0.1 Consumer Resource Models in microbial ecology

Biology and Physics have always been tightly intertwined. Especially the years following the end of World War II saw many famous physicists getting interested in the blooming field of Biology [1], Leo Szilard or Erwin Schroedinger and his *What is Life? The Physical Aspect of the Living Cell* [2] among others. That exodus is no surprise, many biological phenomena at different scales are well modelled with Physics weaponry: from the use of Statistical Physics to solve protein folding problems [3] and find phase transitions in ecological communities [4] to the application of Hamiltonian dynamics to describe the movement of starling flocks [5].

However, Physics has not solved every problem yet: the study of microbial communities remains one of the biggest and most interesting challenges of contemporal microbiology. Indeed microbes and their complex interactions have a substantial, non trivial and very large impact on humans and their environment in various ways: we only start to understand the role of microbiological interactions in verterbrates' guts [6], or how they shape our soils [7] and seas [8].

Population dynamics in ecological communities are often approximated by variations of the Lotka-Volterra model [9]. This approach works well when the mediators of the competitive interaction between species reach a steady state fast enough such that their own dynamics can be eliminated [10]. However, such an assumption is not always true and one must in general always ask themselves whether it may be applied [11]. For microbial communities, previous literature shows that the population dynamics are not always well captured by a Lotka-Volterra model [10], which explains the need of a more mechanistic approach, where the dynamics of both the microbes and their resources are explicitly modelled. Robert MacArthur is one of the first ecologists to establish and study such a *Consumer Resource Model* (CRM) [12], launching a field still active today [13].

In the light of recent developments in the microbiology literature [14], we propose here a CRM[1] which explicitly takes into account syntrophy. This process, which is largely observed in microbial communities [14], by definition occurs when microbes release, through a metabolic process, byproducts that are consumed by some members of the microbial community. In short, we want to know what happens when consumers are also allowed to release resources.

0.2 Establishing the model and the goals

We want to write down a consumers-resources model (CRM) which describes the coupled evolution of N_S different species of biomass (denoted S_i with $i = 1, \dots, N_S$) and their N_R resources (denoted R_μ with $\mu = 1, \dots, N_R$) in a chemostat[2].

[1] One could argue that Flux Balance Analysis (FBA) [15] would be well suited for such a study. We ruled it out because it is known to scale badly [16] with system size and we do not want to be hindered by this limitation.

[2] In a chemostat, new nutrients are continuously added, while at the same time microorganisms and resources are removed in order to keep the culture volume constant [17].

The coupled evolution of the set of variables $\{R_\mu, S_i\}$ is given by:

$$\begin{cases} \dfrac{dR_\mu}{dt} = l_\mu - m_\mu R_\mu - \sum_j \gamma_{j\mu} R_\mu S_j + \sum_j \alpha_{\mu j} S_j & \text{(1a)} \\ \dfrac{dS_i}{dt} = \sum_\nu \sigma_{i\nu}\gamma_{i\nu} R_\nu S_i - d_i S_i - \sum_\nu \alpha_{\nu i} S_i & \text{(1b)} \end{cases}$$

We will call the variables $\{l_\mu, m_\mu, \gamma_{i\mu}, \alpha_{\mu i}, \sigma_{i\mu}, d_i\}$ *parameters of the model* **add R^* and S^***. On the other hand, because they may dynamically evolve, $\{R_\mu, S_i\}$ will be refered to as *dynamical variables* or simply *variables*. Note that there are in this model a lot of different symbols that may be easy to confuse. We will at least try to keep the following conventions:

- Quantities related to resources have subscripts in greek alphabet (*e.g.* the resource μ abundance R_μ). Quantities related to species have subscripts in latin alphabet (*e.g.* the species i abundance S_i). Finally, quantities related to both have both indices.
- Vectors (*i.e.* quantities with one index) are written with the latin alphabet (*e.g.* the resource μ death rate m_μ).
- Matrices (*i.e.* quantities with two indices, usually relating resources and species) are written with the greek alphabet (*e.g.* the rate $\gamma_{i\mu}$ at which species i consumes resource μ).

Our model tries to take numerous phenomena into account and it therefore may be helpful to take the time to explain the different terms of each differential equation. The temporal evolution of a resource R_μ is essentially driven by the following processes:

- Constant input from an outsider experimenter: this corresponds to the constant $+l_\mu$ term,
- Natural diffusion/deterioration at rate m_μ: this corresponds to the $-m_\mu R_\mu$ term,
- Consumption by the biomass species S_j at a rate $\gamma_{j\mu}$. In total this corresponds to the Lotka-Volterra style term **[insert ref]** $-\sum_j \gamma_{j\mu} R_\nu S_j$,
- Production coming from the species S_j at a rate $\alpha_{\mu j}$: $+\sum_j \alpha_{\mu j} S_j$. This is essentially what makes this model different from traditional CRMs **[insert ref]**.

On the other hand, biomass of species S_i changes because of the following processes **[insert figure]**:

- Consumption of resource R_ν at a rate $\gamma_{i\nu}$. Only a fraction $\sigma_{i\nu}$ of this is allocated to biomass growth: $+\sum_\nu \sigma_{i\nu}\gamma_{i\nu} R_\nu S_i$.
- Cell death/diffusion at rate d_i: this is the $-d_i S_i$ term.
- Release of resource R_ν at rate $\alpha_{\nu i}$ (this is the syntrophic interaction). In total $-\sum_\nu \alpha_{\nu i} S_i$.

We will mostly focus on the case where no resource coming from syntrophy is lost, *i.e.* $\tau_{\mu i} = \alpha_{\mu i}$.

The aim of the project is to study equilibria points of this model and their stability. In particular, we are interested in how syntrophy changes the robustness of the equilibria.

0.2.1 Equilibria of the model

We are interested in studying the stability of the equilibrium points of our model Eqs.(1). We say that $\{R_\mu^*, S_j^*\}$ is an *equilibrium*[3] of our model if they are fixed points of it, that means if the following equations are fulfilled:

$$\begin{cases} 0 = l_\mu - m_\mu R_\mu^* - \sum_j \gamma_{j\mu} R_\mu^* S_j^* + \sum_j \alpha_{\mu j} S_j^* & \text{(2a)} \\ 0 = \sum_\nu \sigma_{i\nu}\gamma_{i\nu} R_\nu^* S_i^* - d_i S_i^* - \sum_\nu \alpha_{\nu i} S_i^* & \text{(2b)} \end{cases}$$

As said above, our main goal is to study the stability of such equilibria. Before introducing different notions around equilibria, we focus on simplifying the problem first as much as we can. Note that we consider R_μ^* and S_i^* as parameters of the model.

0.2.2 Attack strategy and important notions

Before jumping right into the matter, it is important to explain how we will study this system of differential equations. Mainly two different but complimenteray approaches will be used: analytical and numerical. Note that the $\sim$ 5'000 lines of code we wrote from scratch and that we use to get the results of Section **??** is available at the address `https://gitlab.ethz.ch/palberto/consumersresources.git`.

Metaparameters and matrix properties

Studying the equilibria of our CRM will lead us to establish and study several relations involving the different parameters of the problem. Namely, these are: $l_\mu, m_\mu, \gamma_{i\mu}, \alpha_{\mu i}, \sigma_{i\mu}, d_i, R_\mu^*$ and S_i^* $\forall i = 1, \dots, N_S; \mu = 1, \dots, N_R$. We define the *parameters space* $\mathcal{P}$ as the space that contains all the parameters:

$$\mathcal{P} \equiv \left\{ p : p = (l_\mu, m_\mu, d_i, \gamma_{i\mu}, \alpha_{\mu i}, \sigma_{i\mu}, R_\mu^*, S_i^*) \right\} \tag{3}$$

Without taking into account the constraints on these parameters, there are $3N_R + 2N_S + 3N_R N_S$ free parameters, so $\mathcal{P} \simeq \mathbb{R}_+^{3N_R+2N_S+4N_R N_S}$. The difficulty of task becomes clear: $\mathcal{P}$ is a huge, complicated space so we need to find a way to simplify it.

Indeed, the goal is to study microbial communities with a large number of consumers and resources, typically $N_R, N_S \approx 25, 50, 100, \dots$ *i.e.* $\sim 2'500$ or more free parameters and so it is clear that a precise study on each element of each matrix is way too tenuous of a job. It's easier to look at it from a statistical point of view, *i.e.* we will write a matrix $q_{i\mu}$ as [18]:

$$q_{i\mu} = \mathfrak{Q} Q_{i\mu} \tag{4}$$

where $\mathfrak{Q}$ is a random variable of mean Q_0 and standard deviation σ_Q. $Q_{i\mu}$ is a binary matrix that, if interpreted as an adjacency matrix, tells about the network structure of the quantity $q_{i\mu}$.

[3]For the sake of brevity, we will sometimes drop the μ and j subscripts when we write $\{R_\mu^*, S_j^*\}$.

We apply this way of thinking to the parameters of our problem, namely we can write:

$$l_\mu = \mathfrak{L} \tag{5a}$$
$$m_\mu = \mathfrak{M} \tag{5b}$$
$$\gamma_{i\mu} = \mathfrak{G} G_{i\mu} \tag{5c}$$
$$\alpha_{\mu i} = \mathfrak{A} A_{\mu i} \tag{5d}$$
$$\sigma_{i\mu} = \mathfrak{S} \tag{5e}$$
$$d_i = \mathfrak{D} \tag{5f}$$
$$R^*_\mu = \mathfrak{R} \tag{5g}$$
$$S^*_i = \mathfrak{S} \tag{5h}$$

Note that we did not put any explicit topological structure on $l_\mu, m_\mu, d_i, R^*_\mu, S^*_i$ and $\sigma_{i\mu}$ because we require these to always be larger than zero (in particular for S^* and R^*, we require positive-valued equilibria [19]).

In order to make computations analytically tractable, we will require a small standard deviation on the parameters involved in the problem. In that regime, every random variable $\mathcal{Q}$ is well approximated by its average value Q_0, that we call a *metaparameter*. While studying things analytically we will hence often come back to the following approximation:

$$l_\mu \approx l_0 \tag{6a}$$
$$m_\mu \approx m_0 \tag{6b}$$
$$\gamma_{i\mu} \approx \gamma_0 G_{i\mu} \tag{6c}$$
$$\alpha_{\mu i} \approx \alpha_0 A_{\mu i} \tag{6d}$$
$$\sigma_{i\mu} \approx \sigma_0 \tag{6e}$$
$$d_i \approx d_0 \tag{6f}$$
$$R^*_\mu \approx R_0 \tag{6g}$$
$$S^*_i \approx S_0 \tag{6h}$$

This is mathematically equivalent to collapsing the parameter space $\mathcal{P}$ to a lower dimensional space. Formally that lower dimensional space is the Cartesian product of $\mathcal{M}$ and $\mathcal{B}_{N_S \times N_R} \times \mathcal{B}_{N_R \times N_S}$, where $\mathcal{M}$ is the *metaparameters space*:

$$\mathcal{M} \equiv \{m : m = (l_0, m_0, d_0, \gamma_0, \alpha_0, \sigma_0, R_0, S_0)\} \tag{7}$$

and $\mathcal{B}_{N \times M}$ is the set of binary matrices of dimensions $N \times M$. To summarize, what we did was simply designing a collapsing procedure $\mathcal{C} : \mathcal{P} \to \mathcal{M} \times \mathcal{B}_{N_S \times N_R} \times \mathcal{B}_{N_R \times N_S}$ in order to simplify our problem.

Mathematically, when we do analytical computations, we will mostly work in the collapsed space $\mathcal{C}(\mathcal{P})$ because it reduces the number of parameters from $3N_R + 2N_S + 3N_R N_S$ (continous) to 8 (continous) + $3N_R N_S$ (binary). And to make the problem even simpler, instead of looking at each entry of the binary matrices G and A individually, we will consider only some globally defined quantities of these matrices. For a matrix M_{ij} the metrics interesting to us are most of all:

- Its **nestedness**: this measures how "nested" the system is, *i.e.* if there are clusters grouped together[4]. It is known [21, 18] that nestedness can play a profound role in the dynamics of ecological communities. Although it is somewhat controversed [22], we will keep the definition of the nestedness $\eta(M)$ of a binary matrix M as it was used in [21]:

$$\eta(M) \equiv \frac{\sum_{i<j} n_{ij}}{\sum_{i<j} \min(n_i, n_j)} \tag{8}$$

where the number of links n_i is simply the degree of the i-th row of M

$$n_i \equiv \sum_k M_{ik}, \tag{9}$$

and n_{ij} is the overlap matrix defined as

$$n_{ij} \equiv \sum_k M_{ik} M_{jk}. \tag{10}$$

- Its **nestedness**[5]: this measure, simply defined as the ratio of non-zero links in a matrix, is central in the study of plants-and-animals systems [18]. It is formally defined for a matrix q_{ij} of size $N \times M$ as:

$$\kappa(q) \equiv \frac{\sum_{ij} Q_{ij}}{NM} \tag{11}$$

where Q is the (binary) network adjacency matrix of q.

Losing complexity – how to gain it back

As explained above, the idea is to simplify the study of a system with a large number of parameters to a system with a manageable number of so-called "metaparameters". Of course, collapsing a very high dimensional space to a low-dimensional space will make us lose information. Losing information – and hence complexity – is convenient when doing analytical computations but it is not when we want to produce precise numerical results.

So, how do we bridge the gap between what we work with analytically, *i.e.* a set of metaparameters and binary matrices, to precise measurements of quantities defined in our model Eq.(1)? The answer is simple: we define a function $\mathcal{A} : \mathcal{M} \times \mathcal{B}_{N_S \times N_R} \times \mathcal{B}_{N_R \times N_S} \to \mathcal{P}$ which brings us from the collapsed space to the parameter space[6]. Numerically what we will do is: from a set of metaparameters $m \in \mathcal{M}$ and binary matrices $B = (G, A) \in \mathcal{B}_{N_S \times N_R} \times \mathcal{B}_{N_R \times N_S}$, produce a (or several) set(s) of parameters $p = \mathcal{A}(m, B) \in \mathcal{P}$ and study properties of it. Section **??** details how $\mathcal{A}$ is constructed.

[4] In typical Lotka-Volterra models, where only species-species interactions are considered, *e.g.* [20], measuring the nestedness of the γ consumption matrix would be in the same spirit as counting how many niches there are in the community.

[5] For the matrix consumption G, we will call it especially the "ecological overlap".

[6] Note that since the collapsed space is lower dimensional than the parameters space, $\mathcal{A}$ is not the inverse of $\mathcal{C}$.