check if, after the expected dilution step, there are still species that are going to extinction or others that keep growing. If that happened it would mean that:

- Even though the *approximate steady state condition* had been met, the system could have been far from reaching stationarity; that is in the step after, the situation could have changed because there would have been still species trying to grow or decrease

- The abundance threshold that we want to set in a simulation to delete at each dilution cycle those species that have been diluted too much could be wrong; we cannot be sure whether a species is really going to extinction if the rate $X_{i,j+1}(T)/X_{i,j}(T)$ is greater than zero so the threshold could introduce a bias in cutting the distribution of abundances in the wrong point.

We consider now two cases: the consumption matrix $a_{i\alpha}$ is sparse with a certain connectivity and the consumption matrix is fully filled.

10.1 Sparse matrix convergence time scales

For the purpose of seeing when the system is supposed to reach an approximate steady state we use a different algorithm choice. Instead of setting beforehand an abundance threshold below which species reach extinction and a condition regarding the dilution factor D we decide to plot the log rates of growth of species abundances with respect to the end of each dilution step.

First of all we check if the total species abundance at the end of each step is the same in order to see if the normalized initial conditions, discussed in section (8.3), have been respected by changing the variability of the consumption matrix, σ. In Figure 26 we see that total final species abundances are constant when changing the dilution step. With the aim of seeing what is changing from one step to another we can plot the relative abundances when moving from one dilution step to another. We expect that when relative abundances remain constant then the steady state is reached. We can compare the number of steps necessary to reach the approximate steady state following the condition in eq.(6.4) and what we see in the relative abundances dilution steps time series. As we saw in Figure 21, we need to have ≈ 200 steps on average over the

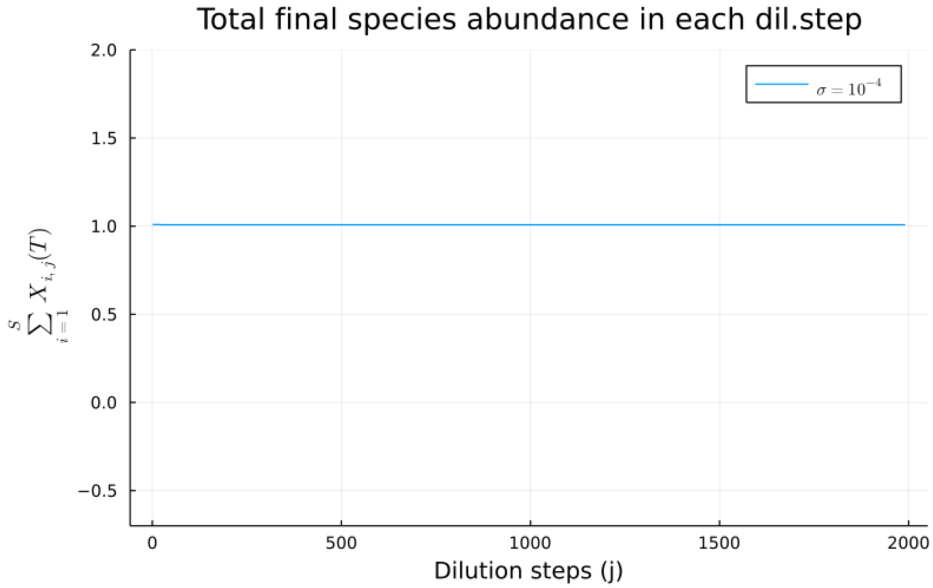


Figure 26: Total species abundances at the end of each dilution cycle versus the dilution steps. σ ranges in $[10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 0.25, 0.5]$. We are displaying only one case. The rest are shown in appendix.

realizations of the process to reach the approximate steady state. By looking at Figure 27 we see that that condition is still met within the same number of steps to have the relative abundances not to change from their stationary values independently on σ. On the other hand what happens to the difference between $X_{i,j+1}(T)$ and $X_{i,j}(T)$ among different steps. In order to see that, we decide to plot the logarithmic rates of species final abundances with respect to the final abundances of those species in each dilution step. Figure 28 is crucial since it allows us to recognize if species have really reached stationarity or not. It is better to take some lines to explain what is happening:

- In the first dilution steps there are some species that are still growing, that is there are points above the zero value. Species final abundances are almost at the same value since they have not been affected by competition and dilution yet. We see also more than a half species that have a negative log growth rate and it suggests that those species are going to extinction in the following steps
- After a few hundred dilution steps those species that were growing reached stationarity in the sense that they are no more changing their relative abundances, as we see in green, pink and red points in Figure 28. That group of species collapsed in zero while the others are moving to smaller relative abundances and also are

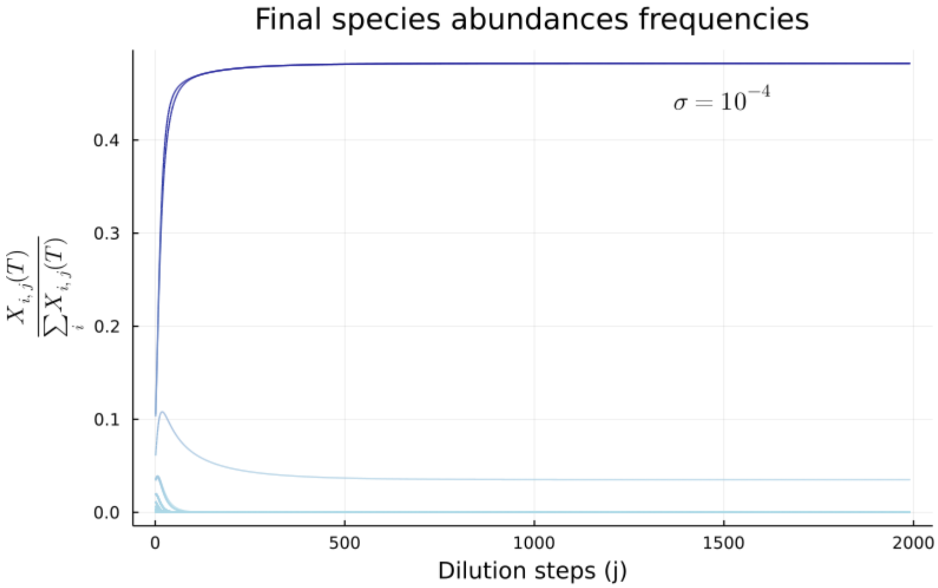


Figure 27: Relative final abundances of species with respect to the dilution steps. The other plots for different σs are shown in appendix since the time convergence time scale is not affected.

decreasing with a lower decrease rate.

- After two thousand dilution steps we see a gap between the group of species that collapsed in zero and those species that are decreasing. However since the latter group is moving to smaller relative abundances but also towards the zero growth rate it means that there could be some species whose decrease rate will slow down more in the subsequent steps until they reach zero.

Eventually we realize that, even if the group of species on the left is ending up with very small abundances, there might be some of them that could reach stationarity with more dilution steps than those that have been computed so far. For the purpose of understanding which of them are really going to extinction we can approach the problem in a smarter way than let the simulation running for more and more dilution steps.

10.1.1 Invasion analysis

Now we inspect the first species that has a negative growth rate. We choose it such that its abundance is second only to the last species with a zero growth rate. We

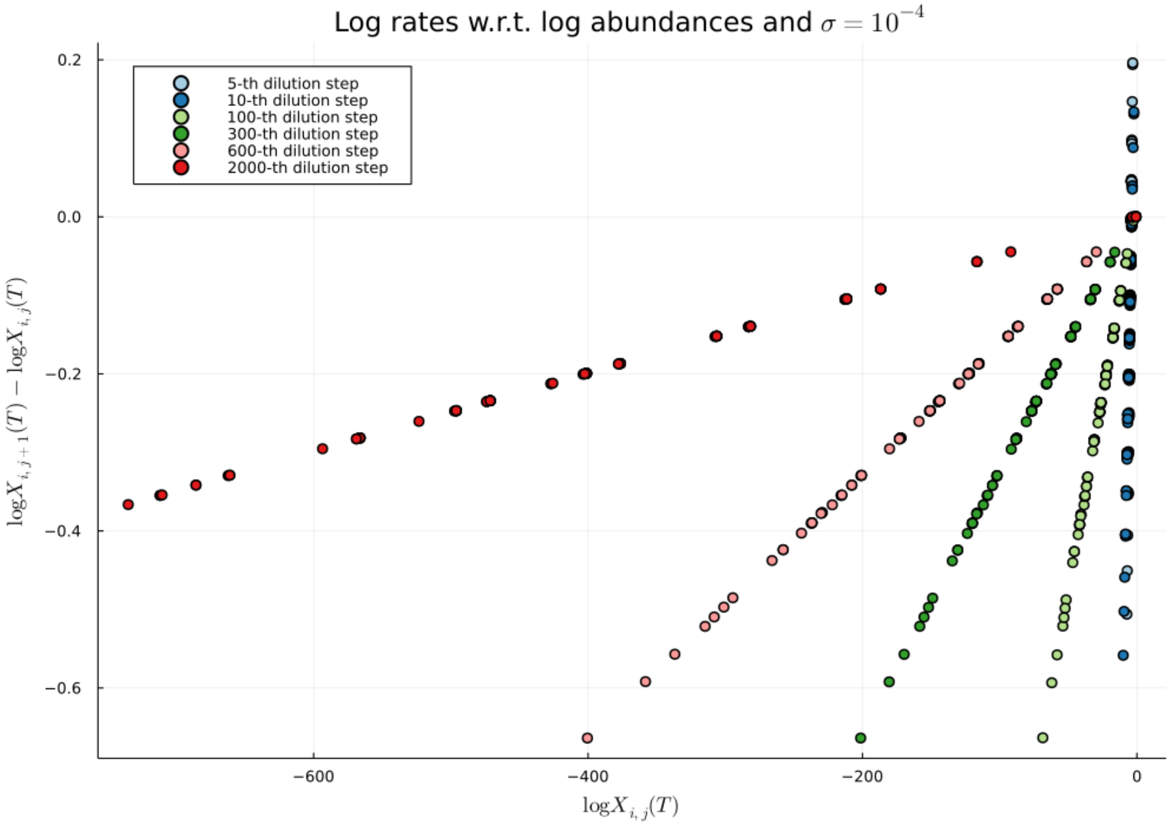


Figure 28: Log rates of species final abundances across subsequent dilution steps versus the log abundances of species at the j-th dilution step. For each dilution step (j) we see, in different colors, as many points as the number of species (one hundred in this simulation). In the case of sparse consumption matrix, when changing the value of σ in the range $[10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 0.25, 0.5]$ we don't see any noticeable differences. The rest of figures for other σs are shown in the appendix.

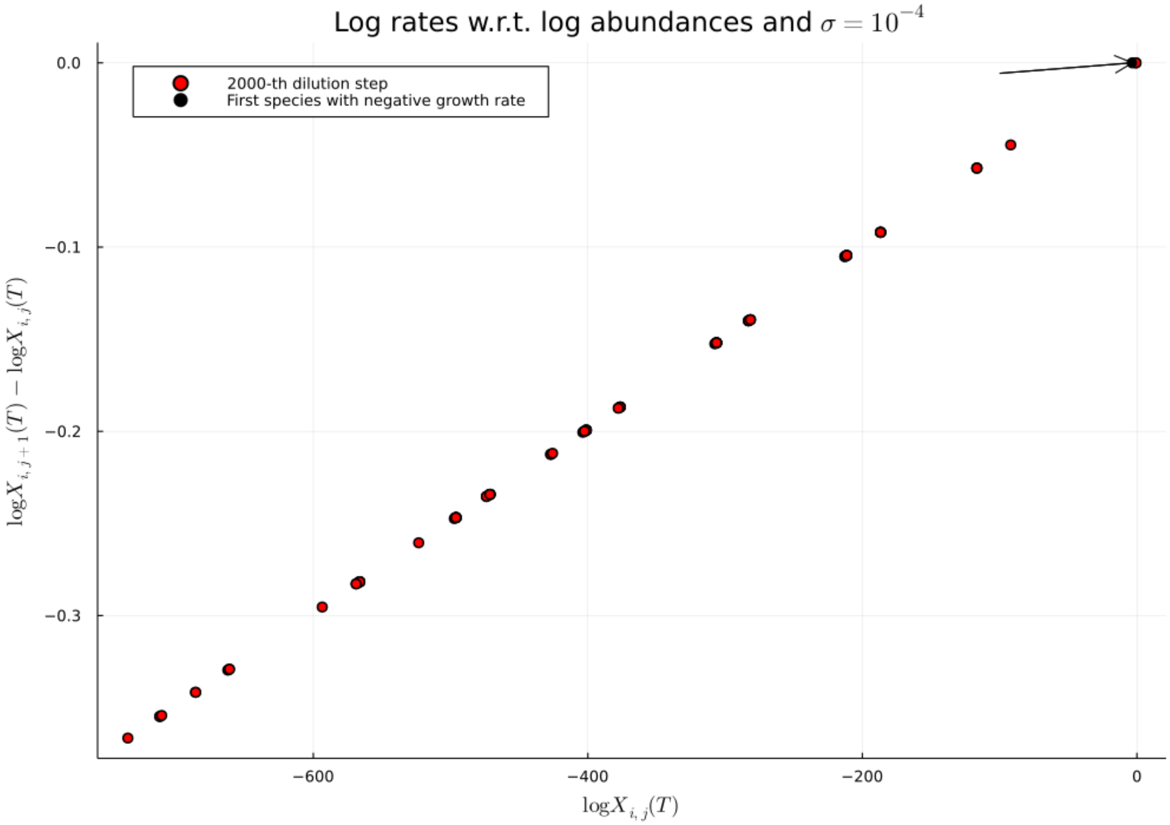


Figure 29: Log rates of species final abundances across subsequent dilution steps versus the log species abundances at j-th dilution step. The arrow is pointing to the species we are looking for in the first step of the invasion analysis. That is the one with negative decrease rate and second in abundance to the last one that has a growth rate equal to zero.

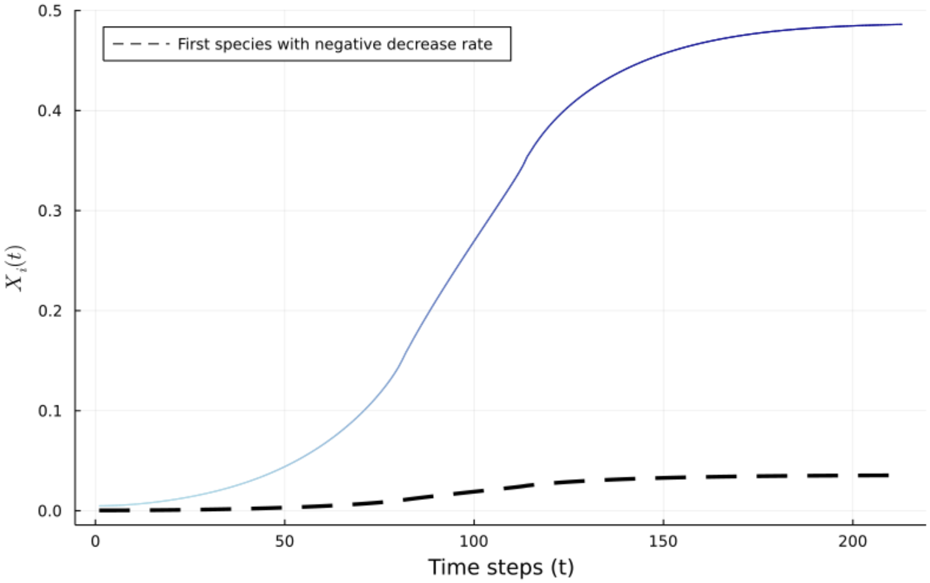


Figure 30: Species abundances versus the evolution time of the last dilution step. We show the growth of the species that we consider in the first step of the invasion analysis. The blue ones (two overlapping) represent those with zero or positive growth rate while the dashed black is the first in abundance that has a negative growth rate

ask ourselves now if the species pointed by the arrow in Figure 29 would be able to invade the system in the next dilution step. We can verify this statement by deleting all the species except the one that will try to invade. We can plot the correspondent abundances in the last dilution step. In Figure 30 every other species than those with zero growth rate and the first one with negative growth rate are deleted from the system and we are plotting the last dilution step of the simulation. Now we set the initial abundance of the invader species for the next dilution step to a small amount. We are only interested in its growth rate and not in its absolute final abundance; indeed by setting its initial abundance to a small value we are neglecting the effect of stealing resources from those species that are already at stationarity. After this step we look at the growth rate of the invader after the additional dilution step. If $X_{\text{inv},k+1}/X_{\text{inv},k} > D$ then we expect this species to slow down its decrease rate in the following steps and finally reach stationarity with zero growth rate after a while. We are not interested in this "while" in terms of dilution steps since we want to cut the simulation earlier than once the system reaches its *real* steady state since we don't have infinite amount of time for running the code. The same procedure can be repeated for all the other species to check whether all the system is going to survive or not. Furthermore, since experimentally, from A.Day 2004 we know that $\approx 10^9$ cells are found in coli cultures, we can set a threshold in the invasion analysis in terms of the log abundances of those species whose capability of invasion we are going to check to -9. [1]

10.1.2 First results on diversity w.r.t D

We start to explore the parameters space in terms of the diversity of the surviving species. We want to see what is the condition where we can find the largest number of survived species in order to record the final species abundance. We therefore apply the invasion analysis to a simple simulation with a 100 species and resources. We change the evolution time of each dilution step together with the dilution factor D for different values of variability of the consumption matrix σ; In Figure 31 we see that the more D increases, the greater the black region for small evolution times corresponding to a few number of survived species for high dilution factors. What's going on in these simulations? We can distinguish two regions:

- The first one is where no species, almost independently on the values of D and t, is surviving; that is the black region;
- The second region is on the right of the plot where increasing the dilution factor, after reaching a steady state, makes the number of species decrease;

In the first region we do not reach stationarity, that means that time span is small enough that species do not have time to grow and exceed the threshold, 10^{-9} absolute abundance. Furthermore, the larger the dilution factor is the slower species will evolve in order to reach the survive threshold, enlarging in terms of t the "extinguishing" region.

The second region displays a gradient only in terms of the value of the dilution factor since stationarity has been reached from forty to the upper limit of the evolution time. It means that if D is larger then the less species are going to survive since they are diluted more in each dilution step of the cycle.

10.2 Dense matrix convergence time scales

As we saw in section 10.1 we can gain useful insights about the convergence time scales of the system by plotting how long it takes for the species growth rates to reach zero among

---

[1] The experiments show that for E.coli the average number of cells per ml of solution are 10^8. Since there are in general more than twenty ml of solution there will be 10^9 cells in total. At most we can find one cell for one species we can set the abundance threshold of species to 10^{-9}

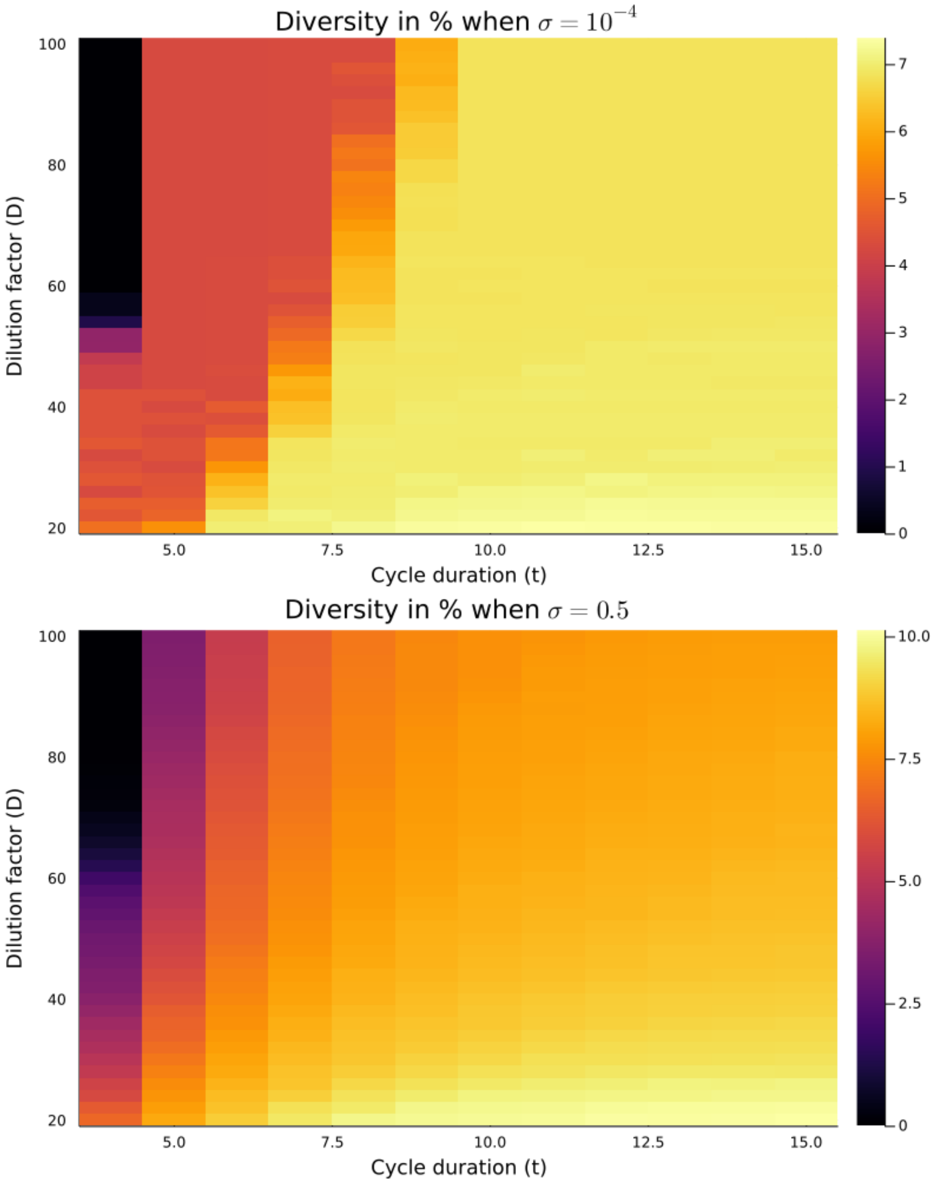


Figure 31: grid of D vs t inspecting the number of consumers and resources that are surviving $\langle \frac{S^*}{M} \rangle$ where the expected value is computed w.r.t the disorder of the interaction matrix $d_{i\alpha}$ realizations. We decided to represent only the upper and lower limit of the variability range of the consumption matrix. The rest are shown in appendix.

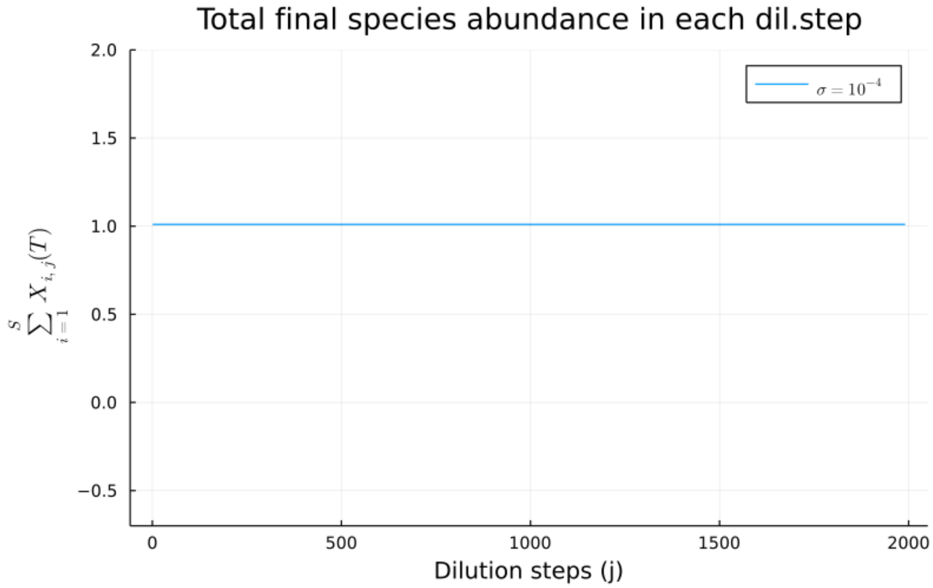


Figure 32: Total final species abundances are constant when changing the dilution step. σ ranges in $[10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 0.25, 0.5]$ but we are displaying only one case. The rest are shown in appendix.

different dilution steps. In this case we decide to consider the consumption matrix $a_{i\alpha}$ with sparsity zero for the purpose of investigating whether there are differences with respect to the previous setting. The total species abundance at the end of each dilution step is correctly conserved and stays at the same value changing also the variability of the consumption matrix, σ as we see in Figure 32.

In order to see whether the system needs more dilution steps to converge to a stationary state we plot the species abundances frequencies with respect to the dilution steps. As stated in section 10.1 we expect to reach the steady state when frequencies do not change anymore. By comparing the number of steps necessary to reach the approximate steady state following the condition in eq.(6.4) and what we see in the relative abundances dilution steps time series we end up with Figure 33 where we see that the case of non sparse consumption matrix differs a lot with respect to the previous one. The change of σ affects the speed of convergence of the system. It looks like that for small sigma species are behaving almost the same. It means that there is no distinct separation of niches and competition is much harder since each species is trying to approach the same resources to eat. For one hundred species and resources the number of dilution steps is not sufficient to allow frequencies to reach a stationary state, as they are still changing among different steps. On the other hand

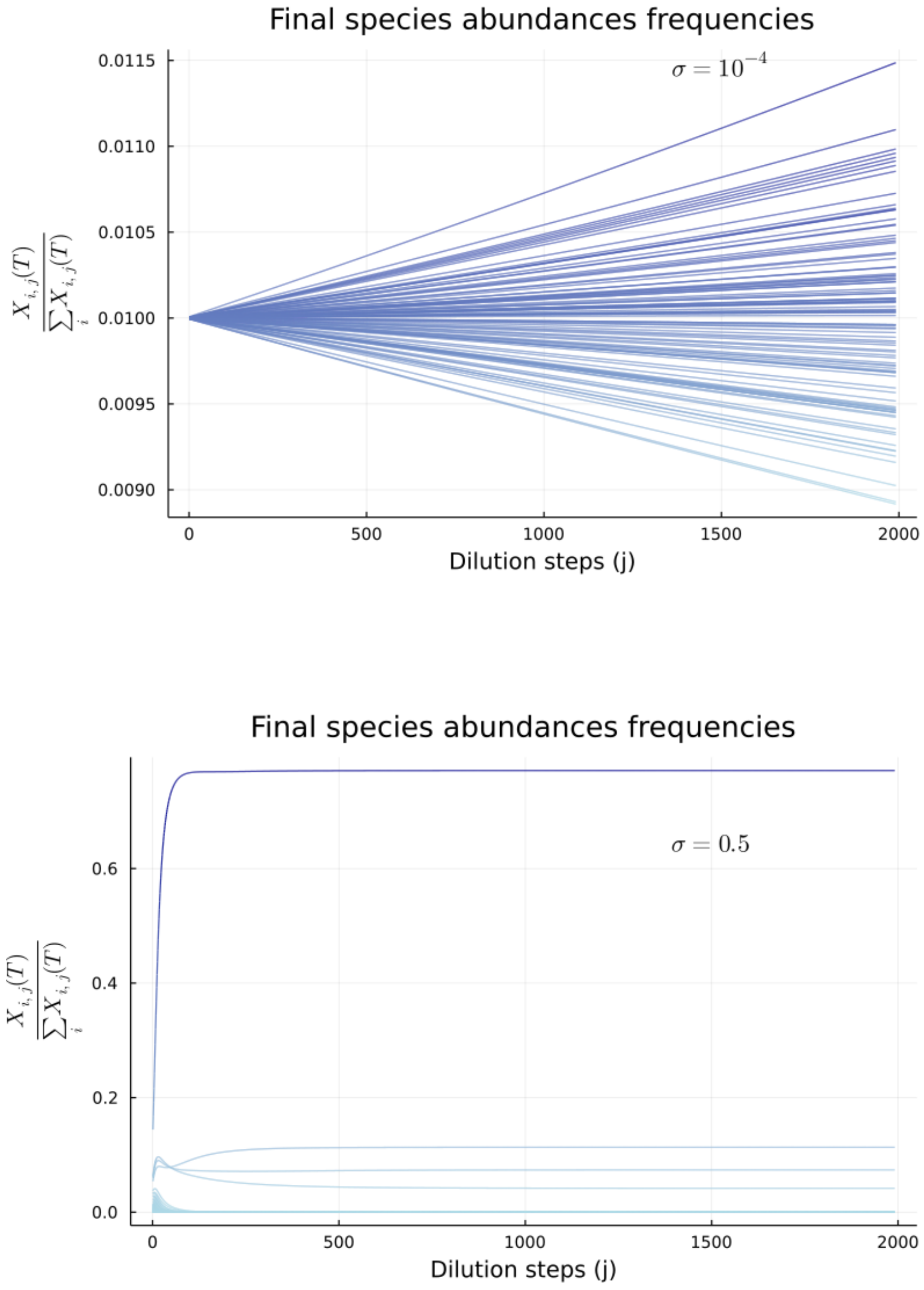


Figure 33: Species relative abundances versus the dilution steps with two different σs.

if we increase the variability it seems that we come back to the sparse consumption matrix case where niches are more separated and steady state is reached within the given number of dilution steps in terms of the change of abundance frequencies. The plots that relate $X_{i,j+1}(T)$ and $X_{i,j}(T)$ among different steps are even more surprising, indeed: As we can see in Figures 34 there are big differences with respect to the case of sparse consumption matrix. Independently on the value of σ there are still some species that are growing in the last dilution step, that is that one i choose beforehand to stop the simulation. Those that are decreasing have instead two different paths to follow:

- They can still decrease in a further dilution step but with less magnitude; it means that they could be able sooner or later to invade the system and belong to the group of surviving species; By using the invasion analysis we will see this situation happens very often that means that the number of dilution steps is not necessary to see a clear gap between those species whose growth rate is zero and those that are going to extinction;
- They can decrease with the same "extinction" rate or even more such that they are going to extinction. However this situation happens rarely and represents exactly that one that corresponds to the system reaching a stationary state after the decided number of dilution steps.

Similar results of section 10.1.2 have been found when plotting diversity with respect to the dilution factor.

Now we try therefore to apply the invasion analysis algorithm, together with the abundance threshold of 10^{-9} suggested empirically by experiments, to show the species abundance distributions and how diversity changes with different choices in the parameter space.

11 Species abundance distribution and statistics

From now on we decide to call $n_i^* = X_i$ the rescaled species abundances since we are interested in their probability distribution and their diversity at the end of the serial dilution process after n dilution steps. For the purpose of exploring the parameters space and see how properties of the serial dilution model vary with respect to them we can split the analysis in two cases: sparse and dense consumption matrices. We computed three different simulations taking into account various number of species and resources: $N = 50, M = 50$, $N = 100, M = 100$, $N = 200, M = 200$. In order to inspect the first results regarding how diversity and the shape of the species abundance

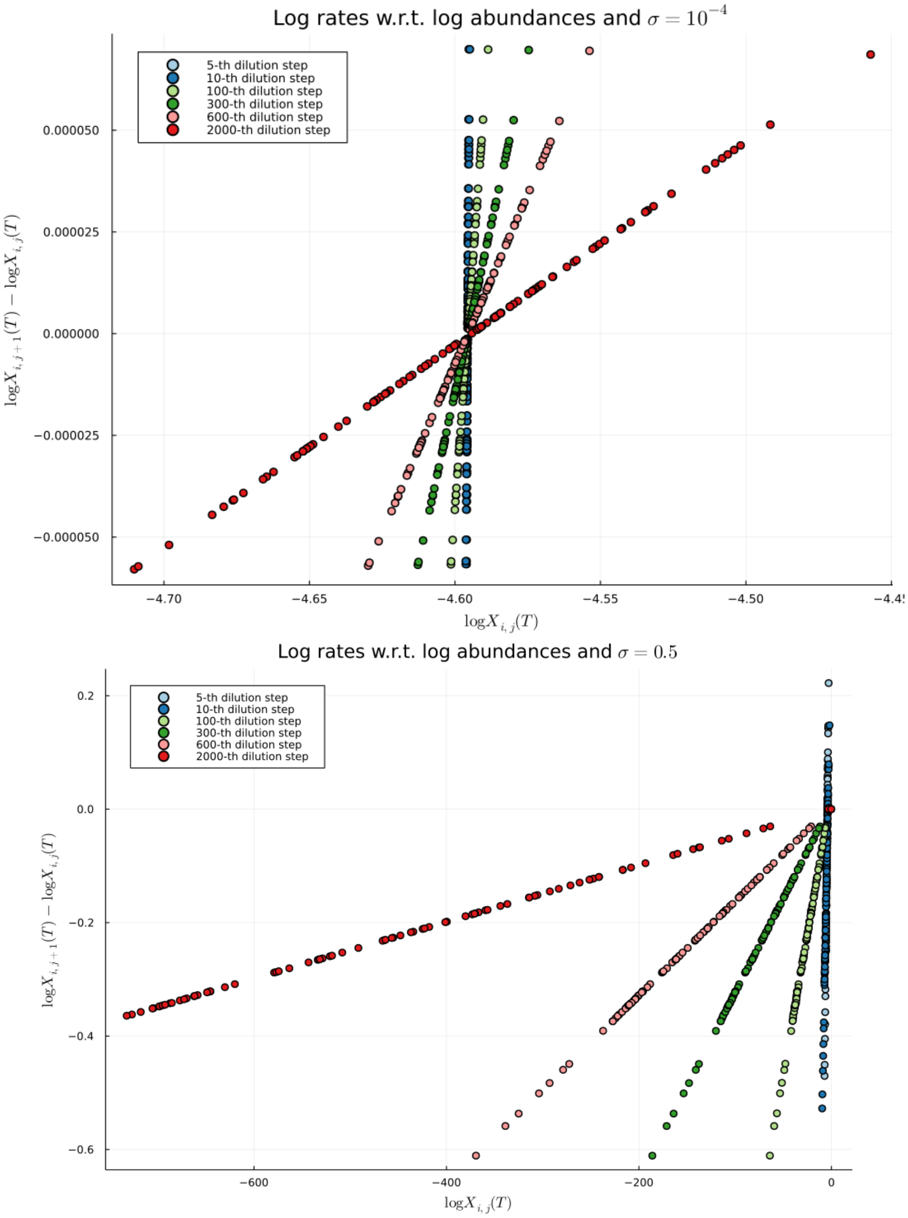


Figure 34: Growth rates for two different values of σ representing the smallest and biggest of the range of analysis. We see the number of species, for the last dilution step when $\sigma = 10^{-4}$, that have still positive growth rates is approximately a half of species. It means that the system is still to reach a steady state.

distribution we will consider only the case with a 100 species and 100 resources. All the simulations have been ran with the following procedure:

- Choice of the sparsity of the consumption matrix $a_{i\alpha}$
- Choice of one value for each the variability σ of $a_{i\alpha}$ and the dilution factor D and let the system evolve for a large number of dilution steps n ($N = 50, n = 100$;$N = 100, n = 200$;$N = 200, n = 300$ where $N = M$).
- Setting of the threshold for species in the end of each run to 10^{-9} as a first filter for species surviving
- Application of the invasion algorithm discussed in section 10.1.1 and save only those species that went through the test correctly
- Repeat the same run for different seeds of the random consumption matrix so that we would have had the possibility of averaging w.r.t. the noise for having more coherent points.
- Analysis of the output data

11.1 Sparse consumption matrix case

Diversity is strictly influenced by different factors like the sparsity of the consumption matrix, its variability and the dilution factor. We can see now what we can observe from simulations ranging in the above mentioned parameters space. In Figure 35 we see that when we introduce sparsity, even if the number of zero entries of the consumption matrix is only ten percent, we see that diversity is only influenced by the dilution factor. As we would expect if we increase the dilution factor then the number of species surviving at the end of the serial dilution procedure is less. Most of them, in fact, have been diluted and have reached an abundance less than the surviving threshold or they have not succeeded in going through the invasion algorithm. To see how the diversity is varying when changing the dilution factor we can choose as an example $\sigma = 0.25$ and plot Φ^* vs $\log D$ Figure 36 tells that the largest diversity $\Phi^* = 80\%$ is obtained when the dilution factor is $D = 1.05$. Then it drops to less than $\Phi^* = 50\%$ when $D = 2$ until the smallest value, $\Phi^* = 8.7\%$, when $D = 100$.

The next question now regards how the species abundance distribution shape changes by exploring the parameters space. First of all we see how the standard deviation of the SAD is affected by D and σ. The shape of the species abundance distributions is not affected neither by the variability σ or the dilution factor D. That is the curves we see

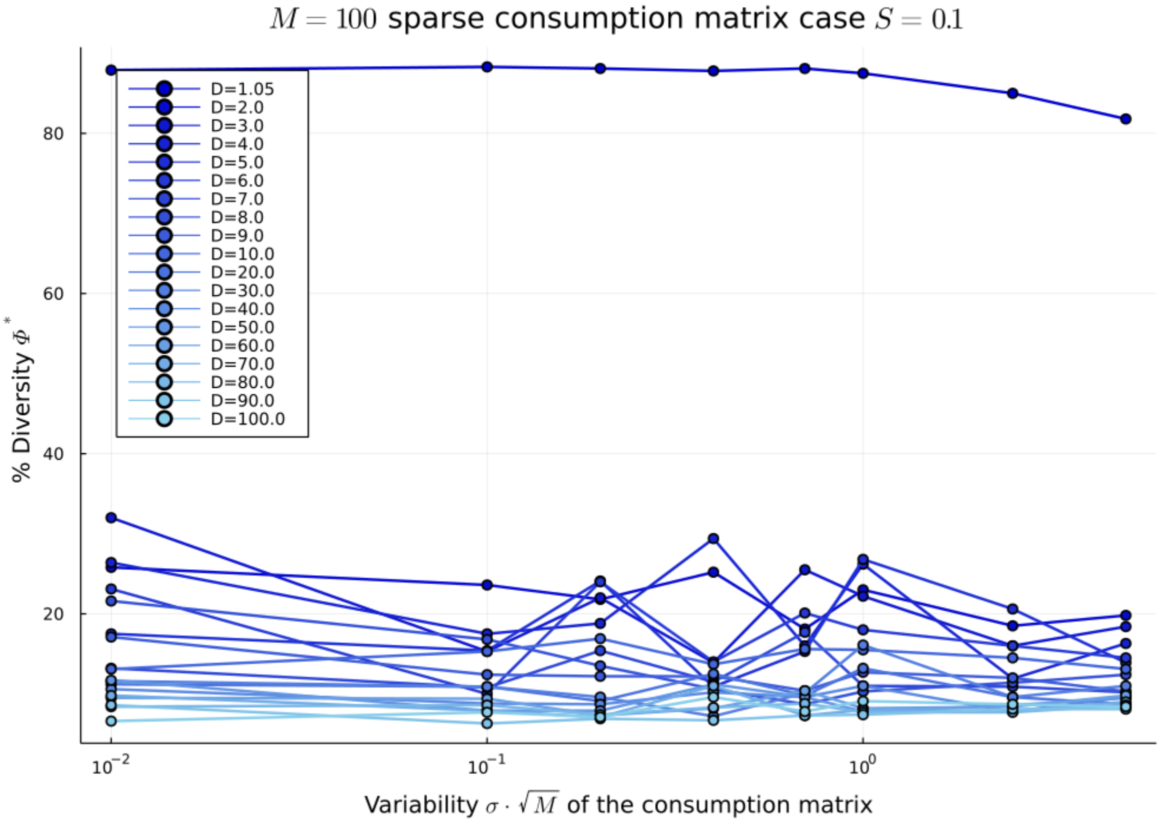


Figure 35: Percentage of diversity with respect to rescaled variability of the consumption matrix. $\sigma \in [0.001, 0.5]$ and $D \in [1.05, 100]$.

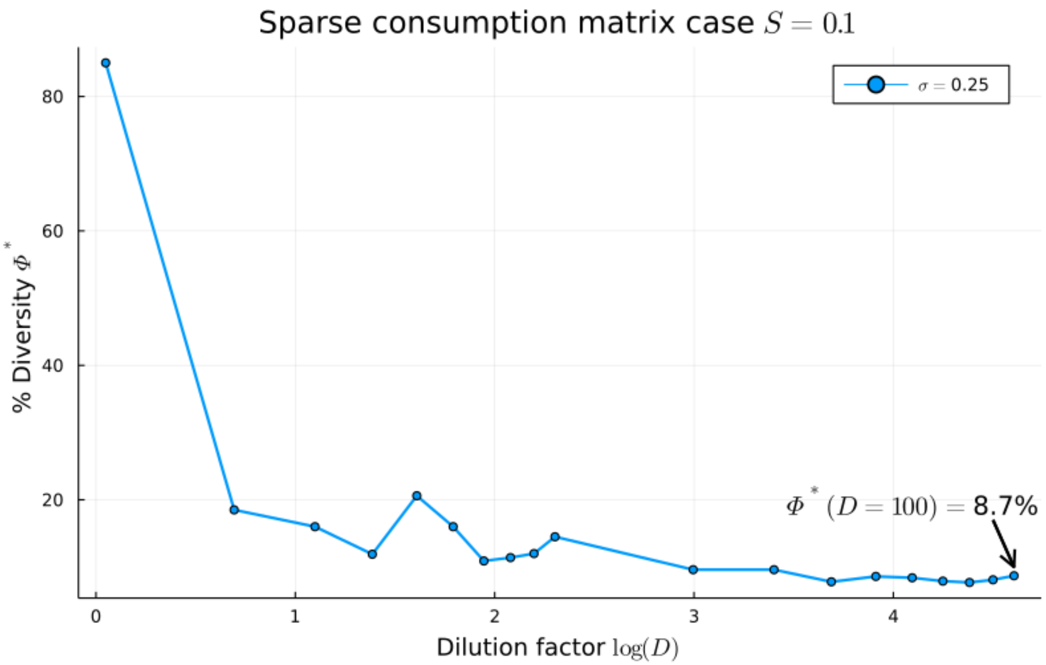


Figure 36: Percentage of diversity versus the natural log of the dilution factor $\log D$.

in Figure 37 are intersecting and do not follow a precise trend. Indeed we can represent two SADs with one value of σ and two different D (the smallest and the largest). Both cases in Figure 38 present a tailed distribution whose width is not dependent on the dilution factor in any distinctive way.

11.2 Dense consumption matrix case

As seen in the sparse $a_{i\alpha}$ case we first examine the diversity of the surviving species. As we expected from section 10.1.1, we notice that curves of diversity in Figure 39 are trying to reach a steady state but are not there yet. In fact we clearly see two trends in the figure. One is for the first curve to approach a large value of diversity, the second is the rest of dilution factors decreasing diversity when increasing σ. Unlike the sparse matrix case where the two trends were already set such that diversity did not change with respect to variability here we realize that there is a strong dependence. Furthermore in the sparse matrix case we notice that by adding just a small value of sparsity we are also adding variability into the system. It means that the "effective" variability is larger than what we are imposing to the consumption matrix σ and this makes the process to converge sooner. We can choose therefore a value of sigma where

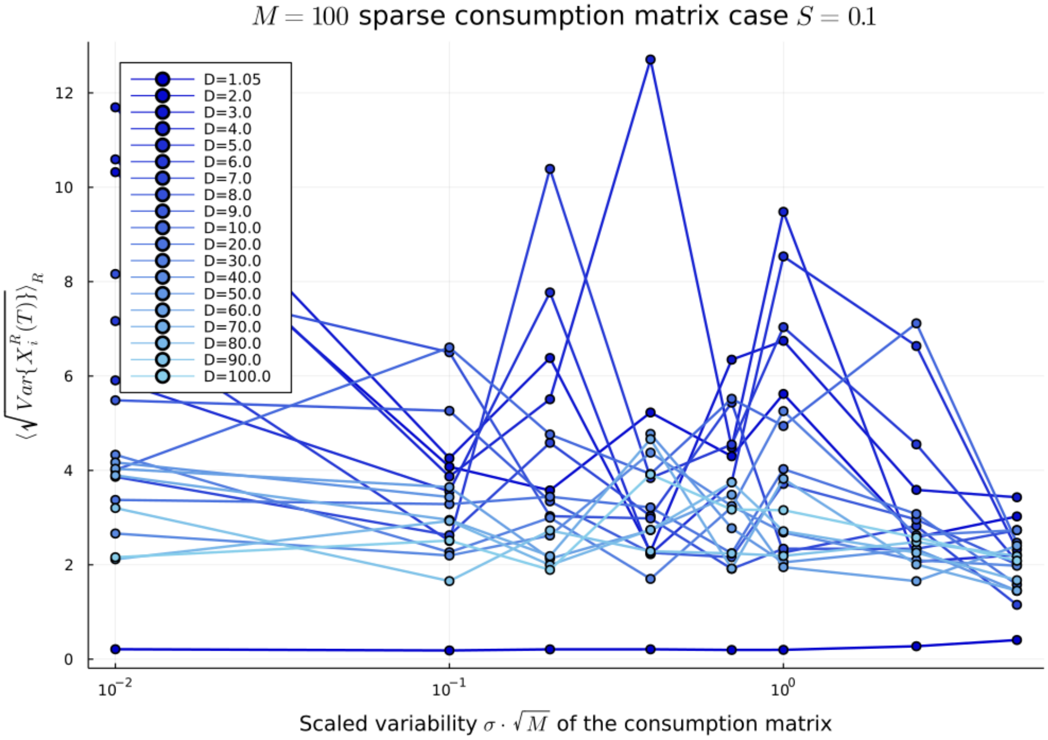


Figure 37: Standard deviation of the rescaled abundances w.r.t the dilution factor D and the variability σ of the consumption matrix. The latter statistical indicator has been mediated across different realizations R of the same simulation for each point in the plot. X axis is in base ten logarithmic scale.

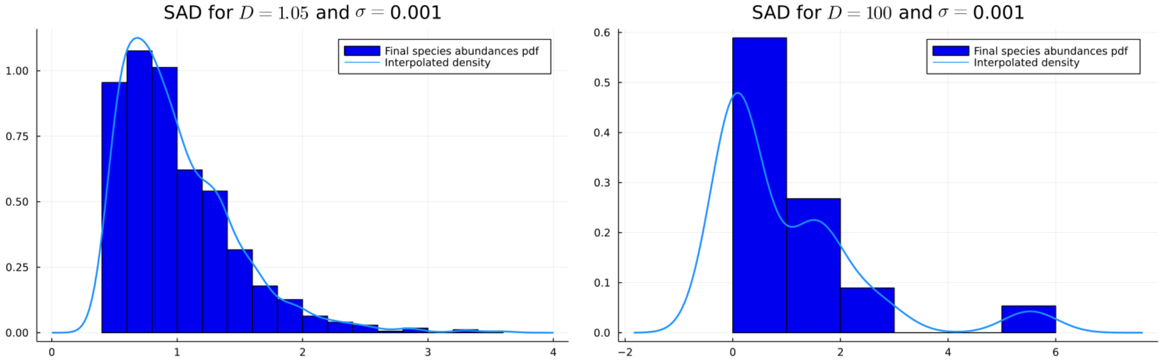


Figure 38: SADs when the consumption matrix is sparse. The dilution factor is the smallest, $D = 1.05$ and the largest, $D = 100$ with the same value of $sigma$.

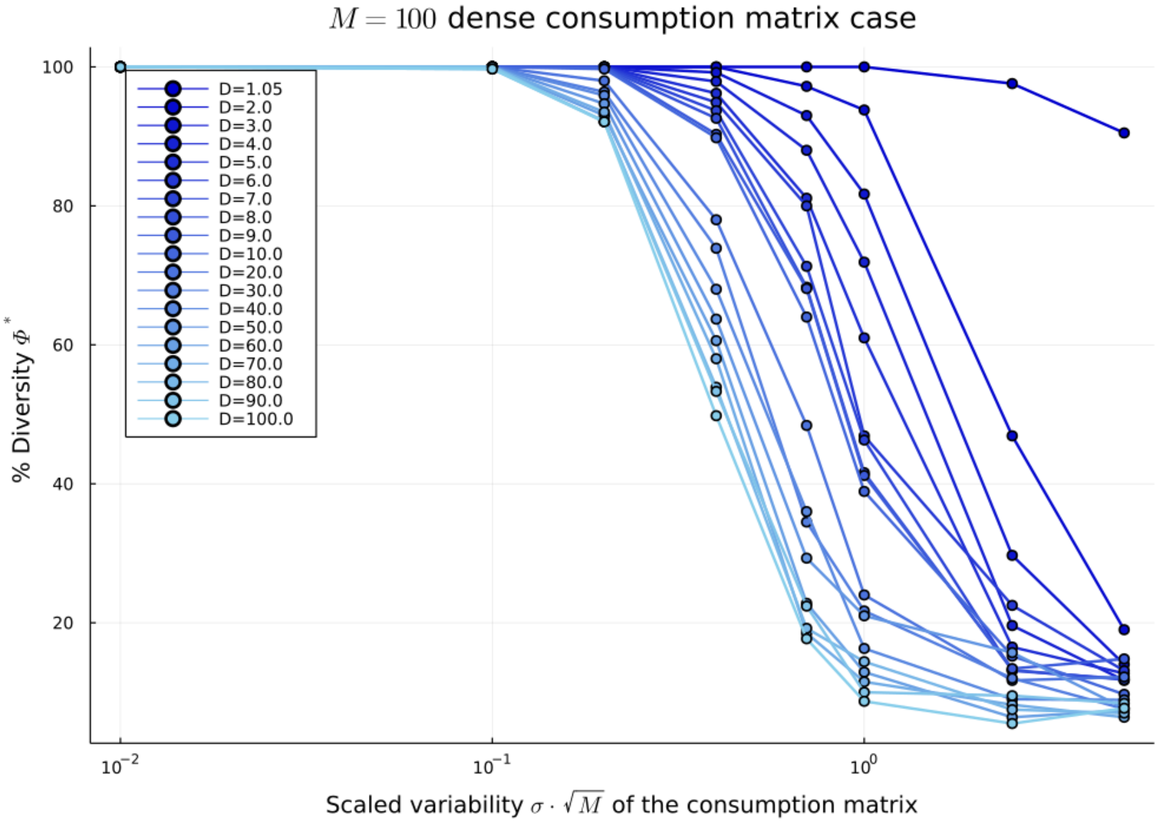


Figure 39: Percentage of diversity with respect to rescaled variability of the consumption matrix. $\sigma \in [0.001, 0.5]$ and $D \in [1.05, 100]$.

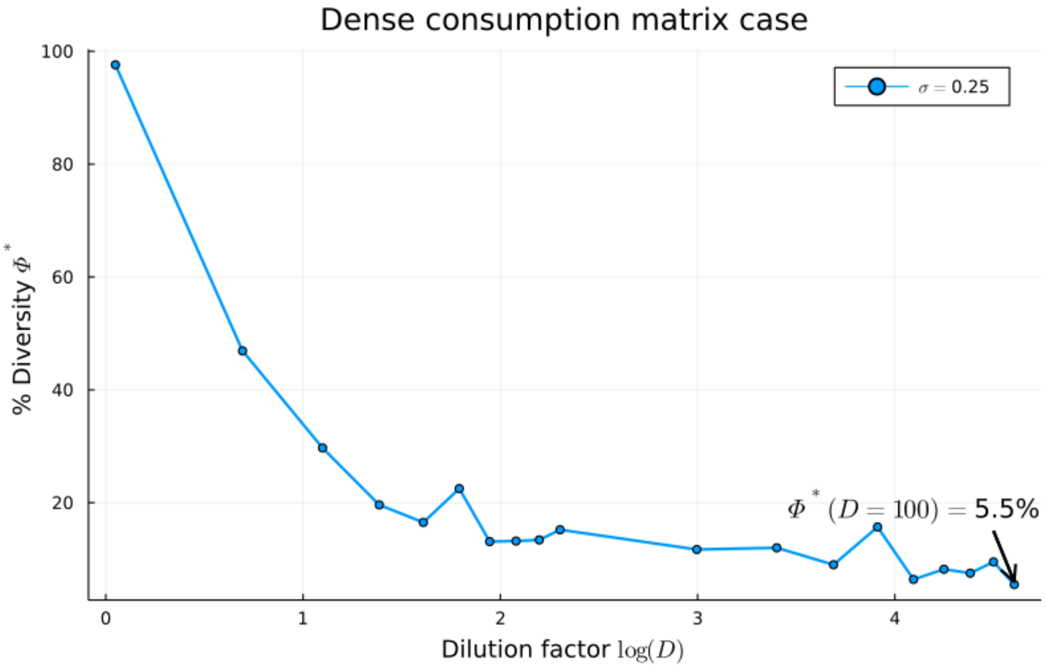


Figure 40: Percentage of diversity versus the natural log of the dilution factor $\log D$.

the separation of curves is happening and plot the percentage of diversity versus $\log D$. The largest diversity $\Phi^* = 100\%$ is obtained when the dilution factor is $D = 1.05$. Then it drops to less than $\Phi^* = 50\%$ when $D = 2$ until the smallest value, $\Phi^* = 5.5\%$, when $D = 100$.

Now we expect that this transition across two "phases" in the diversity plot must be reflected in the statistical indicators of the shape of final abundances distributions. Indeed Figure 41 shows that the "steady state" is reached when the transition across the two "phases" happens and curves are separated into two groups as before. Furthermore now we look at the species abundances distributions for various σ and D. In Figure 42, when σ is small, we see that independently on D, either the largest or the smallest makes no difference, the SADs are mostly Gaussian distributed with a small variance as we saw in Figure 41. When σ increases of an order of magnitude we have two different SADs, respectively still a Gaussian and a long tailed distribution. It means that SADs are becoming wider and therefore a tail appears where species have largest final abundances. In this case by changing D we are moving across the two above mentioned "phases".

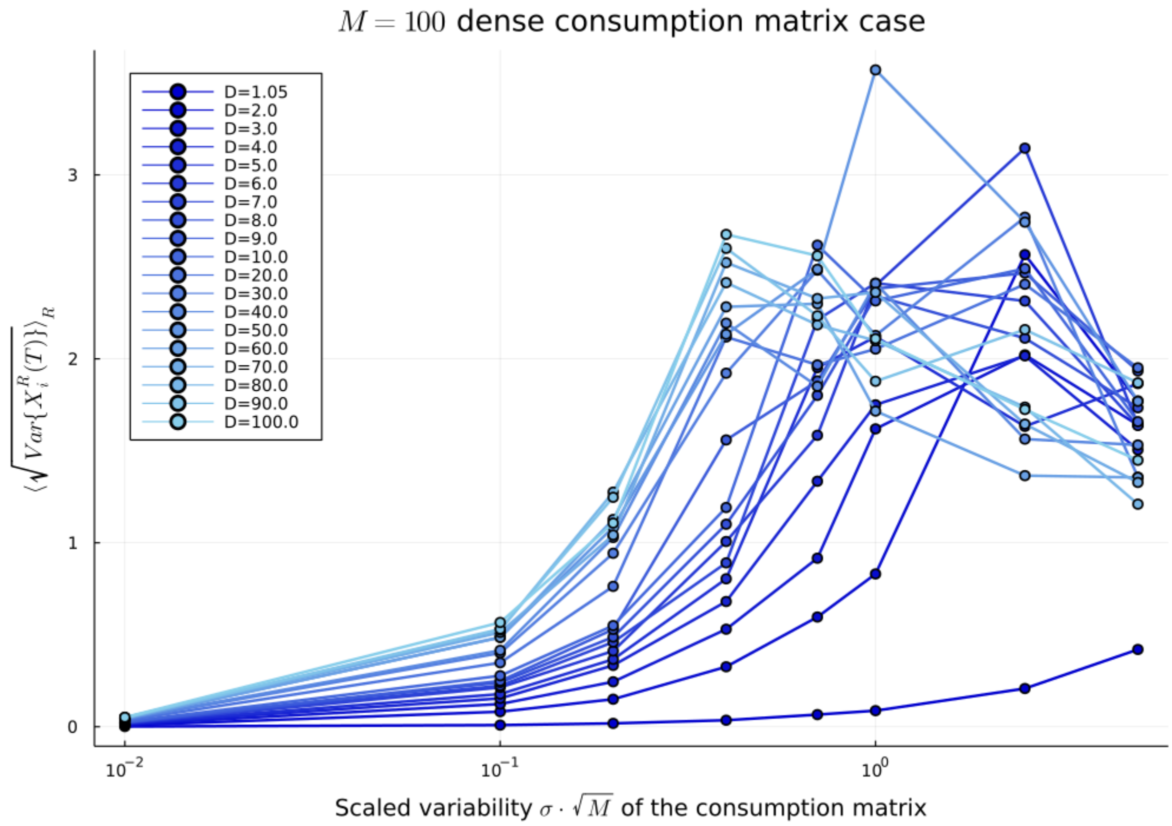


Figure 41: Standard deviation of the rescaled abundances w.r.t the dilution factor D and the variability σ of the consumption matrix. The latter statistical indicator has been mediated across different realizations R of the same simulation for each point in the plot. X axis is in base ten logarithmic scale.

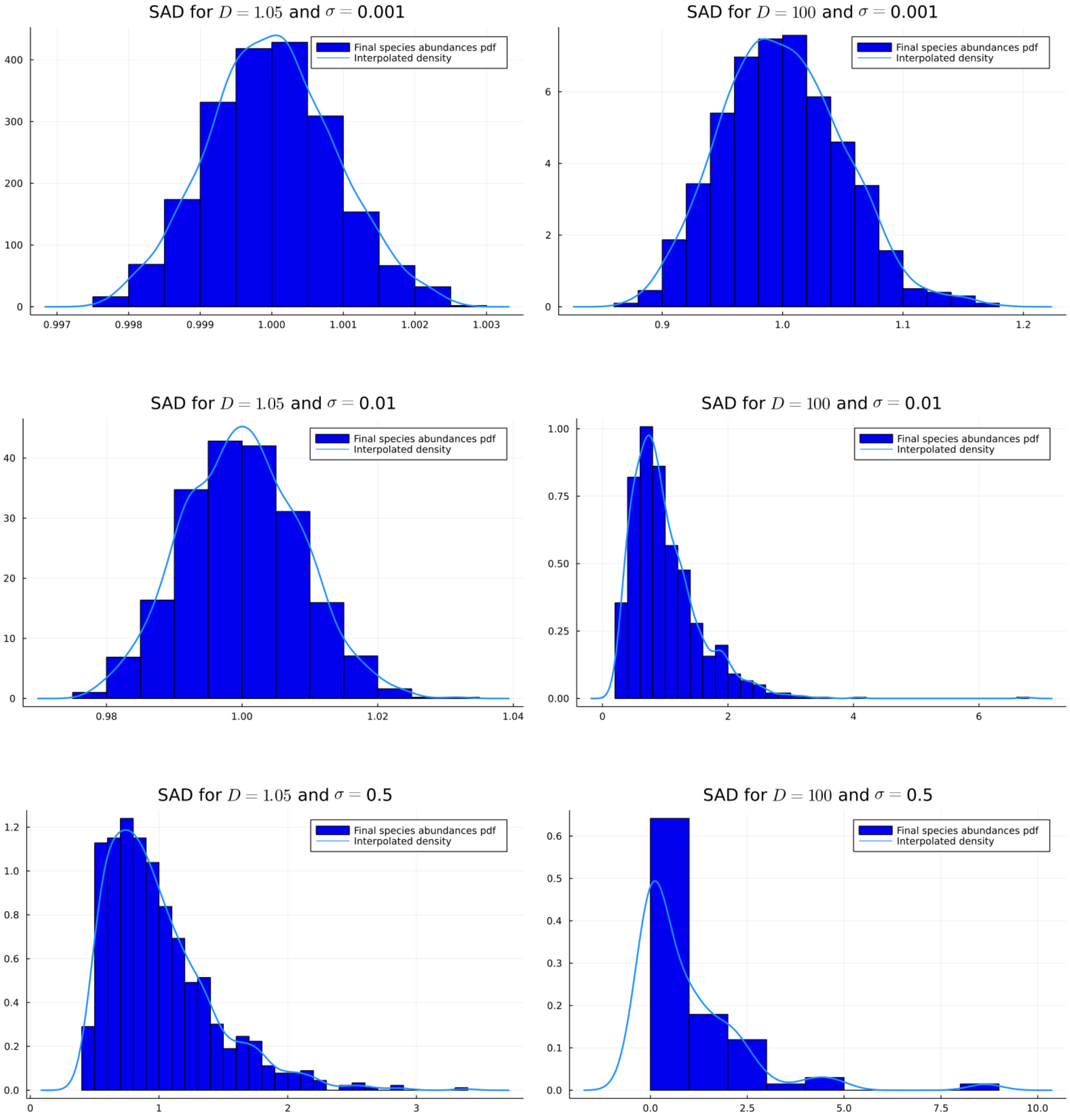


Figure 42: SADs when the consumption matrix is dense. Dilution factors are the smallest, $D = 1.05$, and the biggest, $D = 100$. σs are the smallest, $\sigma = 0.001$, and the largest, $\sigma = 0.5$.

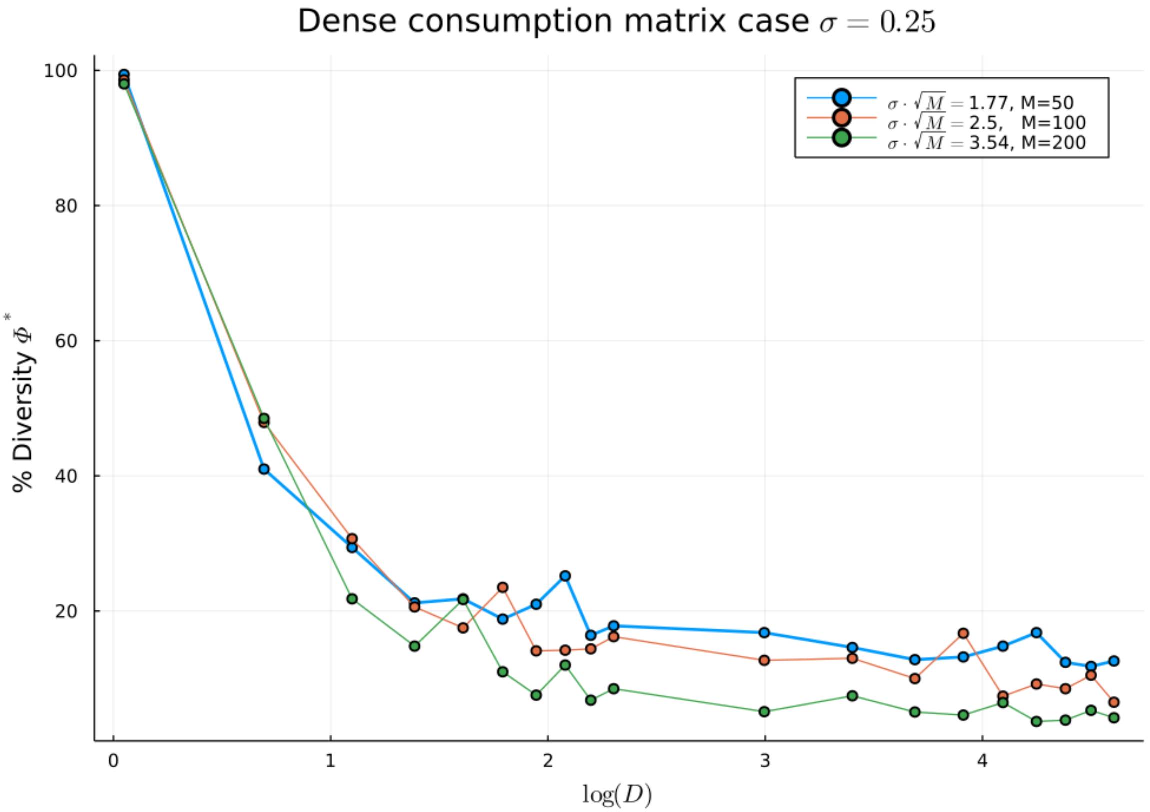


Figure 43: Percentage of diversity when $\sigma = 0.25$ and the number of species and resources changes vs $\log D$.

11.3 Dense matrix case deeper analysis

The case when the consumption matrix is dense shows some peculiarities as we have already noticed in sections 10.2 and 11.2. A deeper analysis of this situation will consist in studying how diversity and width of SAD scale with the number of species and resources and trying to interpret the already found relations by starting from plots.

11.3.1 Scaling with the number of species and resources

We can plot both the diversity and the width of SADs when changing the number of species and resources and see whether curves are going to scale in some particular pattern or not. Since the values of σ as variance of the consumption matrix are not scaling explicitly with respect to the number of species and resources we decide to

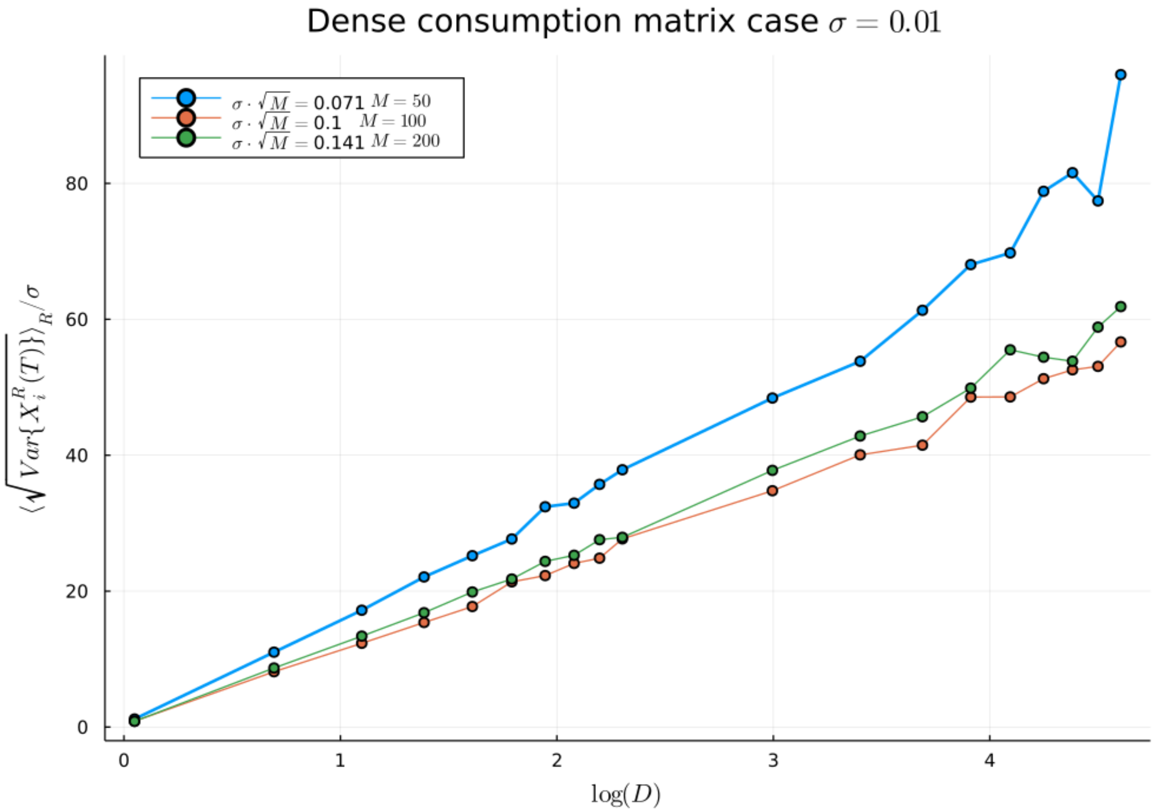
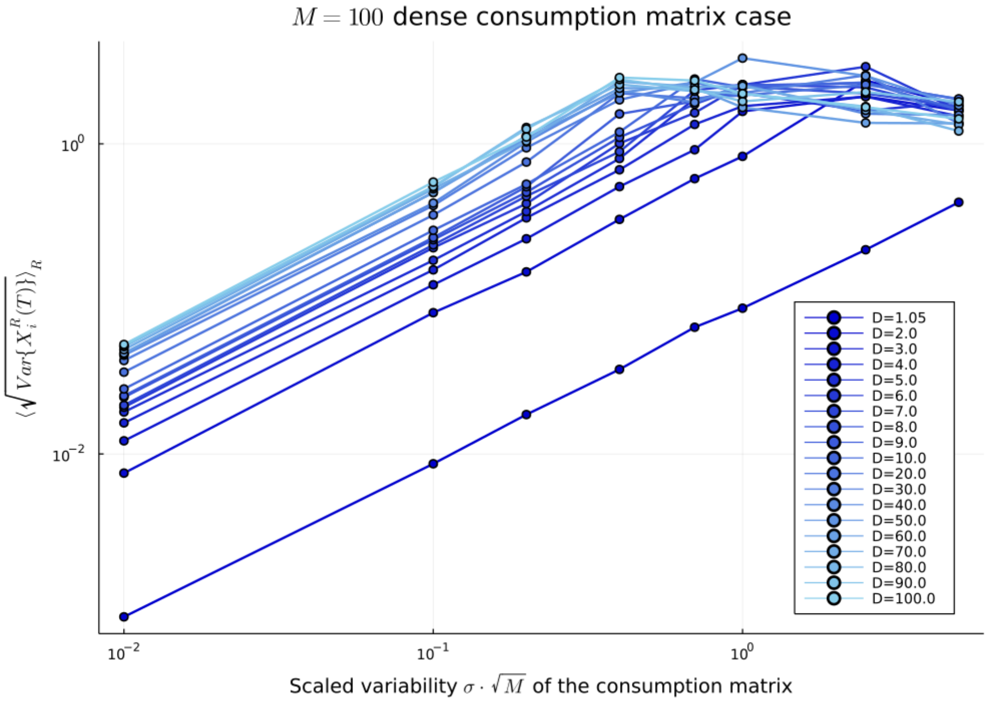


Figure 44: Scaled standard deviation of the species abundance distributions versus $\log D$.

represent diversity for the same values of σ but different M. That means that in Figure 43 different curves correspond to see how diversity scales with M. Here we see that by increasing the number of species diversity tends to decrease. Diversity lines are therefore lowering with the number of species increasing. The widths curves of SADs in Figure 44, on the other hand, are converging to a certain slope when changing the number of species.

11.3.2 Analytical progresses

By looking at Figure 41 we notice that there could be a functional relation between the standard deviation of the final species abundances mediated on the realizations of the process $\langle\sqrt{Var\{X_i^R\}}\rangle_R$ and the rescaled variability of the consumption matrix $\sigma\sqrt{M}$. Indeed we want now to consider the log-log plot. Up to a certain σ, when the previously



mentioned "transition" happens, the two quantities are linearly dependent. In order to better see that we can try to plot $\langle\sqrt{Var\{X_i^R\}}\rangle_R/\sigma\sqrt{M}$ vs $\sigma\sqrt{M}$. Figure 45 shows that $\langle\sqrt{Var\{X_i^R\}}\rangle_R/\sigma\sqrt{M}$ is independent of σ. What we find is that the wideness of the species abundance distributions increases when the number of species and the variability of the consumpption matrix do. Now we want to make the statement general for other realizations of the process with different dilution factors D. We want to find a way such that we can include the dependence on $\log D$ into the spread change of the species abundance distributions. One possible solution is to represent the diversity predictions versus $\sigma\sqrt{M}\log D$ instead of just versus $\sigma\sqrt{M}$. Surprisingly in Figure 46 we find out that different D curves collapse into a single one except for the smallest dilution $D = 1.05$. It means that the width dependence on $\sigma\sqrt{M}$ is also linear with respect to $\log D$. Furthermore it is possible to generalize an analytical form of the SAD for the parameters space we have explored so far. Figure 47 shows also that if we consider the three different simulations with $M = 50, 100, 200$ then the curves regarding diversity almost collapse on the same one, except again for $D = 1.05$. We realize also that the dependence on the number of species is irrelevant in writing the prediction for the diversity. The width and skewness of the SADs depend on $\sigma, \log D$; if we change M instead we expect to converge to a certain standard deviation as we saw in Figure

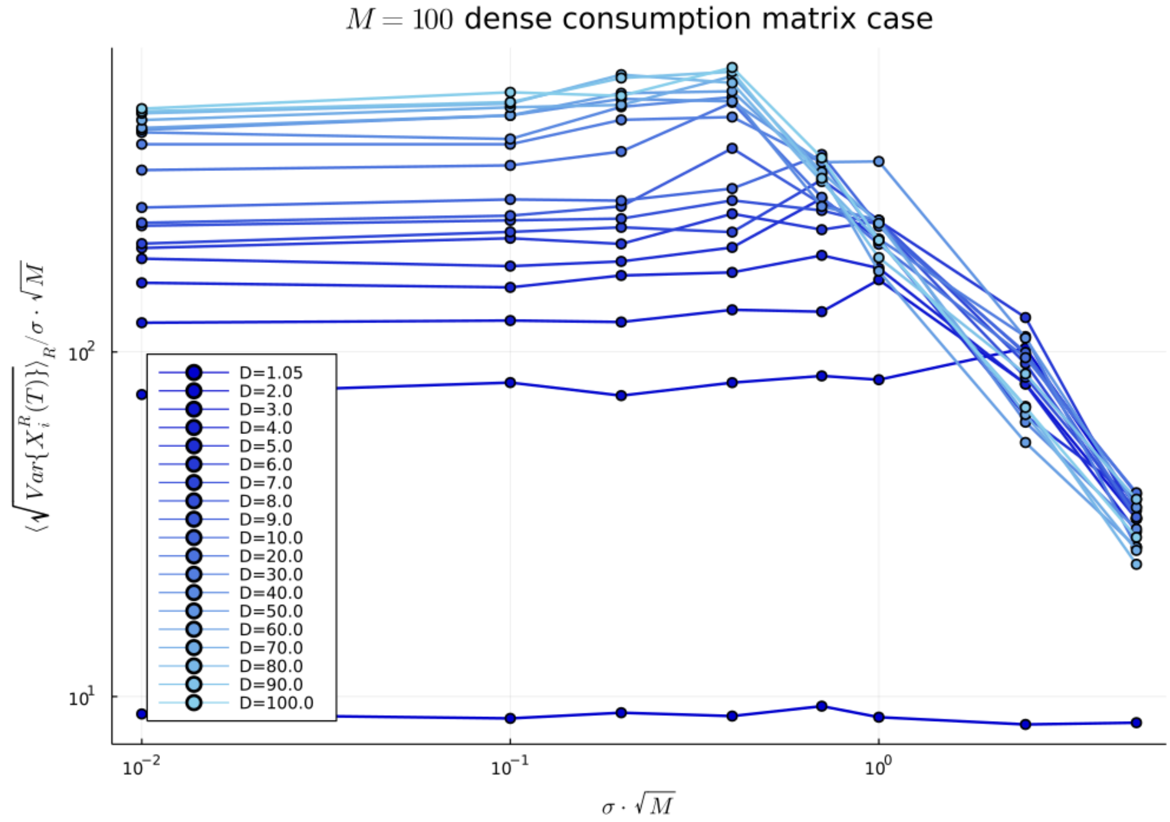


Figure 45: Rescaled standard deviation of SADs versus the rescaled variability of the consumption matrix. Both axis are in logarithmic scale and curves represent different dilution factors D.

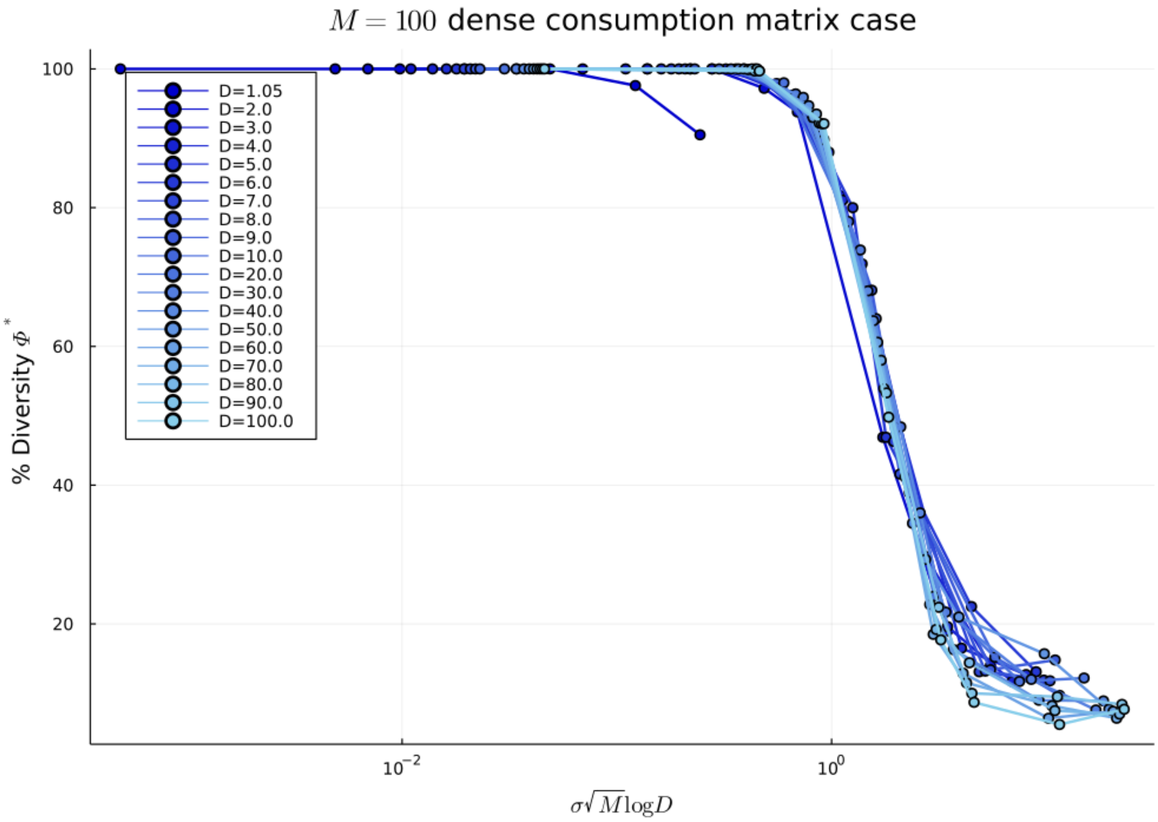


Figure 46: Percentage of diversity versus $\sigma\sqrt{M}\log D$. Since different D curves are represented we see their ending points are different.

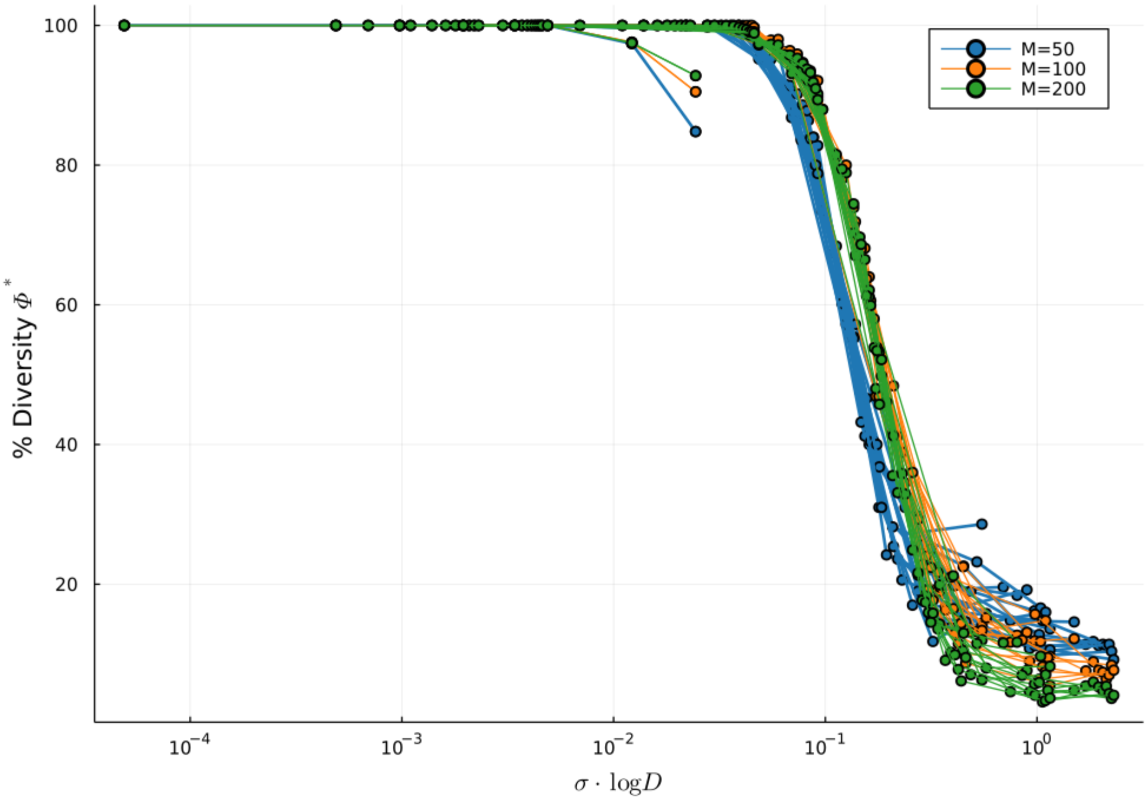


Figure 47: Percentage of diversity versus $\sigma \log D$. Three colors represent different number of species while in the same color there are all the simulated dilution factors D.

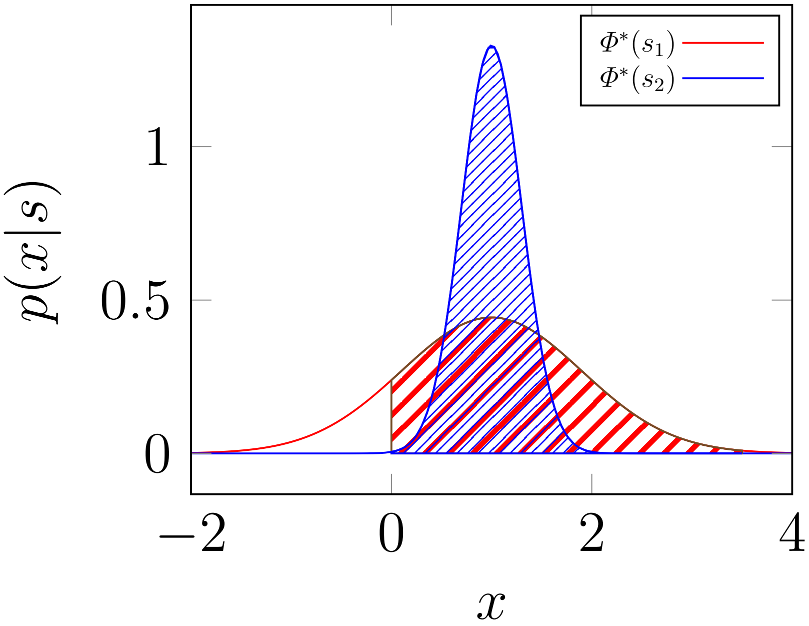


Figure 48: There are two plausible shapes of the final species abundance distribution where $s_2 > s_1$. The areas under the curves represent the diversity predictions $\Phi^*(s_1)$ and $\Phi^*(s_2)$. The blue curve shows that all species are surviving while in the red one only a fraction of them succeed.

44. Based on the previous results we know that the standard deviation is proportional to $s = \sigma \log D$ and therefore the full dependency on σ, and D of the SAD can be written as a dependency on s. There will exist a value of s such that the distribution has negative entries and, inspired from G. P.Mehta M. 2018, we can determine when not all the species survive and some of them are going to extinction. An example is shown in Figure 48 Therefore a prediction for the number of species surviving in the end of the dilution process would be described by the integral of the species abundance distribution from zero to infinity.

$$\Phi^*(s) = \int_0^\infty dx \; p(x|s) \tag{11.1}$$

12 Conclusion and further developments

The main results of this work aim to analyze the differences between the chemostat model and the serial dilution process. At first we found that the chemostat setting is not well defined to describe natural environments like the gut microbiota. Therefore we chose a simple serial dilution model with stochastic replenishing of resources. By looking at one species and one resource we verified some properties of the time series statistics like the Taylor law. Furthermore by changing the integration time in each dilution step we found out that the shape of the final abundances distribution at each dilution step moves from a Gaussian to a log-normal. However in order to dig deeper into the properties of the serial dilution process we chose a consumer resource approach with disordered consumption matrix, multiple species and constant replenishing resources. We discovered that the convergence time scales are highly influenced by the variability of the consumption matrix and its sparsity. Indeed when we introduce even a small sparsity to the system the latter converges much faster than the case when the consumption matrix is fully connected. The difference in convergence time scales has been numerically approached by introducing a novel algorithm that predicts, through an invasion analysis, how many species are going to extinction and how many are surviving in the end of the simulations. Furthermore the difference in convergence time scales are reflected in different properties of both diversity and the species final abundances distributions. Diversity tends to be on average higher when sparsity is added into the consumption network and SADs are strictly dependent on the variability of the $a_{i\alpha}$; in particular a transition in terms of the variance of these distributions takes place when we consider a dense matrix and increase its variability while no transition happens when the consumption matrix is sparse. That is because a steady state has been already reached in the sparse case while we see the approach to that state in the dense matrix case. After this work one possible approach to the problem could be trying to solve it analytically through dynamical mean field calculations by exploiting some properties of the system that are suggested by data. In this way it would be interesting to see whether simulations predictions are coherent with analytical prescriptions.

13 Appendix

We will show some additional figures regarding simulations of the models with different parameters choices:

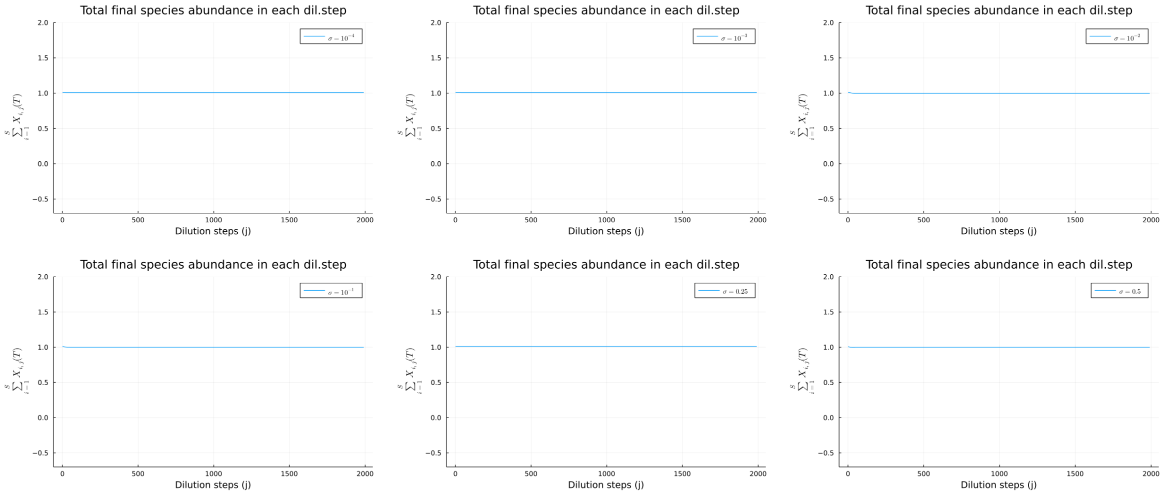


Figure 49: Sparse matrix case $S = 0.1$.

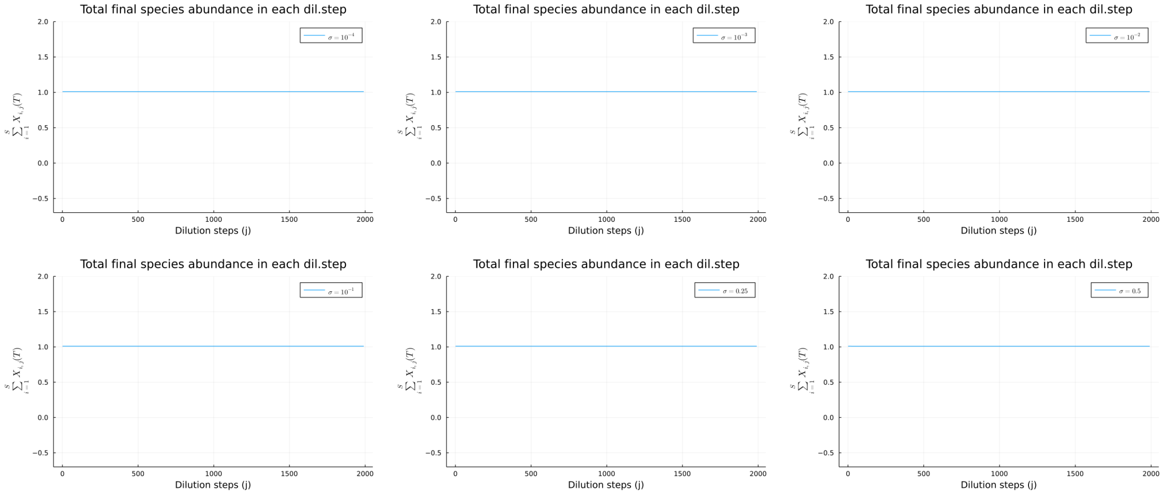


Figure 50: Dense matrix case.

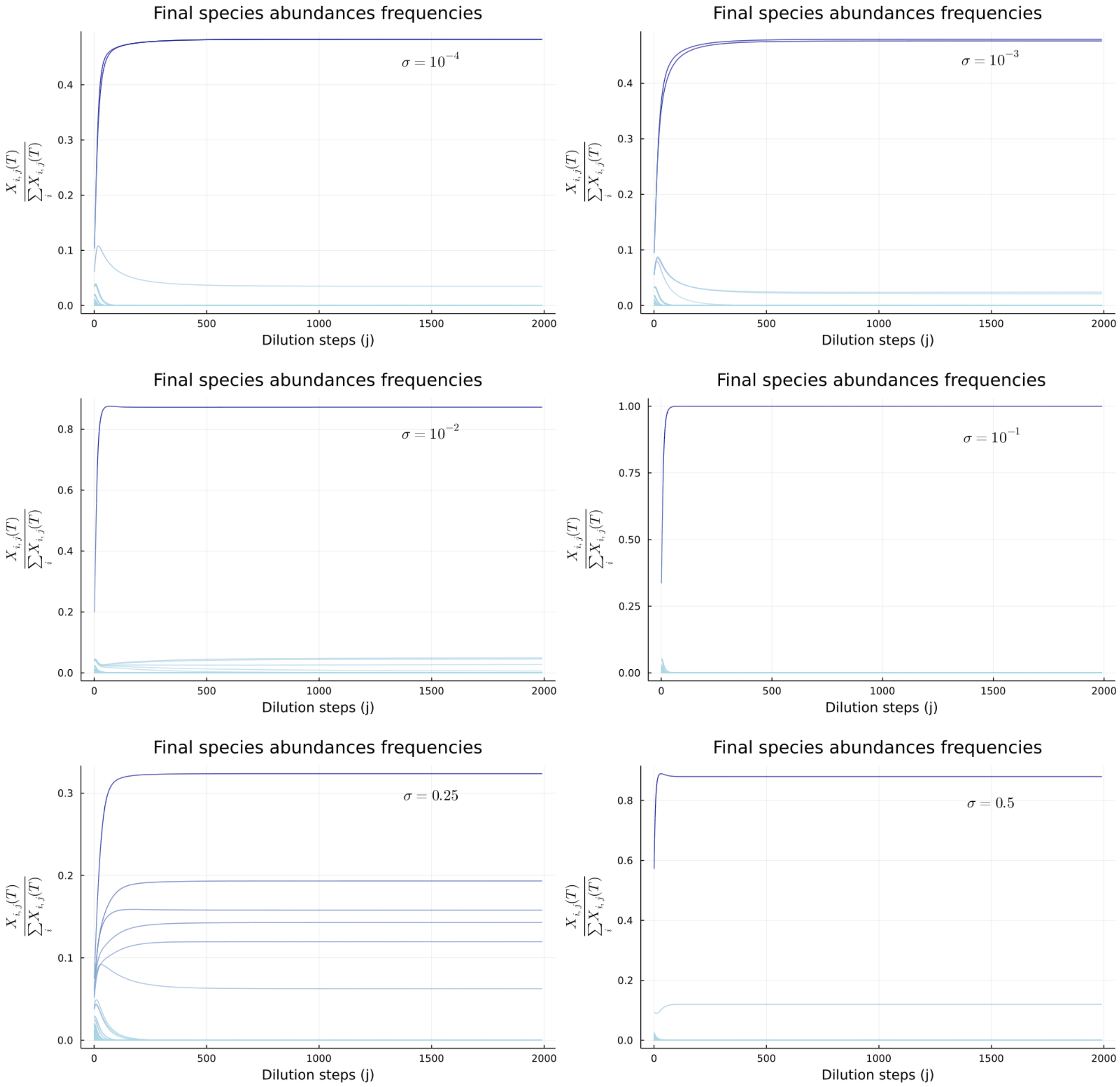


Figure 51: Sparse matrix case $S = 0.1$.

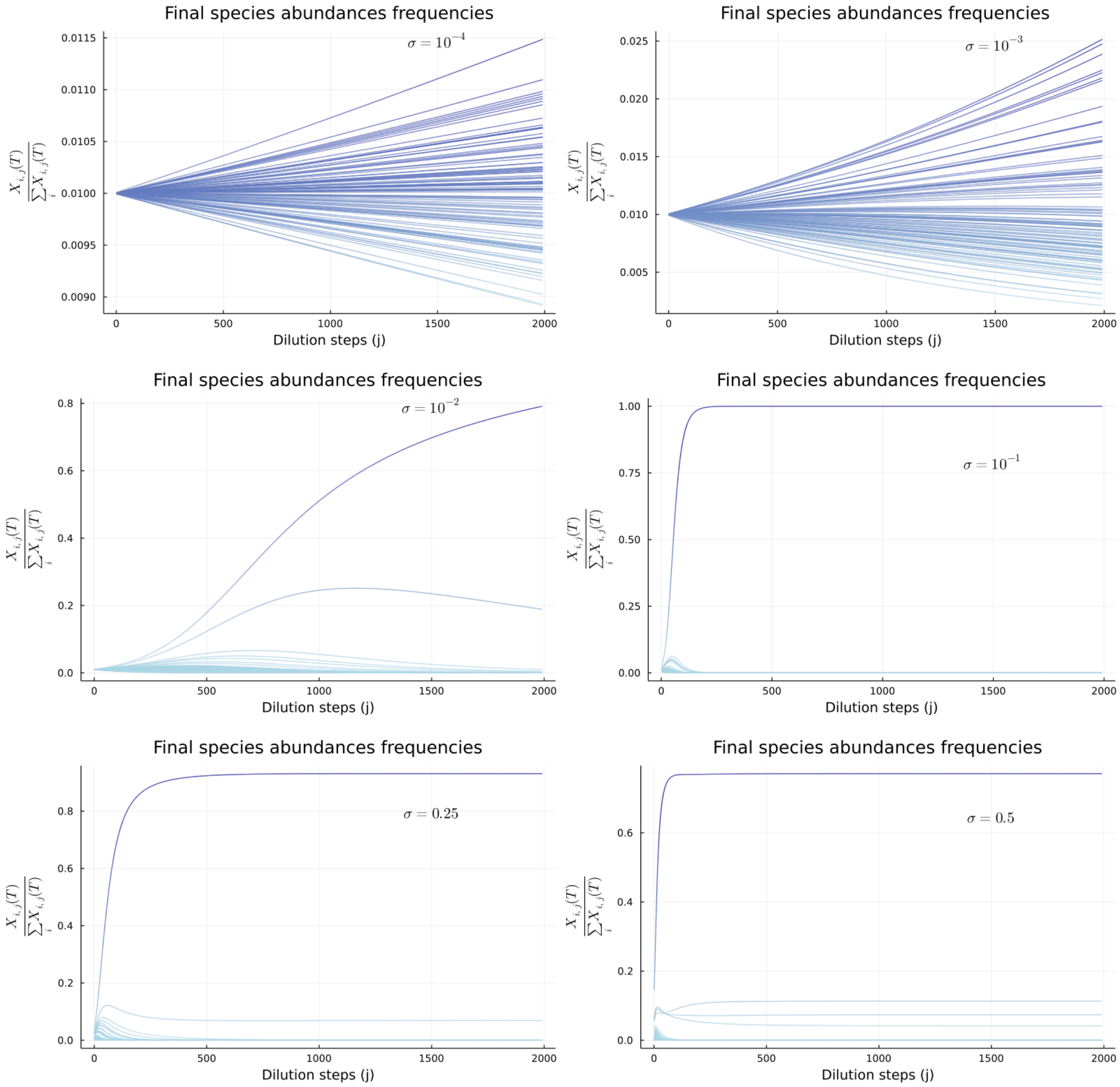
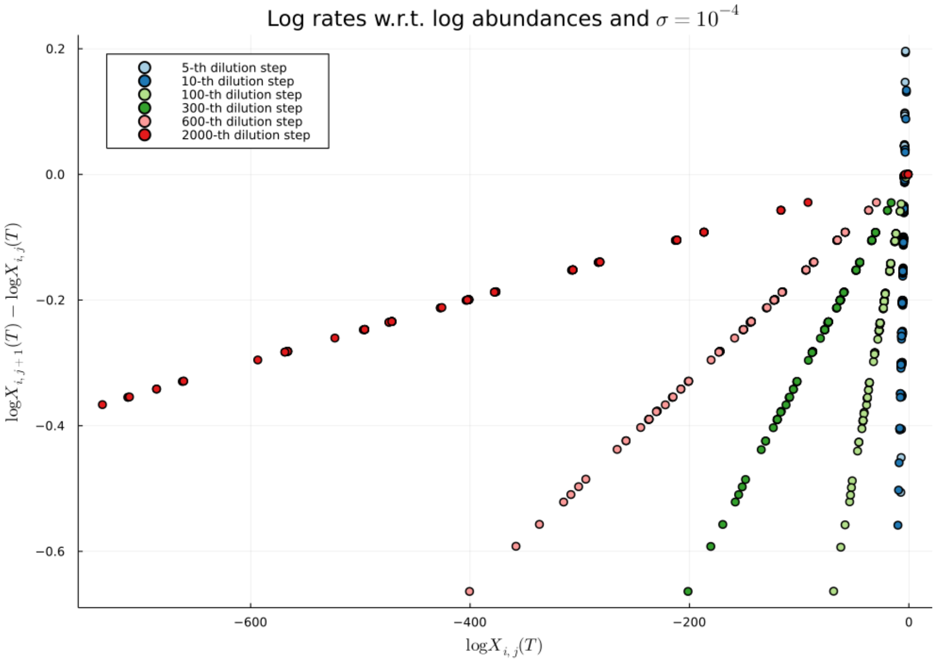
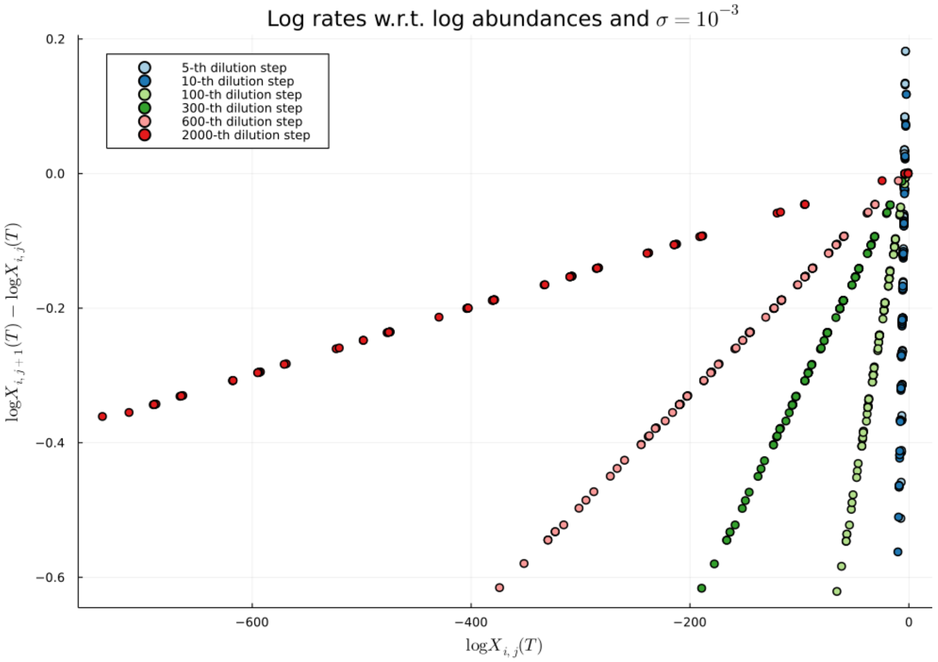
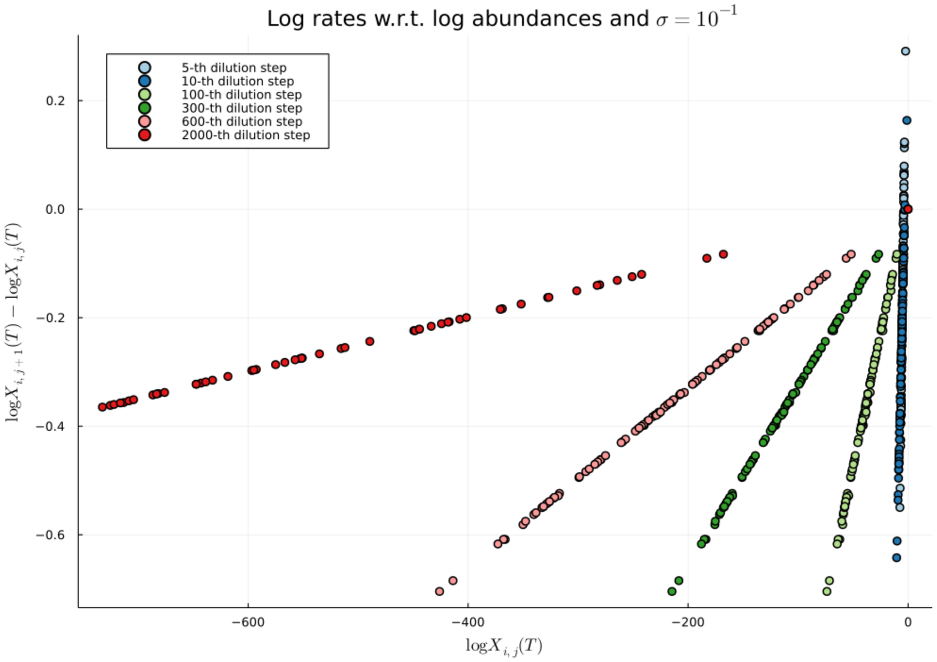


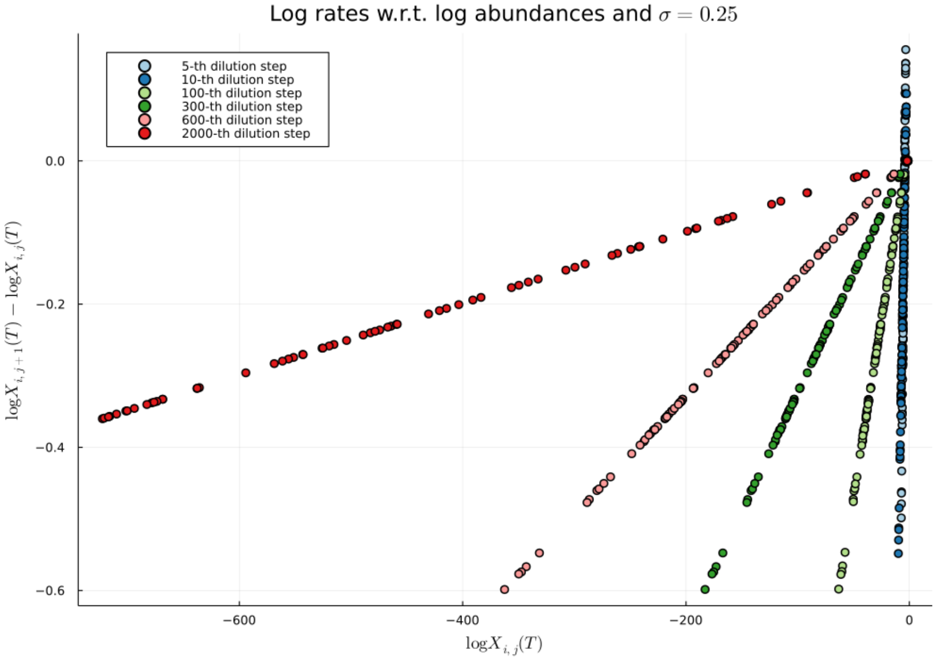
Figure 52: Dense matrix case.


Log rates w.r.t. log abundances and $\sigma = 10^{-4}$


Log rates w.r.t. log abundances and $\sigma = 10^{-3}$



Log rates w.r.t. log abundances and $\sigma = 10^{-1}$



Log rates w.r.t. log abundances and $\sigma = 0.25$

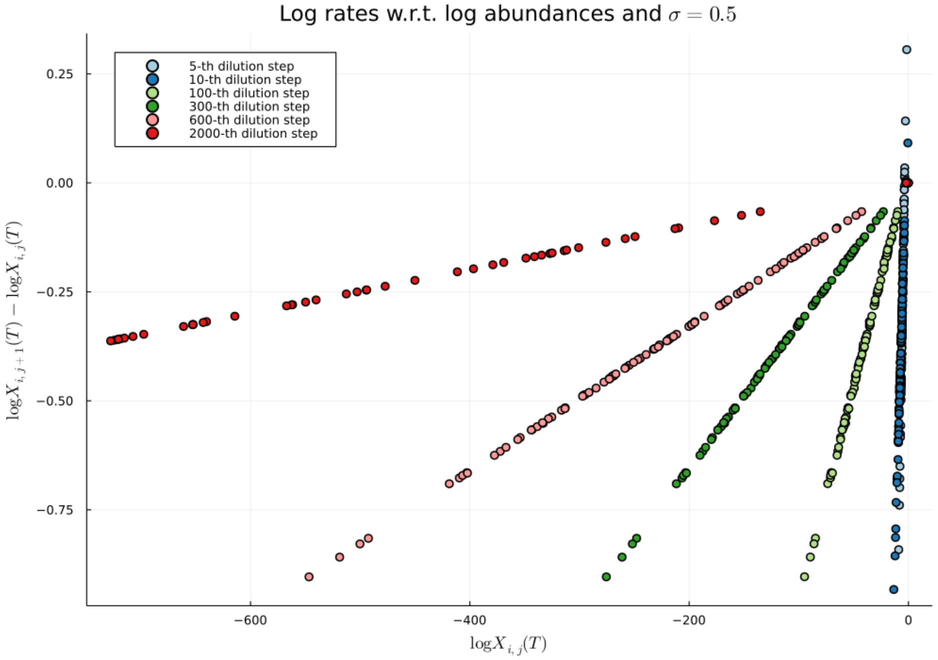
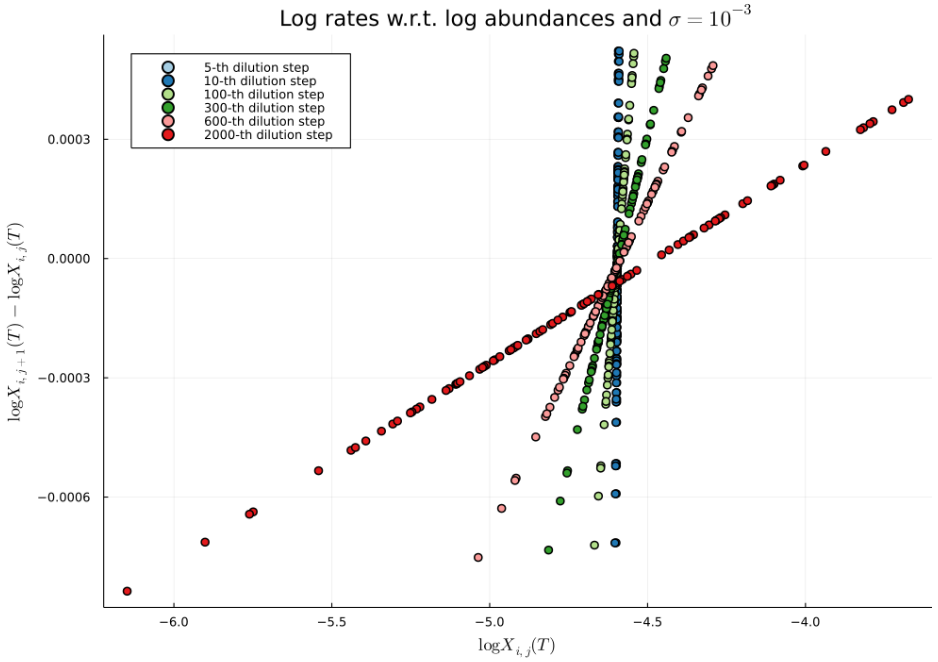
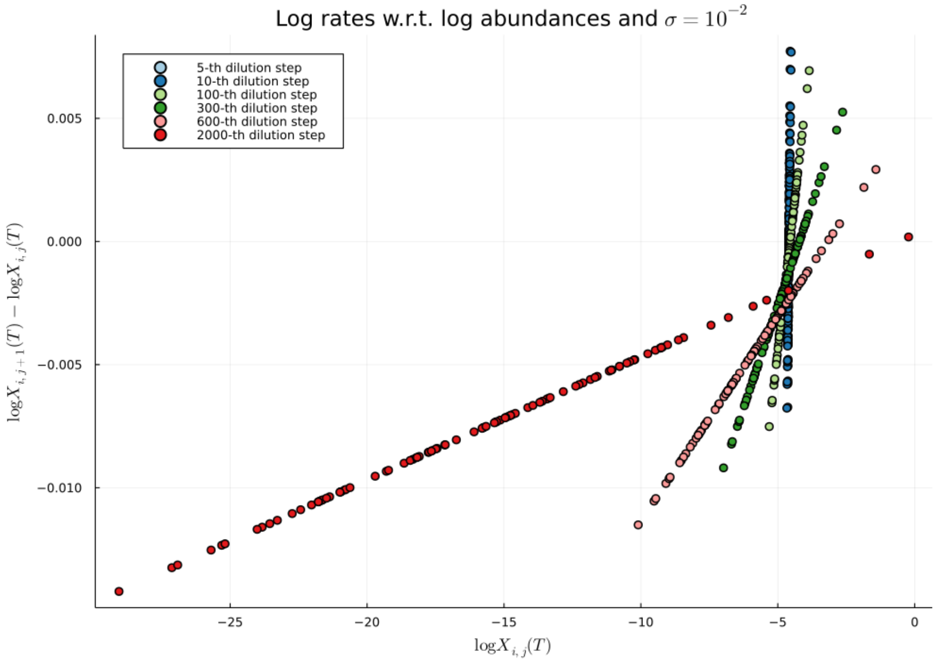
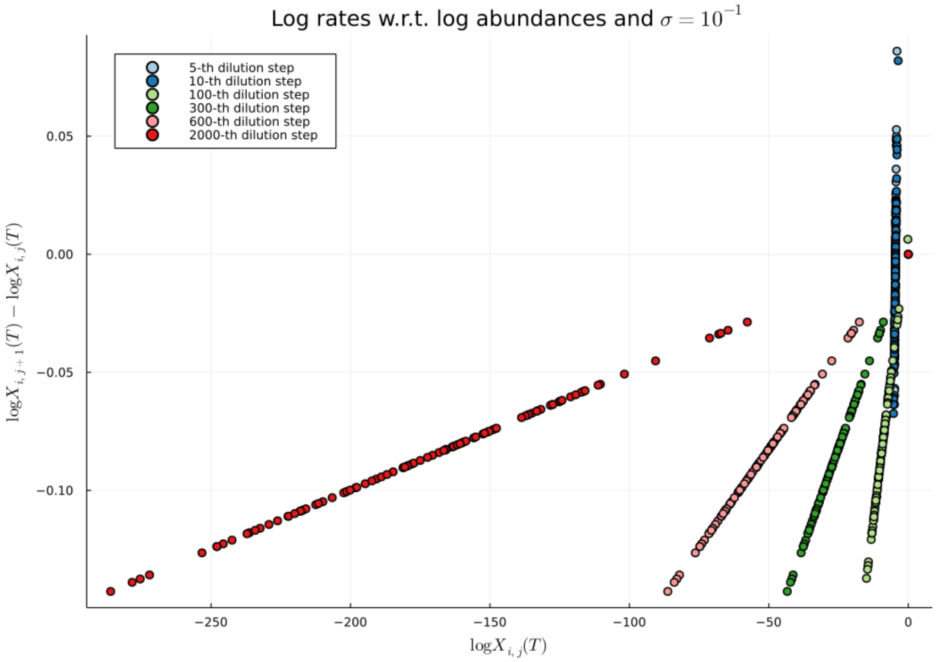


Figure 53: Sparse matrix case $S = 0.1$.




Log rates w.r.t. log abundances and $\sigma = 10^{-2}$


Log rates w.r.t. log abundances and $\sigma = 10^{-1}$

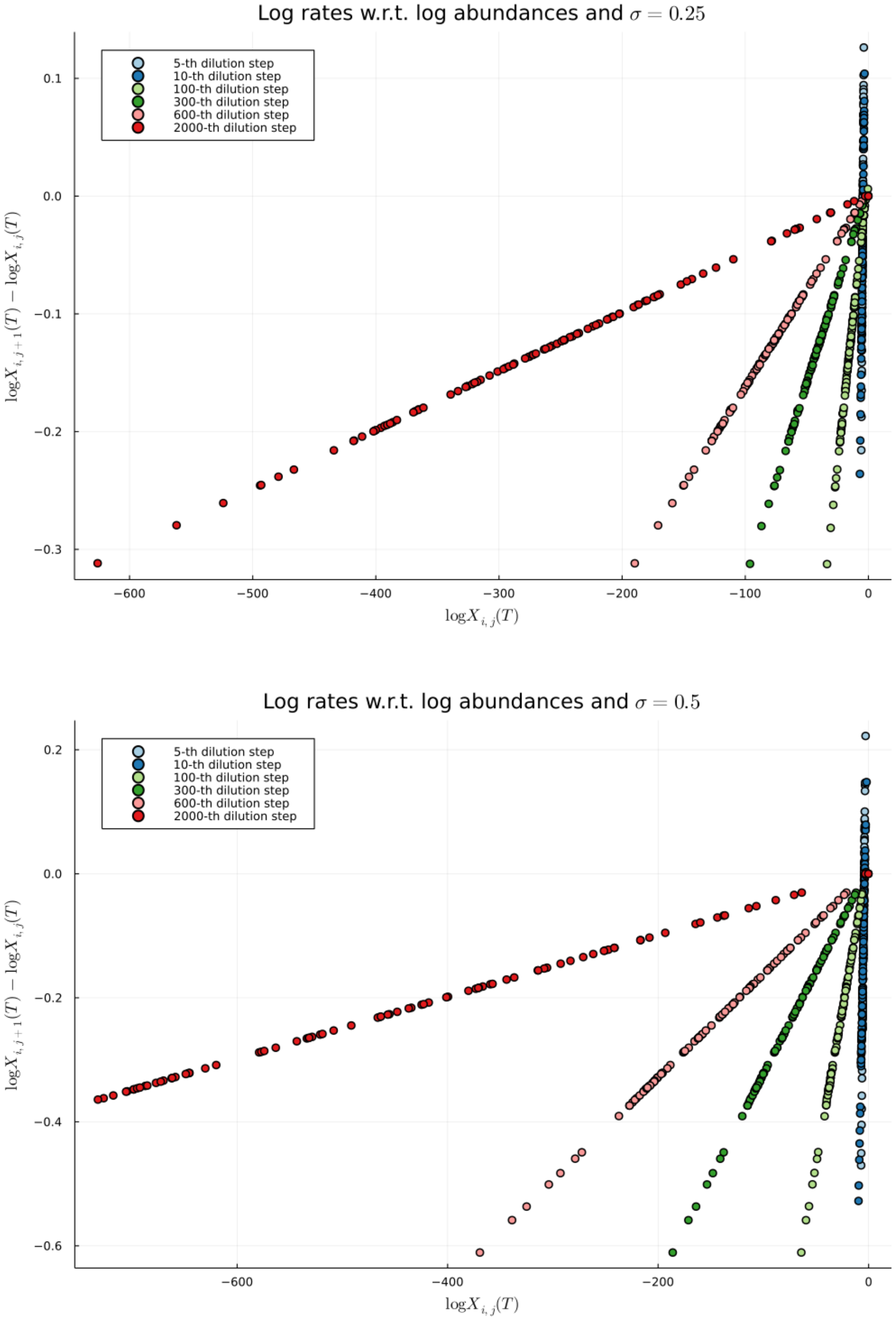


Figure 54: Dense matrix case.

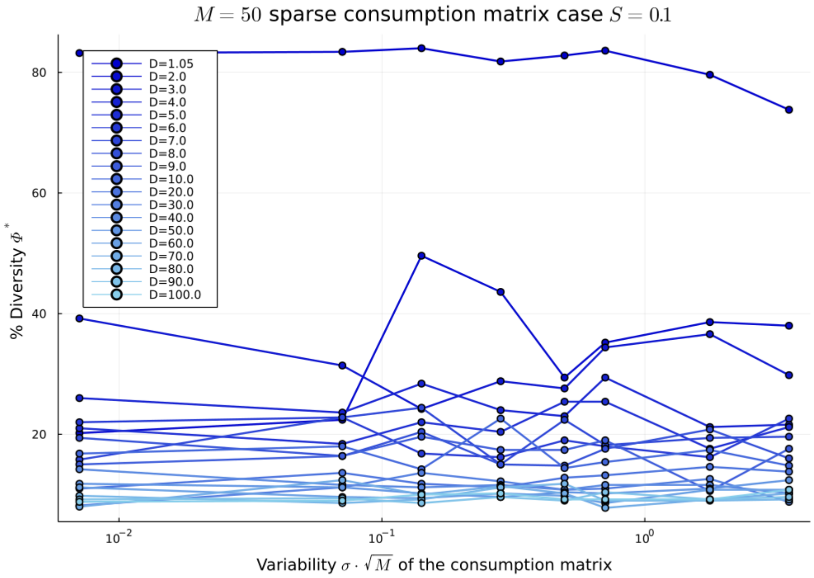


Figure 55: Sparse matrix case $S = 0.1$.

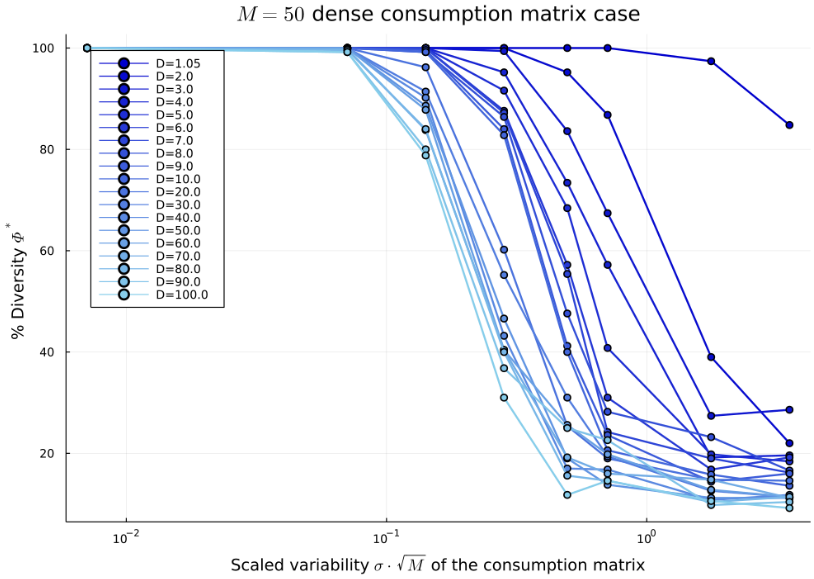


Figure 56: Dense matrix case

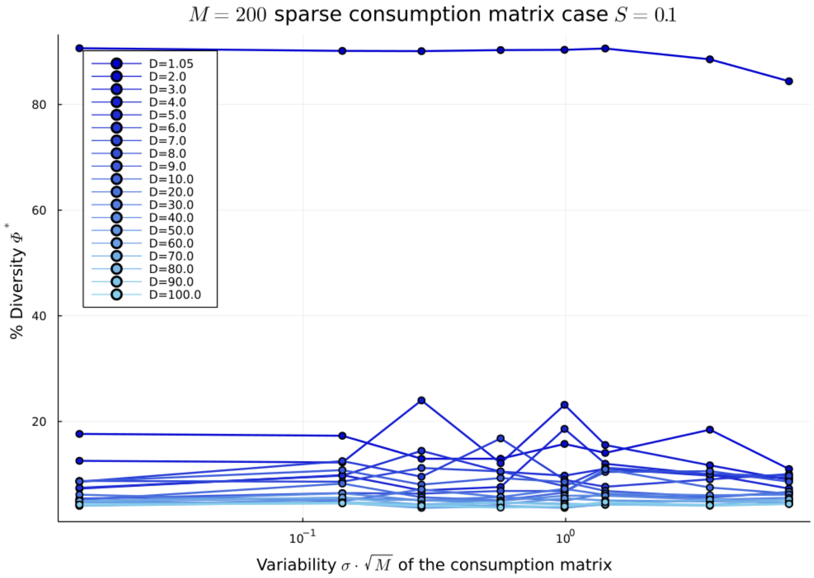


Figure 57: Sparse matrix case $S = 0.1$.

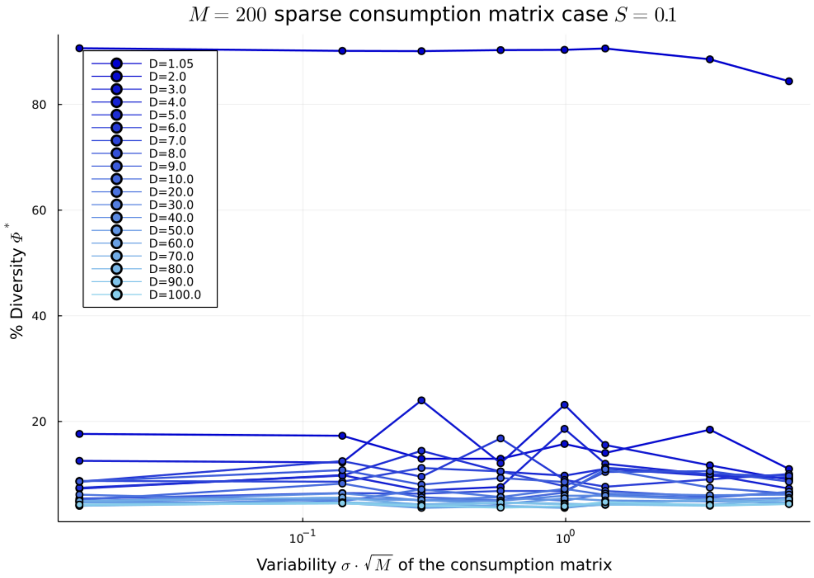
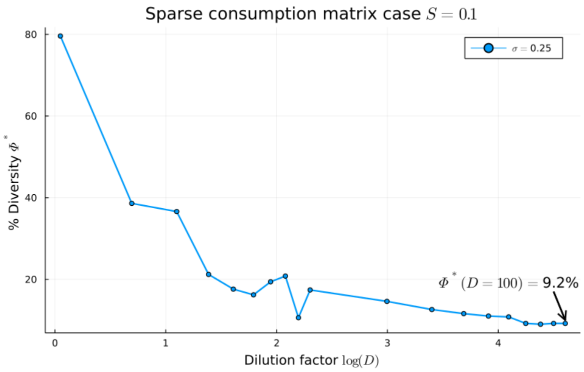
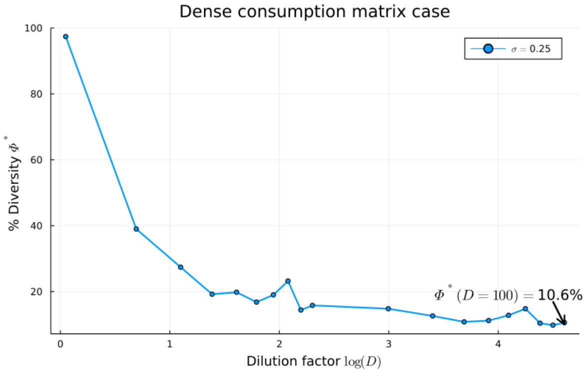


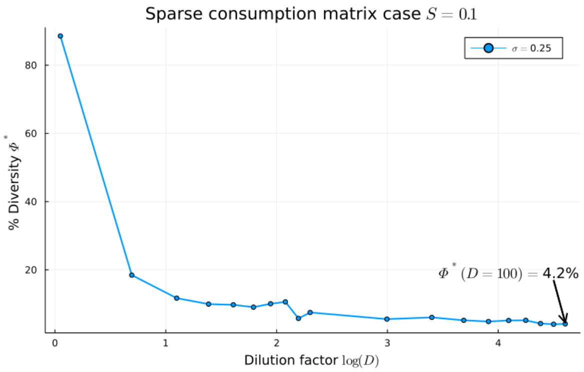
Figure 58: Sparse matrix case $S = 0.1$.



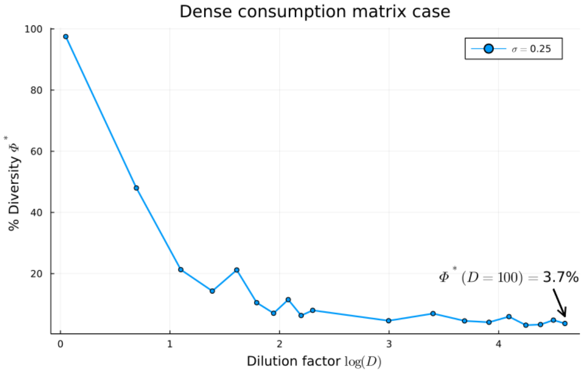
(a) Sparse matrix case $S = 0.1$.



(b) Dense matrix case



(a) Sparse matrix case $S = 0.1$.



(b) Dense matrix case

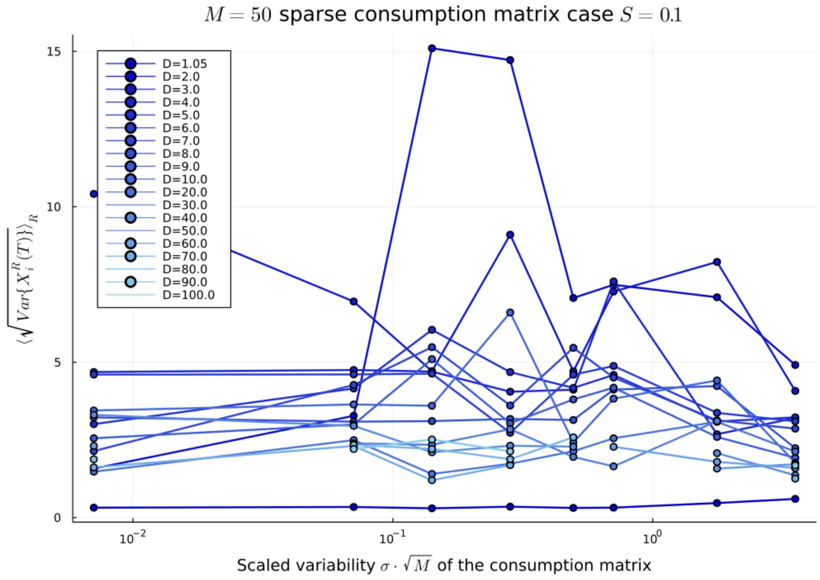


Figure 61: Sparse matrix case $S = 0.1$.

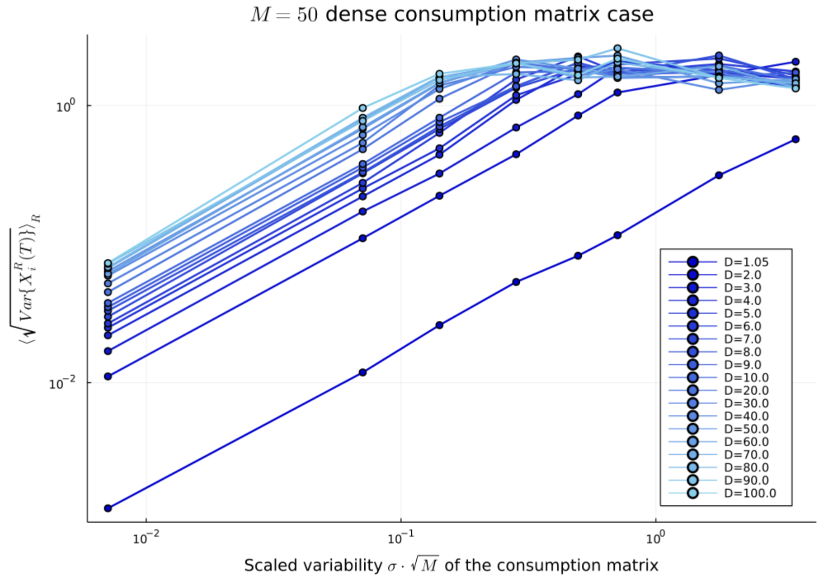


Figure 62: Dense matrix case

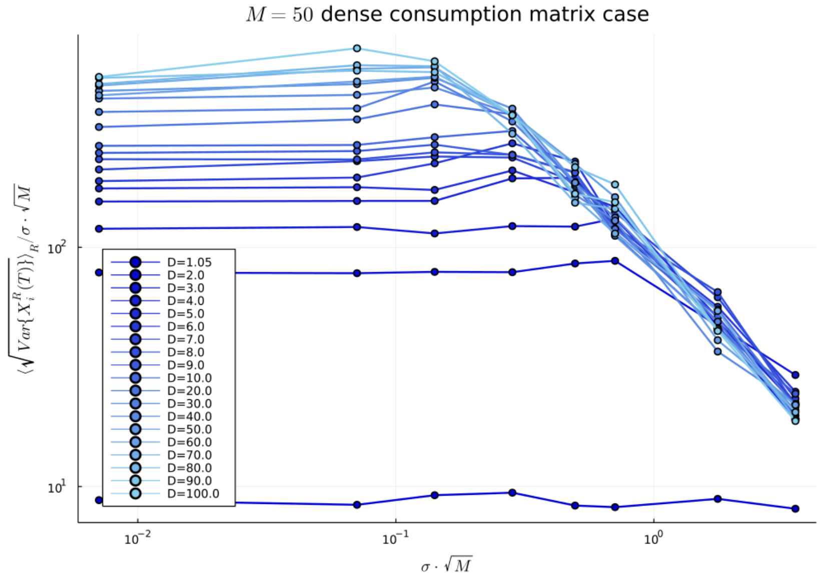


Figure 63: Dense matrix case

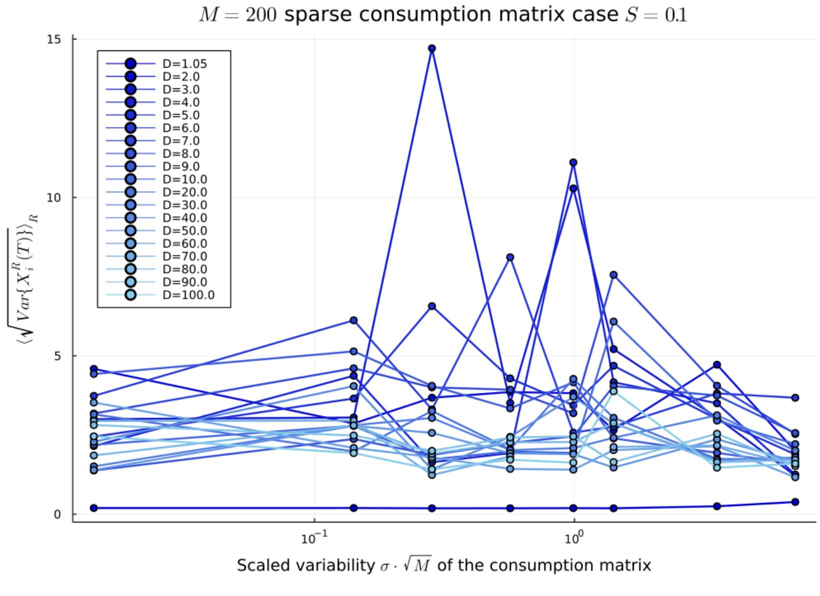


Figure 64: Sparse matrix case $S = 0.1$.

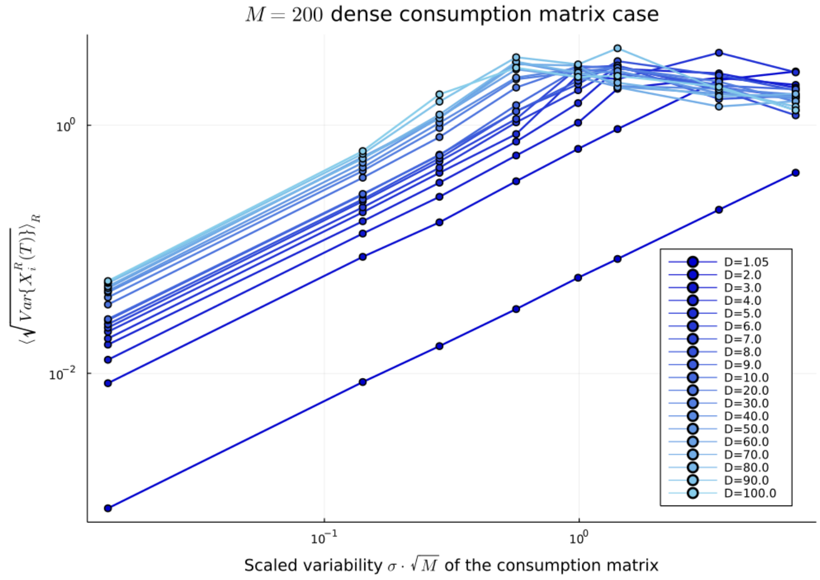


Figure 65: Dense matrix case

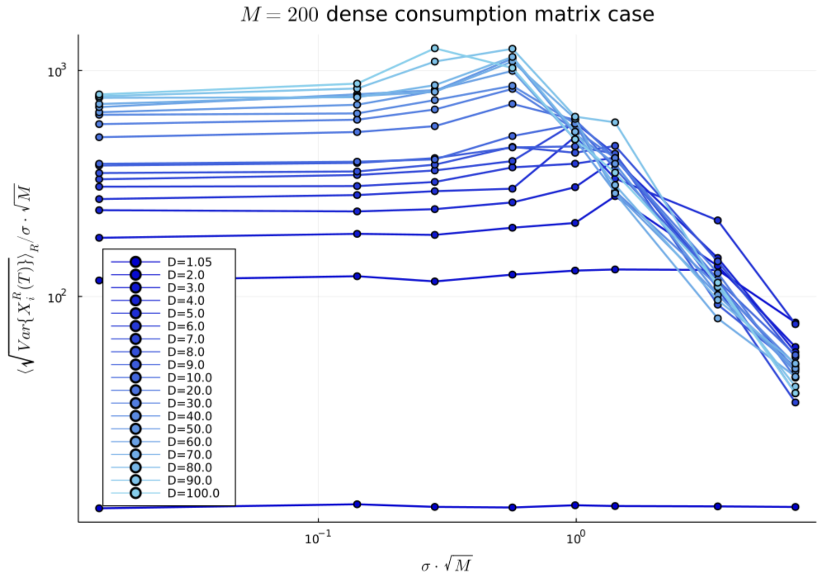


Figure 66: Dense matrix case

References

A.David, et al (2013). “Diet rapidly and reproducibly alters the human gut microbiome”. In: *Nature*. DOI: https://doi.org/10.1038/nature12820.

A.Day C.Schneider, B.L.Schneider (2004). *Yeast Cell Synchronization*. URL: https://bionumbers.hms.harvard.edu/bionumber.aspx?id=100986&ver=3#:~:text=One%5C%20OD600%5C%20is%5C%20equivalent%5C%20to,cells%5C%2FmL%5C%20(3) (visited on 11/14/2022).

A.Erez J.G.Lopez, B.G.Weiner Y.Meir N.S.Wingreen (2020). “Nutrient levels and trade-offs control diversity in a serial dilution ecosystem”. In: *eLife*. DOI: https://doi.org/10.7554/eLife.57790.

A.Posfai T.Taillefumier, N.S.Wingreen (2017). “Metabolic Trade-Offs Promote Diversity in a Model Ecosystem”. In: *Physical review letters*. DOI: 10.1103/PhysRevLett.118.028103.

B.Good P.Ho, K.C. Huang (2022). “Competition for fluctuating resources reproduces statistics of species abundance over time across wideranging microbiotas”. In: *eLife*.

C.A.Servan J.A.Capitan J.Grilli, K.E.Morrison S.Allesina (2018). “Coexistence of many species in random ecosystems”. In: *Nature*.

C.J.Krebs (1994). *Ecology: The Experimental Analysis of Distribution and Abundance*. HarperCollins College Publishers.

D.Tilman (1982). *Resource Competition and Community Structure*. Monographs in population ecology.

H.G.Andrewartha (1961). “An analysis of the idea of ’resources’ in animal ecology”. In: *Journal of theoretical ecology*. DOI: 10.1016/0022-5193(61)90028-5.

J.Faith, et al (2013). “The long-term stability of the human gut microbiota”. In: *Science*. DOI: 10.1126/science.1237439.

J.G.Caporaso C.L.Lauber, W.A.Walters and R.Knight (2011). “Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample”. In: *PNAS*. DOI: https://doi.org/10.1073/pnas.1000080107.

J.Grilli (2020a). “Macroecological laws describe variation and diversity in microbial communities”. In: *Nature*. DOI: https://doi.org/10.1038/s41467-020-18529-y.

— (2020b). “Macroecological laws describe variation and diversity in microbial communities”. In: *Nature Communications*. DOI: https://doi.org/10.1038/s41467-020-18529-y.

J.Kuenen (2019). “Continuous Cultures (Chemostats)”. In: *Encyclopedia of Microbiology*. DOI: https://doi.org/10.1016/B978-0-12-801238-3.02490-9.

L.Descheemaeker, S.de Buyl (2020). “Stochastic logistic models reproduce experimental time series of microbial communities”. In: *eLife*. DOI: `https://doi.org/10.7554/eLife.55650`.

M.Barbier, J.F.Arnoldi (2017). “The cavity method for community ecology”. In: *BioRxiv*. DOI: `https://doi.org/10.1101/147728`.

M.G.Bowler, C.K.Kelly (2010). “The general theory of species abundance distributions”. In: *arXiv*.

P.Chesson (1989). “MacArthur’s Consumer- Resource Model”. In: *Theoretical population ecology*.

P.Mehta M.Advani, G.Bunin (2018). “Statistical physics of community ecology: a cavity solution to MacArthur’s consumer resource model”. In: *Sissa*. DOI: `https://doi.org/10.1088/1742-5468/aab04e`.

P.Mehta W.Cui, R.Marsland (2020). “Effect of Resource Dynamics on Species Packing in Diverse Ecosystems”. In: *BioRxiv*.

P.Turchin (2001). “Does Population Ecology Have General Laws?” In: *OIKOS*. DOI: `https://doi.org/10.1034/j.1600-0706.2001.11310.x`.

— (2003). *Complex Population Dynamics: A Theoretical/Empirical Synthesis*. URL: `http://www.jstor.org/stable/j.ctt24hqkz` (visited on 11/15/2022).

R.A.Armstrong, R.McGehee (1978). “Competitive Exclusion”. In: *The American Naturalist*.

S.Allesina, J.M.Levine (2011). “A competitive network theory of species diversity”. In: *PNAS*. DOI: `www.pnas.org/lookup/suppl/doi:10.1073/pnas.1014428108/-/DCSupplemental`.

S.Azaele S.Suweis J.Grilli, I.Volkov J.R.Banavar and A.Maritan (2015). “Statistical Mechanics of Ecological Systems: Neutral Theory and Beyond”. In: *arXiv*.

S.Gupta S.M Mortensen, S.Schjørring (2019). “Amplicon sequencing provides more accurate microbiome information in healthy children compared to culturing”. In: *Nature*. DOI: `https://doi.org/10.1038/s42003-019-0540-1`.

S.Levin (1970). “Community equilibria and stability, and an extension of the competitive exclusion principle”. In: *The American Naturalist*.

S.Strogatz (1994). “Nonlinear Dynamics and Chaos”. In: *Advanced book program Perseus books Reading*. DOI: `https://doi.org/10.1063/1.4823332`.

W.Cui P.Mehta, R.Marsland (2021). “Diverse communities behave like random ecosystems”. In: *BioRxiv*.

W.Shoemaker, et al (2017). “A macroecological theory of microbial biodiversity”. In: *Nature*. DOI: `https://doi.org/10.1038/s41559-017-0107`.