

Since its very inception [1], the study of ecological interactions has been and still is tightly close to the one of random matrices [2, 3, 4]. Usually, the procedure is we assume a feasible equilibrium point, where some matrix of the model (*e.g.* the species-interaction matrix or the jacobian) is approximated as random, and then study the dynamical or structural stability of said feasible point.

That framework is not satisfying for the study we would like to conduct, because it does not take time to study whether random parameters make sense in the first place. Indeed, before studying whether a microbial community can sustain perturbations, we need to know if said community actually *exists*. Biological systems, like any other natural systems, are constrained by laws, whether they arise from physical or biological considerations. For instance, it would not make sense to consider microbial communities that *e.g.* violate the laws of thermodynamics. In the following section, we explain how such considerations can help determining the answer to the *feasibility* question:

Can microbial communities arising from a random set of parameters make sense on a physical and biological level? If not, what are the conditions that should be imposed and how are these translated mathematically?

0.0.1 Basic concepts

As explained above, we want to impose conditions such that we only study systems that are compatible with biological and physical laws. Choosing such restrictions is a crucial task : we want to be as close to nature as possible but we also need to stay simple enough such that the model remains mathematically tractable. Our choice is the following : any system deemed as feasible must have “biological” model parameters and conserve biomass.

Asking for the model parameters to be “biological” means we want them to carry their intended biological interpretation. This means *e.g.* that any syntrophic interaction has to be non-negative $\alpha_{\mu i} \geq 0$ otherwise it cannot be interpreted as a syntrophic interaction anymore! More generally, the values of the parameters will be restricted. Namely, we are looking for positive-valued equilibria. Also, we require that every consumer can allocate some of each resource it consumes to growth¹ : zero efficiencies are forbidden. Finally every resource external feeding rate should be non-zero in order to avoid resource depletion and every resource and consumer must eventually die out in the absence of interaction. Mathematically, these considerations are equivalent to:

$$\boxed{R_{\mu}^*, S_i^*, \sigma_{i\mu}, l_{\mu}, d_i, m_{\mu}, \sigma_{i\mu} > 0 \text{ and } \gamma_{i\mu}, \alpha_{\mu i} \geq 0.} \quad (1)$$

That condition already greatly restricts the choice of parameters $p \in \mathcal{P}$. However, additional complexity arises from the relationships parameters have to follow by definition. Indeed, the $3N_R + 2N_S + 3N_R N_S$ parameters are constrained by the $N_R + N_S$ equations (??). So if we choose $2N_R + N_S + 3N_R N_S$ parameters, the remaining $N_R + N_S$ are instantly determined. Traditionally, we would solve for R^* and S^* and choose the rest of the parameters, but for reasons explained in Appendix ??, we will solve for the consumers death rates d_i and

¹It wouldn’t make sense to say that species i eats resource μ with efficiency 0, since this is equivalent to species i not eating resource μ , and this is already encoded in the network structure.

the resources diffusion rate m_μ . This means that if we *choose* non-negative $\gamma, \alpha, \sigma, \tau, l, R^*$ and S^* , Eqs.(1) can be combined with Eqs.?? into:

$$\begin{cases} d_i = \sum_{\nu} (\sigma_{i\nu} \gamma_{i\nu} R_{\nu}^* - \alpha_{\nu i}) > 0 \quad \forall i = 1, \dots, N_S \\ m_{\mu} = \frac{l_{\mu} - \sum_j (\gamma_{j\mu} R_{\mu}^* - \alpha_{\mu j}) S_j^*}{R_{\mu}^*} > 0 \quad \forall \mu = 1, \dots, N_R \end{cases} \quad (2a)$$

$$\quad (2b)$$

In addition to these constraints, any feasible system should conserve biomass *at equilibrium*²: no species should be able to produce more biomass than it physically can. More specifically, a consumer i attains, from resources consumption, a total biomass of $\sum_{\nu} \gamma_{i\nu} R_{\nu}^* S_i^*$. From this available biomass, only $\sum_{\nu} \sigma_{i\nu} \gamma_{i\nu} R_{\nu}^* S_i^*$ is devoted to growth. Out of the remaining $\sum_{\nu} (1 - \sigma_{i\nu}) \gamma_{i\nu} R_{\nu}^* S_i^*$, a part $\sum_{\nu} \alpha_{\nu i} S_i^*$ is given back to the resources as a syntrophic interaction. We simply impose that the syntrophic interaction is smaller than or equal to the available remaining biomass :

$$\sum_{\nu} (1 - \sigma_{i\nu}) \gamma_{i\nu} R_{\nu}^* \geq \sum_{\nu} \alpha_{\nu i} \quad \forall i = 1, \dots, N_S. \quad (3)$$

From now on, we will say that **a parameter set p is *feasible* if it satisfies Eqs.(2) and (3)**. This is completely deterministic, in the sense that for a given parameters set $p \in \mathcal{P}$ one can without a doubt say whether it is feasible or not. Hence we define the *parameters set feasibility function* $\mathfrak{F} : \mathcal{P} \rightarrow \{0, 1\}$, which takes a parameter set as an input and tells you whether this parameter set is feasible or not:

$$\mathfrak{F}(p) = \begin{cases} 1 & \text{if } p \text{ is feasible,} \\ 0 & \text{else.} \end{cases} \quad (4)$$

However as explained in the introduction we will usually not work with a parameter set $p \in \mathcal{P}$ directly – because there are too many variables to keep track of – but with a metaparameter set $m \in \mathcal{M}$ and a consumption-syntrophy network $(G, A) \in \mathcal{B}_{N_S \times N_R} \times \mathcal{B}_{N_R \times N_S}$ instead. We can define a corresponding *metaparameters set feasibility function* $\mathcal{F} : \mathcal{M} \rightarrow [0, 1] \times \mathcal{B}_{N_R \times N_S}$ which is the probability that a given set of metaparameters $m \in \mathcal{M}$ coupled with binary matrices $B = (G, A)$ gives rise – through the algorithmic procedure \mathcal{A} – to a feasible parameter set :

$$\boxed{\mathcal{F}(m, B) = \text{Probability} \{ \mathfrak{F}(\mathcal{A}(m, B)) = 1 \}} \quad (5)$$

In practice $\mathcal{F}(m, B)$ is estimated numerically by generating N parameters sets from (m, B) and calculating the number of feasible ones :

$$\mathcal{F}(m, B) = \lim_{N \rightarrow \infty} \sum_{i=1}^N \frac{\mathfrak{F}(\mathcal{A}(m, B))}{N} \approx \sum_{i=1}^N \frac{\mathfrak{F}(\mathcal{A}(m, B))}{N} \quad \text{for } N \gg 1. \quad (6)$$

²This weak condition should hold only at equilibrium : we allow transition periods where biomass may not be conserved.

0.0.2 The feasibility region

Appendix ?? explains the algorithmic procedure $\mathcal{A}(m, B) \in \mathcal{P}$ which allows us to build feasible parameters out of a set of metaparameters and a consumption-syntrophy network. However, in order to work properly, the combination of metaparameters used as an input of the algorithm must most of the time lead to the realisation of feasible systems. We hence need to find what region of the metaparameters space lead to a high feasibility : this is precisely the idea behind the notion of the feasibility region discussed below.

But first, let's see how our study can be made simpler. Feasibility conditions discussed above tell us that we may choose six metaparameters³ : $\gamma_0, \alpha_0, l_0, \sigma_0, S_0$ and R_0 . However, following the analysis of [4], we notice that our model Eqs.(??) still possesses some freedom. Indeed we can choose the set of units we work in to fix the values of some metaparameters. There are two physical quantities at stake here : biomass and time, and we may choose, however we want it, a specific set of units describing both of them. We measure biomass in units of the average resource abundance at equilibrium⁴:

$$\langle R_\mu \rangle = R_0 = 1. \quad (7)$$

Similarly, we measure time such that the average external resource uptake rate is one, that is:

$$\langle l_\mu \rangle = l_0 = 1. \quad (8)$$

After this manipulation, the number of metaparameters is reduced from six to four : only γ_0, S_0, α_0 and σ_0 remain.

For the sake of simplicity, we keep the same σ_0 throughout our whole study. We take a value close to the efficiency of real microbial systems [**insert ref**], that is $\sigma_0 = 0.25$.

Overall, we need to choose the last three remaining metaparameters: α_0, γ_0 and S_0 . We will modify these metaparameters throughout the study. Since the remaining eight are fixed, we sometimes will elude them in the notation and will write instead of $m = (\gamma_0, S_0, \alpha_0, \sigma_0, R_0, l_0, d_0, m_0) \in \mathcal{M}$ simply $m = (\gamma_0, S_0, \alpha_0)$.

Formally, we can define for a consumption matrix G coupled with a syntrophy adjacency matrix A the x -feasible volume $\mathcal{F}_x^{G,A}$ of the metaparameters space \mathcal{M} that will lead to at least a ratio x of feasible systems *i.e.* :

$$\mathcal{F}_x^{G,A} \equiv \{m \in \mathcal{M} : \mathcal{F}(m, (G, A)) \geq x\}. \quad (9)$$

It is clear that $\mathcal{F}_0^{G,A} = \mathcal{M} \forall G$ and $\text{Vol}(\mathcal{F}_x^{G,A}) \leq \text{Vol}(\mathcal{F}_y^{G,A}) \forall x > y, G$. We can similarly define for a set $S = \{(G_1, A_1), (G_2, A_2), \dots, (G_N, A_N)\}$ of N couples of matrices their *common feasibility* region \mathcal{F}_x^S , which is the region of the metaparameters space where feasibility is at least x for every couple in the set:

$$\mathcal{F}_x^S \equiv \bigcap_{(G,A) \in S} \mathcal{F}_x^{G,A}. \quad (10)$$

³Indeed, we saw that d_i and m_μ are set by the other parameters, so we cannot freely choose d_0 and m_0 .

⁴That choice is not completely innocent. Indeed we will see later that the matrix $\alpha_{\nu i} - \gamma_{i\nu} R_\nu^*$ is a crucial quantity here. Setting $\langle R^* \rangle = 1$ allows us to simply study the impact of γ against α instead of the more complicated γR^* versus α .

We also define for a matrix set S , its critical feasibility $f^*(S)$, which is the largest feasibility we can get while still having a non-zero common volume :

$$f^*(S) \equiv \max_{x \in [0,1]} \{x : \text{Vol}(\mathcal{F}_x^S) > 0\}. \quad (11)$$

For actual computations, we will choose a matrix set S_M , stick to it during the whole thesis, and work in its critical feasibility region \mathcal{F}^* , defined as :

$$\boxed{\mathcal{F}^* \equiv \mathcal{F}_{f^*(S_M)}^{S_M}}. \quad (12)$$

Our hope is that we may find a fully feasible common region, *i.e.* $f^*(S) = 1$.

0.0.3 Estimating the fully feasible region $\mathcal{F}_1^{G,A}$

Now that we defined the x -feasible volume of a given couple consumption-syntrophy network (G, A) in Eq.(9), we would like to know what regions of the metaparameters space lead to fully feasible systems. We imposed two conditions that characterise the set of feasible parameters: Eqs.(2) and (3). We use them as a start to get corresponding metaparameters equations that describe $\mathcal{F}_1^{G,A}$.

Biomass conservation

As stated above, we require that biomass is conserved in our model. This is equivalent to fulfilling Eq.(3), which we rewrite here:

$$\sum_{\nu} (1 - \sigma_{i\nu}) \gamma_{i\nu} R_{\nu}^* \geq \sum_{\nu} \alpha_{\nu i} \quad \forall i = 1, \dots, N_S. \quad (13)$$

Eqs.(??) can be used to estimate the RHS of this equation:

$$\sum_{\nu} \alpha_{\nu i} \approx \deg(A, i) \alpha_0, \quad (14)$$

where $\deg(A, i)$ is the degree of the i -th column of the α matrix :

$$\deg(A, i) = \sum_{\nu} A_{\nu i}. \quad (15)$$

Similarly,

$$\sum_{\nu} (1 - \sigma_{i\nu}) \gamma_{i\nu} R_{\nu}^* \approx (1 - \sigma_0) R_0 \sum_{\nu} \gamma_{i\nu} \approx \deg(G, i) (1 - \sigma_0) R_0 \gamma_0, \quad (16)$$

Energy conservation Eq.(3) is then equivalent to

$$\deg(A, i) \alpha_0 \lesssim \deg(G, i) (1 - \sigma_0) R_0 \gamma_0 \quad \forall i = 1, \dots, N_S \quad (17)$$

Since $\deg(G, i) > 0$, we have⁵:

$$\frac{\deg(A, i)}{\deg(G, i)} \alpha_0 \lesssim (1 - \sigma_0) R_0 \gamma_0 \quad \forall i = 1, \dots, N_S \quad (18)$$

⁵Indeed, $\deg(G, i)$ is the number of resources species i eats. We of course ask every consumer to at least consume something, otherwise they would not be part of the microbial community.

This is fulfilled if :

$$\boxed{\max_i \left\{ \frac{\deg(A, i)}{\deg(G, i)} \right\} \alpha_0 \lesssim (1 - \sigma_0) R_0 \gamma_0.} \quad (19)$$

Systems where the ratio $\frac{\# \text{resources released to}}{\# \text{resources consumed}}$ is small for each species allow for a larger individual syntrophy interaction (which is very intuitive).

Biological interpretation of the parameters

Additionally, the consumers death rates d_i have to be positive. This implied Eq.(2a), which may be recast as :

$$\sum_{\mu} \sigma_{i\mu} \gamma_{i\mu} R_{\mu}^* > \sum_{\mu} \alpha_{\mu i} \quad (20)$$

Using a reasoning similar to above, we get a corresponding metaparameters inequality:

$$\boxed{\max_i \left\{ \frac{\deg(A, i)}{\deg(G, i)} \right\} \alpha_0 \lesssim \sigma_0 R_0 \gamma_0.} \quad (21)$$

Also, the resources diffusion rates m_{ν} need to be positive:

$$l_{\nu} + \sum_j \alpha_{\nu j} S_j^* > \sum_j \gamma_{j\nu} R_{\nu}^* S_j^* \quad \forall \nu = 1, \dots, N_R \quad (22)$$

Which is equivalent to

$$l_0 + \deg(A, \nu) \alpha_0 S_0 \gtrsim \deg(G, \nu) \gamma_0 R_0 S_0 \quad \forall \nu \quad (23)$$

Since $\deg(G, \nu) > 0$, we⁶ can divide the above equations by $\deg(G, \nu) > 0$ and then recast these N_R equations into a single condition:

$$\boxed{\min_{\nu} \left\{ \frac{l_0}{\deg(G, \nu) S_0} + \frac{\deg(A, \nu)}{\deg(G, \nu)} \alpha_0 \right\} \gtrsim \gamma_0 R_0} \quad (24)$$

This says that systems where the ratio $\frac{\# \text{number of species that release to me}}{\# \text{number of species that consume me}}$ is large for every resource are more feasible. The strategy should be then to have γ 's that have large $\deg(G, \nu)$ (*i.e.* resources are consumed by many species) and large $\deg(G, i)$ (*i.e.* species consume a lot of species), and the other way around for α (not sure about this for the last one).

Combining both conditions

The two upper bounds Eqs.(19)-(21) on α_0 can be combined in a single inequality :

$$\max_i \left\{ \frac{\deg(A, i)}{\deg(G, i)} \right\} \alpha_0 \lesssim \min(1 - \sigma_0, \sigma_0) \gamma_0 R_0 \quad (25)$$

⁶Similarly to a previous footnote, we require that every resource ν is eaten by at least one consumer, *i.e.* $\deg(G, \nu) > 0$, otherwise it does not belong to the community.

Note that when $\alpha_0 > 0$, we will trivially require that the syntrophy matrix is not empty, *i.e.* there exists at least an i for which $\deg(A, i) \geq 1$. Also, the largest value $\deg(G, i)$ can get (for any i) is N_R . Hence,

$$\max_i \left\{ \frac{\deg(A, i)}{\deg(G, i)} \right\} \geq \frac{1}{N_R}, \quad (26)$$

and we can find the largest allowed theoretical non-zero α_0 :

$$\boxed{\alpha_0 \lesssim \min(1 - \sigma_0, \sigma_0) \gamma_0 R_0 N_R.} \quad (27)$$

Finally, Eq.(24) and (25) can be combined into a single one, which characterises the fully feasible region $\mathcal{F}_1^{G,A}$:

$$\boxed{\max_i \left\{ \frac{\deg(A, i)}{\deg(G, i)} \right\} \alpha_0 \lesssim \min(1 - \sigma_0, \sigma_0) \gamma_0 R_0 \lesssim \min(1 - \sigma_0, \sigma_0) \min_{\nu} \left\{ \frac{l_0}{\deg(G, \nu) S_0} + \frac{\deg(A, \nu)}{\deg(G, \nu)} \alpha_0 \right\}} \quad (28)$$