UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia "Galileo Galilei"

CORSO DI LAUREA IN FISICA

TESI DI LAUREA

Modello eco-evolutivo per l'evoluzione sintetica del DNA basata sull'affinità

Eco-Evolutionary Model for Affinity-based DNA Synthetic Evolution

Relatore
Prof. Samir Simon Suweis

Laureando
Augusto Piccolin

Anno Accademico 2024/2025

Riassunto in italiano

Nella seguente tesi andremo ad indagare il contesto della competizione ecologica e, in particolare, il concetto di coesistenza delle specie.

Viene proposto un modello analitico, basato sui modelli Consumer-Resource e Generalized Lotka-Volterra, appositamente costruito e adattato per riprodurre le condizioni ambientali e sperimentali adottate in Casiraghi et al. (2024) [2].

In tale esperimento, un ecosistema di individui sintetici, rappresentati da filamenti di DNA a singola elica, viene fatto evolvere in un ambiente dove è presente una singola risorsa, rappresentata anch'essa da un filamento di DNA singolo.

La selezione avviene tramite affinità tra le sequenze, che permette agli individui di legare e venire selezionati. Una volta che la risorsa è esaurita, il sistema viene ripristinato proporzionalmente allo stato raggiunto e la dinamica riparte, in una struttura a cicli.

Dopo aver introdotto i modelli teorici principali per la trattazione dei sistemi ecologici in cui più specie competono per una risorsa e interagiscono, studiamo le proprietà del particolare modello scelto (eq. 3.10) ed effettuiamo una serie di simulazioni numeriche.

Per rendere più pratico lo studio, partiamo da una versione più semplice del modello e aggiungiamo via via complessità. Ad ogni passaggio, cerchiamo di caratterizzare la dinamica analiticamente descrivendone gli equilibri e le principali proprietà, per poi confrontare tali predizioni coi risultati numerici.

In particolare, analizziamo gli equilibri delle equazioni, che rappresentano la dinamica di un singolo ciclo, e poi passiamo alla dinamica multi-ciclo cercando di capire quando essa riproduce i risultati precedenti e quando invece influenza la sopravvivenza delle specie.

Tale dinamica è caratterizzata dal processo di ricampionamento, con cui le specie alla fine di un ciclo vengono selezionate in base alla loro abbondanza e poi amplificate fino a tornare a un totale fissato di individui. Il ricampionamento viene inizialmente implementato con una semplice soglia, con cui si eliminano tutte le specie meno abbondanti di una certà quantità fissata. Ciò velocizza molto la dinamica e porta a uno stato stazionario che differisce, seppur di poco, da quello predetto analiticamente (figure 12a e 13a).

Per l'ultimo caso, dove le equazioni comprendono le mutue interazioni e corrispondo perciò al modello completo, si è svolta un'analisi più approfondita.

Per quanto riguarda la dinamica del singolo ciclo, si è trovata una soluzione analitica risolvendo un Linear Complementarity Problem equivalente al sistema, confrontata poi con quella trovata tramite simulazione (figure 15b e 16b).

Analizzando invece la dinamica multi-ciclo, troviamo gli stessi risultati dei casi precedenti con il ricampionamento in soglia (figura 20a). Poiché questa non è descrittiva del reale processo sperimentale, implementiamo un nuovo metodo di ricampionamento, basato sulla selezione casuale dei superstiti secondo una distribuzione multinomiale, dove le probabilità di estrazione sono le abbondanze relative con cui le specie arrivano alla fine del ciclo precedente.

Con questo modello, le abbondanze delle specie sopravvissute oscillano casualmente a causa della stocasticità del processo, e non è perciò possibile definire un equilibrio deterministico come nei casi precedenti (figure 23a e 23b).

L'analisi effettuata si concentra perciò su come la probabilità di estinzione associata al processo stocastico influenzi la distribuzione delle abbondanze (figura 25), riuscendo in particolare a dare una stima del numero di specie sopravvissute in base al numero del ciclo (figure 27a e 27b).

Contents

1 Introduction **3**

2 DNA Experiment **3**

3 Mathematical models and methods **4**
3.1 Modeling species population dynamics and interactions with resources 4
3.1.1 Consumer Resource Model 4
3.1.2 Lotka-Volterra equations and species interaction 5
3.2 Model for the single-cycle dynamics 6
3.3 Implementation of the serial cycle dynamics 6

4 Simulation of a system with non-interacting species **7**
4.1 Model properties 7
4.2 Simulation of multi-cycle dynamics with no sampling 8
4.3 Simulation of multi-cycle dynamics with sampling 9

5 Simulation of a system with self-interacting species **10**
5.1 Model Properties 10
5.2 Simulation of multi-cycle dynamics with no sampling 11
5.3 Simulation of multi-cycle dynamics with sampling 13

6 Simulation of a system with self and mutual interactions between species **14**
6.1 Model properties 14
6.2 Simulation of single cycle dynamics 16
6.3 Simulation of multi-cycle dynamics with no sampling 18
6.4 Simulation of multi-cycle dynamics with sampling 19
6.4.1 Sampling with threshold 19
6.4.2 Stochastic multinomial sampling 21
6.4.3 Stochastic process and species abundance distribution 22

7 Conclusion and further developments **24**

1 Introduction

In theoretical biology and ecology, when it comes to understanding the properties and evolution of a particular ecosystem, one of the most fundamental aspects is the coexistence of different species [9].

Several attempts have been made to explain how many different species can coexist together in the same environment, despite competing for the same limited resources. The traditional explanation relies on the concept of niches, which can be expressed as the set of environmental conditions and resources that allow a species to survive, grow, and reproduce; including abiotic factors and biotic interactions [4].

In this framework, species that coexist in the same environment simply occupy different niches, which means that their biological role in the environment is different. If two species compete for the same resource, the fittest will eventually outcompete the other and be the only one to survive. This rationale is consistent with a large number of theoretical models, such as the consumer-resource model originally proposed by MacArthur[1]: as we will show in this work, the number of species that can coexist at stationarity is limited by the number of different resources in the system. This result is also known as Competitive Exclusion Principle.

However, several natural environments have been observed in which a large number of consumers coexist despite competing for the same limited resources. One example is plankton, where thousands of different species live in an environment with a dozen of different resources: this is also known as the Plankton paradox [8].

It is generally difficult to study and model systems with a large variety of species, due to the inherent complexity of ecosystems, the large number of potentially relevant variables of difficult experimental access, and the role of stochasticity [3].

In this context, it is important to develop new experimental and theoretical methods to explore eco-evolutionary models with accessible and easily computable parameters, in order to better understand evolutionary processes and the role of fitness in species coexistence. To this aim, recently a molecular platform based on DNA to study the synthetic evolution of species has been proposed [2].

In this thesis, we propose a theoretical framework to study the dynamics of competing species that can be used to describe the experiment of Casiraghi et al. (2024) [2], where a synthetic eco-evolutionary scheme is used to analyze the role of fitness in the evolution of a spatially homogeneous ecosystem.

In section 2, we present the experiment by explaining the ecosystem's features and the protocol used. In section 3 we describe the main theoretical models used to characterize species interaction and growth, namely the Consumer-Resource Model (CRM) and the Generalized Lotka-Volterra equation (GLV).

Then, in Sections 4 to 6, we try to use those models to build a theoretical framework that reproduces the experimental setting. In each of these sections, we focus on characterizing the equilibria and on predicting the number of species that survive or go extinct.

2 DNA Experiment

The synthetic eco-evolutionary scheme proposed in the experiment [2] is based on the direct interaction between single-strand DNA sequences, with no connection with the molecular biology of the cell: no coding sequences, no translation, no proteins involved.

The ecosystem consists of a large number ($\sim 10^{15}$) of different individuals, each characterized by its unique sequence of $L = 50$ monomers, competing for a fixed amount of resources.

The resource consists of $\sim 10^{12}$ single-strand DNA oligonucleotides, all of the same sequence of length $L = 20$, to which individuals can bind according to their affinity and level of complementarity.

The system starts from this initial setting and evolves, with DNA individuals (DNAi) binding to the resource's strands until they are all occupied. Then, the resource is extracted and the remaining individuals are discarded, while the "survived" species are PCR-amplified to the initial total abundance. A small fraction of the amplified sample ($1 - 3 \times 10^6$ molecules) is analyzed to determine the properties of the survived pool. Since this fraction is negligible compared to the total abundance, we can ignore the fact that some individuals are missing in the following cycles.

The described process constitutes a cycle, that is then repeated a fixed number of times. The protocol is shown in figure (1).

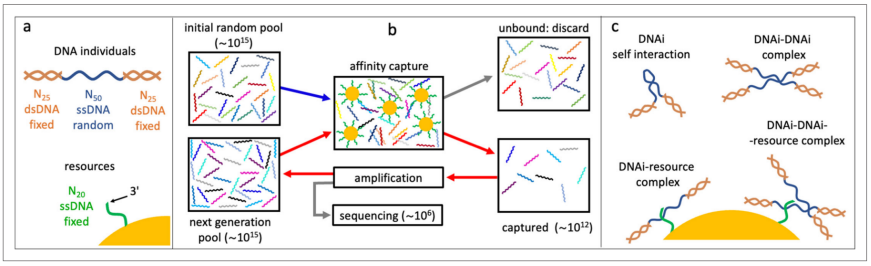


Figure 1: Evolutionary protocol of the DNA ecosystem, with the description of the cycle dynamics. The image is directly taken from [2].

To actually enable sequencing and amplification, the individuals are built with two 25 base-long fixed sequences on the side, which are made inactive by hybridization with the complementary oligonucleotide.

During the evolution process, the species show, in addition to the predicted interaction with the resource, also self and mutual binding that influence their survival ability.

3 Mathematical models and methods

3.1 Modeling species population dynamics and interactions with resources

3.1.1 Consumer Resource Model

To describe the evolution of an ecosystem, where a certain number of consumers interact with a certain amount of resources, we can use a set of differential equations that model their growth and interaction.

A powerful model that we can use for this ecological relationship is the Consumer-Resource Model [11]. Using $N_i(t)$ as the number of individuals of the consumer species i and $R_\sigma(t)$ as the amount of the resource σ, we can write:

$$\begin{aligned} \frac{dN_i}{dt} &= N_i g_i(R_1,\dots,R_M) \\ \frac{dR_\sigma}{dt} &= f_\sigma(R_1,\dots,R_M,N_1,\dots,N_S) \end{aligned} \tag{3.1}$$

where $g(R)$ is the growth rate of species i, and f_σ is the growth rate of resource σ. This general form contains a limitation of the model, namely the absence of explicit interactions between different species.

One of the first proposed Consumer Resource Models is the one named after Robert H. MacArthur [1], which serves as a basis for a large variety of ecological system descriptions:

$$\begin{aligned} \frac{dN_i}{dt} &= N_i\left(\sum_{\sigma=1}^{M} v_\sigma \alpha_{i\sigma} R_\sigma - d_i\right) \\ \frac{dR_\sigma}{dt} &= \mu(R_\sigma) - \sum_{i=1}^{S} R_\sigma \alpha_{i\sigma} N_i \end{aligned} \tag{3.2}$$

where v_σ summarizes consumption efficiency, $\alpha_{i\sigma}$ is the metabolic strategy of species i for resource σ and d_i is the death rate of species i. The function μ depends on the environment and the type of resources; it is a constant supply rate when the aim is to model a chemostat environment [12]; it is modeled as a logistic function with carrying capacity k, when the resource is biotic [13].

This model is linked to a fundamental concept of ecology, the so-called Competitive Exclusion Principle. It states that two species competing for the same limited resource cannot coexist, i.e., one will eventually outcompete the other [6]. We can derive it directly from equation (3.2), imposing the equilibrium

condition:

$$\dot{N}_i = 0 \implies \sum_{\sigma=1}^{M} \underbrace{v_\sigma \alpha_{i\sigma}}_{\hat{\alpha}_{i\sigma}} R_\sigma = d_i \implies \sum_{\sigma=1}^{M} \hat{\alpha}_{i\sigma} R_\sigma = d_i \quad i = 1, \dots, S. \tag{3.3}$$

These are S equations in M variables, which means that it is solvable $\iff S \leq M$. So, in Mac Arthur's model, only a number of consumers at most equal to the number of resources can survive.

3.1.2 Lotka-Volterra equations and species interaction

Another important step in precisely describing the complex ecological relationships between species is to model their direct interactions.

To do so, we can use the Generalized Lotka-Volterra equations, which can be written as:

$$\frac{dN_i}{dt} = \left(a_i - \sum_{j=1}^{S} \alpha_{ij} N_j \right) N_i \tag{3.4}$$

where a_i represents the intrinsic growth rate of species i. The coefficients α_{ij} encode the interaction network between species: if $\alpha_{ij} > 0$, the j^{th} species preys the i^{th} one, while if $\alpha_{ij} < 0$, the species j^{th} is predated by the i^{th} one; if both are positive, their relationship is simply of competition.

The system is at equilibrium, other than in the trivial case $\mathbf{N}^* = 0$, when $\mathbf{N}^* > 0$ and satisfies

$$a_i = \sum_{j=1}^{S} \alpha_{ij} N_j^*. \tag{3.5}$$

We can achieve important information about the system by studying the stability of this equilibrium. We first compute the Jacobian:

$$J_{ij} = \frac{\partial \dot{N}_i}{\partial N_j}\Big|_{N^*} = N_i^* \alpha_{ij}. \tag{3.6}$$

Since we are dealing with populations and therefore the entries of $\mathbf{N}^*$ must be positive, the sign of the eigenvalues of α and J is the same, and the stability of the equilibrium depends entirely on the adjacency matrix.

In some simple cases the stability can be immediately deduced. For example, when the adjacency matrix is antisymmetric ($\alpha_{ij} = -\alpha_{ji}$) its eigenvalues are purely imaginary and the solutions oscillate around the equilibrium point. In general, equilibrium is not guaranteed to be stable and, even if it is, we cannot predict which and how many species will eventually be present or extinct [7].

It can also be shown that within certain hypotheses the CRM and GLV descriptions are equivalent. For example, let us consider a system described by eq. (3.2) with the logistic growth for resources:

$$\mu(R_\sigma) = r_\sigma R_\sigma \left(1 - \frac{R_\sigma}{k} \right). \tag{3.7}$$

We can then use the Quasi-Stationary Approximation, where we suppose that resources react quickly to changes in abundance of species, and thus replace $R_\sigma = R_\sigma^*$ in the equations for N_i:

$$\begin{aligned} \dot{R}_\sigma = 0 &\implies r_\sigma R_\sigma \left(1 - \frac{R_\sigma}{k} \right) = \sum_{i=1}^{S} R_\sigma \alpha_{\sigma i} N_i \\ &\implies R_\sigma^* = k \left(1 - \frac{1}{r_\sigma} \sum_{i=1}^{S} \alpha_{\sigma i} N_i \right), \end{aligned} \tag{3.8}$$

from which we get:

$$\begin{aligned} \dot{N}_i &= N_i \left(\sum_{\sigma=1}^{M} v_\sigma \alpha_{i\sigma} R_\sigma^* - d_i \right) = N_i \left[\sum_{\sigma=1}^{M} v_\sigma \alpha_{i\sigma} k \left(1 - \frac{1}{r_\sigma} \sum_{j=1}^{S} \alpha_{\sigma j} N_j \right) - d_i \right] \\ &= N_i \left[\underbrace{\sum_{\sigma=1}^{M} v_\sigma \alpha_{i\sigma} k - d_i}_{\gamma_i} - \sum_{j=1}^{S} \underbrace{\sum_{\sigma=1}^{M} \frac{k v_\sigma}{r_\sigma} \alpha_{i\sigma} \alpha_{\sigma j}}_{\hat{\alpha}_{ij}} N_j \right] = N_i \left[\gamma_i - \sum_{j=1}^{S} \hat{\alpha}_{ij} N_j \right], \end{aligned} \tag{3.9}$$

which is a set of differential equations equivalent to the GLV one given by eq. (3.4).

3.2 Model for the single-cycle dynamics

We now aim to model the DNAi experiment, by describing the evolution of the single resource and the large number of species competing for it.

The resource dynamics is simple: during the interaction period (i.e. a cycle) there is no supply and its dynamics is only described by consumption from the species, which causes its eventual depletion.

For consumers, we need to account for the resource consumption and the inter- and intra-species interactions. Because the survival mechanism and mutual interactions are based on DNA hybridization, affinity with the resource or other individuals is the only factor that affects the species growth rate. In addition, species do not have a mortality rate.

Therefore, the model chosen to describe this ecosystem's dynamics is the following:

$$\begin{aligned} \frac{dn_i}{dt} &= n_i \left(a_i r(c) - \sum_{j=1}^{S} A_{ij} n_j \right) \\ \frac{dc}{dt} &= -r(c) \sum_{i=1}^{S} a_i n_i \end{aligned} \tag{3.10}$$

where we used the Monod function

$$r(c) = \frac{c}{K+c} \tag{3.11}$$

to model resource consumption by species, since they are abiotic [10].
In this simple formulation, the consumption strength of the S species is encoded by coefficients a_i, which are all positive, while the matrix A represents the interaction network between species. The entries of A are chosen to be all positive, to mimic self-limitation and ensure stability, therefore the interaction network displays only competition.

In the following sections, we are going to study the system by starting from a simpler case, where the species only interact with the resource, and then gradually building up complexity by introducing self and mutual interactions between species. This way, we can better understand its properties and explain the obtained results.

3.3 Implementation of the serial cycle dynamics

The model introduced above will be used to describe the dynamics of a single cycle, where consumers compete and grow until the resource runs out. If we want to mimic the serial approach of the experiment and study how it affects species survival, we need to define a new inter-cycle dynamics.

At the end of each cycle, the abundance of the resource is restored to its initial value c_0, and the dynamics start again for another cycle.
When this process is performed without considering any kind of species sampling, we have simple relationships for the species abundance dynamics over cycles:

$$n_i^{k+1}(0) = n_i^k(T) \; ; \;\; f_i^{k+1}(0) = f_i^k(T) \;\; \forall i. \tag{3.12}$$

In this simpler case, the dynamics of cycle k + 1 begins simply from that at the end of cycle k: we can think of the serial dynamics as the same as that of a single cycle with $c_0^{TOT} = c_0 k_{max}$.

A more correct approach consists of introducing a sampling process at the end of each cycle, which models the experimental selection of surviving species.

In the experiment, the total abundance of species is $\sim 10^{15}$, while the abundance of the resource is $\sim 10^{12}$. At the end of each cycle, the surviving 10^{12} individuals are amplified, with constant f_i, back to the original quantity.

We can model this mechanism by selecting a threshold under which a species is considered extinct; then we can obtain the relative abundances for the S' survivors as

$$f^k_{i,surv}(T) = \frac{n^k_{i,surv}(T)}{\sum_{j=1}^{S'} n^k_{j,surv}(T)} = f^{k+1}_i(0). \tag{3.13}$$

For the following cycle, we simply use

$$n^{k+1}_i(0) = f^{k+1}_i(0) N_{tot} \tag{3.14}$$

with $N_{tot} = \sum_i^S n^0_i(0)$, fixed to a value greater than c_0, to mimic limited resource availability.

For the non-interacting and the self-interacting systems, the threshold chosen is the following:

$$f_i(T) \geq \alpha \;\text{ with }\; \alpha = \min\left\{\frac{1}{S}, 0.005\right\}, \tag{3.15}$$

which increases over time, since the number of surviving species becomes smaller, and eventually saturates at 0.5% of the total.

The number of cycles must be large enough for the system to reach the approximate steady state, that is, the state at which

$$f^k_i(T) = f^{k+1}_i(T) \;\; \forall i \implies f^k_i(0) = f^k_i(T) \;\; \forall i. \tag{3.16}$$

Numerically, we implement this condition through an upper bound:

$$\left|\frac{f^k_i(T)}{f^k_i(0)} - 1\right| = \left|\frac{n^k_i(T)}{n^k_i(0)} - 1\right| < \varepsilon \;\; \forall i. \tag{3.17}$$

For example, choosing $\varepsilon = 0.001$ we ask all species to have a relative growth smaller than 0.1%.

When dealing with the case of self- and mutually interacting species, these can naturally go extinct, and the threshold will be used only to accelerate the dynamics.

4 Simulation of a system with non-interacting species

4.1 Model properties

We first consider a system of 1 resource and S consumers with the following growth dynamics:

$$\begin{aligned} \frac{dn_i}{dt} &= r(c) a_i n_i, \\ \frac{dc}{dt} &= -r(c) \sum_{i=1}^{S} a_i n_i. \end{aligned} \tag{4.1}$$

Although there is no closed solution to this ODE system, we can derive some interesting properties:

$$\frac{dc}{dt} = -\sum_{i=1}^{S} r(c) a_i n_i = -\sum_{i=1}^{S} \frac{dn_i}{dt}, \tag{4.2}$$

that leads to

$$\Delta c = -\sum_{i=1}^{S} \Delta n_i. \tag{4.3}$$

This means that during the cycle, the entire biomass of the resource is gained by consumers, which grow in abundance.

For species i we can write

$$\dot{n}_i = a_i r n_i \implies \int_{n_i(0)}^{n_i(T)} \frac{1}{n_i} dn_i = \int_0^T a_i r(c(t)) dt \tag{4.4}$$

$$\iff \ln \frac{n_i(T)}{n_i(0)} = a_i \int_0^T r(c(t))dt = a_i\gamma \implies g_i = \frac{n_i(T)}{n_i(0)} = e^{\gamma a_i},$$

where g_i is the growth rate of species i. It can be seen that the growth of species with greater a_i is exponentially favored.

We can also derive the species relative abundance dynamics

$$f_i = \frac{n_i}{\sum_j n_j} \implies \dot{f}_i = \frac{\dot{n}_i}{\sum_j n_j} - \frac{n_i \sum_j \dot{n}_j}{\left(\sum_j n_j\right)^2} = \frac{r a_i n_i}{\sum_j n_j} - \frac{n_i r \sum_j a_j n_j}{\left(\sum_j n_j\right)^2} = r f_i \left(a_i - \sum_{j=1}^{S} a_j f_j \right). \tag{4.5}$$

To find the equilibrium, we must find the solutions to $\dot{f}_i = 0$, i.e., the solutions of the equation

$$\mathbf{f} \odot (\mathbf{a} - (\mathbf{a} \cdot \mathbf{f})\,\mathbf{1}) = \mathbf{0}, \tag{4.6}$$

with conditions $f_i > 0 \; \forall i$ and $\sum_i f_i = 1$. Since, in general, coefficients a_i are all different from one another, the only non-trivial solution is $\mathbf{f} = \mathbf{e}_i$ for some i, meaning that at equilibrium only 1 species has survived.

It must be noted that this state cannot be actually reached, since $\dot{n}_i > 0$ and thus species can't actually go extinct. If we do not manually remove species under a certain threshold, we can only observe that every species' relative abundance, besides one, decreases and eventually becomes negligible; while the number of species in the system remains the same.

Of course, this steady state is approached only when the resource is present; if at time t^* the resource runs out, we have $\dot{n}_i = 0$, which means $n_i(t > t^*) = n_i(t^*)$. Since in our serial evolution scheme the resource abundance is replenished by the same amount after each cycle, the species dynamics simply continue from where they stopped, eventually reaching equilibrium after a certain number of cycles.

We can also describe the dynamics of g_i between cycles (we already know that in a cycle $g_i = \exp\{a_i\gamma\}$). Since the biomass of species can only grow, for each cycle we have $n_i^{k+1}(0) = n_i^k(T) > n_i^k(0)$, but since $\Delta c = -\sum_i \Delta n_i$ we have $\Delta n_i < c_0 \; \forall i$. Therefore, we can write

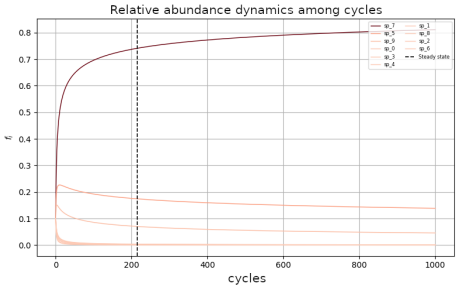
$$g_i^{k+1} = \frac{n_i^{k+1}(T)}{n_i^{k+1}(0)} = \frac{n_i^{k+1}(0) + \Delta n_i^{k+1}}{n_i^{k+1}(0)} = 1 + \frac{\Delta n_i^{k+1}}{n_i^{k+1}(0)} \leq 1 + \frac{c_0}{n_i^{k+1}(0)} \xrightarrow[k \to \infty]{} 1, \tag{4.7}$$

since the growth of species abundance is not bounded.

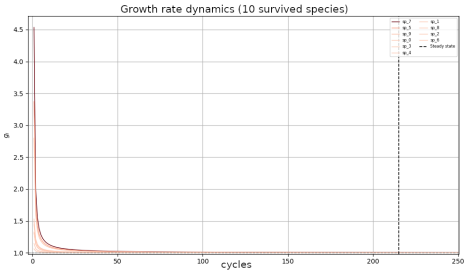
4.2 Simulation of multi-cycle dynamics with no sampling

We run a simulation with $S = 10$ species, $n_i(0) = 1$ and $c(0) = 10$, while the parameter $K = 0.05$ was fixed. We extracted the parameters a_i from a Lognormal distribution with $\mu = 2.5$ e $\sigma = 0.8$, .

We analyze the time series of f_i and g_i, considering their value at the end of each cycle, as well as the relationship between g_i and a_i:



(2a) Species relative abundances among cycles. It can be seen how one species dominates and the system approaches the predicted equilibrium.



(2b) Growth rate of species among cycles. Again, the trend follows the expected behavior.

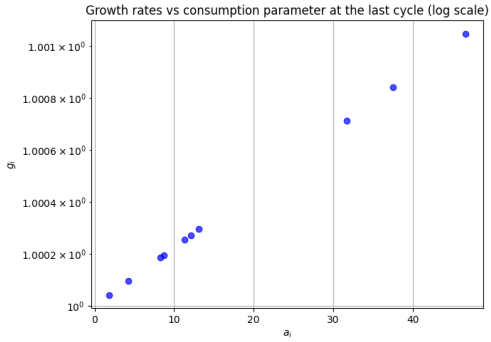


Figure 3: Plot of g_i as a function of a_i, in log scale. The linear trend is consistent with our hypothesis of $\log(g_i) \propto a_i$.

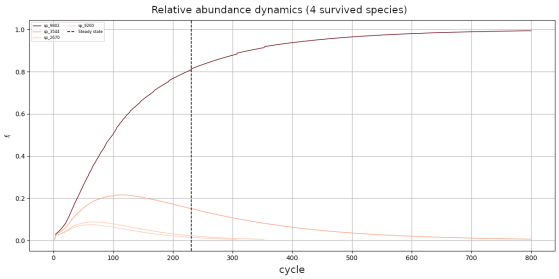
Figures (2a) and (2b) show that the system follows the predicted behavior; while in figure (3) we can see how the expected trend for g_i is followed, i.e. species with higher a_i are exponentially advantaged.

Let us now introduce the sampling process, which gradually excludes from the simulations the species whose relative abundance is smaller than the fixed threshold.

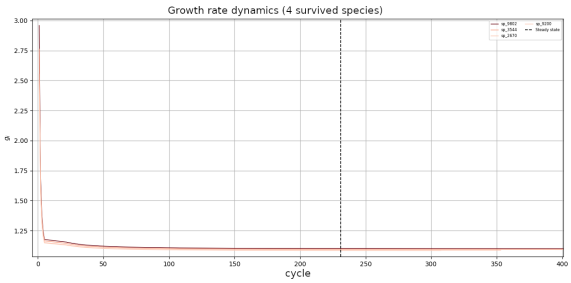
4.3 Simulation of multi-cycle dynamics with sampling

We now introduce the sampling process, as described in section (3.3), and run a simulation with $S = 10^4$ species, $c_0 = 10^4$, $K = 50$, $N_{tot} = 10^5$ and $n_i^0(0) = 10\ \forall i$.

From the properties of the model, we know that after a certain number of cycles, only one species survives, and the steady state is reached.



(4a) Species relative abundances time series. The vertical line shows the cycle at which the approximate steady state is reached, which in this case does not correspond to a true equilibrium.



(4b) Growth rates time series. This time, since the total amount of species is fixed at every cycle, the growth rate does not approach 1, but the only surviving species reach a value that depends on c_0 and N_{tot}.

As we can see in figure (4b), when introducing the sampling process g_i does not approach 1, but a higher value. This happens because the total abundance of the species at the beginning of every cycle is fixed, so the fraction $\Delta n_i^k / n_i^k(0)$ is not negligible at high k.

In fact, we can evaluate the new limit for g_i, which is reached when there is only one species left:

$$g_i^{end} = 1 + \frac{\Delta n_i}{n_i(0)} = 1 + \frac{c_0}{N_{tot}}, \tag{4.8}$$

which in our case corresponds to $g_i^{max} = 1 + \frac{10^4}{10^5} = 1.10$.

Also, from figure (5) we can see how the theoretical relationship between g_i and a_i is still respected.

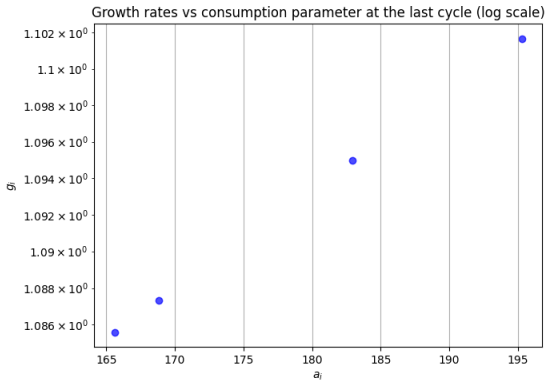


Figure 5: Plot of g_i vs a_i, at the approximate steady state. The trend still reflects the exponential advantage of species with greater a_i

5 Simulation of a system with self-interacting species

5.1 Model Properties

We now want to include in our model the interaction that a species has with itself, by modifying the equations in the following way:

$$\begin{aligned} \frac{dn_i}{dt} &= n_i\left(a_i r(c) - A_i n_i\right), \\ \frac{dc}{dt} &= -r(c)\sum_{i=1}^{S} a_i n_i, \end{aligned} \tag{5.1}$$

where A_i is the self-interaction coefficient.

The only equilibrium of the system is $(c = 0, \mathbf{n} = \mathbf{0})$, but for our purposes, it is interesting to study the equilibrium of f_i. To do that, we write the relative abundances dynamics equations:

$$\dot{f}_i = \frac{\dot{n}_i}{\sum_j n_j} - \frac{n_i \sum_j \dot{n}_j}{\left(\sum_j n_j\right)^2} = f_i(a_i r - n_i A_i) - \frac{f_i \sum_j n_j (a_j r - n_j A_j)}{\sum_j n_j} = f_i\left[a_i r - n_i A_i - \sum_j f_j (a_j r - n_j A_j)\right], \tag{5.2}$$

and

$$\dot{f}_i = 0 \iff f_i = 0 \vee a_i r - n_i A = \frac{\sum_j n_j (a_j r - A_j n_j)}{\sum_j n_j}. \tag{5.3}$$

Writing $R = \sum_j n_j (a_j r - A_j n_j)$ and $N = \sum_j n_j$, independent from i, we have

$$a_i r - A_i n_i = \frac{R}{N} = Q \ \forall i, \tag{5.4}$$

which means that $a_i r - A_i n_i$ has the same value for all species, at equilibrium. This leads to

$$n_i^* = \frac{a_i r - Q}{A_i} \implies f_i^* = \frac{a_i r - Q}{A_i N}. \tag{5.5}$$

We can now analyze two separate cases:

- If the abundance of the resource is much greater than that of the species, we can assume $r \approx$ cost.. Thus equation (5.1) gives us directly the non-trivial equilibrium $n_i^* = \frac{a_i r}{A_i}$, obtaining for f_i

$$f_i^* = \frac{\frac{a_i r}{A_i}}{\sum_j \frac{a_j r}{A_j}} = \frac{\frac{a_i}{A_i}}{\sum_j \frac{a_j}{A_j}}, \tag{5.6}$$

  which is independent of r(c(t)) and valid only while the resource is greatly abundant. Comparison of eq. (5.6) and eq. (5.1) shows that this equilibrium corresponds to $Q = 0$.

- If the resource runs out we have $r = 0$ and the equilibrium becomes

$$n_i^* = \frac{a_i r - Q}{A_i} = -\frac{Q}{A_i} \implies f_i^* = -\frac{Q}{A_i}\frac{1}{\sum_j \left(-\frac{Q}{A_j}\right)} = \frac{\frac{1}{A_i}}{\sum_j \frac{1}{A_j}}. \tag{5.7}$$

The same result can be derived from $\dot{n}_i = -A_i n_i^2$, by searching for an equilibrium for f_i.

It is easy to see that, in this environment, species do not naturally go extinct, and we need to again implement the sampling mechanism in order to mimic the selection process.

Since we defined our time-series equilibrium as $f_i^{k+1}(0) = f_i^k(0)$, we expect the growth rate to reach 1 both with and without the sampling mechanism:

$$g_i^k = \frac{f_i^k(T)}{f_i^k(0)} = \frac{f_i^{k+1}(0)}{f_i^k(0)} = 1. \tag{5.8}$$

Of course, the condition shown in eq. (3.12) is not true for all cycles when we implement the sampling mechanism, but becomes accurate once we approach the steady state (and species no longer go extinct). Thus, we can expect the growth rate to approach 1 more "smoothly" in the case without sampling, but nevertheless to reach it in both cases.

5.2 Simulation of multi-cycle dynamics with no sampling

We now run a simulation with $S = 10$, $c_0 = 10^4$, $K = 50$ and $n_i^0(0)$=1.

Since condition $c >> n_i$ is met, we expect the dynamics to reach the equilibrium described in eq. (5.6) and to stay there until the resource runs out. When this happens, the system approaches the second equilibrium for f_i, while the species abundances tend to 0. This behavior can be seen in Fig. (6).

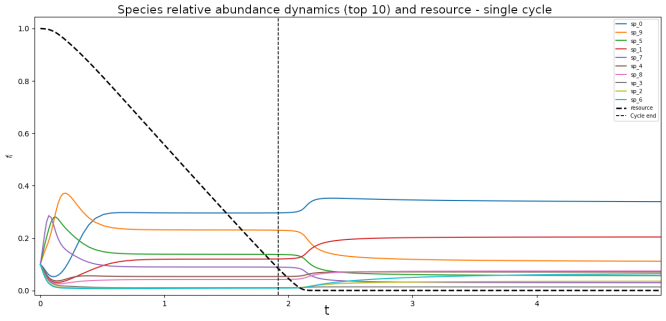
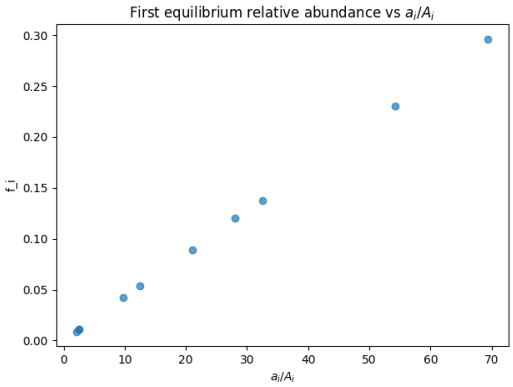
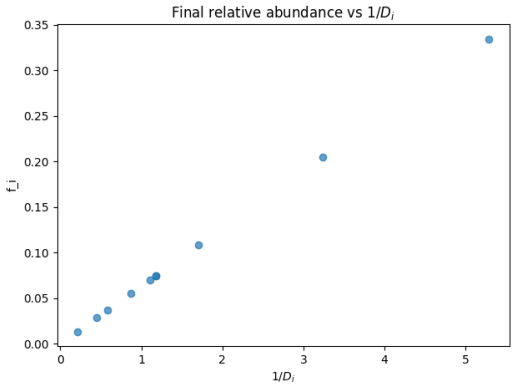


Figure 6: Simulation of a single cycle with large resource abundance. It can be seen how the system quickly reaches the first equilibrium, while the resource is still present, and then, once it runs out, it changes again, approaching the second one.

We can verify that the relative abundance of the species follows the two trends we expect by plotting them as functions of a_i/A_i when the resource is present and of $1/A_i$ when it is not:

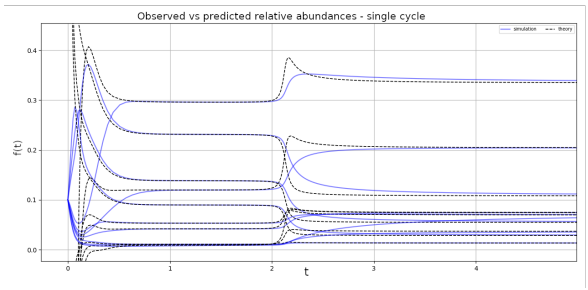


(7a) Relative abundances vs a_i/A_i, when the resource is present. The trend confirms our hypothesis for the first equilibrium.

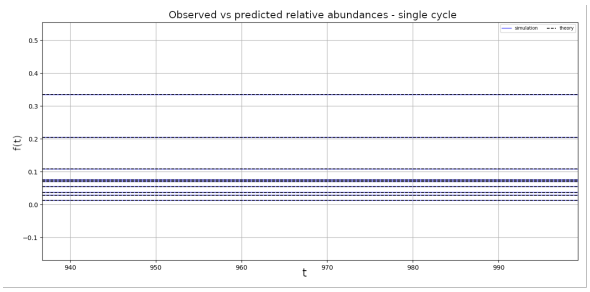


(7b) Relative abundances vs $1/A_i$, when the resource is exhausted. Again, the trend follows the expected one.

We then evaluated Q(t) and S(t) to plot the theoretical values of f_i, described in eq. (5.5). We expect the simulated values to match the predicted ones only at equilibrium, since the expression used is true only when $\dot{f}_i = 0$.

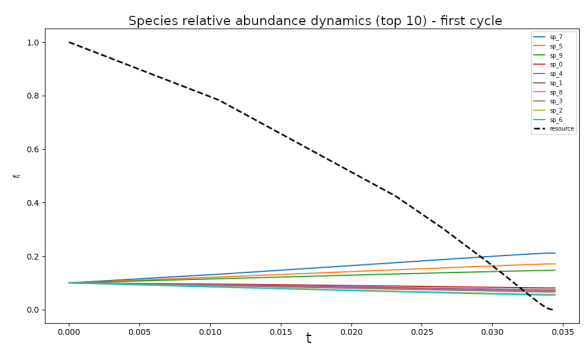


(8a) The graph shows the simulated (blue lines) vs theoretical (black dotted lines) relative abundances. As we can see, they match while the system is at equilibrium.

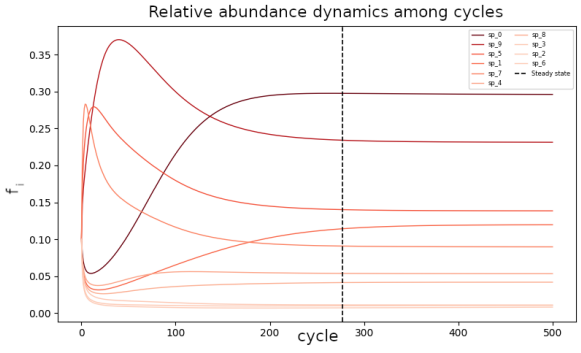


(8b) The two sets of curves match again only after a long time, when the system approaches the new equilibrium for f_i, while the absolute abundances tend to 0.

If c_0 is similar to the abundance of the species, the system does not reach equilibrium in one cycle. Since the resource is replenished at the beginning of every cycle and there is no sampling, the multi-cycle time series is exactly the same as the previous one, and the steady state is reached after a finite number of cycles.



(9a) Relative abundance dynamics of the first cycle. The resource starts at $c_0 = 10$ and runs out before the species can reach the steady state.



(9b) By replenishing the resource to its original value, the dynamics start again and the species evolve towards the steady state, eventually reaching it after a finite number of cycles.

This can also be seen by plotting f_i^{eq} vs a_i/A_i, which is identical to the previous trend and matches the theoretical prediction.

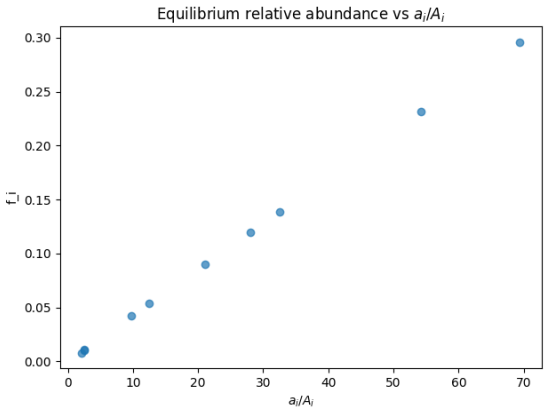
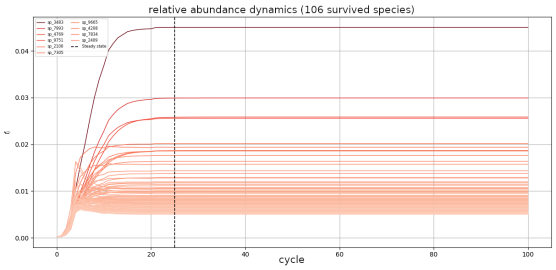


Figure 10: f_i vs a_i/A_i at equilibrium. The relative abundances follow the theoretical trend, and the equilibrium is equivalent to the one obtained in a single cycle with plenty of resources.

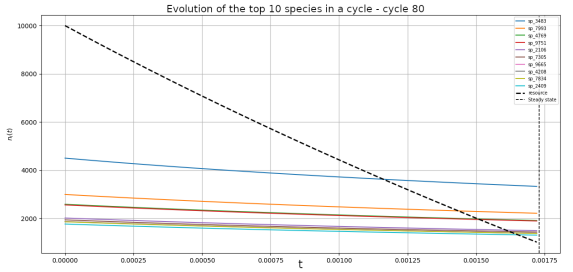
5.3 Simulation of multi-cycle dynamics with sampling

If we now introduce the sampling mechanism, the number of species will decrease over time. When the remaining species all have their relative abundance greater than the threshold, they will no longer go extinct, and we expect the system to evolve towards the usual equilibrium.

We run a simulation with $S = 10^4$, $c_0 = 10^4$, $K = 50$, $n_i^0(0) = 10$, and the dynamical threshold described by eq. (3.15).



(11a) Relative abundance dynamics between cycles, for the survived species. The steady state is reached after 24 cycles, and it holds indefinitely.

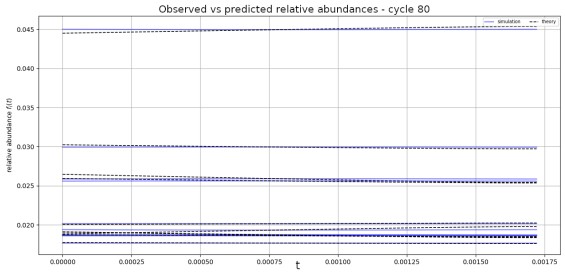


(11b) Species dynamics for cycle 80, well after the steady state is reached. The species abundance varies over time, but the relative abundance is constant, giving the equilibrium.

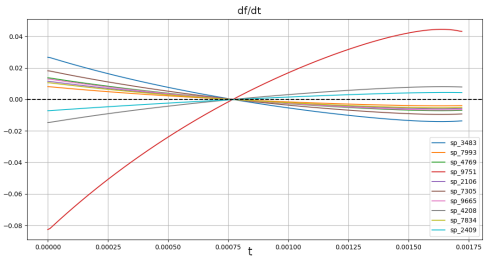
From figure (11a) it can be seen that the surviving species quickly reach the steady state for f_i, while the single-cycle dynamics for n_i are not in equilibrium.

We would expect that the single-cycle dynamics for f_i were also at equilibrium, since for our time series to be considered in the steady state, we need $f_i^{k+1}(0) = f_i^k(T) = f_i^k(0)$, but we find that actually this is not the case.

From figure (12a), it can be seen that the calculated and predicted relative abundances are not the same, suggesting that the equations do not satisfy $\dot{f}_i = 0$. This can also be seen in figure (12b), where we computed df_i/dt for the top 10 survived species: since they are not identically zero during the cycle, the system is not at equilibrium.



(12a) Observed vs theoretical relative abundances after the steady state reached. From the fact that they do not overlap, we can conclude that the system is not at equilibrium.

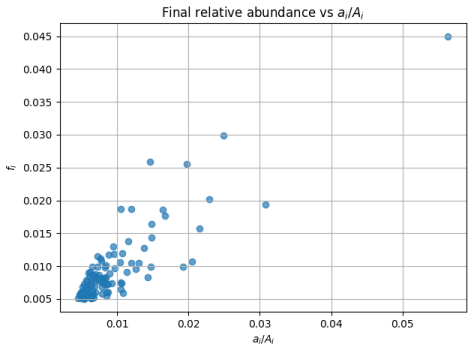


(12b) Computing the derivatives of the relative abundances, we see that they are not identically zero, thus the system is not at equilibrium.

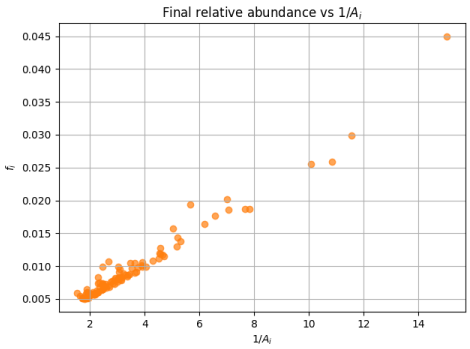
What happens is that, since the total abundance at the beginning of every cycle is fixed, with fewer species remaining, the abundance of a single species increases, and the hypothesis $c_0 \gg n_i(t)$ does not hold.

What we actually see is that the system lies in the very slow transition state between the two equilibria, but the absolute change in relative abundance Δf_i is so small (order of $\sim 10^{-6} - 10^{-7}$ for the top 10 species) that it is neglected in the sampling process, leading to the steady state we observe for the multi-cycle dynamics.

Therefore, by looking at the distribution of f_i at the steady state, we find that it is different from the ones described by eq. (5.6) and eq. (5.7), with the second being more similar to the observed trend:



(13a) f_i vs a_i/A_i at equilibrium. It can be seen that this "series equilibrium" is different from the expected dynamics equilibrium.



(13b) f_i vs $1/A_i$ at equilibrium. Again, this is different from the expected dynamics equilibrium with no resource, but in this case the trend is closer to the theoretical one.

6 Simulation of a system with self and mutual interactions between species

6.1 Model properties

We now discuss the complete case, where we also consider mutual interactions between different species. The equations become:

$$\begin{aligned} \frac{dn_i}{dt} &= n_i \left(a_i r - \sum_{j=1}^{S} A_{ij} n_j \right) \\ \frac{dc}{dt} &= -r \sum_{i=1}^{S} a_i n_i \end{aligned} \tag{6.1}$$

where A_{ij} encodes the strength of the interaction of the i species with the jth one.

Again, the only equilibrium of eq.(6.1) is ($c = 0$, $\mathbf{n} = \mathbf{0}$), but we are interested in studying the case where the resource does not run out, allowing us to assume $r = cost$ and leading to a non-trivial equilibrium for our system.

To study this problem, we must solve the equation

$$\mathbf{n} \odot (\mathbf{a}\, r - A\,\mathbf{n}) = \mathbf{0}, \tag{6.2}$$

which is analogous to solving the corresponding Linear Complementarity Problem (LCP), which can be expressed as

$$\mathbf{n} \geq \mathbf{0},\ \ \lambda := A\mathbf{n} - r\,\mathbf{a} \geq \mathbf{0},\ \ n_i \lambda_i = 0\ \forall i. \tag{6.3}$$

This problem can have an internal or a boundary solution:

- The internal solution corresponds to the equilibrium in which all species survive. This can be written as

  $$\mathbf{n} > \mathbf{0},\ \ \lambda = \mathbf{0} \implies \mathbf{n}^* = rA^{-1}\mathbf{a}. \tag{6.4}$$

  In practice, we can evaluate $\mathbf{n}^*$ and check if it follows the condition $\mathbf{n} > \mathbf{0}$; if this is the case, every species survives and $\mathbf{n}^*$ gives us the equilibrium abundances.

- The boundary solution corresponds to the equilibrium in which only a subset of species survives. This means that for species i there are two possibilities:

  - if the species survives, $n_i > 0$ and $\lambda_i = 0 \implies (An)_i = ra_i$.
  - if the species goes extinct, $n_i = 0$ and $\lambda_i > 0 \implies (An)_i > ra_i$.

  What this means is that, in general, of the S species, only a subset $\tilde{S}$ survives at equilibrium, and their abundance is given by

  $$\tilde{\mathbf{n}}^* = r\tilde{A}^{-1}\tilde{\mathbf{a}} \tag{6.5}$$

  where $\tilde{\mathbf{n}}^*$, $\tilde{A}^{-1}$ and $\tilde{\mathbf{a}}$ indicate the vectors with the sole indices corresponding to the surviving species.

Solving the LCP would mean checking every subset $\tilde{S}$ of species, by computing $\tilde{\mathbf{n}}^*$ and making sure that $n_i^* > 0\ \forall i$. Since the number of subsets is 2^S, a more efficient way is to numerically find $\mathbf{n}$ that minimizes

$$\mathbf{n} \odot (\mathbf{a}\, r - A\,\mathbf{n}) \geq \mathbf{0}. \tag{6.6}$$

With this method, we can predict which species will survive at equilibrium and even their abundances, which we can then compare with the results obtained by the simulation. Of course, it is not guaranteed that the system has only one equilibrium (it would be if the matrix were positive symmetric or a P-matrix, which is not the case for our system), so solving the LCP could give the wrong one. Despite that, we can always evaluate $\tilde{\mathbf{n}}^*$ with eq. (6.5) and compare it with the numerical solution.

This discussion is sensible only assuming the resource to be constant; since this is not the case in our model, the equilibrium will not be reached in one cycle, and we must rely on the multi-cycles time series.

As we did previously, we now focus on the relative abundance dynamics:

$$\dot{f}_i = \frac{\dot{n}_i}{\sum_j n_j} - \frac{n_i \sum_j \dot{n}_j}{\left(\sum_j n_j\right)^2} = f_i \left[\left(a_i r - \sum_{j=1}^{S} A_{ij} n_j \right) - \sum_{k=1}^{S} f_k \left(a_k r - \sum_{j=1}^{S} A_{kj} n_j \right) \right]. \tag{6.7}$$

We can now define the instant growth and its weighted average:

$$h_i(n) = a_i r - \sum_{j=1}^{S} A_{ij} n_j\,; \tag{6.8}$$

$$\bar{h} = \sum_{k=1}^{S} f_k h_k \tag{6.9}$$

and we get:

$$\dot{f}_i = f_i(h_i - \bar{h}) = 0 \iff f_i = 0 \vee h_i = \bar{h} \ \forall i \tag{6.10}$$

So, for any number of surviving species, $\bar{h}$ has the same value and the species are at equilibrium for f_i when all their h_i match that value. We can then find an expression for f_i of the survived species as a function of parameter $\bar{h}$:

$$h_i = \bar{h} = a_i r - \sum_{j=1}^{S} A_{ij} n_j \iff \sum_{j=1}^{S} A_{ij} n_j = a_i r - \bar{h} \implies \tilde{A}\tilde{\mathbf{n}} = \tilde{\mathbf{a}} r - \bar{h}\mathbf{1}, \tag{6.11}$$

where again the tilde is used to refer to the vectors and matrix whose elements are only of the surviving species. Therefore, we can write

$$\tilde{\mathbf{n}} = \tilde{A}^{-1}(\tilde{\mathbf{a}} r - \bar{h}\mathbf{1}) \implies \tilde{\mathbf{f}} = \frac{\tilde{A}^{-1}(\tilde{\mathbf{a}} r - \bar{h}\mathbf{1})}{\mathbf{1}^\top \tilde{A}^{-1}(\tilde{\mathbf{a}} r - \bar{h}\mathbf{1})} \tag{6.12}$$

Similarly to the case discussed in section (5.1), while the resource is present, the system will tend to the equilibrium for n_i we found before, which corresponds to $\bar{h} = 0$:

$$\tilde{\mathbf{f}} = \frac{\tilde{A}^{-1}\tilde{\mathbf{a}} r}{\mathbf{1}^\top \tilde{A}^{-1}\tilde{\mathbf{a}} r} = \frac{\tilde{A}^{-1}\tilde{\mathbf{a}}}{\mathbf{1}^\top \tilde{A}^{-1}\tilde{\mathbf{a}}} \implies f_i = \frac{\sum_{j\in I} \tilde{A}^{-1}_{ij} \tilde{a}_j}{\sum_{k,j\in I} \tilde{A}^{-1}_{kj} \tilde{a}_j} \tag{6.13}$$

With this reasoning, we can also find the equilibrium when the resource is exhausted, using $r = 0$:

$$\tilde{\mathbf{f}} = \frac{-\bar{h}\tilde{A}^{-1}\mathbf{1}}{-\bar{h}\mathbf{1}^\top \tilde{A}^{-1}\mathbf{1}} = \frac{\tilde{A}^{-1}\mathbf{1}}{\mathbf{1}^\top \tilde{A}^{-1}\mathbf{1}} \implies f_i = \frac{\sum_{j\in I} \tilde{A}^{-1}_{ij}}{\sum_{k,j\in I} \tilde{A}^{-1}_{kj}} \tag{6.14}$$

We can also notice how eq. (6.13) and (6.14) reduce to eq. (5.6) and (5.7) respectively, if we consider the matrix A to be diagonal.

In this environment, we saw that species can naturally go extinct, thus we do not need to artificially select a surviving threshold to discriminate between surviving and extinct species. However, we must remember that, since solutions of our ODE are locally unique, they cannot actually coincide with the equilibrium at a finite time. Extinct species will therefore have extremely small abundances, which we have to manually set to 0.

We can then expect the dynamics to be identical between the case of a single cycle with great resource abundance and the case of serial cycles with resource replenishment.

A different behavior can arise only when we fix the initial abundance of every cycle, as we have seen in the previous section.

We again expect the growth rate of the survived species to approach 1, since again their abundances evolve to a state where $\dot{n}_i \approx 0$ and therefore $n_i(T) \approx n_i(0)$.

6.2 Simulation of single cycle dynamics

To test our hypothesis for equilibrium, we first simulate the dynamics in a single cycle with $dc/dt = 0$, so that the resource is constant and we can easily compare the results with the predictions of the model.

We run a simulation with $S = 10^2$, $n_i(0) = 1$, $c_0 = 10^4$, and $K = 50$.

Since, in general, mutual interactions have different strengths compared to self-interactions, we can explore different scenarios by modifying the ratio of inter- to intra-species interaction coefficients, $\rho_{ij} = A_{ij}/A_{ii}$, as done in Gupta et al., 2021 [5]. The off-diagonal entries of the matrix A are extracted from the same distribution as the diagonal ones and then multiplied by the chosen factor.

We expect the dynamics to approach the one described in the previous chapter as the off-diagonal entries of A decrease, and to achieve maximum competition when they are of the same magnitude.

With a similar reasoning, we can suppose that a given species will not interact with every other one, but rather with a small fraction of the total. If we think about the DNAi experiment, it is realistic to consider

mutual interaction an uncommon characteristic, and therefore, we need to limit the number of interactions a species can have.

To reproduce this behavior in our model, we introduce the sparsity p in the matrix A, which means that each A_{ij} is set to 0 with probability p. For example, choosing $p = 0.1$, the number of interactions for a single species is $\sim pS = 10$.

This process is applied only to off-diagonal entries, since it is plausible that every species experiences self-limitation through their self-interaction. Since the more sparse the matrix, the more it tends to the diagonal case, we can expect the number of surviving species at equilibrium to grow as p decreases.

We now run multiple simulations and analyze the dependence of the fraction of surviving species at equilibrium with the two aforementioned parameters.

We expect the system to reproduce the results of the case discussed in section (5), where all species survive, when the ratio and the density are smaller. Similarly, we can predict that only a few species will survive if the interactions are all of the same magnitude.

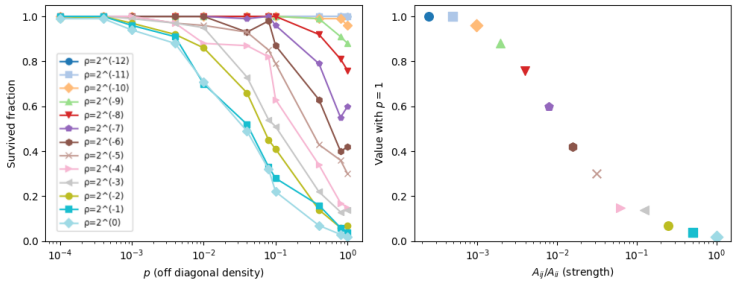
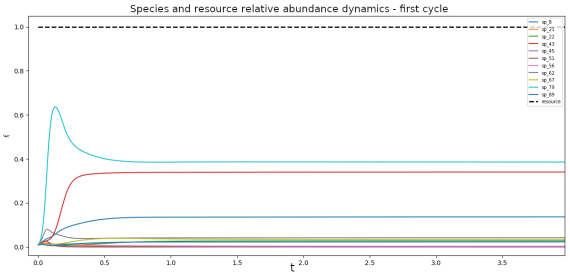


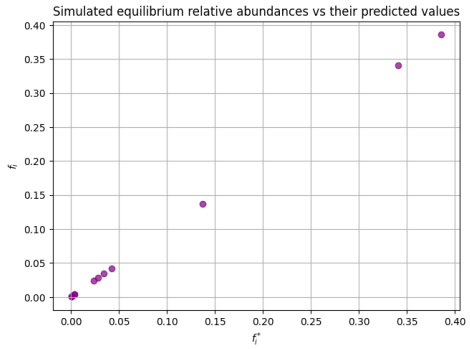
Figure 14: On the left panel, every curve corresponds to a fixed value of ρ_{ij}. When the density is small, every species survives regardless of the interaction ratio. As it grows, the fraction of surviving species is higher for smaller values of the interaction ratio.
On the right panel, we can see the survival fraction for $p = 1$, as a function of the interaction ratio. The trend shows that species coexistence is favored when mutual interactions are weaker.

The result shown in figure (14) is consistent with our prediction and with the results obtained in Gupta et al., 2021, where a similar dependence was found.

We now focus on a single simulation, with fixed values of p and ρ_{ij}. In every simulation, we numerically solved the LCP to find the expected equilibrium and then compared it with the results obtained.

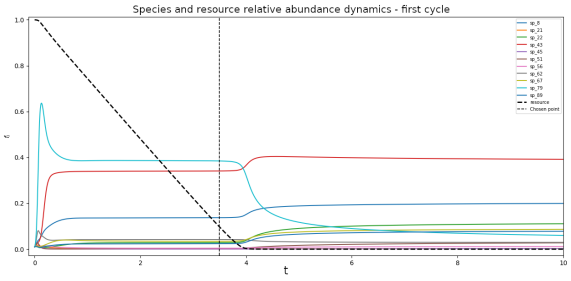


(15a) Evolution of species relative abundances over time, with a fixed amount of resource. The system quickly reaches the equilibrium, where the majority of consumers naturally become extinct. The simulation is run with $\rho_{ij} = 2^{-4}$ and $p = 1$.

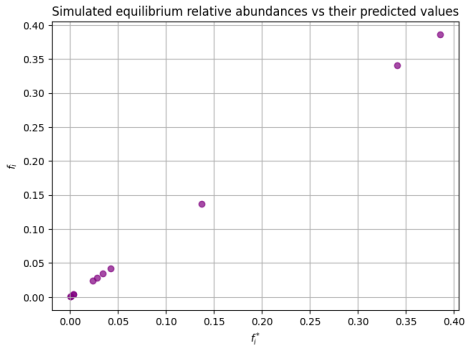


(15b) Plot of simulated versus predicted relative abundances, at equilibrium. The system reaches the state predicted by the LCP.

If we now consider a single-cycle with resource depletion, we expect the species to reach the usual equilibrium, and then, when the resource runs out, to approach the second one, described by eq.(6.14).



(16a) Evolution of species relative abundances over time. The system quickly reaches the same equilibrium as before, and then tends towards the second one when the resource runs out. The simulation is run with $\rho_{ij} = 2^{-4}$ and $p = 1$.



(16b) Plot of simulated versus predicted relative abundances, at equilibrium. In this case, the values of f_i are chosen when the resource is 10% of the maximum, so we are sure that the system is still at equilibrium. Again, the system reaches the state predicted by the LCP.

Figures (16a) and (16b) show how the simulation gives the expected result, where the equilibrium reached when the resource is present follows again the prediction of the LCP.

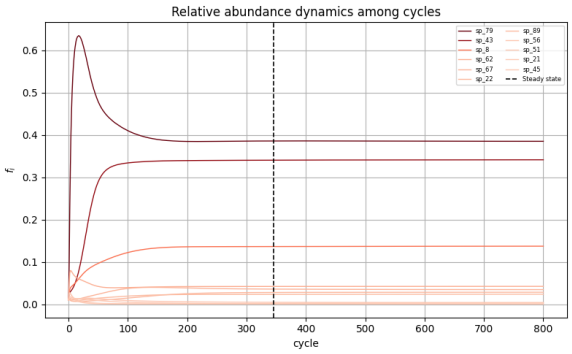
6.3 Simulation of multi-cycle dynamics with no sampling

If we now limit the amount of initial resources, the system will not reach its steady state in one cycle. In this case, the abundances of surviving species at the end of a cycle are not rescaled with a fixed N_{tot}, so we simply have

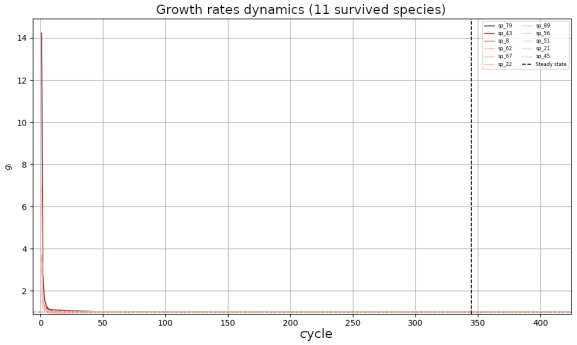
$$\begin{cases} n^{k+1}_{i,surv}(0) = n^{k}_{i,surv}(T) \\ n^{k+1}_{i,dead}(0) = 0 \end{cases}, \tag{6.15}$$

where a species is considered extinct if, at the end of a cycle, its abundance is numerically equivalent to zero.

The simulation is performed with $c_0 = 10^2$ and $K = 0.5$, $S = 10^2$ and $n^0_i(0) = 1$.



(17a) Evolution of species relative abundances among cycles. The system reaches the same equilibrium as the single cycle dynamics with high resource abundance. The simulation is run with $\rho_{ij} = 2^{-4}$ and $p = 1$.



(17b) Growth rate time series over cycles. Similarly to the other simulations, since species abundance approaches an equilibrium, the growth rate of surviving species approaches 1.

It can be seen that the dynamics between cycles reproduce the one in a single cycle, and the equilibrium reached is the same. Again, this equilibrium corresponds to the one predicted by the LCP.

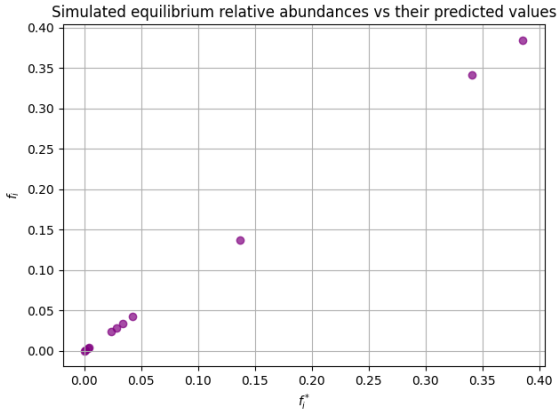


Figure 18: Species relative abundances at the steady state versus the predicted ones. The plot is identical to the one in figure (1616b).

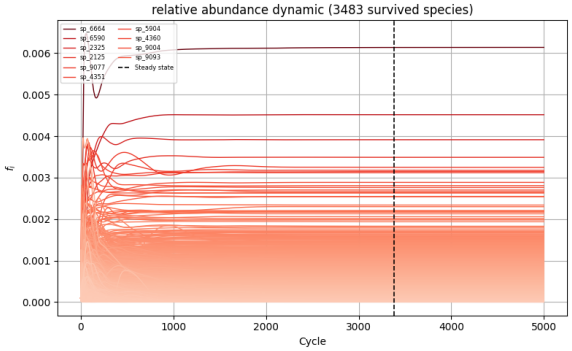
6.4 Simulation of multi-cycle dynamics with sampling

6.4.1 Sampling with threshold

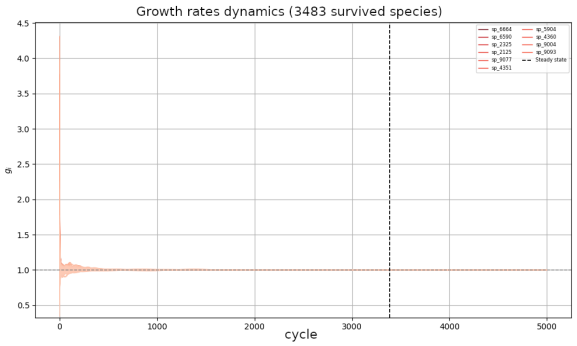
We now simulate the system dynamics with $S = 10^4$, $n_i^0(0) = 10$, $c_0 = 10^4$ and $K = 50$.

Since the number of species in the system is high, we can expect that a given species will not interact with every other one, and therefore we can assume that the matrix has only a small fraction of its entries $\neq 0$.

We first run a simulation with $p = 0.01$ and $\rho_{ij} = 2^{-4}$, which highlighted some difficulties.



(19a) Evolution of species relative abundances among cycles. After 3383 cycles the steady state is reached, but this does not correspond to a true equilibrium for the system.



(19b) Growth rate time series over cycles. We can see how, for very species, the growth rate approaches 1, indicating that the dynamics slowed.

Because the number of surviving species is high, the dynamics near equilibrium are extremely slow, but the system is not stationary. This can be deduced from the fact that, despite the steady state being considered reached, a few species still go extinct in the following hundreds of cycles.

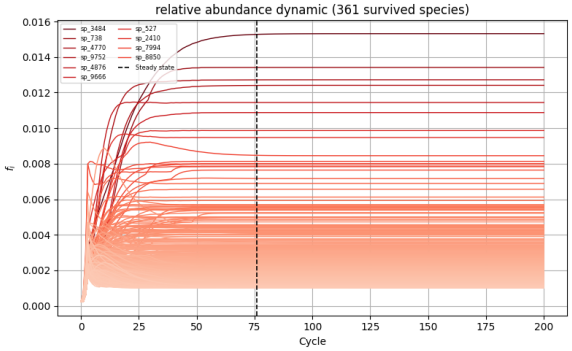
This also means that, at the approximate steady state, species are present with extremely small relative abundances, enough, for example, to be considered extinct if we adopt the criteria discussed in Section (5.3).

In that case, we decided to use the sampling process to characterize the cycle dynamics and overcome the absence of natural extinction of species, so we can simply introduce it in this system and study its effects alongside the extinction process that derives from the equations.

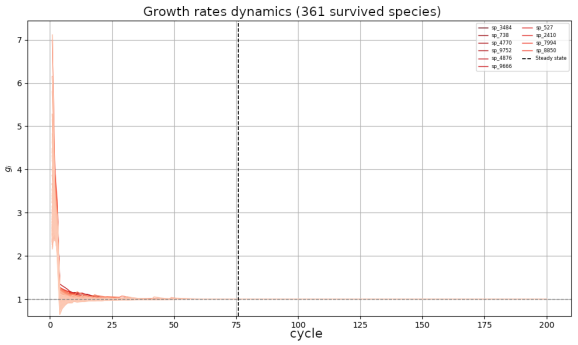
This also means that the theoretical prediction of the LCP cannot describe our new multi-cycle equilibrium, nor can we hope that this new state will be described by eq. (6.12) and its subcases, since it is not guaranteed that, for a generic subset of surviving species, condition

$$\tilde{A}^{-1}\tilde{\mathbf{a}} > 0 \tag{6.16}$$

will be met.
We now analyze the results of the simulations where the sampling process is described by eq. (3.15) is used, with $\alpha_{max} = 0.001$.



(20a) Evolution of species relative abundances among cycles. After 75 cycles the steady state is reached, and the selected species have constant relative abundance. The simulation is performed with $p = 0.01$ and $\rho_{ij} = 2^{-4}$



(20b) Growth rate time series over cycles. Every surviving species' growth rate reaches 1 when the dynamics reach the equilibrium.

It is easy to see that by using the sampling mechanism, the number of surviving species is drastically reduced and a true steady state is reached in a limited number of cycles.

We can now explore the dependence of the fraction of surviving species at equilibrium with p and ρ_{ij}, as we did in the previous section.

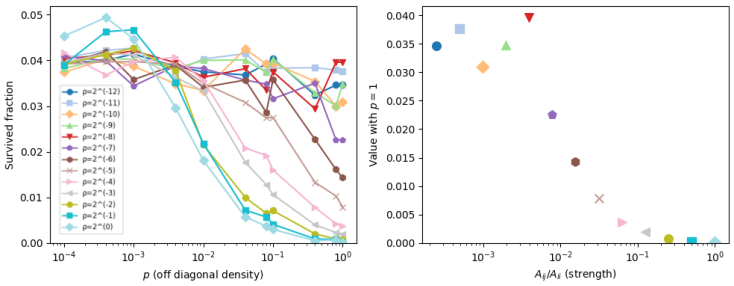


Figure 21: On the left panel, every curve corresponds to a fixed value of ρ_{ij}; on the right panel we can see the survival fraction for $p = 1$, as a function of the interaction ratio. This time, the presence of the sampling process affects the dynamics, specifically when the system is similar to the diagonal case.
The trend is identical to the one in figure (14) for $\rho_{ij} \gtrapprox 10^{-2}$.

From figure (21) we can see that, for parameters that allow few species to survive, the trend is identical to the one obtained without the sampling process. With a choice of parameters that would, without the sampling process, give a survival fraction close to 1, the system behaves differently.

This is consistent with the results we observed for $p = 0.01$ and $\rho_{ij} = 2^{-4}$, where the number of survived species with the sampling process is 10 times smaller than that without it.

6.4.2 Stochastic multinomial sampling

Another way of implementing the sampling process, which is closer to the real case, is to randomly choose N_{tot} individuals, each with probability f_i of belonging to the species i.

Formally, at the end of each cycle, the S' surviving species are present with a relative abundance $f_i(T)$, with $\sum_j f_j = 1$. Then, the number of individuals assigned to the species is randomly extracted, following the multinomial distribution:

$$P(n_1, n_2, \dots, n_{S'}) = \frac{N_{tot}!}{n_1! n_2! \dots n_{S'}!} \prod_{i=1}^{S'} f_i^{n_i}. \tag{6.17}$$

This process is equivalent to assigning every individual to a species with probability f_i, independently of each other, and it adds stochastic fluctuations to the species dynamics, which are smaller the greater N_{tot} is. The expected value for each species is

$$\mathbb{E}[n_i] = N_{tot} f_i, \tag{6.18}$$

with the following variance and covariance (for $i \neq j$):

$$\begin{aligned} \mathrm{Var}(n_i) &= N_{tot} f_i (1 - f_i) \\ \mathrm{Cov}(n_i, n_j) &= -N_{tot} f_i f_j. \end{aligned} \tag{6.19}$$

Adding stochasticity to the system, we observe different behaviors depending on the number of surviving species, and therefore on different values of p and ρ_{ij}.

If competition is strong, and only a small fraction of the initial species survives, their abundance is relatively high and the system reaches a stochastic steady state: species abundance has a constant mean, and the number of species in the system is fixed.

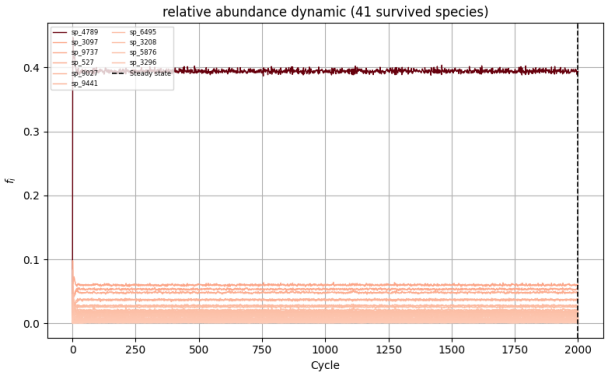
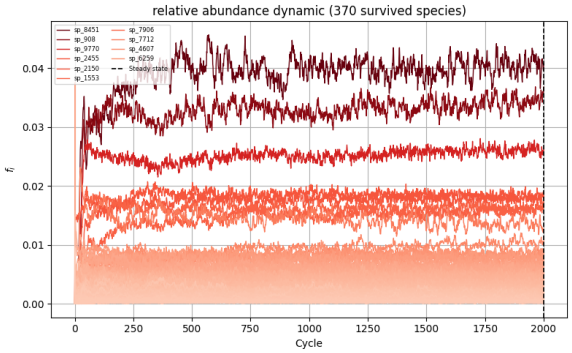
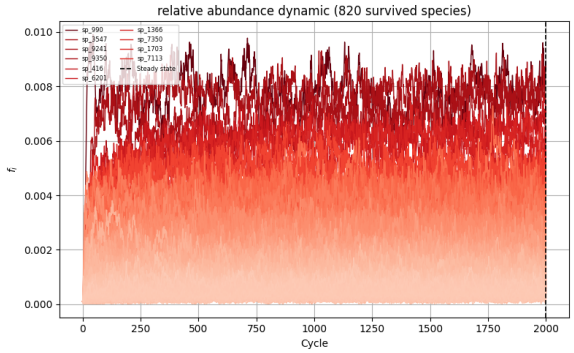


Figure 22: Species relative abundance over time, with stochastic sampling. The simulation is performed with $p = 1$ and $\rho_{ij} = 2^{-4}$, so the number of surviving species is very low.

When the number of species increases, their abundance decreases, and the stochastic noise becomes more relevant.



(23a) Species relative abundance over time, with stochastic sampling. The simulation is performed with $p = 0.1$ and $\rho_{ij} = 2^{-4}$, and the number of surviving species is therefore higher.



(23b) Species relative abundances for $p = 0.01$ and $\rho_{ij} = 2^{-4}$. This time, the number of surviving species is even higher, and the stochastic oscillations are extremely high compared to the abundances.

In this case, the probability for a species with low relative abundance to go extinct because of the casual oscillations is nonzero and becomes higher for even more species.

We can still explore the dependence between the number of surviving species and the parameters p and ρ_{ij}, by running a set of simulations for a great number of cycles (2000), so we can analyze the system after a long time.

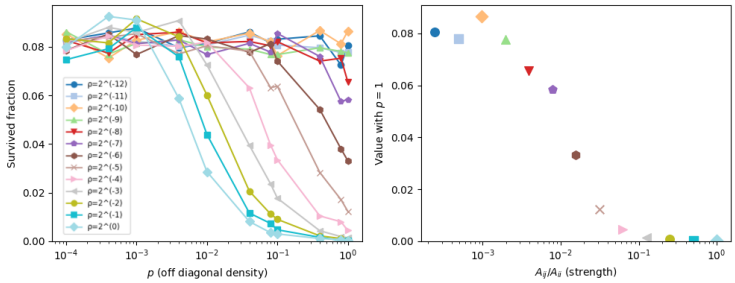


Figure 24: On the left panel, every curve corresponds to a fixed value of ρ_{ij}; on the right panel, we can see the survival fraction for $p = 1$, as a function of the interaction ratio.
With the stochastic sampling, the trend appears more similar to figure (14), only saturating at $F = 0.08$ instead of $F = 1$, while in the previous case, the expected dependence was lost when the sampling process was predominant.

We can see how the dependence observed in the case without sampling is reproduced, this time even better than the one with the threshold sampling.

6.4.3 Stochastic process and species abundance distribution

As we have seen in the previous section, by introducing stochastic multinomial sampling, the system does not approach the deterministic equilibrium predicted by the LCP. Instead, even species that should survive can, if their relative abundance is small enough, go extinct because of the casual oscillations.

At the end of every cycle, the probability that species i will be sampled to zero is

$$\mathbb{P}(n_i = 0) = (1 - f_i(T))^{N_{tot}}. \tag{6.20}$$

Thus, the probability of being present in the system after k cycles is

$$P_i(k) = \prod_{\eta=1}^{k} (1 - (1 - f_i^{\eta}(T))^{N_{tot}}), \tag{6.21}$$

where $f_i^{\eta}(T)$ is the relative abundance of species i, at the end of cycle η but before sampling.

In order to compare the predicted and observed species abundance distributions (SAD), we need to evaluate the solution to the LCP before running the simulation. Therefore, we have to limit the number of species in the system since the solver used to find the solution cannot work with large matrices.

We decided to settle at $S = 500$ and $n_i^0(0) = 10$, so $N_{tot} = 5000$, $c_0 = 500$ and $K = 2.5$. We run a simulation with $p = 1$ and $\rho_{ij} = 2^{-6}$, analyzing the SAD at $k = 500$.

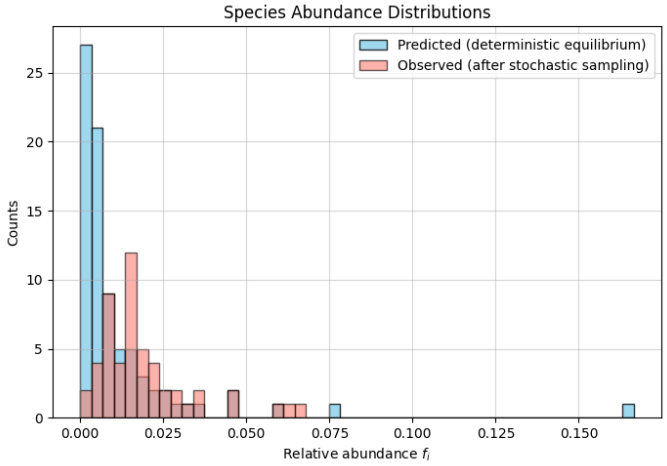


Figure 25: Histograms for the theoretical and observed species abundance distribution. After the sampling, the species with lower f_i are more probable to go extinct, and therefore the two distributions are radically different.

More useful is to compute the predicted number of surviving species after the stochastic sampling and compare it with the actual number of species present in the system.

We define the indicator function that tells us whether the species is present or not at cycle k:

$$X_i(k) = \begin{cases} 1, & \text{if } n_i^k(T) > 0 \\ 0, & \text{if } n_i^k(T) = 0 \end{cases} \tag{6.22}$$

so the number of species in the system will be

$$S_{obs}(k) = \sum_{i=1}^{S} X_i(k). \tag{6.23}$$

To give an estimate of this quantity, we can write

$$\begin{aligned} \mathbb{E}[S_{obs}(k)] &= \sum_{i=1}^{S} \mathbb{E}[X_i(k)] = \sum_{i=1}^{S} P_i(k) \\ &= \sum_{i=1}^{S} \prod_{\eta=1}^{k} (1 - (1 - f_i^{\eta}(T))^{N_{tot}}). \end{aligned} \tag{6.24}$$

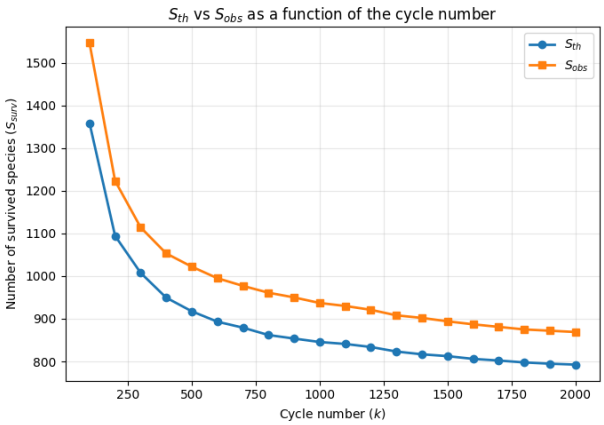
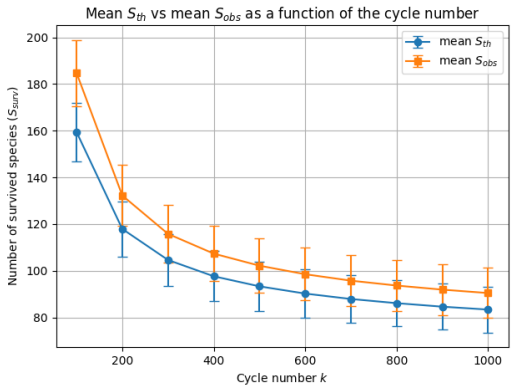


Figure 26: Theoretical and simulated number of survived species, among cycles. The two trends are identical, but the expected number is lower.

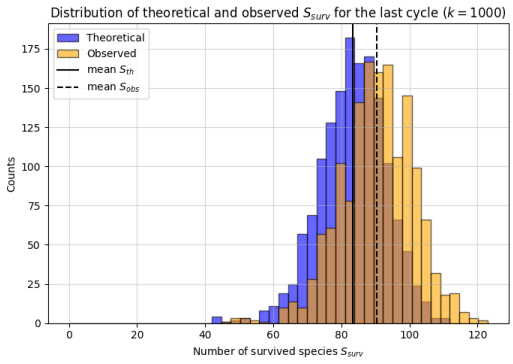
We now plot the expected and observed number of surviving species as a function of the cycle number. From figure (26), we can see that the two quantities have the same dependence, but the expected value is systematically lower. This happens because eq. (6.24) is true for the expected value of the observed number, not for a single realization.

We thus expect that if we run multiple realizations of the same system (same number of species and same parameters), the two trends will eventually coincide, as we can see in figures (27a) and (27b).

These iterated simulations were carried out with $S = 1000$, $n_i^0(0) = 10$, $c_0 = 1000$, $K = 5$, $\rho_{ij} = 2^{-6}$ and $p = 0.1$, for $k = 1000$ cycles and $R = 1500$ times.



(27a) Theoretical and simulated mean number of survived species among cycles, with their error bar. The two trends are the result of $R = 1500$ iterations of the same set of parameters. It can be seen that the two are compatible and, as the number of iterations increases, they get closer.



(27b) Distribution of predicted and observed number of survived species at cycle $k = 1000$, among different iterations. We can see how the two distributions largely overlap, and we can predict they will coincide for $R \to \infty$.

7 Conclusion and further developments

In this work, we proposed a theoretical model, based on the consumer-resource and the generalized Lotka-Volterra frameworks, to describe the evolution of a particular DNA ecosystem. After studying the properties of such a model and its equilibrium, focusing on the conditions needed to achieve species coexistence, we developed a serial approach and investigated the behavior of this multi-cycle dynamics.

Whether the dynamics were purely deterministic or combined with the stochastic sampling, we found that species coexistence is favored when inter-species interactions are weaker. We also found that the multi-cycle equilibrium can be different from the theoretical one, depending on the number of surviving species and their abundances compared to the resource, and that the dynamics is faster when a threshold-based sampling is used, since species are eliminated from the system at a higher rate.

Finally, we analyzed how stochastic multinomial sampling influences the species abundance distribution, with the less abundant species disappearing from the system as the dynamics evolve; and we found a correct way to estimate the number of surviving species after k cycles.

To further improve the model and its predictions, a direct comparison with the experimental results could be carried out to extrapolate the right parameters and the functional relationships of species interactions. Also, by analytically computing the deterministic SAD, we could combine it with the survival probability of stochastic sampling to evaluate the real SAD at a given cycle.

References

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

[2] L. Casiraghi, F. Mambretti, A. Tovo, E. M. Paraboschi, S. Suweis, and T. Bellini. Synthetic eco-evolutionary dynamics in simple molecular environment. *Elife*, 12:RP90156, 2024.

[3] P. Catalán, C. F. Arias, J. A. Cuesta, and S. Manrubia. Adaptive multiscapes: an up-to-date metaphor to visualize molecular adaptation. *Biology Direct*, 12(1):7, 2017.

[4] J. M. Chase and M. A. Leibold. *Ecological niches: linking classical and contemporary approaches*. University of Chicago Press, 2009.

[5] D. Gupta, S. Garlaschi, S. Suweis, S. Azaele, and A. Maritan. Effective resource competition model for species coexistence. *Physical review letters*, 127(20):208101, 2021.

[6] G. Hardin. The competitive exclusion principle: an idea that took a century to be born has implications in ecology, economics, and genetics. *science*, 131(3409):1292–1297, 1960.

[7] J. Hofbauer and K. Sigmund. *Evolutionary games and population dynamics*. Cambridge university press, 1998.

[8] G. E. Hutchinson. The paradox of the plankton. *The American Naturalist*, 95(882):137–145, 1961.

[9] W. KINDS. Homage to santa rosalia. *The American Naturalist*, 93(870):145–159, 1959.

[10] O. Levenspiel. The monod equation: a revisit and a generalization to product inhibition situations. *Biotechnology and bioengineering*, 22(8):1671–1687, 1980.

[11] W. W. Murdoch, C. J. Briggs, and R. M. Nisbet. *Consumer-resource dynamics (MPB-36)*. Princeton University Press, 2013.

[12] H. L. Smith and P. Waltman. *The theory of the chemostat: dynamics of microbial competition*, volume 13. Cambridge university press, 1995.

[13] D. Tilman. *Resource competition and community structure*. Number 17. Princeton university press, 1982.