Mathematical Models of Microbial Community Dynamics in Continuous Cultures

Pablo Ugalde

THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE MONTPELLIER SUPAGRO

En Biotechnologie et Microbiologie

École doctorale GAIA – Biodiversité, Agriculture, Alimentation, Environnement, Terre, Eau
Portée par l'Université de Montpellier

Unité de recherche Laboratoire de Biotechnologie de l'Environnement (LBE, INRAE UR050)

Mathematical Models of Microbial Community Dynamics in Continuous Cultures

Présentée par Pablo UGALDE SALAS
Le 2 novembre 2020

Sous la direction de Jérôme Harmand
et Théodore Bouchez

Devant le jury composé de

Denis DOCHAIN, Professeur d'Université, Université Catholique de Louvain	Rapporteur
Orkun SOYER, Professeur d'Université, Warwick University	Rapporteur
Jérôme HARMAND, Directeur de Recherche, LBE INRAE	Directeur de thèse
Elie DESMOND-LE QUEMENER, Directeur de Recherche, LBE INRAE	Examinateur
Beatrice LAROCHE, Directeur de Recherche, MaIAGE INRAE	Examinateur
Matthew WADE, Professeur associé, Newcastle University	Examinateur
Marc HERAN, Professeur d'Université, Université de Montpellier	Président du jury

UNIVERSITÉ DE MONTPELLIER

Montpellier SupAgro

Entendemos x ecologismo
Un movimiento socio-económico
Basado en la idea de la armonía
Del ser humano con la naturaleza
Que lucha x una vida lúdica
Creativa pluralista igualitaria
libre de explotación
Y basada en la comunicación
de grandes y chicos
Muchos los problemas
Una la solución
Economía mapuche de subsistencia
Nicanor Parra, 2001

In loving memory of my father.

Contents

List of Figures

List of Tables

Acknowledgements

I would like to start by thanking my supervisors Dr. Elie Desmond Le-Quéméner, Dr. Jerome Harmand, and Dr. Théodore Bouchez for the constant help and feedback given to me in the last three years. Thanks to Dr. Héctor Ramirez, for having me on a 3 month internship at the Center for mathematical modelling in Universidad de Chile. Thanks to Dr. Alain Rapaport for his kind and stimulating discussions. Thanks to Dr. Eric Trably for the scientific exchanges. Thanks to Dr. Alexandra Elbakyan for liberating knowledge in the true spirit of science.

Since this is one of the highest degrees one can get from formal education, I can not help but honor some of the teachers that have marked me through my studies: Paola Miño, Lucía Castrucci, Claudia Carrasco, Heidi Michaelis, Daniella Frigerio, Macarena Bajas, Tania Mella, Rossana Moraga, Jacqueline Pereira, Víctor Cruz, and Ignacio Fuenzalida.

I would like to give thanks to all the people I have encountered and shared at LBE. Wendy and Helene who welcomed me and helped me from day one, Hicham and his big heart, Esmee with her joyful spirit, Felipe and Yolanda who even shared their home with me, Florian and his frankness, Julie who has always been there for me, Alice for being Alice, Clement and his good will, and Lucía for her very dear friendship. Mohammed, Vicente, Sego, and Mathilde we had a lot of fun together on my first year, I'll always cherish those moments. To my office partners Kevin and Alex, for sure you did work more tolerable, as well as Gaetano for our post lunch discussions. I also have to thank Farouk, Quentin, Doha, Lluis, Amine, Nico, Roman, Virginie, Julien, Fernanda, Angie, Lorraine, Kevin, Anaïs, Ulysses, Roland, Celine, Pierre, Antoine, Lucie, and Raffaello.

To my housemates: Noemie, Clemence, Marie, Bastien, and Morand for their good vibe, it was great spending the last two years with you. Thanks to David Fernández and his family, they have been the most kind to me.

Thanks to Annie, Sylvie, Ouasila, Alexandra, and Nadine, who make everything possible. To Eric Latrille and Domi who are always worried for the well being of the community. To Nico for all the scientific discussions and help. To all those with whom I shared a meal or a good laugh: Hélène, Robert, Michel, Marjolaine, Jérôme, Kim, Jean-Phi, Stephanie, Gaëlle, Audrey, Diana, Denis, Philippe, Nathalie, and Isabelle.

To all my people who have kept constantly in touch during this experience: Mari, Pancho, Efe, Joaco, Diego Ortiz, Tomi Ortega, Cata Hernández, Cata Cuevas, Pato Foncea, Pipe Campos, Garrido, Checo, Javi, Pipers, Juangui, Seba, Piter, Daniel, Negro, Marce, Eddy, Pato Vega, Dani, Carlos, Papo, Cristobal, Kuky, Natacha, Rosi, Clau, Cami,Felipe, Coni, Goussy, Alex, Benjie, Charly, Dany.

Thanks to my aunt Alicia and uncle Fernando for being there whenever I needed it. Finally, thanks to those who have constantly supported and encouraged me during my life: Clari, Tati, Chica, Tomás, Yiyo, my father, and especially my mother, I cannot express how much I owe her.

Résumé

La thèse traite de la modélisation des communautés microbiennes dans les réacteurs homogènes: elle explore les implications des récents progrès techniques et théoriques de la microbiologie sur la modélisation du taux de croissance. Elle aborde deux axes de recherche principaux. Le premier est lié à l'intégration des apports de l'approche moléculaire en écologie microbienne, tandis que le second vise à intégrer les avancées de la bioénergétique. Ces approches apportent un éclairage nouveau sur la modélisation des interactions entre espèces dans les modèles déjà existants, en particulier dans la théorie du chemostat. Ces axes, bien que complémentaires, peuvent être considérés indépendamment l'un de l'autre.

Pour intégrer les apports de l'écologie microbienne, l'objectif est de réaliser le passage d'une modélisation fonctionnelle (approche utilisée traditionnelle en ingénierie microbiologique) à une modélisation intégrant l'unité taxonomique opérationnelle (OTU). Ce sujet est exploré en exploitant des données issues d'un chemostat dans lequel a lieu un processus de nitrification. Des questions concrètes telles que l'identification du métabolisme associé à chaque OTU [125], l'étude des interactions microbiennes et la reconstruction de leurs fonctions de croissance [124] sont abordées. Les outils mathématiques utilisés sont principalement la programmation linéaire en nombres entiers mixtes et la théorie du contrôle.

L'intégration de la bioénergétique apporte, quant à elle, des informations sur la relation entre l'énergie dissipée, les rendements de conversion du substrat en biomasse et des paramètres cinétiques. Ces relations sont dérivées de recherches théoriques fondées sur la théorie de l'état de transition microbien [123] et les résultats sont utilisés pour

revisiter des données provenant de cultures pures en chemostats consommant du glucose en conditions aérobies.

Abstract

The thesis deals with the modelling of microbial communities in homogeneous reactors: it explores the implications of the recent technical and theoretical advances in microbiology on the growth rate expression. Two main axis of research compose this thesis. The first one is linked with the use of genetic sequencing advances coming from microbial ecology, while the second one looks at bioenergetics. Both of them bring more complexity in already existing models, particularly in chemostat theory. Though complementary to each other, each axis of research can be regarded as an independent subject.

From a microbial ecology perspective the passage from functional group to operative taxonomic unit (OTU) is the main concern. This topic is explored with data coming from a nitrification process. More concrete questions such as identifying the functionality of each OTU [125] and reconstructing their growth expressions are tackled [124]. The mathematical tools used are mainly mixed integer programming and control theory.

From bioenergetics, insights are drawn on the relationship between dissipated energy, biomass to substrate conversion yields and other kinetics parameters. These relationships are derived from theoretical enquiries and are strongly based on microbial transition state theory [123]. Data coming from chemostats with glucose as substrate is revisited under this approach.

Chapter 1

Introduction

In the field of bioprocess, the measurement of variables either physical, chemical, or biological, within reactors has reached a point where almost a continuum of data can be obtained in real time [14]. The benchmark mathematical models for wastewater treatment [49] and anaerobic digestion [8] have been built over these observational capabilities to offer excellent plant operation, nevertheless unexpected breakdowns in reactors still occur [134]. A new window of opportunity emerges as data obtained through genetic sequencing of the microbial community (MC) offers a new proxy in order to better understand and pilot bioprocesses [138]. In that context there is a rising demand for mathematical models which take into account the MC.

Ecology, as a scientific discipline studying the interactions between organisms and abiotic factors, is evidently much older than microbial ecology. It is the case that concepts in classical ecology find their way in microbial ecology, but it is by no means an easy or direct transition. To begin with, the distinction of species in the macro world, as to the criteria of reproductive capabilities between individuals, from species in the micro world is not evident. Indeed, microbial cells multiply themselves by division, and might acquire genes from other cells through horizontal gene transfer [15] (as opposed to vertical transfer from DNA replication during division). Taxonomically, microbes have been classified, but it is not easy to extend the concept of species in a complex microbial community. In 1963 Sokal and Sneath defined classification as "the ordering of organisms into groups on the

basis of their relationships by descent or similarity or both". They used the concept of Operational Taxonomic Unit (OTU) [114], to simply enumerate the species being studied at hand, in order to set the principles of numerical taxonomy given the rising use of electrical and computational devices to analyse observations. However, to pinpoint who are these OTU in a microbial community was an arduous task of microscopic observation and experiments to determine characteristics such as the presence of a cell wall. It was in 1977 that the identification of the 16S ribosomal RNA as a slow varying genetic marker was used to establish and reconstruct phylogeny in the prokariotic domain [140], making a revolution in the field of microbial ecology, and thus enabling an easier determination of the OTU within an ecosystem. The microbial OTU take a central place in the models studied in this thesis.

Cells are open systems exchanging energy in the form of heat and mass with their environment and other cells, constantly pushing away from thermodynamic equilibrium. Prigogine and Nikolis in their book of non-equilibrium thermodynamics refer to such systems as being dissipative structures [89]. They established connections between them and stable limit cycles from differential equations, and proposed that periodicity, or self sustained oscillations must lie at the heart of such structures. On the other hand, without making use of non-equilibrium thermodynamics and dynamical systems theory, Mitchell proposed the chemiosmotic theory, that explains how an ever-turning wheel of the in and out flow of protons through the membrane of mitochondria maintains an energy gradient responsible for ATP formation [82]. This wheel, however, is sustained by the high energy yielding molecule that is acetyl-coA, and in the case of non-photosynthetic organisms, it comes from the break down of nutrients captured by the cell. Bioenergetics deals at the same time with the intricate mechanisms sustaining energy transformations in life, but also with the available energy gradients to do so, which is one of the focus of the thesis.

This manuscript can be seen as a dialogue between the three introduced topics, bio-processes, microbial ecology, and bioenergetics, mediated by mathematical modelling.

1.1 Microbial Growth in Bioprocess

Mathematical models aiming to grasp ecological dynamics of a population can be traced back to Verlhust (1838) [128]. More than a hundred years later Jacques Monod using data from continuous cultures proposed a growth law that bears his name, to model bacterial growth under substrate limitation [85] (see table 1.1). The system of differential equations corresponding to a chemostat is expressed in equation 1.1, x represents the biomass, s the substrate concentration, D the dilution rate, s_{in} the concentration of the substrate at the entry of the reactor, $y_{x/s}$ is the yield of substrate to biomass conversion, and $\mu(s)$ is the growth rate depending on substrate s. The fifties and the sixties were marked by the use of the chemostat [92] to study microbial development [54], which led some to question the validity of Monod's growth law under certain circumstances and consequently to the proposal of new growth expressions. Table 1.1 shows some expressions for μ, the constants in these expressions are usually referred to as kinetic parameters. In general terms, these expressions all retained the key aspect of substrate limitation and added another important experimental observation or phenomena, such as a density dependent effect [23], inhibitory effects of the substrate [4], inhibitory effects of the products [2], and slow transient dynamics [86].

$$\begin{aligned} \dot{x} &= (\mu(s) - D)x \\ \dot{s} &= (s_{in} - s)D - \frac{1}{y_{x/s}}\mu(s)x \end{aligned} \tag{1.1}$$

All of the cited growth expressions are empirical in nature, and thus the physical interpretation of their parameters remains, in many cases, open. Furthermore these interpretations are sometimes inherited from their analogues in enzyme kinetics. Monod growth rate is an analogue to the Michaelis-Menten [80] expression, and Andrews inhibitory expression is analogue to that of Haldane [42] (moreover, in the literature it is customary to call this expression Haldane growth). This has attracted some interest in deriving these expressions from first principles or mechanistic approaches, in order to

Name	Expression
Monod(1942)	$\mu_{max}\frac{s}{K_s+s}$
Contois(1959)	$\mu_{max}\frac{s}{K_x x+s}$
Haldane(Andrews 1968)	$\mu_{max}\frac{s}{s+K_s+\frac{s^2}{K_I}}$
Hill or Moser (1958)	$\mu_{max}\frac{s^\alpha}{K_s+s^\alpha}$, $\alpha>0$

Table 1.1: Some common expressions used to model growth rate $\mu(\cdot)$ in response to substrate s and concentration x. An exhaustive list can be found in [7].

gain insight on these parameters. For example the density dependent Contois equation has been theoretically derived by Wang [136], considering that microbes may attach to suspended particles and obstruct other microbes access to substrate.

The Monod equation, being the most popular, has certainly been subject to theoretical enquiries, specially in light of the affinity constant K_s, whose interpretation is usually very dependent on the hypothesis used to derive the model [71].

Tan *et al.* used statistical mechanics based on the number of ligand sites and encounters with molecules to derive the Moser equation [118] (which generalizes Monod Equation see table 1.1). Two years later they used the same approach to derive Haldane and other inhibitory expressions [117]. In a more recent statistical mechanics use, Desmond-Le Quemener and Bouchez [25] derived an expression on the limiting resource introducing the concept of harvest volume, which is the volume in which a microbe has access to substrate, the theory is named Microbial Transition State theory (MTS) for its analogue in chemical reactions.

The concept of relaxation time has been used in the approach proposed by Merchuk and Asenjo [79] to derive the Monod equation. Suppose a system at an equilibrium E_0 is perturbed, name the new equilibrium E_1, the relaxation time, as defined by Roels [105], is the time τ that a system takes to pass from E_0 to $\left(1-\frac{1}{e}\right)(E_1-E_0)$. The process rate μ can be estimated by $\frac{1}{\tau}$. The cellular process of growth, encompasses several steps that go from transport of substrate from the medium, to catabolism and finally biomass formation or anabolism. As in electrical circuits, the relaxation time of processes in series

are added linearly, thus, the relaxation time of cell growth τ_g is the sum of the relaxation times of the previously mentioned stages, namely $\tau_g = \tau_T + \tau_{cat} + \tau_{an}$ or in process rates $\frac{1}{\mu_g} = \frac{1}{\mu_T} + \frac{1}{\mu_{cat}} + \frac{1}{\mu_{an}}$. By supposing that the transport process rate is proportional to the difference of intracellular substrate concentration and other assumptions they derived a Monod expression.

Heijnen and Romein [47] defined a cellular process by dividing it as catabolism and anabolism, and derived rates of each process depending on the ADP and ATP concentrations, respectively. Essentially considering that the sum of ADP and ATP is constant and giving particular forms to the rates of catabolism and anabolism, they derived a Monod expression.

Yet in another mechanistic, but much simpler than the above approaches (perhaps too simple), Lobry deducted the Monod equation from a differential equation with delay, coming from the fact that substrate is not immediately transformed into biomass [73]. Interestingly, from its derivation it follows that $K_s \propto \mu_{max}$.

In each of the mentioned approaches, the K_s value ends up being a function of several other intermediate coefficients. In the case of Tan *et al.* [118] it depends on the energy state of binding sites, in the case of Merchuck and Asenjo it depends on the yield and the mass transfer coefficient of the cell wall [79], and in the case of Heijnen on the particular coefficients taken for the catabolism and anabolism laws. Thus, it makes very unlikely to find a unique interpretation on the nature of K_s.

If the situation for the affinity constant seems complicated from a theoretical standpoint, it does not get any better in terms of measurements. Roques *et al.* [106] found that in activated sludge systems the constant depended on the age of the sludge, as well as on the difference of the input s_{in} and residual substrate s. For that reason, he extended the Monod model to $\frac{s}{s+M+b(s_{in}-s)}$, where M is interpreted as the affinity constant. The term $b(s_{in} - s)$ is interpreted as the inhibitory effect of metabolites produced during biomass formation, because $s_{in} - s$ is proportional to the formed biomass, with b being a dimensionless constant.

In a review of modelling approaches [91], Nielsen pointed out that, to measure tran-

sient kinetics, chemostat systems are better suited. Particularly the K_s constant is not easy to measure in batch experiments, as opposed to μ_{max}, for which the abundance of nutrients of a batch experiment makes the measure easy. Years earlier, Owens had concluded the same [94], but from an experimental point of view.

Button [18] collected extensive data regarding the kinetic parameters of chemostat and batch experiments, particularly he found more than ten different expressions (and methods to estimate them) that referenced affinity. He claimed that μ_{max} and K_s are specific characteristics for each organism and subject to adaptability, as illustrated by an experiment in which, by slowly raising the dilution rate in a chemostat, μ_{max} doubled for a marine yeast [19]. He stressed, therefore, the importance of adaptation of microorganisms to new media in order to obtain good results in measurements. However, he did not question the usefulness of a single parameter to model microbial growth. Jannasch and Egli, in their historical review of microbial growth cultures, argued that a set of parameters for describing a particular microbial culture would be of interest [58].

Grady *et al.* [39], made a review of the possible causes for the variability of parameters. Three causes were identified: culture history, parameter identifiability, and in the case of batch experiments the ratio between initial substrate and biomass. Kovarova and Egli [67] also discussed this variability and noted that for a same strain the pairs μ_{max}, K_s can go from low μ_{max} and low K_s to high μ_{max} and high K_s. They proceeded to argue that certain properties related to growth would be better expressed according to the environmental stimulus: If there is no food shortage a rapid growth (high μ_{max}) will be expressed after some generations, on the contrary, under food limitation its ability to scavenge efficiently will flourish (low K_s). This argument implies that it would be unusual to find low μ_{max} and high K_s, since the two experimental devices to measure growth, select for either trait. Under this line of reasoning Ferenci [35] made the case for an uncertainty principle, as in quantum mechanics, regarding the pair (μ_{max}, K_s) (and in general parameters of equations describing growth rates). The inherent adaptability of life, makes it paradoxical to define it in terms of constants, instead one should strive to consider them as variables, or at least study the bounds in which these parameters may

lie.

Mathematical analysis of growth models in continuous cultures also offers important insights to the theoretical understanding of microbial growth, this is known as chemostat theory [44], and more generally it is an important part of mathematical ecology [65]. Continuous cultures are often described by a set of differential equations representing the mass balance of the inlet, outlet, and reactions taking place inside the reactor. The solution of a system of differential equations represents the concentration in time of the chemicals and microorganisms. Lyapunov stability theory, or simply stability analysis, provides information of the long-term run of the solution of a system of differential equations [60].

The stability analysis of any system of differential equations is very likely to depend on its parameters. In the case of continuous cultures a reactor can be perturbed by varying the input concentration (s_{in}) or the dilution rate (D) (or equivalently, the inverse of the hydraulic retention time). A common way to illustrate the stability analysis of the system are operating diagrams, because it shows in a $D - s_{in}$ plane the state in which a reactor will end in the long run (such as the washout state) [95]. Operating diagrams are well known to process engineers, and have been developed empirically without the need of mathematical models, for example in the nitrification removal of wastewater [111], or to show the outcome of three OTUs interacting in a reactor [59]. More recently operating diagrams have shown the optimal zone for biogas production in anaerobic digestion models [61].

As a closing comment, it is interesting to note that most of the literature on parameters of growth expressions, either dealing with the interpretation, identifiability, or measurements, comes mainly from the 80s to the early 2000s. More recent articles with the same depth on its inherent and unresolved problems seem to be rare. It seems as if, with the turn of the century, the increase of models complexity (ADM1, ASM1, systems biology), computational capabilities and bigger data collection, the most appealing issue is whether a model is able to predict accurately or not.

1.2 Mathematical Modelling in Microbial Ecology

As previously stated in 1838 [128] Verhulst established a model to study population growth known as the logistic equation shown in (1.2), where N is the species abundance, r is the growth rate, and K is the carrying capacity. When N is small (compared to K), the species exponentially grows at a rate r, and as N approaches its carrying capacity its growth rate approaches zero.

$$\dot{N} = rN\left(1 - \frac{N}{K}\right) \tag{1.2}$$

Model (1.2) is a textbook standard in mathematical ecology [65], and serves to illustrate the basis of $r - K$ selection theory in ecology [98]. The paradigm essentially states that two extreme strategies of survival are able to explain the traits of a species [103]. Either that of fast growth (big r), but inefficient resource uptake (conceptually a small carrying capacity), or slow growth but efficient resource uptake. Organisms can then be classified in an $r - K$ continuum of both strategies, and environments characteristics will select for them. This theory has proved consistent in certain domains, notably in botany [78]. Interestingly, even though the logic of $r - K$ selection theory spawns from the logistic equation (1.2), there are no cases of animals or plants that actually follow the predictions of the model [33].

The application of this paradigm in microbial ecology has been defended [5]. It has been studied how microorganisms strategies are reflected by their operon copy numbers [63] and affected by resource scarcity [6]. Naturally, no organism can master all strategies, and in a theoretical approach, models have predicted a metabolic tradeoff between the rate of growth versus yield of biomass [83] and how the different metabolic strategies might explain diversity in homogeneous mediums [11].

In the case of bioreactors, one can think that a batch reactor is a very rich nutrient medium, thus r traits (high μ_{max}) should be selected, while in chemostat experiments nutrients are limited and the environment should select for K traits (low K_s). Velicer

and Lenski conducted an experiment to test this hypothesis by sequentially adapting two different strains to batch and chemostats independently, and then making them compete jointly, [127]. Clearly if the $r-K$ paradigm holds, consistency should be observed in the competition experiments (e.g. a batch adapted bacteria should outperform a chemostat adapted strain in a new batch experiment), nevertheless inconsistency was found in this sense for most of the experiments.

Once again, a paradigm of general ecology does not seem to easily fit in microbial ecology. In general terms the $r-K$ paradigm concentrates too much on the environmental pressure exerted on organisms. As West *et al.* summarized, microbes seem to have a very rich social life, phenomena such as quorum sensing, bacteriocin production, and kin selection might also drive evolution and traits of microorganisms [137].

Which takes us to the point of microbial interactions from a modelling perspective. Lotka [74], and Volterra [129], 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 [65]. However, in its original form, the LV model has its limitations as not being able to model mutualistic behaviour without the species growing unbounded. Taking that into account, Hoek *et al.* [51] proposed a modified version of it to model how available resources shape the cooperative or competitive behaviour in the case of an experiment where two yeast strains delivered an essential amino acid to one another. Extensions of the Lotka-Volterra model have derived what is now known as generalized Lotka-Volterra (gLV) models shown in equation (4.1):

$$\dot{x_i} = \mu_i \left(1 + \sum_{j=1}^{n} a_{ij} x_j\right) x_i \quad i \in \{1, \ldots, n\} \tag{1.3}$$

the terms a_{ij} reflect the effect of OTU j on the growth of OTU i. 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

$(+,-)$, competition $(-,-)$, and ammensalism $(-,0)$ can be recovered [97]. Model (4.1) has been thoroughly analysed, even when the coefficients μ_i and a_{ij} are time dependent and exhibit periodicity (which models seasonal traits) [37], [3]. Usually the condition of negative intra-species interactions is used to show that the system can exhibit periodicity. 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* [17]. 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. Besides, pairwise interactions limit the possibility of accounting for more complex situations, furthermore the linearity fails to capture known pairwise interactions, for example the secretion of a metabolite from one species known to be toxic to another species [84]. 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.

LV models have certainly found their way in the chemostat setting. The predator prey relationships of an amoeba and *E. coli* in a chemostat where glucose was the sole carbon source were investigated by Tsuchiya [122]. The resulting three dimensional dynamical system was able to reproduce the oscillatory behaviour observed in the experiments. Also, a general theoretical account of interactions in the form of LV model coupled to a chemostat system with one substrate was performed by Saito *et al.* [107].

Chemostat theory naturally deals with interactions among microorganisms. A cornerstone of the theory is known as the competitive exclusion principle (CEP). It roughly states that in a microbial community of n OTUs, whose growth rates are increasing functions of one limiting substrate, then only one or none will survive. A proof can be found as early as 1977 [55] for the Monod function, and broad generalizations can be found [141], even to more abstract mathematical settings [56]. Experimental demonstration of the principle was accomplished in 1980 [57]: By independently cultivating strains and fitting a Monod expression to each of them, they predicted the outcome of the two strains

growing together, proving the usefulness of rigorous mathematical analysis in bioprocess. Particularly, from the CEP one already knows that more complex growth laws or model structures are needed in bioreactors, in order to have the observed diversity and functional redundancy one observes in engineered ecosystems [112]. Interactions as a hypothesis for coexistence in a chemostat setting have readily been pointed out in the literature [72], which has been studied under very general hypothesis on the growth function for two OTU and one limiting substrate [34]. Alternatively, coexistence can also be achieved if the growth laws are non-monotonic (such as the Haldane function presented before) on the substrate [44].

Di and Yang studied generic combinations of OTUs (up to three), substrate (up to three), and growth functions (monotonic and inhibited by either OTU or substrates) in order to explore the outcome of a reactor in terms of stability (presence of oscillations) and productivity, by the use of simulation [26]. More classically, detailed mathematical analysis of a system composed of two OTU and a single substrate, in which one of the growth rates is non-monotonic, shows that coexistence can be achieved (under certain operational conditions) [44]. In a more complex model a chemostat system composed of three bacteria x_1, x_2, x_3 performing denitrification, that is the conversion of nitrite and nitrates into nitrogen gas, was studied by Vasiliadou *et al.* [126]. They studied three chemostat models, for different combinations of the involved OTUs, namely $x_1 - x_2$, $x_2 - x_3$, and $x_1 - x_2 - x_3$, where the growth laws were not all monotonic. They constructed operating diagrams for each model to compare the different denitrification rates. The diagrams showed a myriad of possible scenarios in terms of the outcome of the reactor, from where they could deduct the most performing combination of bacteria, depending on the operating parameters.

In other works that derive coexistence from a chemostat system, Pilyugin and Waltman analytically proved the existence of limit cycles occurring in a chemostat system conformed by 1 OTU and 1 substrate with a monod law, and variable yield depending on substrate s [99]. Then they proceeded by simulations to show that when another OTU is added, with a fixed yield and a monod law, coexistence can be achieved. However, the

yield substrate dependence is completely artificial, in the sense it does not attempt to model any phenomena.

A paradigm in biology is that evolutionary time scales are larger than ecological time scales, however the situation is being reconsidered under what is known as eco-evolutionary dynamics [96]. For example, in a chemostat experiment it was shown how the acquisition of traits in some generations of algae (evolution) to protect against a predator shifted the ecological dynamics. This implies that these two time scales, ecological and evolutionary, might be more intertwined than what is classically thought of [12], specially in the microbial world. It is in that sense that Gorter *et al.* argues that interspecies interactions (such as the $a_{ij}\, i \neq j$ coefficients) should not be thought of as constants [38].

This section concludes by noting that the division between ecology and microbial ecology and the insistence on the top-down approach of introducing ecological concepts in microbiology might not be the best way to go [101], and perhaps it has created more theoretical problems than it has solved. Ecological theory should, by definition, include microbial ecology insights. Widder *et al.* make the case for an integrated collaboration of theory and modelling along with data collection and method developments in order to better predict and understand microbial communities [138]. Surely such an effort should benefit general ecology as well. Thermodynamics, or more precisely bioenergetics, offer some solid ground to develop microbial theory and is the basis of the next chapter.

1.3 Bioenergetics Models for Microbial Growth

The highly ordered state that life represents can only be sustained by the continuous use of the available energy gradients in the environment. It is rather appropriate then, to search for relationships between the physiochemical properties of the environment and the growth rates and yields of substrate to biomass conversion $y_{s/x}$ of a microbial community. Implicit in the following methods to be presented is the conceptualisation of a molecule of biomass to describe the mean contents of a cell and associated thermodynamic properties such as entropy and enthalpy of formation. Hoover and Allison in 1940 had already

proposed a formula for biomass composition of $C_{19}H_{32}N_2O_9$ obtained from the analysis of organic dry matter of batch cultures of *Rhizobium meliloti* [53]. Edwin Battley in the 50's went a step further on measuring besides a mean formula, also the thermodynamic properties of yeast cells growing anaerobically and aerobically. At the time he concluded that the mean composition depends on the substrate used, but in a more recent account of his [9], he revised this claim arguing for the measurement errors given the technical capabilities at the time. The direct chemical analysis of lyophilized cells after centrifugation is preferred now, compared to the direct combustion of dry biomass; the difference in both methods is specially reflected in the oxygen content of the cell. In the same article he argued that the fabric of cells of a given microbial species grown exponentially on a substrate are expected to have the same composition regardless of the substrate used, but one should expect them to have a different formula when grown on limited nutrient medium such as continuous cultures. The case is also made for the use of C-mol of biomass as a formula, that is, normalizing the biomass formula to contain one atom of carbon, a convention adopted in this manuscript.

A good starting point to understand the progress and paradigms in these estimation methods is the 1991 review of Heijnen and Van Dijken [48]. They distinguished between metabolic approaches which use the description of reactions within the cell and black box models which deals with the carbon source, electron donor and acceptor and the N-source. The advantage of black box methods is that one does not require specific knowledge on the biochemistry of particular organisms, the trade-off being less accurate predictions. Regardless the methods label, all of them aim to establish a macrochemical equation describing the process of biomass formation. Several approaches in either category have been developed, the relationship between the environmental variables and the yield were done by the coupling of the set of chemical reactions to be considered. The coupling of chemical reactions is a linear combination of them, so in order to choose the correct combination, all methods relied on a parameter calibrated from experimental and observed data. Various of these parameters made reference to energy efficiencies: ratios between the energies of different reactions that describe the process. It was observed

that even though each method arrived at interesting correlations some of them suffered from not being able to represent the second law of thermodynamics appropriately (already pointed out in detail by Roels [104]) or that their value depended on the frame of reference on which the thermodynamic properties were taken from (inconsistency). Taking into account the observation made by Roels [105], they ended up proposing a black box method that relies on the dissipated Gibbs free energy per C-mole of biomass of the machrochemical equation which seemed to be the same for similar C-sources (in terms of the length of its carbon chain and its reduction state), they had a mean error of 20% for the prediction of yields with their method.

Curiously, Heijnen did not survey the Thermodynamic Electron Equivalents Method (TEEM) presented in 1965 by McCarty to estimate yields [77] which can be classified as a black box method. More than 42 years later the method was republished under new considerations on the half reactions involved, after Xiao and Vanbriesen [142], a year earlier, had modified the original method to take into account the carbon degree of reduction and nitrogen balances. As a consequence, Kleerebezem and Van Loosdrecht [64] discussed three methods to estimate $y_{x/s}$: the Gibbs energy dissipation method by Heijnen, McCarty's updated TEEM method, and the ATP balancing method (which is a metabolic approach). They concluded that for most of the cases the TEEM method and Gibbs dissipated energy predictions behave similarly- their equations being even equivalent in some cases- and that if one wishes to open either black box, a method quite similar to the ATP balancing method should result. The Gibbs energy dissipation method has been further expanded to account for environments where the energy gradients are more limited [113].

A theoretical explanation on why the dissipated Gibbs energy might be such a good predictor has been provided by Von Stockar and Liu [130]. A description of the non-equilibrium system that a cell represents is given in terms of differential equations describing the entropy flowing in and out of the cell. They elucidated the fact that cells are able to remain in a non steady state, by either exchange of chemical entropy, or heat, even archea that take up heat were later found [70]. They concluded that cells growth

is driven by Gibbs free energy, which englobes both heat and chemical entropy, Battley had reached the same conclusions years earlier from experimental measurements of *S. cerevisae* growth [9]. The theoretical framework of Van Stockar and Liu also allows to explain the decrease of yields of cultures at slow growth rates [120].

Taking a larger view of MC modelling and bioenergetics, the use of an already sequenced genome allows to link the enzymatic activity to the genes and to the metabolic network. A list of r metabolic reactions concerning m metabolites can then be constructed, which is represented in a stoichiometric matrix $S \in \mathbb{R}^{m \times r}$, where S_{ij} stand for the stoichiometric coefficient of metabolite i in reaction j. Note that coordinate-wise positive vectors v solving the equation $Sv = 0$ represent possible combinations of the identified reactions; any model searching for such a vector can be classified as a constraint based model (CBM). Flux balance analysis (FBA) is a type of CBM that adds a linear objective function in order to find such a v [116]. Typical functions include maximization of growth rate, or minimizing or maximizing the production of ATP. Unfortunately, CBM methods suffer from lack of standardization, and the construction of matrix S usually requires expert knowledge to fill the gap from the missing enzymes in order to complement metabolic pathways.

This introduction ends by noting that contrary to the prolific literature on methods able to predict yields based on energy gradients of the environment, much less has been said about other kinetic parameters from both empirical and theoretical enquiries [131]. As mentioned, Desmond-Le Quemener and Boucher proposed a new growth rate that relates a threshold energy to be attained in order for a cell two divide [25], but even in their approach no light is shed on μ_{max}. In more recent findings, using a FBA approach, growth maximization and a universal upper limit on the rate of Gibbs energy dissipation- that is the dissipated Gibbs energy times the growth rate- is presented as an underlying principle describing microbial growth [90]. It is able to explain the metabolic shift from respiration to aerobic fermentation as well as earlier works on the Pareto frontier of kinetic parameters [109].

1.4 Thesis outline

Chapter 2 states some definitions and technical concepts that are used throughout the thesis. Particularly, the Gibbs energy dissipation method is explained. It was the preferred black box method, because the TEEM method relies on different equations depending on the carbon source, while the Gibbs energy method uses one unique equation, but it is the Gibbs dissipated energy that changes depending on the carbon source.

Chapter 3 studies the problem of functional classification of OTU in a continuous reactor. Widder *et al.* [138] reviewed the current challenges in microbial ecology with a focus on modelling demands. It is crucial to link functionality to the members of a microbial community through models taking into account stoichiometry with an FBA formulation, this can be seen as the spirit of Chapter 3, even though the model does not qualify as a CBM, it is based the abundances of the OTU in the reactor and a stoichiometric matrix. It strongly uses the notion of asymptotic observer [27] in order to bypass the knowledge of the growth function. First a discussion is given on the use of the bioenergetics yield methods and asymptotic observers to obtain the effective yields of functional groups in a chemostat system where only the total biomass and substrates are measured. The ideas are applied to a nitrification process. In a second movement, a mixed integer linear program (MILP) can be constructed for the classification of the OTU in functional groups of the same nitrification process.

Widder *et al.* [138] also argued that kinetic models to better represent the growth of each OTU in an MC are strived for, chapters 4 and 5 go in that direction. Chapter 4 deals with the growth functions of the classified OTU on the previous chapter. First a mathematical analysis of a model involving pairwise linear interactions (as in a gLV model) coupled with substrate limitation for a two step nitrification process results in operating diagrams revealing unintuitive behaviour of such a system. Operating diagrams were also done in the course of this thesis as a technical aid for the work of Hannaki *et al.* [43]. The concept of interaction function generalizing pairwise interactions is then introduced. A method to reconstruct the shape of such an interaction function that may vary through time is presented.

In the line of understanding the growth rates of OTU chapter 5 takes a radical twist to dive into the nature and origins of substrate limitation from the scope of MTS theory. Chapter 5 begins by noting that a first order approximation of the MTS expression one gets a Monod function, so a new interpretation of the affinity constant is discussed. The chapter goes on to extend the MTS theory in two senses: showing how its formulation can naturally entail in density dependence and product inhibition. The latter is used to explain coexistence in a real experimental setting on one limiting substrate.

Chapter 6 ends the thesis with conclusions and perspectives of the studied topics.

Chapter 2

Technical framework

The key tools and definitions used throughout the manuscript are stated and exemplified in this chapter. The dynamical systems section was taken from [44], the chemical reactions section from [105], and the metabolism section from [64].

Some notation used:

- Given a vector $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$, the function $\text{diag}(v)$ stands for:

$$\begin{aligned} \text{diag} : \mathbb{R}^n \to & \qquad \mathbb{M}_{n\times n}(\mathbb{R}) \\ v \to & \begin{pmatrix} v_1 & 0 & \ldots & 0 \\ 0 & v_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & v_n \end{pmatrix} \end{aligned} \tag{2.1}$$

- Let $n \in \mathbb{N}$ then $[n] := \{1, \ldots, n\}$.

- For a scalar $a \in \mathbb{R}$ the term $a_{m\times n}$ denotes the $m \times n$ matrix containing a in all its entries.

- Let M be a matrix, then $M_{i\bullet}$ denotes the i-th row, analogously, $M_{\bullet j}$ denotes de j-th column.

2.1 Dynamical Systems and the Chemostat Model

The main types of models used in this thesis are dynamical systems in the form of differential equations. To begin consider the system (2.2).

$$\begin{aligned} \dot{x} &= f(x) \\ x(t_0) &= x_0 \end{aligned} \tag{2.2}$$

It is classical to show that under conditions on function f there is a unique function $t \mapsto x(t)$ that solves the system (2.2) the proof can be found in any standard textbook on differential equations, for example [60], the following definition and theorem were taken from the same book.

Definition 1. *x^{eq} is an equilibrium point of system (2.2) if $f(x^{eq}) = 0$. They can be:*

- *Stable if for each $\epsilon > 0$ there exists $\delta > 0$ such that $\|x_0 - x_{eq}\| < \delta \Rightarrow \|x(t)\| < \epsilon$, $\forall t \geq 0$.*
- *Unstable if it is not stable.*
- *Asymptotically stable if stable, and δ can be chosen such that $\|x_0 - x_{eq}\| < \delta \Rightarrow \lim_{t\to\infty} x(t) = x^{eq}$*
- *Globally asymptotically stable if it is asymptotically stable, and for every $x_0 \in dom(f) \Rightarrow \lim_{t\to\infty} x(t) = x^{eq}$*

One can characterize a stable point in terms of the jacobian matrix of function f

Theorem 1. *Let x^{eq} be an equilibrium. And let $A = \left.\frac{\partial f}{\partial x}\right|_{x=x^{eq}}$ be the Jacobian matrix of function f evaluated at x^{eq}. Let λ_i $i \in [n]$ be the eigenvalues of matrix A Then:*

- *x^{eq} is asymptotically stable if $Re(\lambda_i) < 0$ for all $i \in [n]$*
- *x^{eq} is unstable if $Re(\lambda_i) > 0$ for some $i \in [n]$.*

The case when $Re(\lambda_i) = 0$ is case dependent of the system at hand.

In this thesis, the models used correspond to the class of chemostat models [44]. The most simple chemostat model consists of two state variables x and s, that represent the concentrations in mass per volume of the biomass of an OTU and the limiting substrate used for growth, respectively, as seen in equations (2.3). $\mu(x, s)$ represents the growth rate (also known as process kinetics) depending on the state variables, D is the dilution rate of the reactor, both in $[time]^{-1}$ units. s_{in} is the concentration of substrate s at the entry of the reactor. $y_{x/s}$ is the yield of biomass per substrate.

$$\begin{aligned} \dot{x} &= (\mu(x, s) - D)x \\ \dot{s} &= (s_{in} - s)D - \tfrac{1}{y_{x/s}}\mu(x, s)x \end{aligned} \tag{2.3}$$

This type of system can be easily extended to multiple substrates and multiple OTU using vector notation and the diag operator defined at the beginning of the chapter. Note $x \in \mathbb{R}^n$ the vector containing the concentration of each OTU. Let $s \in \mathbb{R}^m$ the vector containing the concentration of measured molecules of the environment and s_{in} the vector containing the concentration of molecules s at the entry of the reactor. $\mu(x, s) \in \mathbb{R}^n$, the vector containing the growth rate of each OTU, the one can write system (2.4).

$$\begin{aligned} \dot{x} &= \operatorname{diag}(\mu(x, s) - D_{n\times 1})x \\ \dot{s} &= (s_{in} - s)D + Y \operatorname{diag}(\mu(x, s))x \end{aligned} \tag{2.4}$$

Eventually instead of $D_{n\times 1}$, one could have different dilution rates for each variable. Matrix $Y \in \mathbb{R}^{m\times n}$ is the stoichiometric matrix. The entries Y_{ij} are denoted y_{s_i/x_j} to recall the classical yield expression of microbiology, however the values y_{s_i/x_j} have a sign to account for production (+) or consumption (-). Finally note that at equilibrium one has $x_i = 0 \ \vee \ \mu_i(x, s) = D \quad \forall i \in [n]$.

The reaction invariant, which is a quantity that is conserved and does not depend on the growth rate, is particularly useful for circumventing the knowledge of process kinetics. It will be exploited in chapters 3 and 4.

Definition 2. *Given system* (2.4). $z := -Yx + s$ *is called the reaction invariant.*

Lemma 1. *If matrix Y is constant, and $D > 0$ then $z(t) \to s_{in}$.*

Proof.

$$\begin{aligned}
\dot{z} &= -(\dot{Y}x) + \dot{s} \\
&= -Y\dot{x} + \dot{s} \\
&= -Y(\operatorname{diag}(\mu(x,s) - D_{n\times 1})x) + (s_{in} - s)D + Y \operatorname{diag}(\mu(x,s))x \\
&= -Y \operatorname{diag}(\mu(x,s)x + Y \operatorname{diag}(D_{n\times 1})x + (s_{in} - s)D + Y \operatorname{diag}(\mu(x,s))x \\
&= (s_{in} - s + Yx)D \\
&= (s_{in} - z)D
\end{aligned}$$

Note that $z = s_{in}$ is the only equilibrium for equation $\dot{z} = f(z) = (s_{in} - z)D$. The jacobian of $f(z)$:

$$D_z f(z) = \operatorname{diag}(-D_{m\times 1}) \tag{2.5}$$

Which shows that the jacobian is a negative definite matrix, thus $z^{eq} = s_{in}$ is a global asymptotic equilibrium. □

2.2 Chemical Reactions

The reason for this section is just to formalize some computations (as done in Roels [105]), to make the link of the stoichiometric matrix presented before and cellular metabolism presented below.

Cells can only take up soluble particles, and thus chemical reactions appearing in this manuscript are assumed to take place in medium where water is the solvent. 95% of a cell's biomass is composed of Carbon (C), Nitrogen (N), oxygen (O), and Hydrogen (H), so in most cases reactions dealing with molecules of the type $\mathrm{C_{n_1}N_{n_2}O_{n_3}H_{n_4}}$ will be the case.

Definition 3. *Consider a set of n molecules composed of m elements $C^1, \ldots, C^m$, rep-*

resented by eq (2.6).

$$\{C^1_{a_{1j}} \dots C^m_{a_{mj}}\}_{j=1}^n \tag{2.6}$$

where a_{ij} is the number of atoms of elements i that compose molecule j. The matrix $m \times n$ matrix A composed of entries a_{ij} is called the atomic matrix.

Definition 3 extends to the case with charged molecules. An extra element e is added to each molecule accounting for its electric charge.

Example 1. *Consider the following set of molecules*

$$\{\mathrm{C_6H_{12}O_6}, \mathrm{O_2}, \mathrm{H_2O}, \mathrm{HCO_3}^-, \mathrm{H}^+\}$$

Its atomic matrix is given by:

	$\mathrm{C_6H_{12}O_6}$	$\mathrm{O_2}$	$\mathrm{H_2O}$	$\mathrm{HCO_3}^-$	H^+
C	6	0	0	1	0
H	12	0	2	1	1
O	6	2	1	3	0
e	0	0	0	-1	1

Given a set of molecules (2.6) a reaction is possible if one can find coefficients $v_1, \dots, v_n$ not all zero, such that:

$$\sum_{i=1}^{n} v_i a_{ij} = 0 \quad \forall j \in [m] \tag{2.7}$$

Or equivalently in vector notation let $v = (v_1, \dots, v_n)^\top$, one searches $v \in \mathbb{R}^n \setminus \{0\}$ such that $Av = 0$. In that sense the kernel of the atomic matrix provides us information concerning the stoichiometry [75].

Definition 4. *Let $v \in \ker(A) \setminus \{0\}$ then v is a possible stoichiometry that defines a reaction on a set of molecules $\{C^1_{a_{1j}} \dots C^m_{a_{mj}}\}_{j=1}^n$. For every $j \in [n]$, such that $v_j < 0$,*

one says that molecule j is a reactant. Inversely for any j such that $v_j > 0$ one says that molecules j is a product of reaction.

Example 2. *In example 1 the kernel of the atomic matrix A is given by:*

$$\ker(A) = \langle(-1, -6, 0, 6, 6)^\top\rangle \tag{2.8}$$

vector $v = (-1, -6, 0, 6, 6)^\top$ defines the following reaction

$$\mathrm{C_6H_{12}O_6} + 6\,\mathrm{O_2} \longrightarrow 6\,\mathrm{HCO_3}^- + 6\,\mathrm{H}^+ \tag{2.9}$$

Note that the stoichiometric coefficient of water is zero in order to balance this reaction.

Since $\ker(A)$ is a vector space, then for any $\lambda \in \mathbb{R} \neq 0$, $\lambda(-1, -1, 0, 6, 6)^\top$ is also a possible stoichiometry. Particularly if $\lambda = -1$ it means that the reactants become products, and the products become reactants.

The concept of Gibbs free energy (G) is of importance in chemical reactions, since the criterion for the most probable state of a system at constant pressure and temperature is the minimum of Gibbs free energy [68]. Like any energy function the importance lies in the difference of energy between two states, rather than its absolute value. The Gibbs free energy $G = H - TS$ unifies two other concepts, namely H the enthalpy and S the entropy, and T the temperature. Thus a change in the Gibbs free energy may be due to a change in the entropy or enthalpy or both.

The Gibbs free energy change ΔG of a reaction is the key to know in which direction the reaction spontaneously takes place. It is calculated from the Gibbs free energy of formation of each molecule and a possible stoichiometry.

Definition 5. *Given a set of molecules $\{C^1_{a_{1j}} \dots C^m_{a_{mj}}\}_{j=1}^n$ and the vector $Gf = (Gf_1, \dots, Gf_n)^\top$ where entry Gf_j represents the Gibbs free energy of formation of molecule j. Let $v \in \ker(A)$ define a reaction. Then the Gibbs free energy change of the reaction v corresponds*

to:

$$\Delta G_v = Gf \cdot v \tag{2.10}$$

If there is no confusion on the reaction being used it will simply be noted ΔG*.*

If $\Delta G < 0$ then the reaction is said to be spontaneous or thermodynamically feasible. On the contrary if $\Delta G \geq 0$ the reaction is non spontaneous, or thermodynamically infeasible. Note that if $v \in \ker(A)$ defines a non-spontaneous reaction, then $-v$ defines a spontaneous reaction (unless $\Delta G = 0$).

The standard state of a substance is a reference point used to calculate certain properties. The standard state in this manuscript is referred to a temperature (T_S) of 298K, with a concentration of 1 molar (M) in the case of soluble particles, and 1 atm of pressure for gases. The superscript 0 denotes the standard state of a given quantity. For example, the standard Gibbs formation energy of a molecule is noted as Gf^0 and the standard Gibbs energy change of a reaction ΔG^0.

Example 3. *In example 1, let* $Gf^0 = (-917.2, 0, -237.2, -586.9, 0)^\top$ *(units in [kJ/mole]). Let* $v \in \ker(A)$ *define a reaction, then* $v = \lambda(-1, -6, 0, 6, 6)^\top$ *for some* $\lambda \in \mathbb{R}$*. Therefore* $\Delta G^0 = -\lambda 2604.2$ *[kJ]. The reaction is then spontaneous if* $\lambda > 0$*. Note that in the previous example, the Gibbs free energy of the reaction is in [kJ], for practical reasons one would like to express this quantity as [kJ/mole]. For example if* $\lambda = 1$ *then one could say that the reaction has a standard Gibbs free energy change of* $\Delta G^0 = -2604.2[kJ/mole\, of\, \mathrm{C_6H_{12}O_6}]$.

Reactions may occur at a different temperature T, and at different chemical activities compared to the standard state (which is the case in most systems to be studied in this manuscript). A general formula [64] for the Gibbs free energy of formation of a chemical species i is presented in equation (2.11).

$$Gf_i = Gf_i^0 \frac{T_S}{T} + \frac{T_S - T}{T} Hf_i^0 + RT \ln(a_i) \tag{2.11}$$

where R is the ideal gas constant, Hf_i^0 and a_i are the standard enthalpy of formation and the chemical activity of chemical species i in the reaction. Equation (2.11) should contain a term $\ln(a_i/a_i^0)$ instead of $\ln(a_i)$ in order to have a dimensionless quantity inside the logarithm, but since it is assumed $a_i^0 = 1$, this will be omitted. The chemical activity in this manuscript, will be the concentration of substance i in molars for diluted chemical species, and the fraction between the partial pressure and the total pressure for gases. Since equations coming from bioprocesses are used, water is the common solvent in reactors, and its chemical activity is always 1 (thus the log term disappears). For insoluble particles the chemical activity is always 1 as well, which is the case of biomass.

Example 4. *In example 3, suppose the medium has pH 7, is saturated in O_2 and HCO_3^-, and a glucose concentration of* $0.1[gr/l]$. *The chemical activities are:*

$$\begin{bmatrix} 0.1/180 & 0.2 & 1 & 0.004/29 & 10^{-7} \end{bmatrix} \tag{2.12}$$

The Gibbs free energy is $\Delta G = -2927[kJ/mole\,of\,\mathrm{C_6H_{12}O_6}]$. *About a 12 % difference with respect to* ΔG^0.

Note: *The chemical activity corresponding to oxygen and HCO_3^- are determined from the atmospheric partial pressure of oxygen and CO_2 (in [atm]) (atmosphere is composed 20% oxygen and 0.04 % CO_2), divided by Henry constant in the case of HCO_3^-, since the Gibbs free energy of formation for HCO_3^- is considered as a diluted molecule. At pH=7, HCO_3^- is the most abundant inorganic carbon form [68].* $R = 0.0083[kJK^{-1}mol^{-1}]$.

In the former examples one could have chosen $\mathrm{CO_2}$ instead of $\mathrm{HCO_3^-}$ (noting that $\mathrm{CO_2 + H_2O \longleftrightarrow HCO_3^- + H^+}$) in the set of molecules. For stressing the importance of the chemical activity of charged species in ecosystems (for example the pH) the latter formulation is preferred throughout this manuscript.

Definition 6. *Given a set of n molecules, a stoichiometry v, and the vector of chemical activities given by vector $a \in \mathbb{R}^n$. The reaction quotient $Q_v(a)$ is defined as*

$$Q_v(a) = \prod_{i=1}^{n} a_i^{v_i} \tag{2.13}$$

When the Gibbs energy change of a reaction becomes zero, the reaction is said to be in ***equilibrium****. The equilibrium constant (at a given temperature and pressure) K_c is the value of the reaction quotient at an equilibrium.*

Given a stoichiometry v and using equation (2.11) the relationship between the reaction quotient, equilibrium and equilibrium constant becomes clear from equations (2.14), (2.15), (2.16), and (2.17).

$$\Delta G_v = 0 \tag{2.14}$$

$$\frac{T_s}{T}\Delta G_v^0 + \frac{T_S - T}{T}\Delta H_v^0 + RT\sum v_i \ln(a_i) = 0 \tag{2.15}$$

$$\frac{T_s}{T}\Delta G_v^0 + \frac{T_S - T}{T}\Delta H_v^0 + RT\ln\left(\prod_{i=1}^{n} a_i^{v_i}\right) = 0 \tag{2.16}$$

$$\frac{T_s}{T}\Delta G_v^0 + \frac{T_S - T}{T}\Delta H_v^0 + RT\ln\left(K_c\right) = 0 \tag{2.17}$$

Equilibrium constant allows to calculate the chemical activity of one of the molecules of the reaction at equilibrium provided the activity of the rest of the molecules involved in the reaction. At a given temperature and pressure, any a satisfying $K_c = Q_v(a)$, also represents a state of equilibrium. For example, this proves useful in calculating the form of nitrogen compounds and inorganic carbon fractions depending on the pH.

There is no general relationship that links the rate of a chemical reaction with its Gibbs free energy change. The only thing known is that if the reaction is at equilibrium, then the rate of the reaction going in one sense is the same as the rate going in the other sense.

Finally, it is well known that one can add two reactions v_1 and v_2 with different molecules just by adding the corresponding stoichiometric coefficients. Even though v_1 and v_2 might be of different sizes, one can write $v_1 + v_2$ without ambiguity (which is the case for the anabolic and catabolic reactions). For the sake of completeness, a mathematical proof of this can be seen in the appendix A.1.

2.3 Cellular metabolism

This section formalizes metabolism as a chemical reaction using the dissipated Gibbs Energy method [64], which is used in chapters 3 and 5.

From an evolutionary perspective, a microbial species will be able to thrive as long as it is able to profit from a thermodynamically feasible reaction of the set of molecules of their environment [76]. This is the rational for determining the possible catabolic reactions of a particular species in a given ecosystem. By definition, a catabolic reaction must then be thermodynamically feasible. Example 4 could represent the catabolism of *E. coli* growing on glucose under aerobic conditions.

The idea of a molecule of biomass (and its formation energy) as characterizing the mean composition of a cell is an abstraction that serves to build a balanced reaction representing anabolism. For example $CH_{1.8}O_{0.5}N_{0.2}$ represents a C-mole of biomass [64], and whether one would like to include other elements such as phosphorus depends on the intentions in mind. Its composition implies that forming biomass requires a nitrogen source and a carbon source as seen in example 5 below. Which carbon and nitrogen sources to use in order to represent anabolism seems unclear from the literature. Ammonium for instance seems to be preferred when no other nitrogen source is present. When no organic carbon source is present in the environment bicarbonate can be used, in other cases it is more likely that the organic carbon source is used, since the reaction is thermodynamically more favourable.

Example 5. *Consider the set of molecules* $\{C_6H_{12}O_6, HCO_3^-, NH_4^+, H^+, H_2O, CH_{1.8}O_{0.5}N_{0.2}\}$ *and* $Gf^0 = (-917.2 - 586.9 - 79.40 - 237.2 - 67)^\top$.

$ker(A) = \langle(-0.175, 0.05, -0.2, 0.25, 0.4, 1)\rangle$ *which gives the reaction:*

$$0.175\,C_6H_{12}O_6 + 0.2\,NH_4^+ \longrightarrow 0.05\,HCO_3^- + 0.25\,H^+ + 0.4\,H_2O + CH_{1.8}O_{0.5}N_{0.2}$$

$$\Delta G^0 = -14.83[kJ/C - mole\,of\,Biomass]$$

Consider the activities $a = \begin{bmatrix} 0.1/180 & 0.004/29 & 0.1/18 & 10^{-7} & 1 & 1 \end{bmatrix}$, *then* $\Delta G = \Delta G_0 + RT\ln(Q_v) = -20[kJ/C - mole\,of\,Biomass]$

Definition 7. *Consider a set of molecules $\mathcal{C}_{cat}$, which does not include a biomass molecule, with A_{cat} its stoichiometric matrix. A catabolic reaction is a feasible reaction of $\mathcal{C}_{cat}$. Consider another set of molecules $\mathcal{C}_{an}$ containing a biomass molecule in its n-th coordinate. Let $v_{cat} \in \ker(A_{cat})$ and $v_{an} \in \ker(A_{an})$ such that $v_{an}(n) = 1$. A metabolic reaction is a reaction corresponding to $v_{met} = v_{cat} + v_{an}$.*

The former definition is just to remind than when referring to metabolism, in this thesis, is always per $C - mol$ of biomass. Also, since it is customary to already have a balanced reaction to describe catabolism v_{cat}, one usually writes the catabolic reaction as a multiple of v_{cat}, that is λv_{cat}, with $\lambda > 0$.

Even though the formation of biomass (the anabolic reaction) can be thermodynamically feasible in itself, recall that it is an abstraction in order to summarize the energetic cost of the mean cell composition that was constructed from calorimetric measurements [9]. The whole process of building all of the life blocks required for a cell to properly function involves a lengthy number of intermediate reactions. At each step not all of the energy can be harnessed and a fraction must be dissipated. Therefore a dissipated energy ΔG_{dis} is added to the anabolic reaction Gibbs free energy to represent the real energetic cost of biomass synthesis $\Delta G_{an} + \Delta G_{dis} > 0$. According to studies, the value of ΔG_{dis} depends mainly on the carbon source and its reduction number in the case of heterotrophic growth [64] and empirical expressions are proposed. In order to overcome the energetic cost of forming a C-mole of biomass, the cell must then recur to an energy yielding catabolic reaction from its environment.

Definition 8. *Let $\Delta G_{dis} > 0$, v_{cat} be a catabolism with Gibbs free energy ΔG_{cat}, v_{an} be an anabolism with Gibbs free energy ΔG_{an}, and $\lambda > 0$. We say a metabolism $v_{met} = \lambda v_{cat} + v_{an}$ dissipates ΔG_{dis} if equation (2.18) holds. λ is referred to as the* ***metabolic coupling coefficient****.*

$$\lambda \Delta G_{cat} + \Delta G_{an} + \Delta G_{dis} = 0 \tag{2.18}$$

Example 6. *Suppose $\Delta G_{dis} = 236[kJ/C - mole\, of\, Biomass]$. Let v_{cat} and v_{an} be as in*

examples 4 and 5. Then compute $\lambda = -\dfrac{\Delta G_{dis} + \Delta G_{an}}{\Delta G_{cat}}$. *By equation* (2.18)*, is clear that* $v_{met} = \lambda v_{cat} + v_{an}$*, dissipates* ΔG_{dis} *(*$\lambda = 0.0737$*). A molecule's stoichiometric coefficient of* v_{met} *is denoted (without sign)* $y_{s/x}$*. It is the yield of substrate consumed per mole of biomass formed. Inversely* $y_{x/s} = \frac{1}{y_{s/x}}$*, denotes the yield of biomass formed per substrate consumed. For example in the case of glucose this gives* $y_{s/x} = 0.249 \dfrac{mol\, Glucose}{C - mol\, Biomass}$.

Finally suppose the metabolism of each OTU i is represented by vector v_i, and let Π be the projection from the space of all molecules, to the coordinates representing s of the chemostat system (2.4). To construct the stoichiometric matrix the columns of matrix Y are formed by: $Y_{\bullet i} = \Pi v_i$. When the Gibbs free energy of the metabolism is corrected by the chemical activity of molecules s, matrix Y depends on s, and lemma 1 no longer holds. A model under this circumstances is explored in chapter 5.

2.4 Optimization Theory

Throughout the thesis several optimization problems are formulated. Given a set S and a function $f : S \to \mathbb{R}$ an optimization problem can be generally formulated as (2.19) where *s.t.* is read as "subject to". The set S is called the constraints, and the function f is called the objective function. However there is no general procedure for solving such a problem [16]. If the set S is non-empty we say that the problem is feasible, and any point in set S is a feasible point. A local optimum x point is a feasible point such that there exists a neighbourhood V of x $f(x) \leq f(y)$ for every $y \in V \cap S$. A global optimum x is a feasible point such that $f(x) \leq f(y)$ for every $y \in S$.

$$\begin{aligned} &\min f(x) \\ &s.t.\ x \in S \end{aligned} \tag{2.19}$$

Optimization problems are classified for particular instances of the function f and set S. A mixed integer linear problem (MILP) is a problem where function f is linear and the set S consists of a linear constraint ($Ax = b$) and some of the entries of x must take

integer values ($x \in \mathbb{R}^m \times \mathbb{Z}^n$) (2.20). When there are no integer values, the problem simply becomes a linear problem, and can be solved by the use of the simplex-dual algorithm. For solving the case with integer values, a common method is the branch and bound algorithm both using the simplex-dual algorithm to its advantage [108]. In several open-source projects these algorithms are already deployed, for example Pyomo [46]. However, we used an academic license of the Gurobi solver [41]. This type of problems are solved in chapter 3.

$$\begin{aligned} f(x) &= c^\top x \\ S &= \{x \in \mathbb{R}^m \times \mathbb{Z}^n |\ Ax = b\} \end{aligned} \tag{2.20}$$

Another common case is that of smooth non-linear problems (2.21). In this case the objective function is a convex differentiable function, and the set S is defined by the points satisfying $g(x) = 0$ and $h(x) \leq 0$) both being differentiable functions (possibly multivalued, where the inequality should be read component-wise). Interior point methods or barrier methods are usually employed [16] and can be found in open source software as well [46]. This type of problem can sometimes have multiple local optimum points. A type of non-linear problem is also solved in chapter 3.

$$\begin{aligned} f(x) &\in C^1(\mathbb{R}^n) \\ S &= \{x \in \mathbb{R}^n |\ g(x) = 0, h(x) \leq 0\} \end{aligned} \tag{2.21}$$

Finally an optimal control problem (2.22). The objective function is now a functional (a function that uses functions as inputs) that takes the solution of a differential equation and a control u, the control u affects the solution of the differential equations. The set of constraints becomes a functional space such as the set of all measurable functions, which is infinite dimensional. This type of problem makes part of control theory [121]. A method for solving them is called the shooting method which is implemented in the open source software bocop [119].

$$f(u) = \int_0^T r(x,u)dt$$
$$S = \left\{ \begin{array}{c} u : [0,T] \to \mathbb{R}^m \text{ measurable} \\ |\dot{x} = g(x,u),\ x(0) = x_0 \end{array} \right\} \tag{2.22}$$

Chapter 3

Quantification and Assignment of Metabolisms in a Microbial Community

This chapter deals with two challenges, the assignment and quantification (in terms of yields of substrate to biomass) of metabolism to the OTU in a MC. A framework is developed that is applied to data coming from the PhD thesis of Dumont *et al.* [28]. Dumont had already explored the problem of functional assignment using asymptotic observers [31], this chapter extends or completes his results in three ways:

- A discussion on how to identify the yields by the aid of the Gibbs energy dissipation method and asymptotic observers.
- A class of asymptotic observers are proposed, that includes the original observer proposed by Dumont.
- The problem of functional assignment is done using mixed integer linear programming instead of statistical likelihood.

3.1 Stoichiometric Matrix and Functional Groups

The functionality of a microbial group corresponds here to the catabolic reaction associated with this group. The term functional group makes reference then to all OTU that perform a given catabolic reaction. The challenge is the assignment of known functionalities to the identified OTU within a reactor. Even if two OTU perform the same catabolic reaction, their yields might differ, however, for an initial simplification, consider a unique yield for all of the OTU of a same functional group. We will come back to this hypothesis at the end of this section.

Biological *a priori* knowledge is required for identifying the known functionalities existing in a given environment. All of the possible thermodynamically feasible reactions of the set of environment molecules do not forcefully imply the existence of a microbial species performing such a reaction. Let n_F be the number of identified functional groups, and consider a system of the type (3.1)

$$\begin{aligned} \dot{x} &= \operatorname{diag}(\mu(x,s) - D_{n_F \times 1})x \\ \dot{s} &= (s_{in} - s)D + Y \operatorname{diag}(\mu(x,s))x \end{aligned} \tag{3.1}$$

where $x \in \mathbb{R}^{n_F}$ is the vector containing the biomass concentration of each functional group, and s represents the concentrations of a subset of the environment molecules that define the metabolism. If v_i defines the metabolism of functional group i, then Π is the projection such that $\Pi v_i = Y_{\bullet i}$. That is, the column of the stoichiometric matrix is composed of the metabolic coefficients.

A problem of parameter identification arises. How to find the best fitting coefficients of matrix Y. The matrix Y is of size $m \times n_F$. To achieve this, the theory of asymptotic observers may be employed [[27], so let us present them at first. In general biochemical reactions they allow to reconstruct the unmeasured state variables, from the measured state variables, independent of the reaction rate, which represents the main source of uncertainty in modelling mass-balance biological systems. In this case suppose the substrates s are measured, represented by $s^{obs}(t)$. The reaction invariant of system (2.4),

$z := -Yx + s$, had already been introduced in chapter 2. Let Y^{-1} be a left inverse of Y (if Y has full column rank one can take, for example, $Y^{-1} = (Y^\top Y)^{-1}Y^\top$), then $x = Y^{-1}(s - z)$. The asymptotic observer is constructed by using the measurement s^{obs}, instead of s, in the former equation.

$$\hat{x} := Y^{-1}(s^{obs} - z). \tag{3.2}$$

$\hat{x}_i$ is called the asymptotic observer of the functional group i. Suppose now that the total biomass X_T is also measured. The following optimization problem arises for finding the best stoichiometric matrix.

$$\begin{aligned} \min & \int_0^T \left| \sum_{i=1}^{n_F} \hat{x}_i(t) - X_T(t) \right| dt \\ st. \quad & Y_{\bullet i} = \Pi v_i \\ & \Delta G_{v_i} + \Delta G_{dis} = 0 \\ & \Delta G_{dis} \in [\Delta G_{dis}^{min}, \Delta G_{dis}^{max}] \end{aligned} \tag{3.3}$$

Note that the constraints are linear. Furthermore the thermodynamic formalism of the metabolism of a species can be omitted if one already has some bounds on the stoichiometric coefficients and model structure for a particular bioprocess. Also it can prove numerically unwise to calculate the pseudoinverse of a matrix in an optimization problem, so it is desirable to derive an algebraic expression of the left inverse.

Lets apply all of the former to the case of a nitrification process in a chemostat, the conditions of the experiments performed by Dumont are used [28]. Two reactors, A and B, were operated continuously for approximately 500 days with variable dilution rate and substrate input. They were inoculated with wastewater sludge and a synthetic mineral medium was used, as a consequence a nitrification process took place. Oxygen injection was maximized so there was no oxygen limitation and pH was regulated and maintained around 7. The activities and the standard Gibbs free energies of formation used are given in Table 3.1.

In the nitrification process two functional groups are known to drive the system, ammonium oxidizing bacteria (AOB) and nitrite oxidizing bacteria (NOB). In both functional groups it is suggested that the dissipated energy per C-mole should be around 3500 kJ/C-mol of biomass, because the reverse electron transport mechanism is used to drive the anabolic reaction [64].

Molecules	$G_f^0\,[kJ \cdot mol^{-1}]$	MW $[g \cdot mol^{-1}]$	Observation	$a_i\,[mol \cdot l^{-1}]$
NH_4^+	-79.4	18	1 $[g \cdot l^{-1}]$	1/18
O_2	0	32	saturated	0.2
NO_2^-	-32.2	46	0.5$[g \cdot l^{-1}]$	0.5/46
H^+	0	1	pH=7	10^-7
H_2O	-237.2	18	-	1
NO_3^-	-111.3	62	0.5 $[g \cdot l^{-1}]$	0.5/62
HCO_3^-	-586.9	61	saturated	0.004/29
$CH_{1.8}O_{0.5}N_{0.2}$	-67	24.6	-	1

Table 3.1: Molecules, standard Gibbs free energies of formation (from [64]), and activities for a nitrification process based on the experiments of Dumont [28].

The AOB and NOB catabolic reactions read as follows:

$$NH_4^+ + 1.5\,O_2 \longrightarrow NO_2^- + 2\,H^+ + H_2O \quad \Delta G_{cat\,AOB} = -264\,[kJ/mol\,NH_4^+] \tag{3.4}$$

$$NO_2^- + 0.5\,O_2 \longrightarrow NO_3^- \quad \Delta G_{cat\,NOB} = -75.5\,[kJ/mol\,NO_2^-] \tag{3.5}$$

A biomass synthesis reaction, representing anabolism, for both of the functional groups:

$$0.2\,NH_4^+ + HCO_3^- + 0.8\,H^+ \longrightarrow CH_{1.8}O_{0.5}N_{0.2} + 1.05\,O_2 + 0.4\,H_2O \tag{3.6}$$

$$\Delta G_{an} = 493\,[kJ/mol\,CH_{1.8}O_{0.5}N_{0.2}] \tag{3.7}$$

Note that oxygen is a product of this reaction, however the global stoichiometry (metabolism) will be oxygen consuming. Based on the analysis made by Von Stockar *et al.* on formalizing anabolic reactions [132], there is no problem in adding molecules from the catabolic reaction (oxygen in this case) to balance the synthesis reaction while using the Gibbs energy dissipation method, only the metabolic coupling coefficient changes, but the global stoichiometry will remain the same. With that in mind, one solves the

following equation:

$$\lambda_{AOB} = -\frac{\Delta G_{dis} + \Delta G_{an}}{\Delta G_{cat\,AOB}} = 15.1$$
$$\lambda_{NOB} = -\frac{\Delta G_{dis} + \Delta G_{an}}{\Delta G_{cat\,NOB}} = 52.8$$

which gives the following metabolisms

$$15.3\,NH_4^+ + 21.6\,O_2 + HCO_3^- \longrightarrow 15.1\,NO_2^- + 29.4\,H^+ + 15.5\,H_2O + CH_{1.8}O_{0.5}N_{0.2}$$
$$0.2\,NH_4^+ + 25.4\,O_2 + 52.8\,NO_2^- + 0.8\,H^+ + HCO_3^- \longrightarrow 0.4\,H_2O + 52.8\,NO_3^- + CH_{1.8}O_{0.5}N_{0.2}$$

In the nitrification process here studied, denote $x = (x_{AOB}, x_{NOB})$, s consists of ammonium (s_1), nitrite (s_2) and nitrate (s_3). From this analysis it can be seen that the stoichiometric coefficients $y_{s_1/x_{AOB}}$ and $y_{s_2/x_{AOB}}$ are almost the same, $\left|\frac{y_{s_1/x_{AOB}}}{y_{s_2/x_{AOB}}}\right| = 1.01$, and approximated by λ_{AOB}, with their respective signs. Conversely, the proportion between $y_{s_1/x_{NOB}}$ and $y_{s_2/x_{NOB}}$, $\left|\frac{y_{s_1/x_{NOB}}}{y_{s_2/x_{NOB}}}\right| = 0.004$, is very small. Evidently, the question remains to what extent does the value ΔG_{dis} affect the results. In table 3.2 the computations of the ratios are shown for other ΔG_{dis} values obtained from fitting ΔG_{dis} to literature values (see table 3.3)

$\Delta G_{dis}[kJ/mol\,CH_{1.8}O_{0.5}N_{0.2}]$	$\Delta G_{dis} = 539$	$\Delta G_{dis} = 2668$	$\Delta G_{dis} = 3500$
λ_{AOB}	3.9	12	15.1
λ_{NOB}	13.7	41.9	52.8
$\left\|\frac{y_{s_1/x_{AOB}}}{y_{s_2/x_{AOB}}}\right\|$	1.05	1.02	1.01
$\left\|\frac{y_{s_1/x_{NOB}}}{y_{s_2/x_{NOB}}}\right\|$	0.01	0.005	0.004

Table 3.2: Sensitivity of the ratio of yields for other ΔG_{dis} values.

The former implies that the usual hypothesis [93] that the yield of ammonium consumption and nitrite production are the same for the AOB community, and the hypothesis that the NOB community does not take up ammonium, are both well justified under this thermodynamic analysis.

For the above reason the stoichiometric matrix is approximated by:

$$Y = \begin{bmatrix} -\lambda_{AOB} & 0 \\ \lambda_{AOB} & -\lambda_{NOB} \\ 0 & \lambda_{NOB} \end{bmatrix} \tag{3.8}$$

In table 3.3 a comparison to the nitrification study performed by Weissmann [139] is shown, the yields are calculated as $\frac{1}{\lambda_{AOB}} \frac{mw\mathrm{CH_{1.8}O_{0.5}N_{0.2}}}{mw\mathrm{NH_4^+ -N}}$ and $\frac{1}{\lambda_{NOB}} \frac{mw\mathrm{CH_{1.8}O_{0.5}N_{0.2}}}{mw\mathrm{NO_2^- - N}}$ where mw stands for molecular weight, in particular, $mw\mathrm{NH_4^+ - N}$ and $mw\mathrm{NO_2^- - N}$ stands for the molecular weight corresponding to nitrogen of the compound, which in both cases is 14. It is interesting that a single ΔG_{dis} value fitted both reported yields, shown to be $\Delta G_{dis} = 2668[kJ/mol\,\mathrm{CH_{1.8}O_{0.5}N_{0.2}}]$, significantly lower that the value proposed in [64] for organisms performing reverse electron transport. For the AOB community even yields as high as $0.45[g\,odm/g\mathrm{NH_4^+ - N}]$ have been reported [13], implying a $\Delta G_{dis} = 474[kJ/mol\,\mathrm{CH_{1.8}O_{0.5}N_{0.2}}]$.

Yield values $y_{x/s}$	AOB $[g\,odm/g\mathrm{NH_4^+ - N}]$	NOB $[g\,odm/g\mathrm{NO_2^- - N}]$
$\Delta G_{dis} = 539[kJ/mol\,\mathrm{CH_{1.8}O_{0.5}N_{0.2}}]$	0.45 (ref: [13])	0.13
$\Delta G_{dis} = 2668[kJ/mol\,\mathrm{CH_{1.8}O_{0.5}N_{0.2}}]$	0.147 (ref: [139])	0.042 (ref: [139])
$\Delta G_{dis} = 3500[kJ/mol\,\mathrm{CH_{1.8}O_{0.5}N_{0.2}}]$	0.12	0.03

Table 3.3: The relationship between ΔG_{dis} and yield for nitrifiers. odm stands for organic dry matter.

Given that data of s_1, s_2 and s_2 is given in grams of nitrogen, the stoichiometric matrix will be written in terms of $y_{x/s}$ in grams of odm (organic dry matter) per gram of nitrogen, as in Table 3.3. Going back to the optimization problem (3.8), denote y_{AOB} and y_{NOB} the yields of AOB and NOB, respectively and rewrite (3.8) as:

$$Y = \begin{bmatrix} -1/y_{AOB} & 0 \\ 1/y_{AOB} & -1/y_{NOB} \\ 0 & 1/y_{NOB} \end{bmatrix} \tag{3.9}$$

$$\Rightarrow Y^{-1} = (Y^\top Y)^{-1} Y^\top = \begin{bmatrix} -\frac{2y_{AOB}}{3} & \frac{y_{AOB}}{3} & \frac{y_{AOB}}{3} \\ -\frac{y_{NOB}}{3} & -\frac{y_{NOB}}{3} & \frac{2y_{NOB}}{3} \end{bmatrix} \tag{3.10}$$

which in turn implies that:

$$\hat{x} = \begin{bmatrix} \frac{y_{AOB}}{3}\left(-2(s_1^{obs} - z_1) + (s_2^{obs} - z_2) + s_3^{obs} - z_3\right) \\ \frac{y_{NOB}}{3}\left(-(s_1^{obs} - z_1) - (s_2^{obs} - z_2) + 2(s_3^{obs} - z_3)\right) \end{bmatrix} \tag{3.11}$$

For this case study the optimization problem (3.3) can be written as:

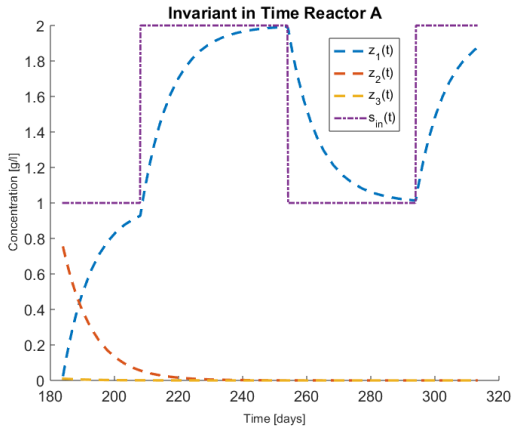
$$\begin{aligned} \min & \int_0^T |\hat{x}_{AOB}(t) + \hat{x}_{NOB}(t) - X_T(t)|\, dt \\ s.t. \quad & y_{AOB} \in [0.12,\, 0.45] \\ & y_{NOB} \in [0.03,\, 0.13] \end{aligned} \tag{3.12}$$

Before solving (3.12) to the data from experiments, a final word on the fact that $\hat{x}$ might be negative, which is physically senseless. It would be better to use $\hat{x}_i(t) = \max\{(Y^{-1}(s^{obs}(t) - z(t)))_i, 0\}$ as the observer, and it is indeed the case in the following . A change in temperature was done in day 183, and only then complete nitrification started, therefore only data from that point onwards was analysed.

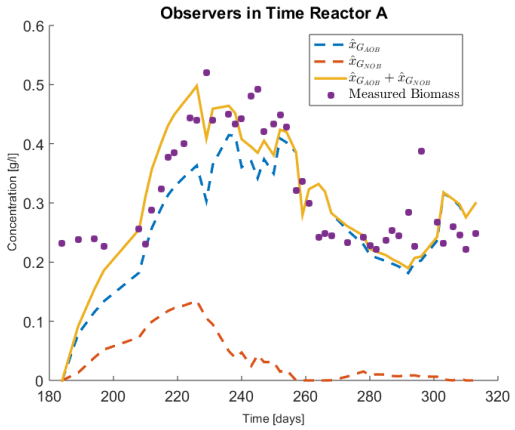
Samples of the total dry biomass, concentrations of ammonium (NH_4^+), nitrite (NO_2^-), and nitrate (NO_3^-), were taken at specific times. Only ammonium (s_1) was entering the reactor, abusing of notation note $s_{in} = \begin{bmatrix} s_{in} & 0 & 0 \end{bmatrix}^\top$

From Reactor A the fitted yields were: $y_{AOB} = 0.213$, $\Delta G_{dis,AOB} = 1531$, $y_{NOB} = 0.085$ $\Delta G_{dis,NOB} = 964$. While in Reactor B fitted yields were: $y_{AOB} = 0.2622$, $\Delta G_{dis,AOB} =$

1127, $y_{NOB} = 0.03$ $\Delta G_{dis,NOB} = 3602$. In the case of the AOB community, the dissipated energies in each reactor are somewhat close. This is not the case for the NOB community, however, note that in Reactor B, y_{NOB} is the lower bound of the optimization problem, which suggests that the NOB community probably did not developed as well as in Reactor A. Figures 3.1 and 3.2 show the results of the solution of problem (3.3) for reactors A and B, respectively. Note how the invariants in figures 3.1a and 3.2a converge quickly to the same values. The observers from figures 3.1b and 3.2b indeed show that the NOB community in reactor B was very low compared to reactor A.

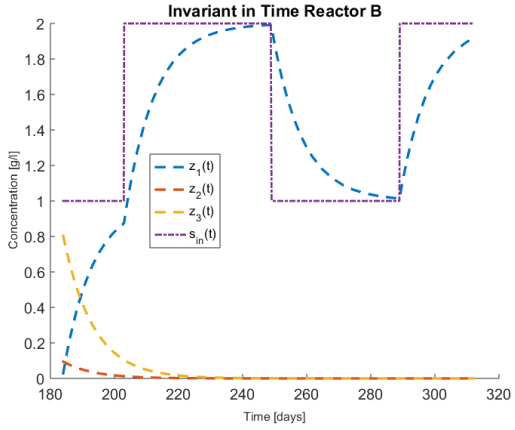


(a) Invariant for reactor A.

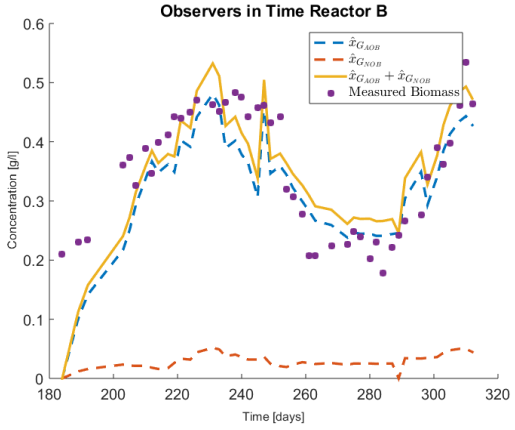


(b) Observer for reactor A.

Figure 3.1: Reactor A results of solving problem (3.12)



(a) Invariant for reactor B.



(b) Observer for reactor B.

Figure 3.2: Reactor B results of solving problem (3.12)

3.2 Assigning functions

Now that the stoichiometry of each functional group has been found, the task at hand is to assign one of the possible n_F stoichiometries to each OTU. In other words one wishes to assign to each of the n OTU a functional group. Suppose $p \in \mathbb{N}$ genetic sequencing measurements were made at timesteps $t_1, \dots t_p$, and denote $\hat{x}(t)$ the observer of the n_F functional groups. Some variables and parameters to pose the problem of assignation as an optimization problem are stated:

- $\alpha_{ik} \in \{0, 1\}$, $i \in [n]$, $k \in [n_F]$. $\alpha_{ik} = 1$ if OTU i is assigned to functional group k, 0 otherwise.
- $\epsilon_{kj} \in \mathbb{R}_+ \, k \in n_F$, $j \in [p]$, error associated to the classification in functional group k of OTU at measurement j.
- $W \in M_{n\times p}(\mathbb{R}_+)$: W_{ij} represents the concentration of OTU i at measurement j.
- $\hat{X} \in M_{n_F\times p}(\mathbb{R}_+)$. $\hat{X}_{kj} = \hat{x}_k(t_j)$, represents the observer of functiona group k evaluated at the time of measurement j.

With this one can define the following constraints:

$$-\epsilon_{kj} \leq W_{\bullet j}^{\top}\alpha_{\bullet k} - \hat{X}_{kj} \leq \epsilon_{kj} \quad \forall k \in [n_F],\, j \in [p] \tag{3.13}$$

$$\sum_{k=1}^{n_f} \alpha_{ik} \leq 1 \quad \forall i \in [n]. \tag{3.14}$$

Constraint (3.13) describes the made done by the assignment of α_{ik} at measurement j. Constraint (3.14) implies that each OTU can be classified to a maximum of one functional group (leaving the possibility to not be classified at all). The problem to be

solved becomes:

$$\begin{aligned} &\min \sum_{k\in[n_F],\, j\in[m]} \epsilon_{kj} \\ s.t.\quad &(3.13), (3.14). \\ &\alpha_{ik} \in \{0,1\},\, i \in [n],\, k \in [n_F] \\ &\epsilon_{kj} \geq 0\, k \in n_F,\, j \in [p] \end{aligned} \tag{3.15}$$

Problem (3.15) is a mixed integer linear problem, however one might like to minimize the quadratic error, in that case the problem becomes a mixed integer quadratic problem. The procedure was applied to the data of Dumont [28], and appeared in the 2019 European Control Conference proceedings [125]. At the moment of its writing, the method was applied with a different left inverse (eq (3.16)), and therefore a different asymptotic observer.

$$Y^{-1} = \begin{bmatrix} -y_{AOB} & 0 & 0 \\ -y_{NOB} & -y_{NOB} & 0 \end{bmatrix} \tag{3.16}$$

The multiplicity of left inverses opens up the question of the best choice. We argue that the best choice comes from equation (3.10), for the simple reason that the resulting observer uses the information of s_3^{obs}, whereas with (3.16) it does not. Moreover, the definition of the observer in the article was done without multiplying the yield, so it may cause confusion to the reader. This was done because the optimization problem explained in the article tries at the same time to find the optimal yields and classify each species, which eliminates the original hypothesis of all of the OTU having the same metabolism in each functional group. This is done by the use of big-M constraints and its technicality would have obscured the presentation of Gibbs energy dissipation method and asymptotic observers for fitting the yield. Without any other warning, the rest of this chapter consists of the aforementioned article [125], which describes in detail the implementation of a mixed integer program.

Asymptotic Observers and Integer Programming for Functional Classification of a Microbial Community in a Chemostat.

Pablo Ugalde-Salas[1], Jérôme Harmand[2] and Elie Desmond-Le Quéméner [3]

[1]LBE, Univ Montpellier, INRA, 102 avenue des Etangs, 11100, Narbonne, France. Ph.D Student. pablo.ugalde-salas@inra.fr

[2] LBE, INRA, Univ Montpellier, 11100, Narbonne, France. jerome.harmand@inra.fr

[3] LBE, INRA, Univ Montpellier, 11100, Narbonne, France. elie.lequemener@inra.fr

Abstract

From genetic sequencing, dry biomass, and metabolites measurements, the assignment of functions to the species present in chemostat experiments was solved by merging chemostat modelling and quadratic mixed integer programming. The method was tested on a nitrification bioprocess where two functions are known to drive the system. Sensitivity of the method, its advantages, and limitations are discussed.

3.3 Introduction

The objective of this manuscript is to present an optimization method based on mixed integer programming and invariants of chemostat dynamical models for the functional classification of microorganisms in bioprocess. One of the first attempts to implement an optimization procedure can be found in Dumont et al. [31].

Measurements based on genetic material have become standard practice in ecological engineered systems (e.g. wastewater treatments plants) [88]. However using these measurements for prediction and control is still at a very early stage. To face such challenges linking functionality to the different members of the community is an important initial task to be tackled [138], [133]. Comparisons of such measurements to current databases often fall short, since the coverage of existing species is very limited compared to reality.

The motivation is therefore to develop tools for incorporating these new measurements in engineering models.

3.4 Materials and Methods

Variables used throughout the article are summarized in Table 3.4. The following conventions are used: Let $n \in \mathbb{N}$ then $[n] := \{1, \ldots, n\}$, $\mathbb{R}_+ := \{t \geq 0 | t \in \mathbb{R}\}$.

3.4.1 Experimental Conditions

The optimization method is applied to data coming from a chemostat experiment. A chemostat is an experimental device used to study microbial growth where a reactor continuously receives a solution containing nutrients for proper microbial development [76]. The reactor has the same inlet and outlet flow, thus a constant volume is maintained inside the reactor.

Two reactors, A and B, were operated continuously for approximately 500 days with variable dilution rate and substrate input. They were inoculated with wastewater sludge and a dilution composed of ammonium and a synthetic mineral medium was used; as a consequence a nitrification process took place. Oxygen injection was maximized so there was no oxygen limitation, and pH was regulated and maintained around 7.

Samples of the total dry biomass, concentrations of ammonium (NH_4^+), nitrite (NO_2^-), and nitrate (NO_3^-), were taken at specific times. Microbial diversity was analyzed using single strand conformation polymorphism (SSCP) at specific times: 44 different Operational Taxonomic Unit (OTU) were identified in total, with most of them being present in both. More details on the experimental conditions can be found in the author's original article [30].

3.4.2 Stoichiometry and Functional Groups

For this article a cascade (bio)reaction process is considered. Suppose n different OTU are present in the chemostat. The cascade reaction refers to the situation where a group

Table 3.4: Notation used throughout the article.

Symbol	Description
$s_1(t)$	Concentration ([g/l]) of ammonium in time.
$s_2(t)$	Concentration ([g/l]) of nitrite in time.
$s_3(t)$	Concentration ([g/l]) of nitrate in time.
$s(t)$	Vector containing $s_1(t), s_2(t), s_3(t)$.
$x_i(t)$	Concentration ([g/l]) of OTU i in time.
$x(t)$	Vector containing $x_1(t), \ldots, x_n(t)$.
$D(t)$	Dilution Rate ([1/day]) in time.
$s_{in}(t)$	Concentration ([g/l]) of input ammonium in time.
$s_1^{obs}(t)$	Measured concentration ([g/l]) of ammonium in time.
$s_2^{obs}(t)$	Measured concentration ([g/l]) of nitrite in time.
$x_i^{obs}(t)$	Measured concentration ([g/l]) of OTU i in time.
$z_1(t)$	Reaction invariant of the dynamical system treated.
$z_2(t)$	Reaction invariant of the dynamical system treated.
$\hat{x}_{G_1}(t)$	Observer of the sum of the biomass of OTU in G_1.
$\hat{x}_{G_2}(t)$	Observer of the sum of the biomass of OTU in G_2.

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 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 it represents the direction of the bioprocess and the proportions (stoichiometric coefficients) of different consumed and formed compounds of interest.

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

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

The terms y_i are known as yields, it represents the number of moles of biomass produced per mole of substrate consumed. However in this work the unit grams of dry biomass per gram of substrate is used for yields. The term M_x represents a molecule of biomass and several expressions can be found in the literature (e.g. $CH_{1.613}O_{0.557}N_{0.158}$ [10]). Furthermore, for each $i \in [n]$, OTU i is characterized by its process rate (also known as growth function or kinetics) $\mu_i(s, x)$, where the first variable s represents a

vector containing the concentration of (s_1, s_2, s_3). The second variable x represents the vector containing the concentration of all OTU.

From an evolutionary perspective there is reason to think that each OTU may have its own yield and growth function.

From biological knowledge it is sometimes known that two different OTU can not belong to the same functional group, that is $G_1 \cap G_2 = \emptyset$. This is the case of nitrification process [139]. The group G_1 is known as ammonia oxidizer Bacteria (AOB), which turn ammonia (s_1) into nitrite (s_2), the group G_2 is known as nitrite oxidizer Bacteria (NOB) which turns nitrite into nitrate (s_3).

3.4.3 Mass-Balanced model

The cascade reaction is modelled as a substrate-coupled dynamical model.

The chemostat has a dilution rate of D and an input of ammonium concentration s_{in}, both of which are operating parameters that can change in time, that is $D = D(t)$ and $s_{in} = s_{in}(t)$, however for alleviating notation the time dependence is dropped. For more details in chemostat modelling the reader may refer to [44].

Denoting each OTU concentration by x_i, ammonium by s_1, nitrite by s_2, nitrate by s_3, and considering reactions (R1), and (R2) the mass balanced model can be formally expressed:

$$\dot{x_i} = (\mu_i(s,x) - D)\, x_i \quad \forall i \in G_1 \tag{3.17}$$

$$\dot{x_i} = (\mu_i(s,x) - D)\, x_i \quad \forall i \in G_2 \tag{3.18}$$

$$\dot{s_1} = (s_{in} - s_1)D - \sum_{i \in G_1} \frac{1}{y_i} \mu_i(s,x) x_i \tag{3.19}$$

$$\dot{s_2} = -\, s_2 D + \sum_{i \in G_1} \frac{1}{y_i} \mu_i(s,x) x_i - \sum_{i \in G_2} \frac{1}{y_i} \mu_i(s,x) x_i \tag{3.20}$$

$$\dot{s_3} = -\, s_3 D + \sum_{i \in G_1} \frac{1}{y_i} \mu_i(s,x) x_i \tag{3.21}$$

At this point we can formally state the problem. Let us assume that we know the

measurements of the abundance of n OTU $x_i^{obs}(t)$ $i \in [n]$, $s_1^{obs}(t)$, $s_2^{obs}(t)$, and $s_3^{obs}(t)$ in a chemostat where s_{in} and $D(t)$ are known. Find two disjoint subsets G_1, $G_2 \subset [n]$, such that the norm of the difference of the observations and the solution of the system (x, s) given by equations (3.17), (3.18), (3.19), (3.20), and (3.21) is minimized, that is

$$\begin{aligned} \min \quad & \sum_{i \in G_1} \|x_i - x_i^{obs}\|_2 + \sum_{i \in G_2} \|x_i - x_i^{obs}\|_2 \\ & + \|s_1 - s_1^{obs}\|_2 + \|s_2 - s_2^{obs}\|_2 + \|s_3 - s_3^{obs}\|_2 \\ s.t \quad & G_1, G_2 \subset [n] \\ & G_1 \cap G_2 = \emptyset \\ & (x, s) \text{ solution of } (3.17), (3.18), (3.19), (3.20), (3.21) \end{aligned} \tag{3.22}$$

The problem relies on (i) not knowing the growth rates and (ii) testing all the possible combinations and simulating is computationally expensive.

3.4.4 Asymptotic Observer

If we knew a priori both sets G_1 and G_2 and the growth rates $\mu_i(x, s)$, then we could directly compare the measurements and the dynamical system given by equations (3.17), (3.18), (3.19), (3.20), and (3.21). However we do not know neither the sets nor the kinetics.

To solve this problem a classic invariant of such types of model are derived. These are called also reaction invariants [135]. They allow the construction of asymptotic observers [27], which are observers in the sense that, whatever the initial conditions are, they converge to a manifold which only depends on the yields and some state variables, thus circumventing the knowledge of process rate; this has been done for general biochemical reactors (equations (49) to (56)) in [27]. The price to pay for such observers is that the convergence rate to the manifold depends on the operating conditions.

Define $z_1 := \sum_{i \in G_1} \frac{1}{y_i} x_i + s_1$ and compute $\dot{z_1}$ using equations (3.17) and (3.19):

$$\dot{z_1} = \sum_{i \in G_1} \frac{1}{y_i} \dot{x_i} + \dot{s_1} \tag{3.23}$$

$$= \sum_{i \in G_1} \frac{1}{y_i} \left(\mu_i(s,x) - D\right) x_i + (s_{in} - s_1)D - \sum_{i \in G_1} \frac{1}{y_i} \mu_i(s,x) x_i \tag{3.24}$$

$$= -D \left(\sum_{i \in G_1} \frac{1}{y_i} x_i + s_1 - s_{in} \right) \tag{3.25}$$

$$= -D(t)(z_1 - s_{in}(t)) \tag{3.26}$$

Analogously another invariant can be derived that allows linking the biomass of G_2 and substrates. Define $z_2 := \sum_{i \in G_2} \frac{1}{y_i} x_i + s_1 + s_2$ and compute $\dot{z_2}$ using equations (3.18), (3.19), and(3.20):

$$\dot{z_2} = \sum_{i \in G_2} \frac{1}{y_i} \dot{x_i} + \dot{s_1} + \dot{s_2} \tag{3.27}$$

$$= \sum_{i \in G_2} \frac{1}{y_i} \left(\mu_i(s_2,x) - D\right) x_i + (s_{in} - s_1)D \tag{3.28}$$

$$- \sum_{i \in G_1} \frac{1}{y_i} \mu_i(s_1,x) x_i - s_2 D + \sum_{i \in G_1} \frac{1}{y_i} \mu_i(s_1,x) x_i$$

$$- \sum_{i \in G_2} \frac{1}{y_i} \mu_i(s_2,x) x_i \tag{3.29}$$

$$= -D \left(\sum_{i \in G_2} \frac{1}{y_i} x_i + s_1 + s_2 - s_{in} \right) \tag{3.30}$$

$$= -D(t)(z_2 - s_{in}(t)) \tag{3.31}$$

Note that z_1 and z_2 satisfy the same dynamics, which does not depend in μ_i. Invariants z_1 and z_2 can be shown to be stable [27].

A simulation of the differential equation (3.26) using the dilution rate and input ammonium of the experiment can be seen in Fig. 3.3, it suggests that the solutions approach rapidly to a similar curve independently of the initial point. Since at the beginning of the

experiment minimal biomass is present one assumes $z_1(0) = \sum_{i \in G_1} \frac{1}{y_i} x_i(0) + s_1(0) \approx s_1(t_1)$. Analogously $z_2(0) = \sum_{i \in G_2} \frac{1}{y_i} x_i(0) + s_1(0) + s_2(0) \approx s_1(t_1) + s_2(t_1)$.

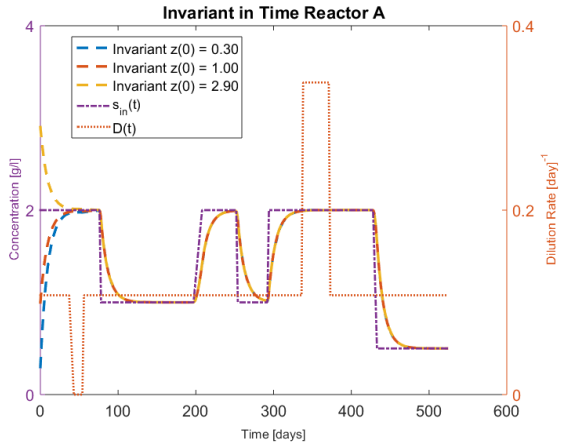


Figure 3.3: Invariant evolution for Reactor A with three different initial points.

The observers are defined as $\hat{x}_{G_1} := z_1 - s_1$ and $\hat{x}_{G_2} := z_2 - s_1 - s_2$, each one of them converges to $\sum_{i \in G_1} \frac{1}{y_i} x_i$ and $\sum_{i \in G_2} \frac{1}{y_i} x_i$, respectively. A simulation of the observers trajectory can be seen from figure 3.4 where s_1 and s_2 were taken as s_1^{obs} and s_2^{obs}, respectively.

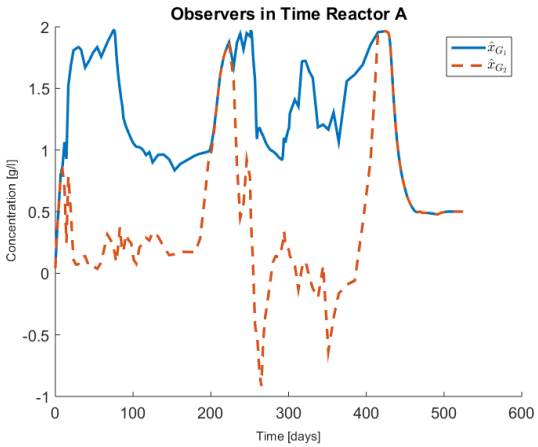


Figure 3.4: Observer evolution for Reactor A using the measurements of s_1 and s_2.

3.4.5 Mixed Integer Program

A mixed integer program is presented in order to classify each OTU as AOB, NOB, or not determined by using observers $\hat{x}_{G_1}$ and $\hat{x}_{G_2}$ as inputs. The objective function is to minimize the error defined as the difference of the observers and $\sum_{i \in G_j} \frac{1}{y_i} x_i$ for $j \in \{1, 2\}$.

Let n be the number of different OTU identified, m the number of measurements, and t_j the time stamp of measurement $j \in [m]$. The variables to decide the classification are:

Variables:

- $a \in \{0,1\}^n$: $a_i = 1$ if OTU i is classified as AOB, 0 otherwise.
- $b \in \{0,1\}^n$: $b_i = 1$ if OTU i is classified as NOB, 0 otherwise.
- $k^A \in \mathbb{R}_+^n$: $k_i^A > 0$ if OTU i is classified as AOB, 0 otherwise. If $k_i^A > 0$ then $k_i^A = y_i^{-1}$.
- $k^B \in \mathbb{R}_+^n$: $k_i^B > 0$ if OTU i is classified as NOB, 0 otherwise. If $k_i^B > 0$ then $k_i^B = y_i^{-1}$.
- $\epsilon \in \mathbb{R}_+^m$: ϵ_j error associated to classification of AOB in measurement j.
- $\eta \in \mathbb{R}_+^m$: η_j error associated to classification of NOB in measurement j.

Parameters of the optimization problem are divided in data as presented in table 3.4, and meta-parameters: $y_{ref}^A, y_{ref}^B, \delta, m_a, M_a, m_b, M_b$, meaning that these parameters come from prior knowledge to the experiment. All together they give bounds for the variables k^B and k^A.

Parameters of the Problem

- $W \in M_{n\times m}(\mathbb{R}_+)$: $W_{ij} = x_i(t_j)$. Column j contain the concentration of each OTU at timestep t_j.
- $s_1(t_j) \in \mathbb{R}^+ \quad \forall j \in [m]$ as defined in table 3.4.
- $s_2(t_j) \in \mathbb{R}^+ \quad \forall j \in [m]$:as defined in table 3.4.
- $\hat{x}_{G_1}(t_j) \in \mathbb{R}^+ \quad \forall j \in [m]$: Observer evaluated at timestamps.
- $\hat{x}_{G_2}(t_j) \in \mathbb{R}^+ \quad \forall j \in [m]$: Observer evaluated at timestamps.
- $y_{ref}^A, y_{ref}^B \in \mathbb{R}_+$: Literature reference value for yields of AOB and NOB, respectively.
- $\delta \in (0,1)$: fraction allowed to deviate from the reference yields.

- $m_A, M_A \in \mathbb{R}_+$: lower and upper bounds for k_i^A, respectively. $m_A := \dfrac{1}{(1+\delta)y_{ref}^A}$, $M_A := \dfrac{1}{(1-\delta)y_{ref}^A}$.
- $m_B, M_B \in \mathbb{R}_+$: lower and upper bounds for k_i^B, respectively. $m_B := \dfrac{1}{(1+\delta)y_{ref}^B}$, $M_B := \dfrac{1}{(1-\delta)y_{ref}^B}$.

In the numerical experiences, the reference yield for AOB is $y_{ref}^A = 0.147$ [gr odm/grNH_4^+], and for the NOB is $y_{ref}^B = 0.042$ [gr odm/grNO_2^-] where odm stands for organic dry matter [139].

Based on the former discussion the following constraints are imposed. The notation $W_{\bullet j}$ is used to represent the j-th column of matrix W.

Constraints:

- AOB mass error classification: The difference between $\hat{x}_{G_1}$ and the assigned mass at each measurement is bounded by ϵ .

$$-\epsilon_j \leq W_{\bullet j}^\top k^A - \hat{x}_{G_1}(t_j) \leq \epsilon_j \quad \forall j \in [m]. \tag{3.32}$$

- NOB mass error classification: The difference between $\hat{x}_{G_2}$ and the assigned mass at each measurement is bounded by η.

$$-\eta_j \leq W_{\bullet j}^\top k^B - \hat{x}_{G_2}(t_j) \leq \eta_j \quad \forall j \in [m] \tag{3.33}$$

- Each species can be classified in only one functional group:

$$a_i + b_i \leq 1 \quad \forall i \in [n] \tag{3.34}$$

- Linking constraint (Big-M Constraints): Activation of k_i^A or k_i^B when a_i or b_i is

active, respectively:

$$m_A a_i \leq k_i^A \leq M_A a_i \quad \forall i \in [n] \tag{3.35}$$

$$m_B b_i \leq k_i^B \leq M_B b_i \quad \forall i \in [n] \tag{3.36}$$

- Objective Function: The minimum of the norms of vector ϵ and η.

$$\|\epsilon\|^2 + \|\eta\|^2 = \sum_{j=1}^{m} (\eta_j^2 + \epsilon_j^2) \tag{3.37}$$

The problem to be solved is:

$$\begin{aligned} \min \quad & \sum_{j=1}^{m} (\eta_j^2 + \epsilon_j^2) \\ s.t. \quad & (3.32), (3.33), (3.34), (3.35), (3.36) \\ & \epsilon, \eta, \in \mathbb{R}_+^m \\ & a, b \in \{0,1\}^n \\ & k^A, k^B \in \mathbb{R}_+^n \end{aligned} \tag{$MIQP$}$$

The problem ($MIQP$) falls in the category of mixed integer quadratic programming, and it can be properly described in terms of the number of OTU (n) and the number of functional groups (r), and the number of observations (m). Note that since if one considers one observer per functional group, the number of variables is calculated as $2 \times n \times r + 2 \times m \times r$. The number of restrictions is calculated as $2 \times m \times r + 3 \times n \times r$. This shows that the number of restrictions and variables grow linearly with the number of OTU (for fixed r).

3.5 Results and discussion

In all cases here presented the computing time was less than a second with zero optimality gap. The solver GUROBI was used within a Matlab interface. The computer was equipped with 8gb of RAM memory and Intel core i3-7100U CPU 2,40 GHz .

By varying δ the classification changes for certain OTU. The method always classifies 35 OTU in the same guild, which represent 87 % of the total biomass found. While 9 of them changed of class by varying the allowed bounds. Not all of the present OTU were participating in the nitrification process. This can be explained by the presence of heterotrophs which are feeding on decayed cell material and or predators.

The obtained yields were plotted for the case $\delta = 0.3$ in figure 3.5. One can see that the classification usually assigns the minimum or maximum yield.

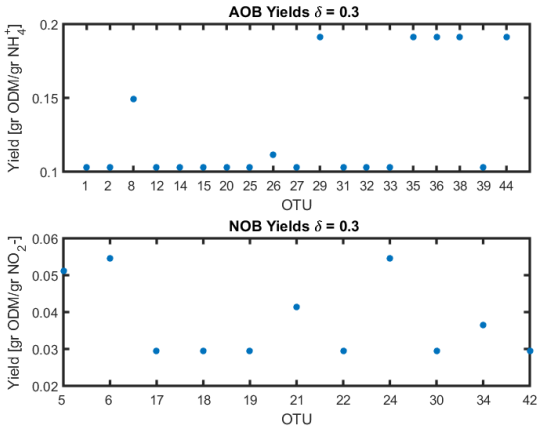


Figure 3.5: Obtained Yields for Reactor A

The study shows that the bounds imposed on the inverse of the yields (k^A, k^B) deserve some attention, since they can change the classification of some OTU. A more precise quantification of the possible variability within a guild escapes the authors' knowledge. In the literature a measurement error of 30 % is usually found, therefore $\delta = 0.3$ was taken in order to compare both reactors.

The results for the comparison of both reactors can be seen in Table 3.5, where one notes that 3 OTU were assigned to a different functional group (highlighted in dark grey); it is very unlikely from a biological point of view that they had changed their function in each chemostat, implying that the classification can be tricked in certain cases. 15 OTU were assigned to a functional group in one case, but to none in the other (highlighted in light grey); which can be explained by their low abundance and presence time in the reactors where they were not assigned to any guild. Finally 26 OTU were assigned to the same guild (highlighted in white), representing a mean abundance of 74 % and 76 % of

Table 3.5: Comparison of reactors and results of previous work.

A 1 under an AOB and NOB columns represent classification in the functional group, 0 means not classified in the functional group. Rows highlighted in light grey show OTU that changed from either AOB or NOB in one reactor to none in the other one. Rows highlighted in dark grey show OTU that changed from either AOB or NOB in one reactor to NOB or AOB, respectively, in the other one. Highlighted in light blue the cases where the classification changed from either AOB or NOB in one work to none in the other. Higlighted in blue the cases where classification changed from either AOB or NOB in one work to NOB or AOB in the other work, respectively. Time present in bold means it was close to clones identified in a database. ND stands for non demonstrated nitrifying capacity.

	Reactor A							Reactor B						
OTU	Time Present (%)	Relative Species Abundance (%)		Functional Assignment				TimePresent (%)	Relative Species Abundance (%)		Functional Assignment			
				This Work		Previous Work					This Work		Previous Work	
		Mean	Max	AOB	NOB	AOB	NOB		Mean	Max	AOB	NOB	AOB	NOB
1	20	0	2	1	0	1	0	9	0	3	1	0	1	0
2	31	1	6	1	0	0	1	20	0	2	1	0	0	0
3	15	0	3	0	0	0	0	3	0	3	0	0	0	0
4	0	0	0	0	0	0	0	16	0	5	1	0	1	0
5	57	1	11	0	1	0	1	**81 NOB**	4	27	0	1	0	1
6	32	1	15	0	1	0	1	25	1	7	0	0	0	1
7	14	0	4	0	0	1	0	24	1	5	1	0	0	0
8	71	3	18	1	0	1	0	17	0	11	0	0	1	0
9	**24 NOB**	1	6	0	0	0	0	36	1	8	0	0	0	1
10	19	1	9	0	0	1	0	16	0	5	1	0	0	1
11	**12 ND**	0	13	0	0	1	0	10	0	7	0	0	0	0
12	19	1	8	1	0	1	0	6	0	9	1	0	1	0
13	31	1	12	0	0	0	1	88	2	8	0	1	0	1
14	**36 ND**	1	9	1	0	0	0	20	1	9	1	0	0	1
15	38	1	12	1	0	0	1	41	1	8	1	0	0	1
16	44	1	4	0	0	0	0	83	2	6	0	0	0	1
17	36	1	4	0	1	0	1	30	1	8	0	1	0	1
18	35	1	10	0	1	0	1	0	0	0	0	0	0	1
19	**11 ND**	0	4	0	1	0	1	20	0	3	0	0	0	0
20	19	0	3	1	0	0	1	31	1	5	1	0	1	0
21	41	1	6	0	1	0	1	18	0	4	0	1	0	0
22	15	0	2	0	1	0	1	31	1	7	0	1	0	1
23	**73 ND**	2	8	0	0	1	0	76	2	8	0	0	1	0
24	23	1	4	0	1	1	0	23	0	5	1	0	0	1
25	**64 ND**	2	11	1	0	1	0	89	3	18	1	0	1	0
26	34	1	5	1	0	0	1	17	0	4	1	0	1	0
27	18	1	12	1	0	0	1	40	3	23	1	0	0	0
28	41	1	7	0	0	1	0	28	1	13	0	0	1	0
29	**45 ND**	2	6	1	0	1	0	15	0	5	0	0	1	0
30	16	1	11	0	1	0	1	**76 ND**	2	10	1	0	0	1
31	23	1	5	1	0	0	1	20	1	13	1	0	0	1
32	60	2	18	1	0	0	1	58	1	6	0	1	0	1
33	50	2	36	1	0	1	0	95	4	49	1	0	1	0
34	17	1	16	0	1	0	0	0	0	0	0	0	0	0
35	**100 ND**	10	41	1	0	1	0	**100 ND**	12	38	1	0	1	0
36	50	3	26	1	0	1	0	96	5	28	0	0	1	0
37	**100 ND**	6	37	0	0	1	0	39	2	10	0	0	1	0
38	**100 AOB**	38	77	1	0	1	0	**100 AOB**	35	65	1	0	1	0
39	30	1	13	1	0	1	0	42	2	20	0	0	0	0
40	0	0	0	0	0	0	0	57	2	11	0	1	0	1
41	25	1	5	0	0	0	0	11	1	28	0	0	1	0
42	17	1	13	0	1	0	0	9	0	12	0	0	1	0
43	93	5	30	0	0	1	0	93	5	29	1	0	0	0
44	11	1	21	1	0	1	0	0	0	0	0	0	0	0

the total biomass of reactors A and B, respectively. The mean abundance corresponding to AOB in reactors A and B was of 71 and 69 % and of the total biomass, respectively. while the abundance of the NOB community in reactors A and B was 9 and 11 % of the total biomass, respectively.

The method used in the work of Dumont *et al.* [31], consisted in generating the total AOB and NOB biomass from an observer. Then they randomly picked 10 OTU and tested all the possible assignments to find the one that best fitted the generated AOB and NOB biomasses. They repeated this process 10000 times and finally assigned probabilities to each OTU to be classified as either AOB, NOB or not determined. It took three days of computing time.

Comparison from the classification obtained from the previous work can be seen from table 3.5 as well. In Reactor A 19 OTU were classified differently, representing 27 % of the total biomass; 7 out of 19 OTU (highlighted in blue) changed functional group while the others (highlighted in light blue) had no functional group assigned in one of the works. In Reactor B, 23 OTU were classified differently, representing 32 % of the total biomass; 5 out of 23 OTU (highlighted in blue) changed functional group while the others (highlighted in light blue) had no functional group assigned in one of the works. The change in functional group (highlighted in blue) was systematically from AOB with the method here presented, to NOB from the previous work. Another point worth noting is that 7 and 9 false positives (assignment as AOB or NOB when database assigned ND) can be seen from this work and previous work, respectively. This suggests that one should consider modelling for heterotrophs, or taking out the rows corresponding to the ND from the mass Matrix W of problem ($MIQP$). The disagreement from the methods may be explained, partially, from the low relative abundances of the OTU which should create difficulties in any method.

Testing the effectiveness of the method would require chemostat experiments with a completely characterized inoculum. However some *a priori* advantages of the method here presented are highlighted: (1) Allows a deviation from a reference yield accounting for variability within a microbial community, (2) the only user-defined parameter (δ) is

suggested from experimental error (3) all OTU can be compared at the same time, and (4) low computing time.

3.6 Conclusions and Perspectives

The classification problem of assigning a functional group to the different microorganisms present in bioreactors was solved from an asymptotic observer that bypasses the choice of the growth function and mixed integer programming.

The extension of the method to other types of bioprocess is currently under development by considering the general invariants described in [27] and biological knowledge for the different functional groups interacting in the process. The complexity of ($MIQP$) offers, at least theoretically, chances that the problem is solvable for a big number of OTU, if the number of functional groups remain low.

The use of mixed integer programming seems more suitable as an engine for classification than testing combinations; it inherently handles the combinatorial nature of the task.

Acknowledgement

The authors thank GUROBI for the license to solve the QMIP. This work was supported by project Thermomic ANR-16-CE04-0003.

Chapter 4

On the Growth Rates of a Microbial Community

The previous chapter dealt with the assignation of OTU to a given function within a complex ecosystem independently of growth rate expressions. This chapter deals with the modelling of growth rates of the same microbial community classified in the previous chapter. It begins with the stability analysis of the gLV model proposed by Dumont *et al* [29]. The structure of that model is generalized by the interaction function that modulates a substrate limited growth expression. We show how microbial interactions drive the nitrification process and a method to measure the influence of interactions on the growth rate is presented. It is based on an article that has not yet been submitted. The method used to identify the growth rates was published in the European Control Conference 2020 proceedings [124], but it is completely retaken in the present article, so it is redundant to include both of them.

Microbial Interactions as Drivers of a Nitrification Process in a Chemostat

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

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.

4.1 Introduction

Microbial communities and their interactions play a central role in the understanding of microbial ecosystems [137], and a current challenge is integrating genetic sequencing data in a deterministic modelling framework [138, 133]. Using the terminology from the thorough review in current methodologies on the deterministic modelling approaches of microbial community dynamics presented by Song *et al.* [115], 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 [15]. 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 [36]. 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 [74] and Volterra [129], 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 [65]. Extensions of the Lotka-Volterra model have derived what is now known as generalized Lotka-Volterra (gLV) models [50] shown in equation (4.1):

$$\dot{x_i} = \mu_i \left(1 + \sum_{j=1}^{n} a_{ij} x_j \right) x_i \quad i \in \{1, \ldots, n\} \tag{4.1}$$

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 $(+,-)$, competition $(-,-)$, and ammensalism $(-,0)$ can be recovered [97]. Model (4.1) has been thoroughly analysed, even when the coefficients μ_i and a_{ij} are time dependent and exhibit periodicity (which models seasonal traits) [37], [**?**]. 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* [17]. 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 his name is that any function with the same shape (monotone, concave, and bounded on the substrate) would have served [85], 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.* [29] 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. 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 [21]. In section 5 the methodology from section 4 is applied to the data from the experiments performed by Dumont *et al.* [30] and not just to the most abundant species as it was the case of the model analysed in section 3 [29]. 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 can not 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.

4.2 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 $[m] := \{1, \ldots, n\}$.

4. x_i : is the concentration of OTU i measured in $[g/l]$. $i \in [n]$.

5. x: vector $(x_1, \ldots, x_n)^\top$.

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 $i \in \{1, \ldots, n\}$.

12. $\mu_i(s, x)$: growth function of OTU $i \in \{1, \ldots, n\}$.

13. $\mu = (\mu_1(s, x), \ldots, \mu_n(s, x))$ 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 $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$, the function diag$(v)$ stands for:

$$\begin{aligned} \text{diag} : \mathbb{R}^n \to & \qquad \mathbb{M}_{n\times n}(\mathbb{R}) \\ v \to & \begin{pmatrix} v_1 & 0 & \ldots & 0 \\ 0 & v_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & v_n \end{pmatrix} \end{aligned} \tag{4.2}$$

4.2.1 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:

H 1. *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 (R G1) and (R G2). 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 (R G1), 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.

$$s_1 \xrightarrow{\mu_i(s,x)} s_2 + y_i x_i \quad \forall i \in G_1 \tag{R G1}$$

$$s_2 \xrightarrow{\mu_i(s,x)} s_3 + y_i x_i \quad \forall i \in G_2 \tag{R G2}$$

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 4.1. This defines the stoichiometry matrix $Y \in \mathbb{R}^{3\times n}$, such that $Y_{ij} = y_{s_i/x_j}$.

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	$\frac{1}{y_j}$

Table 4.1: Relationship of y_{s_i/x_j} with 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) [111].

4.2.2 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 (4.3) represents this situation. A specific case of $\mu_i(s, x)$ is given in the next subsection.

$$\begin{aligned}
\dot{x_i} &= (\mu_i(s,x) - D)\, x_i \quad \forall i \in [n] \\
\dot{s_1} &= (s_{in} - s_1)D - \sum_{i \in G_1} \frac{1}{y_i}\mu_i(s,x)x_i \\
\dot{s_2} &= -s_2 D + \sum_{i \in G_1} \frac{1}{y_i}\mu_i(s,x)x_i - \sum_{i \in G_2} \frac{1}{y_i}\mu_i(s,x)x_i \\
\dot{s_3} &= -s_3 D + \sum_{i \in G_2} \frac{1}{y_i}\mu_i(s,x)x_i
\end{aligned} \tag{4.3}$$

System (4.3) can also be written in a more compact form using the stoichiometric matrix Y and the diag operator.

$$\dot{x} = \mathrm{diag}(\mu(x,s) - D_{n\times 1})x \tag{4.4}$$

$$\dot{s} = \left(\begin{bmatrix} s_{in} & 0 & 0 \end{bmatrix}^\top - s \right) D + Y\,\mathrm{diag}(\mu(x,s))x \tag{4.5}$$

4.2.3 Kinetic Equations

In the work of Dumont *et al.* [29] the growth rates seen in equations (4.6) were calibrated against experimental data for the two most abundant OTU of each functional group.

$$\begin{aligned}
\mu_i(s,x) &= \bar{\mu}_i \frac{s_1}{K_i + s_1} \left(1 + \sum_{j \in [n]} a_{ij} x_j \right) \quad \forall i \in G_1 \\
\mu_i(s,x) &= \bar{\mu}_i \frac{s_2}{K_i + s_2} \left(1 + \sum_{j \in [n]} a_{ij} x_j \right) \quad \forall i \in G_2
\end{aligned} \tag{4.6}$$

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 [85]. 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 (4.3) with growth rates given by (4.6).

4.3 Mathematical Analysis

The system of equations (4.3) 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 $t \geq 0$ [60].

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.

4.3.1 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 2. *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* (4.3) *satisfy the following inequalities:*

$$\sum_{i \in G_1} \frac{1}{y_i} x_i + s_1 \leq M_1 \tag{4.7}$$

$$\sum_{i \in G_2} \frac{1}{y_i} x_i + s_1 + s_2 \leq M_2 \tag{4.8}$$

$$s_1 + s_2 + s_2 \leq M_3 \tag{4.9}$$

The proof can be seen in the supplementary material. 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 3. *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* (4.3) *with growth rates given by* (4.6)*, remain in* Ω *and are bounded.*

The proof can be seen in the supplementary material.

The importance of Lemma 3 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.

4.3.2 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 the appendix.

Let $f(s)$ be such that,

$$f_i(s) = \begin{cases} \bar{\mu}_i \dfrac{s_1}{K_i + s_1} & \forall i \in G_1 \\ \bar{\mu}_i \dfrac{s_2}{K_i + s_2} & \forall i \in G_2 \end{cases} \tag{4.10}$$

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

Thus, system (4.3) is rewritten as follows.

$$\dot{x} = \text{diag}(\mu(x,s) - D_{n\times 1})x \tag{4.11}$$

$$\dot{s_1} = (s_{in} - s_1)D + Y_{1\bullet}\,\text{diag}(\mu(x,s))x \tag{4.12}$$

$$\dot{s_2} = -\,s_2 D + Y_{2\bullet}\,\text{diag}(\mu(x,s))x \tag{4.13}$$

$$\dot{s_3} = -\,s_3 D + Y_{3\bullet}\,\text{diag}(\mu(x,s))x \tag{4.14}$$

Definition 9. *An equilibrium point (or steady state) is a point $(x^{eq}, s^{eq}) \in \Omega$ so that the right hand side of equations* (4.11),(4.12), (4.13), *and* (4.14) *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 10. *Given an equilibrium point (x^{eq}, s^{eq}) of system* (4.3), *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.

H 2. *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 (4.11):

$$x^{act} = (A^{act})^{-1}(\text{diag}(f^{act}(s))^{-1}D_{n^{act}\times 1} - 1_{n^{act}\times 1}) \tag{4.15}$$

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 (4.12), (4.13), and (4.14) which gives :

$$s_{in} = s_1 + s_2 + s_3 \tag{4.16}$$

Both functional groups are present

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

H 3. *The set $\mathcal{J}$ satisfies $G_1 \not\subset \mathcal{J}$, $G_2 \not\subset \mathcal{J}$.*

By replacing equation (4.15) in equation (4.12) s_2 can be written as a function of s_1:

$$s_2 = \frac{s_1}{b_1 s_1^2 + b_2 s_1 + b_3} \tag{4.17}$$

Then by replacing (4.17) in equation (4.13), one gets a fourth degree polynomial for s_1.

$$a_4 s_1^4 + a_3 s_1^3 + a_2 s_1^2 + a_1 s_1 + a_0 = 0 \tag{4.18}$$

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

The equilibrium point can be calculated from the solutions of the system of equations (4.15), (4.16), (4.17), and (4.18) with non negative coordinates. If the system only

provides solutions with at least one negative entry then the set $\mathcal{J}$ can not define an equilibrium point.

Washout of G_2

The washout of G_2 is equivalent to hypothesis 4.

H 4. *$G_2 \subset \mathcal{J}$ and $G_1 \not\subset \mathcal{J}$*

Under this case note that $f^{act}(s)$ depends only on s_1. Therefore when equation (4.15) is replaced in (4.12), one obtains a quadratic equation for s_1:

$$a_2' s_1^2 + a_1' s_1 + a_0' = 0 \tag{4.19}$$

Where a_i' can be found in the appendix.

Since $x_i = 0 \, \forall i \in G_2$ then from equation (4.14).

$$\dot{s_3} = 0 = -s_3 D \tag{4.20}$$

$$\Rightarrow s_3 = 0 \tag{4.21}$$

In this case the equilibrium point can be calculated from the solutions of the system of equations (4.15), (4.16), (4.19), and (4.21) with non-negative coordinates. If the system only provides solutions with at least one negative entry then the set $\mathcal{J}$ can not define an equilibrium point.

Washout

The washout equilibria means $x_i = 0$ for every $i \in \{1, \ldots, n_1\}$. 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.

H 5. $\mathcal{J} = G_1 \cup G_2$

From equation (4.12), one gets

$$s_{in} = s_1$$

then (4.16) 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:

$$\overbrace{(2^{n_1}-1)}^{\substack{\text{nonempty subsets}\\ \text{of } G_1}} \Big(\underbrace{4}_{\substack{\text{possible}\\ \text{solutions of}\\ \text{equation (4.18)}}} \cdot \overbrace{(2^{n_2}-1)}^{\substack{\text{nonempty subsets}\\ \text{of } G_2}} + \underbrace{2}_{\substack{G_2\\ \text{Washout}}} \Big) + \overbrace{1}^{\text{Washout}} \tag{4.22}$$

4.3.3 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 (4.3) are identified [95].

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 the supplementary material (A.70). Algorithm 1 summarizes this procedure.

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 hypothesis 3, 4, or 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

Algorithm 1: Algorithm for evaluating the possible equilibrium points of system (4.3).

Operating and ecological diagrams are created by running algorithm 1 for different pairs (s_{in}, D). In the case of operating diagrams [**?**] (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 [62]. 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.

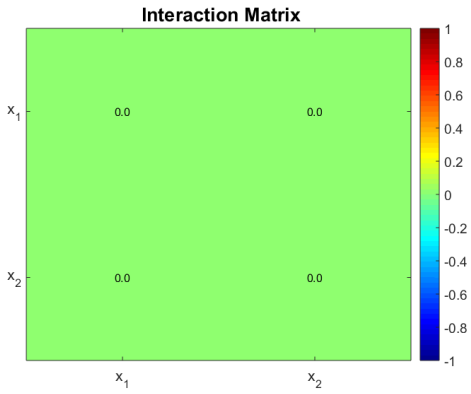
Case study 1: 1 AOB and 1 NOB

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 supplementary material section. The interaction matrices are shown in Figures 4.1a and 4.1b, 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 [137], 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 4.2.

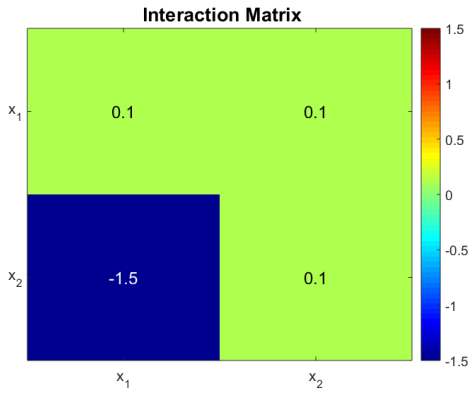
The operating diagrams can be seen in Figure 4.2, note how partial nitrification (washout of G_2) of Figure 4.2b is much bigger when compared to Figure 4.2a. The shape of the PN region in 4.2b is somewhat unintuitive, because at a constant dilution rate ($0.24\ day^{-1}$ for example) and an increasing 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 ($\frac{s_2}{s_2+K_2}$).

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

Table 4.2: A set of kinetic parameters of model (4.3).

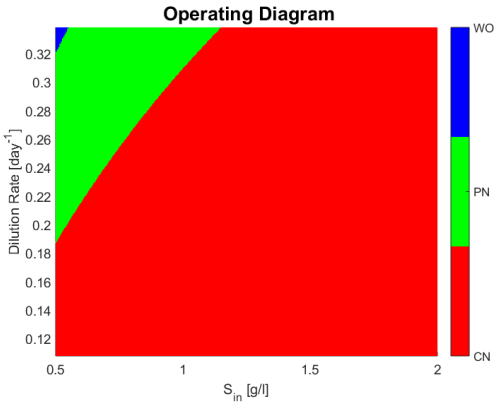


(a) Interaction matrix of model (4.3) with no interactions.

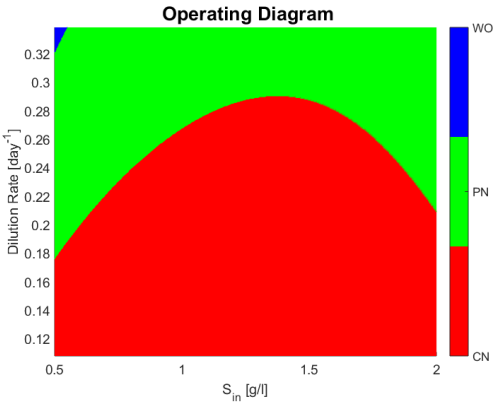


(b) A non-zero interaction matrix of model (4.3).

Figure 4.1: Interaction matrices. Note how the presence of x_1 affects very negatively x_2 in Figure 4.1b, 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) Operating diagram of model (4.3) with no interactions (interaction matrix represented by figure 4.1a).



(b) Operating diagram of model (4.3) with interactions represented by figure 4.1b.

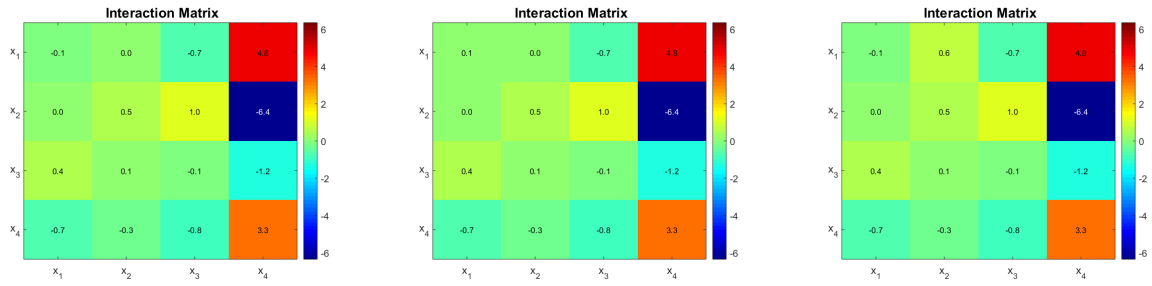
Figure 4.2: Note how 4.2b has a much larger zone where partial nitrification takes place. This is due to the negative interaction of x_1 on x_2.

Case study 2: 2 AOB and 2 NOB

The case study 2 is based on Dumont *et al.* [29] 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 4.3. The estimated interaction matrix is shown in Figure 4.3a. A second matrix is presented, which is obtained by the sign change of coefficient a_{11} (Figure 4.3b), and finally a third one is obtained by using a positive value for a_{12} (Figure 4.3c). The idea is to show that qualitatively different outcomes can be obtained by changing one interaction at a time.

Case study 2 kinetic parameters	$\mu_i\,[1/day]$	$K_i\,[mg/L]$	$\frac{1}{y_i}\,[gr/gr]$
$x_1 \in G_1$	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

Table 4.3: Kinetic parameters of model (4.3) from Dumont *et al.* [29]



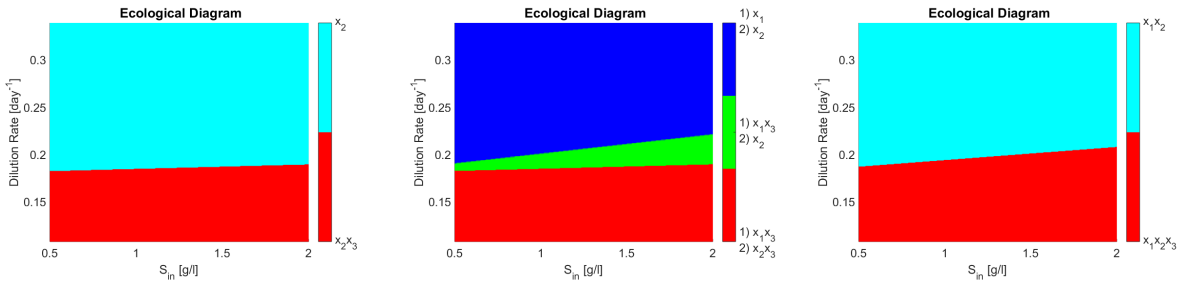
(a) Originally calibrated interaction matrix.

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

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

Figure 4.3: 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 figures (4.3b) and (4.3c), respectively.

The ecological diagrams are presented in Figure 4.4, where the legend indicates the species surviving in the zone of the respective colour. In Figure 4.4b, 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 4.4b has two stable equilibria, meaning that the



(a) Original parameters ecological diagram.

(b) ED from interaction matrix on Figure 4.3b. In the legend 1) and 2) represent the two different stable equilibria in each zone.

(c) ED from interaction matrix on Figure 4.3c.

Figure 4.4: Ecological diagrams. 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. Note that in figure 4.4b two stable equilibria exist for each zone.

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 4.4a, 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.

One can see that coexistence in the same functional group is never attained in Figure 4.4a and 4.4b, whereas in figure 4.4c x_1 and x_2, both AOB, coexist in either partial or complete nitrification. That means that the competitive exclusion principle [44] (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 [134], 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 (4.3) 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 [29], the next section proposes a new approach to study interactions.

4.4 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:

$$\begin{aligned} \mathcal{I} : \mathbb{R}_+^n \to & \qquad \mathbb{R}_+^n \\ v \to & \begin{pmatrix} I_1(v) \\ I_2(v) \\ \vdots \\ I_n(v) \end{pmatrix} \end{aligned} \tag{4.23}$$

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

$$\mu_i(s, x) = f_i(s) I_i(x) \tag{4.24}$$

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 can not 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:

H 6. *The interaction function $\mathcal{I}$ previously defined satisfies:*

1. $\mathcal{I}\left((0, \ldots, 0)^\top\right) = (1, \ldots, 1)^\top$

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 (equations (4.6)) implying that $J_I(\bar{0})$ can be seen as the interaction matrix from model (4.3).

4.4.1 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 (4.25), with $y(t) = x(t)$ being the output, because we are observing measurements coming from genetic sequencing.

$$
\begin{aligned}
\dot{x_i} &= (f_i(s)u_i(t) - D)\, x_i \quad \forall i \in G_1 \\
\dot{x_i} &= (f_i(s)u_i(t) - D)\, x_i \quad \forall i \in G_2 \\
\dot{s_1} &= (s_{in} - s_1)D + \sum_{i \in G_1} y_{s_1/x_i} f_i(s)u_i(t)x_i \\
\dot{s_2} &= -s_2 D + \sum_{i \in G_1 \cup G_2} y_{s_2/x_i} f_i(s)u_i(t)x_i \\
\dot{s_3} &= -s_3 D + \sum_{i \in G_2} y_{s_3/x_i} f_i(s)u_i(t)x_i \\
y &= x
\end{aligned}
\tag{4.25}
$$

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:

$$
\begin{aligned}
\min \quad & \int_0^T \|y - z\|_Q + \|(u - \vec{1})\|_R dt \\
s.t. \quad & (x, s_1, s_2, s_3) \text{ solution of (4.25)} \\
& u_i(t) \in [0, \bar{u}]
\end{aligned}
\tag{4.26}
$$

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 (4.25), 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 [45]. Since the ob-

jective 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.* [22, 21] 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 (4.25) 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.* [22] 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.

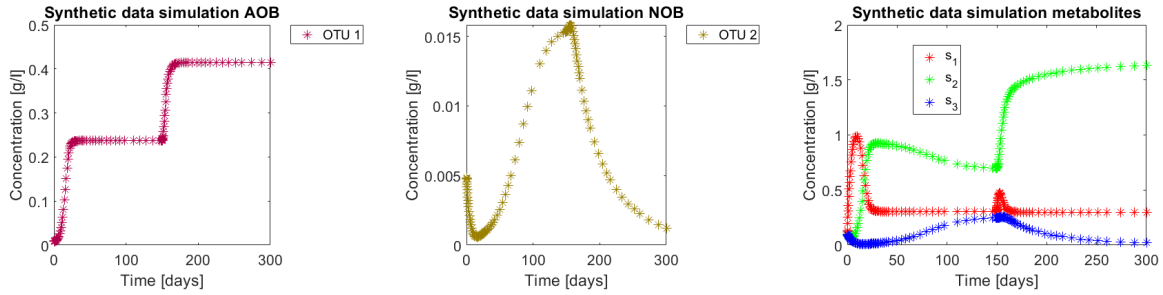
Another departure from their method is that in their formulation an approximation of the dynamics for calculating the trajectories is used. They proved that such a linearisation converges to the original dynamics. In the case here presented the linearisation of the dynamics was not necessary.

The change of variable $u_i(t) = v_i(t) + 1$ is used for technical reasons explained in the appendix. This in turn implies $v_i(t) \in [-1, \bar{u} - 1]$. The feedback control $v(t)$, will be of the form $v(t) = -R^{-1}\tilde{B}^{\top}(t)\left(\tilde{P}(t)x(t) - s_f(t)\right)$. Where matrix $P(t)$ and vector $s_f(t)$ solve differential equations. In the supplementary material 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.

4.4.2 Proof of concept

The approach was tested with data generated by simulating model (4.3) using the parameters of the case study 1 (Table 4.2) with interaction matrix given by (4.1b). In the operating diagram of the same case (figure (4.2b)) 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 4.5. Note how from day 150 the NOB population (OTU 2) represented in figure (4.5b) decreases, which in turn implies a decrease in s_3, as seen in figure (4.5c). In the case where no interactions take place, the OD seen in Figure 4.2a 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 4.5: Synthetic data generated by model (4.3), 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 4.2) 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

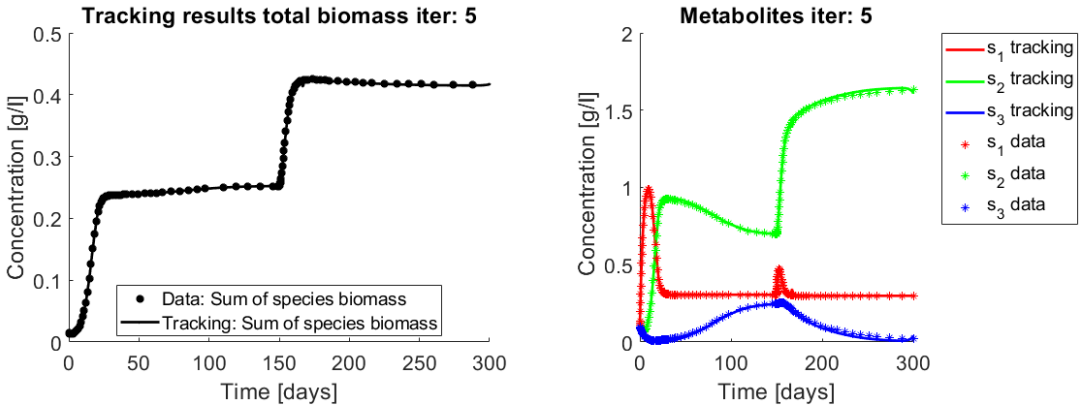
the optimal control becomes $u = 1$, thus no tracking is performed. Further diminishing the value from 10^{-5} adds too much noise to the control, without significant gains on the quality of the tracking.

The results of the procedure to identify interactions can be seen in figures 4.6 and 4.7. Figures 4.6a, 4.7a, and 4.7b 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 4.6b are in accordance with the simulated: The method is able to reconstruct the metabolites trajectories from the community measurements.

Figures 4.8 and 4.9 show the controls and the corrected growth rate for each functional group. The control for each functional group can be seen in figures 4.8a and 4.9a. 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 4.8b and 4.9b show the resulting growth rates for AOB and NOB, respectively, without the control $u_i(t)$. Figures 4.8c and 4.9c 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 4.8b and 4.9b, 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 $f_i(s(t))x_i(t)$.

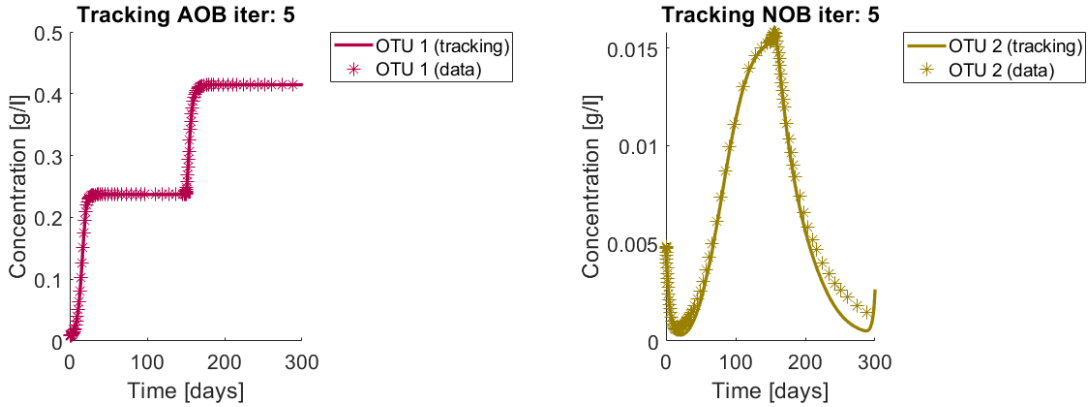
A final comment on the identifiability of the interaction terms. Even though one might propose a growth rate with the tracking control $u(t)$ that accurately replicates the OTU trajectory $x(t)$, retrieving the original interaction coefficients from the obtained control for this example was not possible. 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 (4.6) in model (4.3) might be very low.



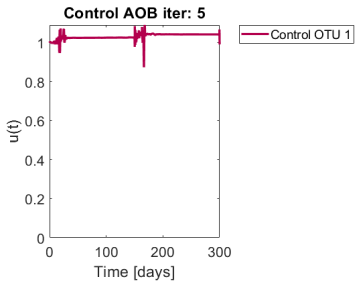
(a) Total biomass. (b) Metabolites.

Figure 4.6: 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.

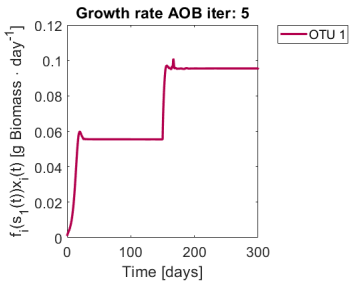


(a) AOB biomass. (b) NOB biomass.

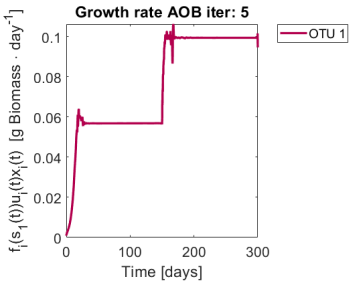
Figure 4.7: Asterisks represent the synthetic data, while the continuous lines represent the method's output. The method reconstructs a continuous trajectory from the synthetic data.



(a) Control for the AOB community.

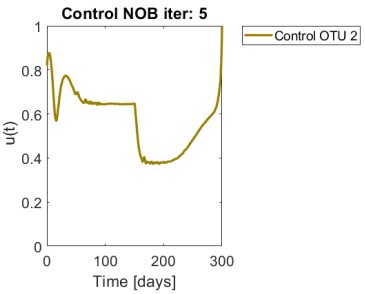


(b) Growth rate "without" the control.

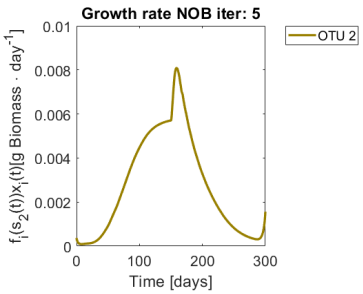


(c) Complete reconstructed growth rate for the AOB community.

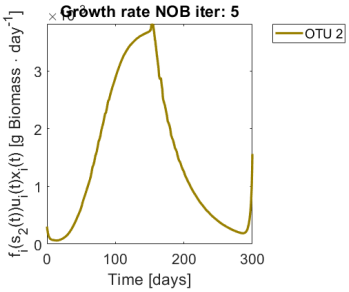
Figure 4.8: Obtained control and reconstructed growth rate for OTU 1 (AOB).



(a) Control for the NOB community.



(b) Growth rate "without" the control.



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

Figure 4.9: Control and reconstructed growth rate for OTU 2 (NOB).

4.5 Application

The tracking problem was applied to data coming from a nitrification process with experimental conditions described in [32]. 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.* [125] 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 [139]. Particularly given the system's ammonium and nitrite concentration a Monod function (eq (4.27)) was used for G_1 and G_2 with parameters given in table 4.4 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 lies in the same order of magnitude than synthetic data.

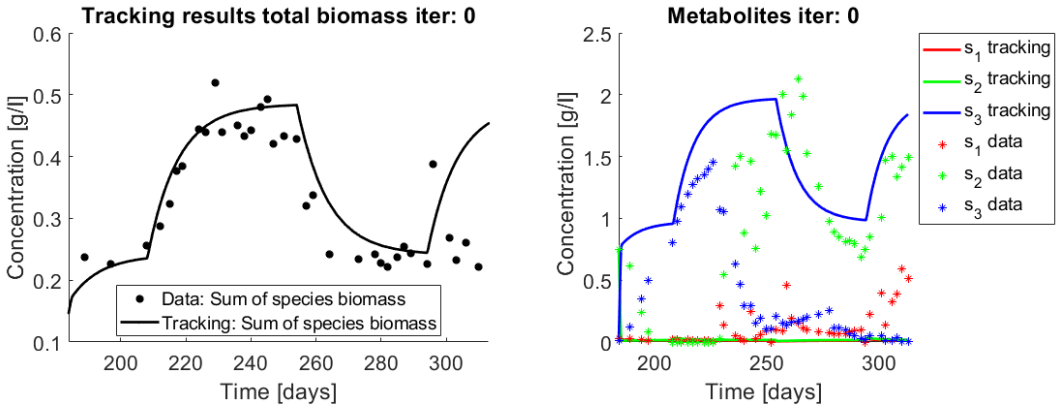
$$f_i(s) = \bar{\mu}_2 \frac{s_1}{K_1 + s_1} \forall i \in G_1 \tag{4.27}$$
$$f_i(s) = \bar{\mu}_2 \frac{s_2}{K_2 + s_2} \forall i \in G_2$$

For the reader to gain understanding of the situation, a simulation of the system

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

Table 4.4: A set of kinetic parameters of model (4.25).

using the experiments operating parameters (D and s_{in}) is presented without control (i.e. $u(t) = 1$) in figure 4.10 nitrate (s_3) accumulates all along the trajectory, but when compared to data it is clear that s_3 stops accumulating after a while.



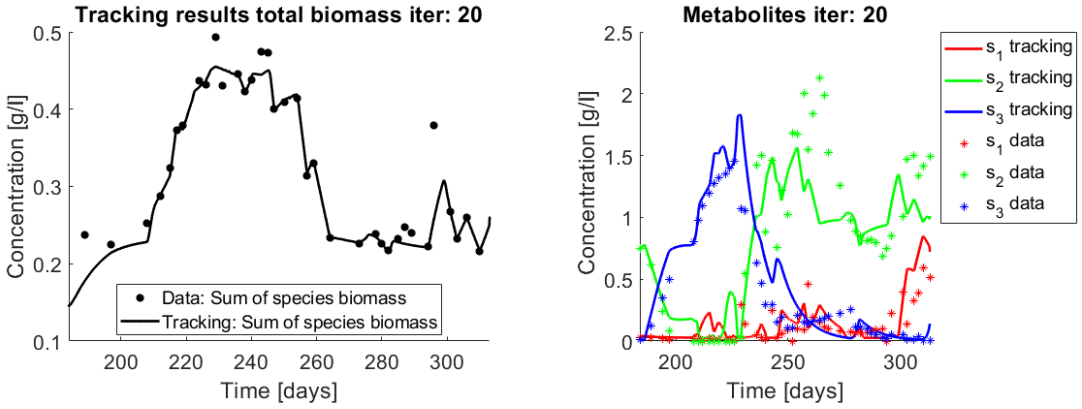
(a) Total Biomass. (b) Metabolites.

Figure 4.10: Simulation of system (4.25) when $u = 1$, with functions as in (4.27). Data points are represented by a star. The continuous line represents the simulation.

When applying the tracking method one obtains the simulation that can be seen in figures 4.11 and 4.12. The method captures the tendencies of the measured substrates as seen in figure 4.11b. The tracking of each functional group G_1 (AOB), and G_2 (NOB) can be seen in figures 4.12a and 4.12b, respectively.

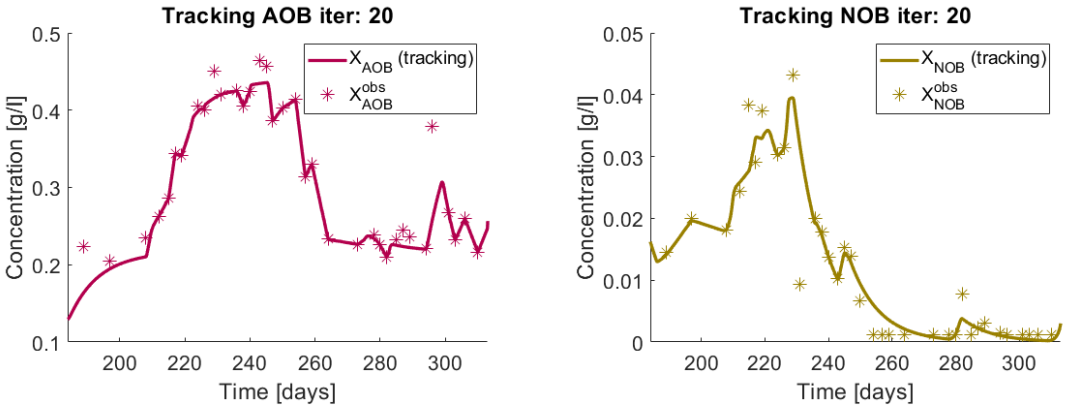
The growth rates of each functional group are shown in figure 4.13. Note in the case of AOB (figure 4.13a) the resulting growth rate shows a noisy curve formed by pulses. The behaviour of the NOB community (figure 4.13b) is qualitatively very similar with somewhat stronger pulses and less noise. The former is to be expected since more OTU were regrouped to compose the AOB biomass, therefore more noise sources were added.

The same procedure is applied without regrouping. The results on total biomass and metabolites are shown in figures 4.14a and 4.14b, respectively. Both patterns still fit the data, but to a lesser degree of precision when compared to figure 4.11. This can be explained by inspecting figures 4.15 to 4.20. First note the absolute error of the tracking



(a) Total Biomass. (b) Metabolites.

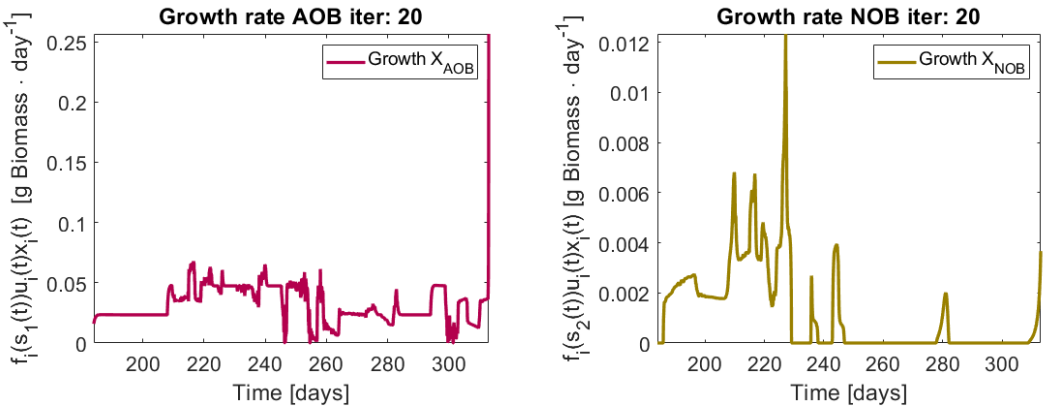
Figure 4.11: Results on applying the tracking method to a nitrification experiment when regrouping OTU in their functional groups. Data points are represented by asterisks. The continuous line represents the tracking procedure results.



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

Figure 4.12: The tracking procedure applied to the observed biomass (asterisks) regrouped in two functional groups.

for each of the OTU in the AOB community (figures 4.15b to 4.19b), 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 4.20b), 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



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

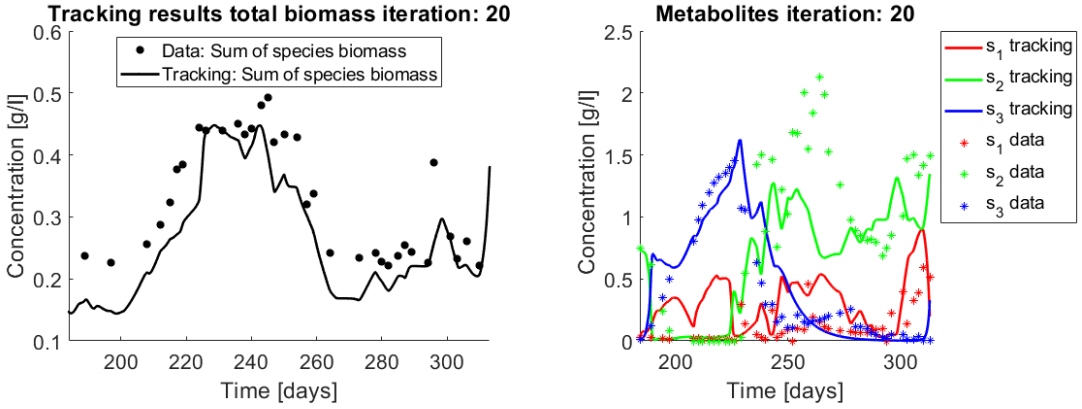
(b) Growth rate of the model for the NOB community.

Figure 4.13: Obtained growth rates when regrouping OTU in their functional groups.

entries of matrix Q) and did not improve the results. When looking at the growth rates (figures 4.15c to 4.20c) one again observes pulses for each OTU. Finally, note that most OTU were present only for a fraction of the experiment's duration.

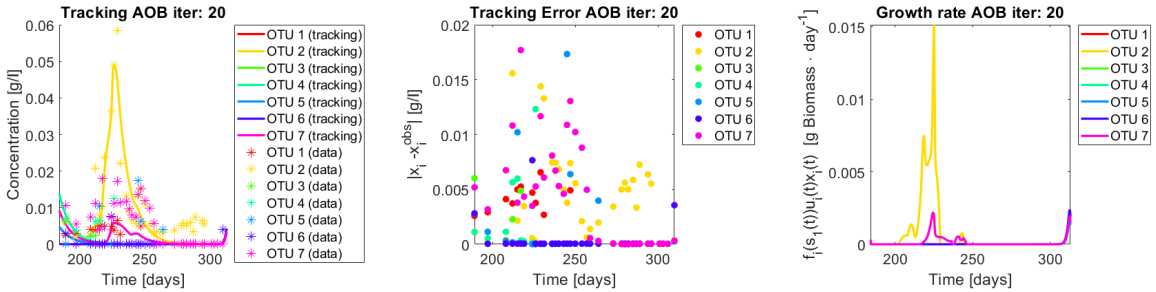
In both cases, the regrouped and individual tracking, the growth rate varies strongly, raising the question whether the observed pulses are emerging from interactions within the microbial community. When growth rates are compared to the proof of concept section it seems doubtful that a linear pairwise 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 can not 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 hypothesis for explaining the observed patterns in the microbial community should be considered as well.



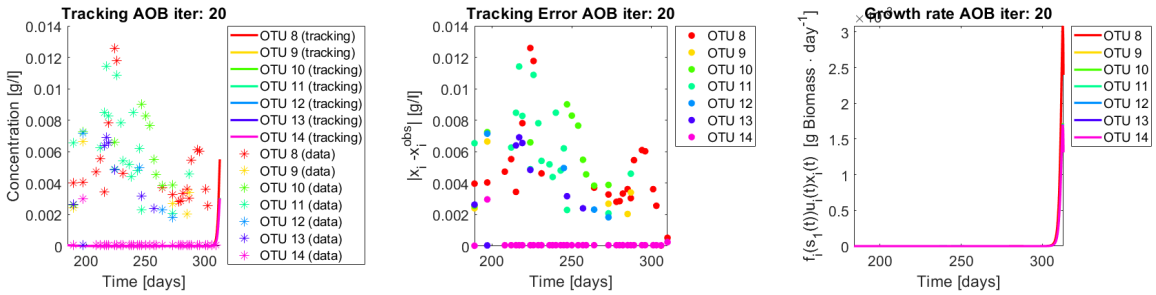
(a) Total Biomass. (b) Metabolites.

Figure 4.14: 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.



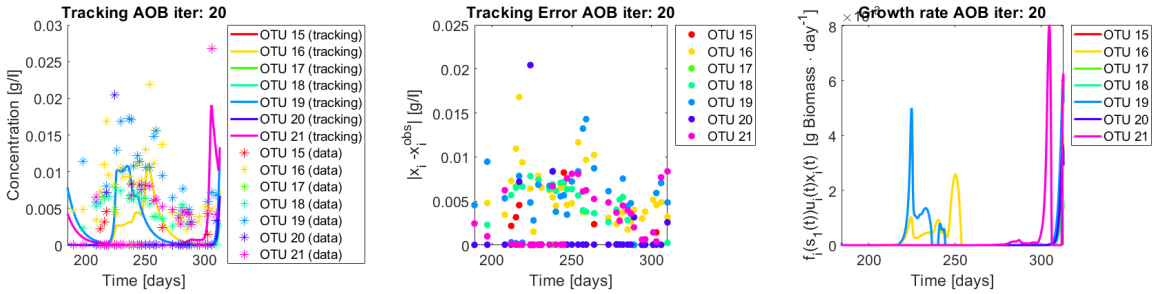
(a) Tracking results and data. (b) Absolute difference between tracking and data. (c) Reconstructed growth rate.

Figure 4.15: Results for OTU 1-7 (AOB)



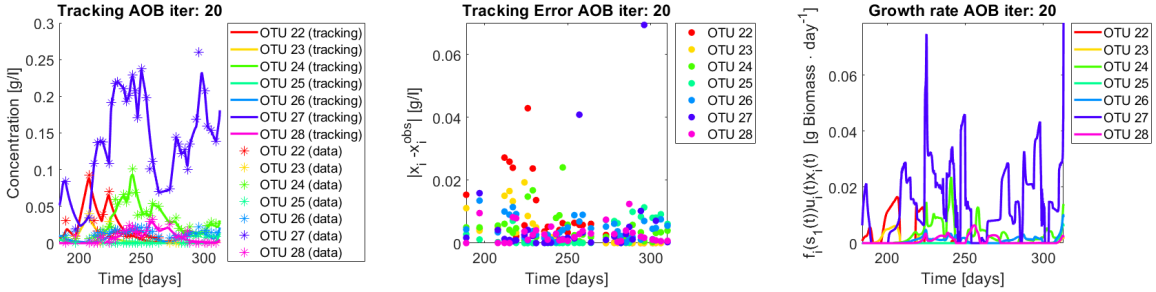
(a) Tracking results and data. (b) Absolute difference between tracking and data. (c) Reconstructed growth rate.

Figure 4.16: Results for OTU 8-14 (AOB)



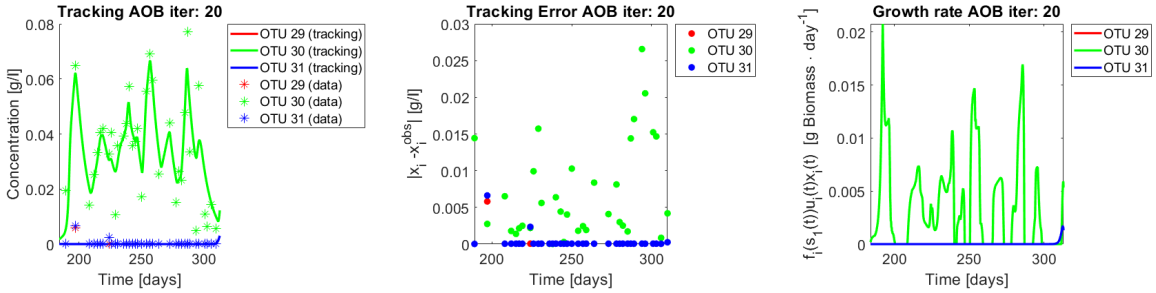
(a) Tracking results and data. (b) Absolute difference between tracking and data. (c) Reconstructed growth rate.

Figure 4.17: Results for OTU 15-22 (AOB)



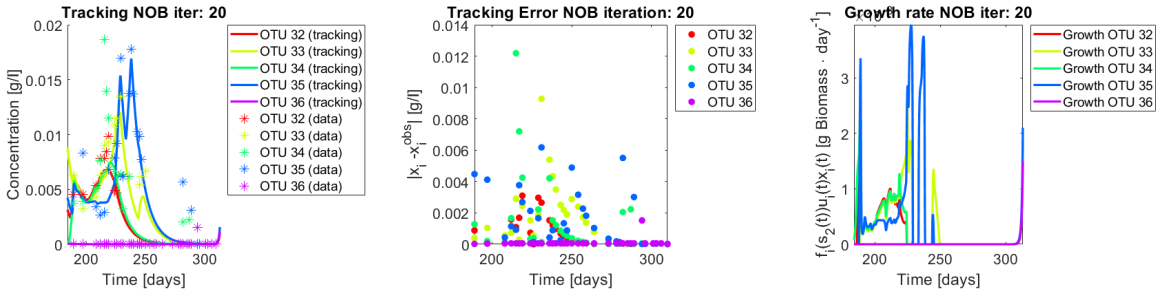
(a) Tracking results and data. (b) Absolute difference between tracking and data. (c) Reconstructed growth rate.

Figure 4.18: Results for OTU 23-28 (AOB)



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

Figure 4.19: Results for OTU 29-31 (AOB)



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

Figure 4.20: Results for OTU 32-36 (NOB)

4.6 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.* [29] 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 since the quality of molecular fingerprints at the moment of the data collection was not the best (techniques for raw data treatment were still discussed at the time), one can not discard the possibility that this is due to noise of the measurements. In spite of the former, 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 [27], integrating genetic sequencing could provide a path for more certainty in model calibration.

4.7 Codes

All the codes can be recovered from the following repository `https://github.com/paus-5/Class-and-Track`.

4.8 Acknowledgements

The authors would like to thank Alain Rapaport for the exchanges and comments. This work was financed by project Thermomic ANR-16-CE04-0003.

Chapter 5

On Microbial Transition State Theory Implications

The final chapter of this thesis studies the growth rates of OTU from a theoretical standpoint. This section explores the recent theory of microbial growth, Microbial Transition State Theory (MTS), proposed by Desmond-Le Quemener and Bouchez [25]. We begin by an article that was published in Scientific Reports that introduces their model and explores the meaning of Monod's affinity constant under this setting. The novelty lies in a link between K_s and the substrate to biomass yield. In the second part further implications are drawn from the MTS theory, dealing with density dependence and variable yields.

5.1 Insights from Microbial Transition State Theory on Monod's Affinity Constant

Insights from Microbial Transition State Theory on Monod's Affinity Constant

Pablo Ugalde-Salas*[1], Elie Desmond-Le Quéméner**[1], Jérôme Harmand[1], Alain Rapaport[2], Théodore Bouchez[3].

[1] LBE, INRAE, Univ Montpellier, 102 avenue des Etangs, 11100, Narbonne, France

[2] MISTEA, INRAE, Univ. Montpellier, Montpellier SupAgro

[3] Université Paris-Saclay, INRAE, UR PROSE, 92160, Antony, France

* Corresponding author: Tel: +33 (0)4 68 42 51 62. Mail : pablo.ugalde-salas@inrae.fr

** Corresponding author: Tel: +33 (0)4 68 42 51 87. Mail : elie.le-quemener@inrae.fr

Abstract

Microbial transition state theory (MTS) offers a theoretically explicit mathematical model for substrate limited microbial growth. By considering a first order approximation of the MTS equation one recovers the well-known Monod's expression for growth, which was regarded as a purely empirical function. The harvest volume of a cell as defined in MTS theory can then be related to the affinity concept, giving a new physical interpretation to it, and a new way to determine its value. Consequences of such a relationship are discussed.

Since the success of Monod's expression (Equation (5.1)) to model substrate-limited microbial growth [85], many expressions have been proposed [7], accounting for a range of phenomena including substrate inhibition [4] and population density effects [23]. All of these expressions rely on empirical rules, differently to enzymology for which analogues of Monod and Haldane expressions have been mathematically derived [87]). Microbial transition state theory [25] recently introduced a new expression for microbial growth based on the statistics of molecules distribution in a medium inspired from chemical transition state theory. In this communication we explore the physical meaning of the affinity concept through the lens of MTS theory and particularly show how it may provide a novel interpretation of Monod's growth function.

$$\mu(s) = \mu_{max} \frac{s}{K_s + s} \tag{5.1}$$

Equation (5.1) represents the Monod growth function, where $\mu_{max}[1/day]$ is the maximal growth rate, $s[g/L]$ represents the substrate concentration of the medium and $K_s[g/L]$ is known as the 'affinity constant'. Earlier works on kinetics [94], [67] show the differences in reported literature values for the affinity constant for the same species: these differences are explained by culture history, quality of the experimental data, and posterior data analysis. However little to no consensus can be found in the literature on its interpretation. Furthermore in a review of theoretical derivations of the Monod growth function [71] the author concludes that no clear interpretation may be given to the affinity constant. A revision of the affinity concept in Microbiology was made by Button [18], where fourteen different expressions for affinity are documented. The concept is largely influenced by the Michaelis-Menten model for enzyme kinetics interpretation of affinity from receptor and ligand binding sites, since Monod's expression for growth is mathematically equivalent to the Michaelis-Menten expression. As stated by Monod himself, Monod's growth function is purely empirical, while Michaelis-Menten expression has a rigorous theoretical justification [87], thus one might wonder if the concept of affinity for representing cell growth has a solid conceptual ground.

MTS theory relates the growth rate to the amount of energy available to perform cellular work. The central idea of bioenergetics is that the energy consuming anabolism can only be thermodynamically feasible if it is coupled with an energy yielding catabolism. The overall reaction resulting from the coupling is known as metabolism [64]. The formulation and complexity of both catabolism and anabolism vary greatly depending on the objective the modeller has in mind. On the one hand, when describing the metabolic pathways within a specific microbial species, the formulation takes into account ATP formation and intra cellular intermediates and quickly becomes a very complex web, e.g. [100]. On the other hand if one is interested in observing the general metabolism of a culture at a macroscopic level then the situation simplifies to just a couple of reactions

[64] . We will focus on the latter.

Let us consider a first reaction representing catabolism (Equation (5.2)), a second reaction representing anabolism (Equation (5.3)) then a linear combination of the two creates metabolism (Equation (5.4)): by completing λ times the catabolism the energy requirements of the global metabolic reaction are fulfilled 13 (its negative free enthalpy constitutes the driving force for growth).

$$\mathrm{E}_d + \mathrm{aE}_a \longrightarrow \mathrm{bP} \quad \Delta_r G_{cat} < 0 / \cdot \lambda > 0 \tag{5.2}$$

$$\mathrm{dP} \longrightarrow \mathrm{B}_x + c\,\mathrm{E}_a \quad \Delta_r G_{an} > 0 \tag{5.3}$$

$$\lambda \mathrm{E}_d + (\lambda \mathrm{a} - \mathrm{c})\,\mathrm{E}_a \longrightarrow \mathrm{B}_x + (\lambda \mathrm{b} - \mathrm{d})\mathrm{P} \quad \Delta_r G_{met} = \lambda \Delta_r G_{cat} + \Delta_r G_{an} < 0 \tag{5.4}$$

where E_d, E_a, and P stand for electron donor, electron acceptor, and products, respectively. B_x represents an equivalent biomass unit, for instance $B_x = \mathrm{CH_{1.8}O_{0.5}N_{0.2}}$ is a generic composition of one C-mole of biomass [10]. a, b, c, d are stoichiometric coefficients. Finally $\Delta_r G$ represents the Gibbs free energy variation for each reaction.

The reader should notice that λ is the inverse of the yield as usually expressed ($y_{x/s}$) in microbiology as shown in the equation (5.5)

$$y_{x/s} = \frac{1}{y_{s/x}} = \frac{1}{\lambda} \tag{5.5}$$

$y_{x/s}$ represents how many moles of biomass are formed per mole of substrate consumed, conversely $y_{s/x} = \lambda$ represents how many moles of substrate are being consumed per mole of biomass formed. The methods reviewed by Kleerebezem et al. [64] allow computing λ from mass balanced reactions with examples coming from a variety of biological process.

MTS theory demonstrates on a theoretically explicit ground a growth rate expression μ of a culture of bacteria limited by an electron donor in perfectly mixed conditions [25]. More precisely, if we denote by s the concentration of the limiting electron donor and x the concentration of the species then these two concentrations are dynamically related

by:

$$\dot{x} = \mu(s)x = \mu_{max} \exp\left(\frac{-\lambda}{V_h s}\right) x \tag{5.6}$$

where V_h, known as the harvest volume, represents the volume to which each microbe has access in order to harvest the substrate s during the time between two cell divisions. It is worth pointing out that the harvest volume is an average characteristic.

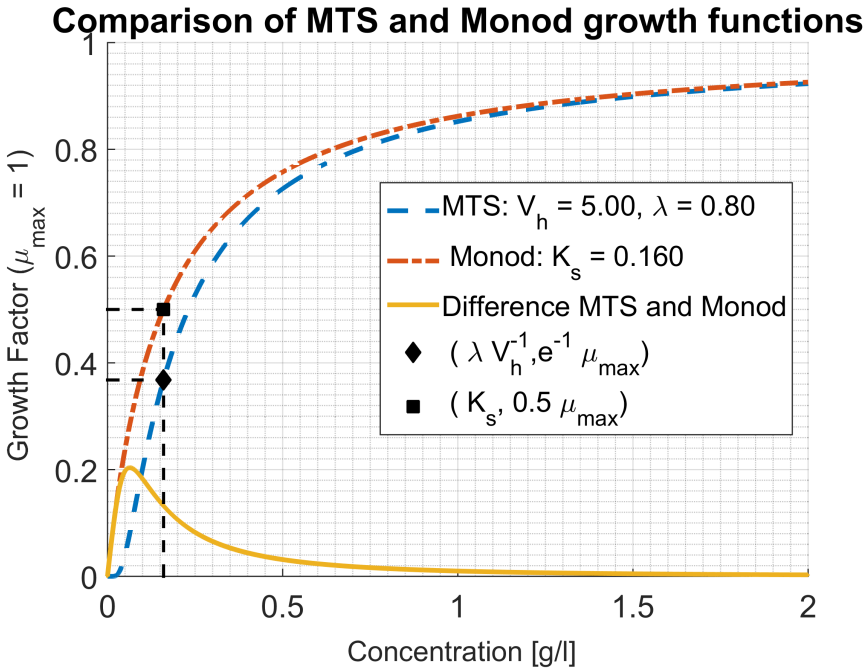


Figure 5.1: Example of plots of equations 1 and 6, with the values chosen such that $K_s = \frac{\lambda}{V_h}$. The measurement of the harvest volume from growth experiments can be obtained in an analogous fashion to the determination of the affinity constant: by noting s^* the value of substrate concentration at which the growth rate is $e^{-1}\mu_{max}$ (represented by the black diamond) one obtains V_h by the formula $V_h = \frac{1}{y_{x/s}s^*1}$, similarly to the K_s value identified as the concentration for which the specific growth rate μ is equal to $\frac{\mu_{max}}{2}$ in the Monod expression, (represented by the black square).

If one considers a first order approximation of the exponential function near zero (see supplementary material) then one recovers Monod's expression of growth:

$$\mu_{max} e^{-\frac{\lambda}{V_h s}} \approx \mu_{max} \frac{s}{s + \frac{\lambda}{V_h}} = \mu_{max} \frac{s}{s + \frac{1}{y_{x/s} V_h}} \tag{5.7}$$

The approximation holds true for high substrate concentrations. More precisely, it can

be shown that the two curves differ by less than 10% for $s \geq 1.92K_s$, (see supplementary material). In Figure 5.1 the graphical comparison of both growth functions can be seen for a given set of parameters. The MTS growth function is approximated very well by the Monod growth function, which is reassuring from a practical point of view: in a re-examination of the kinetics of *Escherichia coli* [110] different empirical substrate limiting expressions- all of them with a Monod-like shape- were compared and no difference was found in the identifiability of their parameters. Note also that in equation (5.7), $\frac{1}{y_{x/sV_h}}$ replaces the K_s parameter of Monod's expression. In that sense the affinity constant can be interpreted as a decreasing function of the harvest volume of the cell and its yield per mole of substrate. On one hand, associating low K_s values to large harvest volumes is well in line with our understanding of the affinity concept, since a cell that can harvest substrate molecules in a more extended region should be less substrate limited. On the other hand, the fact that a low K_s value could be due to a higher conversion yield of substrate to biomass sheds a new light on the affinity concept. The order of magnitude of V_h can be seen from Table 1 for some literature references for *E. coli* ML 30. In the cases where no yield was reported the energy dissipation method 11 can be used as illustrated in table 1 and supplementary material. For computing the yield a unique biomass formula was used ($CH_1 \cdot 8\,O_0 \cdot 5\,N_0 \cdot 2$). However, for each case, the biomass composition could be different and, consequently, the yield, thus contributing to the explanation of the observed variability of K_s.

K_s reported $[\mu g/l]$ for E. Coli ML 30	λ [gS/gX] * Estimated by Energy dissipation method ** Measured during experiment	$V_h[l/gX]$	$V_h[\mu m^3/cell]$ (cell dry weight: $2.8 \cdot 10^{-13}[gr/cell]$ Ref: BNID 103904 [81])	Radius $[\mu m]$ of a sphere of Volume V_h.
33 (Ref: [66])	$\lambda^* = 1.89$	$4.91 \cdot 10^4$	$1.6 \cdot 10^7$	156
33 (Ref: [66])	$\lambda^* = 1.88$	$4.85 \cdot 10^4$	$1.6 \cdot 10^7$	156
53 (Ref: [110])	$\lambda^* = 1.88$	$3.02 \cdot 10^4$	$9.94 \cdot 10^6$	133
72 (Ref: [110])	$\lambda^* = 1.88$	$2.22 \cdot 10^4$	$7.32 \cdot 10^6$	120
76 (Ref: [69])	$\lambda^{**} = 2.22$	$2.92 \cdot 10^4$	$8.19 \cdot 10^6$	125
90 (Ref: [69])	$\lambda^{**} = 2.22$	$2.47 \cdot 10^4$	$6.91 \cdot 10^6$	118
100 (Ref: [69])	$\lambda^{**} = 2.22$	$2.22 \cdot 10^4$	$6.22 \cdot 10^6$	114
132 (Ref: [69])	$\lambda^{**} = 2.22$	$1.68 \cdot 10^4$	$4.71 \cdot 10^6$	104
125 (Ref: [69])	$\lambda^{**} = 2.22$	$1.77 \cdot 10^4$	$4.98 \cdot 10^6$	105

Table 5.1: Literature values of K_s and calculation of $V_h = \frac{\lambda}{K_s}$, for different chemostat experiments using hexoses as substrates.

On a more conceptual ground, the MTS approach proposes a way to revisit our current perception of the "affinity-concept" of a microbial culture for a given substrate. It offers an alternative view of the microbial affinity notion than its enzymatic analogue related to Michaelis Menten theory. It unravels a contribution that is related to the yield (mole of biomass formed per mole of substrate consumed) from another that represents the capacity of the microbial culture to explore a fraction of its surroundings in order to harvest substrate (V_h term). To this extent, it allows to compute the affinity constant from the knowledge of the yield and the harvest volume, which is a completely new approach to determining this constant. This analysis thus plants a seed towards a more physically grounded view of affinity than earlier proposals made from attempts to theoretically derive Monod's equation[71]. The physical interpretation of the affinity concept raises new opportunities to analyse and experimentally challenge the meaning of the V_h parameter. Particularly interesting would be to assess to which extent V_h constitutes an intrinsic trait of the microbial culture, or if extrinsic attributes associated to the culture conditions (such as agitation, viscosity or ionic force) could also significantly influence its value. Such questions remain open and obviously await further studies.

Conflict of Interest The authors declare no conflict of interest.

Acknowledgements This work was supported by project Thermomic ANR-16-CE04-0003. The authors would like to thank Roman Moscoviz for the fruitful discussions and exchanges.

In the appendix A.3 the details in the approximation of the MTS expression are given.

5.2 Extension of the model

The MTS model was derived from statistical mechanics as mentioned earlier. This section changes the hypothesis that the harvest volumes of different organisms do not intersect, in order to account for the case of high population densities that may arise in ecosystems.

Consider a system consisting of an isolated physical space of volume V_{tot} in which a clonal population of N microbes is consuming substrate s and dividing. Each individual

microbe has access to a volume V_h in which it can harvest the substrate. The probability to find oneself in one volume V_h is $p = \frac{V_h}{V_{tot}}$. The probability of not being in any harvest volume is then $(1-p)^N$. Therefore, the probability of finding oneself in the union of all harvest volumes becomes $1-(1-p)^N$. This allows to define the apparent harvest volume, (5.8), which essentially represents the volume in which substrate is available to microbes.

$$V_{app} = \frac{1 - \left(1 - \frac{V_h}{V_{tot}}\right)^N}{N} V_{tot} \tag{5.8}$$

The microbes concentration in the medium x is defined as $x = \frac{N}{V_{tot}}$. Therefore

$$V_{app}(x) = \frac{1 - \left(1 - \frac{V_h}{V_{tot}}\right)^{xV_{tot}}}{x} \tag{5.9}$$

We redefine in the same framework of statistical mechanics of [25] the elementary volume $n_{app} := \frac{V_{tot}}{V_{app}}$, all their computations follow exactly the same which gives a growth function of

$$\mu(x,s) = \mu_{max} \exp\left(\frac{-\lambda}{V_{app}(x)s}\right) \tag{5.10}$$

As done in the previous section, function (5.10), can be approximated by:

$$\mu(x,s) \approx \mu_{max} \frac{s}{s + \frac{\lambda}{V_{app}(x)}} \tag{5.11}$$

Note $V_{app}(x)$ decreases monotonically with respect to x, which implies that $\mu(x,s)$ decreases monotonically with respect to x, which can be interpreted as a decreasing growth rate from density dependence.

Now consider the order 2 approximation centred at 0 of $f(y) = (1-y)^N$. That is $f(y) = f(0) + f'(0)y + 1/2f''(0)y^2 + o(|y|^2) = 1 - Ny + \frac{1}{2}N(N-1)y^2 + o(|y|^2)$. Replaced

in (5.8) it gives:

$$V_{app} = \frac{1 - \left(1 - N\frac{V_h}{V_{tot}} + \frac{1}{2}N(N-1)(\frac{V_h}{V_{tot}})^2 + o(|\frac{V_h}{V_{tot}}|^2)\right)}{N} V_{tot} \tag{5.12}$$

$$\approx V_h \left(1 - \frac{1}{2}xV_h\right) \tag{5.13}$$

Which implies one can further approximate (5.10), by:

$$\mu(x,s) \approx \mu_{max} \frac{s}{s + \frac{\lambda}{V_h\left(1 - \frac{1}{2}xV_h\right)}} \tag{5.14}$$

$$\approx \mu_{max} \frac{s}{s + \frac{\lambda}{V_h}\left(1 + \frac{1}{2}xV_h\right)} \tag{5.15}$$

$$= \mu_{max} \frac{s}{s + \frac{\lambda}{V_h} 1 + \frac{\lambda}{2}x} \tag{5.16}$$

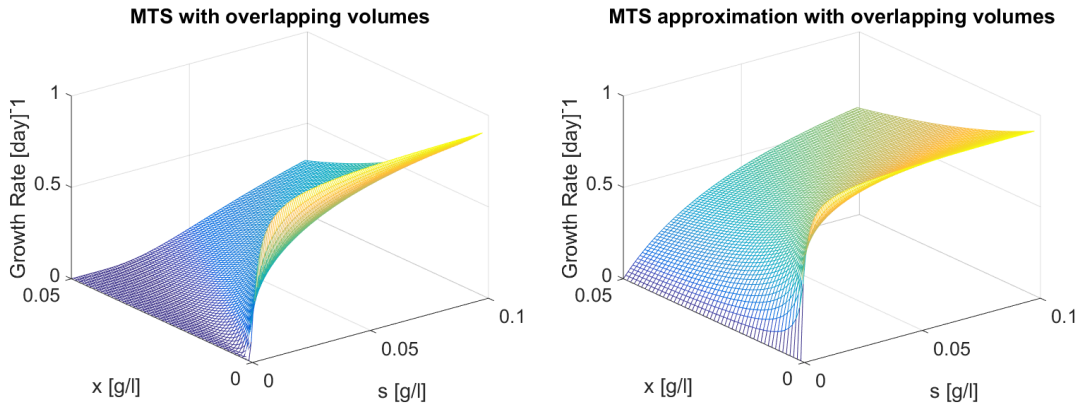
Where the approximation $\frac{1}{1-\epsilon} \approx 1 + \epsilon$, was made from the passage of equation (5.14) to (5.15). Expression (5.16) generalizes at the same time Contois and Monod's growth functions. Unexpectedly, approximation (5.16) has the form of one of the growth expressions proposed by Roques *et al* [106]. They studied the validity of Monod's expression in activated sludge systems and experimentally it was shown that the K_s constant of the Monod expression depends linearly on the difference of the input substrate and the substrate concentration, namely: $K_s = M + b(s_{in} - s)$. Then they approximated the term $s_{in} - s$ by $\frac{x}{y_{x/s}}$. From the former section, if one remembers that $\lambda = \frac{1}{y_{x/s}}$, one recovers their formula with $b = \frac{1}{2}$ and $M = \frac{\lambda}{V_h}$.

This section ends by revisiting the parameters of the article of Lendenmann *et al.* [69], they measured the maximum growth rate μ_{max}in batch experiments, and the K_s value for chemostat experiments for *E. coli* ML30. (The same last entries as in table 5.1). In figure 5.2 one can see a comparison of expression (5.10) and (5.16) using the set of parameters of table 5.2, and by doing the identification of $K_s = \frac{\lambda}{V_h}$. Clearly the approximation has a much softer density inhibition than the original function.

Moreover let us see if the density dependence postulated by the MTS theory pre-

μ_{max}	K_s	yield ($y_{x/s}$)	s_{in}	Dilution Rate	Reactor volume
$0.92[day]^{-1}$	$90[\mu g/l]$	$0.45[gX/gS]$	$0.1[g/l]$	$0.14 - 0.71[day]^{-1}$	$0.93[l]$

Table 5.2: Experimental conditions and fitted parameters of a Monod growth function to E. Coli ML30 growing on a continuous culture fed with galactose [69]. μ_{max} was calculated from batch experiments.



(a) Plot of equation (5.10) using parameters of table 5.2.

(b) Plot of equation (5.16) using parameters of table 5.2.

Figure 5.2: Comparison of the MTS function and its approximation considering density dependence.

dicts their experimental data. Figure 5.3 shows a chemostat simulation using a Monod growth rate with the original parameters (dotted line) and the the extended MTS growth rate with the identification $K_s = \frac{\lambda}{V_h}$ (continuous line), the blue lines represent biomass, while the red lines represent the substrate, galactose in this case. The green dot shows the residual galactose measured in the reactor ($107\,[\mu g/l]$), evidently, since the Monod expression was fitted to the data, the red dotted line ends at the residual galactose, while the continuous red line ends at $6000\,[\mu g/l]$, one order of magnitude higher than the residual galactose. This suggests, perhaps, that the implicitly assumed hypothesis of independence in the computing of the probability of not being in any harvest volume ($(1-p)^N$) was too strong. The spontaneous formation of flocks, biofilms and granules in microbial ecosystems indicates that some microbes arrange preferentially in aggregates, and not independently from each other. It would be interesting from a mechanical statistics approach to numerically compute the probability distribution of not being in any harvest volume from chemostat experiments data, in order to gain insight on microbes tendency of organizing themselves on a given continuous culture.

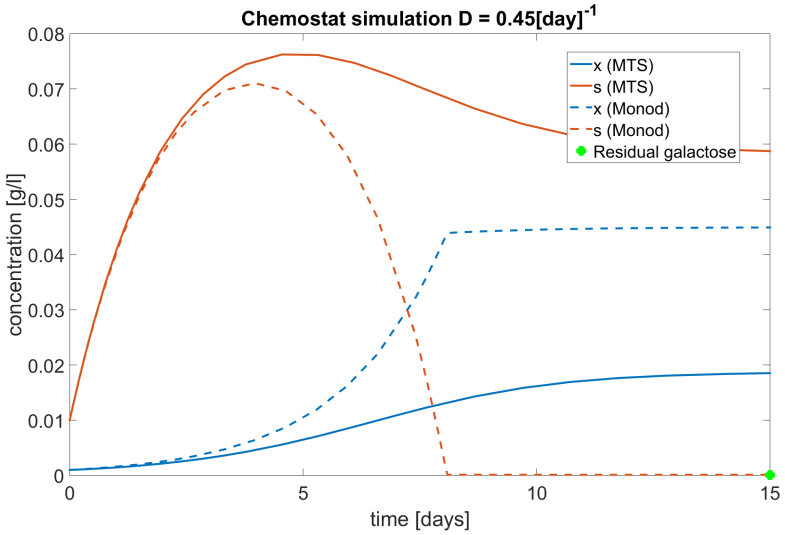


Figure 5.3: Simulation of a chemostat model with MTS growth and Monod growth considering the parameters in 5.2. The density inhibitory effects are too strong and the resulting residual galactose is extremely elevated.

5.3 Diversity arising from MTS Theory

This section was inspired by the article of Großkopf and Soyer [40]. They achieved coexistence of two species feeding on the same substrate in a chemostat model by using a growth rate that corrects a Monod function by an expression involving the Gibbs free energy of the metabolic reaction, as introduced by Hoh [52]. Davidson *et al.* had already studied, in both theoretical and experimental ways, how product inhibition in the growth function can imply coexistence of two species feeding on the same substrate [24]. Note that the Gibbs free energy ΔG of a feasible reaction (negative Gibbs free energy), is decreasing with respect to the chemical activity of its reactants, and increasing with respect to the chemical activity of the products. And as such if a growth rate $\mu(s, \Delta G)$ is decreasing with respect to the Gibbs free energy, it means it is decreasing with respect to the concentration of the product of the reaction.

Here, the coupling coefficient λ is calculated from the Gibbs free energy of the catabolic and anabolic reaction. The stoichiometric coefficients of metabolism are an affine function of λ, and in MTS theory the stoichiometric coefficient corresponding to the limiting substrate appears in the growth expression. Therefore one can expect that the same effect as in [40] might arise when correcting the Gibbs free energies of the involved reactions.

In anaerobic systems, it is the case that the activities of molecules can affect the Gibbs free energy of the reaction, thus it is only natural to test the MTS expression in that context. Data coming from the experiments of Yan Rafrafi was revisited [102]. Seven experiments were carried out using different sources of inoculum. The data from the experiment corresponding to the anaerobic sludge inoculum without heat treatment was used in a continuous stirred tank reactor, with a working volume of 1.5 L. pH was controlled at 5.5. A batch period of 24h after inoculation took place before continuous operation. The hydraulic retention time (HRT) was fixed at 6h. The reactor was heated at 37° C and the gas flow rate was monitored using an electronic gas volumeter. Experiment lasted 10 days. A glucose solution of $10[g/l]$ was used as sole carbon source. The gas phased was initially composed solely of nitrogen gas, and after less than a day of operation it was composed of 65% H_2 and 35 % CO_2 at atmospheric pressure. Two species were

present: *Clostridium pasteurianum* and *Bacillus racemilacticus*, which were identified via a CE-SSCP fingerprinting method. The following possible catabolic reactions were proposed by Rafrafi in the same article.

The catabolism of *Clostridium* is supposed to be a linear combination of the two following reactions:

$$\mathrm{C_6H_{12}O_6 + 4\,H_2O \longrightarrow 2\,C_2H_3O_2^- + 4\,H^+ + 2\,HCO_3^- + 4\,H_2} \tag{5.17}$$

$$\mathrm{C_6H_{12}O_6 + 2\,H_2O \longrightarrow C_4H_7O_2^- + 3\,H^+ + 2\,HCO_3^- + 2\,H_2} \tag{5.18}$$

The coupling was chosen *a priori*, such that the proportion of acetate and butyrate observed in the steady state is attained. Which gave the following catabolism:

$$\mathrm{C_6H_{12}O_6 + 2.5\,H_2O \longrightarrow 0.5\,C_2H_3O_2^- + 0.75\,C_4H_7O_2^- + 3.25\,H^+ + 2\,HCO_3^- + 2.5\,H_2} \tag{5.19}$$

For *Bacillus* that lactate pathway was assumed:

$$\mathrm{C_6H_{12}O_6 \longrightarrow 2\,C_3H_5O_3^- + 2\,H^+} \tag{5.20}$$

The synthesis reaction, representing anabolism, is modelled as follows (for both):

$$\mathrm{0.2\,NH_4^+ + 0.175\,C_6H_{12}O_6 \longrightarrow 0.4\,H_2O + 0.05\,HCO_3^- + 0.25\,H^+ + CH_{1.8}O_{0.5}N_{0.2}} \tag{5.21}$$

Note from table 5.3, the activities from molecules and which are chosen to be variables of the model, as usual, let $x = (x_1, x_2)$ and $s = (s_1, s_2, s_3, s_4, s_5)$. Let $v_{cat,1}$, $v_{cat,1}$, and v_{an} be the vectors representing reactions (5.19), (5.20), and (5.21), respectively. Let $Q_{cat,1}(s)$, $Q_{cat,2}(s)$, and $Q_{an}(s)$ be the reaction quotients of each reaction, note their dependence

Name	Formula	$G_f^0\,[kJ \cdot mol^{-1}]$	Variable $[mol/l]$	$a_i\,[mol \cdot l^{-1}]$
Glucose	$C_6H_{12}O_6$	-917.2	s_1	s_1
Acetate	$C_2H_3O_3{}^-$	-369.4	s_2	s_2
Butyrate	$C_4H_7O_2{}^-$	-352.6	s_3	s_3
Lactate	$C_3H_5O_3{}^-$	-517.1	s_4	s_4
Hydrogen	H_2	0	s_5	0.65
Ammonium	$NH_4{}^+$	-	-	0.05/18
Proton	H^+	0	-	$10^{-5.5}$
Water	H_2O	-237.2	-	1
Carbonate	$HCO_3{}^-$	-394.4*	-	0.35
Clostridium	$CH_{1.8}O_{0.5}N_{0.2}$	-67	x_1	1
Bacillus	$CH_{1.8}O_{0.5}N_{0.2}$	-67	x_2	1

Table 5.3: Molecules, Gibbs free energies of formation, chemical activities, and variables of the model. *Gibbs free energy of formation of CO_2.

on s. The Gibbs free energy of each reaction can be written as:

$$\Delta G_{cat,1}(s) = \Delta G^0_{cat,1} + RT\ln(Q_{cat,1}(s)) \tag{5.22}$$

$$\Delta G_{cat,2}(s) = \Delta G^0_{cat,2} + RT\ln(Q_{cat,2}(s)) \tag{5.23}$$

$$\Delta G_{an}(s) = \Delta G^0_{an} + RT\ln(Q_{an}(s)) \tag{5.24}$$

Let $\Delta G_{dis,1}$, and $\Delta G_{dis,2}$ be the dissipated energy of each metabolism. The coupling coefficient of each metabolism is then calculated as

$$\lambda_i(s) = -\frac{\Delta G_{dis,i} + \Delta G_{an,i}(s)}{\Delta G_{cat,i}(s)} \quad i \in 1,2 \tag{5.25}$$

the metabolic reaction of each species is then $v_{met,i}(s) = \lambda_i(s)v_{cat,i} + v_{an}$. Let Π be the projection into the coordinates of molecules corresponding to variables $(s_1, s_2, s_4, s_4, s_5)$. So one can write:

$$Y(s) = \begin{bmatrix} \Pi v_{met,1}(s) & \Pi v_{met,1}(s) \end{bmatrix} \tag{5.26}$$

Let $\mu_{max,1}$ and $\mu_{max,2}$, the maximum growth rates of x_1 and x_2, and $V_{h,1}$ and $V_{h,2}$ their respective harvest volumes. Let $(v_{met,i}(s))_{s_1}$ be the coordinate corresponding to the glucose

stoichiometric coefficient of metabolism i. Then the growth rate is written as:

$$\mu_i(s) = \mu_{max,i} \exp\left(\frac{(v_{met,i}(s))_{s_1}}{V_{h,i} s_1}\right) \quad i \in \{1,2\} \tag{5.27}$$

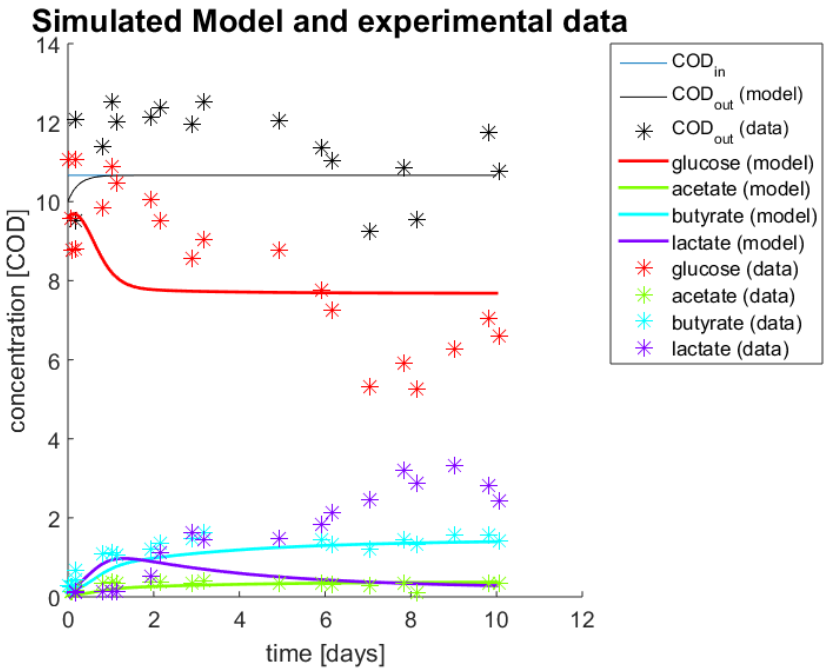
with $\mu(s) = \begin{bmatrix}\mu_1(s) & \mu_2(s)\end{bmatrix}^\top$. Let (once again, abusing of notation) $s_{in} = \begin{bmatrix}s_{in} & 0 & 0 & 0 & 0\end{bmatrix}^\top$ From all the above one can write the the chemostat model as:

$$\begin{aligned}\dot{x} &= \operatorname{diag}(\mu(s) - D_{2\times 1})x \\ \dot{s} &= (s_{in} - s)D + Y(s)\operatorname{diag}(\mu(s))x\end{aligned} \tag{5.28}$$

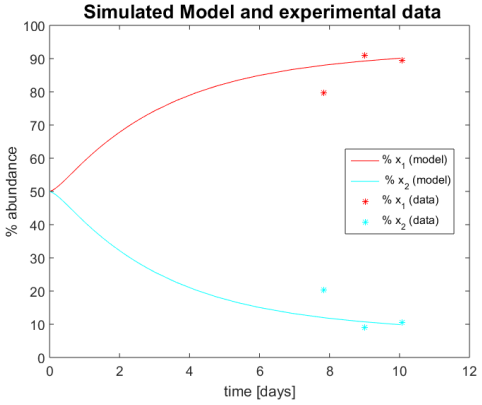
In the following some simulations are presented for 5 sets of parameters shown in table 5.4. The idea is to observe the qualitative behaviour of the model over reasonable values of its output compared to the experiment. For the first set, the dissipated energy corresponds to the value suggested for glucose as the carbon source [64]. The μ_{max} values are taken from other study of hydrogen production [1]. Only the V_h value was used to "fit" the model. In sets 2, 3, and 4, just one parameter was changed with respect to set 1. No optimization procedure was used, because the reactor was not perturbed neither in its dilution rate, nor in its input glucose, so identifiability is probably low. The former is shown with Set 5, the parameters are completely different, particularly the $\Delta G_{dis,1}$ was estimated in order to have the same hydrogen yield as the experiment ($1.9\,[mol\mathrm{H}_2/mol\mathrm{C}_6\mathrm{H}_{12}\mathrm{O}_6]$), this implies $\lambda = 0.76$. The activity used to estimate ΔG_{dis} in that case was the observed data at the steady state.

Parameters	Set 1	Set 2	Set 3	Set 4	set 5
$\Delta G_{dis,1}[kJ/C-mol\,Biomass]$	236	236	236	236	**198**
$\Delta G_{dis,2}[kJ/C-mol\,Biomass]$	236	**210**	236	236	**170**
$\mu_{max,1}[day]^{-1}$	35	35	35	35	**20**
$\mu_{max,2}[day]^{-1}$	35	35	35	35	**25**
$V_{h,1}[mol\,Biomass/l]$	13	13	13	13	**17.5**
$V_{h,2}[mol\,Biomass/l]$	14.5	14.5	**15**	14.5	**15**
$x_1(t=0)[mol/l]$	10^{-3}	10^{-3}	10^{-3}	$\mathbf{10^{-5}}$	$\mathbf{5}\cdot 10^{-4}$
$x_2(t=0)[mol/l]$	10^{-3}	10^{-3}	10^{-3}	10^{-3}	$\mathbf{5}\cdot 10^{-4}$

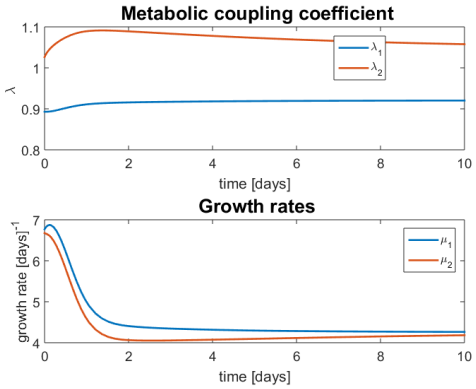
Table 5.4: 5 sets of parameters for model (5.28). Bold numbers indicate a difference with respect to set 1.



(a) Metabolites resulting from model simulation and data.



(b) Proportion of OTU measured by molecular fingerprints.



(c) Coupling coefficients $\lambda_i(s(t))$ and growth rates $\mu_i(s(t))$ through time.

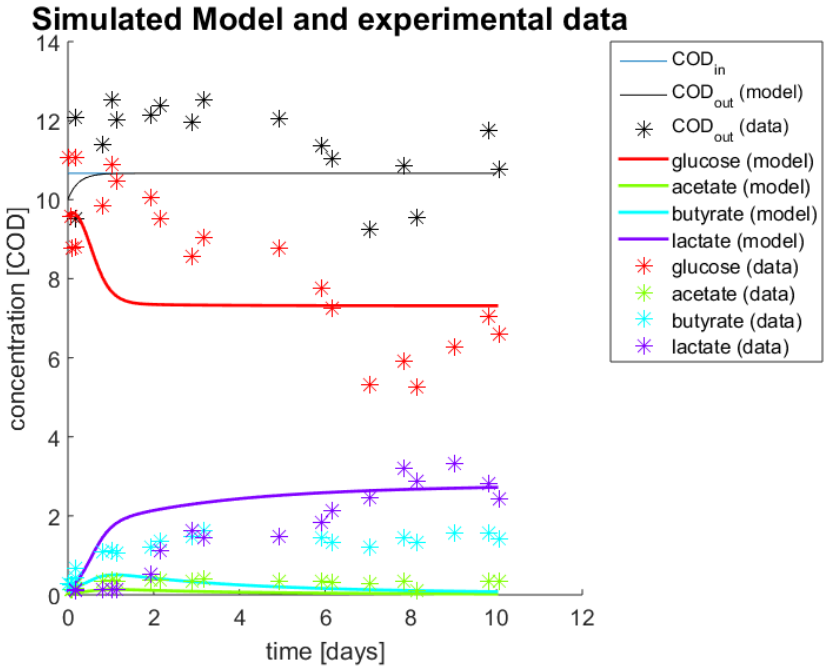
Figure 5.4: Model simulation for parameter set 1. Note how the growth rate converge, implying that coexistence was attained.

The model seems to replicate the results of coexistence found in [40], to estimate if coexistence is achieved, longer simulation times were used to observe the species concentration, in all cases shown here coexistence occurred, but for discussion only the simulation corresponding to the experiments duration are shown. Evidently, washout of one of the two species can be achieved by, for example, doubling the μ_{max} of one of the species (not shown in here). The growth rates are shown in all cases to converge to D. In all figures, the COD balance of the model was taken as the sum of the COD equivalent of vector $s(t)$, the COD_{in} represents the glucose input concentration, finally the

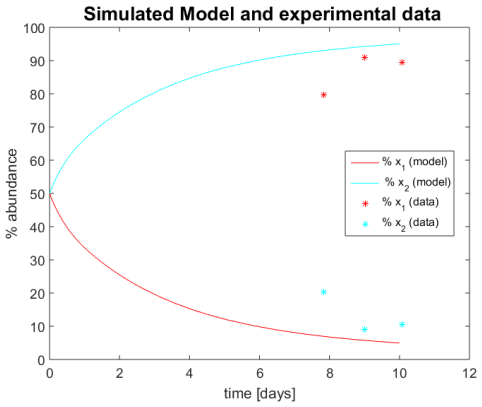
$COD_{out}(data)$ was taken only as the sum of s_1 to s_4 (remaining glucose + VFA). It seems from simulations that one can not replicate at the same time the proportion of x_1 and x_2 and the lactate concentration (compare figures 5.4 and 5.5 where $\Delta G_{dis,2}$ was lowered). In figure 5.6 a different initial point was used (x_1 initial concentration was divided by a hundred), and the system appears to reach the same equilibria as in 5.4.

The behaviour of the glucose transient concentrations could neither be captured in any of the simulations done while studying this system. However, given the 20 % error of the COD mass balance, the simulations can be considered acceptable in terms of capturing the qualitative dynamics of the system. The metabolic coupling coefficient appears to be dynamic in all five simulations, particularly in figure 5.8. Under the hypothesis that the dissipated energy of a cell is constant this would imply a very quick adaptability from the OTU.

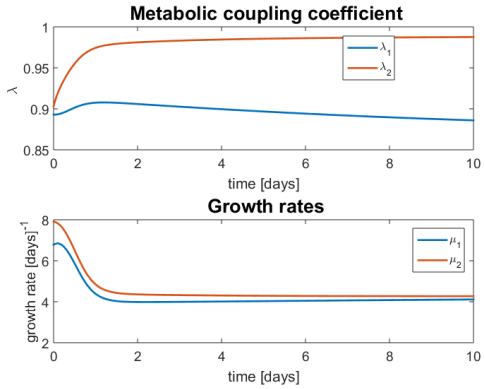
It seems that thermodynamics corrections in anaerobic systems are necessary in order to explain the ecological dynamics. The growth function dependecy on catabolic products is essential to achieve coexistence, so the CEP can no longer hold.



(a) Metabolites resulting from model simulation and data.

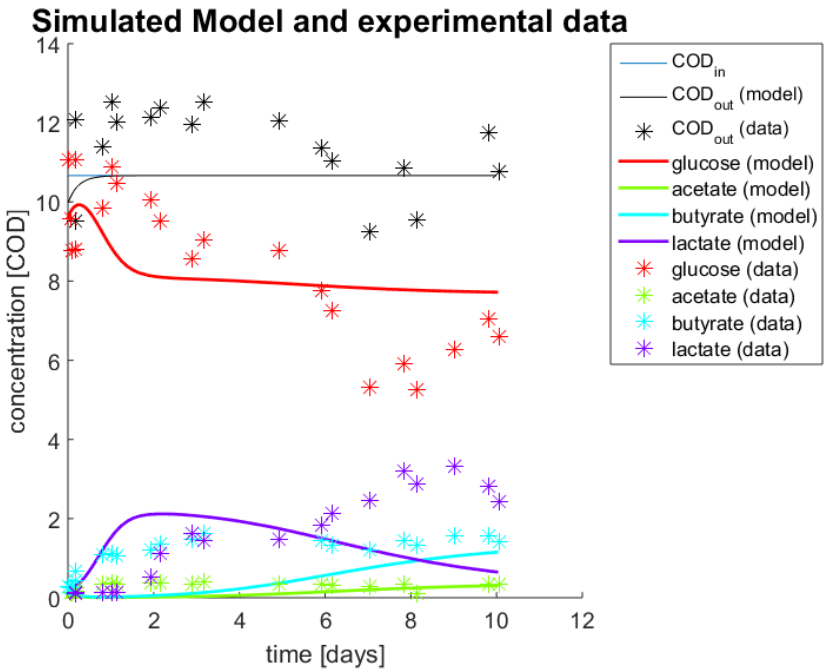


(b) Proportion of OTU measured by molecular fingerprints.

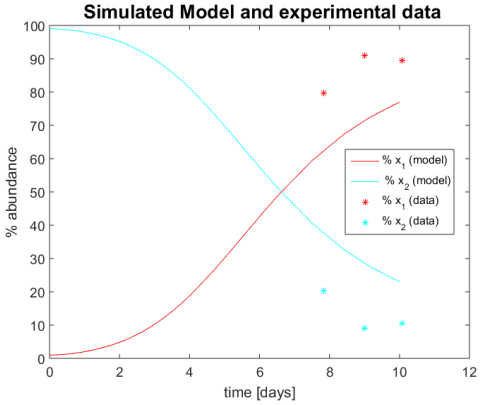


(c) Coupling coefficients $\lambda_i(s(t))$ and growth rates $\mu_i(s(t))$ through time.

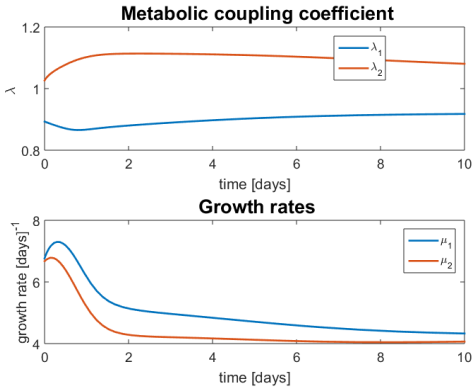
Figure 5.5: Model simulation for parameter set 2. $\Delta G_{dis,2}$ was lower than set 1, thus the yield of glucose consumed per biomass for x_2 became lower, consequently its "K_s" (as in the identification shown at the beginning of the chapter) also diminished, giving x_2 the upper hand in substrate acquisition.



(a) Metabolites resulting from model simulation and data.

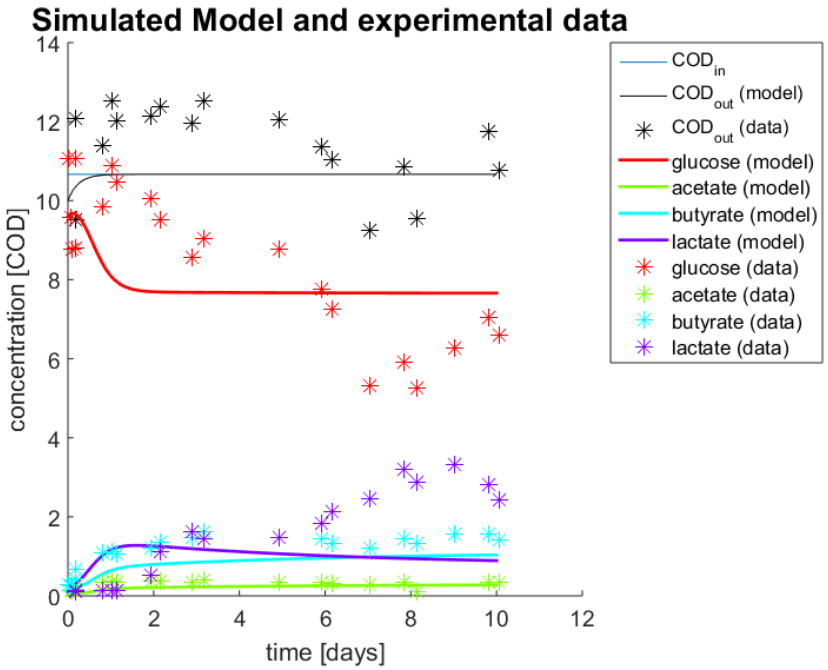


(b) Proportion of OTU measured by molecular fingerprints.

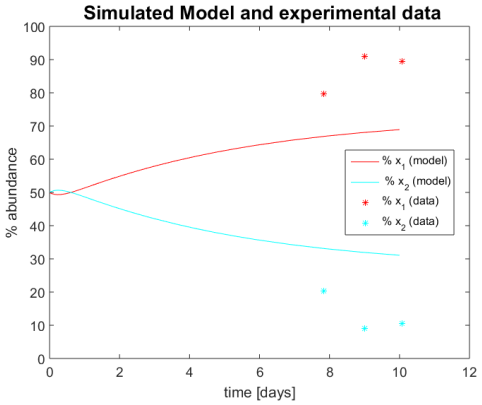


(c) Coupling coefficients $\lambda_i(s(t))$ and growth rates $\mu_i(s(t))$ through time.

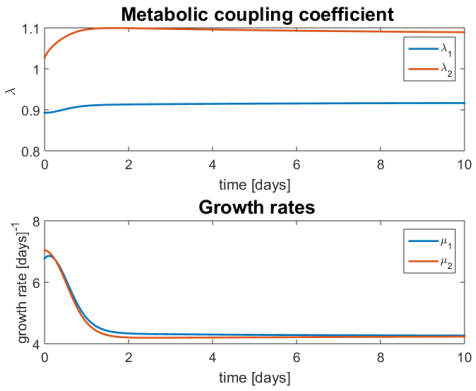
Figure 5.6: Model simulation for parameter set 3. The initial concentration for x_1 was divided by a hundred, but the system still arrives to a similar steady state as in case 1.



(a) Metabolites resulting from model simulation and data.

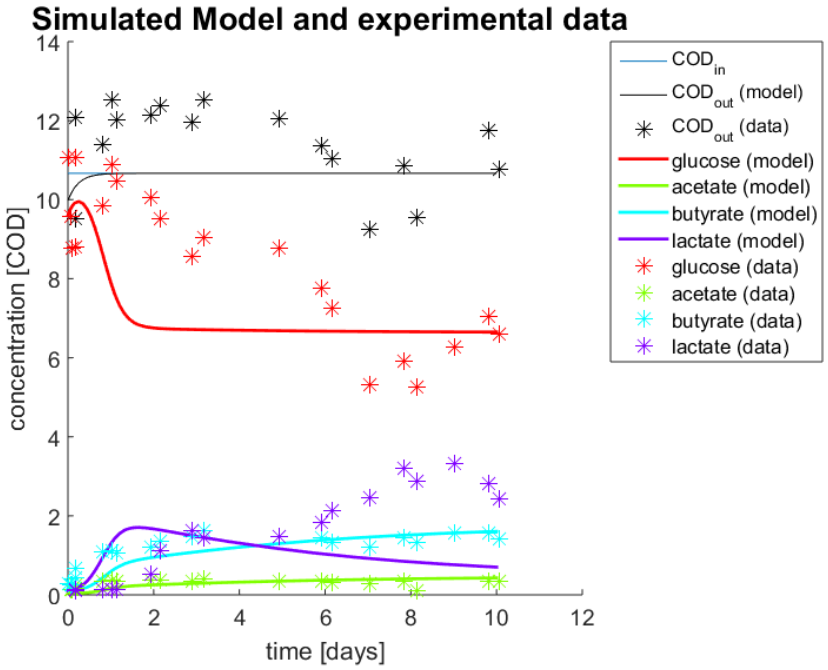


(b) Proportion of OTU measured by molecular fingerprints.

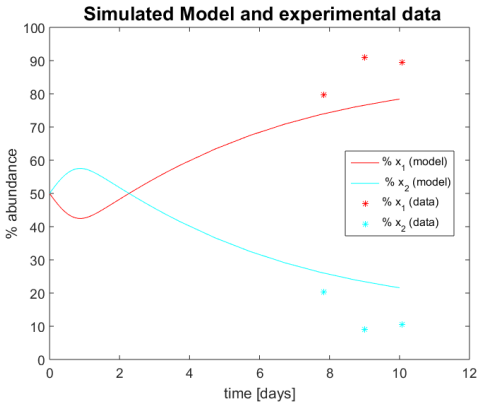


(c) Coupling coefficients $\lambda_i(s(t))$ and growth rates $\mu_i(s(t))$ through time.

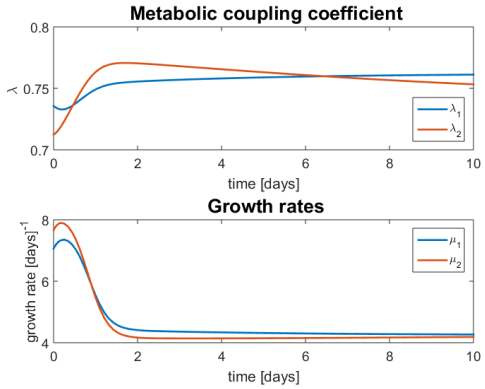
Figure 5.7: Model simulation for parameter set 4. $V_{h,2}$ was lowered, and consequently "K_s", the system arrives at a coexistence equilibrium as well, but the proportion of x_2 augmented with respect case 1.



(a) Metabolites resulting from model simulation and data.



(b) Proportion of OTU measured by molecular fingerprints.



(c) Coupling coefficients $\lambda_i(s(t))$ and growth rates $\mu_i(s(t))$ through time.

Figure 5.8: Model simulation for parameter set 5. $\Delta G_{dis,1}$ was calculated to fit the hydrogen yield experimentally found. The other parameters were tinkered to more or less fit data. The models parameters probably can not be identified to a great degree of accuracy.

Chapter 6

Conclusions and Perspectives

In the previous pages some mathematical models on the community dynamics in chemostat settings were developed and discussed. The main aspects that one might say have contributed to the current state of modelling in bioprocess are three: 1) A method to assign functionality, or more precisely, a metabolic reaction to OTU present in a reactor 2) A framework to study the interactions of OTU that can be directly integrated in already existing mass-balance models and 3) Theoretical connections between growth rates of microbial development under substrate limitation and the energy gradients available in the environment.

The first contribution, as already mentioned, was motivated by the challenges posed by Widder *et al.* [138]. Even though a CBM formulation was not used, the method is robust, in the sense that it does not need expert knowledge on the particular metabolic pathways of each microorganism, but rather general biological knowledge on the catabolic reactions that are known to take place in a particular environment. A very important part of the classification method relies on the reference yield used. For assessing the yield (that is known to vary from one reactor to another), a previous analysis based on the Gibbs energy dissipation method was coupled to asymptotic observers.

Developing an asymptotic observer that takes into account the activity corrections of the stoichiometric matrix requires finding a new invariant. Note that the reaction invariant as defined in this work would imply, $\dot{z} = \dot{s} - \dot{Y}(s)x - Y(s)\dot{x} = (s_{in} - z)D - \dot{Y}(s)x$,

so the deduction is not as straightforward as before.

The framework developed to integrate interactions as drivers of a reactor outcome was originally inspired from a gLV model, both of the models exposed what can be gained from combining population-based models as used in ecology with functional group based approaches as used in bioengineering. However, gLV models suffer from quadratic growth of parameters that clearly renders them of low potential application, even though some successful cases do exist. Contrarily the framework presented grows linearly in terms of parameters per OTU (the weights on the control), and also grows linearly in terms of computing time per OTU. How to go from studying interactions to using them for prediction is clearly the next step. In that sense, using the control terms to fit a new particular growth function seems a good idea (instead of trying to fit the whole system of differential equations again), and searching for functions with a reduced number of parameters (as opposed to linear pairwise interactions) is also a good guideline.

In these times, when data-driven approaches are on the rise, our capacity to understand phenomena is being relegated to the performance of prediction of algorithms. "Do not try to understand, compute," seems to be the general paradigm. It is worth asking, then, the value in the task of deriving microbial growth expressions from theoretical enquiries. From the author's point of view it is not only a useful but a necessary exercise. In a certain sense, one cannot advance without the other, and as the power of algorithms predictability increases, the ability to understand the causality of the input-output of the corresponding algorithm decreases. This just creates a new scenario in which theory should include these predictions as new observations.

Moreover, data is not born out of thin air. It is theory that pushes for the search and development of measurement instruments to gather information. The chemostat did not exist before microbial growth; however, as a tool to measure it, both of them became entangled, and it is hard to discuss microbial growth without using the affinity concept. Even though the meaning of the affinity constant has not yet been unveiled (if it ever is), it has been measured in thousands of experiments on different settings. What does it mean to measure a constant without a clear interpretation? A very practical and

concrete answer to that question is that a function has been fitted to data in a particular condition. To a certain extent, the meaning comes from the model and the theory itself. Once a model works, we strive to adjust the model to different situations other than the one in which it originally worked. When these adjustments become a list of cases (all of the existing growth expressions), then, perhaps a new theory is required that wraps up all of the above cases.

Certain theoretical consequences of the MTS model seem to be reflected in observed phenomena. The concept of harvest volume is an intellectual edifice that intuitively encapsulates the notion of density dependence and affinity, so it is no surprise that its approximations appear to be in line with some documented experimental growth rates. Since it also takes into account energetic principles governing cellular growth it forces a link between stoichiometry and the growth rate which came to be around Monod's affinity constant.

The need for a thermodynamic framework in order to gain insight on the trade-offs of classical kinetic parameters is sought for. In here the affinity K_s was shown to have a theoretical relationship to the yield, any insight on μ_{max} still remains a mystery. The recent findings in an upper limit of gibbs energy dissipation rate could be tested in a chemostat setting [90]. A good starting point would be to compile data coming from chemostat experiments where dilution rate D was varied, and calculate $D\Delta G_{dis}$ in order to obtain an approximation of the dissipated energy rate.

Appendix A

Appendix

A.1 Technical Framework

Theorem 2. *Consider two sets of molecules $\mathcal{C}_1$ and $\mathcal{C}_2$, with n_1 and n_2 elements, A_1 and A_2 atomic matrices, and Gf_1 and Gf_2 formation energies, respectively. Let $\mathcal{C} = \mathcal{C}_1 \cup \mathcal{C}_2$ be the set of all molecules, with n elements, atomic matrix A, and formation energy G_f. Let $\mathcal{J}_i$ the set of indices corresponding to molecules $\mathcal{C}_i$ in set $\mathcal{C}$, and define the matrix $M_i \in \mathbb{R}^{n \times n_i}$ $(M_i)_{\bullet j} = e_j$, $j \in \mathcal{J}_i$, with e_j canonical vector j of $\mathbb{R}^n$. Then $\ker(A_i) \cong \ker(A) \cap \mathrm{Im}(M_i)$. Using this identification for $v_1 \in \ker(A_1)$ and $v_2 \in \ker(A_2)$ we can write $v = v_1 + v_2$ without ambiguity. Furthermore $Gf = M_1 Gf_1 + M_2 Gf_2$ and $\Delta G_v = v \cdot Gf = v_1 \cdot Gf_1 + v_2 \cdot Gf_2$.*

Proof. Basic matrix manipulation yields $AM_i = A_i$. Note that $\pi_i : x \in \ker(A_i) \to M_i x \in \ker(A) \cap \mathrm{Im}(M_i)$, defines a surjective homomorphism, and $\ker(\pi_i) = \{x : M_i x = 0\} = \{0\}$, since M_i is composed of linearly independent columns, thus π_i is an isomorphism. □

Note that $\mathrm{Im}\, M_i = \{x \in \mathbb{R}^n, x_j = 0, \forall j \notin \mathcal{J}_i\}$. Basically the isomorphism injects the elements $v \in \ker(A_i)$ into the coordinates of the corresponding molecules of a vector of $\mathbb{R}^n$, with zeros in the other coordinates.

A.2 On the Growth Rates of a Microbial Community

A.2.1 Proofs of Properties of the System

Lemma 4. *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* (4.3) *satisfy the following inequalities:*

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

$$\sum_{i \in G_2} \frac{1}{y_i} x_i + s_1 + s_2 \leq M_2 \tag{A.2}$$

$$s_1 + s_2 + s_2 \leq M_3 \tag{A.3}$$

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

$$\dot{w} = D(\bar{s} - w); \quad w(0) = \sum_{i \in G_1} \tfrac{1}{y_i} x_i(0) + s_1(0) \tag{A.4}$$

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 (A.4) 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

$$\dot{z_1} = D(s_{in} - z_1); \quad z_1(0) = \sum_{i \in G_1} \tfrac{1}{y_i} x_i(0) + s_1(0) \tag{A.5}$$

From a comparison lemma (see chapter 3 [60]), the solution of (A.5) 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 5. *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 (4.3) with growth rates given by (4.6), 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 2.

Note that for any $i \in [n]$, if $x_i = 0$ then $\dot{x_i} = 0$, therefore all solutions of system (4.3) 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 (2) 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 $M := \dfrac{\bar{k}}{M_1 + M_2}$, 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:

$$(1 + A_{i\bullet}x) \geq 0 \quad \forall i \in [n] \tag{A.6}$$

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 (A.6) 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 2 that the sum of positive elements is bounded, therefore each element is bounded. □

A.2.2 Deduction of Equilibrium points

Recall that $f_i(s) = \begin{cases} \bar{\mu}_i \dfrac{s_1}{K_i + s_1} & \forall i \in G_1 \\ \bar{\mu}_i \dfrac{s_2}{K_i + s_2} & \forall i \in G_2 \end{cases}$, then $\mu(x,s) = \text{diag}(f(s))(1_{n\times 1} + Ax)$ thus, system (4.3) is rewritten as follows.

$$\dot{x} = \text{diag}(\mu(x,s) - D_{n\times 1})x \tag{A.7}$$

$$\dot{s_1} = (s_{in} - s_1)D + Y_{1\bullet}\,\text{diag}(\mu(x,s))x \tag{A.8}$$

$$\dot{s_2} = -s_2 D + Y_{2\bullet}\,\text{diag}(\mu(x,s))x \tag{A.9}$$

$$\dot{s_3} = -s_3 D + Y_{3\bullet}\,\text{diag}(\mu(x,s))x \tag{A.10}$$

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:

$$x^{eq} = Mx^{act};\ f^{act}(s) = M^\top f(s);\ \mu^{act}(x,s) = M^\top \mu(x,s);\ Y^{act} = YM;\ M^\top M = I_{n^{act}} \tag{A.11}$$

From equation (A.7) equilibrium points satisfy:

$$x_i = 0 \vee (\text{diag}(f(s))(1_{n\times 1} + Ax) - D_{n\times 1})_i = 0 \quad \forall i \in [n] \tag{A.12}$$

$$\Rightarrow f_i(s)(1 + A_{i\bullet}x) - D = 0 \quad \forall i \in [n] \setminus \mathcal{J} \tag{A.13}$$

note that $\forall j \in \mathcal{J},\ x_j^{eq} = 0$ therefore the coefficients a_{ij} play no role in equation (A.13) and one can rewrite equation (A.13) as

$$\mu^{act}(x,s) = \mathrm{diag}(f^{act}(s))(1_{n^{act}\times 1} + A^{act}x^{act}) = D_{n^{act}\times 1} \tag{A.14}$$

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

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

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 (A.8), (A.9), and (A.10) which gives:

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

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 (A.8) yields

$$(s_{in} - s_1)D + Y_{1\bullet}^{act}\,\mathrm{diag}(\mu^{act}(x,s))x^{act} = 0 \tag{A.17}$$

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 (A.15) in equation (A.8) reads as follows:

$$(s_{in} - s_1) + Y_{1\bullet}^{act}(A^{act})^{-1}\underbrace{M^{\top}M}_{I_{n^{act}}}(\mathrm{diag}(f^{act}(s))^{-1}D_{n^{act}\times 1} - 1_{n^{act}\times 1}) = 0 \tag{A.18}$$

$$s_{in} - s_1 + \sum_{i\in G_1} B_i\left(\frac{K_i + s_1}{\bar{\mu}_i s_1}D - 1\right) + \sum_{i\in G_2} B_i\left(\frac{K_i + s_2}{\bar{\mu}_i s_2}D - 1\right) = 0 \quad / \cdot s_1 s_2 \tag{A.19}$$

$$s_2\left(-s_1^2 + s_1\left(s_{in} + \sum_{i\in G_1\cup\in G_2} B_i\left(\frac{D}{\bar{\mu}_i} - 1\right)\right) + \sum_{i\in G_1}\frac{DB_iK_i}{\bar{\mu}_i}\right) = -s_1\sum_{i\in G_2}\frac{DB_iK_i}{\bar{\mu}_i} \tag{A.20}$$

and so we get to formula (4.17) of the main text:

$$s_2 = \frac{s_1}{b_1 s_1^2 + b_2 s_1 + b_3} \tag{A.21}$$

where

$$b_1 = \left(\sum_{i \in G_2} \frac{DB_i K_i}{\bar{\mu}_i} \right)^{-1} \tag{A.22}$$

$$b_2 = -\left(s_{in} + \sum_{i \in G_1 \cup \in G_2} B_i \left(\frac{D}{\bar{\mu}_i} - 1 \right) \right) \left(\sum_{i \in G_2} \frac{DB_i K_i}{\bar{\mu}_i} \right)^{-1} \tag{A.23}$$

$$b_3 = -\sum_{i \in G_1} \frac{DB_i K_i}{\bar{\mu}_i} \left(\sum_{i \in G_2} \frac{DB_i K_i}{\bar{\mu}_i} \right)^{-1} \tag{A.24}$$

$$\tag{A.25}$$

The same computations must be done with equation (A.9), which is structurally very similar to (A.8). By replacing $x^{eq} = Mx^{act}$ in equation (A.9) yields

$$-s_2 + Y_{2\bullet}^{act} (A^{act})^{-1} (\mathrm{diag}(f^{act}(s))^{-1} D_{n^{act} \times 1} - 1_{n^{act} \times 1}) = 0 \tag{A.26}$$

It is again useful to define:

$$C := Y_{2\bullet}^{act} (A^{act})^{-1} M^{\top}. \tag{A.27}$$

$$-s_2 + Y_{2\bullet}^{act} (A^{act})^{-1} \underbrace{M^{\top} M}_{I_{n^{act}}} (\mathrm{diag}(f^{act}(s))^{-1} D_{n^{act} \times 1} - 1_{n^{act} \times 1}) = 0 \tag{A.28}$$

$$-s_2 + \sum_{i \in G_1} DC_i \frac{K_i + s_1}{\bar{\mu}_i s_1} + \sum_{i \in G_2} DC_i \frac{K_i + s_2}{\bar{\mu}_i s_2} - \sum_{i \in G_1 \cup G_2} C_i = 0 \quad / \cdot s_1 s_2 \tag{A.29}$$

$$s_1 \left(-s_2^2 + s_2 \sum_{i \in G_1 \cup G_2} C_i \left(\frac{D}{\bar{\mu}_i} - 1 \right) + \sum_{i \in G_2} \frac{DC_i K_i}{\bar{\mu}_i} \right) = -s_2 \sum_{i \in G_1} \frac{DC_i K_i}{\bar{\mu}_i} \tag{A.30}$$

and one gets to expression:

$$s_1 = \frac{s_2}{c_1 s_2^2 + c_2 s_2 + c_3} \tag{A.31}$$

where

$$c_1 = \left(\sum_{i \in G_1} \frac{DC_i K_i}{\bar{\mu}_i} \right)^{-1} \tag{A.32}$$

$$c_2 = - \sum_{i \in G_1 \cup \in G_2} C_i \left(\frac{D}{\bar{\mu}_i} - 1 \right) \left(\sum_{i \in G_1} \frac{DC_i K_i}{\bar{\mu}_i} \right)^{-1} \tag{A.33}$$

$$c_3 = - \sum_{i \in G_2} \frac{DC_i K_i}{\bar{\mu}_i} \left(\sum_{i \in G_1} \frac{DC_i K_i}{\bar{\mu}_i} \right)^{-1} \tag{A.34}$$

$$\tag{A.35}$$

Then by replacing (A.21) in equation (A.31), and after some reordering, one gets a fifth degree polynomial for s_1.

$$s_1 \left(c_1 \left(\frac{s_1}{b_1 s_1^2 + b_2 s_1 + b_3} \right)^2 + c_2 \left(\frac{s_1}{b_1 s_1^2 + b_2 s_1 + b_3} \right) + c_3 \right) = \frac{s_1}{b_1 s_1^2 + b_2 s_1 + b_3} \quad / \cdot \left(b_1 s_1^2 + b_2 s_1 + b_3 \right)^2 \tag{A.36}$$

$$a_4 s_1^4 + a_3 s_1^3 + a_2 s_1^2 + a_1 s_1 + a_0 = 0 \quad \vee \quad s_1 = 0 \tag{A.37}$$

where

$$a_0 = b_3 + c_3 b_3^2 \tag{A.38}$$

$$a_1 = c_2 b_3 + 2 c_3 b_3 b_2 - b_2 \tag{A.39}$$

$$a_2 = c_1 + c_2 b_2 + c_3 (b_2^2 + 2 b_1 b_3) - b_1 \tag{A.40}$$

$$a_3 = c_2 b_1 + 2 c_3 b_1 b_2 \tag{A.41}$$

$$a_4 = c_3 b_1^2 \tag{A.42}$$

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 (A.7), one obtains the same computations as in the previous section, so retake equation (A.19)

$$s_{in} - s_1 + \sum_{i \in G_1} B_i \left(\frac{K_i + s_1}{\bar{\mu}_i s_1} D - 1 \right) + \sum_{i \in G_2} B_i \left(\frac{K_i + s_2}{\bar{\mu}_i s_2} D - 1 \right) = 0 \tag{A.19}$$

Since $G_1 \subset \mathcal{J}$ then $B_i = 0 \forall i \in G_2$ thus (A.19) becomes:

$$s_{in} - s_1 + \sum_{i \in G_1} B_i \left(\frac{K_i + s_1}{\bar{\mu}_i s_1} D - 1 \right) = 0 \quad / \cdot s_1 \tag{A.43}$$

$$s_1 s_{in} - s_1^2 + \sum_{i \in G_1} B_i \frac{D B_i K_i}{\bar{\mu}_i} + s_1 \sum_{i \in G_1} B_i \left(\frac{D}{\bar{\mu}_i} - 1 \right) = 0 \tag{A.44}$$

And so a quadratic equation for s_1 is obtained.

$$a_2' s_1^2 + a_1' s_1 + a_0' = 0 \tag{4.19}$$

where

$$a_2' = -1 \tag{A.45}$$

$$a_1' = s_{in} + \sum_{i \in G_1} B_i \left(\frac{D}{\bar{\mu}_i} - 1 \right) \tag{A.46}$$

$$a_0' = \sum_{i \in G_1} B_i \frac{D B_i K_i}{\bar{\mu}_i} \tag{A.47}$$

A.2.3 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 $[x - a]$ [20]. 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:

$$h'(x, y)[a, b] = h_1'(\text{diag}(x), y) \circ (\text{diag}'(x),\ I_n)[a, b] \tag{A.48}$$

$$= h_1'(\text{diag}(x), y)[\text{diag}(a), b] \tag{A.49}$$

$$= \text{diag}(a)y + \text{diag}(x)b \tag{A.50}$$

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

$$\dot{x} = g_1(x, s) = \text{diag}(\mu(x, s) - D_{n\times 1})x \tag{A.51}$$

$$\dot{s} = g_2(x, s) = \left(\begin{bmatrix} s_{in} & 0 & 0 \end{bmatrix}^\top - s \right) D + Y \,\text{diag}(\mu(x, s))x \tag{A.52}$$

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}':

$$g_{1s}'(x)[e_j] = h'(\mu_s(x) - D_{n\times 1}, x) \circ (\mu_s'(x), I_n)[e_j, e_j] \tag{A.53}$$

$$= h'(\mu_s(x) - D_{n\times 1}, x)[\text{diag}(f(s))Ae_j, e_j] \tag{A.54}$$

$$= \text{diag}(f(s))\,\text{diag}(x)A_{\bullet j} + \text{diag}(\mu_s(x) - D_{n\times 1})e_j \tag{A.55}$$

$$\Rightarrow g_{1s}'(x) = \text{diag}(f(s))\,\text{diag}(x)A + \text{diag}(\mu_s(x) - D_{n\times 1}) \tag{A.56}$$

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

$$g'_{2s}(x)[e_j] = Yh'(\mu_s(x), x) \circ (\mu'_s(x), I_n)[e_j, e_j] \tag{A.57}$$

$$= Y\left(\operatorname{diag}(f(s)) \operatorname{diag}(x) A_{\bullet j} + \operatorname{diag}(\mu_s(x)) e_j\right) \tag{A.58}$$

$$\Rightarrow g'_{2s}(x) = Y(\operatorname{diag}(f(s)) \operatorname{diag}(x) A + \operatorname{diag}(\mu_s(x))) \tag{A.59}$$

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

$$f(s) = \begin{pmatrix} f_{G_1}(s_1) \\ f_{G_2}(s_2) \end{pmatrix} \tag{A.60}$$

One can see then that:

$$f'(s) = \begin{bmatrix} f'_{G_1}(s_1) & 0_{n_1 \times 2} \\ 0_{n_2 \times 1} & f'_{G_2}(s_2) \quad 0_{n_2 \times 1} \end{bmatrix} \tag{A.61}$$

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

$$g'_{1x}(s)[e_j] = h'(\mu_x(s) - D_{n \times 1}, x) \circ (\mu'_x(s), 0)[e_j, e_j] \tag{A.62}$$

$$g'_{1x}(s)[e_j] = h'(\mu_x(s) - D_{n \times 1}, x)[\mu'_x(s) e_j, 0] \tag{A.63}$$

$$g'_{1x}(s)[e_j] = \operatorname{diag}(x) \mu'_x(s) e_j \tag{A.64}$$

$$\Rightarrow g'_{1x} = \operatorname{diag}(x) \operatorname{diag}(1 + Ax) f'(s) \tag{A.65}$$

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}:

$$g'_{2x}(s)[e_j] = -DI_n e_j + Yh'(\mu_x(s), x) \circ (\mu'_x(s), 0)[e_j, e_j] \tag{A.66}$$

$$g'_{2x}(s)[e_j] = -DI_n e_j + Yh'(\mu_x(s), x)[\mu'_x(s)e_j, 0] \tag{A.67}$$

$$g'_{2x}(s)[e_j] = -DI_n e_j + Y \operatorname{diag}(x)\mu'_x(s)e_j \tag{A.68}$$

$$\Rightarrow g'_{2x}(s) = -DI_n + Y \operatorname{diag}(x) \operatorname{diag}(1 + Ax)f'(s) \tag{A.69}$$

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:

$$J(x, s) = \begin{bmatrix} g'_{1s}(x) & g'_{1x}(s) \\ g'_{2s}(x) & g'_{2x}(s) \end{bmatrix} \tag{A.70}$$

A.2.4 Tracking Problem reformulation and details

For applying the methods developed in [22]. 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 (4.25). The system may be rewritten then as:

$$\begin{aligned}
\dot{x_i} &= (f_i(s)(1 + v_i(t)) - D)\, x_i \quad \forall i \in G_1 \\
\dot{x_i} &= (f_i(s)(1 + v_i(t)) - D)\, x_i \quad \forall i \in G_2 \\
\dot{s_1} &= s_1 \left(\frac{s_{in}}{s_1} - 1\right) D + \sum_{i \in G_1} y_{s_1/x_i} f_i(s)(1 + v_i(t)) x_i \\
\dot{s_2} &= -s_2 D + \sum_{\substack{i \in G_1 \cup G_2 \\ n_1 + n_2}} y_{s_2/x_i} f_i(s)(1 + v_i(t)) x_i \\
\dot{s_3} &= -s_3 D + \sum_{i \in G_2} y_{s_3/x_i} f_i(s)(1 + v_i(t)) x_i \\
y(t) &= x(t)
\end{aligned} \tag{A.71}$$

Define:

$$A\left(X\right)=\begin{bmatrix}A_{11}(X) & A_{12}(X)\\ A_{21}(X) & A_{22}(X)\end{bmatrix};\ B\left(X\right)=\begin{bmatrix}B_1(X)\\ B_2(X)\end{bmatrix} \tag{A.72}$$

with

$$A_{11}(X)=\operatorname{diag}(f(s)-D_{n\times 1});\ A_{12}(X)=0_{n\times 3} \tag{A.73}$$

$$A_{21}(X)=Y\operatorname{diag}(f(s));\ A_{22}(X)=\begin{bmatrix}\left(\frac{s_{in}}{s_1}-1\right)D & 0 & 0\\ 0 & -D & 0\\ 0 & 0 & -D\end{bmatrix} \tag{A.74}$$

and

$$B_1(X)=\operatorname{diag}(f(s))\operatorname{diag}(x);\ B_2(X)=Y\operatorname{diag}(f(s))\operatorname{diag}(x);\ C(X)=\begin{bmatrix}\operatorname{diag}(I_n) & 0_{n\times 3}\end{bmatrix} \tag{A.75}$$

Then the system (A.71) can be rewritten as:

$$\dot{X}=A(X)X+B(X)v;\ y=C(X)X \tag{A.76}$$

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

$$J(v)=\left(z(t_f)-C(X)X(t_f)\right)^{\top}F\left(z(t_f)-C(X)X(t_f)\right) \tag{A.77}$$

$$+\int_{t_0}^{t_f}\left(z(t)-C(X)X(t)\right)^{\top}Q\left(z(t)-C(X)X(t)\right)+v(t)^{\top}Rv(t) \tag{A.78}$$

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 :

$$\dot{X}^{[i]} = A(X^{[i]})X^{[i]} + B(X^{[i]})v^{[i]} \quad i \in \mathbb{N} \tag{A.79}$$

$$y^{[i]} = X^{[i]} \quad i \in \mathbb{N} \tag{A.80}$$

$$X^{[i]}(t_0) = X_0 \quad i \in \mathbb{N} \tag{A.81}$$

and for $i = 0$ define $X^{[0]}(t)$ as the solution of (A.71) with $v(t) \equiv 0$. The control law is given by

$$v^{[i]}(t)_j = \max\left\{-1, \min\left\{0, \left(-R^{-1}B^\top\left(X^{[i-1]}(t)\right)\left(P^{[i]}(t)X^{[i]}(t) - s_f^{[i]}(t)\right)\right)_j\right\}\right\} \forall j \in [n] \tag{A.82}$$

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:

$$\dot{P}^{[i]} = -C^T\left(X^{[i-1]}(t)\right)QC\left(X^{[i-1]}(t)\right) - P^{[i]}A\left(X^{[i-1]}(t)\right) - A^\top\left(X^{[i-1]}(t)\right)P^{[i]} \tag{A.83}$$

$$+ P^{[i]}B\left(X^{[i-1]}(t)\right)R^{-1}B^\top\left(X^{[i-1]}(t)\right)P^{[i]} \tag{A.84}$$

$$P^{[i]}(t_f) = C^\top\left(X^{[i-1]}(t_f)\right)FC\left(X^{[i-1]}(t_f)\right) \tag{A.85}$$

$$\dot{s_f^{[i]}} = -C^\top\left(X^{[i-1]}(t)\right)Qz(t) - \left[A\left(X^{[i-1]}(t)\right) - B\left(X^{[i-1]}(t)\right)R^{-1}B^\top\left(X^{[i-1]}(t)\right)P^{[i]}(t)\right]^\top s_f^{[i]} \tag{A.86}$$

$$s_f^{[i]}(t_f) = C^\top\left(X^{[i-1]}(t_f)\right)Fz(t_f) \tag{A.87}$$

Replacing the matrices of our problem

$$\dot{P}^{[i]}(t) = -\begin{bmatrix} Q & 0_{n\times 3} \\ 0_{3\times n} & 0_{3\times 3} \end{bmatrix} - P^{[i]} A\left(X^{[i-1]}(t)\right) - A^{\top}\left(X^{[i-1]}(t)\right) P^{[i]} \tag{A.88}$$
$$+ P^{[i]} B\left(X^{[i-1]}(t)\right) R^{-1} B^{\top}\left(X^{[i-1]}(t)\right) P^{[i]}$$

$$P^{[i]}(t_f) = \begin{bmatrix} 0_{n\times n} & 0_{n\times 3} \\ 0_{3\times n} & 0_{3\times 3} \end{bmatrix} \tag{A.89}$$

$$\dot{s_f}^{[i]}(t) = -\begin{bmatrix} Qz(t) \\ 0_{3\times 1} \end{bmatrix} - \left[A\left(X^{[i-1]}(t)\right) - B\left(X^{[i-1]}(t)\right) R^{-1} B^{\top}\left(X^{[i-1]}(t)\right) P^{[i]}(t)\right]^{\top} s_f^{[i]} \tag{A.90}$$

$$s_f^{[i]}(t_f) = \begin{bmatrix} 0_{n\times n} & 0_{n\times 3} \end{bmatrix}^{\top} z(t_f) \tag{A.91}$$

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:

$$P^{[i]}(t) = \begin{bmatrix} \tilde{P}^{[i]}(t) & 0_{n\times 3} \\ 0_{3\times n} & 0_{3\times 3} \end{bmatrix};\ s_f^{[i]}(t) = \begin{bmatrix} \tilde{s}_f^{[i]}(t) \\ 0_{1\times 3} \end{bmatrix} \tag{A.92}$$

$$A\left(X^{[i-1]}(t)\right) = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix};\ B\left(X^{[i-1]}(t)\right) = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \tag{A.93}$$

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

$$\dot{\tilde{P}}^{[i]}(t) = -Q - \tilde{P}^{[i]} A_{11} - A_{11}^{\top} \tilde{P}^{[i]} + \tilde{P}^{[i]} B_1 R^{-1} B_1^{\top} \tilde{P}^{[i]};\ \tilde{P}^{[i]}(t_f) = 0 \tag{A.94}$$

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. And therefore, once again, by existence and uniqueness they should be constantly zero. Hence only the diagonal entries should be calculated.

$$\dot{\tilde{P}}_{jj}^{[i]}(t) = -Q_{jj} - 2\left(f_j\left(s^{[i-1]}\right) - D\right)\tilde{P}_{jj}^{[i]} + R^{-1}\left(\tilde{P}_{jj}^{[i]}\right)^2 f_j\left(s^{[i-1]}\right)^2\left(x_j^{[i-1]}\right)^2;\ \tilde{P}_{jj}^{[i]}(t_f) = 0 \tag{A.95}$$

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

$$\dot{\tilde{s}}_f^{[i]}(t) = -z(t) - \left[A_{11} - R^{-1}B_1B_1^\top\tilde{P}^{[i]}(t)\right]^\top\tilde{s}_f^{[i]};\ \tilde{s}_f^{[i]}(t_f) = 0 \tag{A.96}$$

And the control law is given by

$$v^{[i]}(t)_j = \max\left\{-1, \min\left\{0, \left(-R^{-1}B_1^\top\left(X^{[i-1]}(t)\right)\left(\tilde{P}^{[i]}(t)x^{[i]}(t) - \tilde{s}_f^{[i]}(t)\right)\right)_j\right\}\right\}\forall j \in [n] \tag{A.97}$$

They were solved using standard backward numerical integration.

A.3 On Microbial Transition State Theory Implications

Supplementary Material
Insights from Microbial Transition State Theory on Monod's Affinity Constant

Pablo Ugalde-Salas[1] Elie Desmond-Le Quéméner[1] Jérôme Harmand[1] Alain Rapaport[2] Théodore Bouchez[3]

[1]LBE, INRAE, Univ Montpellier, Narbonne, France

[2] MISTEA, INRAE, Univ. Montpellier, Montpellier SupAgro

[3] INRAE, UR PROSE, Antony, Centre d'Antony, Antony, France

Approximation

Consider the first order approximation of the exponential function:

$$\exp(x) = 1 + x + o(x) \tag{A.98}$$

and the first order approximation:

$$\frac{1}{1+x} = 1 - x + o(x) \tag{A.99}$$

Then rewrite MTS expression as follows:

$$\mu_{max} exp\left(\frac{-\lambda}{V_h s}\right) \approx \mu_{max}\left(1 - \frac{\lambda}{V_h s}\right) \tag{A.100}$$

$$\approx \mu_{max}\frac{1}{1 + \frac{\lambda}{V_h s}} \tag{A.101}$$

$$= \mu_{max}\frac{s}{s + \frac{\lambda}{V_h}} \tag{A.102}$$

Comparison of both expressions

The substrate limitation range can be studied through the ratio of both growth functions, shown in expressions (A.103) and (A.104), respectively.

$$\mu_{max}\frac{s}{s + \frac{\lambda}{V_h}} \tag{A.103}$$

$$\mu_{max} exp\left(\frac{-\lambda}{V_h s}\right) \tag{A.104}$$

Noting $K_s := \frac{\lambda}{V_h}$ One then considers the ratio:

$$R(s) = \frac{\exp\left(\frac{-K_s}{s}\right)}{\frac{s}{s + K_s}} \tag{A.105}$$

Note that ratio (A.105) does not depend on μ_{max}, It can be shown that $R(s) \in (0,1)$, by using the well known inequality $\exp(x) < \frac{1}{1-x}$ for $x < 1$. Which is valid since the term inside the exponential is negative:

$$R(s) \leq \frac{1}{1 - \frac{-K_s}{s}} \frac{K_s + s}{s} = 1 \tag{A.106}$$

The change of variables $u = \frac{s}{K_s}$ is used to analyse the expression $R(s)$, which gives equation (A.107).

$$F(u) = \exp\left(-\frac{1}{u}\right) \frac{u+1}{u} \tag{A.107}$$

$$F'(u) = \exp\left(-\frac{1}{u}\right) \frac{1}{u^2} \frac{u}{u+1} + \exp\left(-\frac{1}{u}\right) \frac{1}{(u+1)^2} > 0 \tag{A.108}$$

Since F is monotonic, one gets that there exists a unique u^* such that $\exp\left(-\frac{1}{u^*}\right) \frac{u^*+1}{u^*} = 0.9$, implying a unique $s^* := u^* K_s$, such that $R(s^*) = 0.9$. From the former it can be seen that for each K_s MTS expression approximates to 90 % of the Monod expression whenever $s \geq u^* K_s$.

The curve $s \mapsto R(s)$ is shown in figure A.1, for different K_s values of table 5.1 of the manuscript. One can see that $u^* \approx 1.92$ therefore $s \geq 1.92 K_s$ then $R(s) \geq 0.9$.

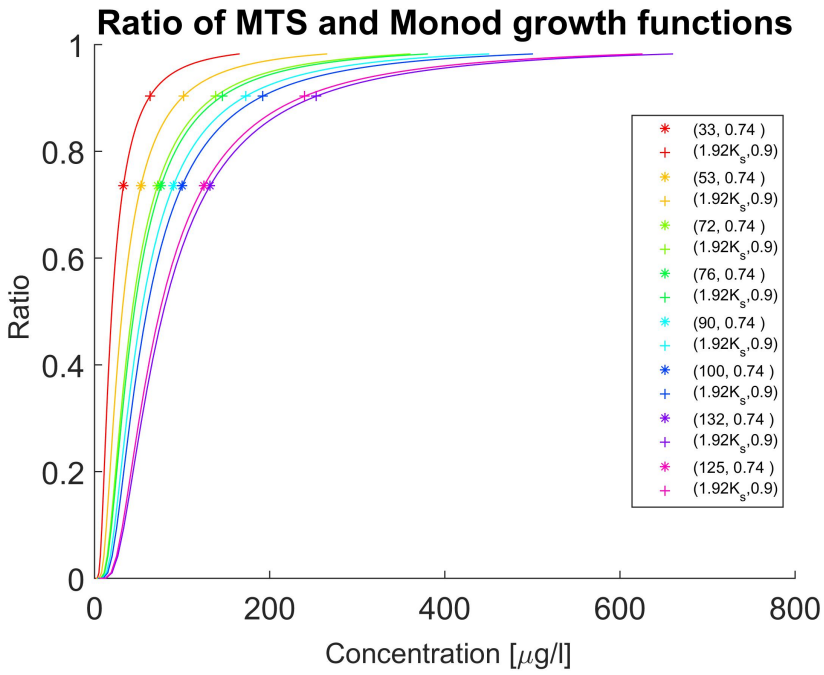


Figure A.1: $R(s)$ curve for several K_s values obtained for E. Coli ML 30. The asterisk $*$ represents K_s values evaluated in expression $R(s)$, note that $R(K_s) \approx 0.74$. The plus sign $+$ represents the point where the $R(s) = 0.9$.

Bibliography

[1] C. A. Aceves-Lara, E. Latrille, N. Bernet, P. Buffière, and J. P. Steyer. A pseudo-stoichiometric dynamic model of anaerobic hydrogen production from molasses. *Water Research*, 2008.

[2] S. Aiba, M. Shoda, and M. Nagatani. Kinetics of product inhibition in alcohol fermentation. *Biotechnology and Bioengineering*, 1968.

[3] A. Y. Aleksandrov and E. B. Aleksandrova. Convergence conditions for some classes of nonlinear systems. *Systems and Control Letters*, 2017.

[4] J. F. Andrews. A mathematical model for the continuous culture of microorganisms utilizing inhibitory substrates. *Biotechnology and Bioengineering*, 10(6):707–723, 1968.

[5] J. H. Andrews and R. F. Harris. r-and k-selection and microbial ecology. *Advances in microbial ecology*, pages 99–147, 1986.

[6] H. Bachmann, M. Fischlechner, I. Rabbers, N. Barfa, F. B. Dos Santos, D. Molenaar, and B. Teusink. Availability of public goods shapes the evolution of competing metabolic strategies. *Proceedings of the National Academy of Sciences of the United States of America*, 2013.

[7] G. Bastin and D. Dochain. *On-line estimation and adaptive control of bioreactors*. Hermann, 1991.

[8] D. J. Batstone, J. Keller, I. Angelidaki, S. V. Kalyuzhnyi, S. G. Pavlostathis, A. Rozzi, W. T. Sanders, H. Siegrist, and V. A. Vavilin. The iwa anaerobic digestion

model no 1 (adm1). *Water science and technology : a journal of the International Association on Water Pollution Research*, 2002.

[9] E. H. Battley. The development of direct and indirect methods for the study of the thermodynamics of microbial growth. *Thermochimica Acta*, 309(1-2):17–37, 1998.

[10] E. H. Battley, R. L. Putnam, and J. Boerio-Goates. Heat capacity measurements from 10 to 300 k and derived thermodynamic functions of lyophilized cells of saccharomyces cerevisiae including the absolute entropy and the entropy of formation at 298.15 k. *Thermochimica Acta*, 298(1-2):37–46, sep 1997.

[11] R. E. Beardmore, I. Gudelj, D. A. Lipson, and L. D. Hurst. Metabolic trade-offs and the maintenance of the fittest and the flattest. *Nature*, 2011.

[12] L. Becks, S. P. Ellner, L. E. Jones, and J. G. Hairston Nelson G. Reduction of adaptive genetic diversity radically alters eco-evolutionary community dynamics. *Ecology Letters*, 2010.

[13] M. Bellucci, I. D. Ofiţeru, D. W. Graham, I. M. Head, and T. P. Curtis. Low-dissolved-oxygen nitrifying systems exploit ammonia-oxidizing bacteria with unusually high yields. *Applied and Environmental Microbiology*, 2011.

[14] P. Biechele, C. Busse, D. Solle, T. Scheper, and K. Reardon. Sensor systems for bioprocess monitoring. *Engineering in Life sciences*, 15(5):469–488, 2015.

[15] Y. Boucher, C. J. Douady, R. T. Papke, D. A. Walsh, M. E. R. Boudreau, C. L. Nesbø, R. J. Case, and W. F. Doolittle. Lateral gene transfer and the origins of prokaryotic groups. *Annual Review of Genetics*, 2003.

[16] S. Boyd, S. P. Boyd, and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[17] V. Bucci, B. Tzen, N. Li, M. Simmons, T. Tanoue, E. Bogart, L. Deng, V. Yeliseyev, M. L. Delaney, Q. Liu, et al. Mdsine: Microbial dynamical systems inference engine for microbiome time-series analyses. *Genome biology*, 17(1):121, 2016.

[18] D. K. Button. Kinetics of nutrient-limited transport and microbial growth. *Microbiological Review*, 1985.

[19] D. K. Button, S. S. Dunker, and M. L. Morse. Continuous culture of rhodotorula rubra: kinetics of phosphate-arsenate uptake, inhibition, and phosphate-limited growth. *Journal of Bacteriology*, 1973.

[20] H. P. Cartan. *Differential calculus*, volume 1. Hermann, 1971.

[21] T. Çimen. Ifac proceedings volumes (ifac-papersonline). 2008.

[22] T. Çimen and S. P. Banks. Nonlinear optimal tracking control with application to super-tankers for autopilot design. *Automatica*, 40(11):1845–1863, 2004.

[23] D. Contois. Kinetics of Bacterial Growth: Relationship between Population Density and Specific Growth Rate of Continuous Cultures. *Journal of General Microbiology*, 21(1):40–50, aug 1959.

[24] B. H. Davison and G. Stephanopoulos. Coexistence of s. cerevisiae and e. coli in chemostat under substrate competition and product inhibition. *Biotechnology and Bioengineering*, 1986.

[25] E. Desmond-Le Quéméner and T. Bouchez. A thermodynamic theory of microbial growth. *ISME Journal*, 8(8):1747–1751, 2014.

[26] S. Di and A. Yang. Analysis of productivity and stability of synthetic microbial communities. *Journal of the Royal Society Interface*, 2019.

[27] D. Dochain. State and parameter estimation in chemical and biochemical processes: A tutorial. *Journal of Process Control*, 2003.

[28] M. Dumont. *Apports de la modélisation des interactions pour une compréhension fonctionnelle d'un écosystème : application à des bactéries nitrifiantes en chémostat*. PhD thesis, Université de Montpellier, 2008.

[29] M. Dumont, J.-J. Godon, and J. Harmand. Species Coexistence in Nitrifying Chemostats: A Model of Microbial Interactions. *Processes*, 2016.

[30] M. Dumont, J. Harmand, A. Rapaport, and J. J. Godon. Towards functional molecular fingerprints. *Environmental Microbiology*, 2009.

[31] M. Dumont, A. Rapaport, J. Harmand, B. Benyahia, and J. J. Godon. Observers for microbial ecology - how including molecular data into bioprocess modeling? In *2008 Mediterranean Conference on Control and Automation - Conference Proceedings, MED'08*, 2008.

[32] M. Dumont, A. Rapaport, J. Harmand, B. Benyahia, and J.-J. Godon. Observers for microbial ecology-how including molecular data into bioprocess modeling? In *Control and Automation, 2008 16th Mediterranean Conference on*, pages 1381–1386. IEEE, 2008.

[33] T. Fagerström and T. Fagerstrom. On theory, data and mathematics in ecology. *Oikos*, 1987.

[34] R. Fekih-Salem, C. Lobry, and T. Sari. A density-dependent model of competition for one resource in the chemostat. *Mathematical Biosciences*, 2017.

[35] T. Ferenci. Growth of bacterial cultures' 50 years on: towards an uncertainty principle instead of constants in bacterial growth kinetics. *Research in microbiology*, 150(7):431–438, 1999.

[36] I. Ferrera and O. Sánchez. Insights into microbial diversity in wastewater treatment systems: How far have we come? *Biotechnology Advances*, 34(5):790 – 802, 2016.

[37] K. Gopalsamy. Global asymptotic stability in a periodic lotka-volterra system. *The Journal of the Australian Mathematical Society. Series B. Applied Mathematics*, 1985.

[38] F. A. Gorter, M. Manhart, and M. Ackermann. Understanding the evolution of interspecies interactions in microbial communities. *Philosophical Transactions of the Royal Society B*, 375(1798):20190256, 2020.

[39] C. L. Grady Jr, B. F. Smets, and D. S. Barbeau. Variability in kinetic parameter estimates: a review of possible causes and a proposed terminology. *Water Research*, 30(3):742–748, 1996.

[40] T. Großkopf and O. S. Soyer. Microbial diversity arising from thermodynamic constraints. *ISME Journal*, 2016.

[41] L. Gurobi Optimization. Gurobi optimizer reference manual, 2021.

[42] J. Haldane. Enzymes longmans. *Green and Co, UK*, 1930.

[43] M. Hanaki, J. Harmand, Z. Mghazli, T. Sari, A. Rapaport, and P. Ugalde. Mathematical study of a two-stage anaerobic model when the hydrolysis is the limiting step. working paper or preprint, 2020.

[44] J. Harmand, C. Lobry, A. Rapaport, and T. Sari. *The Chemostat: Mathematical Theory of Microorganism Cultures*. John Wiley & Sons, 2017.

[45] J. Harmand, C. Lobry, A. Rapaport, and T. Sari. *Optimal Control in Bioprocesses: Pontryagin's Maximum Principle in Practice*. John Wiley & Sons, 2019.

[46] W. E. Hart, C. D. Laird, J.-P. Watson, D. L. Woodruff, G. A. Hackebeil, B. L. Nicholson, and J. D. Siirola. *Pyomo–optimization modeling in python*, volume 67. Springer Science & Business Media, second edition, 2017.

[47] J. J. Heijnen and B. Romein. Derivation of Kinetic Equations for Growth on Single Substrates Based on General Properties of a Simple Metabolic Network. *Biotechnology Progress*, 1995.

[48] J. J. Heijnen and J. P. Van Dijken. In search of a thermodynamic description of biomass yields for the chemotrophic growth of microorganisms. *Biotechnology and Bioengineering*, 1992.

[49] M. Henze, W. Gujer, T. Mino, and M. van Loosedrecht. Activated sludge models asm1, asm2, asm2d and asm3. *Water Intelligence Online*, 2015.

[50] B. Hernández-Bermejo and V. Fairén. Lotka-volterra representation of general nonlinear systems. *Mathematical Biosciences*, 1997.

[51] T. A. Hoek, K. Axelrod, T. Biancalani, E. A. Yurtsev, J. Liu, and J. Gore. Resource availability modulates the cooperative and competitive nature of a microbial cross-feeding mutualism. *PLoS Biology*, 2016.

[52] C. Y. Hoh and R. Cord-Ruwisch. A practical kinetic model that considers end-product inhibition in anaerobic digestion processes by including the equilibrium constant. *Biotechnology and Bioengineering*, 2000.

[53] S. R. Hoover and F. E. Allison. The growth metabolism of rhizobium, with evidence on the interrelations between respiration and synthesis. *Journal of Biological Chemistry*, 134(1):181–192, 1940.

[54] P. A. Hoskisson and G. Hobbs. Microbiology. 2005.

[55] S.-B. Hsu, S. Hubbell, and P. Waltman. A mathematical theory for single-nutrient competition in continuous cultures of micro-organisms. *SIAM Journal on Applied Mathematics*, 32(2):366–383, 1977.

[56] S. B. Hsu, H. L. Smith, and P. Waltman. Competitive exclusion and coexistence for competitive systems on ordered danach spaces. *Transactions of the American Mathematical Society*, 1996.

[57] S. Hubbell et al. Single-nutrient microbial competition: qualitative agreement between experimental and theoretically forecast outcomes. *Science*, 207(4438):1491–1493, 1980.

[58] H. W. Jannasch and T. Egli. Microbial growth kinetics: a historical perspective. *Antonie van Leeuwenhoek*, 1993.

[59] J. Jost, J. Drake, A. Fredrickson, and H. Tsuchiya. Interactions of tetrahymena pyriformis, escherichia coli, azotobacter vinelandii, and glucose in a minimal medium. *Journal of Bacteriology*, 113(2):834–840, 1973.

[60] H. K. Khalil and J. W. Grizzle. *Nonlinear systems*, volume 3. Prentice hall Upper Saddle River, NJ, 2002.

[61] Z. Khedim, B. Benyahia, B. Cherki, T. Sari, and J. Harmand. Effect of control parameters on biogas production during the anaerobic digestion of protein-rich substrates. *Applied Mathematical Modelling*, 61:351 – 376, 2018.

[62] T. Khin and A. P. Annachhatre. Novel microbial nitrogen removal processes. *Biotechnology Advances*, 22(7):519 – 532, 2004.

[63] J. A. Klappenbach, J. M. Dunbar, and T. M. Schmidt. rrna operon copy number reflects ecological strategies of bacteria. *Applied and Environmental Microbiology*, 66(4):1328–1333, 2000.

[64] R. Kleerebezem and M. C. Van Loosdrecht. A generalized method for thermodynamic state analysis of environmental systems. *Critical Reviews in Environmental Science and Technology*, 40(1):1–54, 2010.

[65] M. Kot. *Elements of mathematical ecology*. Cambridge University Press, 2001.

[66] K. Kovářová, A. J. Zehnder, and T. Egli. Temperature-dependent growth kinetics of escherichia coli ml 30 in glucose-limited continuous culture. *Journal of Bacteriology*, 178(15):4530–4539, 1996.

[67] K. Kovárová-Kovar and T. Egli. Growth Kinetics of Suspended Microbial Cells: From Single-Substrate-Controlled Growth to Mixed-Substrate Kinetics. *Microbiology and Molecular Biology Reviews*, 62(3):646–666, 1998.

[68] J. E. Leffler and E. Grunwald. *Rates and equilibria of organic reactions: as treated by statistical, thermodynamic and extrathermodynamic methods*. Courier Corporation, 2013.

[69] U. Lendenmann, M. Snozzi, and T. Egli. Growth kinetics of escherichia coli with galactose and several other sugars in carbon-limited chemostat culture. *Canadian Journal of Microbiology*, 46(1):72–80, 2000.

[70] J. S. Liu, I. W. Marison, and U. von Stockar. Microbial growth by a net heat up-take: A calorimetric and thermodynamic study on acetotrophic methanogenesis by methanosarcina barkeri. *Biotechnology and Bioengineering*, 2001.

[71] Y. Liu. Overview of some theoretical approaches for derivation of the monod equation. *Applied microbiology and biotechnology*, 73(6):1241–1250, 2007.

[72] C. Lobry and J. Harmand. A new hypothesis to explain the coexistence of n species in the presence of a single resource. *Comptes Rendus - Biologies*, 2006.

[73] J. R. Lobry, J. P. Flandrois, G. Carret, and A. Pave. Monod's bacterial growth model revisited. *Bulletin of Mathematical Biology*, 1992.

[74] A. J. Lotka. Elements of physical biology. *Journal of the American Statistical Association*, 1925.

[75] T. R. Maarleveld, R. A. Khandelwal, B. G. Olivier, B. Teusink, and F. J. Bruggeman. Basic concepts and principles of stoichiometric modeling of metabolic networks. *Biotechnology Journal*, 8(9):997–1008, 2013.

[76] M. T. Madigan, J. M. Martinko, D. A. Stahl, and D. P. Clark. *Brock Biology of Microorganisms*. 2012.

[77] P. L. McCarty. Thermodynamics of biological synthesis and growth. *Air and water pollution*, 9(10):621–639, 1965.

[78] S. McNaughton. r-and k-selection in typha. *The American Naturalist*, 109(967):251–261, 1975.

[79] J. Merchuk and J. Asenjo. The monod equation and mass transfer. *Biotechnology and Bioengineering*, 45(1):91–94, 1995.

[80] L. Michaelis and M. L. Menten. Die kinetik der invertwirkung. *Biochemische Zeitschrift*, 1913.

[81] R. Milo, P. Jorgensen, U. Moran, G. Weber, and M. Springer. Bionumbers the database of key numbers in molecular and cell biology. *Nucleic Acids Research*, 38(SUPPL.1), oct 2009.

[82] P. Mitchell. Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. *Nature*, 1961.

[83] D. Molenaar, R. Van Berlo, D. De Ridder, and B. Teusink. Shifts in growth strategies reflect tradeoffs in cellular economics. *Molecular systems biology*, 5(1):323, 2009.

[84] B. Momeni, L. Xie, and W. Shou. Lotka-volterra pairwise modeling fails to capture diverse pairwise microbial interactions. *eLife*, 6, 2017.

[85] J. Monod. Recherches sur la croissance des cultures bacteriennes. 1942.

[86] H. Moser et al. *The dynamics of bacterial populations maintained in the chemostat.* Carnegie Institution of Washington, 1958.

[87] J. D. Murray. *Mathematical Biology : I . An Introduction , Third Edition*, volume 1. 2002.

[88] S. Narayanasamy, E. E. Muller, A. R. Sheik, and P. Wilmes. Integrated omics for the identification of key functionalities in biological wastewater treatment microbial communities. *Microbial Biotechnology*, 2015.

[89] G. Nicolis and I. Prigogine. *Self-Organization in Nonequilibrium Systems: From Dissipative Structures to Order Through Fluctuations.* A Wiley-Interscience publication. Wiley, 1977.

[90] B. Niebel, S. Leupold, and M. Heinemann. An upper limit on gibbs energy dissipation governs cellular metabolism. *Nature Metabolism*, 2019.

[91] J. Nielsen and J. Villadsen. Modelling of microbial kinetics. *Chemical Engineering Science*, 47(17-18):4225–4270, 1992.

[92] A. Novick and L. Szilard. Description of the chemostat. *Science*, 1950.

[93] G. S. Ostace, V. M. Cristea, and P. Agachi. Extension of activated sludge model no 1 with two-step nitrification and denitrification processes for operation improvement. *Environmental Engineering and Management Journal*, 2011.

[94] J. D. Owens and J. D. Legan. Determination of the Monod substrate saturation constant for microbial growth. *FEMS Microbiology Letters*, 46(4):419–432, 1987.

[95] S. Pavlou. Computing operating diagrams of bioreactors. *Journal of Biotechnology*, 1999.

[96] F. Pelletier, D. Garant, and A. P. Hendry. Eco-evolutionary dynamics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 2009.

[97] O. Perez-Garcia, G. Lear, and N. Singhal. Metabolic network modeling of microbial interactions in natural and engineered environmental systems. *Frontiers in Microbiology*, 2016.

[98] E. R. Pianka. On r- and k-selection. *The American Naturalist*, 1970.

[99] S. S. Pilyugin and P. Waltman. Multiple limit cycles in the chemostat with variable yield. *Mathematical Biosciences*, 2003.

[100] L. Poughon, C. G. Dussap, and J. B. Gros. Energy model and metabolic flux analysis for autotrophic nitrifiers. *Biotechnology and Bioengineering*, 72(4):416–433, feb 2001.

[101] J. I. Prosser, B. J. Bohannan, T. P. Curtis, R. J. Ellis, M. K. Firestone, R. P. Freckleton, J. L. Green, L. E. Green, K. Killham, J. J. Lennon, A. M. Osborn, M. Solan, C. J. van der Gast, and J. P. W. Young. The role of ecological theory in microbial ecology. *Nature Reviews Microbiology*, 2007.

[102] Y. Rafrafi, E. Trably, J. Hamelin, E. Latrille, I. Meynial-Salles, S. Benomar, M. T. Giudici-Orticoni, and J. P. Steyer. Sub-dominant bacteria as keystone species in microbial communities producing bio-hydrogen. *International Journal of Hydrogen Energy*, 2013.

[103] D. Reznick, M. J. Bryant, and F. Bashey. r-and k-selection revisited: the role of population regulation in life-history evolution. *Ecology*, 83(6):1509–1520, 2002.

[104] J. Roels. Application of macroscopic principles to microbial metabolism. *Biotechnology and Bioengineering*, 22(12):2457–2514, 1980.

[105] J. A. Roels. Energetics and kinetics in biotechnology. *The Quarterly Review of Biology*, 1984.

[106] H. Roques, S. Yue, S. Saipanich, and B. Capdeville. Is monod's approach adequate for the modelisation of purification processes using biological treatment? *Water Research*, 1982.

[107] Y. Saito and T. Miki. Species coexistence under resource competition with intraspecific and interspecific direct competition in a chemostat. *Theoretical Population Biology*, 2010.

[108] A. Schrijver. *Theory of linear and integer programming*. John Wiley & Sons, 1998.

[109] R. Schuetz, N. Zamboni, M. Zampieri, M. Heinemann, and U. Sauer. Multidimensional optimality of microbial metabolism. *Science*, 2012.

[110] H. Senn, U. Lendenmann, M. Snozzi, G. Hamer, and T. Egli. The growth of escherichia coli in glucose-limited chemostat cultures: a re-examination of the kinetics. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1201(3):424–436, 1994.

[111] B. Sharma and R. Ahlert. Nitrification and nitrogen removal. *Water Research*, 11(10):897 – 925, 1977.

[112] S. Siripong and B. E. Rittmann. Diversity study of nitrifying bacteria in full-scale municipal wastewater treatment plants. *Water Research*, 41(5):1110 – 1120, 2007.

[113] C. M. Smeaton and P. Van Cappellen. Gibbs Energy Dynamic Yield Method (GEDYM): Predicting microbial growth yields under energy-limiting conditions. *Geochimica et Cosmochimica Acta*, 2018.

[114] R. R. Sokal. The principles and practice of numerical taxonomy. *TAXON*, 1963.

[115] H.-S. Song, W. Cannon, A. Beliaev, and A. Konopka. Mathematical modeling of microbial community dynamics: A methodological review. *Processes*, 2014.

[116] A. Succurro and O. Ebenhöh. Review and perspective on mathematical modeling of microbial ecosystems. *Biochemical Society Transactions*, 2018.

[117] Y. Tan, Z. X. Wang, and K. C. Marshall. Modeling substrate inhibition of microbial growth. *Biotechnology and Bioengineering*, 1996.

[118] Y. Tan, Z. X. Wang, R. P. Schneider, and K. C. Marshall. Modelling microbial growth: A statistical thermodynamic approach. *Journal of Biotechnology*, 1994.

[119] I. S. Team Commands. Bocop: an open source toolbox for optimal control, 2017.

[120] L. Tijhuis, M. C. Van Loosdrecht, and J. J. Heijnen. A thermodynamically based correlation for maintenance gibbs energy requirements in aerobic and anaerobic chemotrophic growth. *Biotechnology and Bioengineering*, 1993.

[121] E. Trélat. *Contrôle optimal: théorie & applications*. Vuibert Paris, 2005.

[122] H. M. Tsuchiya, J. F. Drake, J. L. Jost, and A. G. Fredrickson. Predator-prey interactions of dictyostelium discoideum and escherichia coli in continuous culture. *Journal of Bacteriology*, 1972.

[123] P. Ugalde-Salas, E. Desmond-Le Quéméner, J. Harmand, A. Rapaport, and T. Bouchez. Insights from Microbial Transition State Theory on Monod's Affinity Constant. *Scientific Reports*, 2020.

[124] P. Ugalde-Salas, J. Harmand, and E. Desmond-Le Quéméner. Modelling interactions of a microbial community through an optimal tracking control approach. In *2020 European Control Conference (ECC)*, pages 587–592. IEEE, 2020.

[125] P. Ugalde-Salas, J. Harmand, and E. D. L. Quemener. Asymptotic observers and integer programming for functional classification of a microbial community in a chemostat. In *2019 18th European Control Conference, ECC 2019*, 2019.

[126] I. A. Vasiliadou, S. Pavlou, and D. V. Vayenas. Dynamics of a chemostat with three competitive hydrogen oxidizing denitrifying microbial populations and their efficiency for denitrification. *Ecological Modelling*, 2009.

[127] G. J. Velicer and R. E. Lenski. Evolutionary trade-offs under conditions of resource abundance and scarcity: Experiments with bacteria. *Ecology*, 1999.

[128] P. F. Verhulst. Notice sur la loi que la population suit dans son accroissement. *Correspondance Mathématique et Physique*, 1838.

[129] V. Volterra. Fluctuations in the abundance of a species considered mathematically. *Nature*, 1926.

[130] U. Von Stockar and J. S. Liu. Does microbial life always feed on negative entropy? thermodynamic analysis of microbial growth. *Biochimica et Biophysica Acta - Bioenergetics*, 1999.

[131] U. Von Stockar, T. Maskow, J. Liu, I. W. Marison, and R. Patiño. Thermodynamics of microbial growth and metabolism: An analysis of the current situation. *Journal of Biotechnology*, 2006.

[132] U. von Stockar, V. Vojinović, T. Maskow, and J. Liu. Can microbial growth yield be estimated using simple thermodynamic analogies to technical processes? *Chemical Engineering and Processing: Process Intensification*, 2008.

[133] M. J. Wade, J. Harmand, B. Benyahia, T. Bouchez, S. Chaillou, B. Cloez, J.-J. Godon, C. Lobry, B. Moussa Boubjemaa, A. Rapaport, M. J. Wade, J. Harmand,

B. Benyahia, T. Bouchez, S. Chaillou, B. Cloez, J.-J. Godon, B. Moussa Boudjemaa, A. Rapaport, T. Sari, R. Arditi, and C. Lobry. Perspectives in mathematical modelling for microbial ecology. *Ecological Modelling*, 2016.

[134] M. Wagner, A. Loy, R. Nogueira, U. Purkhold, N. Lee, and H. Daims. Microbial community composition and function in wastewater treatment plants. *Antonie van Leeuwenhoek*, 2002.

[135] K. V. Waller and P. M. Mäkilä. Chemical reaction invariants and variants and their use in reactor modeling, simulation, and control. *Industrial and Engineering Chemistry Process Design and Development*, 1981.

[136] Z. W. Wang and Y. Li. A theoretical derivation of the contois equation for kinetic modeling of the microbial degradation of insoluble substrates. *Biochemical Engineering Journal*, 2014.

[137] S. A. West, S. P. Diggle, A. Buckling, A. Gardner, and A. S. Griffin. The Social Lives of Microbes. *Annual Review of Ecology, Evolution, and Systematics*, 2007.

[138] S. Widder, R. J. Allen, T. Pfeiffer, T. P. Curtis, C. Wiuf, W. T. Sloan, O. X. Cordero, S. P. Brown, B. Momeni, W. Shou, et al. Challenges in microbial ecology: building predictive understanding of community function and dynamics. *The ISME journal*, 10(11):2557–2568, 2016.

[139] U. Wiesmann. Biological nitrogen removal from wastewater. *Biotechnics/Wastewater*, 1994.

[140] C. R. Woese and G. E. Fox. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proceedings of the National Academy of Sciences of the United States of America*, 1977.

[141] G. S. Wolkowicz and Z. Lu. Global dynamics of a mathematical model of competition in the chemostat: general response functions and differential death rates. *SIAM Journal on Applied Mathematics*, 52(1):222–233, 1992.

[142] J. Xiao and J. M. VanBriesen. Expanded thermodynamic model for microbial true yield prediction. *Biotechnology and Bioengineering*, 2006.