Microbial Interactions as Drivers of a Nitrification Process in a Chemostat

Pablo Ugalde-Salas, Héctor Ramírez C., Jérôme Harmand, Elie Desmond-Le Quéméner

bioengineering

MDPI

Article

Microbial Interactions as Drivers of a Nitrification Process in a Chemostat

Pablo Ugalde-Salas [1,*], Héctor Ramírez C. [2], Jérôme Harmand [1] and Elie Desmond-Le Quéméner [1]

[1] LBE, INRAE, Université de Montpellier, 11100 Narbonne, France; jerome.harmand@inrae.fr (J.H.); elie.le-quemener@inrae.fr (E.D.-L.Q.)
[2] Departamento de Ingeniería Matemática, Centro de Modelamiento Matemático (CNRS UMI 2807), Universidad de Chile, Santiago, Chile; hramirez@dim.uchile.cl
* Correspondence: pablo.ugalde.s@gmail.com

Abstract: This article deals with the inclusion of microbial ecology measurements such as abundances of operational taxonomic units in bioprocess modelling. The first part presents the mathematical analysis of a model that may be framed within the class of Lotka–Volterra models fitted to experimental data in a chemostat setting where a nitrification process was operated for over 500 days. The limitations and the insights of such an approach are discussed. In the second part, the use of an optimal tracking technique (developed within the framework of control theory) for the integration of data from genetic sequencing in chemostat models is presented. The optimal tracking revisits the data used in the aforementioned chemostat setting. The resulting model is an explanatory model, not a predictive one, it is able to reconstruct the different forms of nitrogen in the reactor by using the abundances of the operational taxonomic units, providing some insights into the growth rate of microbes in a complex community.

Keywords: microbal interactions; microbial growth rate; bifurcation analysis; generalized Lotka–Volterra; chemostat theory; optimal control

Citation: Ugalde-Salas, P.; Ramírez C., H.; Harmand, J.; Desmond-Le Quéméner, E. Microbial Interactions as Drivers of a Nitrification Process in a Chemostat. *Bioengineering* **2021**, *8*, 31. https://doi.org/10.3390/bioengineering8030031

Academic Editor: Dongda Zhang, Matthew Wade and Sovanna Tik

Received: 20 November 2020
Accepted: 9 February 2021
Published: 25 February 2021

1. Introduction

Microbial communities and their interactions play a central role in the understanding of microbial ecosystems [1], and a current challenge is integrating genetic sequencing data in a deterministic modelling framework [2,3]. Using the terminology from the thorough review in current methodologies on the deterministic modelling approaches of microbial community dynamics presented by Song et al. [4], this articles deals with population-based approaches where species are taken as the interacting units.

The classical ecological concept of species and niche in the microbial world is an elusive one: in the macro world one can clearly differentiate one species from another for reproductive reasons and their ability to give birth to offspring. In the case of bacteria and archea, reproduction goes simply by binary fission and exchange of some functional genes (e.g., the ability to synthesize or metabolize substances) can be acquired in evolutionary scale through lateral gene transfer [5]. Therefore as an ecological problem is hard to define precisely the 'niche' of 'microbial species'. These obstacles can be circumvented by considering the microbiologist concept of operational taxonomic unit (OTU) based on the clustering of organisms sharing similar sequences of the 16S rDNA marker gene. In the past years considerable efforts have been made to measure the bacterial community composition. Tests such as fluorescence in situ hybridization (FISH), polymerase chain reaction (PCR) dependent techniques, and PCR independent techniques for the analysis of DNA have become a standard tool for studying microbial diversity [6]. The contribution of this article is a new method to integrate the microbial community measurements in chemostat models, based on any sequencing or fingerprinting technique that can quantify the species abundances over time. In other words, while most models used in bioengineering are functional—in

the sense they consider only one species per biological reaction considered—this work is an attempt to merge classical population-based models used in ecology and those used by engineers in biotechnology.

Interactions lie at the heart of ecology. Lotka [7] and Volterra [8], independently, presented a 2 dimensional dynamical system to model prey-predator relationships, now known as the Lotka–Volterra (LV) equations. The model is very rich from a mathematical standpoint, and is also a classic equation to study in Mathematical Ecology [9]. Extensions of the Lotka–Volterra model have derived what is now known as generalized Lotka–Volterra (gLV) models [10] shown in Equation (1):

$$\dot{x}_i = \mu_i \left(1 + \sum_{j=1}^{n} a_{ij} x_j \right) x_i \quad i \in \{1, \ldots, n\} \tag{1}$$

where x_i represents the species abundance, μ_i the intrinsic growth rate of the species, and the terms a_{ij} reflect the effect of OTU j on the growth of OTU i. The equation states that the growth rate of x_i is proportional to x_i, but this proportionality constant depends on its intrinsic growth rate multiplied by the sum of all interactions affecting it. Note that if there are no type of interactions ($a_{ij} = 0$), one recovers n uncoupled linear differential equations, and thus the solution becomes $x_i(t) = x_i(0) \exp(\mu_i t)$, that is exponential growth on time.

The diagonal terms a_{ii} are known as intraspecies interaction, while the off diagonal terms are known as the pairwise interspecies interactions. Noting the signs of pairs (a_{ij}, a_{ji}), the classical ecological relationships of mutualism or cooperation $(+,+)$, commensalism $(+,0)$, predation or parasitism $(+,-)$, competition $(-,-)$, and ammensalism $(-,0)$ can be recovered [11]. Model (1) has been thoroughly analysed, even when the coefficients μ_i and a_{ij} are time dependent and exhibit periodicity (which models seasonal traits) [12,13]. The gLV model has been used in microbial ecology to some degree of success to study the gut microbiome of mice infected with *C. difficile* [14]. However, the quadratically growing number of parameters to describe interactions naturally entails problems of identifiability if the data set is not large enough, or the system has not been sufficiently perturbed. On a more conceptual ground, the interaction coefficients of a gLV model do not represent mechanistically anything, so even if a model correctly predicts the microbial community dynamics, it might not add to the understanding of what could be physically or biologically taking place. These observations led us to develop what can be considered the core contribution of this article, which is to study the growth rate of each species in a mixed culture: we reconstruct the shape of their growth rate, instead of trying to fit a particular function (such as the gLV equations). As Monod himself commented when developing the growth law that bears their name is that any function with the same shape (monotone, concave, and bounded on the substrate) would have served [15], in this spirit we formulate the question: what is the shape of the growth functions of multiple species developing together?

As a departing point, the work of Dumont et al. [16] is presented in Section 2. They modelled a chemostat experiment where nitrification takes place by considering a glV model coupled with a substrate limited growth expression (μ_i is no longer a constant, but the classic Monod expression) and fitted their model using absolute abundances of the major OTU identified by molecular fingerprints obtained by single-strand conformation polymorphism (SSCP). Section 3 inspects the model through a mathematical analysis. Some interesting outputs of this analysis are that the number of possible equilibrium points grows exponentially with the number of species, coexistence can be achieved within the same functional group, and bi-stability may arise. In Section 4, the concept of interaction function is developed such that it generalizes the gLV model. For approximating the interaction function a method of optimal control theory was adapted: The growth rate of each species is modulated by a constrained regular control of the system, thus the growth rate of each OTU is corrected in order to fit the experimental data. The regular control is composed of a feedback part on the species state variable, and a feed forward part, or tracking, on the measurement of abundances of each species; the method involves solving state-dependent

Ricatti equations [17]. In Section 5, the methodology from Section 4 is applied to the data from the experiments performed by Dumont et al. [18] and not just to the most abundant species as it was the case of the model analysed in Section 3 [16]. This approach explicitly assumes that dynamics of complex ecosystems are driven by interactions, that are the results of feedback loops of each species on the growth rate of others. The method shows that by following the community dynamics one can propose a growth rate that reconstructs the substrates dynamics, however this cannot be considered a predictive model, but rather a explicatory model. The article ends with a discussion on the scope of applicability and perspectives of the method.

2. Model Definition

Notations used throughout the article:

1. n: the number of OTU considered.
2. n_i, $i \in \{1,2\}$: the number of OTU in functional group G_i. In the example G_1 corresponds to ammonia oxidizing bacteria (AOB) and G_2 corresponds to nitrite oxidizing bacteria (NOB).
3. Let m be an interger then $[m] := \{1,\dots,n\}$.
4. x_i: is the concentration of OTU i measured in $[g/l]$. $i \in [n]$.
5. x: vector $(x_1,\dots,x_n)^\top$.
6. s_1: concentration of substrate 1 in $[g/l]$. In the example s_1 represents ammonium.
7. s_2: concentration of substrate 2 in $[g/l]$. In the example s_2 represents nitrite.
8. s_3: concentration of substrate 3 in $[g/l]$. In the example s_3 represents nitrate.
9. s_{in}: entry concentration of substrate 1 in $[g/l]$. May depend on time $s_{in} = s_{in}(t)$.
10. s: vector $(s_1,s_2,s_3)^\top$. Referred to as metabolites.
11. $\mathcal{I}_i(t,x)$: Interaction function of OTU $i \in \{1,\dots,n\}$.
12. $\mu_i(s,x)$: growth function of OTU $i \in \{1,\dots,n\}$.
13. $\mu = (\mu_1(s,x),\dots,\mu_n(s,x))$ vector containing the growth function of every OTU.
14. D: dilution rate of the continuous reactor in $[1/day]$. May depend on time $D = D(t)$.
15. y_i: yield of grams of OTU i formed per gram of substrate consumed.
16. y_{s_i/x_j}: yield of grams of substrate s_i consumed/produced per gram of OTU j formed. If negative it represents consumption, if positive it represents production.
17. Y: matrix containing all yields such that $Y_{ij} = y_{s_i/x_j}$.
18. For integers m_1 and m_2 and $a \in \mathbb{R}$, $a_{m_1\times m_2}$ represents a matrix of m_1 rows and m_2 columns with a in every entry.
19. Let m be an integer then I_m is the identity matrix of size m.
20. Let M be a matrix, then $M_{i\bullet}$ represents the i-th row of matrix M.
21. Let S be a finite set with $m \in \mathbb{N}$ elements. Then $|S| := m$.
22. Given a vector $v = (v_1,\dots,v_n) \in \mathbb{R}^n$, the function $\mathrm{diag}(v)$ stands for:

$$\begin{aligned} \mathrm{diag} : \mathbb{R}^n &\to \mathbb{M}_{n\times n}(\mathbb{R}) \\ v &\to \begin{pmatrix} v_1 & 0 & \dots & 0 \\ 0 & v_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & v_n \end{pmatrix} \end{aligned} \tag{2}$$

2.1. Stoichiometric Equations

A cascade (bio)reaction process is considered. Suppose n different OTU are present in the chemostat. A two step cascade reaction refers to the situation where a group of microorganisms ($G_1 \subset [n]$) consumes a substrate s_1 and produces s_2 and biomass, while another group of microorganisms ($G_2 \subset [n]$) consumes s_2 and produces s_3 and biomass. G_1 and G_2 are called functional groups. The number of organisms in each functional will

be denoted n_1 and n_2, respectively, that is $|G_1| = n_1$ and $|G_2| = n_2$. This work treats the case when G_1 and G_2 are disjoint sets:

Hypothesis 1 (H1). *Sets G_1 and G_2 satisfy: $G_1 \cap G_2 = \emptyset$ and $G_1 \cup G_2 = [n]$.*

The situation is described as simplified Reactions (R1) and (R2). The reactions are simplified in the sense that they do not attempt to represent a balanced chemical reaction, rather they represent the direction of the bioprocess and the proportions of different consumed and formed compounds of interest. The terms y_i are known as yields, they represent the quantity of g of biomass produced per g of substrate consumed by OTU i. For example, in the case of reaction (R1), one gram of s_1 is consumed, one gram of s_2 and y_{x_i/s_1} grams of dry biomass of OTU i are produced.

$$s_1 \xrightarrow{\mu_i(s,x)} s_2 + y_i x_i \quad \forall i \in G_1 \tag{R1}$$

$$s_2 \xrightarrow{\mu_i(s,x)} s_3 + y_i x_i \quad \forall i \in G_2 \tag{R2}$$

However for expressing the system of differential equations further below, the terms y_{s_i/x_j} are used. They express the grams of substrate s_i consumed (negative sign) or produced (positive sign) per gram of OTU j formed. They are related to y_i as seen in Table 1. This defines the stoichiometry matrix $Y \in \mathbb{R}^{3\times n}$, such that $Y_{ij} = y_{s_i/x_j}$.

Table 1. Relationship of y_{s_i/x_j} with y_j.

Yields per Biomass Formed	$j \in G_1$	$j \in G_2$
y_{s_1/x_j}	$-\frac{1}{y_j}$	0
y_{s_2/x_j}	$\frac{1}{y_j}$	$-\frac{1}{y_j}$
y_{s_3/x_j}	0	$\frac{1}{y_j}$

Furthermore, for each $i \in [n]$, OTU i is characterized by its process rate (also known as growth function) $\mu_i(s,x)$. Notice that for being as generic as possible, the growth rate may be a function of the whole state in order to model the influence of all OTU on the growth rates of others.

An example of this process is the nitrification process where group G_1 is known as Ammonia oxidizing Bacteria (AOB), and group G_2 is known as Nitrite oxidizing Bacteria (NOB) [19].

2.2. Mass Balance Equations

Consider the scenario of a continuous and homogeneous reactor: the input flow is the same as the output flow, with a dilution rate D. The input flow contains a concentration s_{in} of substrate s_1. Each OTU grows at a rate $\mu_i(s,x)$. System (3) represents this situation. A specific case of $\mu_i(s,x)$ is given in the next subsection.

$$\begin{aligned}
\dot{x_i} &= (\mu_i(s,x) - D)x_i \quad \forall i \in [n] \\
\dot{s_1} &= (s_{in} - s_1)D - \sum_{i\in G_1} \frac{1}{y_i}\mu_i(s,x)x_i \\
\dot{s_2} &= -s_2 D + \sum_{i\in G_1} \frac{1}{y_i}\mu_i(s,x)x_i - \sum_{i\in G_2} \frac{1}{y_i}\mu_i(s,x)x_i \\
\dot{s_3} &= -s_3 D + \sum_{i\in G_2} \frac{1}{y_i}\mu_i(s,x)x_i
\end{aligned} \tag{3}$$

System (3) can also be written in a more compact form using the stoichiometric matrix Y and the diag operator.

$$\dot{x} = \text{diag}(\mu(x,s) - D_{n\times 1})x \tag{4}$$

$$\dot{s} = \left(\begin{bmatrix} s_{in} & 0 & 0 \end{bmatrix}^\top - s\right)D + Y\,\text{diag}(\mu(x,s))x \tag{5}$$

2.3. Kinetic Equations

In the work of Dumont et al. [16], the growth rates seen in Equation (6) were calibrated against experimental data for the two most abundant OTU of each functional group.

$$\begin{aligned} \mu_i(s,x) &= \bar{\mu}_i \frac{s_1}{K_i + s_1}\left(1 + \sum_{j\in[n]} a_{ij}x_j\right) \quad \forall i \in G_1 \\ \mu_i(s,x) &= \bar{\mu}_i \frac{s_2}{K_i + s_2}\left(1 + \sum_{j\in[n]} a_{ij}x_j\right) \quad \forall i \in G_2 \end{aligned} \tag{6}$$

The term $\left(1 + \sum_{j\in[n]} a_{ij}x_j\right)$ accounts for pairwise interactions affecting the growth rate of each OTU, while the term $\bar{\mu}_i \frac{s_j}{K_i + s_j}$ is a Monod growth expression, where $\bar{\mu}_i$ represents the maximum growth rate, and K_i the half saturation constant [15]. Note that if every $a_{ij} = 0$, then one recovers a classic substrate limited growth. Let A denote the matrix with entries a_{ij} hereafter referred to as the interaction matrix. Dumont et al. did not analyse their model but simply provided several simulations using parameter values identified from experimental data. The following section of this article deals with the mathematical analysis of model (3) with growth rates given by (6).

3. Mathematical Analysis

The system of Equation (3) is defined in the region

$$\Omega := \{(x_1, \ldots, x_n, s_1, s_2, s_3) \in \mathbb{R}^{n+3} | x_1, \ldots, x_n, s_1, s_2, s_3 \geq 0\}$$

First, sufficient conditions on the interaction matrix for the system to be well posed are established: meaning that solutions remain bounded and non-negative in time, this ultimately implies that the solution exists for every $t \geq 0$ [20].

Second, the equilibria of the system are derived. Possible equilibrium points for this system grow exponentially with the number of OTU considered (n). Stability is not analytically addressed, a numerical scheme calculating every equilibrium point and the system's Jacobian eigenvalues at the equilibrium point was implemented for studying the system.

3.1. Properties of the System

A bound on the norm of the interaction matrix that depends on the initial conditions and parameters one establishes that solutions will remain positive and bounded.

Lemma 1. *For initial conditions* $(x_1(0), \ldots, x_n(0), s_1(0), s_2(0), s_3(0))) \in \Omega$, *there exists positive scalars* M_1, M_2, *and* M_3 *such that solutions to* (3) *satisfy the following inequalities:*

$$\sum_{i\in G_1} \frac{1}{y_i} x_i + s_1 \leq M_1 \tag{7}$$

$$\sum_{i\in G_2} \frac{1}{y_i} x_i + s_1 + s_2 \leq M_2 \tag{8}$$

$$s_1 + s_2 + s_2 \leq M_3 \tag{9}$$

The proof can be seen in Appendix A. A bound on the norm of A is found such that every matrix A respecting the bound, guaranties that Ω is a positively invariant set.

Lemma 2. *For initial conditions* $(x_1(0), \ldots, x_n(0), s_1(0), s_2(0), s_3(0))) \in \Omega$, *there exists a constant* $M > 0$ *such that for every matrix* A *satisfying* $\|A\|_\infty \leq M$, *the solutions of system* (3) *with growth rates given by* (6), *remain in* Ω *and are bounded.*

The proof can be seen in Appendix A.

The importance of Lemma 2 is that by restricting the norm of matrix A the system is well-posed, meaning that the solutions can have biological and physical sense (there is no such thing as negative concentrations). Particularly, one has that $\|A\|_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^{n} |a_{ij}|$, which in this context implies that a bound on the sum of the absolute value of the interaction terms that affects each species allows the system to be well-posed. Note, however, this is a sufficient condition, thus the range of values matrix A can sustain for the system to remain well-posed may be considerably larger.

3.2. Equilibrium Points

In this section, analytical expressions for equilibrium points are shown. However, no analytic expression concerning the stability of such points is presented. In the following pages the reader will appreciate that the expressions of the equilibrium points are not simple, consequently replacing them in a 5×5 block matrix and calculating eigenvalues resisted an algebraic treatment. To answer the question of stability a numeric scheme is used by evaluating the Jacobian at the equilibrium point. At the end of the section an algorithm is provided for exploring all the possible equilibria. All the computations for deriving the equations of this section can be found in Appendix B.

Let $f(s)$ be such that,

$$f_i(s) = \begin{cases} \bar{\mu}_i \dfrac{s_1}{K_i + s_1} & \forall i \in G_1 \\ \bar{\mu}_i \dfrac{s_2}{K_i + s_2} & \forall i \in G_2 \end{cases} \tag{10}$$

Then $\mu(x, s) = \operatorname{diag}(f(s))(1_{n \times 1} + Ax)$

Thus, system (3) is rewritten as follows.

$$\dot{x} = \operatorname{diag}(\mu(x, s) - D_{n \times 1})x \tag{11}$$

$$\dot{s_1} = (s_{in} - s_1)D + Y_{1\bullet} \operatorname{diag}(\mu(x, s))x \tag{12}$$

$$\dot{s_2} = -s_2 D + Y_{2\bullet} \operatorname{diag}(\mu(x, s))x \tag{13}$$

$$\dot{s_3} = -s_3 D + Y_{3\bullet} \operatorname{diag}(\mu(x, s))x \tag{14}$$

Definition 1. *An equilibrium point (or steady state) is a point* $(x^{eq}, s^{eq}) \in \Omega$ *so that the right hand side of Equations* (11)–(14) *equals zero.*

Observe that equilibrium points are by definition non-negative so the state variables can have physical meaning. For studying the cases where x^{eq} contains zero valued entries, the set of non-active coordinates is defined as follows:

Definition 2. *Given an equilibrium point* (x^{eq}, s^{eq}) *of system* (3), *then the set of non-active coordinates* $\mathcal{J} \subset \{1, \ldots, n\}$ *is defined as:* $\mathcal{J} = \{j_1, \ldots, j_m : x^{eq}_{j_i} = 0,\ i \in [m]\}$. n_1^{act} *and* n_2^{act} *denote the number of positive entries of* x^{eq} *of functional groups* G_1 *and* G_2, *respectively.* $n^{act} = n - m$ *denotes the total number of positive entries of* x^{eq}. *The active point* $x^{act} \in \mathbb{R}^{n^{act}}$ *is defined by the positive entries of* x^{eq}. *Analogously, the functions* $f^{act}(s)$ *and* $\mu^{act}(x, s)$ *are defined by the positive entries of* x^{eq}. *The active interactions* A^{act} *is defined as the matrix* A *without the* $\mathcal{J}$ *rows and columns. The active stoichiometry matrix* Y^{act} *is the matrix* Y *without the* $\mathcal{J}$ *columns.*

In order to derive the equilibrium points, it is desirable an invertible A^{act} matrix. Therefore in what follows of the work it is assumed that matrix A and some of its submatrices have an inverse, this is stated properly in Hypothesis 2.

Hypothesis 2 (H2). *Let A be the interaction matrix of size $n \in \mathbb{N}$ and S be a proper subset of $[n]$ with $|S| = m$. Then the matrix $B \in \mathbb{R}^{(n-m)\times(n-m)}$ defined by taking out the S rows and columns of matrix A is invertible.*

Assuming Hypothesis 2 a formula for the active points is derived from Equation (11):

$$x^{act} = (A^{act})^{-1}(\text{diag}(f^{act}(s))^{-1}D_{n^{act}\times 1} - 1_{n^{act}\times 1}) \tag{15}$$

Note as well that at the equilibrium, s_3 can be defined in terms of s_1, s_2 and s_{in}. This is done by adding Equations (12)–(14) which gives:

$$s_{in} = s_1 + s_2 + s_3 \tag{16}$$

3.2.1. Both Functional Groups Are Present

The case where in each functional group remains at least one OTU is represented by Hypothesis 3.

Hypothesis 3 (H3). *The set $\mathcal{J}$ satisfies $G_1 \not\subset \mathcal{J}$, $G_2 \not\subset \mathcal{J}$.*

By replacing Equation (15) in Equation (12) s_2 can be written as a function of s_1:

$$s_2 = \frac{s_1}{b_1 s_1^2 + b_2 s_1 + b_3} \tag{17}$$

Then by replacing (17) in Equation (13), one gets a fourth degree polynomial for s_1.

$$a_4 s_1^4 + a_3 s_1^3 + a_2 s_1^2 + a_1 s_1 + a_0 = 0 \tag{18}$$

Formulae for coefficients $b_1, b_2, b_3, a_0, a_1, a_2, a_3, a_4$ can be found in Appendix B.

The equilibrium point can be calculated from the solutions of the system of Equations (15)–(18) with non negative coordinates. If the system only provides solutions with at least one negative entry then the set $\mathcal{J}$ cannot define an equilibrium point.

3.2.2. Washout of G_2

The washout of G_2 is equivalent to Hypothesis 4.

Hypothesis 4 (H4). *$G_2 \subset \mathcal{J}$ and $G_1 \not\subset \mathcal{J}$.*

Under this case note that $f^{act}(s)$ depends only on s_1. Therefore when Equation (15) is replaced in (12), one obtains a quadratic equation for s_1:

$$a_2' s_1^2 + a_1' s_1 + a_0' = 0 \tag{19}$$

where a_i' can be found in Appendix B.

Since $x_i = 0 \, \forall i \in G_2$ then from Equation (14).

$$\dot{s_3} = 0 = -s_3 D \tag{20}$$

$$\Rightarrow s_3 = 0 \tag{21}$$

In this case the equilibrium point can be calculated from the solutions of the system of Equations (15), (16), (A44) and (21) with non-negative coordinates. If the system only provides solutions with at least one negative entry then the set $\mathcal{J}$ cannot define an equilibrium point.

3.2.3. Washout

The washout equilibria means $x_i = 0$ for every $i \in \{1, \dots, n_1\}$. This is equivalent to Hypothesis 5. Note that the structure of a cascade reaction implies that if G_1 gets washed out, then so is G_2.

Hypothesis 5 (H5). $\mathcal{J} = G_1 \cup G_2$.

From Equation (12), one gets

$$s_{in} = s_1$$

then (16) implies

$$s_2 = s_3 = 0$$

The equilibrium is then given by $\begin{pmatrix} 0_{1\times n} & s_{in} & 0 & 0 \end{pmatrix}^\top$.

All the former discussion leads to a potential number of $(2^{n_1} - 1) \cdot (4 \cdot (2^{n_2} - 1) + 2) + 1$ different equilibria. Indeed:

$$\overbrace{(2^{n_1} - 1)}^{\substack{\text{nonempty subsets} \\ \text{of } G_1}} \Big(\underbrace{4}_{\substack{\text{possible} \\ \text{solutions of} \\ \text{Equation (18)}}} \cdot \overbrace{(2^{n_2} - 1)}^{\substack{\text{nonempty subsets} \\ \text{of } G_2}} + \underbrace{2}_{\substack{G_2 \\ \text{Washout}}} \Big) + \overbrace{1}^{\text{Washout}} \tag{22}$$

3.3. Stability: Operating and Ecological Diagrams

In this subsection the stability of the equilibrium points is addressed. Operating and ecological diagrams are created from this stability analysis. Both are an illustrative way of representing the long term behaviour of a reactor depending on operating parameters, namely D and s_{in}: In a $D - s_{in}$ plane different zones representing the stability properties of system (3) are identified [21].

For checking local asymptotic stability of the equilibrium points, the Jacobian of the system is provided and evaluated at each of these points. The resulting matrix's eigenvalues must have negative real part. A general formula for this Jacobian is presented in Appendix B (see (A70)).

Algorithm 1 summarizes this procedure.

Algorithm 1: Algorithm for Evaluating the Possible Equilibrium Points of System (3).

Data: $A \in M_{n\times n}(\mathbb{R})$, $D, s_{in}, \bar{\mu}_i, K_i, k_i, \in \mathbb{R}\ i \in [n],, n_1, n_2 \in \mathbb{N}$
Result: Set $\mathcal{P}$ containing all Positive Stable Equilibrium Points
$\mathcal{P} = \emptyset$
for $S \subset [n]$ **do**
  Calculate equilibrium (x, s_1, s_2, s_3) when $\mathcal{J} := S$ according to Hypotheses 3–5.
  if $(x, s_1, s_2, s_3) \geq 0$ **then**
    eig = Eigenvalues of $J(x, s_1, s_2, s_3)$
    if $Real(eig) < 0$ **then**
      $\mathcal{P} = \mathcal{P} \cup (x, s_1, s_2, s_3)$
    end
  end
end

Operating and ecological diagrams are created by running Algorithm 1 for different pairs (s_{in}, D). In the case of operating diagrams [22] (OD) all the pairs (s_{in}, D) are regrouped such that the points of the set $\mathcal{P}$ represent when partial nitrification (PN), complete nitrification (CN), washout (WO), or a combination of them may arise [23]. PN refers to the

state when nitrite (s_2) accumulates because the OTU of G_2 are washed out and thus no conversion from s_2 to s_3 takes place. On the contrary CN is when nitrate (s_3) accumulates because of the presence of OTU of G_2.

In the case of ecological diagrams (ED) the pairs are regrouped such that the points in the set $\mathcal{P}$ have the same non-active coordinates. In other words, instead of representing areas where either CN, PN or WO take place, ranges of pairs (s_{in}, D) where species coexist are represented. ED provide more information than OD, in the sense that one can deduce the latter from the former.

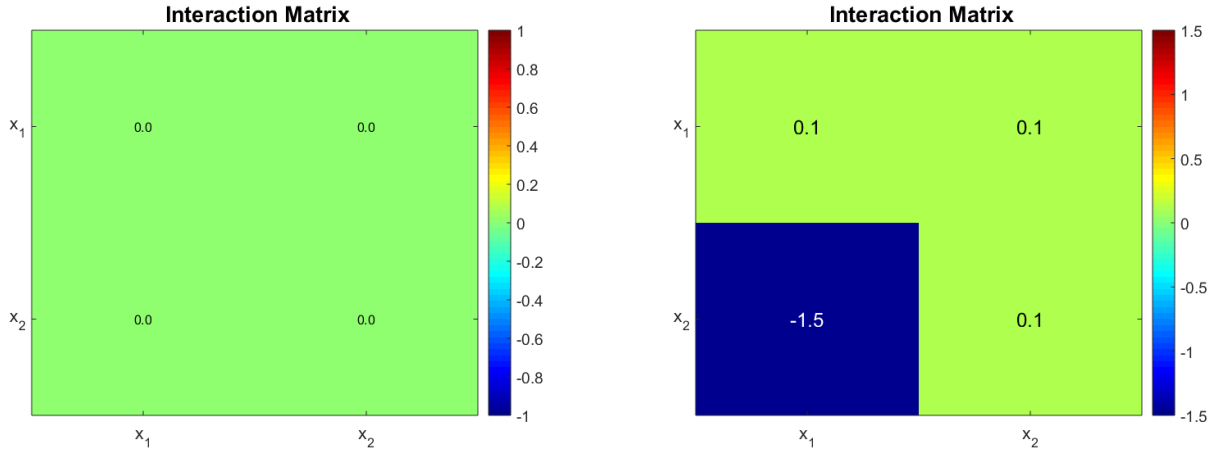
A first example using operating diagrams is presented to illustrate how adding interactions in a model consisting of 1 OTU in G_1 and 1 OTU in G_2 may lead to very different outcomes.

The important question of the existence of limit cycles was not resolved in this work. In the numerical analysis of this model at least one stable equilibrium was found for any choice of parameters. This obviously does not exclude the existence of limit cycles, but to what concerns the authors' intuitions there always seems to be at least one stable equilibrium point.

3.3.1. Case Study 1: 1 AOB and 1 NOB

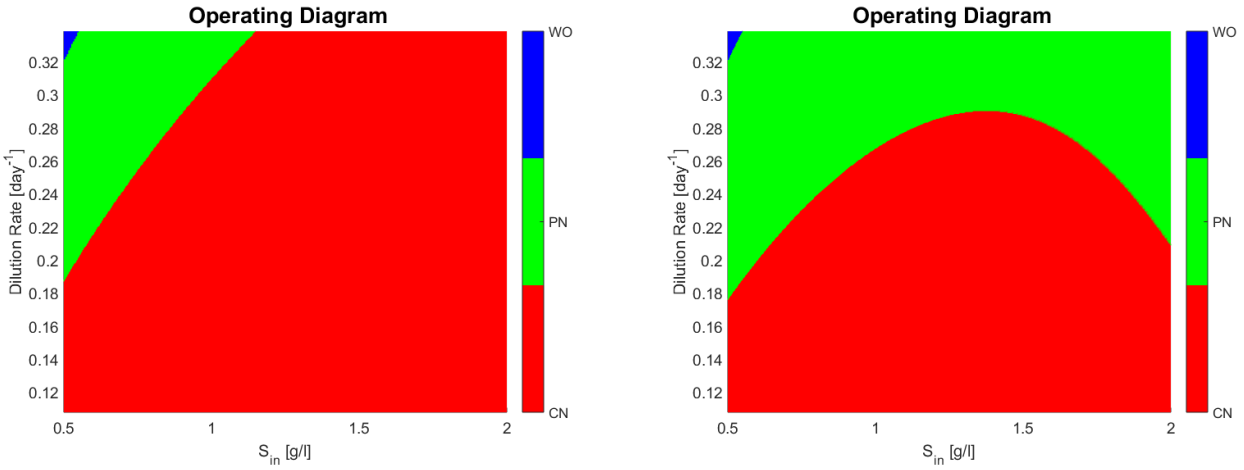
Consider the case where $n_1 = 1$ and $n_2 = 1$. The operating diagrams when no interactions take place ($A = 0$) and a non-zero interaction matrix are presented. When $A = 0$ Algorithm 1 is no longer valid (A is not invertible), nevertheless the stability analysis is much simpler and is given in the Appendix B section. The interaction matrices are shown in Figure 1a,b, the rational behind the second choice was to force a very strong interaction of x_1 on x_2 and observe its effects. The biological reason behind a negative microbial interaction might be the release of a toxin by x_1 that affects x_2 [1], or in this case it might represent competition for oxygen; we stress the fact that gLV interactions do not explicitly account for mechanistically anything they just try to represent an ecological relationship taking place. The rest of parameters can be seen in Table 2.

The operating diagrams can be seen in Figure 2, note how partial nitrification (washout of G_2) of Figure 2b is much bigger when compared to Figure 2a. The shape of the PN region in Figure 2b is somewhat unintuitive, because at a constant dilution rate (0.24 day^{-1} for example) and an increasing s_{in}, one passes from a PN zone, to a CN zone, and then back again to a PN zone. The mathematical explanation lies in the fact that x_1 also increases with s_{in}, and the affine part $(1 + a_{21}x_1 + a_{22}x_2)$ of the growth function of x_2 plays a bigger role than substrate limitation ($\frac{s_2}{s_2+K_2}$).



(**a**) (**b**)

Figure 1. Interaction matrices. Note how the presence of x_1 affects very negatively x_2 in (**b**), with respect to other interactions. The terms in the diagonal entries of the matrix represent intraspecies interactions, while the terms off the diagonal represent the interspecies interactions. (**a**) Interaction matrix of model (3) with no interactions. (**b**) A non-zero interaction matrix of model (3).



(**a**) (**b**)

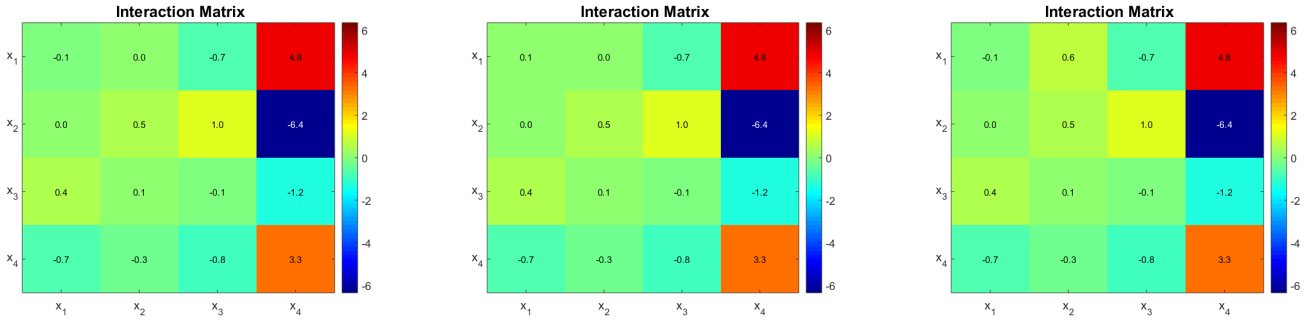
Figure 2. (**a**) Operating diagram of model (3) with no interactions (interaction matrix represented by Figure 1a). (**b**) Operating diagram of model (3) with interactions represented by Figure 1b. Note how (**b**) has a much larger zone where partial nitrification takes place. This is due to the negative interaction of x_1 on x_2.

Table 2. A set of kinetic parameters of model (3).

Kinetic Parameters	$\bar{\mu}_i$ [1/day]	K_i [g/L]	$\frac{1}{y_i}$ [gr/gr]
$x_1 \in G_1$	0.77	0.7	3.98
$x_2 \in G_2$	1.07	0.3	16.12

3.3.2. Case Study 2: 2 AOB and 2 NOB

Case study 2 is based on Dumont et al. [16] model parameters. They proposed a distribution of parameters obtained from a Bayesian estimation method. Their fit describes well the dynamics of the two most abundant OTU in each functional group, but it still fails to capture the measured substrates dynamics. Kinetic parameters of case study 2 can be seen in Table 3. The estimated interaction matrix is shown in Figure 3a. A second matrix is presented, which is obtained by the sign change of coefficient a_{11} (Figure 3b), and finally a third one is obtained by using a positive value for a_{12} (Figure 3c). The idea is to show that qualitatively different outcomes can be obtained by changing one interaction at a time.



(**a**) Originally calibrated interaction matrix.

(**b**) Modified interaction matrix with positive intraspecies interaction $a_{11} > 0$.

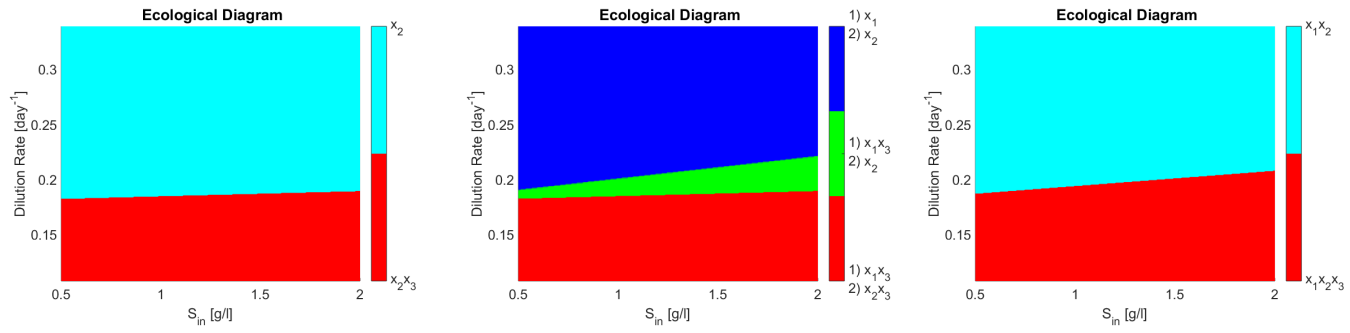
(**c**) Modified interaction matrix with positive interspecies interaction $a_{12} > 0$.

Figure 3. Interaction matrices for each case for a consortia of 4 bacterial species where x_1 and x_2 are AOB and x_3 and x_4 are NOB. Parameters a_{11} and a_{12} were modified in (**b**,**c**), respectively.

Table 3. Kinetic parameters of model (3) from Dumont et al. [16].

Case Study 2 Kinetic Parameters	μ_i [1/day]	K_i [mg/L]	$\frac{1}{y_i}$ [gr/gr]
$x_1 \in G_1$	0.828	0.147	3.85
$x_2 \in G_1$	0.828	0.147	3.85
$x_3 \in G_2$	0.18	0.026	100
$x_4 \in G_2$	0.18	0.026	100

The ecological diagrams are presented in Figure 4, where the legend indicates the species surviving in the zone of the respective colour. In Figure 4b, the system exhibits bi-stability (it is represented in numbering as (1) and (2) of the different possible equilibriums). Note how every zone in Figure 4b has two stable equilibria, meaning that the outcome of the system is determined by its initial conditions, particularly interesting is the green zone where either x_1, x_3 coexist or only x_2 remains, because in operational terms this means that either PN or CN may take place. When compared to Figure 4a, one can see that this change in the interactions of the microbial community can dramatically change the outcome of the reactor in a large operating zone.



(**a**) (**b**) (**c**)

Figure 4. Ecological diagrams (ED). The different zones represent the combination of surviving species in the steady state. PN takes place when neither x_3 nor x_4 are present. CN takes place if x_3 or x_4 are present. (**a**) ED from interaction matrix on Figure 3a. (**b**) ED from interaction matrix on Figure 3b. In the legend (1) and (2) represent the two different stable equilibria in each zone. (**c**) ED from interaction matrix on Figure 3c. Note that in (**b**) two stable equilibria exist for each zone.

One can see that coexistence in the same functional group is never attained in Figure 4a,b, whereas in Figure 4c x_1 and x_2, both AOB, coexist in either partial or complete nitrification. That means that the competitive exclusion principle [22] (CEP) does not hold. The CEP roughly states that if two species are growing on the same limiting resource, and their growth laws only depend non decreasingly on the resource, then only one of them will survive in the long run. This is interesting in light of reports on wastewater treatment plants where coexistence between species in nitrifying reactors has been shown [24], thus implying that a more complex growth law (as shown in here) or model structure involving other biological processes is required to include microbial diversity in mathematical models.

Remarks

Model (3) serves to illustrate that by considering a more complex growth rate that tries to model ecological interactions one might explain differences in reactors operating under similar conditions. It also shows a new mechanism by which the CEP no longer holds and which explains how multiple stable equilibria may appear. Since the gLV model discussed fails to completely capture the dynamics observed in the chemostat experiments [16], the next section proposes a new approach to study interactions.

4. Generalized Approach for Modelling Interactions

In the following an explanatory model (as opposed to a predictive model) is developed based on the hypothesis that interactions might be driving the nitrification process. In the previous sections interactions were modelled as an affine function of the OTU concentration that multiplies a substrate dependent growth equation. More generally the interaction function represents how the growth rate of species i is affected by the concentration of other species, x:

Given a vector $(v_1, \ldots, v_n)^\top$ the interaction function $\mathcal{I}$ is denoted as:

$$\begin{aligned} \mathcal{I} : \mathbb{R}_+^n \rightarrow & \quad \mathbb{R}_+^n \\ v \rightarrow & \quad \begin{pmatrix} I_1(v) \\ I_2(v) \\ \vdots \\ I_n(v) \end{pmatrix} \end{aligned} \tag{23}$$

Let $f_i(s)$ be a bounded, positive, and continuous function of s (e.g., Monod, Haldane). The growth equation of OTU i becomes:

$$\mu_i(s, x) = f_i(s) I_i(x) \tag{24}$$

Note $f(s) := (f_1(s), \ldots, f_n(s))^\top$.

Since the growth of a single strain in batch experiments is driven by the substrate concentration, when no interactions are present one should recover expression $f_i(s)$. Therefore if there are no interactions then $I_i(x) = 1$. From this hypothesis, note that $\lim_{x \to 0} I_i(x) = 1$ since if there is minimal presence of OTU, interactions cannot exist. Furthermore for this study it is assumed that $I_i(\cdot)$ is a continuously differentiable function on x. For making explicit all of the former:

Hypothesis 6 (H6). *The interaction function $\mathcal{I}$ previously defined satisfies:*

1. $\mathcal{I}\big((0, \ldots, 0)^\top\big) = (1, \ldots, 1)^\top$
2. *There is an open set $\Omega \subset \mathbb{R}^n$ such that $\mathcal{I} \in C^1(\Omega)$.*

Note $J_I(x)$ the Jacobian matrix of function $\mathcal{I}$, then a first order approximation of $\mathcal{I}(\cdot)$ centred at $\bar{x}$ gives: $\mathcal{I}(x) = \mathcal{I}(\bar{x}) + J_I(\bar{x})(x - \bar{x}) + o(\|x - \bar{x}\|)$. When $\bar{x} = 0$ one recovers the growth expression from the previous section (Equation (6)) implying that $J_I(\bar{0})$ can be seen as the interaction matrix from model (3).

4.1. Unravelling the Interaction Function

Suppose that the functions $f_i(s)$, and the yields y_i are well-known. By using experimental measurements of x, represented by $z(t)$, the objective is to reconstruct function $I(x)$. For doing so, the terms $I_i(x)$ are replaced by controls $u_i(t)$, thus $I_i(x(t)) = u_i(t)$. A control law is obtained by solving a nonlinear optimal tracking problem.

Consider the observable system (25), with $y(t) = x(t)$ being the output, because we are observing measurements coming from genetic sequencing.

$$\begin{aligned}
\dot{x_i} &= (f_i(s)u_i(t) - D)x_i \quad \forall i \in G_1 \\
\dot{x_i} &= (f_i(s)u_i(t) - D)x_i \quad \forall i \in G_2 \\
\dot{s_1} &= (s_{in} - s_1)D + \sum_{i \in G_1} y_{s_1/x_i} f_i(s)u_i(t)x_i \\
\dot{s_2} &= -s_2 D + \sum_{i \in G_1 \cup G_2} y_{s_2/x_i} f_i(s)u_i(t)x_i \\
\dot{s_3} &= -s_3 D + \sum_{i \in G_2} y_{s_3/x_i} f_i(s)u_i(t)x_i \\
y &= x
\end{aligned} \tag{25}$$

Consider the weighted norms defined by positive definite matrices Q and R, represented by $\|\cdot\|_Q$ and $\|\cdot\|_R$, respectively, and $\bar{u} > 0$. The optimal tracking problem is defined as:

$$\begin{aligned}
\min \quad & \int_0^T \|y - z\|_Q + \|(u - \vec{1})\|_R dt \\
s.t. \quad & (x, s_1, s_2, s_3) \text{ solution of (25)} \\
& u_i(t) \in [0, \bar{u}]
\end{aligned} \tag{26}$$

The control $u(t)$ is intended to drive the system to be near a desired output $z(t)$, which in this context are the measurements of the concentrations of OTU. The term $\|(u - \vec{1})\|_R$, was added for two reasons:

- First, because the interest is testing the idea that interactions could be driving the system. Therefore adding a penalization in the objective function for each control to remain near 1 can be seen as an attempt to explain data without any interaction. In other words, if the control terms are found to drift from 1, it means that interactions are necessary to explain the system dynamics.
- Second, to force a regularized control. Otherwise note that u is linear in (25), therefore if the integral cost does not have a non-linear expression of u the optimal control will be of a bang-bang type with possibly singular arcs [25]. Since the objective is to find a differentiable expression of $I(x)$ the addition of the regularization term is deemed necessary.

The problem of approximating the solution of the system to a desired reference (z in this case) is called the optimal tracking problem. For solving such a problem the approach developed by Cimen et al. [17,26] was adapted to our problem. The method proposed involves the resolution of Approximating Sequences of Ricatti Equations (ASRE). It consists of iteratively calculating trajectories of System (25) with a certain control law to later feed a non-autonomous Ricatti differential equation with the resulting trajectory. Then, a new control law that uses the solution of the Ricatti equation is proposed and a new trajectory is calculated. The iteration is stopped when a convergence in the output or the control is observed.

The control term should remain positive for the system to be well posed (no negative states), and an upper bound was added to represent the fact that life cannot grow infinitely fast. The tracking problem does not consider a constrained control. Nevertheless, the methods of Cimen et al. [26] were directly used with an explicit constraint in the synthesis of the control. Even though this is probably suboptimal when the control reaches its bounds, one at least is certain about its optimality when the control never reaches its constraints.

Another departure from their method is that in their formulation an approximation of the dynamics for calculating the trajectories is used. They proved that such a linearisation converges to the original dynamics. In the case here presented the linearisation of the dynamics was not necessary.

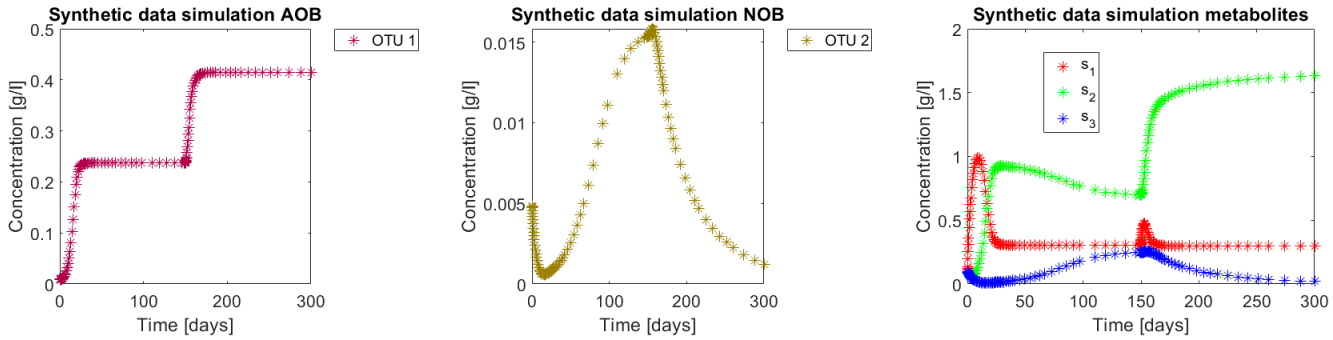
The change of variable $u_i(t) = v_i(t) + 1$ is used for technical reasons explained in Appendix C.

This in turn implies $v_i(t) \in [-1, \bar{u} - 1]$. The feedback control $v(t)$, will be of the form $v(t) = -R^{-1}\tilde{B}^\top(t)\Big(\tilde{P}(t)x(t) - s_f(t)\Big)$. Where matrix $P(t)$ and vector $s_f(t)$ solve differential equations. In Appendix C it is proved that, thanks to the structure of the system, one only needs to calculate $2n$ differential equations for the synthesis of the control, instead

of $(n+3)^2 + (n+3)$ that would imply the direct application of the method, which renders the method—at least theoretically—scalable for a growing number of OTU.

4.2. Proof of Concept

The approach was tested with data generated by simulating model (3) using the parameters of the case study 1 (Table 2) with interaction matrix given by (Figure 1b). In the operating diagram of the same case (Figure 2b) the red zone implies complete nitrification, while the green zone means partial nitrification. For integrating the former phenomena in the simulation, the system was simulated for 300 days, and perturbed at day 150 from the CN zone ($(s_{in}, D) = (1.25, 0.24)$) to a PN zone ($(s_{in}, D) = (1.95, 0.24)$). Simulations can be seen in Figure 5. Note how from day 150 the NOB population (OTU 2) represented in Figure 5b decreases, which in turn implies a decrease in s_3, as seen in Figure 5c. In the case where no interactions take place, the OD seen in Figure 2a implies that s_3 would have accumulated all along the trajectory, since the perturbation still remains in the CN zone.



(**a**) AOB. (**b**) NOB. (**c**) Metabolites.

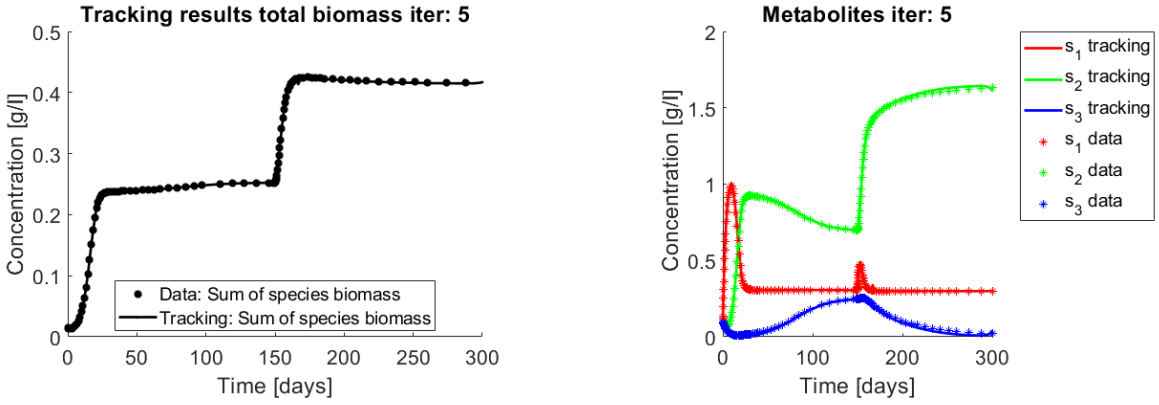
Figure 5. Synthetic data generated by model (3), with parameters from case study 1. Note the effects of the increased input s_{in} generated in day 150.

For the tracking procedure the functions $f_i(s)$ and the yields y_i were the same as those used for simulating the synthetic data (parameters in Table 2) and the control is meant to account for the interaction term. The Q and R matrices were $\begin{bmatrix} \lambda_1 I_{n_1} & 0 \\ 0 & \lambda_2 I_{n_2} \end{bmatrix}$ and I_n, respectively, with $\lambda_1 = 10^{-4}$ and $\lambda_2 = 10^{-5}$ in order to better track the NOB trajectories, since they are less abundant. The values were obtained by trial and error, by using a single λ for both functional groups, beginning with $\lambda = 1$, in which case one can see how the optimal control becomes $u = 1$, thus no tracking is performed. Further diminishing the value from 10^{-5} adds too much noise to the control, without significant gains on the quality of the tracking.

The results of the procedure to identify interactions can be seen in Figures 6 and 7. Figures 6a and 7a,b show the total biomass concentration, and the trajectories for the OTU belonging to G_1 and G_2, respectively. It can be seen that the method approaches well the trajectories of the OTU, with a better result for the AOB community, which can be explained by the one order of magnitude difference in their concentrations (which in turn is a consequence of the one order of magnitude difference in their yields). The metabolites concentration represented in Figure 6b are in accordance with the simulated: The method is able to reconstruct the metabolites trajectories from the community measurements.

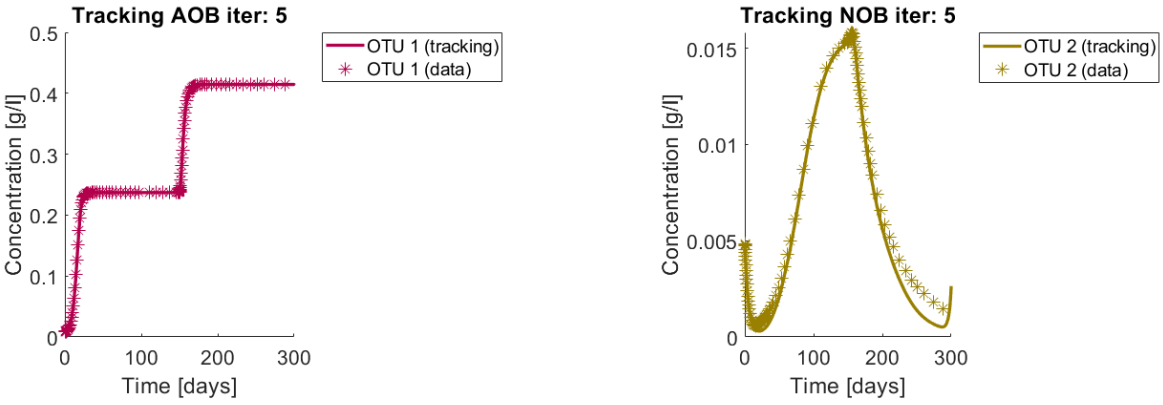
Figures 8 and 9 show the controls and the corrected growth rate for each functional group. The control for each functional group can be seen in Figures 8a and 9a. Note that from the structure of a quadratic regulator, since there is no cost in the final state, the end value is always 1. Figures 8b and 9b show the resulting growth rates for AOB and NOB, respectively, without the control $u_i(t)$. Figures 8c and 9c are the complete expression that determines growth rate, that is $f_i(s(t))u_i(t)x_i(t)$. Note how little the shape changes with respect to Figures 8b and 9b, which might mislead the reader to conclude that the control

had reduced effects in the dynamics. The way out of this conundrum is to remember that the control's effects are already included in $x_i(t)$ and $s_i(t)$, and thus in expression $f_i(s(t))x_i(t)$.



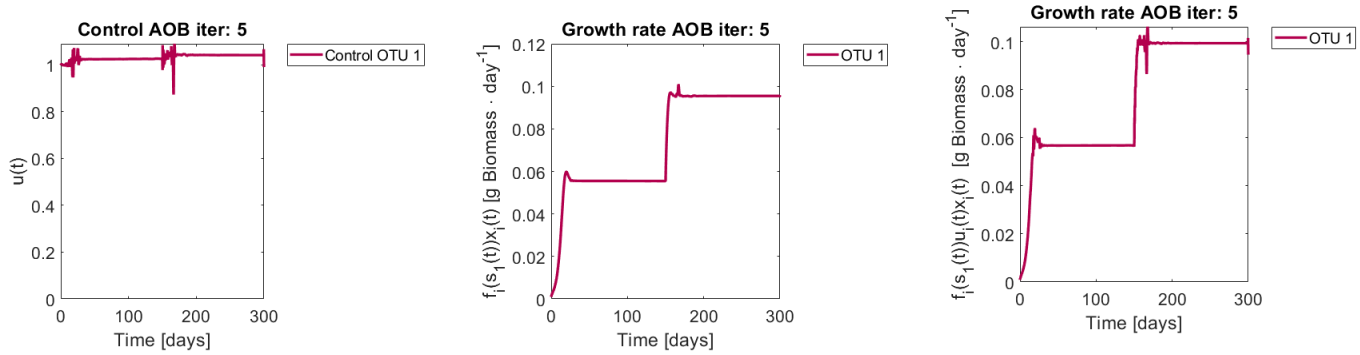
(**a**) Total biomass. (**b**) Metabolites.

Figure 6. Asterisks represent the synthetic data, while the continuous lines represent the method's output. The method is able to reconstruct the metabolites pattern, from the biomasses concentrations.



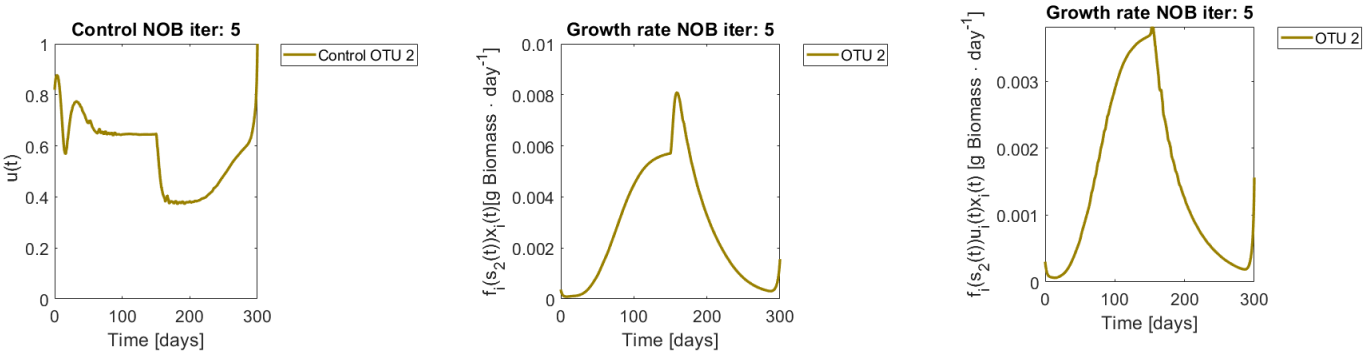
(**a**) AOB biomass. (**b**) NOB biomass.

Figure 7. Asterisks represent the synthetic data, while the continuous lines represent the method's output. The method reconstructs a continuous trajectory from the synthetic data.



(**a**) Control for the AOB community. (**b**) Growth rate "without" the control. (**c**) Complete reconstructed growth rate for the AOB community.

Figure 8. Obtained control and reconstructed growth rate for OTU 1 (AOB).



(**a**) Control for the NOB community.

(**b**) Growth rate "without" the control.

(**c**) Complete reconstructed growth rate for the NOB community.

Figure 9. Control and reconstructed growth rate for OTU 2 (NOB).

A final comment on the identifiability of the interaction terms. Even though one might propose a growth rate with the tracking control $u(t)$ that accurately replicates the OTU trajectory $x(t)$, retrieving the original interaction coefficients from the obtained control for this example was not possible. The former was tried by minimizing function $f(A) = \left\| \int_0^T u(t) - (1 + Ax(t))dt \right\|$ with a non linear optimization solver for a 1000 initial random guesses for matrix A. If one also takes into account $\bar{\mu}_i$ and K_i as parameters to fit this adds even more degrees of freedom, thus suggesting that the identifiability of growth functions (6) in model (3) might be very low.

5. Application

The tracking problem was applied to data coming from a nitrification process with experimental conditions described in [27]. For exploring the hypothesis of interactions as drivers of bioreactors performance environmental conditions should be kept as constant as possible. Therefore only data from day 183 onwards was used because a change in the operating temperature happened at that point, which is known to have an effect on kinetics. For choosing which species belong in which functional group, the procedure described of Ugalde-Salas et al. [28] was used. From day 183 to day 315, 31 OTU were identified in the G_1 group (AOB) and 5 in the G_2 functional group (NOB).

A first example of the procedure is performed when the classified OTU are regrouped in their assigned functional groups by adding their concentrations. A 5 dimensional dynamical system is obtained, thus there are only two interacting functional biomasses: this case is structurally the same as in the proof of concept, but here a real dataset is used. The same procedure is applied where no regrouping occurs and the system state grows to 39.

The knowledge of functions $f_i(s)$ was based on a study of nitrification's kinetic parameters [29]. Particularly given the system's ammonium and nitrite concentration a Monod function (Equation (27)) was used for G_1 and G_2 with parameters given in Table 4 calculated from the equation of Table 2 of the same article. The yields were fitted to match the nitrogen mass balances. The Q and R matrices were the same as in the proof of concept section, that is $\begin{bmatrix} \lambda_1 I_{n_1} & 0 \\ 0 & \lambda_2 I_{n_2} \end{bmatrix}$ and I_n, respectively, with $\lambda_1 = 10^{-4}$ and $\lambda_2 = 10^{-5}$, because data lie in the same order of magnitude than synthetic data.

$$f_i(s) = \bar{\mu}_2 \frac{s_1}{K_1 + s_1} \; \forall i \in G_1 \qquad (27)$$
$$f_i(s) = \bar{\mu}_2 \frac{s_2}{K_2 + s_2} \; \forall i \in G_2$$

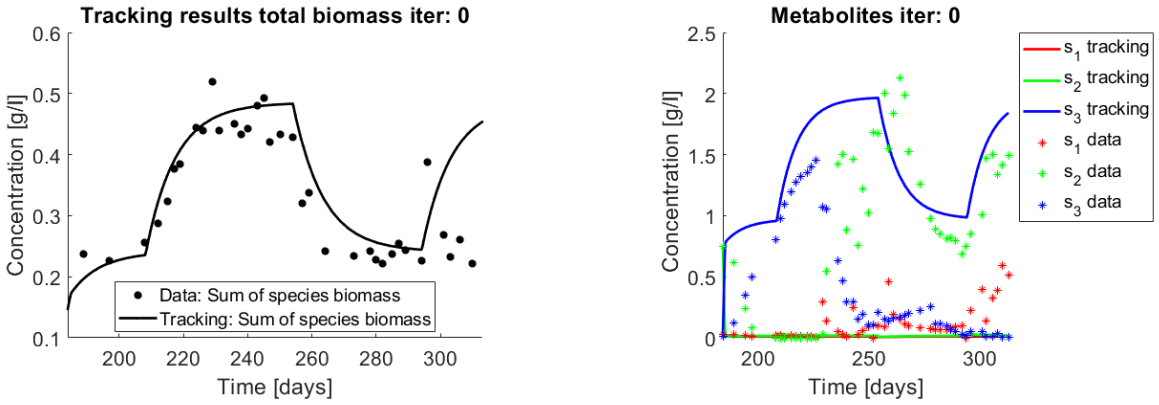
Table 4. A set of kinetic parameters of model (25).

Kinetic Parameters	μ_i [1/day]	K_i [g/L]	$\frac{1}{y_i}$ [gr/gr]
$x_1 \in G_1$	1.97	7×10^{-1}	4.49
$x_2 \in G_2$	1.87	5.4×10^{-1}	45.51

For the reader to gain understanding of the situation, a simulation of the system using the experiments operating parameters (D and s_{in}) is presented without control (i.e., $u(t) = 1$) in Figure 10 nitrate (s_3) accumulates all along the trajectory, but when compared to data it is clear that s_3 stops accumulating after a while.

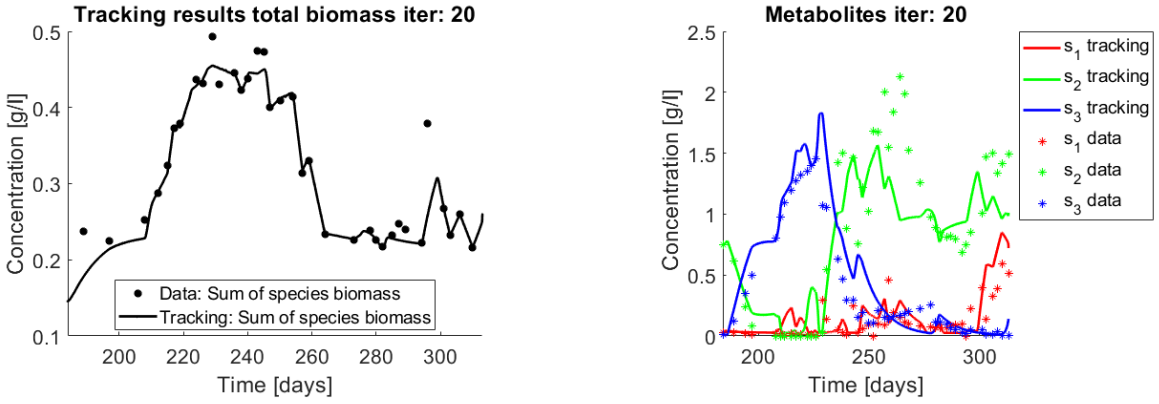
When applying the tracking method one obtains the simulation that can be seen in Figures 11 and 12. The method captures the tendencies of the measured substrates as seen in Figure 11b. The tracking of each functional group G_1 (AOB), and G_2 (NOB) can be seen in Figure 12a,b, respectively.

The growth rates of each functional group are shown in Figure 13. Note in the case of AOB (Figure 13a) the resulting growth rate shows a noisy curve formed by pulses. The behaviour of the NOB community (Figure 13b) is qualitatively very similar with somewhat stronger pulses and less noise. The former is to be expected since more OTU were regrouped to compose the AOB biomass, therefore more noise sources were added.



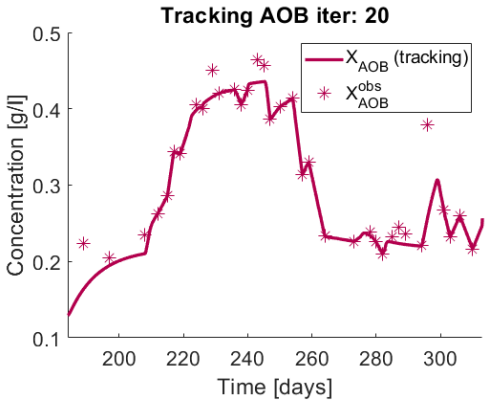
(**a**) Total Biomass. (**b**) Metabolites.

Figure 10. Simulation of system (25) when $u = 1$, with functions as in (27). Data points are represented by a star. The continuous line represents the simulation.

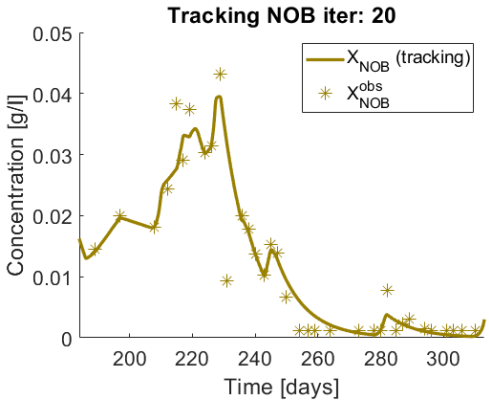


(**a**) Total Biomass. (**b**) Metabolites.

Figure 11. Results on applying the tracking method to a nitrification experiment when regrouping OTU in their functional groups. Data points are represented by asterisks. The continuous line represents the tracking procedure results.

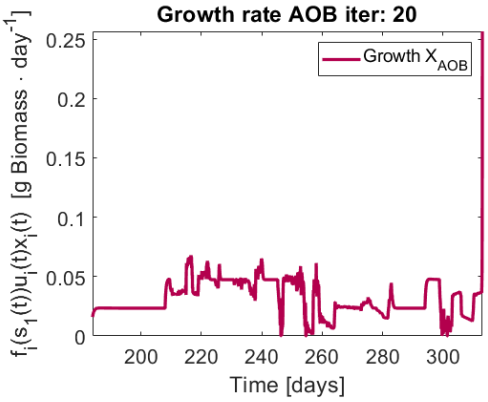


(**a**) AOB biomass.

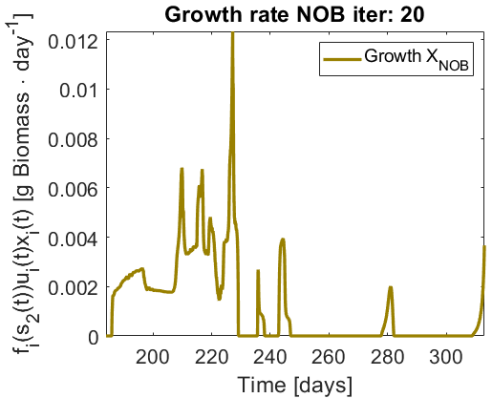


(**b**) NOB biomass.

Figure 12. The tracking procedure applied to the observed biomass (asterisks) regrouped in two functional groups.



(**a**) Growth rate of the model for AOB community.



(**b**) Growth rate of the model for the NOB community.

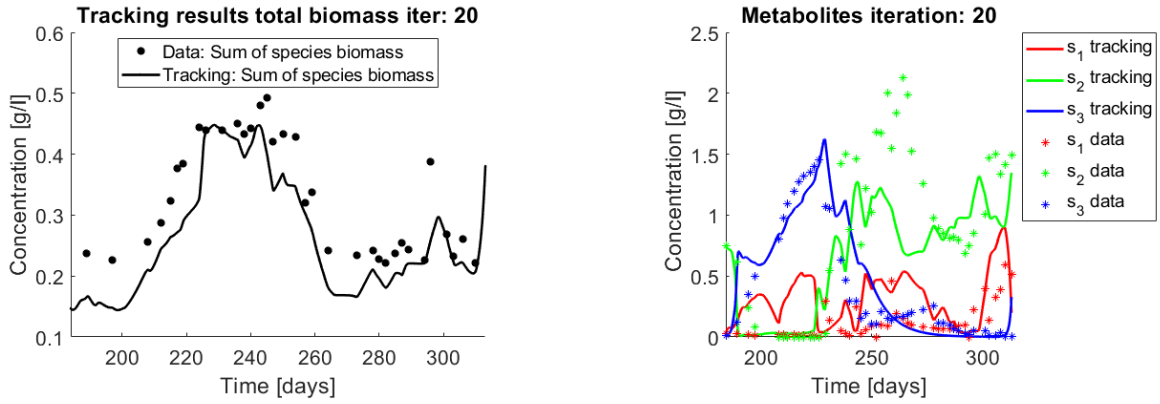
Figure 13. Obtained growth rates when regrouping OTU in their functional groups.

The same procedure is applied without regrouping. The results on total biomass and metabolites are shown in Figure 14a,b, respectively. Both patterns still fit the data, but to a lesser degree of precision when compared to Figure 11. This can be explained by inspecting Figures 15–20. First note the absolute error of the tracking for each of the OTU in the AOB community (Figures 15b–19b), almost every point lies below 0.015 [g/L], implying that the method might not be able to track below that threshold for the members of the AOB community. The former ultimately implies that the most abundant OTU are better tracked, thus the information contained in the least abundant species is not integrated in the model. Notice that the error for the NOB community is lower (Figure 20b), almost every point lies below 0.005 [g/L], this can be explained in the one order of difference in the entry of matrix Q for the AOB and the NOB community. It may be the case that using appropriate weight matrices that account for the difference between OTU abundances could help in this aspect; in that sense only one rational was tested (inverse of the mean abundance of each OTU in the diagonal entries of matrix Q) and did not improve the results. When looking at the growth rates (Figures 15c–20c) one again observes pulses for each OTU. Finally, note that most OTU were present only for a fraction of the experiment's duration.

In both cases, the regrouped and individual tracking, the growth rate varies strongly, raising the question whether the observed pulses are emerging from interactions within the microbial community. When growth rates are compared to the proof of concept section it seems doubtful that a linear pairwise interaction model such as the gLV model could capture the complexity of the particular chemostat analysed. Perhaps these interactions are not constant through time (as opposed to the gLV model) or a different interaction function should be thought of. However the former questions cannot be fully clarified here,

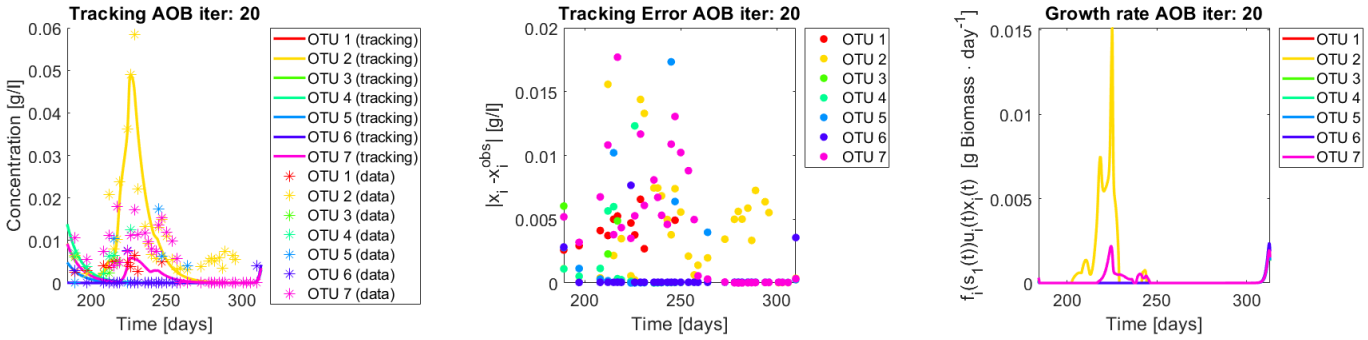
because the quality of the genetic sequencing from molecular fingerprints might not be the best when compared to more recent techniques, thus it is unclear if the pulses are due to noise of the measurements.

The interpretation of the correction term as interactions is not the only possible reading. In other contexts the correction term might also be interpreted as a non accounted phenomena ranging from environmental factors (e.g., temperature, pH) to other biological factors (viruses, flock formation, pathogens). Alternative hypotheses for explaining the observed patterns in the microbial community should be considered as well.



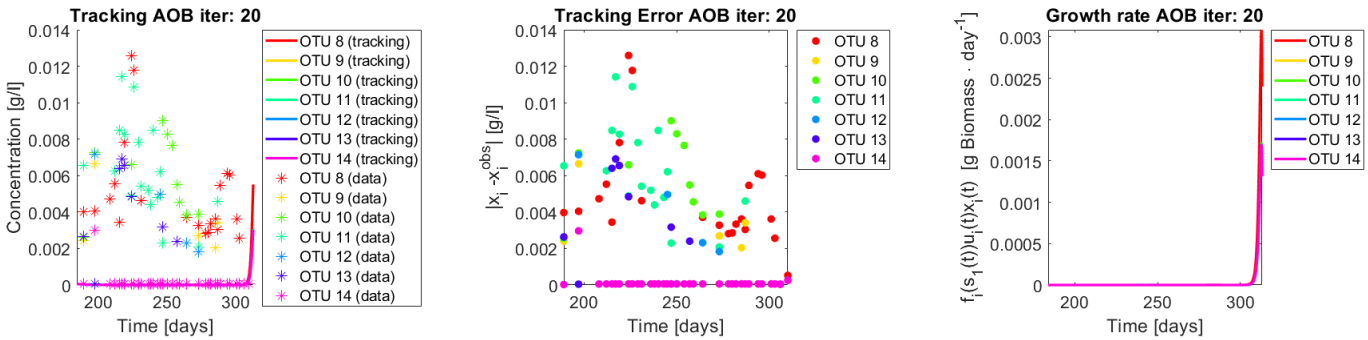
(**a**) Total Biomass. (**b**) Metabolites.

Figure 14. Results on applying the tracking method to a nitrification experiment when all OTU are tracked independently. Data points are represented by a star. The continuous line represents the tracking procedure results.



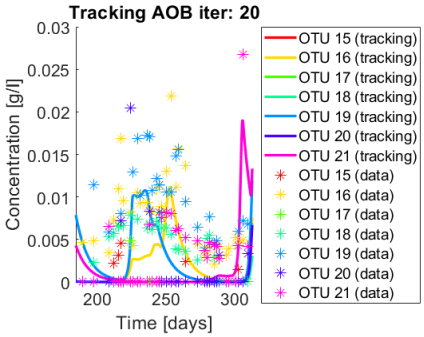
(**a**) Tracking results and data. (**b**) Absolute difference between tracking and data. (**c**) Reconstructed growth rate.

Figure 15. Results for OTU 1-7 (AOB).

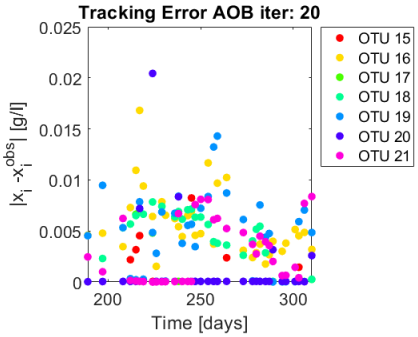


(**a**) Tracking results and data. (**b**) Absolute difference between tracking and data. (**c**) Reconstructed growth rate.

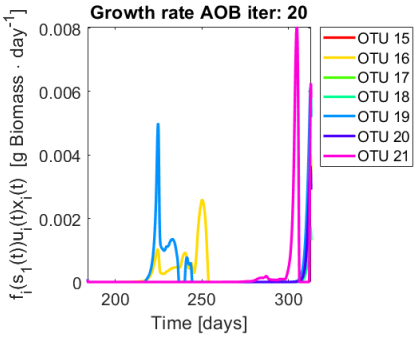
Figure 16. Results for OTU 8-14 (AOB).



(**a**) Tracking results and data.

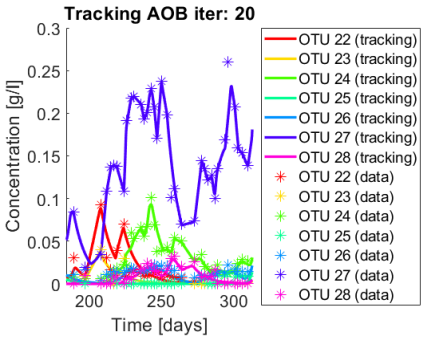


(**b**) Absolute difference between tracking and data.

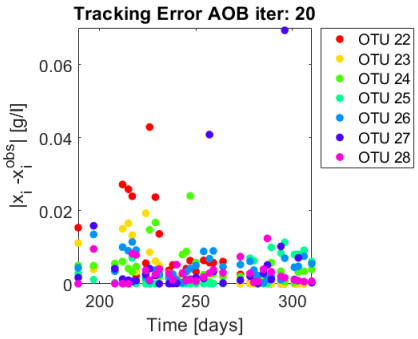


(**c**) Reconstructed growth rate.

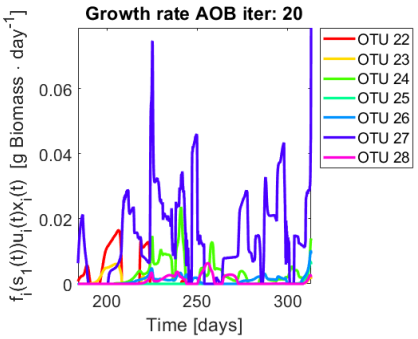
Figure 17. Results for OTU 15-22 (AOB).



(**a**) Tracking results and data.

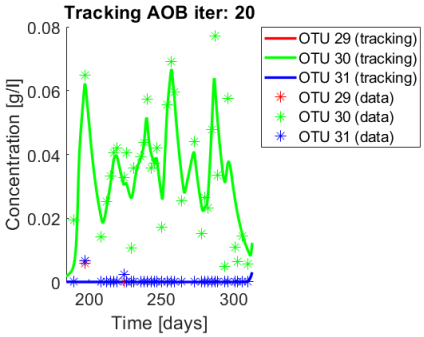


(**b**) Absolute difference between tracking and data.

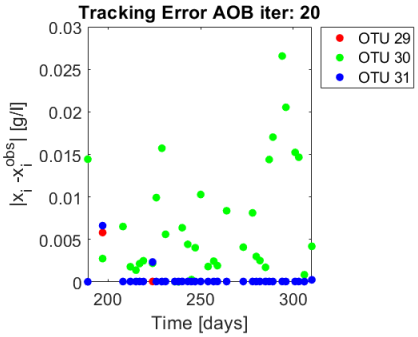


(**c**) Reconstructed growth rate.

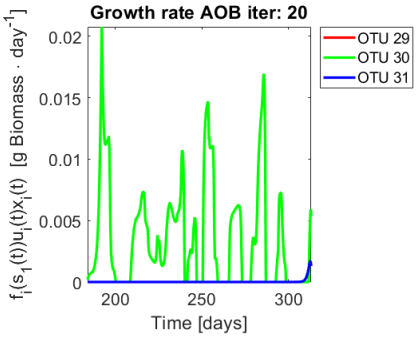
Figure 18. Results for OTU 23-28 (AOB).



(**a**) Tracking results and data.

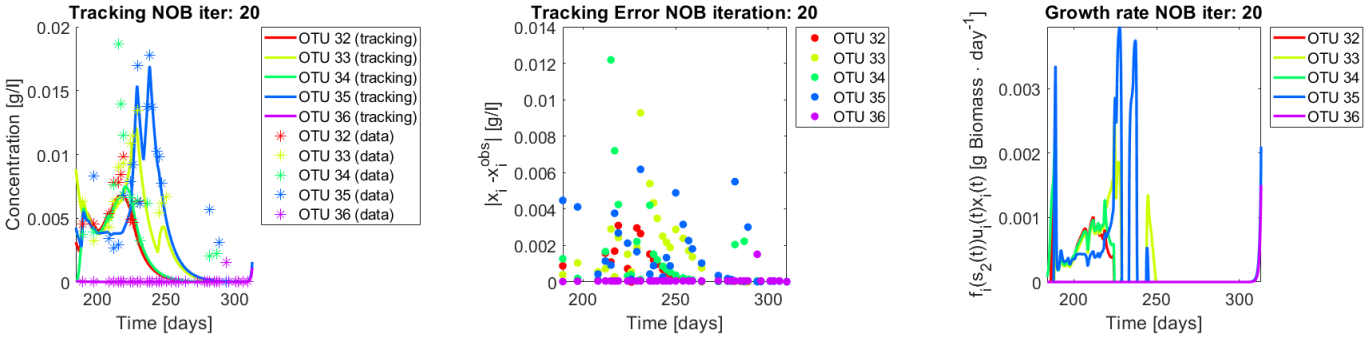


(**b**) Absolute difference between tracking and data.



(**c**) Reconstructed growth rate.

Figure 19. Results for OTU 29-31 (AOB).



(**a**) Tracking results and data.

(**b**) Absolute difference between tracking and data.

(**c**) Reconstructed growth rate.

Figure 20. Results for OTU 32-36 (NOB).

6. Conclusions and Perspectives

Over the last decades, advances in genetic sequencing and microbial ecology have opened a gap for modellers in biochemical processes to integrate this valuable information. Considering the success of mass balance models to predict and pilot bioreactors, new models should be built upon them, or at least be compared against them. This article exposed what can be gained from combining population-based models as used in ecology with functional group based approaches as used in bioengineering. The analysis of the gLV model proposed by Dumont et al. [16] already shows that such a combination can give way to models that include bi-stability, coexistence within a functional group, and unintuitive operational insights such as raising the input ammonium s_{in} to achieve partial nitrification. The increased number of parameters of this particular model obviously hinders its potential application, but it surely helps to illustrate what can be gained by joining both types of models. The mathematical analysis focused on the particular case of pairwise interactions, which can be seen as a first order approximation of the introduced concept of interaction function. This opens the question for a broader class of interaction functions that could well represent complex microbial ecosystems, particularly bioreactors.

With that line of reasoning, in order to understand what this interaction function should look like, a data-driven approach was presented. It can be simply described as correcting the growth rate expression of each individual species in a mass balance model by explicitly assuming a control loop on the growth rate depending on the species state variables and the measured abundances. The reconstructed growth rates seem to consist of pulses, suggesting that a form of a possible interaction function should reproduce this behaviour. However, the former questions cannot be fully clarified here, because the identification and quantification of OTU from single-strand conformation polymorphism (SSCP) as performed by Dumont et al. [18] might not be optimal. Indeed, the determination of OTU from SSCP profiles rely on the identification and quantification of peaks and miss many OTU that appear only as background noise [30]. This type of fingerprinting is thus less accurate than more recent methods such as the sequencing of the 16S ribosomal RNA. In spite of this, one is able to recover the substrates dynamics, implying that the hypothesis of microbial interactions as drivers of a bioprocess, in the form of feedback loops affecting each others growth rates, is not far-fetched.

The use of the tracking technique can be applied in a straightforward manner to already existing models in mixed homogeneous bioreactors, under the condition that the microbial species have already been identified with a particular functionality of the system. Even though the tracking model was proposed for a chemostat setting as a way to correct a substrate limited expression, the method could be used in contexts where less information on the growth function of microbes is known. If one supposes nothing on the growth expression but the fact that is bounded (life cannot grow infinitely fast) the model becomes a linear model and one recovers a classic quadratic regulator for linear systems. This was tested for the data presented in this article and, unfortunately, negative substrate appeared

as an output, suggesting that the substrate limitation term is crucial for the model to be well-posed. Nevertheless, even in the former case, the synthesis of the optimal bounded control remains an open theoretical challenge. One might bypass this issue of the current control scheme by, for example, a very thorough use of the Pontryagin maximum principle for the synthesis of the control. In a more general view the reconstruction of the growth function in chemostat systems is already subject to problems of identifiability [31], integrating genetic sequencing could provide a path for more certainty in model calibration.

Author Contributions: Conceptualization J.H., P.U.-S. and E.D.-L.Q., Formal analysis P.U.-S. and H.R.C., Supervision J.H. and E.D.-L.Q. Writing—review & editing J.H., P.U.-S. and E.D.-L.Q. Software P.U.-S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was financed by project Thermomic ANR-16-CE04-0003 (France) and FONDECYT grant 1200355 and Basal Program CMM-AFB 170001, both from ANID (Chile).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data can be recovered from the following repository https://github.com/paus-5/Class-and-Track (accessed on 24 February 2021).

Acknowledgments: The authors would like to thank Alain Rapaport for the enriching exchanges and comments.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AOB	Ammonia oxidizing bacteria
NOB	Nitrite oxidizing bacteria
OTU	Operational Taxonomic Unit

Appendix A. Proofs of Properties of the System

Lemma A1. *For initial conditions* $(x_1(0),\dots,x_n(0),s_1(0),s_2(0),s_3(0))) \in \Omega$, *there exists positive scalars* M_1, M_2, *and* M_3 *such that solutions to* (3) *satisfy the following inequalities:*

$$\sum_{i\in G_1}\frac{1}{y_i}x_i + s_1 \leq M_1 \tag{A1}$$

$$\sum_{i\in G_2}\frac{1}{y_i}x_i + s_1 + s_2 \leq M_2 \tag{A2}$$

$$s_1 + s_2 + s_2 \leq M_3 \tag{A3}$$

Proof. Define $z_1 := \sum_{i\in G_1}\frac{1}{y_i}x_i + s_1; z_2 := \sum_{i\in G_2}\frac{1}{y_i}x_i + s_1 + s_2; z_3 := s_1 + s_2 + s_3$. Computing $\dot{z_1}$ one gets:

$$\dot{z_1} = \sum_{i\in G_1}\frac{1}{y_i}\dot{x_i} + \dot{s_1} = D\left(-\sum_{i\in G_1}\frac{1}{y_i}x_i - s_1 + s_{in}\right) = D(s_{in} - z_1)$$

Define $\bar{s} = \max\{s_{in}(t)|t \geq 0\}$ and consider the differential equation:

$$\dot{w} = D(\bar{s} - w); \quad w(0) = \sum_{i\in G_1}\tfrac{1}{y_i}x_i(0) + s_1(0) \tag{A4}$$

If there is a time interval H such that $t^* \in H \Rightarrow D(t^*) = 0$, then z_1 is constant and therefore bounded by the value of the solution in $z(t^*)$. If $D(t) > 0$, then $w = \bar{s}$ is

a stable asymptotic equilibrium. Define $M_1 := \max\{w(0), \bar{s}\}$. If $w(0) > \bar{s}$ then by the Picard-Lindeloff theorem $\forall t\ \dot{w}(t) < 0$, otherwise it would cross the solution of the initial value problem (A4) with starting point $\bar{s}$, therefore $w(t) \geq \bar{s}$. The same reasoning may be applied if $w(0) < \bar{s}$. One concludes that $w(t) \leq M_1$.

Now consider

$$\dot{z_1} = D(s_{in} - z_1); \quad z_1(0) = \sum_{i \in G_1} \frac{1}{y_i} x_i(0) + s_1(0) \tag{A5}$$

From a comparison lemma (see chapter 3 [20]), the solution of (A5) is bounded by w and therefore $z_1(t) \leq M_1$.

Define $M_2 = \max\left\{ \sum_{i=n_1+1}^{n} \frac{1}{y_i} x_i(0) + s_1(0) + s_2(0), \bar{s} \right\}$, and $M_3 = \max\{s_1(0) + s_2(0) + s_3(0), \bar{s}\}$ and noting that $z_2(t)$ and $z_3(t)$ satisfies the same differential equations as z_1 the above reasoning may be applied and one has the desired bounds. □

Lemma A2. *For initial conditions $(x_1(0), \ldots, x_n(0), s_1(0), s_2(0), s_3(0))) \in \Omega$, there exists a constant $M > 0$ such that for every matrix A satisfying $\|A\|_\infty \leq M$, the solutions of system (3) with growth rates given by (6), remain in Ω and are bounded.*

Proof. If any coordinate of the solution becomes zero, then it's derivative is either zero, or positive. The bound will follow from Lemma 1.

Note that for any $i \in [n]$, if $x_i = 0$ then $\dot{x_i} = 0$, therefore all solutions of system (3) with initial condition $x_i(t_0) = 0$ remain in these planes. By the Picard Lindeloff theorem a solution starting in $int(\Omega)$ cannot cross these planes, therefore $x_i(t) \geq 0$ for $t \geq 0$. If $s_1 = 0$, then $\dot{s_1} = Ds_{in} > 0$. Therefore $s_1(t) \geq 0$. If $s_2 = 0$, let $\bar{k} = \min\left\{ \frac{1}{y_i} : i \in [n] \right\}$, by adding both inequalities from Lemma (1) and since $s_1 \geq 0$ one has $\bar{k} \sum_{i=1}^{n} x_i \leq M_1 + M_2$, which in turn implies: $\bar{k}\|x\|_\infty \leq M_1 + M_2$. Define $M := \frac{\bar{k}}{M_1 + M_2}$, and let A be a matrix such that $\|A\|_\infty \leq M$. It follows that $\|Ax\|_\infty \leq \|A\|_\infty \|x\|_\infty \leq M\frac{1}{M} = 1$, and therefore:

$$(1 + A_{i\bullet}x) \geq 0 \quad \forall i \in [n] \tag{A6}$$

Computing $\dot{s_2} = \sum_{i \in G_1} \frac{1}{y_i} \bar{\mu}_i f_i(s)(1 + A_{i\bullet}x)x_i \geq 0$, and thus $s_2(t) \geq 0$. Note that since $s_2 \geq 0$, bound (A6) is valid for any time, and not only when $s_2 = 0$. If $s_3 = 0$ then $\dot{s_3} = \sum_{i \in G_2}^{n} \frac{1}{y_i} \mu_i(s_2, x) \geq 0$. Therefore $s_3 \geq 0$. For the boundedness it suffices to notice from Lemma 1 that the sum of positive elements is bounded, therefore each element is bounded. □

Appendix B. Deduction of Equilibrium Points

Recall that $f_i(s) = \begin{cases} \bar{\mu}_i \frac{s_1}{K_i + s_1} & \forall i \in G_1 \\ \bar{\mu}_i \frac{s_2}{K_i + s_2} & \forall i \in G_2 \end{cases}$, then $\mu(x, s) = \text{diag}(f(s))(1_{n\times 1} + Ax)$ thus, system (3) is rewritten as follows.

$$\dot{x} = \text{diag}(\mu(x, s) - D_{n\times 1})x \tag{A7}$$

$$\dot{s_1} = (s_{in} - s_1)D + Y_{1\bullet}\,\text{diag}(\mu(x, s))x \tag{A8}$$

$$\dot{s_2} = -s_2 D + Y_{2\bullet}\,\text{diag}(\mu(x, s))x \tag{A9}$$

$$\dot{s_3} = -s_3 D + Y_{3\bullet}\,\text{diag}(\mu(x, s))x \tag{A10}$$

Recall $\mathcal{J}$ the set of non-active coordinates. Let M be the matrix defined by taking out the $\mathcal{J}$ columns of the identity matrix of size n. When M multiplies from the left it adds rows of zeros to the multiplied matrix in the $\mathcal{J}$ coordinates. When matrix M multiplies from the right it takes out the $\mathcal{J}$ columns of the multiplied matrix. When $M^\top$ multiplies from the left it takes out the $\mathcal{J}$ rows of the multiplied matrix. When $M^\top$ multiplies from the right it adds columns of zeros to the multiplied matrix in the $\mathcal{J}$ coordinates. This gives the following relationships:

$$x^{eq} = Mx^{act};\ f^{act}(s) = M^\top f(s);\ \mu^{act}(x,s) = M^\top \mu(x,s);\ Y^{act} = YM;\ M^\top M = I_{n^{act}} \tag{A11}$$

From Equation (A7) equilibrium points satisfy:

$$x_i = 0 \vee (\mathrm{diag}(f(s))(1_{n\times 1} + Ax) - D_{n\times 1})_i = 0 \quad \forall i \in [n] \tag{A12}$$

$$\Rightarrow f_i(s)(1 + A_{i\bullet}x) - D = 0 \quad \forall i \in [n] \setminus \mathcal{J} \tag{A13}$$

note that $\forall j \in \mathcal{J},\ x_j^{eq} = 0$ therefore the coefficients a_{ij} play no role in Equation (A13) and one can rewrite Equation (A13) as

$$\mu^{act}(x,s) = \mathrm{diag}(f^{act}(s))(1_{n^{act}\times 1} + A^{act}x^{act}) = D_{n^{act}\times 1} \tag{A14}$$

From Hypothesis 2 matrix A^{act} has an inverse. This gives the following formula

$$x^{act} = (A^{act})^{-1}(\mathrm{diag}(f^{act}(s))^{-1}D_{n^{act}\times 1} - 1_{n^{act}\times 1}) \tag{A15}$$

Note as well that at the equilibrium, s_3 can be defined in terms of s_1, s_2 and s_{in}. This is done by adding Equations (A8)–(A10) which gives:

$$s_{in} = s_1 + s_2 + s_3 \tag{A16}$$

Appendix B.1. Both Functional Groups Are Present

The case where in each functional group remains at least one OTU is represented by Hypothesis 3. By replacing $x^{eq} = Mx^{act}$ in Equation (A8) yields

$$(s_{in} - s_1)D + Y_{1\bullet}^{act}\,\mathrm{diag}(\mu^{act}(x,s))x^{act} = 0 \tag{A17}$$

For notation and indexing purposes it is useful to define $B := Y_{1\bullet}^{act}(A^{act})^{-1}M^\top$ (note $B_j = 0\,\forall j \in \mathcal{J}$). Replacing (A15) in Equation (A8) reads as follows:

$$(s_{in} - s_1) + Y_{1\bullet}^{act}(A^{act})^{-1}\underbrace{M^\top M}_{I_{n^{act}}}(\mathrm{diag}(f^{act}(s))^{-1}D_{n^{act}\times 1} - 1_{n^{act}\times 1}) = 0 \tag{A18}$$

$$s_{in} - s_1 + \sum_{i\in G_1} B_i\left(\frac{K_i + s_1}{\bar{\mu}_i s_1}D - 1\right) + \sum_{i\in G_2} B_i\left(\frac{K_i + s_2}{\bar{\mu}_i s_2}D - 1\right) = 0 \quad / \cdot s_1 s_2 \tag{A19}$$

$$s_2\left(-s_1^2 + s_1\left(s_{in} + \sum_{i\in G_1\cup G_2} B_i\left(\frac{D}{\bar{\mu}_i} - 1\right)\right) + \sum_{i\in G_1}\frac{DB_iK_i}{\bar{\mu}_i}\right) = -s_1\sum_{i\in G_2}\frac{DB_iK_i}{\bar{\mu}_i} \tag{A20}$$

and so we get to formula (17) of the main text:

$$s_2 = \frac{s_1}{b_1 s_1^2 + b_2 s_1 + b_3} \tag{A21}$$

where

$$b_1 = \left(\sum_{i \in G_2} \frac{DB_iK_i}{\bar{\mu}_i} \right)^{-1} \tag{A22}$$

$$b_2 = -\left(s_{in} + \sum_{i \in G_1 \cup G_2} B_i \left(\frac{D}{\bar{\mu}_i} - 1 \right) \right) \left(\sum_{i \in G_2} \frac{DB_iK_i}{\bar{\mu}_i} \right)^{-1} \tag{A23}$$

$$b_3 = -\sum_{i \in G_1} \frac{DB_iK_i}{\bar{\mu}_i} \left(\sum_{i \in G_2} \frac{DB_iK_i}{\bar{\mu}_i} \right)^{-1} \tag{A24}$$

The same computations must be done with Equation (A9), which is structurally very similar to (A8). By replacing $x^{eq} = Mx^{act}$ in Equation (A9) yields

$$-s_2 + Y_{2\bullet}^{act} (A^{act})^{-1} (\mathrm{diag}(f^{act}(s))^{-1} D_{n^{act} \times 1} - 1_{n^{act} \times 1}) = 0 \tag{A25}$$

It is again useful to define:

$$C := Y_{2\bullet}^{act} (A^{act})^{-1} M^{\top}. \tag{A26}$$

$$-s_2 + Y_{2\bullet}^{act} (A^{act})^{-1} \underbrace{M^{\top} M}_{I_{n^{act}}} (\mathrm{diag}(f^{act}(s))^{-1} D_{n^{act} \times 1} - 1_{n^{act} \times 1}) = 0 \tag{A27}$$

$$-s_2 + \sum_{i \in G_1} DC_i \frac{K_i + s_1}{\bar{\mu}_i s_1} + \sum_{i \in G_2} DC_i \frac{K_i + s_2}{\bar{\mu}_i s_2} - \sum_{i \in G_1 \cup G_2} C_i = 0 \quad / \cdot s_1 s_2 \tag{A28}$$

$$s_1 \left(-s_2^2 + s_2 \sum_{i \in G_1 \cup G_2} C_i \left(\frac{D}{\bar{\mu}_i} - 1 \right) + \sum_{i \in G_2} \frac{DC_iK_i}{\bar{\mu}_i} \right) = -s_2 \sum_{i \in G_1} \frac{DC_iK_i}{\bar{\mu}_i} \tag{A29}$$

and one gets to expression:

$$s_1 = \frac{s_2}{c_1 s_2^2 + c_2 s_2 + c_3} \tag{A30}$$

where

$$c_1 = \left(\sum_{i \in G_1} \frac{DC_iK_i}{\bar{\mu}_i} \right)^{-1} \tag{A31}$$

$$c_2 = -\sum_{i \in G_1 \cup G_2} C_i \left(\frac{D}{\bar{\mu}_i} - 1 \right) \left(\sum_{i \in G_1} \frac{DC_iK_i}{\bar{\mu}_i} \right)^{-1} \tag{A32}$$

$$c_3 = -\sum_{i \in G_2} \frac{DC_iK_i}{\bar{\mu}_i} \left(\sum_{i \in G_1} \frac{DC_iK_i}{\bar{\mu}_i} \right)^{-1} \tag{A33}$$

Then by replacing (A21) in Equation (A30), and after some reordering, one gets a fifth degree polynomial for s_1.

$$s_1 \left(c_1 \left(\frac{s_1}{b_1 s_1^2 + b_2 s_1 + b_3} \right)^2 + c_2 \left(\frac{s_1}{b_1 s_1^2 + b_2 s_1 + b_3} \right) + c_3 \right) = \frac{s_1}{b_1 s_1^2 + b_2 s_1 + b_3} \quad / \cdot \left(b_1 s_1^2 + b_2 s_1 + b_3 \right)^2 \tag{A34}$$

$$a_4 s_1^4 + a_3 s_1^3 + a_2 s_1^2 + a_1 s_1 + a_0 = 0 \quad \vee \quad s_1 = 0 \tag{A35}$$

where

$$a_0 = b_3 + c_3 b_3^2 \tag{A36}$$
$$a_1 = c_2 b_3 + 2c_3 b_3 b_2 - b_2 \tag{A37}$$
$$a_2 = c_1 + c_2 b_2 + c_3 (b_2^2 + 2b_1 b_3) - b_1 \tag{A38}$$
$$a_3 = c_2 b_1 + 2c_3 b_1 b_2 \tag{A39}$$
$$a_4 = c_3 b_1^2 \tag{A40}$$

Appendix B.2. Washout of G_2

The washout of G_2 is represented in Hypothesis 4.

Under this case note that $f^{act}(s)$ depends only on s_1. By replacing again $x^{eq} = Mx^{act}$ in Equation (A7), one obtains the same computations as in the previous section, so retake Equation (A41)

$$s_{in} - s_1 + \sum_{i \in G_1} B_i \left(\frac{K_i + s_1}{\bar{\mu}_i s_1} D - 1 \right) + \sum_{i \in G_2} B_i \left(\frac{K_i + s_2}{\bar{\mu}_i s_2} D - 1 \right) = 0 \tag{A41}$$

Since $G_1 \subset \mathcal{J}$ then $B_i = 0 \forall i \in G_2$ thus (A41) becomes:

$$s_{in} - s_1 + \sum_{i \in G_1} B_i \left(\frac{K_i + s_1}{\bar{\mu}_i s_1} D - 1 \right) = 0 \quad / \cdot s_1 \tag{A42}$$
$$s_1 s_{in} - s_1^2 + \sum_{i \in G_1} B_i \frac{D B_i K_i}{\bar{\mu}_i} + s_1 \sum_{i \in G_1} B_i \left(\frac{D}{\bar{\mu}_i} - 1 \right) = 0 \tag{A43}$$

Furthermore, so a quadratic equation for s_1 is obtained.

$$a_2' s_1^2 + a_1' s_1 + a_0' = 0 \tag{A44}$$

where

$$a_2' = -1 \tag{A45}$$
$$a_1' = s_{in} + \sum_{i \in G_1} B_i \left(\frac{D}{\bar{\mu}_i} - 1 \right) \tag{A46}$$
$$a_0' = \sum_{i \in G_1} B_i \frac{D B_i K_i}{\bar{\mu}_i} \tag{A47}$$

Appendix B.3. Jacobian of the System

Recall that $f : X \to Y$ is a function, then its derivative is $f' : X \to L(X, Y)$ where $L(X, Y)$ denotes the set of continuous linear mappings from X to Y, such that $\|f(x) - f(a) - f'(a)[x - a]\| = o(\|x - a\|)$, where the linear mapping $f'(a)$ is evaluated at $[x - a]$ [32]. In the case where $X = \mathbb{R}^n$ and $Y = \mathbb{R}^m$, $f'(x)$ is the Jacobian matrix evaluated at point x, furthermore $f'(x)[e_j]$ (where e_j is the canonical j-th vector) is the j-th column of the Jacobian matrix evaluated at point x.

If f is linear then $f'(a) = f$ for any $a \in X$. In the case of the diag operator, we observe it is linear, therefore $\text{diag}'(a)[x] = \text{diag}(x)$

In the case of bilinear mappings another formula holds, let $h : X_1 \times X_2 \to Y$ be bilinear. Then $h'(a_1, a_2)[x_1, x_2] = h(x_1, a_2) + h(a_1, x_2)$. In the case of the function $h_1 : \mathcal{M}_{n \times n}(\mathbb{R}) \times \mathbb{R}^n \mapsto \mathbb{R}^n$ such that $h_1(M, x) = Mx$, one can see that $h_1'(A, x)[B, y] = Bx + Ay$.

The chain rule states that $f : X \to Y$ and $g : Y \to Z$, if $h := g \circ f$ then $h'(a) = g'(f(a)) \circ f'(a)$. Take the expression $h : (x, y) \in \mathbb{R}^n \times \mathbb{R}^n \mapsto \text{diag}(x)y$. It is clear that $h = h_1(\text{diag}(x), y)$. By the chain rule:

$$h'(x,y)[a,b] = h_1'(\text{diag}(x),y) \circ (\text{diag}'(x),\ I_n)[a,b] \tag{A48}$$
$$= h_1'(\text{diag}(x),y)[\text{diag}(a),b] \tag{A49}$$
$$= \text{diag}(a)y + \text{diag}(x)b \tag{A50}$$

Note finally that function h is symmetric, i.e., $h(x, y) = h(y, x)$. Then going back to our system.

$$\dot{x} = g_1(x,s) = \text{diag}(\mu(x,s) - D_{n\times 1})x \tag{A51}$$
$$\dot{s} = g_2(x,s) = \Big(\begin{bmatrix} s_{in} & 0 & 0 \end{bmatrix}^\top - s\Big)D + Y\,\text{diag}(\mu(x,s))x \tag{A52}$$

where $\mu(x, s) = \text{diag}(f(s))(1_{n\times 1} + Ax)$ For a fixed s, let $\mu_s(x) : x \mapsto \text{diag}(f(s))(1 + Ax)$ then, $\mu_s'(x) = \text{diag}(f(s))A$. Let $g_{1s}(x) := h(\mu_s(x) - D_{n\times 1}, x) = g_1(x, s)$, then compute g_{1s}':

$$g_{1s}'(x)[e_j] = h'(\mu_s(x) - D_{n\times 1}, x) \circ (\mu_s'(x), I_n)[e_j, e_j] \tag{A53}$$
$$= h'(\mu_s(x) - D_{n\times 1}, x)[\text{diag}(f(s))Ae_j, e_j] \tag{A54}$$
$$= \text{diag}(f(s))\,\text{diag}(x)A_{\bullet j} + \text{diag}(\mu_s(x) - D_{n\times 1})e_j \tag{A55}$$
$$\Rightarrow g_{1s}'(x) = \text{diag}(f(s))\,\text{diag}(x)A + \text{diag}(\mu_s(x) - D_{n\times 1}) \tag{A56}$$

Again, for a fixed s, let $g_{2s}(x) := \Big(\begin{bmatrix} s_{in} & 0 & 0 \end{bmatrix}^\top - s\Big)D + Yh(\mu_s(x), x) = g_2(x, s)$, then compute g_{2s}':

$$g_{2s}'(x)[e_j] = Yh'(\mu_s(x), x) \circ (\mu_s'(x), I_n)[e_j, e_j] \tag{A57}$$
$$= Y\big(\text{diag}(f(s))\,\text{diag}(x)A_{\bullet j} + \text{diag}(\mu_s(x))e_j\big) \tag{A58}$$
$$\Rightarrow g_{2s}'(x) = Y(\text{diag}(f(s))\,\text{diag}(x)A + \text{diag}(\mu_s(x))) \tag{A59}$$

Let $f_{G_1}(s_1)$ the function containing the first n_1 components of function $f(s)$ and $f_{G_2}(s_2)$ the function containing the last n_2 components of function $f(s)$ so one can write

$$f(s) = \begin{pmatrix} f_{G_1}(s_1) \\ f_{G_2}(s_2) \end{pmatrix} \tag{A60}$$

One can see then that:

$$f'(s) = \begin{bmatrix} f_{G_1}'(s_1) & 0_{n_1\times 2} \\ 0_{n_2\times 1} & f_{G_2}'(s_2) \quad 0_{n_2\times 1} \end{bmatrix} \tag{A61}$$

Now for a fixed x let $\mu_x(s) : s \mapsto \text{diag}(f(s))(1_{n\times 1} + Ax)$, therefore $\mu_x'(s) = \text{diag}(1 + Ax)f'(s)$. Let $g_{1x}(s) := h(\mu_x(s) - D_{n\times 1}, x) = g_1(x, s)$, then compute g_{1x}':

$$g_{1x}'(s)[e_j] = h'(\mu_x(s) - D_{n\times 1}, x) \circ (\mu_x'(s), 0)[e_j, e_j] \tag{A62}$$
$$g_{1x}'(s)[e_j] = h'(\mu_x(s) - D_{n\times 1}, x)[\mu_x'(s)e_j, 0] \tag{A63}$$
$$g_{1x}'(s)[e_j] = \text{diag}(x)\mu_x'(s)e_j \tag{A64}$$
$$\Rightarrow g_{1x}' = \text{diag}(x)\,\text{diag}(1 + Ax)f'(s) \tag{A65}$$

Again, for a fixed x let $g_{2x}(s) := \left(\begin{bmatrix} s_{in} & 0 & 0 \end{bmatrix}^\top - s\right)D + Yh(\mu_s(x), x) = g_2(x, s)$, then compute g'_{1x}:

$$g'_{2x}(s)[e_j] = -DI_n e_j + Yh'(\mu_x(s), x) \circ (\mu'_x(s), 0)[e_j, e_j] \tag{A66}$$
$$g'_{2x}(s)[e_j] = -DI_n e_j + Yh'(\mu_x(s), x)[\mu'_x(s)e_j, 0] \tag{A67}$$
$$g'_{2x}(s)[e_j] = -DI_n e_j + Y \operatorname{diag}(x)\mu'_x(s)e_j \tag{A68}$$
$$\Rightarrow g'_{2x}(s) = -DI_n + Y \operatorname{diag}(x) \operatorname{diag}(1 + Ax)f'(s) \tag{A69}$$

Finally note that:

$$f'_{G_1} := \left(\frac{\partial f_1}{\partial s_1}, \ldots, \frac{\partial f_{n_1}}{\partial s_1}\right)^\top = \left(\frac{\bar{\mu}_1 K_1}{(K_1 + s_1)^2}, \ldots, \frac{\bar{\mu}_{n_1} K_{n_1}}{(K_{n_1} + s_1)^2}\right)^\top \in \mathbb{R}^{n_1}$$
$$f'_{G_2} := \left(\frac{\partial f_{n_1+1}}{\partial s_2}, \ldots, \frac{\partial f_n}{\partial s_2}\right)^\top = \left(\frac{\bar{\mu}_{n_1+1} K_{n_1+1}}{(K_{n_1+1} + s_2)^2}, \ldots, \frac{\bar{\mu}_n K_n}{(K_n + s_2)^2}\right)^\top \in \mathbb{R}^{n_2}$$

Then the Jacobian of the system may be expressed as:

$$J(x, s) = \begin{bmatrix} g'_{1s}(x) & g'_{1x}(s) \\ g'_{2s}(x) & g'_{2x}(s) \end{bmatrix} \tag{A70}$$

Appendix B.4. Stability Analysis with no interactions

In this case the growth functions are

$$\mu_1(s, x) = \bar{\mu}_1 \frac{s_1}{K_1 + s_1}; \quad \mu_2(s, x) = \bar{\mu}_2 \frac{s_2}{K_1 + s_2}$$

Coexistence:

$$s_1^{eq} = \frac{K_1 D}{\bar{\mu}_1 - D}; \quad x_1^{eq} = \frac{s_{in} - s_1^{eq}}{k_1}; \quad s_2^{eq} = \frac{K_2 D}{\bar{\mu}_2 - D};$$
$$x_2^{eq} = \frac{s_{in} - s_1^{eq} - s_2^{eq}}{k_2}; \quad s_3^{eq} = s_{in} - s_1^{eq} - s_2^{eq}$$

Washout of x_2

$$s_1^{eq} = \frac{K_1 D}{\bar{\mu}_1 - D}; \quad x_1^{eq} = \frac{s_{in} - \frac{K_1 D}{\bar{\mu}_1 - D}}{k_1}; \quad s_2^{eq} = s_{in} - s_1^{eq}; \quad x_2^{eq} = 0; \quad s_3^{eq} = 0$$

Washout

$$s_1^{eq} = s_{in}; \quad x_1^{eq} = 0; \quad s_2^{eq} = 0; \quad x_2^{eq} = 0; \quad s_3^{eq} = 0$$

Jacobian of the system:

$$\begin{bmatrix} \bar{\mu}_1 \frac{s_1}{K_1 + s_1} - D & 0 & \bar{\mu}_1 \frac{K_1}{(K_1 + s_1)^2} x_1 & 0 & 0 \\ 0 & \bar{\mu}_1 \frac{s_2}{K_2 + s_2} - D & 0 & \bar{\mu}_2 \frac{K_1}{(K_1 + s_2)^2} x_2 & 0 \\ -k_1 \bar{\mu}_1 \frac{s_1}{K_1 + s_1} & 0 & -D - k_1 \bar{\mu}_1 \frac{K_1}{(K_1 + s_1)^2} x_1 & 0 & 0 \\ k_1 \bar{\mu}_1 \frac{s_1}{K_1 + s_1} & -k_2 \bar{\mu}_2 \frac{s_2}{K_2 + s_2} & k_1 \bar{\mu}_1 \frac{K_1}{(K_1 + s_1)^2} x_1 & -D - k_2 \bar{\mu}_2 \frac{K_2}{(K_2 + s_2)^2} x_2 & 0 \\ 0 & 0 & 0 & 0 & -D \end{bmatrix}$$

Appendix C. Tracking Problem Reformulation and Details

For applying the methods developed in [26]. Define the system state $X = (x, s)$. Make the change of variables $v_i = u_i - 1$ with $v = (v_1, \ldots, v_n)$ are applied to system (25). The system may be rewritten then as:

$$\begin{aligned}
\dot{x_i} &= (f_i(s)(1 + v_i(t)) - D)x_i \quad \forall i \in G_1 \\
\dot{x_i} &= (f_i(s)(1 + v_i(t)) - D)x_i \quad \forall i \in G_2 \\
\dot{s_1} &= s_1\left(\frac{s_{in}}{s_1} - 1\right)D + \sum_{i \in G_1} y_{s_1/x_i} f_i(s)(1 + v_i(t))x_i \\
\dot{s_2} &= -s_2 D + \sum_{i \in G_1 \cup G_2} y_{s_2/x_i} f_i(s)(1 + v_i(t))x_i \\
\dot{s_3} &= -s_3 D + \sum_{\substack{i \in G_2 \\ n_1+n_2}} y_{s_3/x_i} f_i(s)(1 + v_i(t))x_i \\
y(t) &= x(t)
\end{aligned} \tag{A71}$$

Define:

$$A(X) = \begin{bmatrix} A_{11}(X) & A_{12}(X) \\ A_{21}(X) & A_{22}(X) \end{bmatrix};\ B(X) = \begin{bmatrix} B_1(X) \\ B_2(X) \end{bmatrix} \tag{A72}$$

with

$$A_{11}(X) = \operatorname{diag}(f(s) - D_{n\times 1});\ A_{12}(X) = 0_{n\times 3} \tag{A73}$$

$$A_{21}(X) = Y \operatorname{diag}(f(s));\ A_{22}(X) = \begin{bmatrix} \left(\frac{s_{in}}{s_1} - 1\right)D & 0 & 0 \\ 0 & -D & 0 \\ 0 & 0 & -D \end{bmatrix} \tag{A74}$$

and

$$B_1(X) = \operatorname{diag}(f(s))\operatorname{diag}(x);\ B_2(X) = Y \operatorname{diag}(f(s))\operatorname{diag}(x);\ C(X) = \begin{bmatrix} \operatorname{diag}(I_n) & 0_{n\times 3} \end{bmatrix} \tag{A75}$$

Then the system (A71) can be rewritten as:

$$\dot{X} = A(X)X + B(X)v;\ y = C(X)X \tag{A76}$$

$z(t) \in \mathbb{R}^n$ is the measured vector containing the OTU concentrations in time. The cost functional is given by

$$J(v) = \Big(z(t_f) - C(X)X(t_f)\Big)^\top F\Big(z(t_f) - C(X)X(t_f)\Big) \tag{A77}$$

$$+ \int_{t_0}^{t_f} (z(t) - C(X)X(t))^\top Q(z(t) - C(X)X(t)) + v(t)^\top R v(t) \tag{A78}$$

where F, Q and R are positive definite matrices. Since there is no interest in the final time $F = 0$. Q and R are taken as diagonal matrices, in that way the system can be reduced as shown below. Particularly after testing the model in the proof of concept and data, the Q and R matrices were $\begin{bmatrix} \lambda_1 I_{n_1} & 0 \\ 0 & \lambda_2 I_{n_2} \end{bmatrix}$ and I_n, respectively, with $\lambda_1 = 10^{-4}$ and $\lambda_2 = 10^{-5}$.

Define the dynamic sequences for $i \in \mathbb{N}$, $\dot{X}^{[i]}$ as:

$$\dot{X}^{[i]} = A(X^{[i]})X^{[i]} + B(X^{[i]})v^{[i]} \quad i \in \mathbb{N} \tag{A79}$$

$$y^{[i]} = X^{[i]} \quad i \in \mathbb{N} \tag{A80}$$

$$X^{[i]}(t_0) = X_0 \quad i \in \mathbb{N} \tag{A81}$$

and for $i = 0$ define $X^{[0]}(t)$ as the solution of (A71) with $v(t) \equiv 0$. The control law is given by

$$v^{[i]}(t)_j = \max\left\{-1, \min\left\{0, \left(-R^{-1}B^\top\left(X^{[i-1]}(t)\right)\left(P^{[i]}(t)X^{[i]}(t) - s_f^{[i]}(t)\right)\right)_j\right\}\right\} \forall j \in [n] \quad \text{(A82)}$$

where $P^{[i]}(t) \in \mathcal{M}_{n+3\times n+3}(\mathbb{R})$ and $s_f^{[i]}(t) \in \mathbb{R}^{n+3}$ are the solution to the differential equations:

$$\dot{P}^{[i]} = -C^T\left(X^{[i-1]}(t)\right)QC\left(X^{[i-1]}(t)\right) - P^{[i]}A\left(X^{[i-1]}(t)\right) - A^\top\left(X^{[i-1]}(t)\right)P^{[i]} \quad \text{(A83)}$$

$$+ P^{[i]}B\left(X^{[i-1]}(t)\right)R^{-1}B^\top\left(X^{[i-1]}(t)\right)P^{[i]} \quad \text{(A84)}$$

$$P^{[i]}(t_f) = C^\top\left(X^{[i-1]}(t_f)\right)FC\left(X^{[i-1]}(t_f)\right) \quad \text{(A85)}$$

$$\dot{s_f^{[i]}} = -C^\top\left(X^{[i-1]}(t)\right)Qz(t) - \left[A\left(X^{[i-1]}(t)\right) - B\left(X^{[i-1]}(t)\right)R^{-1}B^\top\left(X^{[i-1]}(t)\right)P^{[i]}(t)\right]^\top s_f^{[i]} \quad \text{(A86)}$$

$$s_f^{[i]}(t_f) = C^\top\left(X^{[i-1]}(t_f)\right)Fz(t_f) \quad \text{(A87)}$$

Replacing the matrices of our problem

$$\dot{P}^{[i]}(t) = -\begin{bmatrix} Q & 0_{n\times 3} \\ 0_{3\times n} & 0_{3\times 3}\end{bmatrix} - P^{[i]}A\left(X^{[i-1]}(t)\right) - A^\top\left(X^{[i-1]}(t)\right)P^{[i]} + P^{[i]}B\left(X^{[i-1]}(t)\right)R^{-1}B^\top\left(X^{[i-1]}(t)\right)P^{[i]} \quad \text{(A88)}$$

$$P^{[i]}(t_f) = \begin{bmatrix} 0_{n\times n} & 0_{n\times 3} \\ 0_{3\times n} & 0_{3\times 3}\end{bmatrix} \quad \text{(A89)}$$

$$\dot{s_f}^{[i]}(t) = -\begin{bmatrix} Qz(t) \\ 0_{3\times 1}\end{bmatrix} - \left[A\left(X^{[i-1]}(t)\right) - B\left(X^{[i-1]}(t)\right)R^{-1}B^\top\left(X^{[i-1]}(t)\right)P^{[i]}(t)\right]^\top s_f^{[i]} \quad \text{(A90)}$$

$$s_f^{[i]}(t_f) = \begin{bmatrix} 0_{n\times n} & 0_{n\times 3}\end{bmatrix}^\top z(t_f) \quad \text{(A91)}$$

For certain entries of the dynamic the constantly zero function is a solution for them, implying by existence and uniqueness that they should be constantly zero. Then $P^{[i]}$ has $n \times n$ non zero entries and $s^{[i]}$ has n non zero entries, explicitly:

$$P^{[i]}(t) = \begin{bmatrix} \tilde{P}^{[i]}(t) & 0_{n\times 3} \\ 0_{3\times n} & 0_{3\times 3}\end{bmatrix};\ s_f^{[i]}(t) = \begin{bmatrix} \tilde{s}_f^{[i]}(t) \\ 0_{1\times 3}\end{bmatrix} \quad \text{(A92)}$$

$$A\left(X^{[i-1]}(t)\right) = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22}\end{bmatrix};\ B\left(X^{[i-1]}(t)\right) = \begin{bmatrix} B_1 \\ B_2\end{bmatrix} \quad \text{(A93)}$$

where $A_{11} = \text{diag}\left(f\left(s^{[i-1]}\right) - D_{n\times 1}\right)$, and $B_1 = \text{diag}\left(f\left(s^{[i-1]}\right)\right)\text{diag}\left(x^{[i-1]}\right)$ The equations for $\tilde{P}^{[i]}(t)$:

$$\dot{\tilde{P}}^{[i]}(t) = -Q - \tilde{P}^{[i]}A_{11} - A_{11}^\top\tilde{P}^{[i]} + \tilde{P}^{[i]}B_1R^{-1}B_1^\top\tilde{P}^{[i]};\ \tilde{P}^{[i]}(t_f) = 0 \quad \text{(A94)}$$

Inspecting the former equation one notices that if $i \neq j$, $P_{ij}(t) = 0$ is a solution of all non diagonal entries when R is diagonal. Furthermore, therefore, once again, by existence and uniqueness they should be constantly zero. Hence only the diagonal entries should be calculated.

$$\dot{\tilde{P}}_{jj}^{[i]}(t) = -Q_{jj} - 2\left(f_j\left(s^{[i-1]}\right) - D\right)\tilde{P}_{jj}^{[i]} + R^{-1}\left(\tilde{P}_{jj}^{[i]}\right)^2 f_j\left(s^{[i-1]}\right)^2\left(x_j^{[i-1]}\right)^2;\ \tilde{P}_{jj}^{[i]}(t_f) = 0 \quad \text{(A95)}$$

For and $\tilde{s_f}^{[i]}(t)$ the system reduces to:

$$\dot{\tilde{s}}_f^{[i]}(t) = -z(t) - \left[A_{11} - R^{-1}B_1B_1^\top\tilde{P}^{[i]}(t)\right]^\top\tilde{s}_f^{[i]};\ \tilde{s}_f^{[i]}(t_f) = 0 \quad \text{(A96)}$$

Furthermore, the control law is given by

$$v^{[i]}(t)_j = \max\left\{-1, \min\left\{0, \left(-R^{-1}B_1^{\top}\left(X^{[i-1]}(t)\right)\left(\tilde{P}^{[i]}(t)x^{[i]}(t) - \tilde{s}_f^{[i]}(t)\right)\right)_j\right\}\right\} \forall j \in [n] \quad \text{(A97)}$$

They were solved using standard backward numerical integration.

References

1. West, S.A.; Diggle, S.P.; Buckling, A.; Gardner, A.; Griffin, A.S. The Social Lives of Microbes. *Ann. Rev. Ecol. Evol. Syst.* **2007**. [CrossRef]
2. Widder, S.; Allen, R.J.; Pfeiffer, T.; Curtis, T.P.; Wiuf, C.; Sloan, W.T.; Cordero, O.X.; Brown, S.P.; Momeni, B.; Shou, W.; et al. Challenges in microbial ecology: Building predictive understanding of community function and dynamics. *ISME J.* **2016**, *10*, 2557–2568. [CrossRef]
3. Wade, M.; Harmand, J.; Benyahia, B.; Bouchez, T.; Chaillou, S.; Cloez, B.; Godon, J.J.; Boudjemaa, B.M.; Rapaport, A.; Sari, T.; et al. Perspectives in mathematical modelling for microbial ecology. *Ecol. Model.* **2016**, *321*, 64–74. [CrossRef]
4. Song, H.S.; Cannon, W.; Beliaev, A.; Konopka, A. Mathematical Modeling of Microbial Community Dynamics: A Methodological Review. *Processes* **2014**, *2*, 711–752. [CrossRef]
5. Boucher, Y.; Douady, C.J.; Papke, R.T.; Walsh, D.A.; Boudreau, M.E.R.; Nesbø, C.L.; Case, R.J.; Doolittle, W.F. Lateral Gene Transfer and the Origins of Prokaryotic Groups. *Ann. Rev. Genet.* **2003**. [CrossRef] [PubMed]
6. Ferrera, I.; Sánchez, O. Insights into microbial diversity in wastewater treatment systems: How far have we come? *Biotechnol. Adv.* **2016**, *34*, 790–802. [CrossRef] [PubMed]
7. Lotka, A.J. Elements of Physical Biology. *J. Am. Stat. Assoc.* **1925**. [CrossRef]
8. Volterra, V. Fluctuations in the abundance of a species considered mathematically. *Nature* **1926**. [CrossRef]
9. Kot, M. *Elements of Mathematical Ecology*; Cambridge University Press: Cambridge, UK, 2001; [CrossRef]
10. Hernández-Bermejo, B.; Fairén, V. Lotka–Volterra representation of general nonlinear systems. *Math. Biosci.* **1997**. [CrossRef]
11. Perez-Garcia, O.; Lear, G.; Singhal, N. Metabolic network modeling of microbial interactions in natural and engineered environmental systems. *Front. Microbiol.* **2016**, *7*, 673. [CrossRef]
12. Gopalsamy, K. Global asymptotic stability in a periodic Lotka–Volterra system. *J. Aust. Math. Soc. Ser. B Appl. Math.* **1985**. [CrossRef]
13. Aleksandrov, A.Y.; Aleksandrova, E. Convergence conditions for some classes of nonlinear systems. *Syst. Control Lett.* **2017**, *104*, 72–77. [CrossRef]
14. Bucci, V.; Tzen, B.; Li, N.; Simmons, M.; Tanoue, T.; Bogart, E.; Deng, L.; Yeliseyev, V.; Delaney, M.L.; Liu, Q.; Olle, B.; et al. MDSINE: Microbial Dynamical Systems INference Engine for microbiome time-series analyses. *Genome Biol.* **2016**, *17*, 121. [CrossRef]
15. Monod, J. *Recherches sur la Croissance des Cultures Bacteriennes*; Hermann: Paris, France, 1942.
16. Dumont, M.; Godon, J.J.; Harmand, J. Species Coexistence in Nitrifying Chemostats: A Model of Microbial Interactions. *Processes* **2016**, *4*, 51. [CrossRef]
17. Çimen, T. State-Dependent Riccati Equation (SDRE) control: A survey. *IFAC Proc. Vol. (IFAC-PapersOnline)* **2008**, *41*, 3761–3775. [CrossRef]
18. Dumont, M.; Harmand, J.; Rapaport, A.; Godon, J.J. Towards functional molecular fingerprints. *Environ. Microbiol.* **2009**, *11*, 1717–1727. [CrossRef] [PubMed]
19. Sharma, B.; Ahlert, R. Nitrification and nitrogen removal. *Water Res.* **1977**, *11*, 897–925. [CrossRef]
20. Khalil, H.K. *Nonlinear Systems*; Prentice Hall: Upper Saddle River, NJ, USA, 1996 [CrossRef]
21. Pavlou, S. Computing operating diagrams of bioreactors. *J. Biotechnol.* **1999**, *71*, 7–16. [CrossRef]
22. Harmand, J.; Lobry, C.; Rapaport, A.; Sari, T. *The Chemostat: Mathematical Theory of Microorganism Cultures*; John Wiley & Sons: Hoboken, NJ, USA, 2017. [CrossRef]
23. Khin, T.; Annachhatre, A.P. Novel microbial nitrogen removal processes. *Biotechnol. Adv.* **2004**, *22*, 519–532. [CrossRef]
24. Wagner, M.; Loy, A.; Nogueira, R.; Purkhold, U.; Lee, N.; Daims, H. Microbial community composition and function in wastewater treatment plants. *Antonie Van Leeuwenhoek* **2002**, *81*, 665–680. [CrossRef]
25. Harmand, J.; Lobry, C.; Rapaport, A.; Sari, T. *Optimal Control in Bioprocesses: Pontryagin's Maximum Principle in Practice*; John Wiley & Sons: Hoboken, NJ, USA, 2019. [CrossRef]
26. Çimen, T.; Banks, S.P. Nonlinear optimal tracking control with application to super-tankers for autopilot design. *Automatica* **2004**, *40*, 1845–1863. [CrossRef]
27. Dumont, M.; Rapaport, A.; Harmand, J.; Benyahia, B.; Godon, J.J. Observers for microbial ecology-How including molecular data into bioprocess modeling? In Proceedings of the 2008 16th Mediterranean Conference on Control and Automation, Ajaccio, France, 25–27 June 2008; pp. 1381–1386.
28. Ugalde-Salas, P.; Harmand, J.; Desmond-Le Quéméner, E. Asymptotic Observers and Integer Programming for Functional Classification of a Microbial Community in a Chemostat. In Proceedings of the 2019 18th European Control Conference (ECC), Naples, Italy, 25–28 June 2019; pp. 1665–1670.

29. Wiesmann, U. Biological nitrogen removal from wastewater. In *Biotechnics/Wastewater*; Springer: Berlin/Heidelberg, Germany, 1994; pp. 113–154. [CrossRef]
30. Loisel, P.; Harmand, J.; Zemb, O.; Latrille, E.; Lobry, C.; Delgenès, J.P.; Godon, J.J. Denaturing gradient electrophoresis (DGE) and single-strand conformation polymorphism (SSCP) molecular fingerprintings revisited by simulation and used as a tool to measure microbial diversity. *Environ. Microbiol.* **2006**, *8*, 720–731. [CrossRef] [PubMed]
31. Dochain, D. State and parameter estimation in chemical and biochemical processes: A tutorial. *J. Process Control* **2003**, *13*, 801–818. [CrossRef]
32. Cartan, H. *Calcul Différentiel*; Hermann: Paris, France, 1967; Volume 1.