How do energy gradients shape microbial communities? Study of the Microbial Transition State (MTS) approach for modelling microbial ecosystems dynamics and its application to environmental biotechnology processes

Hadrien Delattre

agreenium
l'institut agronomique, vétérinaire & forestier de France

AgroParisTech

irstea

NNT° : 2018IAVF0011

THESE DE DOCTORAT

préparée à l'Institut des sciences et industries du vivant et de l'environnement (AgroParisTech)

pour obtenir le grade de

Docteur de l'Institut agronomique vétérinaire et forestier de France

Spécialité : Ecologie

École doctorale n°581
Agriculture, alimentation, biologie, environnement et santé (ABIES)

par

Hadrien DELATTRE

How do energy gradients shape microbial communities? Study of the Microbial Transition State (MTS) approach for modelling microbial ecosystems dynamics and its application to environmental biotechnology processes

Directeur de thèse : **Théodore Bouchez**

Co-encadrant de thèse : **Elie Le Quéméner**

Thèse présentée et soutenue à IRSTEA Antony, le 05/07/18 :

Composition du jury :

Béatrice LAROCHE, Directrice de Recherche, INRA-MaIAGE, Jouy-en-Josas, France	Présidente
Thomas MASKOW, PD Dr, Head of the Ecothermodynamics group, UFZ-Leipzig, Germany	Rapporteur
Jean-Philippe STEYER, Directeur de Recherche, INRA-LBE, Narbonne, France	Rapporteur
Robbert KLEEREBEZEM, Associate Professor, TU Delft, Netherlands	Examinateur
Patrick PERRE, Professeur, CentraleSupélec-LGPM, Gif sur Yvette, France	Examinateur
Elie LE QUEMENER, Chargé de recherche, INRA-LBE, Narbonne, France	Co-encadrant
Théodore BOUCHEZ, Ing. en Chef des Ponts des Eaux et des F., Irstea-HBAN, Antony, France	Directeur de thèse

Equipe PROSE
Unité HBAN, IRSATEA Antony, 1 Rue Pierre Gilles de Gennes, 92160 Antony

Remerciements / Thanks

Je remercie les membres de mon jury de thèse, Thomas Maskow, Jean-Philippe Steyer, Robbert Kleerebezem, Patrick Perré, Béatrice Laroche, Théodore Bouchez, Elie Le Quéméner.

Je remercie ma famille et mes amis qui m'ont toujours soutenu dans ce projet et m'ont permis d'aller jusqu'au bout.

Je remercie l'ensemble des membres passés et présents de l'équipe BIOMIC pour m'avoir accueilli et permis de progresser; Théodore Bouchez, Elie Le Quéméner, Christian Duquennoi, Ariane Bize, Oliver Chapleur, Laurent Mazéas, Christine Bureau, Céline Madigou, Cédric Midoux, Jianghao Tian, Thomas Jouen, Simon Poirier, Laëticia Cardonna, Maxime Allioux, Daniela Patinier, Lilia Ahmed-Zaïd et Arnaud Bridier. Je remercie aussi les membres de l'équipe EPURE, avec qui j'ai pû avoir d'intéressantes collaborations et discussion ; je remercie en particulier Ahlem Filali et Jean-Jacques Pernelle pour m'avoir co-encadré. Je remercie plus largement tous les membres de l'unité HBAN pour m'avoir accueilli. Je remercie également les participants de mon comité de thèse ainsi que les membres du projet thermomic, pour les fructueuses discussions que nous avons eu et qui m'ont permis de m'orienter au cours ma thèse, en particulier Bart Haegeman, Romain Girault, Jérôme Harmand, Franck Jabot, Etienne Paul et Jean-Philippe Steyer.

Je remercie enfin toutes les personnes, Christian Duquennoi, Gaelle Tallec, Lénaïck Rouillac, Matthieu Dorel et Thomas Jouen en particulier, avec qui j'ai pu avoir de nombreuses discussions scientifiques audacieuses. Je remercie également toutes les personnes qui ont supporté ces discussions alors qu'elles n'avaient rien demandé.

I also would like to thank the members of TU Delft's Environmental Biotechnology group for hosting me in winter 2016.

Publications

- **DELATTRE H.**, LE QUEMENER E., KLEEREBEZEM R., BOUCHEZ T. Modelling the energy dependence of microbial growth. Submitted to FEMS microbiology (chapter 2 of this memoir)
- **DELATTRE H.**, LE QUEMENER E., DUQUENNOI C., FILALI A., BOUCHEZ T. Consistent microbial dynamics and functional community patterns derived from first physical principles. in Accepted in ISME Journal (IF=9.66) (chapter 4 of this memoir)
- **DELATTRE H.**, LE QUEMENER E., FILALI A., PAUL E., BOUCHEZ T. Consistent dynamics of simplified activated sludge ecosystem obtained by the calibration of a parameter-parsimonious thermodynamic model. In preparation (chapter 5 of this memoir)
- **DELATTRE H.**, LE QUEMENER E., KLEEREBEZEM R., BOUCHEZ T. Predicting microbial growth yield from the nature of metabolic reactions. In preparation (chapter 6 of this memoir)

International communications

- **DELATTRE H.**, LE QUEMENER E., DUQUENNOI C., FILALI A., BOUCHEZ T. A Thermodynamic theory of microbial growth and its perspectives for modelling environmental biotechnology processes. Poster presentation at ISME 2016 (Montréal, August 21-26)
- **DELATTRE H.**, LE QUEMENER E., DUQUENNOI C., FILALI A., BOUCHEZ T. Consistent dynamics and community assembly patterns directly arising from thermodynamic theory of microbial growth. Poster presentation and oral presentation at FEMS 2017 (Valencia, July 7-13)
- **DELATTRE H.**, LE QUEMENER E., BOUCHEZ T. A generic thermodynamic model to predict the dynamics of a microbial community according to mass and energy balance. Oral presentation at the “Bridging theory and practice in ecological engineering” workshop (Narbonne, 25-26th June 2018)

Contents

Chapter 1:

Introduction

1.1 Microbial ecology as a discipline

The term "ecosystem" was coined in 1935 by the botanist Tansley; its most consensual acceptation designates the set of all the individuals (also called community, or biocenosis) living in a delimited environment called "biotope". Microbial ecology is then the study of a microbial community, and the interactions between its members and with its biotope. The origin of this disciplin can be traced back to the work of the microbiologist Winogradsky, starting in the late XIXth century. Indeed he formulated what is now envisioned as microbial ecology problems, by emphasizing the in-situ study of microbes and their relationship with their geochemical environment (Zavarzin, 2006). However, the development of this discipline was hindered by a technical deadlock until the end of the XXth century; culture was the only way to study microorganisms. Although some coculture experiments run in this context led to insights regarding species interactions (such as the formulation of the competitive exclusion principle (Gause, 1934)), microbiology mainly studied microbes under the perspective of pure culture.

The development of metagenomics approaches at the very end of the XXth century totally changed this persective. It was observed that populations in microbial communities interacts in ways comparable as those observed in macrobiology; competition for resources, predation, parasitism and mututalism occur (Little et al., 2008). Moreover, while unicellular organisms have initially been thought of as individualistic and disorganized, high levels of self-organization have also been evidenced through inter and intra specie interactions. Indeed, multiple observations shows that a microbe collective can self-organize, thus making large scale patterns to emerge to the point their behavior have been compared to this of a multicellular organism (Velicer et al., 1998; Velicer, 2003; Jacob et al., 2004; Gloag et al., 2015). To summarise, those observations suggest that microbes can assemble into complex communities harboring interactions already described in macrobial community ecology. Those similarities foster the use of tools and concepts developed for macrobial ecology in microbial ecology.

However microbial communities also carry specificies which make it a field distinct from its macrobial counterpart. A list of differences between microbes and macrobes which are expected to impact community assembly is given by Nemergut and collaborators. This list includes the higher dispersal potential of microbes, the (potentially) vast genetic reservoir induced by their ability to stay dormant for long time periods, their important genetic plasticity. This latter feature implies that microbes are able to share genetic material even with distantly related

kins, and also to adapt their metabolism to adapt to their environment far more easily than most macrobes could (Nemergut et al., 2013).

The study of microbial ecology is particularly relevant in multiple regards; the understanding of the behavior of biogeochemical fluxes in the environment, the understanding of the dynamics of human-associated microbiota such as this of the intestinal tract, and the improvement of industrial processes based on microbial communities. The work undertaken during this thesis particularly focuses on the latter.

1.2 The role of microbes in the geochemical fluxes

Microbial communities plays a crucial role in the regulation of biogeochemical processes, literally controlling the habitability conditions of the earth (Falkowski et al., 2008; Rousk and Bengtson, 2014; Treseder et al., 2012). It has been estimated that the quantity of carbon stored in microorganisms correspond to 60-100% of this stored in plant biomass, and prokaryotes are estimated to constitute by far the largest pool of nitrogen and phosphorus of the biosphere (Whitman et al., 1998).

The fact that microorganisms modify their environment and that such phenomena is important at a global scale has also been intuited long before it could be measured, by microbiologists such as Winogradsky (Zavarzin, 2006) or Baas Becking (De Wit and Bouvier, 2006). Most of microbes are growing by carrying out redox reactions. While every living organisms community performs biochemical interactions with its environment, microbes are performing chemical interactions in the most direct way (even sometimes interacting directly with electrons (Strycharz-Glaven et al., 2011)). Therefore the functional and taxonomic structure of a microbial community is directly influenced, and directly influencing, the physicochemistry of its environment. Moreover, the microbes consitutes an important part of the total biosphere.

The human population recently realized that its own influence on earth's geochemical cycles may impede the development of its future generations (Griggs et al., 2013). The study of the interactions between microbes and their environment is then of paramount importance for the sustainable evolution of the biosphere.

1.3 The use of microbial ecosystems in bioprocesses

Microbial ecosystems being ubiquitous, they have been serendipituously used by humankind to perform biochemical transformations since at least the neolithic. Indeed, the chemical analysis of potteries estimated to date from 7000 years B.C. revealed they contained a mixed fermented beverage of rice, honey and fruit. (McGovern et al., 2004). The first unicellular organisms have been described during the XVIIth century by van Leeuwenhoek and naturalists such as Swammerdam, however, the meaning to give to these observations, the notion of cell, and the actual relevance of the study of biology at microscopic level was debatted until the XIXth century. At that time, vitalism progressively became obsolete, industrial processes actually involving microbes gradually became perceived as such, and the transformations carried out by microbes became perceived as chemical reactions. Some important events marking this paradigm shift in biology are the synthesis of urea by Wöhler in 1828, the discovery of the first enzyme (diastase) by Payen and Perzoz in 1833, and the dispelling of spontaneous generation by Pasteur's swan-necked bottles experiment in 1859.

The main industrial bioprocesses relying on a community of microbes are wastewater treatment and anaerobic digestion. Other examples of bioprocesses are related to food industry, such as the production of wine, beer, cheese and vinegar, and the production of fodder through silaging. Another non-food related industrial use of fermentation is the exploitation of the acetone-butanol-ethanol metabolic pathway (Moon et al., 2016). This specific fermentation has been used to produce acetone until world war II when petrochemistry-based acetone production became cheaper. The possibilities offered by microbial ecosystems engineering also goes beyond the improvement of the current bioprocesses; the use of this discipline is also envisioned as a mean to cure microbiota-related diseases (Shen et al., 2015) or to depollute an environment (Crawford and Crawford, 1996).

Many bioprocesses tend to use pure culture or low diversity cultures of microbes, because it makes the process more controlable. However in some cases the sterility of the process cannot be assured at viable costs; mixed culture then appears as an attractive solution despite its complexity (Kleerebezem and van Loosdrecht, 2007). This is notably the case with wastewater treatment and anaerobic digestion, where the input is non-sterile. In this case, diverse microbial ecosystems have to be used.

1.4 Addressing the technical deadlock on microbial community engineering

The development of novel technologies based on mixed microbial cultures, as well as the improvement of existing bioprocesses, is currently uneasy because of the lack of a solid theoretical basis to understand and predict the trajectories of microbial communities. All the models used to predict the behavior of a microbial community (such as ADM1 (Batstone, 2001)) are based on empirical equations calibrated after years of experiments. Moreover, those models do not provide knowledges useful to the modelling of other bioprocesses. Consequently the development of new bioprocesses, or the modification of existing ones, requires experimental work costly in time, money and engineering effort.

On the other hand, products such as cheese or beer have been consistently reproduced from rudimentary and varied material, with few to none microbiological knowledge. This observation evidences that environmental factors (such as temperature and concentrations) exert an important constraint on microbial communities' structure.

Evidences of the deterministic effect of environmental factors, and more specifically physicochemical factors, on the communities' expressed functions, can also be found in natural environments. Indeed, remarkable spatial structuration patterns of microbial communities have been explained by thermodynamics in the case of Winogradsky columns (Zavarzin, 2006), aquifers (Chapelle and Lovley, 1992) and lake hypolimnia (Boehrer and Schultze, 2008; Müller et al., 2012). The aforementioned cases feature ecological successions reflecting a gradient of available energy in the environment, this is also called a "redox tower". Even in the absence of an obvious gradient, recent studies evidenced that physicochemical factors have a deterministic influence on the functional structure (Louca et al., 2016b,a; Louca and Doebeli, 2017). Those results are in line with other results showing microbial communities' functional stability over taxonomic unstability (Fernandez et al., 2000; Burke et al., 2011). These results suggest that the functional structure of microbial communities is more sensitive to selection and less sensitive to invasion processes when compared to the taxonomic structure. Those evidences lead some authors to argue that functions may be a descriptive unit more relevant than taxon in the search for generic principles explaining microbial community structuration (Burke et al., 2011; Louca et al., 2016b; Lemanceau et al., 2017).

Thermochemistry, which is the study of energy transfers in chemical systems, then appears as a good starting

point in the search for generic organisational principles in microbial functional communities' structuration.

1.5 The emergence of microbial thermodynamics

In 1960, Bauchop and Elsden suggested that since microbial growth was constrained by thermodynamics, the study of its energetics could lead to an effective method to predict growth yields (Bauchop and Elsden, 1960). This fostered two decades of attempts to empirically correlate growth yield with physicochemical parameters of metabolisms (Mayberry et al., 1967, 1968; Prochazka et al., 1970; Minkevich and Eroshin, 1973; Stouthamer, 1973; Linton and Stephenson, 1978). Those attempts later cristallized in models accounting for growth energetics in order to predict growth yields and rates. This emerging field will later be referred to as "microbial thermodynamics". These are presented in the bibliography review of this thesis (chapter 2), in order to provide the reader with a firm grasp on the state of the art in microbial thermodynamics. This review is then concluded by a presentation of the objectives of the thesis (last section of chapter 2).

1.6 Objectives and organization of this memoir

The work undertaken during this thesis is purposed at improving the current theoretical knowledge on microbial ecosystem structuration. It is done by studying the capability of a model, the Microbial Transition State (MTS) model (Desmond-Le Quéméner and Bouchez, 2014; Wade et al., 2016).

This model has been used during this thesis to simulate the growth of multiple microbial populations catalyzing different metabolisms to test to which extent the theory behind the MTS model is able to explain the functional structuration of microbial communities.

The next chapter is the bibliography review of this memoir. It is followed by the materials and methods, which describe the implemention details of the MTS model. Then two articles based on predictions made with the model are presented. The first article is a demonstration of trophic and ecological patterns predictable by the MTS model through increasingly complex implementations. The second papers features an implementation of a microbial community related to a bioprocess (activated sludge), along with the calibration of the model's parameters from experimental data. The third article leverages on a body of experimental data collected from

the literature to question the link between the growth yield and physicochemical properties of the metabolisms. The final chapter is a general discussion which summarises what insight has been earned on the structuration of microbial communities and which further developments are required to step forward in the development of microbial community engineering.

Chapter 2:

Modeling the energy dependence of microbial growth

This chapter is the bibliography synthesis of the memoir. It corresponds to a bibliography review made at the invitation of the FEMS Microbiology Ecology journal. The purpose of this review is to introduce some of the models proposed to describe the role of energy in microbial growth. This review focuses on models predicting growth yields and dynamics at population scale.

The review is split into two parts; the first part concerns the growth yield prediction models, which make predictions about the energy balance of microbial growth and do not provide a formulation for dynamic variables such as growth rate. The second part concerns population dynamics models, which express dynamic variables (such as the growth rate) from energy-based considerations on microbial growth.

As a chapter of this memoir, this review presents the MTS model in the context of other microbial thermodynamics models. Firstly, this helps to understand the relevance of the contribution of the MTS theory to the field of microbial thermodynamics, as the properties of the MTS theory in terms of population dynamics prediction are studied in the chapter 4 and 5 of this memoir. Secondly, this emphasizes the contribution of previous models to the implementations of the MTS model used in this memoir. Indeed, since the MTS theory, in itself, does not define some variable (dissipated energy) required to close the energy balance of growth, it borrows an empirical expression allowing to compute this variable from a previous model (Heijnen's energy dissipation model, introduced in this chapter). The calibration of a new empirical relationship to close the energy balance of microbial growth is attempted in chapter 6.

Modeling the energy dependence of microbial growth

Hadrien Delattre[1], Elie Le Quéméner[2], Robbert Kleerebezem[3], Théodore Bouchez[1]

[1] Irstea, UR HBAN, F-92761, Antony, France.

[2] LBE, Univ Montpellier, INRA, Narbonne, France

[3] Department of Biotechnology, Delft University of Technology, Van der Maasweg 9, 2629 HZ Delft, The Netherlands

Correspondence: T Bouchez, Irstea, UR HBAN, 1 rue Pierre-Gilles de Gennes CS 10030, F-92761 Antony, France

E-mail: theodore.bouchez@irstea.fr

Abstract

Energy plays a key role in microbial community dynamics and structure, as illustrated by the commonly called “redox towers” of microbial metabolism observed in numerous engineered or natural ecosystems such as aquifers. This article presents and discusses how the influence of chemical energy is captured in microbial growth models. On one hand, it is now firmly established from thermodynamic balance analysis of microbial growth that energy constraints largely determine microbial growth yields. Not surprisingly, microbes have to comply with physical rules, and we present methods that enable accurate prediction of the growth yields of microbial cultures solely based on the thermodynamic properties of redox reactions. On the other hand, flux-force relationships in microbial growth are still under debate. We review the different approaches currently available, including a recently proposed dynamic growth theory -the microbial transition state (MTS) theory- in which microbial dynamics and stoichiometry are intrinsically derived from the dissipation of energy gradients. The review ends by identifying prospects for the development of microbial thermodynamics.

Introduction

Microbial communities are ubiquitous and play a crucial role in the regulation of biogeochemical processes, literally controlling habitability conditions on earth (Falkowski et al 2008). Moreover, they are used in many bioprocesses such as wastewater treatment and many food production processes. Despite their importance, studies on microbial growth communities only recently gained momentum, with the development of molecular microbial ecology methods in the late 1990s, which made possible to study microbial communities in situ. It was observed that the populations in microbial communities interact in ways that are comparable to those observed in macrobiology, including competition for resources, predation, mutualism (Little et al 2008). However microbial communities also have specificities that make the field distinct from its macrobial counterpart, including higher dispersal of individuals and important genetic plasticity (Nemergut et al 2013).

Before the study of microbial communities took off, models were designed predicting their mass balance and kinetics, notably for the use of mixed culture waste management bioprocesses (Activated Sludge Model (Henze et al 1987, Henze et al 2000), Anaerobic Digester Model 1 (Batstone et al 2001)). They use an approach similar to the study of pure cultures, where the dependency between the system's variables (biomass, growth rate as a function of substrate concentration etc.) is captured in the parameters of empirical equations calibrated from observations. However, a model designed using such an approach is only to a limited extend applicable if the culture conditions differ from those in which the model was calibrated. Moving outside the calibrated area of the model may result in a complete change in microbial community structure and corresponding stoichiometric and kinetic properties. Notably, the calibration of the biomass composition may be a key limitation. Indeed, the ecological structure reproduced by such models is the consequence of expert knowledge implemented into it, but does not emerge from the model itself. In fact, the mechanistic approach to modelling microbial communities is not very advanced

as ecology in general did not yield many generic rules, to the point that its ability as a discipline to produce such rules has been called into question (Lawton 1999).

These engineering models do not aim at providing a theory in which invariants of microbial community structure can be formulated. Consequently, such theory is still lacking. Meanwhile, the amount of data that can be collected from the study of microbial communities is increasing exponentially (Nature Methods Editorial 2009). To quote Henry Poincarré; “Science is built up with facts, as a house is with stones. But a collection of facts is no more a science than a heap of stones is a house”. Microbial ecology as a science has indeed reached a stage at which a firmer theoretical footing is needed to integrate its numerous observations into a coherent picture.

One reason for the apparent lack of generic rules in microbial community structure is the use of taxon as the descriptive unit for microbial communities. Indeed, multiple experimental reports provide evidence that microbial communities have stable functions despite an unstable taxonomic profile (Burke et al 2011, Fernández et al 1999, Huttenhower et al 2012). Moreover, experimental observations suggest that physicochemical factors (pH, temperature, salinity etc.) exert a determining influence on the functional structure of microbial communities (Louca et al 2016a, Louca et al 2016b, Raes et al 2011). According to those results, the functional structure is more sensitive to selection and less sensitive to invasion than the taxonomic structure.

Thermochemistry appears to be a good starting point to build a theoretical framework to integrate generic assembly rules for microbial communities. Indeed, thermochemistry makes it possible to model the influence of physicochemical factors on growth. While the above-mentioned results were obtained using metagenomics, for decades, remarkable spatial patterns of microbial communities have been explained by thermodynamics (like Winogradsky columns (Zavarzin 2006) or aquifers (Chapelle and Lovley 1992)). From the 1960s on, the observation of invariants in the structure of microbial community

encouraged the development of models linking variables such as growth rate and yield to physicochemical factors (Bauchop and Elsden 1960, McCarty 1965). Since then, thermodynamic theories of microbial growth have been published in a somewhat disseminated fashion. The ensuing problems of redundancy, notation mismatch and internal inconsistency (Heijnen and Dijken 1991) make it hard to draw a clear map of the existing theories on the subject.

The aim of this review is consequently to provide a clear map of the discipline, hereafter referred to as "microbial thermodynamics". We focus on models intended to enable conclusions to be drawn at population scale, although models also exist at the intracellular scale (González-Cabaleiro et al. 2013), and also at the scale of whole ecosystems (Ludovisi 2009, Odum 1969, Svirezhev 2004). In next section, we present the classical differential equation framework to model microbial population growth, along with ideas concerning the formalization of microbial metabolism. In the following section, we review models intended to predict microbial growth yield by accounting for the energy and matter balance in growth. We then present models that apply thermodynamic considerations for the prediction of microbial population growth dynamics.

1. The framework used to describe the model

Each microbial thermodynamic model introduced its own semantics and concepts. In order link the models and hence to present the state of the art of this field in a unified overview, in this section, we introduce semantic and mathematic conventions used to present the models in the two following sections.

Symbol	**Meaning**
$Y_{i/j}^{r}$	Stoichiometric coefficient of reagent i in reaction r, expressed as $mol_i.mol_j^{-1}$. "/j" is optional and denotes the normalization of the coefficient

	$Y^r_{i/j} = {}^1/_{Y^r_{j/i}}$. For example, $Y^{metabolism}_{X/S}$ is the stoichiometric coefficient of biomass X in the metabolic reaction, expressed per mole of substrate S, in other word, the growth yield on substrate S.
Y^r	Vector of stoichiometric coefficients for all the reagents of reaction r
ΔG^0_r	Gibbs energy change of reaction r in standard conditions (temperature of 298 K, pressure of 1 hPa and concentration of 1 M for all solute reagents and 1 bar for all gasses)
$\Delta G^{0\prime}_r$	Gibbs energy change of reaction r in standard conditions except that [H^+] = 1e-7 M
ΔG_r	Gibbs energy change of reaction r in non-standard conditions
m_G	Energy flow to fulfill the maintenance cost for growth (kJ.C-mol$_{biomass}$.hour^{-1})
m_i	Flow of chemical species i to fulfill the maintenance cost for growth (mol$_i$. biomass.hour^{-1})
q_i	Absorption rate of chemical species i by a microbial population
γ_i	Reduction degree of chemical species i using CO2, NH4+1 and H2O as reference state for C, N, and O respectively ($\frac{4\cdot C+1\cdot H-2\cdot O-3\cdot N-charge}{C}$) (mol$_e$.C-mol$_i^{-1}$). If the specie i has no carbon, the denominator is 1
μ	Specific growth rate of a population (time^{-1})

As an illustration, we use an 'aerobic glucosotroph population', consuming glucose as its sole carboneous substrate. Such an idealized set of cells treated as a single population and identified by their catalytic activity is called a 'guild'. Examples of such calculations are provided in indented, italicized paragraphs throughout the descriptions of the models.

1.1. Stoichiometric formulation of the metabolism

Most microbial growth models that account for thermodynamic effects require balancing energy and matter exchanges during metabolic reactions. It is usually assumed that cell activity can be reduced to

anabolism (biomass synthesis), maintenance and catabolism, whose purpose is to fulfill the energy requirement of the former two processes. Cells are then visualized as "energy converters", capturing chemical energy from their environment and transforming it into biomass. The estimation of the energy involved in anabolism and catabolism is made possible by reducing them to chemical reactions with explicit stoichiometry, whereas maintenance generally does not require a specific chemical reaction, but is simply assumed to be fulfilled by some part of the catabolism's energy.

Basically any exergonic chemical reaction can be considered as a possible catabolic reaction. However, catabolic reactions of respiratory metabolisms are usually considered as electron transfer between a donor D and an acceptor A. Such reactions can then be built easily by combining pre-computed electron donation and acceptance reactions.

For example, the catabolic reaction of the glucosotroph population can be devised by combining electron donation by glucose (-1 $C_6H_{12}O_6$ -6 H_2O +6 CO_2 +24 H^+ + 24 e^-) with electron acceptance by oxygen (-4 e^- -1 O_2 -4 H^+ + 2 H_2O), which gives:

$-6\ O_2\ -1\ C_6H_{12}O_6\ +6\ H^+\ +6\ CO_2^-\ (\Delta G^{0'}_{cat} = -2917.8\ kJ.mol_D^{-1})$	1

The negative stoichiometric coefficients denote the species consumed while the positive ones denote the species produced. Here the reaction has been normalized per mole of donor; however the chemical species used for normalization is a convention specific to each model.

In order to represent anabolism as a single reaction, most microbial thermodynamic models use a fictional generic "biomass molecule", traditionally denoted "X", as proxy for the variety of molecules of which the biomass is actually made. Examples of the biomass molecules used in the models described below are:

- $C_1H_{1.4}O_{0.4}N_{0.2}$ ($C_5H_7O_2N$): mean composition of a mixed microbial culture growing on dairy waste (Hoover et al 1953). Widely used in bioprocess modeling (such as in ASM (Henze et al 1987)), and also in microbial thermodynamics models such as McCarty's (McCarty 1965) (reviewed below). Its Gibbs energy of formation is estimated at 18.8 kJ.mol^{-1}.
- $C_1H_{1.8}O_{0.5}N_{0.2}$: mean composition of multiple biomass formulas compiled from the literature (Minkevich and Eroshin 1973). Used in Heijnen's model (Heijnen et al 1992) (described below). Its Gibbs energy of formation is estimated at -67 kJ.mol^{-1}.
- $C_1H_{1.613}O_{0.557}N_{0.158}$: composition of a pure culture of Saccharomyces *cerevisiae* (Battley et al 1997). Its Gibbs energy of formation was directly measured by calorimetry as -82.16 kJ.mol^{-1}.

The models described in detail below use different methods to build the anabolic reaction.

As a simple example, if we consider $C_1H_{1.8}O_{0.5}N_{0.2}$ biomass, and that the glucosotroph population uses glucose as carbon source and ammonium as nitrogen source to synthesize biomass, its anabolic reaction can be devised as:

$-0.175\ C_6H_{12}O_6 - 0.2\ NH_4^+ + 0.05\ CO_2 + 0.2\ H^+ + 0.45\ H_2O + 1\ C_1H_{1.8}O_{0.5}N_{0.2}\ (\Delta G^{0\prime}_{an} = 24.18\ kJ.mol_X^{-1})$	2

It is worth noting that a detailed model of anabolism was recently proposed (LaRowe and Amend 2016).

Assuming that metabolism consists only in those two reactions, let us define the vector Y^{met} of the stoichiometric coefficients of the overall metabolic reaction, as being a linear combination of the vectors of catabolic and anabolic reaction; Y^{cat} and Y^{an}. The coefficients of this linear combination will be one crucial point of divergence between the models presented in the following subsection.

1.2. Conventions in microbial population kinetics

Since the 1950s, microbial population dynamics has often been modeled as an ordinary differential equation (ODE) system formulated as:

$$\begin{cases} \dot{X} &= X \cdot \mu(\cdot) \\ \dot{S} &= -\dfrac{1}{Y_{X/S}} \cdot X \cdot \mu(\cdot) \end{cases} \qquad (3)$$

where X is the biomass concentration, S the substrate concentration, $\mu(\cdot)$ any biomass specific growth rate function and $Y_{X/S}$ the growth yield, expressed in units of X produced per unit of S consumed. In order to account for the concept of microbial maintenance (Herbert and Elsworth 1956); (Pirt 1965); (Pirt 1982), the system known as the "Herbert-Pirt" growth model is expressed as:

$$\begin{cases} q_S = q_S^{max} \dfrac{S}{K_S + S} = \dfrac{1}{Y_{X/S}} \cdot \mu + m_S \\ \mu = Y_{S/X}(q_S - m_S) \\ \dot{X} = X \cdot \mu \\ \dot{S} = -\dfrac{1}{Y_{X/S}} \cdot X \cdot \mu \end{cases} \qquad (4)$$

where m_S is the flow rate of S needed for maintenance (positive), q_S the biomass specific absorption rate of S (positive) and $Y_{X/S}^{max}$ the maximum growth yield, that is, the value of $Y_{X/S}$ if $m_S = 0$. Here it is assumed that the relationship between S and q_S follows Monod's law, but any other growth law could have been used. Interestingly, if the substrate concentration S is lower than $\frac{m_S \cdot K_S}{q_S^{max} - m_S}$, the growth rate μ becomes negative, which reverses the direction of the growth reaction.

Microbial population growth models that account for thermodynamics often fall into this ODE framework and amount to providing an energy-dependent expression of $Y_{S/X}$, q_S and/or m_S.

2. Thermodynamic prediction of microbial growth yield

2.1. TEEM model

The thermodynamic electron equivalent model (TEEM) developed by McCarty is the earliest model to link the yields of microbial populations to thermodynamic variables. Its first formulation (retrospectively called TEEM1) dates back to 1965 (McCarty 1965), while its most recent version (called TEEM2) dates from 2001 (McCarty 2007). This model was intended to be applied to microbial bioprocesses such as activated sludge. It has also been used in an individual-based model to study denitrifying pathways (Araujo et al 2015, Araujo et al 2016), and coupled with chemical speciation computations to predict equilibria in biogeochemical systems (Istok et al 2010).

The TEEM model normalizes metabolic reactions per electron transferred; so they are said to be expressed in electron-equivalent (eeq), hence the name of the model. Reactions involved in the TEEM model are graphically illustrated in figure 1a.

Using our previously defined glucosotroph population as an illustration, its catabolic reaction (called "R") can be formulated by combining electron donation and acceptance "d" and "a";

$d: -\frac{1}{4} CO_2 - 1\ H^+ - 1\ e^- + \frac{1}{24} C_6H_{12}O_6 + \frac{1}{4} H_2O\ (\Delta G^{0'}_d = 38.83\ kJ.eeq^{-1})$	5

$a: -\frac{1}{4} O_2 - 1\ H^+ - 1\ e^- + \frac{1}{2} H_2O\ (\Delta G^{0'}_a = -78.72\ kJ.eeq^{-1})$	6

(Conventionally, both "d" and "a" are expressed as reduction reactions)

$R = a - d: -\frac{1}{4} O_2 - \frac{1}{24} C_6H_{12}O_6 + \frac{1}{4} CO_2\ (\Delta G^{0'}_R = -117.55\ kJ.eeq^{-1})$	7

The R reaction is assumed to provide energy to the anabolism with a constant "energy transfer efficiency" ϵ, which is the proportion of ΔG_R used in anabolism, while the other portion is used for

maintenance or dissipated out of the biomass. By comparing the model with experimental yield measurements in both aerobic and anaerobic metabolisms, McCarty and collaborators estimated ϵ to lie between 0.37 and 0.39 for heterotroph metabolisms, and between 0.5 and 0.6 for autotrophs and anaerobic denitrifiers (McCarty 2007). Support for the value interval for heterotrophs was later provided by another study involving 123 experimental yield measurements on 38 different metabolisms (Roden and Jin 2011).

In the TEEM model, the anabolic reaction (called "S") is formulated as the sum of two sub-reactions; "S = IC + PC". The "IC" reaction is the conversion of the carbon source into an intermediary component (IC), while the "PC" (pyruvate to cell) reaction is the conversion of the intermediary component into the biomass molecule. The IC reaction is formulated as "in – d" where "in" is the half-reaction of reduction of the intermediate compound. The intermediary molecule is chosen as a central molecule in any cell metabolic network; in the first version of the TEEM model, pyruvate ($\Delta G^{0\prime}_{in}$ = 35.09 kJ.eeq-1) was considered, later acetyl-coA ($\Delta G^{0\prime}_{in}$ = 30.9 kJ.eeq^{-1}) was used instead, to increase theoretical accuracy rather than to improve prediction (McCarty 2007).

Concerning the glucosotroph population, we then have $\Delta G^{0\prime}_{IC} = 30.9 - 38.83 = -7.93$ kJ.eeq^{-1}.

The $\Delta G^{0\prime}_{PC}$ only depends on the intermediary component and on the biomass molecule considered. Considering acetyl-coA as intermediary component and $C_5H_7O_2N$ as biomass, we have $\Delta G^{0\prime}_{PC}$ = 18.8 kJ.eeq^{-1}(the reduction degree of the intermediary component must be the same as this of biomass for this relation to be valid) (McCarty 2007).

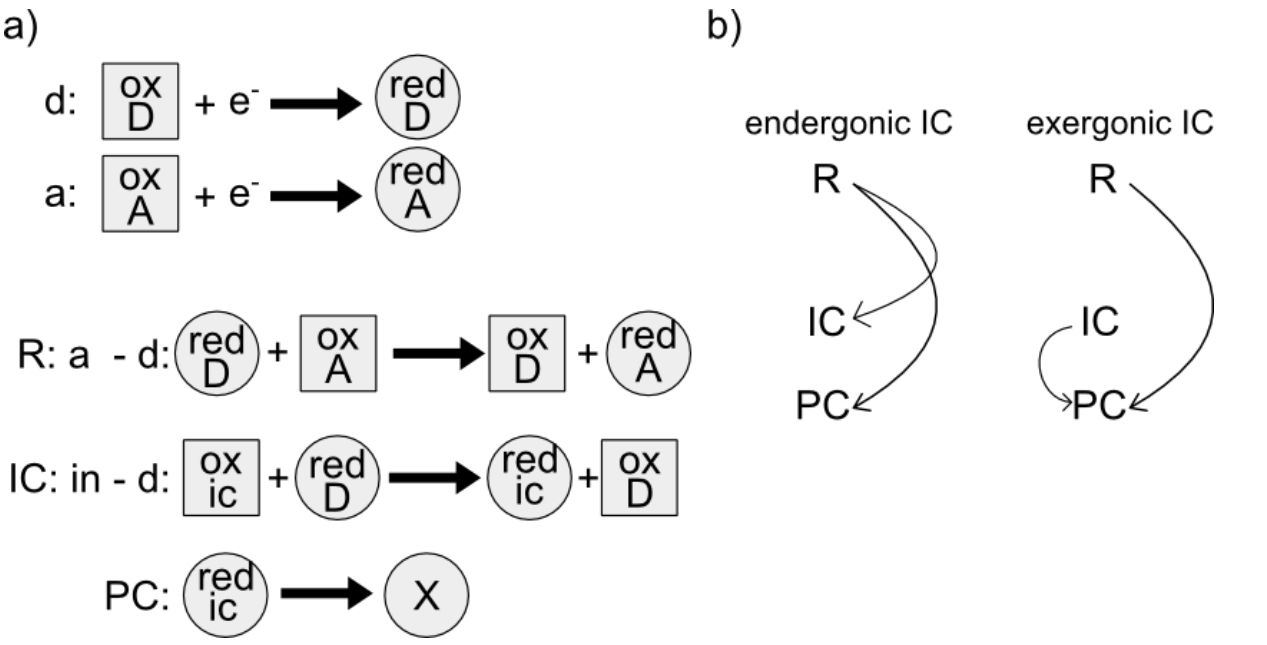


Figure 1: Graphical illustration of some concepts of the TEEM model. a) graphical description of reactions involved in McCarty's model. "D" stands for electron donor, "A" for electron acceptor, "ic" for intermediary compound "pc" for "pyruvate to cell", "ACo" for acetylCoA and "X" for biomass. Squares denote oxidized molecules while circles denote reduced molecules. b) Illustration of the energy transfers from the catabolic reaction "R" to the anabolic reactions ("IC" and "PC") in the two different anabolism configurations according to the TEEM model. If IC is endergonic, the R reaction fuels both IC and PC, and if IC is exergonic, both IC and R fuel the PC reaction.

The point in calculating sub-reactions of S in the TEEM model is that the expression of ΔG_S depends on the sign of the Gibbs energy change associated with the IC reaction, as illustrated by figure 1b; if it is positive (endergonic reaction) then the catabolic reaction (R) provides energy for both the IC and the PC reaction. However, if it is negative (exergonic reaction), it provides energy for the PC reaction, along with R, with an efficiency of ϵ. The Gibbs energy change of anabolism S is then expressed as:

$$\begin{cases} \Delta G_S = \dfrac{\Delta G_{IC} + \Delta G_{PC}}{\epsilon} & if \Delta G_{IC} > 0 \\ \Delta G_S = \epsilon \Delta G_{IC} + \dfrac{\Delta G_{PC}}{\epsilon} & if \Delta G_{IC} < 0 \end{cases} \quad (8)$$

Following this scheme, supplementary refinements of the TEEM2 broke the anabolic reaction down into sub-reactions for metabolisms involving oxygenase or single-carbon organic compounds in order to improve the predictive accuracy of the model. VanBriesen and Rittman developed another extension to take intermediate metabolites into account in the model (VanBriesen and Rittmann 2000). As the TEEM model uses electron transfer as a unit and improves its predictive power by further detailing the description of the metabolic pathways. It can be used to investigate the effect of the structure of metabolic pathways on microbial growth yield, but requires biochemical knowledge and therewith cannot be regarded as a black box model anymore.

As our glucosotroph population is a simple use case, it does not need such refinements. Assuming an efficiency ϵ of 0.39, $\Delta G^{0\prime}{}_S$ is 45.11 kJ.eeq^{-1}.

Once ΔG_R, ΔG_S and ϵ are known, we can compute the number A of times the R reaction must be performed for S to run once:

$$A = -\frac{\Delta G_S}{\epsilon \cdot \Delta G_R}$$	9

This value is 0.98 for the glucosotroph population under standard conditions and a pH of 7.

A represents the ratio of the proportion f_e^0 of electron donor channeled into catabolism over the proportion f_s^0 of electron donor channeled into biomass synthesis:

$$A = \frac{f_e^0}{f_s^0}, f_s^0 = \frac{1}{1+A}, f_e^0 = \frac{A}{1+A}, f_e^0 + f_s^0 = 1$$	10

The proportions f_e^0 and f_s^0 are both proportions of electron equivalent (eeq) of the electron donor. This ratio varies as the energy per eeq depends on the metabolism considered.

For glucosotrophs it is then 0.50 for f_s^0 and 0.50 for f_e^0, which means approximately one electron is transferred in the anabolic pathway for each electron transferred in the catabolic pathway.

The growth yield $Y_{X/S}$ (in $mol_X.mol_D^{-1}$) can be computed from the f_S^0 proportion by multiplying it by the ratio of the degree of reduction of the electron donor and the biomass:

$$Y_{X/S} = f_s^0 \cdot \frac{\gamma_D}{\gamma_X}$$	11

The growth yield predicted for the glucosotroph population is then $0.5 \cdot \frac{4}{5} = 0.4$ $mol_X.mol_D^{-1}$, or 2.0 C-$mol_X.mol_D^{-1}$ as this biomass has 5 carbons.

This method is the same for both autotroph and heterotroph populations.

While this energy-based method is successful at predicting the yield of many metabolisms (Roden and Jin 2011), and is still used today, it has some intrinsic limitations, such as the need for intracellular information to model some metabolisms. Moreover, while the efficiency parameter ϵ was initially assumed to have the same value for every metabolism, this was later disproved by experimental observations (McCarty 2007). The model introduced in the following subsection is an alternative to the TEEM model, for which only the input and outputs of the cell need to be known. In this model, the relationship between dissipated energy (and therefore, implicitly, energy transfer efficiency) is linked to physicochemical properties of the metabolism.

2.2. Gibbs energy dissipation method

In the early 1990s, Heijnen and collaborators reviewed many previous methods used to predict microbial growth yield and found that they could not be applied to all metabolisms as they had internal inconsistencies, or required knowledge about the way metabolism occurs inside the cells, or could work only for a small subset of all possible chemotrophic metabolisms. By analyzing various growth data on chemotrophic microbial cultures collected from the literature, the authors came to the conclusion that the Gibbs energy dissipated by the biomass produced is the most appropriated thermodynamic quantity to use to predict the yield of a microbial population, compared with previous approaches (see (Heijnen and Dijken 1991)) for an in-depth comparative review between those methods and the method developed by Heijnen and collaborators).

The dissipated energy is the Gibbs energy change of an overall growth reaction (including anabolism and catabolism). If all the energy of catabolism was invested into anabolism, the overall Gibbs energy change would be zero. However, such reaction would not be spontaneous, as the Gibbs energy change being negative is the thermodynamic criterion of feasibility for a reaction. Consequently, some of the catabolism's energy must be dissipated by the cell, for example as heat.

Heijnen and collaborators concluded that the Gibbs energy dissipated per biomass produced mostly depends on the number of carbons and on the degree of reduction of the carbon source. This discovery led to a formulation of the energy balance of microbial growth based solely on the chemical definition of the metabolism catalyzed (Heijnen and Dijken 1991, Heijnen et al 1992, Heijnen 2010). Later, Liu and collaborators calibrated a more accurate relationship on a larger experimental dataset, based solely on the reduction degree γ_D of the electron donor ($mol_e.C\text{-}mol_D^{-1}$), and applicable to aerobic metabolisms:

$\begin{cases} \Delta G_{dis} = \dfrac{-666.2}{\gamma_D} - 243.1 \; if \; (\gamma_D \leq 4.67) \\ \Delta G_{dis} = -157 \cdot \gamma_D + 339 \; if \; (\gamma_D > 4.67) \end{cases}$	12

ΔG_{dis} is a negative value expressed in in kJ.C-mol$_X^{-1}$.

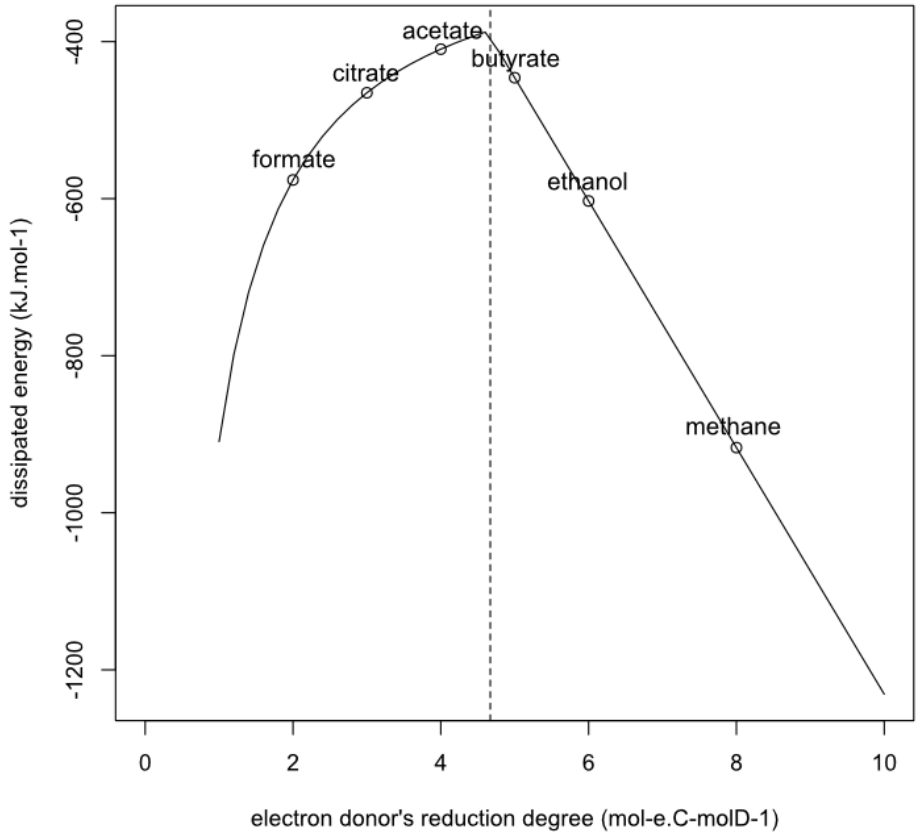


Figure 2: Relationship between ΔG_{dis} (dissipated energy per biomass produced) and the degree of reduction of the electron donor

When applying this relation to the glucosotroph population, considering glucose as its carbon source, its ΔG_{dis} is -409.65 kJ.C-mol$_X^{-1}$.

The λ factor representing the number of times the catabolic reaction must run per anabolic reaction in order to close the energy balance of growth is then expressed as:

$$\lambda = \frac{\Delta G_{dis} - \Delta G_{an}}{\Delta \mathrm{G}_{cat}}$$	13

λ is similar to the "A" variable of the TEEM model described above, but using a different normalization. It can be used to adjust the stoichiometry of the overall growth reaction to fit the energy balance:

$$Y_{i/X}^{max} = Y_{i/X}^{an} + \lambda \cdot Y_{i/D}^{cat}$$	14

where Y_i^{max} is the stoichiometric coefficient of reagent i in an energy-balanced metabolism with zero maintenance. This relation holds true for both reagents and Gibbs energy; $\Delta G_{dis} = \Delta G_{an} + \lambda \cdot \Delta \mathrm{G}_{cat}$, hence

As an example, the λ factor of the glucosotroph population in standard conditions accounting for a pH of 7 would then be $\frac{-409.65+24.18}{-2917.8} = 0.132$ $mol_D.mol_X^{-1}$. Knowing the stoichiometric coefficients of catabolism and anabolism, the growth yield of the glucosotroph population on glucose can be computed ($\frac{1}{0.132+0.175} = 3.26$ $mol_X.mol_D^{-1}$), along with the yield on any consumed species, such as the yield on oxygen ($\frac{1}{0.132 \cdot 6} = 1.26$ $mol_X.mol_{O2}^{-1}$).

Heijnen and collaborators noted that their ΔG_{dis} expression does not hold for some categories of autotrophs. Instead, they explain that when the electron donor used in autotrophic metabolism is thermodynamically strongly unfavorable (which holds true for electron donors like Fe^{+2}, NH_4^+ and NO_2^-), the cell must invest energy in a process called "reverse electron transfer". The upkeep of this process sets ΔG_{dis} to -3500 kJ.mol_X^{-1}. The ΔG_{dis} of the other types of autotroph microorganisms is deemed to be -986 kJ.mol_X^{-1}. Liu and collaborator also apply the same exception when using their own formula (equation 12).

Later, the experimental data collected by Heijnen and collaborators to draw their relation between the dissipated Gibbs energy and the characteristics of the carbon source of the metabolism was enriched by other authors who came to the conclusion that considering a dissipated energy per biomass produced of 500 $kJ.mol_X^{-1}$ explains experimental results almost as well as using the previously defined relation (equation 12) (Liu et al 2007). While the two methods are formally different, it has been demonstrated that they mathematically boils down to very comparable calculations, and thus that the yields predicted from the dissipated energy and the TEEM approach have been shown to fit equally well on the sets of experimental data for various metabolisms (Kleerebezem and Van Loosdrecht 2010).

In the next section we present dynamic thermodynamic growth models, including the kinetic extension of Heijnen's model, where the maintenance energy requirements are taken into account.

3. Dynamic thermodynamic models of microbial growth

3.1. Heijnen's dynamic model

This dynamic model was proposed by Heijnen and collaborators in association with their work on the energy balance of microbial growth (Heijnen 2010) described in subsection 2.2. This model consists in using the energy dissipation model to compute the values of maximum growth yield $Y_{X/S}^{max}$ ($mol_X.mol_S^{-1}$) and the maintenance flux m_S ($mol_S.mol_X^{-1}.hour^{-1}$) in the Herbert-Pirt model (see equation 4) so the microbial dynamics account for the energy balance.

The description of the energy dissipation model in the subsection 2.2. provided an empirical formulation for $Y_{X/S}^{max}$ based on the dissipated energy. In this subsection, we detail the computation of the maintenance substrate consumption rate m_S and how all the terms are included in the Herbert-Pirt framework.

Heijnen and collaborators' energy balance model states that the Gibbs energy change associated with the overall growth reaction ΔG_{ogr} (in kJ.mol$_X^{-1}$) equals the dissipated Gibbs energy of growth ΔG_{dis} plus the maintenance energy $\frac{m_G}{\mu}$ that depends on the growth rate μ of the population:

$$\Delta G_{ogr} = \Delta G_{dis} + \frac{m_G}{\mu}$$	15

where m_G is the Gibbs energy flow required for maintenance in kJ.mol$_X^{-1}$.hour^{-1}, and μ is the growth rate of the population in mol$_X$.mol$_X^{-1}$.hour^{-1}.

Having calibrated their parameters on experimental data, Heijnen and collaborators consider that m_G follows an Arrhenius law, which means it only dependent on temperature, and equals on average -4.5 kJ.mol$_X^{-1}$.hour^{-1} at 298 K. The rate m_D of electron donor absorption for the purpose of maintenance (in mol$_D$.mol$_X^{-1}$.hour^{-1}) can then be expressed as:

$$m_D = \frac{m_G}{\Delta G_{cat}}$$	16

With ΔG_{cat} in kJ.mol$_D^{-1}$. Applying this formula to compute the glucosotroph population's m_D, we find $\frac{-4.5}{-2917.8} = 1.54e-3$ *mol$_D$.mol$_X$.hour^{-1}.*

The default Herbert-Pirt ODE framework uses the empirical Monod growth formula to describe the link between the substrate concentration and its absorption rate. In addition, it proposes a way to estimate the upper limit of the substrate absorption rate (q_D^{max}) and the maximum growth rate (μ^{max}) required by this mathematical description of growth. Indeed, based on prior observations (Andersen and Von Meyenburg 1980), Heijnen and collaborators suggest that a maximum electron absorption rate exists q_e^{max} (in mol$_{electron}$.mol$_X^{-1}$.hour^{-1}), which follows an Arrhenius law ($q_e^{max} = -3 \cdot \left(\frac{-69 \cdot 10^3}{R} \cdot \left(\frac{1}{T} - \frac{1}{298}\right)\right)$)

and equals -3 $mol_{electron}.mol_X^{-1}.hour^{-1}$ at 298 K. This sets the maximum Gibbs energy absorption rate q_G^{max}:

$$q_G^{max} = q_e^{max} . \frac{-\Delta G_{cat}}{\gamma_D^*}$$	17

where γ_D^* is the number of electrons transferred during the catabolic reaction (in $mol_{electron}.mol_D^{-1}$). In the case of a respiration, γ_D^* is the degree of reduction of the electron donor multiplied by its number of carbon (if any). In the case of a fermentation it depends on the degree of reduction of the products. This consequently sets μ^{max} as:

$$\mu^{max} = \frac{q_G^{max}}{\Delta G_{dis} - \Delta G_{an}}$$	18

where ΔG_{dis} is the dissipated energy computed in equation 12 and finally this sets the maximum donor absorption rate as:

$$q_D^{max} = Y_{D/X}^{max} \cdot \mu^{max} + m_D$$	19

In the case of the glucosotroph population, $q_G^{max} = -3 \cdot \frac{2917.8}{24} = -364.72$ *$mol_D.mol_X^{-1}.hour^{-1}$, then* $\mu^{max} = \frac{-364.72}{-409.65+24.18} = 0.95$ *$hour^{-1}$, and finally* $q_D^{max} = \frac{1}{3.26} \cdot 0.95 + 1.54e-3 = 0.29$ *$mol_D.mol_X.hour^{-1}$.*

The ΔG_{dis} and m_D terms can be used in Herbert-Pirt's ODE system (equation 4):

$$\begin{cases} q_D = q_D^{max}\dfrac{[D]}{K_D + [D]} = \dfrac{1}{Y_{X/D}^{max}} \cdot \mu + m_D \\ \mu = Y_{X/D}^{max}(q_D - m_D) \\ [\dot{X}] = [X] \cdot \mu \\ [\dot{D}] = -\dfrac{1}{Y_{X/D}} \cdot [X] \cdot \mu \end{cases}$$	20

(ΔG_{dis} is not directly included in the equation, but is used to compute $Y_{X/D}^{max}$)

To summarize, in this model, the growth rate of a population is expressed provided an estimation of the power of the population maintenance (in $kJ.mol_X^{-1}.hour^{-1}$) and dissipation (both based on empirical fittings on experimental data) are available, and assuming a Monod relationship between the population's energy source D present in the environment and its absorption rate by the population.

One advantage of this approach is that it makes it possible to simulate the dynamics of any chemical species included in the metabolic equations.

The model was implemented with some extensions by González-Cabaleiro and collaborators to support the energy-based explanation of patterns observed in wastewater microbial communities (Gonzalez-Cabaleiro et al 2015). González-Cabaleiro and collaborators devised all possible anaerobic metabolic reactions in wastewater, and then computed the growth yield and rate associated with those metabolisms using the method of Heijnen and collaborators described above with some modifications. The modifications consisted in computing the absorption rate q_S of every chemical species S that can be consumed and using the lowest q_S in the expression of μ. The purpose is to apply Liebig's rule of the single limiting substrate (which states that, at any given instant of time, population growth is limited by the scarcest available resource).

The approach proposed by Heijnen and collaborators only requires a description of the inputs and outputs of metabolism to predict microbial growth yield and growth rate, which means it defines

parameters that only characterize the growth process in terms of input and output. Since then, alternative approaches have been developed to model the influence of thermodynamics on microbial population dynamics.

3.2. Equilibrium-based model

The model proposed by Hoh and Cord-Ruwisch, named "equilibrium-based model" by its authors, associates enzyme kinetics modeling and thermodynamic considerations (Hoh and Cord-Ruwisch 1996). It does not comply with the "cell-as-energy-converter" presented in subsection 1.1. In particular, it does not consider an anabolic reaction and thus focuses on modeling catabolism.

The authors make the assumption that the growth rate of a microbial population mostly depends on the kinetics of an enzyme-catalyzed reaction. Their reference to enzymology is based on a theoretical consideration of microbial growth, as opposed to Monod's growth law, whose resemblance to enzyme kinetics is purely coincidental (Monod 1949). The forward rate ν of a product-inhibited enzyme-catalyzed reaction is known in enzymology (Haldane 1930) as:

$$\nu = \frac{\nu_{max} \cdot (S - \frac{P}{K})}{k_{m(S)} + S + \frac{k_{m(S)}}{k_{m(P)}}} \quad (21)$$

ν_{max} being the maximum growth rate, $k_{m(S)}$ the affinity constant of the enzyme for the substrate, $k_{m(P)}$ the affinity constant of the enzyme for the product (for the reverse-reaction) and K the equilibrium constant of the reaction. K allows for the inclusion of thermodynamic considerations in this model as it can be linked to the Gibbs energy differential of the reaction through the following relationship:

$$\Delta G' = RTln\frac{\Gamma}{K} \quad (22)$$

where $\Delta G'$ is the Gibbs energy differential of the reaction corrected for non-standard concentrations, and Γ is the mass action ratio of the reaction. Therefore $\frac{\Gamma}{K}$ expresses the distance of the reaction from equilibrium. The authors reformulate the growth rate as:

$$\nu = \frac{\nu_{max} \cdot (S - \frac{P}{K})}{k_{m(S)} + (S + \frac{P}{K})}$$	23

The advantage of this formulation is that it explains the $\frac{P}{K}$ threshold for the substrate concentration S above which $\left(S - \frac{P}{K}\right) = 0$, i.e. growth is thermodynamically impossible. In this model, this impossibility is caused by product inhibition, itself caused by departure from thermodynamic equilibrium. The existence of this non-zero concentration threshold for growth, while not captured by Monod's growth law, is more in agreement with experimental observations (Kovarova-Kovar and Egli 1998).

This model was used by Großkopf and Soyer (Großkopf and Soyer 2016) to provide a thermodynamic refutation of the competitive exclusion principle (two species that compete for the same limiting resource cannot coexist) (Gause 1934, Hardin 1960). In their article, the authors demonstrate that a model taking the thermodynamic effect of inhibition on microbial growth rate into account allows for the coexistence of multiple guilds in competition for the same substrate but that break down it into different end products.

As a side note, the hypothesis that microbial growth rate is regulated by a single enzyme-catalyzing reaction is also implemented by another microbial thermodynamic model by Corkrey and collaborators (Corkrey et al 2012, Corkrey et al 2016). This model makes it possible to predict the temperature-dependence of the growth rate of microorganisms from each life domain.

3.3. Jin and Bethke's model

In the 2000s, Jin and Bethke proposed an expression of the rate of microbial reactions based on a model of the microbe's respiratory chain and nonlinear nonequilibrium thermodynamics (Jin and Bethke 2002, Jin and Bethke 2003, Jin and Bethke 2005). An adaptation of the model to fermentative metabolisms exists (Liu et al 2007) but is not detailed here. Similar to the "equilibrium-based model" presented in subsection 3.2., this model does not consider the "cell-as-energy-converter" paradigm but only models the catabolic reaction.

The catabolic reaction is modeled as a combination of multiple, reversible sub-reactions, each one being able to occur multiple times independently from the others. The overall electron transfer reaction can be written similarly to those in the other models presented in this review. Let ΔG_{cat} be its Gibbs energy change and n the number of electrons transferred during the reaction. The reaction is coupled to the reaction of phosphorylation of ADP into ATP. The Gibbs energy change ΔG_P of this reaction is assumed to be 50 kJ.mol^{-1} (Jin and Bethke 2005). Let m be the ratio of ATP molecules produced per electron transferred during the overall electron transfer, the overall catabolic reaction is then:

$$\nu_D D + \nu_A A + m\ ADP + mP_i \rightleftharpoons \nu_{D^+} D^+ + \nu_{A^-} A^- + m\ ATP$$	24

where ν_D, ν_A, ν_{D^+} and ν_{A^-} are the stoichiometric coefficients of the electron donor D, acceptor A, and their oxidized and reduced forms respectively. When implementing their model for simulations, the authors consider constant, guild-specific, calibrated values for m (ranging from $^1/_4$ to $^5/_4$ $mol_{ATP}.mol_D^{-1}$ (Jin, Bethke et al. 2011)). However they admit that microbes are able to adjust the value of m to a certain extent, which would modify their growth yield/rate ratio (Jin and Bethke 2003).

The authors consider the "thermodynamic drive" of this reaction to be:

$$f = -\Delta G_{cat} - m\,\Delta G_P$$	25

which is the inverse of the Gibbs energy change of this reaction. They then express the "thermodynamic potential factor" F_T of the reaction as;

$$F_T = 1 - exp\left(-\frac{f}{\chi RT}\right)$$	26

where χ is the "average stoichiometric number", a factor between 1 and +∞ accounting for the repetition of individual electron transfer steps during the overall electron transfer reaction. The F_T factor takes a value between -∞ and 1 (between 0 and 1 as long as $\Delta G_{cat} < 0$). Its sign indicates whether the respiration reaction is forward (+) or backward (-).

The authors also define F_D a kinetic factor accounting for the concentration of electron donors in their oxidized and reduced forms:

$$F_D = \frac{\prod_D [D]^{\beta_D}}{\prod_D [D]^{\beta_D} + K_D \cdot \prod_{Do} [\mathrm{Do}]^{\beta_{\mathrm{Do}}}}$$	27

with β_D and β_{D+}calibrated factors. They also define F_A, a kinetic factor accounting for the concentrations of electron acceptors, in the same way.

The expression of the respiration rate r of the microbial population is then ultimately linearly correlated with F_T , F_A and F_D;

$$r = k \cdot [X] \cdot F_D \cdot F_A \cdot F_T$$	28

where k is a calibrated constant and $[X]$ is the biomass concentration.

This model implies that some part of the energy produced by the electron transfer reaction is captured by the biomass as ATP (which covers cell synthesis and maintenance expenses), while the other part is (implicitly) dissipated and drives the catabolism. Interestingly, this model then implies a compromise between cell synthesis yield and rate, however the authors do not state whether metabolisms tend to

maximize growth yield, rate, or a compromise between both. Such tradeoff is also included in other models such as the one proposed by Kleerebezem and collaborators (cf box 1 of (Rodríguez et al 2008)).

This model has notably been used to simulate the kinetics of catalytic activities and gene expression in denitrification by a mixed culture microbial community (Li et al 2017).

The authors consider that the catabolism rate r can be used to produce an expression of the growth rate of a population only when growth is "energy limited". Indeed, this model does not provide an expression of the growth rate for a growth limited by a nutriment that is not related to catabolism (Jin and Bethke 2003). The model presented in the following section results from a different approach and provides an expression of the growth rate encompassing energy-limitation and stoichiometric limitation.

3.4. MTS theory

In contrast to other kinetic models described in this review, the microbial transition state (MTS) model (Desmond-Le Quéméner and Bouchez 2014) derives a new growth rate formula as an alternative to the empirical Monod equation, from the "cell-as-energy-converter" hypothesis and a probabilistic reasoning about microbial division. This model is analogous to the classical transition state theory, hence its name (Eyring 1935) (Figure 3a).

In this model, the cells are surrounded by a virtual "harvest volume" (V_H, $m^3.mol_X^{-1}$) in which the substrate is accessible to them. As pictured in figure 4a, the catabolic energy fuels the anabolic reaction and the dissipated energy (estimated using Heijnen's model, see subsection 2.2.), so the stoichiometry of the metabolism is adjusted to close the energy balance using a λ factor, as in Heijnen's model; $Y^{met} = Y^{an} + \lambda \cdot Y^{cat} = Y^{an} + \frac{\Delta G_{dis} - \Delta G_{an}}{\Delta G_{cat}} \cdot Y^{cat}$ (see subsection 2.2.).

Considering a set (s_1, s_2, ... s_i) of substrates of known concentrations, cells and substrates as particles randomly positioned in the culture medium (Figure 3b), it is possible to compute the probability of

finding enough of each substrate in a harvest volume V_H. From this probability, the ratio $\frac{N^{\ddagger}}{N}$ of cells that have enough of each substrate in their harvest volume to reach the "activated state", and thus to be able to divide, can be deduced and expressed as a function of the negative stoichiometric coefficient $Y_{S_i}^{met}$ of each substrate:

$$\frac{N^{\ddagger}}{N} = \prod_i exp\left(\frac{Y_{S_i}^{met}}{V_H \cdot [S_i]}\right) \quad (29)$$

where $[s]$ is the concentration of substrate s in mol.m^{-3}. The growth rate of the biomass is deduced from this proportion of “activated” cells (i.e. ready for division):

$$\mu = \mu_{max} \cdot \frac{N^{\ddagger}}{N} = \mu_{max} \cdot \prod_i exp\left(\frac{Y_{S_i}^{met}}{V_H \cdot [S_i]}\right) \quad (30)$$

where μ_{max} is a theoretical value in time^{-1}, related to theoretical microscopic kinetics, and not related to Monod’s law as the latter aggregates many different phenomena to which a precise fundamental, physical meaning has not yet been attributed (Liu 2007). The growth rate generated by equation 30 follows a response curve that differs slightly from that of Monod’s law, as shown in figure 4b. Indeed, its sinusoidal shape is more in line with experimental observations (Kovarova-Kovar and Egli 1998).

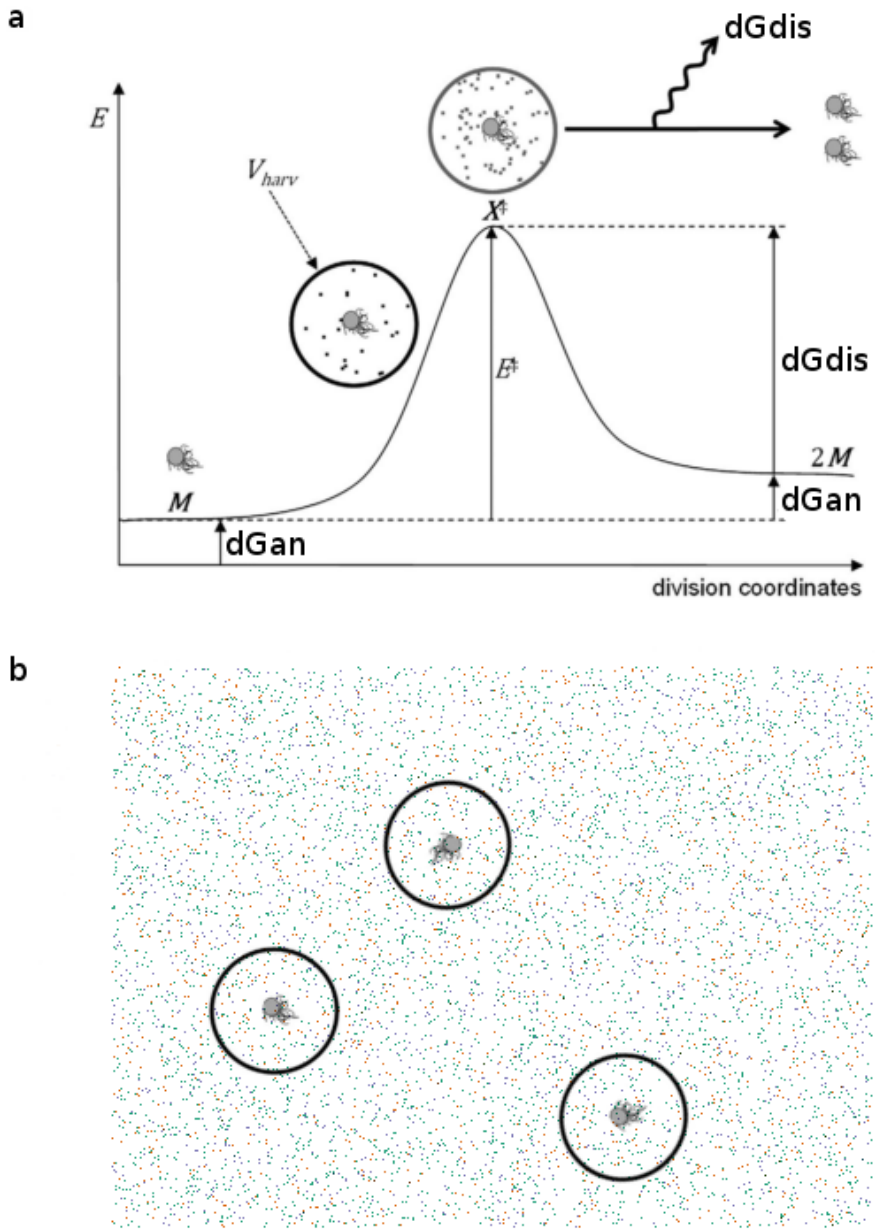
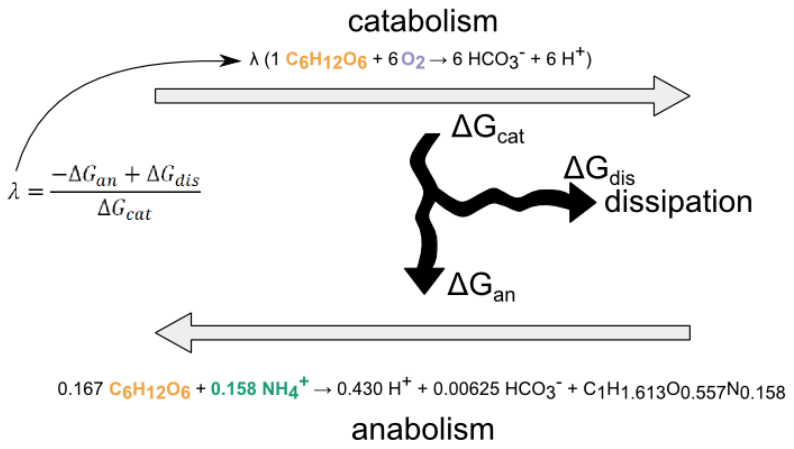


Figure 3 a) Graphical representation of microbial energy levels along division coordinates. Anabolic energy is augmented by the catabolic exergy within the harvesting volume symbolized as dots in a circle surrounding the microbe. When it reaches the threshold catabolic energy, the microbe is activated and an irreversible division process is triggered, associated with energy dissipation resulting in two microbes. b) Microbes in their culture environment, from the viewpoint of the MTS model. The different substrates are particles around the microbes and the microbes are surrounded by their harvesting volume, here symbolized as a black circle.

The λ factor (see equation 13) (and therefore the adjustment of the stoichiometry of the metabolism) varies dynamically. Thus the variations in concentrations during a dynamic simulation may affect the value of a guild's yield ($mol_X.molD^{-1}$) (though experimental investigations on Heijnen's energy balance model suggest that this effect should be negligible (Kleerebezem and Van Loosdrecht 2008)), transcribing the thermodynamic effect of the rarefaction of substrate or the accumulation of product on the energy balance of growth. In the MTS approach, microbial dynamics are therefore intrinsically dependent and coupled to the elemental and energy balance of microbial growth (Figure 4a).

a



$$\mu = \mu_{max} \cdot e^{\frac{\lambda \cdot Y^{cat}_{glucose} + Y^{an}_{glucose}}{V_H \cdot [glucose]}} \cdot e^{\frac{\lambda \cdot Y^{cat}_{oxygen}}{V_H \cdot [oxygen]}} \cdot e^{\frac{Y^{an}_{ammonium}}{V_H \cdot [ammonium]}}$$

b

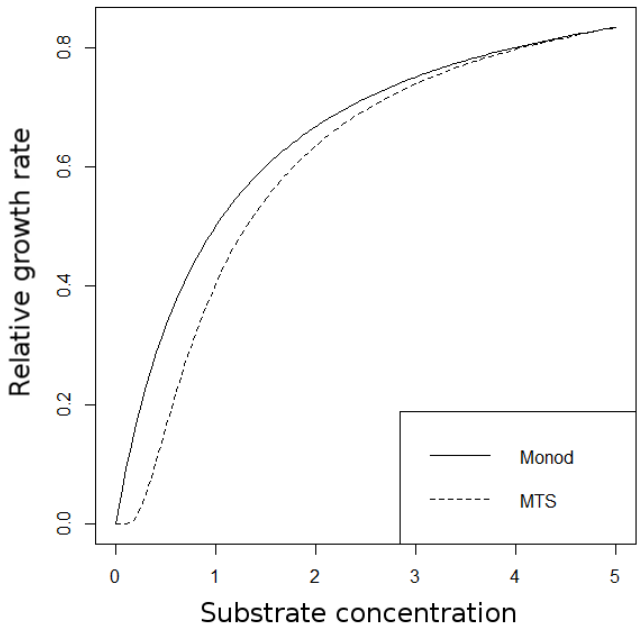


Figure 4: a) Visual representation of the cell-as-energy-converter model applied to the glucosotroph population. One portion of the energy ΔG_{cat} generated by the catabolic reaction is fueled into the energy requirement of anabolism (ΔG_{an}) while the remaining portion (ΔG_{dis}) is dissipated. The stoichiometry of metabolism is adjusted using the λ factor resulting from this energy balance model. The adjusted stoichiometric factor of each consumed chemical species is then taken into account in

the growth function. b) Comparison between Monod's growth rate function and the MTS growth rate function depending on the concentration of a single substrate. The Y axis represents the growth rate relative to the maximum growth rate ($\frac{\mu}{\mu_m}$) for both growth functions. (growth rate as a function of substrate concentration).

Practical implementations of the MTS model show that while starting from a simple microscopic description of microbial division, this model qualitatively predicts phenomena observed at the scale of communities such as ecological successions in guild population densities and time (cf chapter 4).

Conclusions and perspectives

This bibliography review introduced some representative microbial population growth models based on thermodynamics. The models aim to capture the influence of energy on microbial growth. Figure 5 shows the interconnection between the different aspects of microbial growth considered in the models presented. Multiple, possibly conflicting, explanations have been proposed for each relationship, as the actual mechanisms behind them are still not known. Investigation of the precise physical-chemical drivers of microbial growth will require careful experimentation. Hopefully, the comparison of thermodynamic growth models will produce precise hypotheses to guide such experiments.

Figure 5 shows the remaining challenges to advancing our understanding of the energy dependency of microbial growth. They are detailed below.

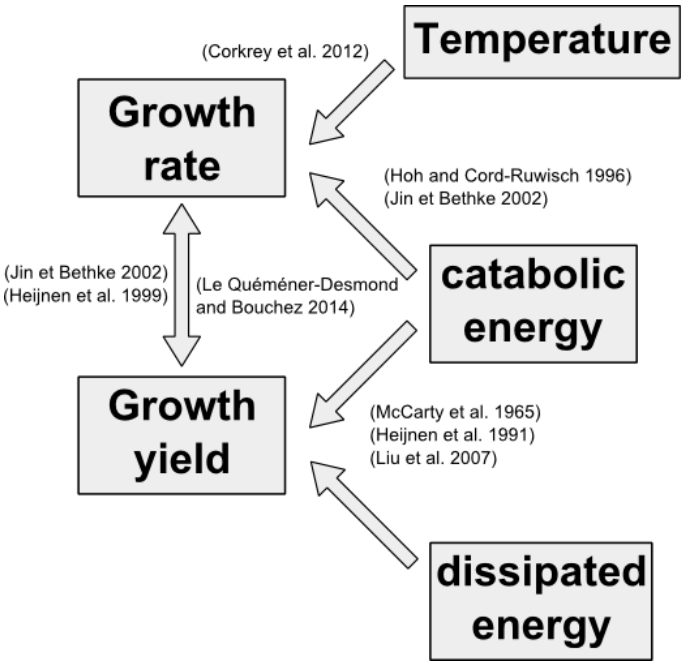


Figure 5: Conceptual links between variables involved in the energy dependence of microbial growth. Each arrow denotes a relationship between two variables. Bibliographic references indicate models proposed to characterize the relationships.

Concerning the relationship between the catabolic energy (ΔG_{cat}) and the growth rate (μ), multiple models came with different explanations. Recent models generally consider that the relationship between ΔG_{cat} and μ is non-linear. However they propose a variety of explanations to link those variables. For instance, the "equilibrium-based model" considers that the growth rate is limited by a single enzyme-catalyzed reaction, so the relationship ΔG_{cat} and μ should depend on enzyme kinetics. As another instance, Jin and Bethke's model describes a relationship between ΔG_{cat} and μ from non-linear non-equilibrium thermodynamics applied to a description of the electron transport chain. All those models involve different aspects of microbial growth (enzymology and equilibria, non-equilibrium thermodynamics, statistical physics). They do not necessarily contradict each other as those phenomena may be simultaneously at play during growth.

Regarding the relationship between growth yield and growth rate, many authors have reported that growth yield can vary and is often negatively correlated with growth rate (*i.e.* a high yield is generally

associated with a low growth rate, and vice versa). Many explanations have been proposed for this relationship (for a review see (Lele and G watve 2014)); one explanation is that microbial cells have some control over their metabolic efficiency. According to this explanation, dissipating more or less energy makes them adjust their yield/rate compromise. This allows them to develop strategies to adapt to their environment. The consequences of this hypothesis in terms of community structure were explored by Pfeiffer and collaborators (Pfeiffer et al 2001, Pfeiffer and Bonhoeffer 2002). According to this hypothesis, equilibria between multiple guilds can be expressed in terms of game theory, which adds subtleties to community structure. Comparatively, simulations of thermodynamic microbial growth models that do not allow such adaptation only reproduce competitive exclusion and the ensuing "redox tower", and fail to capture mutualistic relationships (Bethke et al 2011). However, most thermodynamic models used to predict growth yield (such as the TEEM model and Heijnen's dissipation model presented in this review) assume that the dissipated energy is determined only by the metabolism. For example, Heijnen's model considers dissipated energy to be broken down into a minimum, constant part ΔG_{dis}, and a dynamic maintenance part, depending on the growth rate μ (see equation 15); in this proposed energy balance, there is no scope for adaptation to the environment. Nevertheless, microbial thermodynamic models able to reproduce this tradeoff do exist. Examples are Jin and Bethke's model described in subsection 3.3. Another example is the model proposed by Kleerebezem and collaborators, which makes the dissipated energy a possibly adjustable variable, being positively correlated to the rate of the anabolism (Rodríguez et al 2008). Such models allow questioning whether the growth rate or the growth yield is the objective to be maximized by a microbial community. While there are many possible explanations for the yield rate tradeoff, satisfactory modeling of dissipated energy adjustment by microbial populations appears to be a necessary feature for microbial thermodynamic models to achieve a better representation of community assembly rules from population-scale principles.

To conclude, the way thermodynamics constrains microbial growth is complex and still under debate. The two above-mentioned challenges are focused on thermodynamic modeling of microbial growth at population scale, however, some rules at this scale have been shown to be able to reproduce community assembly patterns (Gonzalez-Cabaleiro et al 2015, Großkopf and Soyer 2016, Van de Leemput et al 2011) (also, see chapter 4). Moreover, the influence of thermodynamics on community structure can also be evidenced from experimental observations (Louca et al 2016a). While the lack of rules expressing invariants is a major problem in ecology (Lawton 1999), microbial thermodynamics as a discipline provides a generic framework for rules about growth and community structure to be proposed and carefully studied.

Acknowledgments

The authors wish to thank Dr Saifuddin Ahmed for very useful comments on the manuscript. The authors are also grateful to the *Région Ile de France* for funding Hadrien Delattre's PhD fellowship in the framework of DIM R2DS project, and to the "*Agence Nationale de la Recherche*" for financial support through the "THERMOMIC" project ANR-16-CE04-0003-01.

Chapter 3:

Material and methods

3.1 Nomenclature

symbol	meaning
$[i]$	concentration of chemical specie i ($\text{mol} \cdot \text{m}^{-3}$)
R	ideal gas constant ($8.314\,459\,8\ \text{J} \cdot \text{mol}^{-1} \cdot \text{K}^{-1}$)
T	system's temperature (K)
$Y^r_{i/j}$	stoichiometric coefficient of chemical specie i in reaction r. "$/j$" (optional) and denotes the chemical specie by which the reaction r is normalized
ΔG^0_r	Gibbs energy change of reaction r under standard conditions (temperature is 298 K and the concentration of every involved chemical specie is $1\ \text{mol} \cdot \text{L}^{-1}$
$\Delta G^{0'}_r$	Gibbs energy change of reaction r in conditions equivalent to standard conviction except that the concentration of protons is $1 \times 10^{-7}\ \text{mol} \cdot \text{L}^{-1}$
ΔG_r	Gibbs energy change of reaction r in any condition
γ_i	reduction degree of chemical specie i in $\text{mol}_\text{e} \cdot \text{C} - \text{mol}_\text{i}^{-1}$

3.2 Specifications of the MTS model

3.2.1 Units and dimensions

- Concentration in the model is expressed in $\text{mol} \cdot \text{m}^{-3}$ for aqueous species and as atmosphere for gaseous species. In articles, concentrations are nonetheless converted to $\text{mol} \cdot \text{L}^{-1}$ for the sake of practicality
- Time is expressed in day
- Energy is expressed in Joules
- Stoichiometric coefficients are positive when the reagent is produced and negative when the reagent is consumed

3.2.2 Thermodynamic standard conditions

The quantitative thermochemical values tabulated in this article are assumed to correspond to the following conditions

- Pressure: $1.013\,25 \times 10^5$ Pa (1 atm)
- Temperature: 298.15 K ($25°C$)

Furthermore, variables assuming that the concentration of every chemical specie is 1 M are denoted by a "0" superscript. This specific setting will be referred to as "standard conditions of concentrations". In this setting, the pH is 0.

3.2.3 Metabolic reactions computations

3.2.3.1 Anabolic reaction

The anabolic reaction of a guild G is a chemical reaction producing 1 molecule of C-normalized biomass per reaction. The biomass molecule considered is the one described by Battley (Battley et al., 1997). Its complete formula is

$$\mathrm{C_1H_{1.613}O_{0.557}N_{0.158}P_{0.012}S_{0.003}K_{0.022}Mg_{0.003}Ca_{0.001}}$$

($26.20\,\mathrm{g \cdot mol^{-1}}$) however it is simplified to

$$\mathrm{C_1H_{1.613}O_{0.557}N_{0.158}}$$

($24.76\,\mathrm{g \cdot mol^{-1}}$) in the simulations. The stoichiometric coefficients of this reaction are expressed as $\mathrm{mol \cdot C - mol^{-1}_{Biomass}}$ and are adjusted so the reaction is balanced for elements and charge. The impact of using the complex formula is evalutated in appendix 9.1.

3.2.3.2 Catabolic reaction

The catabolic reaction of a guild G is a chemical reaction meant to fuel the guild's energy expense. Most of the time it involves an electron transfer between an electron donor D and an electron acceptor A molecules. The

coefficients of this reaction and expressed as mol $\cdot$ mol$_{\text{D}}^{-1}$ are adjusted so the reaction is balanced for elements and charge.

3.2.4 Microbial dynamics

Microbial dynamics for chemical specie C under the action of a guild G is;

$$[\dot{C}] = Y_{C/G}^{met} \cdot \mu_G \cdot [G] \tag{3.1}$$

With $Y_{C/G}^{met}$ being the metabolic stoichiometric coefficient (mol$_{\text{C}}$ $\cdot$ mol$_{\text{biomass}}^{-1}$) of the guild G for the chemical specie C, μ_G being the growth rate of the guild G (day^{-1}) and $[G]$ being the concentration (mol $\cdot$ m^{-3}) of guild G's biomass in the reactor. We posit as an hypothesis that the concentration of water and protons is unaffected by microbial dynamics. $Y_{C/G}^{met}$ is;

$$Y_{C/G}^{met} = Y_{C/G}^{an} + \lambda_G \cdot Y_{C/D}^{cat} \tag{3.2}$$

With $Y_{C/G}^{met}$ being the stoichiometric coefficient (mol$_{\text{C}}$ $\cdot$ mol$_{\text{biomass}}^{-1}$) of the guild G's anabolic reaction for the chemical specie C, $Y_{C/D}^{cat}$ the stoichiometric coefficient (mol$_{\text{C}}$ $\cdot$ mol$_{\text{D}}^{-1}$, D being the guild's electron donor) of the guild G's catabolic reaction for the chemical specie C and λ_G the lambda factor (mol$_{\text{D}}$ $\cdot$ mol$_{\text{biomass}}^{-1}$) of the metabolism of guild G. The lambda factor of a guild G is;

$$\lambda_G = \frac{-\Delta G_{an} + \Delta G_{dis}}{\Delta G_{cat}} \tag{3.3}$$

With ΔG_{an} being the Gibbs energy change (J $\cdot$ mol$_{\text{biomass}}^{-1}$) of the anabolic reaction of guild G, ΔG_{cat} the Gibbs energy change (J $\cdot$ mol$_{\text{D}}^{-1}$) of the catabolic reaction of guild G and ΔG_{dis} the Gibbs energy change (J $\cdot$ mol$_{\text{biomass}}^{-1}$) of one full metabolic reaction. If $-\Delta G_{an} + \Delta G_{dis}$ or ΔG_{cat} is positive, then it is set to 0 instead. The Gibbs energy change ΔG_{dis} of the whole reaction is computed according to the empirical formula devised by Heijnen and collaborators;

$$\Delta G_{dis} = -(200 + 18 \cdot (6 - NoC_{C_S})^{1.8} + exp(((3.8 - \gamma_{C_S})^2)^{0.16} \cdot (3.6 + 0.4 \cdot NoC_{C_S}))) \tag{3.4}$$

With NoC_{C_S} the number of carbons in the carbon source C_S of the guild's anabolic reaction, and γ_{C_S} the degree of reduction of C_S, divided by its number of carbons.

The reduction degree of a molecule reflects the amount of available electrons of the molecule upon electron donation reaction, divided by its number of carbon (or 1 if there is no carbon in the molecule). It is identical to the opposite of oxydation degree. The reduction degree of a molecule is computed as a function of its elemental composition:

$$\frac{C*4+H*1+O*(-2)+N*(-3)+P*5+S*6+Fe*3}{C} \tag{3.5}$$

Negative charges count as 1 and positive charges count as (-1). For example, the reduction degree of acetate ($\mathrm{C_2H_3O_2^{-1}}$) is $2*4+3*1+2*(-2)+1*1=8$ electrons per acetate molecule.

Although seldomly encountered in such calculations, alkaline metals (Li, Na, K) counts as $+1$ and alkaline earth metals (Be, Ca, Mg) counts as $+2$ (Pidello, 2014).

The Gibbs energy change of a reaction r under standard conditions of temperature (298 K) and concentrations (1 mol $\cdot$ $\mathrm{L^{-1}}$ for every chemical species, including $\mathrm{H^+}$) is

$$\Delta G_r^0 = \sum_i Y_i^r \cdot \Delta G_{f,i}^0 \tag{3.6}$$

Where $\Delta G_{f,i}^0$ is the Gibbs energy change of formation of the i^{th} chemical specie involved in the reaction r, as recorded in table 3.1.

The correction of the Gibbs energy change ΔG_r^0 of a chemical reaction r for non-standard concentrations is;

$$\Delta G_r = \Delta G_r^0 + R \cdot T \cdot ln(Q) \tag{3.7}$$

With ΔG_r^0 the Gibbs energy change of the reaction in standard conditions of temperature, pressure and concentrations, R the ideal gas constant (8.314 459 8 J $\cdot$ $\mathrm{mol^{-1}}$ $\cdot$ $\mathrm{K^{-1}}$), T the temperature (K) and Q the mass action ratio (unitless) of the reaction.

Standard conditions of concentration are defined such that the concentration of every chemical specie is 1 M. Those standard concentrations are sometimes modified to assume that the pH is 7 (that is, the concentration of protons is 1×10^{-7} M);

$$\Delta G_r^{0'} = \Delta G_r^0 + R \cdot T \cdot Y_{H^+}^r \cdot ln(1e-7) \tag{3.8}$$

With $Y_{H^+}^r$ the stoichiometric coefficient for protons in the reaction r.

The value of the Gibbs energy change ΔG_r^0 can be computed from the values of enthalpy and entropy changes of the reaction, or by summing the Gibbs energies of formation of every chemical species, weighted by their stoichiometric coefficients;

$$\Delta G_r^0 = \Delta H_r^0 - T \cdot \Delta S_r^0 = \sum_i Y_i^r \cdot \Delta G_{f,i}^0 \tag{3.9}$$

With ΔH_r^0 and ΔS_r^0 respectively the enthalpy and the entropy change of the reaction r under standard conditions of temperature and pressure, Y_i^r the stoichiometric coefficient of the reaction for the i^{th} chemical specie involved in the reaction r, and $\Delta G_{f,i}^0$ its Gibbs energy change of formation as recorded in table 3.1.

ΔG_r^0 is corrected for non-standard temperature conditions using the following relationship (Hanselman, 1991);

$$\Delta G_r^0(T) = \Delta G_r^0(T_S) \cdot \frac{T}{T_S} + \Delta H_r^0 \cdot \frac{T_S - T}{T_S} \tag{3.10}$$

With T being the system's temperature (K), T_S the standard temperature (298.15 K) and $\Delta H_{reaction}^0$ the enthalpy of the reaction in standard conditions of temperature and concentrations ($J \cdot mol^{-1}$), as it is assumed that the reaction's enthalpy do not vary much in the biological range of temperature and concentrations. The mass action ratio Q of the reaction is the product of the concentration of each reagent of the reaction to the power of its own stoichiometric coefficient. Concentrations in the mass action ratio are expressed in $mol \cdot L^{-1}$. The machine epsilon is added to the concentration of each reagent in order to avoid errors caused by the computation of the logarithm of 0. Water and biomasses are excluded from the calculation of the mass action ratio as they are not dilute chemical species.

The growth rate of a guild G is;

$$\mu_G = \mu_{max} \cdot \prod_i exp(\frac{Y_{S_i/G}^{met}}{V_h \cdot [S_i]}) \tag{3.11}$$

With μ_{max} being the maximum growth rate (day^{-1}), $Y_{S_i/G}^{met}$ the stoichiometric coefficient of consumed chemical specie S_i $mol_S \cdot mol_{G-biomass}^{-1}$, V_h the harvest volume ($m^3 \cdot mol_{G-biomass}^{-1}$) and $[S_i]$ the concentration ($mol_{S_i} \cdot m^{-3}$)

of consumed chemical specie S_i. Produced chemical species are not included in the exponential product.

The values of Gibbs change of formation for the chemical species used in the simulations conducted with the MTS model are (in $\mathrm{J \cdot mol^{-1}}$)

The values of enthalpy change of formation for the chemical species used in the simulations conducted with the MTS model are (in $\mathrm{J \cdot mol^{-1}}$)

3.2.5 Additional dynamics

3.2.5.1 Aeration dynamics

Aeration dynamics for aerated chemical specie C is

$$[\dot{C}] = k_L a \cdot ([C]_{sat} - [C]) \tag{3.12}$$

With $k_L a$ being the global exchange coefficient of C ($\mathrm{m \cdot day^{-1}}$), a the specific exchange area of C ($\mathrm{m^2 \cdot m^{-3}}$) and $[C]_{sat}$ being the saturation concentration of C ($\mathrm{mol \cdot m^{-3}}$). In most cases, the $k_L a$ a product ($\mathrm{day^{-1}}$) and the $[C]_{sat}$ value are deemed to be constant during a simulation.

The $[C]_{sat}$ value is computed using the Henry law corrected for the temperature;

$$[C]_{sat} = p_C \cdot H_C \cdot exp(\frac{\Delta H_{sol}}{R}(\frac{1}{T} - \frac{1}{T_S})) \cdot 1e3 \tag{3.13}$$

With p_C being the partial pressure (in atm) of C in the gas bubbled into the reactor, H_C being the Henry's constant of C ($\mathrm{mol \cdot L^{-1} \cdot atm^{-1}}$), ΔH_{sol} being the enthalpy of solution of C ($\mathrm{J \cdot mol^{-1}}$), R the ideal gas constant ($8.314\,459\,8\ \mathrm{J \cdot mol^{-1} \cdot K^{-1}}$), T being the system's temperature (K) and T_S being the standard temperature (298.15 K). The 1e3 factor is here to perform the conversion to $\mathrm{L \cdot m^{-3}}$. The value of the partial pressures are listed in table 3.3. The Henry constants H_C are listed in table 3.4. The values of $\frac{\Delta H_{sol}}{R}$ are listed in table 3.5.

3.2.6 Chemostat dynamics

Chemostat dynamics for solute chemical specie C is;

$$[\dot{C}] = \frac{Q}{V}([C]_{feed} - [C]) \tag{3.14}$$

With Q being the overall flow of the chemostat ($m^3 \cdot day^{-1}$), V the volume of the reactor (m^3) and C_{feed} the concentration of C in the feed bottle of the chemostat ($mol \cdot m^{-3}$). Biomass is considered to be a solute specie in this setting and is therefore affected by this dynamics.

3.2.7 Point-settler dynamics

The "point-settler" dynamics emulates the flow dynamics of a reactor equipped with a clarifier evacuating the soluble chemical species but recirculating a fraction of the biomass back into the reactor. The point-settler dynamics for solute chemical specie C is

$$[\dot{C}] = \frac{Q_i}{V}([C]_{feed} - [C]) \tag{3.15}$$

While the point-settler dynamics for biomass (and eventual other non-solute chemical specie) is

$$[\dot{C}] = \frac{1}{V}(Q_r \frac{Q_i + Q_r}{Q_e + Q_r} - Q_i - Q_r) \cdot [C] \tag{3.16}$$

With Q_i being the main inlet and outlet flow rate of the reactor ($m^3 \cdot day^{-1}$), Q_r being the the non-solute chemical species recirculation flow rate of the clarifier, Q_e being the extraction flow rate of the clarifier (the flow discarding non-solute matter from the recirculated flow), V the volume of the reactor (m^3) and $[C]_{feed}$ the concentration of C in the feed bottle ($mol \cdot m^{-3}$).

3.3 Description of the MTS simulation framework

3.3.1 Notation

The following describes the program developped using the Matlab language in order to simulate the dynamics of the MTS model. The code was deposited at the Agence pour la Protection des Programmes, under the reference number IDDN.FR.001.250026.000.S.C.2017.000.31500. It features some sample code and references to variables of the program. For clarity, explicit references to variables, litteral values (such as dictionary keys) or file names of the program will be written in `Courier` font. Code samples will be written in specific text boxes. Metasyntactic variables inside names will be surrounded by $< >$ symbols. As the program prominently relies on matrix data

structures, the shorthand notation NxM will be used to describe the matrix' dimensions, where N is the matrix' height and M is the matrix' width. When a variable is described, its dimensions, class (according to the Matlab language's data model) and units are written between parentheses after the formal description of the variable. If the dimensions of the variable are provided it is assumed that the variable's class is double.

3.3.2 Files organisation

The files of the program are dispatched among the following directories;

- `src/`: source files of the program
- `hooks/`: additional functions describing supplementary processes affecting the simulation
- `usr/`: user-defined simulation conditions
- `out/`: output of the program; logs and errors
- `doc/`: documentation
- `test/`: functions used for unit testing of the program

3.3.3 Basic principle

The aim of the program is to simulate the evolution of the concentrations of a set of chemical species through time, according to a system of Ordinary Differential Equations (ODE) describing the chemical processes happening in a reactor wherein a user-defined microbial community is growing. The chemical species are the ODE system's variables. The concentration of the chemical species along time are affected by the growth dynamics of the defined microbial guilds according to the specifications of the MTS model. The concentrations can also optionally be affected by other processes such as aeration or chemostat fluxes.

The code of the program is to be executed with the Matlab program. It has been written using the R2014a version of the Matlab language. The Matlab language is made mainly for matrix operations. Therefore, most of the computations done inside the integrated function of the ODE system are formulated as such. As most of the matrices used during the computations are derived from a small set of fundamental vectors, the following constants will be further used for the sake of convenience;

- let C be the total number of chemical species whose concentration is tracked during the simulation of the system
- let G be the number of guilds, that is, the number of simulated microbial populations of the system
- let P the number of implemented processes affecting the dynamics of the chemical species of the system during the simulation

While the growth law of the MTS model itself requires two "parameters", namely the maximum growth rate μ_m and the harvest volume V_h, the MTS simulation program require a greater number of informations also called "parameters". Those parameters will be further called "parameters" while the MTS parameters will be more specifically refered to as "kinetic parameters of the MTS model".

Among the set of all parameters required by the MTS simulation program to work, most of them can be derived from a smaller set of fundamental informations. The program is designed so the user has to provide only this set of informations, without the possibility to input redundant informations or informations which could be derived from computation.

All the computation needed to produce a functional set of parameters for the program in order to perform the simulation of a specified system is done at the beginning of the execution of the program, before the integration of the ODE system.

Figure 3.1 is a sequence diagram illustrating the order into which scripts are called during the initialization of a simulation.

The program can also log some temporary variables of the integration, such as the concentration of each chemical species in the system or the lambda factor of each guild at a given time. This behavior can be enabled or disabled by the user. The generated data is written into a set of csv files stored into a specifically created directory, named with a timestamp in order to avoid confusion between multiple simulation results, and containing the metadata of the simulation. The action of writing the simulation's data will further be refered to as "logging".

The code is structured in order to limit as much as possible to keep unnecessary variables in the namespace (called "workspace" in Matlab's semantics). This is achieved through currying.

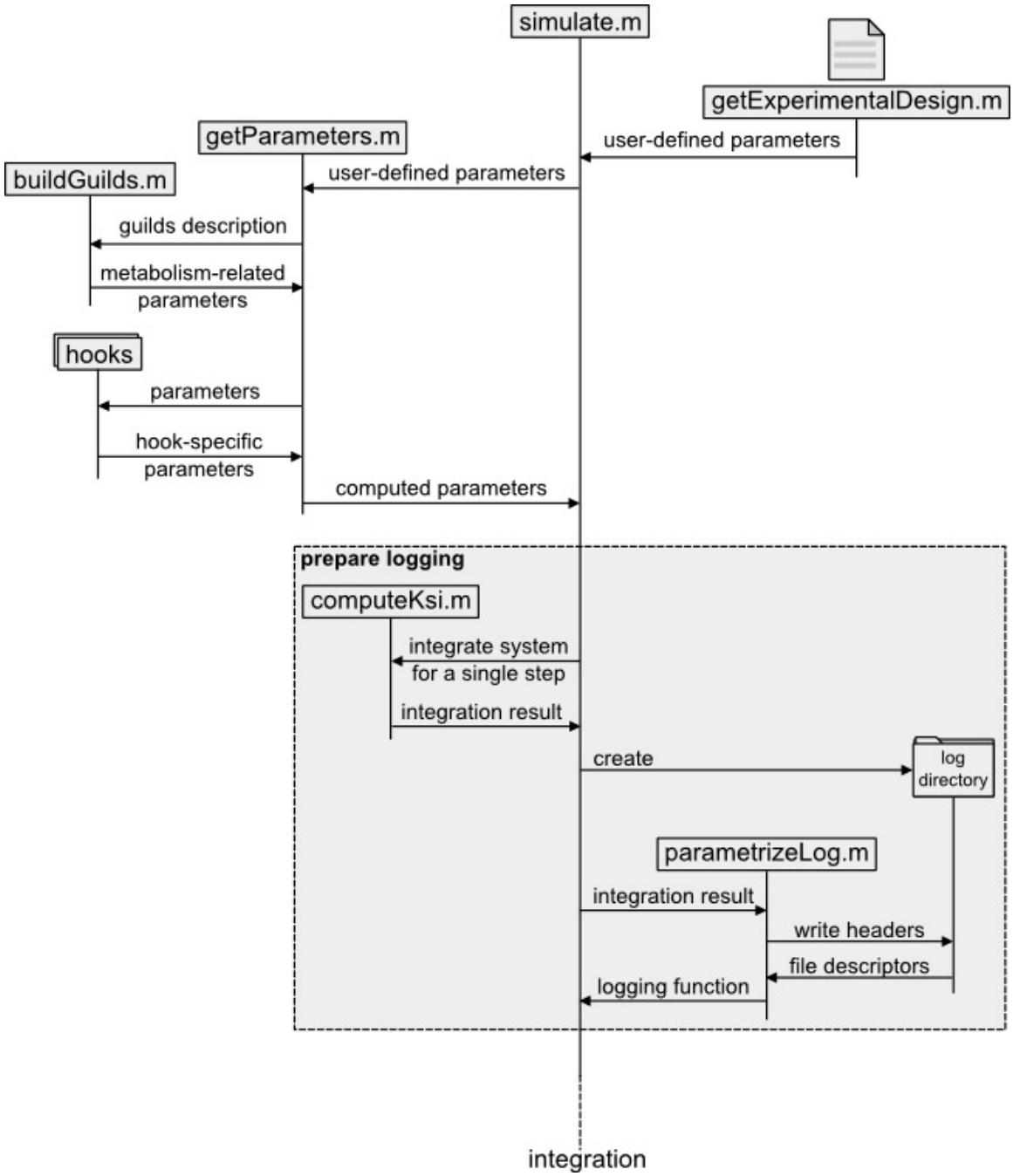


Figure 3.1: Sequence diagram of the initialization of the simulation program, upon call to `simulate.m`, until integration. Read direction is top to bottom

3.3.4 Parameters computation

The computation of the parameters from fundamental informations inputted by the user is done by the `getParameters.m` function. This function is executed once before the integration of the system (cf figure 3.1). It takes a structure containing parameters values set by the user as argument, and returns the complete parameters structure. This structure contains a field for each parameter of the system. To the contrary, the user do not have to provide a value for each parameter. The body of the `getParameters.m` function consists in reviewing each user-definable parameter, setting it to the user-defined value if provided or otherwise to its default value (pre-recorded in the function), then computing the value for each parameter depending on this user-definable parameter. For example, the parameter `T` stands for the system's temperature, it is used to compute other parameters, and if the user do not provide a value for `T`, it is assumed to be standard temperature (298.15 K).

The `getParameters.m` function returns a single structure containing the value of all the simulation's parameters. This structure is refered to in the code as `p` or parameters, and will here be refered to as `p`.

3.3.5 User-definable parameters

The user needs to input a few parameters to the program. The program's parameters that the user can define before a simulation constitutes the fundamental information from which all the other parameters are computed. For example, the user have to input concentrations and temperatures, while saturation concentrations are automatically computed from the latter informations.

This subsection details those parameters, and gives more details for some of them in specific subsections.

- `T`: temperature during the simulation (1x1 K)
- `standardTemperature`: temperature under standard conditions (for thermodynamic data) (1x1 K)
- `guilds`: description of each guilds to include into the simulation. Each guild is described into a struct with fields "donor", "acceptor" and "anabolism". The strings to put in those fields can be found in `buildGuild.m`. (cell array of structs).
- `trifle`: computation epsilon (see the "integration" subsection for more details) (1x1)

- `tankVolume`: tank volume (1x1 m3)
- `initialStateDictionary`: concentration of chemical species in the reactor when the simulation starts. A special value 'biomass' can be used in this dictionary in order to set the concentration of each biomass at once (Cx1 mol.m-3)
- `computeLambda`: function to use to compute lambda factors. The function must take Em, Ecat and Edis as arguments (1x1 function handle)
- `mumaxharv`: μ_m of each guilds (1xG day-1)
- `vh`: V_h of each substrate and each guild (CxG m3.Cmol-1)
- `massActionRatioContributorsDictionary`: mapping of component name to state in mass action ratio contributors matrix (0: the component doesnt contribute, 1: the component does contribute, default is 1). A special value 'biomass' can be used in this dictionary in order to set the concentration of each biomass (containers.Map)
- `metabolizableSpeciesDictionary`: mapping of component name to state in metabolizable components matrix (0: the component's concentration isnt affected by biological reactions, 1: the contrary). (containers.Map)
- `feedBottleDictionary`: mapping of component name to concentration in feed bottle (containers.Map)
- `bioticProcessMask`: mask of biological reactions (=guilds) (1xG boolean)
- `atmosphereDictionary`: mappings of component name to partial pressure (atm) (cell array of containers.Map)
- `supernumeraryComponents`: list of components (cell array of str) to add to the system even though they aren't involved in the metabolism. The supernumerary components are accounted for as non-living components, and if they were are actually involved in the metabolism, not error is raised. The supernumerary components need not to be recorded in `buildGuild.m`. If it doesnt; its dG and dH will be set to 0 and its standard concentration will be set to 1e3 mol.m-3.

- `hooks`: cell array of handles to additional differential functions
- `computeOxygenSaturationAsInWWTP`: use a polynomial relationship to compute oxygen saturation concentration instead of the Henry constant. The polynomial relationship used is $\frac{(14.65-0.41\cdot T+0.00799\cdot T^2-7.78e-5\cdot T^3)}{32}$ with T being the system's temperature in Celsius degrees (1x1 boolean)
- `pH`: initial concentration of protons in the simulation (default pH is 7) (1x1 double)
- `MTSBioticProcessEnabled`: boolean flag stating whether the MTS growth process derivative must be added to the ODE system's derivative or ignored. Disabling MTS biotic processes is likely needed when an alternative growth process is used in a hook (default is 1) (1x1 boolean)

3.3.5.1 Guilds

The guilds are the microbial populations of the simulated system. Each guild is associated with a specific biomass concentration and a growth rate affecting this biomass variable. As the chemical reactions mainly intended to be simulated are microbial growth reactions, and that other implementable reaction concerns chemical species with which microbes directly interact (such as oxygen for aeration), the list of all the tracked chemical species is computed from the metabolic equations of the guilds.

Guilds are inputted as a cell array of structures, each structure defining a guild. The guild-defining structure accepts the following fields;

- `donor`: the name of the electron donation reaction (string). This field is mandatory.
- `acceptor`: the name of the electron acceptation reaction (string). This field is mandatory.
- `anabolism`: the name of the anabolic reaction (string). This field is mandatory.
- `dGdis`: the value of the dissipated energy of the guild (1x1 J). If this field is not set, considerations of bioenergetics are used in order to infer this value (cf equation 3.4 in section 3.2).
- `name`: the name of the guild in the program (string). This defines the name of the guild's biomass in the program's internal variables, and sequently the name of the biomass in the result's log. If this field is not set, a name is automatically generated by combining the name of the acceptor, donor and anabolic reactions.

Guilds in the MTS model are defined by an anabolic and a catabolic reaction. In order to prevent any elemental balance error during the simulation of any microbial community by the program, the stoichiometry of those reactions are not inputted by the user. Instead, reactions and half-reactions' stoichiometry are pre-recorded in the program as dictionary-like data structures (Matlab's containers.Map) and pointed at by their name during guild definition by the user. As the stoichiometry of those reactions is fixed, their standard Gibbs energy and enthalpy change is also automatically computed.

The stoichiometric coefficients of each reactions are computed by solving the following linear equation;

$$\sum_i \nu_i \cdot E_i = 0 \tag{3.17}$$

for every element involved in the reaction, with ν_i the stoichiometric coefficient of the i^th chemical specie involved in the reaction and E_i the number of occurences of the element E in the i^th chemical specie involved in the reaction. Charge is also considered as an element for the purpose of this equation system. These linear equation systems are not solved by hand but are solved automatically using mathematical programming tools such as GNU MathProg.

Whole anabolic reactions are thus pre-recorded. The stoichiometric coefficients of all anabolic reactions are adjusted so that the stoichiometric coefficient of biomass is 1, so that the stoichiometric coefficients of the anabolic reaction can be expressed as mol.molBiomass-1.

The pre-recorded anabolic reactions are;

- `autotrophNitrite`:

$$1\mathrm{HCO_3^-} + 0.82883\mathrm{NH_4^+} \longrightarrow 0.67083\mathrm{NO_2^-} + 1.1013\mathrm{H_2O} + 0.49967\mathrm{H^+} + 1\mathrm{C_1H_{1.1613}O_{0.557}N_{0.158}}$$

- `autotrophNitrate`:

$$1\mathrm{HCO_3^-} + 0.66113\mathrm{NO_2^-} \longrightarrow 0.50313\mathrm{NO_3^-} + 0.93363\mathrm{H_2O} + 0.16425\mathrm{H^+} + 1\mathrm{C_1H_{1.1613}O_{0.55}N_{0.158}}$$

- `glucostroph`:

$$-0.167\mathrm{C_6H_{12}O_6} + 0.158\mathrm{NH_4^+} \longrightarrow 0.43050\mathrm{H_2O} + 0.00625\mathrm{HCO_3^-} + 0.164\mathrm{H^+} + 1\mathrm{C_1H_{1.1613}O_{0.55}N_{0.158}}$$

- `acetotroph`:

$$0.50313\mathrm{C_2H_3O_2^-} + 0.158\mathrm{NH_4^+} + 0.33887\mathrm{H^+} \longrightarrow 0.4305\mathrm{H_2O} + 0.00625\mathrm{HCO_3^-} + 1\mathrm{C_1H_{1.1613}O_{0.55}N_{0.158}}$$

- `propionotroph`:

$$0.2875\mathrm{C_3H_5O_2^-} + 0.158\mathrm{NH_4^+} + 0.267\mathrm{H^+} + 0.1375\mathrm{HCO_3^-} \longrightarrow 0.4305\mathrm{H_2O} + 1\mathrm{C_1H_{1.1613}O_{0.55}N_{0.158}}$$

- `butyrotroph`:

$$0.20125\mathrm{C_4H_7O_2^-} + 0.158\mathrm{NH_4^+} + 0.23825\mathrm{H^+} + 0.195\mathrm{HCO_3^-} \longrightarrow +0.4305\mathrm{H_2O} + 1\mathrm{C_1H_{1.1613}O_{0.55}N_{0.158}}$$

- `valerotroph`:

$$0.15481\mathrm{C_5H_9O_2^-} + 0.158\mathrm{NH_4^+} + 0.22277\mathrm{H^+} + 0.22596\mathrm{HCO_3^-} \longrightarrow 0.4305\mathrm{H_2O} + 1\mathrm{C_1H_{1.1613}O_{0.55}N_{0.158}}$$

In order to allow more flexibility, the catabolic reaction is created from the combination of the electron donor and the electron acceptor reaction. Both half-reactions are pre-recorded with the number of electrons they respectively donate and accept, and their stoichiometry is factorized in order to close the charge balance of the overall reaction, while giving a stoichiometric coefficient of 1 for the electron donor molecule. This way the stoichiometric coefficients of the catabolic reaction can be expressed as mol.molElectronDonor-1.

The pre-recorded electron donation half-reactions are;

- `acetate-oxidizing`:

$$1C_2H_3O_2^- + 4H_2O \longrightarrow 2HCO_3^- + 9H^+ + 8e^-$$

- `propionate-oxidizing`:

$$1C_3H_5O_2^- + 7H_2O \longrightarrow 3HCO_3^- + 16H^+ + 14e^-$$

- `butyrate-oxidizing`:

$$1C_4H_7O_2^- + 10H_2O \longrightarrow +4HCO_3^- + 23H^+ + 20e^-$$

- `valerate-oxidizing`:

$$1C_5H_9O_2^- + 13H_2O+ \longrightarrow 5HCO_3^- + 30H^+ + 26e^-$$

- `glucose-oxidizing`:

$$1C_6H_{12}O_6 + 12H_2O \longrightarrow 6HCO_3^- + 30H^+ + 24e^-$$

- `hydrogen-oxidizing`:

$$1H_2 \longrightarrow 2H^+ + 2e^-$$

- `glc2ace-oxidizing`:

$$1C_6H_{12}O_6 + 8H_2O \longrightarrow 4HCO_3^- + 20H^+ + 1C_2H_3O_2^- + 16e^-$$

- `glc2pro-oxidizing`:

$$1C_6H_{12}O_6 + 5H_2O \longrightarrow 3HCO_3^- + 13H^+ + 1C_3H_5O_2^- + 10e^-$$

- `glc2but-oxidizing`:

$$1C_6H_{12}O_6 + 2H_2O \longrightarrow 2HCO_3^- + 6H^+ + 1C_4H_7O_2^- + 4e^-$$

- `ammonium-oxidizing`:

$$1NH_4^+ + 2H_2O \longrightarrow 8H^+ + 1NO_2^- + 6e^-$$

- `nitrite-oxidizing`:

$$1NO_2^- + 1H_2O \longrightarrow 2H^+ + 1NO_3^- + 2e^-$$

- `iron-oxidizing`:

$$1Fe^{+2} \longrightarrow 1Fe^{+3} + 1e^-$$

The pre-recorded electron acceptation half-reactions are;

- `nitrate-reducing`:

$$1e^- + 0.2NO_3^- + 0.6H_2O + 1.2H^+ \longrightarrow 0.1N_2$$

- `sulfate-reducing`:

$$1e^- + 0.125SO_4^{-2} + 0.5H_2O + 1.125H^+ \longrightarrow 0.125HS^-$$

- `iron3hydroxide-reducing`:

$$1e^- + 1Fe(OH)_3 + 3H^+ \longrightarrow 3H_2O + 1Fe^{+2}$$

- `iron-reducing`:

$$1e^- + 1Fe^{+3} \longrightarrow 1Fe^{+2}$$

- `methanogen`:

$$1e^- - 1.125H^+ + 0.125HCO_3^- \longrightarrow 0.375H_2O + 0.125CH_4$$

- `aerobic`:

$$1e^- - 1H^+ + 0.25O_2 \longrightarrow 0.5H_2O$$

- `manganese-reducing`:

$$1e^- + 0.5Mn^{+4} \longrightarrow 0.5Mn^{+2}$$

- `acetate-reducing`:

$$1e^- + 0.111C_2H_3O_2^- + 1H^+ \longrightarrow 0.222CH_4 + 0.222H_2O$$

- `CO2-reducing`:

$$1e^- + 0.125CO_2(g) + 1H^+ \longrightarrow 0.125CH_4 + 0.25H_2O$$

- `CO2-acetogenesis`:

$$1\mathrm{e}^- + 0.25\mathrm{CO}_2(\mathrm{g}) + 0.875\mathrm{H}^+ \longrightarrow 0.125\mathrm{C}_2\mathrm{H}_3\mathrm{O}_2^- + 0.25\mathrm{H}_2\mathrm{O}$$

The choice of the inorganic form of carbon is a still unresolved issue in metabolic modelling. In the simulation of the MTS model presented in this memoir, HCO_3^- has been considered as the inorganic form of carbon most of the time. However, this working hypothesis is not constitutive of the MTS model. This is why some half-reactions have a variant using CO_2 as inorganic carbon form (for example `CO2-reducing` and `CO2-acetogenesis`).

3.3.5.2 Hooks

The only kind of process that is defined in the core sources of the program is the metabolic processes. They are defined in the `src/` directory and they are the only process affecting the system's dynamic during a simulation unless specified otherwise by the user. All the other possible processes are specified in the hook/ directory. Defining those supplementary processes outside of the main source code allow for the extension of the code with new process without modifying the main source code. During a simulation, those processes are stored in a specific list and executed one after another in a loop at the end of the derivative computation, thus adding their specific effect on the system's variables derivative. As this system allow to modularily add arbitrary processes to the execution loop of the program they correspond to the definition of "hook" in informatics semantics. Hence those processes will further be refered as such.

Each hook is written in a specific directory in `hook/`. It contains two functions and a documentation of the parameters of the hook. The function `<hook>_getSpecificParameters.m` is executed during the parameters computation, before the integration. Its role is to compute the specific parameters of the process from the user's input, just as done by the `getParameters.m` function of the main source code (cf figure 3.1). All the parameters computed from those hook-specific functions are stored in a specific structure in the hook field of the main parameter structure. In addition to the parameters used during the integration of the process, this structure also contains a field defining of the hook specific variables should be logged. The `<hook>_process.m` function is the function executed during the integration of the system.

3.3.6 Logging

Recording the variables of the system at each integration step is a behavior which can be enabled or disabled by the user. It should be noted that when logging is enabled, all loggable variables will be logged; it is not possible to log only a set of the loggable variables. The logging is performed by the `logState.m` function which is called at specific points of the integration (more informations about those specific times in the Integration subsection).

3.3.6.1 Logged data format

The logged data is dispatched in multiple csv (comma separated value) files all stored in a specific timestamped directory. Each csv file tracks the value of a single variable over time. It should be emphasized that variable is considered here in Matlab's sense, which imply that the values of a single variable being a multidimensional array are logged in a single file. For example, the concentration of each chemical species in the system are grouped in a single variable (a Cx1 vector), and consequently are logged in a single file.

The csv files' format follows the "tidy" format defined by Hadley Wickham (Wickham, 2014). According to the tidy convention, each line corresponds to a single observation. When logging the value of a multidimensional variable, each individual value of the array corresponds to an observation. For example when logging the concentrations variable, the concentration of each individual chemical specie consists in an observation and must be logged in an individual line. The csv file containing the record of this variable would then have three columns; namely time, chemical specie name and concentration value.

The benefit of the use of this special data format over the drawback of its verbosity is that the data files can then be readily processed by a set of R packages called "tidyverse" used for data manipulation and display. The packages most prominently required for the manipulation and display of the output of the MTS simulation program are `ggplot2` and `dplyr`.

3.3.6.2 Parametrization of the logging function

The `logState.m` function needs some informations to work. For each variable to be logged, those informations are the filehandle to which data must be written, and the levels of each factors into which the data is splitted into. Those data are obtained before the integration of the system. The program performs a preliminary "blank"

integration of the system over one timestep, which allows to obtain the bundle structure, which is a list of all the loggable variables generated by the derivative function of the program (`computeKsi.m`) (more informations on the bundle structure in the Integration subsection). This preliminary integration also allow to raise eventual obvious errors in the code before to produce log files.

The variables to log are a set of the variables contained in the fields of the bundle structure. This structure is passed to the `parametrizeLog.m` function, into which the list of the variables to log and how to log them (recorded informations are the name of the file under which to log the data, and the name of the levels of the factors of the data) is pre-recorded. The `parametrizeLog.m` function creates the csv file and writes the header for each logged variable and returns a structure associating to each logged variable name the file handle to its csv file and the list of the name of the levels of the factors or the variable. This structure is passed to `logState.m` by currying.

3.3.7 Integration

3.3.7.1 Simulation time

Simulation time is recorded in days. The integration of the ODE system is not performed once from time 0 to the defined end time of the simulation. The integration is rather performed through multiple adjoining time chunks. The ODE solver function is called on each time chunks. The `logState.m` function records the logged variables after each time chunks. The number of times the derivative function of the program (`computeKsi.m`) is called depends on the integration algorithm and on the ODE system's state. In contrast, the number of time chunks is set by the user when parametrizing the simulation by defining the length of the time chunk in simulation time.

A supplementary subtlety in the way simulation time is divided into chunk is added by the existence of the user-defined chunkEasingCoefficient parameter. During the integration, the end time of a chunk is given by the relationship $totalTime \cdot advance^{chunkEasingCoefficient}$ where $totalTime$ is the total time of the simulation in simulation time (day), $advance$ is a number between 0 and 1 indicating the end position of the chunk relative to the total time of the simulation. The total number of time chunks is independant of the $chunkEasingCoefficient$ parameter. If $chunkEasingCoefficient$ is 1, all the chunks do have the same length. If $chunkEasingCoefficient$ is between

0 and 1, the chunks' length will start big and decrease with time. On the contrary, if $chunkEasingCoefficient$ is superior to 1, the chunks' length will start small and increase with time. This parameter allows to focus the precision of the logging on the beginning or the end of the simulation, while eventually accelerating the integration of the other part by diminishing the number of time chunks.

3.3.7.2 Computation of the variables' derivatives

The function integrated by the solver function and representing the ODE system in the program is `computeKsi.m`. This function first call the `MTSBioticProcess.m` function, which computes the derivatives of the concentrations of the system's chemical species according to the growth processes of the simulated microbial community. Then the `computeKsi.m` function collects and sums all the derivatives resulting from the hooks. The sum of all derivatives is then returned by the `computeKsi.m` function.

In the `MTSBioticProcess.m` function, the exergy of the anabolic and catabolic reactions are computed and corrected for non-standard concentrations and temperature conditions. The lambda factor is computed from those values and this of the dissipated energy associated with each metabolism. The stoichiometric coefficients of the metabolic reactions are then computed as a CxG matrix. The CxG matrix of the stoichiometric coefficients for substrate is obtained by filtering the negative values from this matrix. The matrix of substrate stoichiometric coefficients is then used to obtain the matrix of $\frac{stoichiometriccoefficientofsubstrateS}{V_h \cdot (concentrationofsubstrateS)}$ factors. Those factors are then exponentiated and multiplied to the maximum growth rate of their respective guild in order to give the guild's specific growth rate value in $day^{-1} \cdot mol_{biomass}^{-1}$. The specific growth rates are then multiplied to the biomass concentration of their guild to obtain the population growth rates, as a 1xG vector. This vector is then multiplied to the previously computed metabolic stoichiometric coefficients matrix in order to get the derivatives of the ODE system according to the growth reactions.

By definition, the exergy is the maximum usable work from a process. Practically speaking, if the Gibbs energy change of a process is negative, its exergy is the absolute value of the Gibbs energy change. And when the Gibbs energy change of a process is positive, its exergy is null. Exergy is considered in the program instead of Gibbs energy change. This implies when the catabolic reaction becomes endergonic, its exergy is forced to 0. Consequently, the lambda value is infinite if the catabolic reaction is endergonic. The code of the function is organized in order to accept this value from the lambda factor and to give a safe and meaningful value to subsequent variables by

substituting the produced `NaN` values (Matlab's special value denoting "Not A Number"). If the lambda factor is infinite, the growth rate of the guild is zero.

In addition with the computation of the derivative of the ODE system, the `computeKsi.m` function initializes a structure called "bundle" in the program. Every function computing the ODE system's derivatives (which means `MTSBioticProcess.m` and the process function of each hook used in the simulation) receive this "bundle" strucure as argument, enrich it with new fields and return it.

Inside the process functions, the bundle structure is used to store some variables of interest, whose value can then be passed outside of the function. This bundle structure is then returned by the `computeKsi.m` function, along with the derivatives of the ODE system. The use of this function is to be passed to the logging function (`logState.m`) in order to print the data it contains into specific log files. It should be emphasized that a variable must be stored inside the bundle structure in order to be logged.

3.3.7.3 Integration parameters

The parameters of the integration of the ODE system itself are defined by the user separately from the system's parameters. The integration's parameters must all be defined by the user. The are namely;

- `chunkSize`: the size of an integration time chunk, assuming that chunkEasingCoefficient is 1. Whatever the value of chunkEasingCoefficient, this parameter determines the number of time chunks of the integration. (1x1 day)
- `totalTime`: the total time of the integration in simulation time (1x1 day)
- `chunkEasingCoefficient`: a coefficient affecting the size of the time chunks over time (1x1 unitless)
- `resultRoot`: the directory into which the log directory of the simulation will be created, if required (str)
- `errorDir`: the directory into which the error log of the simulation will be written, if required (str)
- `solver`: the solver function to use. Matlab natively propose multiple solver functions such as ode45, ode113 or ode23 (Matlab function handle)
- `solverOptions`: Matlab's ODE solvers are parametrized by a specific structure returned by Matlab's

function odeset. This function allow to define parameters such as the absolute error tolerance of the solver (Matlab's odeset structure)

The variables in the ODE systems integrated by the MTS simulation framework are chemical species concentrations. This raises two issues regarding the integration of such ODE system.

First, concentration cannot meaningfully be negative. However, most ODE solvers assumes that the ODE system's variables value belongs to the set of real numbers. Consequently, the solver must be parametrized in order to handle this concern. With Matlab's solvers, it is done by using the NonNegative parameter of odeset. The NonNegative parameter is a vector specifying the index of each variable that cannot be negative during the integration. This parameter must then include the index of all of the ODE system's variables. Consequently, this parameter can be set only after the system's definition have been parsed by the program, as the total number of variables of the system depends on the microbial community definition.

Second, concentration cannot be meaningfully null. If a concentration is exactly zero, mass action ratio is not meaningful anymore, and division by zero occurs in the MTS model growth rate formula if the chemical specie is a substrate. In order to avoid this case, a very small quantity is added to the concentrations in every computation where a zero value would yield irrelevant values (Matlab's `Inf` or `NaN` special values). As this small quantity is used during the computation of the derivatives and that the derivative functions do not have access to the integration-specific parameters, this quantity is stored as a system's parameter in the p structure as the field name `trifle`. The default value of trifle is Matlab's `realmin` value, defined as the "Smallest positive normalized floating-point number" in Matlab's documentation (2.2251e-308).

3.3.7.4 Error logging

Errors raised during the integration of the ODE system are caught in order to save the last value of the bundle structure in the directory specified by the `errorDir` parameter. This logging behavior is independent of the choice of the user to log or not the results of the simulation.

3.3.8 Definition of simulations's conditions

The specific system's and integration's parameters of a simulation are defined in a user-written function whose name is mandatorily `getExperimentalDesign.m`. This function must have its own directory in the `usr/` directory, as this directory is purposed for user-generated content. The directory into which the `getExperimentalDesign.m` file is written will further be called "simulation directory".

The integration of a simulation is launched by the user by calling the simulate.m function, which takes the path to the simulation directory as sole argument. The `simulate.m` function first clears the warnings, console, namespace and the executable paths, leaving only `src/` and the path to the simulation path as executable paths.

The `getExperimentalDesign.m` function is then executed by `simulate.m`. It returns two variables; a structure containing system's and integration's parameters, and a boolean flag stating whether the data resulting from the simulation must be logged or not. The simulation's parameters can either be a structure or a cell array of structures. In the case of a cell array of structures, each structure is used for the integration of a simulation (see "integration of multiple simulations" below). In the structure defining a simulation, the integration-specific parameters are stored directly as fields of the structure, while the system's parameters are grouped into a specific structure stored in the field `fixedParameters`.

The definition of some parameters by the user constitutes a circular dependency problem. The most obvious case is the dependency of some parameters to the `C` constant (number of tracked chemcial species in the simulation); this constant is determined from the parsing of the metabolisms by the program (inside `getParameters.m` function). The circular dependency problems are solved in `getExperimentalDesign.m` by defining the metabolisms definition parameter first, then calling the `getParameters.m` function on a mock parameter set containing only the metabolisms definition in order to get the `C` and `G` constants.

Examples of `getExperimentalDesign.m` files are given in the annex 9.2.

3.3.9 Calibration of the parameters of the MTS model

The parameters of the program can be calibrated. As the `simulate` function calls Matlab's `clearvars` function, the current state of the parameters calibration function have to be saved as a file, otherwise it will be erased at the beginning of each simulation. The `getExperimentalDesign.m` function must then get part of

its parameters from a text file.

The calibration functions of the MTS simulation program implements the following procedure;

```
    get the values of the parameters to calibrate as argument
    write the values of the parameters to calibrate in a text file
    perform the simulation
    get the watched data from the simulation's log
    delete the simulation log
    compare the simulation's data with a reference
    returns a distance score
```

A function implementing such procedure can then be used for the calibration of the MTS model's parameters with Matlab's `fminsearch` function.

Chemical specie name	ΔG_f^0 ($kJ \cdot mol^{-1}$)
H_2O	-237.2
H^+	0
H_2	0
$CO_2(gas)$	-394.4
$C_2H_3O_2^-$	-369.4
$C_3H_5O_2^-$	-361.4
$C_4H_7O_2^-$	-352.6
$C_5H_9O_2^-$	-344.3
HCO_3^-	-586.9
CH_4	-50.7
$C_6H_{12}O_6$	-919.8
O_2	0
NO_3	-111.3
NO_2	-32.2
$N_2(g)$	0
NH_4^+	-79.4
SO_4^{+2}	-744.6
HS^-	12.1
H_2S	-27.8
Fe^{+2}	-78.9
Fe^{+3}	-4.6
Mn^{+4}	-85.4
Mn^{+2}	-228.0
Biomass	-82.16

Table 3.1: Gibbs energy of formation of the chemical species used in the MTS simulation program

Chemical specie name	ΔH_f^0 (kJ.mol-1)
H_2O	-285.8
H^+	0
H_2	0
$CO_2(gas)$	-393.5
$C_2H_3O_2^-$	-485.8
$C_3H_5O_2^-$	-510.4
$C_4H_7O_2^-$	-535.0
$C_5H_9O_2^-$	-560.0
HCO_3^-	-692.0
CH_4	-74.8
$C_6H_{12}O_6$	-1264.2
O_2	0
NO_3^-	-173.0
NO_2^-	-107.0
N_2	0
NH_4^+	-133.3
SO_4^{+2}	-909.6
HS^-	-17.6
H_2S	-20.5
Fe^{+2}	-48.5
Fe^{+3}	-89.1
Mn^{+4}	-100.4
Mn^{+2}	-220.7
Biomass	-126.83

Table 3.2: Enthalpies of formation of the chemical species used in the MTS simulation program

Chemical specie name	partial pressure (atm)
O_2	0.21
H_2	5.5e-7
N_2	0.78
CO_2	4e-4

Table 3.3: Partial pressures of chemical species used in the MTS simulation program

Chemical specie name	Henry constant ($mol \cdot L^{-1} \cdot atm^{-1}$)
O_2	1.3e-3
H_2	7.8e-4
N_2	6.1e-4
CO_2	3.4e-2

Table 3.4: Partial pressures of chemical species used in the MTS simulation program

table:henry constants

Chemical specie name	$\frac{\Delta_{H_{sol}}}{R}$ (K)
O_2	1700
H_2	500
N_2	1300
CO_2	2400

Table 3.5: Adjustment factor of Henry's law for nonstandard temperature

Chapter 4:

Consistent microbial dynamics and functional community patterns derived from first principles

This chapter corresponds to an article accepted by The ISME journal. The aim of this article is to investigate the properties of the MTS model's growth function. Simulations of incremental complexity are performed in order to assess the implications of this growth function, from the scale of a single population to this of a simplified community.

As the focus of this article is the properties of the growth function itself, the effect of the value of its kinetic parameters (the maximum growth rate μ_{max} and the harvest volume V_h) on the MTS model's predictions have not been tested in this article. Instead, propositional values are affected to them. Indeed, the calibration of the value of those kinetic parameters on experimental data, as well as the study the MTS model's predictions' sensitivity to the value of the kinetic parameters, is treated as a separate subject, addressed in the article presented as the next chapter of this memoir.

In a nutshell, the results presented in the current chapter show that the MTS model is able to account for the limitation of growth by multiple resources at once, and to generate consistent growth patterns from these limitations.

Consistent microbial dynamics and functional community patterns derived from first principles

Hadrien Delattre[1], Elie Desmond-Le Quéméner[1,2], Christian Duquennoi[1], Ahlem Filali[1], Théodore Bouchez[1]*.

[1]Irstea, UR HBAN, F-92761, Antony, France.

[2]LBE, Univ Montpellier, INRA, Narbonne, France

Correspondence: T Bouchez, Irstea, UR HBAN, 1 rue Pierre-Gilles de Gennes CS 10030, F-92761 Antony, France

E-mail: theodore.bouchez@irstea.fr

Abstract

Microbial communities are key engines that drive earth's biogeochemical cycles. However, existing ecosystem models have only limited ability to predict microbial dynamics and require the calibration of multiple population specific empirical equations. In contrast, we build on a new kinetic "Microbial Transition State" (MTS) theory of growth derived from first principles. We show how the theory coupled to simple mass and energy balance calculations provides a framework with intrinsically important qualitative properties to model microbial community dynamics. We first show how the theory can simultaneously account for the influence of all the resources needed for growth (electron donor, acceptor and nutrients) while still producing consistent dynamics that fulfill the Liebig rule of a single limiting substrate. We also show consistent patterns of energy dependent microbial successions in mixed culture without the need for calibration of population specific parameters. We then show how this approach can be used to model a simplified activated sludge community. To this end, we compare MTS derived dynamics with those of a widely used activated sludge model and show that similar growth yields and overall dynamics can be obtained using two parameters instead of twelve. This new kinetic theory of growth grounded by a set of generic physical principles parsimoniously gives rise to consistent microbial population and community dynamics, thereby paving the way for the development of a new class of more predictive microbial ecosystem models.

Introduction

Microbes are the most abundant living things on earth (Whitman et al 1998) and are the key engines that drive earth's biogeochemical cycles (Falkowski et al 2008). Developing models able to capture and predict their dynamics and community assembly patterns is therefore of the outmost importance for the study of global earth ecological equilibria and the development of innovative microbial biotechnology processes (Rodríguez et al 2008, Verstraete 2007, Widder et al 2016). However, current microbial growth models are based on empirical equations, such as those from Monod (Monod, 1949), Contois (Contois, 1959) or Haldane-Andrews (Andrews, 1968), and require extensive parameter calibration based on experimental data. Calibrated models are bound to specific experimental conditions, thus limiting their predictive abilities to a narrow domain. In order to build more generic models, there is a need to more thoroughly capture the fundamental drivers of microbial growth and to mathematically express how they contribute to the emergence of the many community assembly patterns observed in nature.

An increasing number of observations show that environmental physical-chemical factors shape the metabolic niches within a given biotope, resulting in stable functional microbial community structures despite random invasions (Louca et al 2016a). This “functional convergence” phenomenon, i.e. the tendency of microbial functional groups to converge towards defined patterns in specific biotopes, has been observed in systems as diverse as sea water (Louca et al 2016b, Raes et al 2011), soil (Kaiser et al 2016, Nelson et al 2016), activated sludge (Ju et al 2014), plant foliage (Louca et al 2016a), cheese (De Filippis 2016) and numerous human body biomes (Huttenhower et al 2012). These observations suggest that mechanistic processes may largely determine the functional patterns of microbial communities. Up to now, the search for mechanistic physics based models of microbial community structure has been somewhat overlooked, as the very existence of general principles governing the structuring of ecosystems is still the subject of debate (Hansson 2003, Lawton 1999, McGill et al 2006, Simberloff 2004). As complex as community dynamics can be, it is nevertheless recognized that an ecosystem

complies with general laws such as those of thermodynamics, and the rule of stoichiometry (Lawton, 1999), and that these laws play a role in ecosystem structure (Odum, 1969). Whether achievable or not, trying to build on these laws to create a general framework that could result in ecologically consistent population and community patterns is a highly desirable research target (Widder et al 2016).

The seeds for such a general framework were sown by the pioneering work of McCarty (McCarty 1965), followed by several other highly detailed studies (Heijnen and Dijken 1991, Heijnen 2010, McCarty 2007, Roels 1980, von Stockar and Liu 1999) of the thermodynamic balance of microbial growth. This framework relies on thermodynamic and stoichiometric rules to describe the yield and output of microbial metabolism in terms of chemical species, heat and entropy. These contributions led to the development of generic methods to predict the stoichiometry and the energetic balance of microbial growth (for a review see (Kleerebezem and Van Loosdrecht 2010) and (von Stockar et al 2008)).

Several attempts have also been made to link thermodynamic balance calculations to the computation of microbial growth rates (reviewed in (Rodríguez et al 2008) and (Kleerebezem and Van Loosdrecht 2010)). For that purpose, some approaches combined balance calculations with heuristic Monod-like relationships between the concentration of the substrate and the absorption rate (Heijnen and Kleerebezem 1999) and were used to simulate virtual microbial ecosystem dynamics (Gonzalez-Cabaleiro et al 2015). Some authors made assumptions about the structure of the metabolic network, the electron transport chain or the organization of the pathway to establish a link between energy balances and rates (González-Cabaleiro et al 2013, González-Cabaleiro et al 2015a, Jin and Bethke 2003, Noguera et al 1998). It was also suggested to rely on the fact that the rate of microbial reactions is governed by enzyme kinetics in order to use the Michaelis-Menten theory and derive microbial rate equations encompassing thermodynamic constraints (Hoh and Cord-Ruwisch 1996). A collision frequency theory for microbial growth was also proposed (Button 1998) and coupled to thermodynamic considerations

(Liu 2006), which was the first attempt to conceptualize the growth phenomena on a physical basis. Recently, a more fundamental kinetic theory of microbial growth grounded by statistical physics principles was introduced (Desmond-Le Quéméner and Bouchez 2014). For the sake of brevity, this theory is hereafter called “Microbial Transition State” (MTS) theory.

Here we investigate using the MTS approach for the modeling of microbial ecosystems. The theory underlying this approach analyzes an elementary microbial division event using first principles. The probability of a single cell to be surrounded by a sufficient amount of resources (electron donors, acceptors and nutrients) to allow division is expressed, considering their microscopic distribution in the culture medium. Thence, a growth rate formula is derived at the population level, as the statistical outcome of the probability of division computed at the level of each individual. This *ab initio* analysis is sufficiently fundamental to be independent of the microbial species and growth conditions. This implies that the growth of any chemotrophic microbial population catalyzing a defined metabolic reaction can be readily implemented with the MTS model. In cases where the model parameters are not calibrated, what is predicted is the dynamics of functional microbial populations as directly derived from the first principles grounding the MTS theory. The resulting equations of growth are seen here as the first fundamental layer determining microbial dynamics, as a result of physical laws, on top of which parametric equations can be added in the future to account for the complexity of biological or ecological phenomena that are beyond the scope of the current analysis. The approaches developed in this work represent an opposite take on modelling compared to many models in biology that are bound to a specific system, and that calibrate the parameters of an empirical expression in order to quantitatively reproduce the behaviors of the experimental system under study. In contrast, the objective of this contribution is to evaluate the qualitative properties of MTS dynamical models as derived solely from the set of fundamental hypotheses grounding the theory. To this end, we focus on generic and idealized situations starting from the simplest one, i.e. a pure culture grown on a minimal medium, and

progressively add more complexity to the simulations, in order to analyze and to question the consistency of the population and community dynamics that arise directly (without parameter adjustment) from microbial transition state theory.

Methods

Modeling ecosystem reaction stoichiometry

The ecosystem includes a biotic and an abiotic component. The abiotic component comprises all chemical reagents, products and spectator species. The biotic component consists of the whole microbial community. In our approach, the microbial community is subdivided into guilds. A guild is defined by the metabolism it catalyzes (aerobic acetotrophs, denitrifying acetotrophs etc., see list below). The population density of each guild is represented by the molar concentration of a generic, C-normalized biomass molecule. The biomass molecule used in the simulations is $C_1H_{1.613}O_{0.557}N_{0.158}$ (Battley 1998). Its enthalpy of formation is -126.83 kJ.mol^{-1} and its Gibbs energy of formation is -82.16 kJ.mol^{-1}. The molecular weight of this molecule is 24.76 g.C-mol-Biomass^{-1}. Assuming a cell volume of 1e-18m^3.cell^{-1} and a cell density of 1.09e6 g.m^{-3} (Milo and Phillips 2015), a factor of 22.7e12 cell.C-mol-Biomass^{-1} will be used for the purpose of illustration in the results. The choice of this molecule does not constitute the MTS model and theoretically, any other biomass formula could be used.

The system stoichiometry is formulated using a vectorial approach as proposed by Roels (Roels 1980), thereby enabling a compact formulation. The generic formulation is described in this section. The reader should refer to the results section for specific examples of implementation. In all the following formulas, the $*$ symbol denotes matrix product. Let r be the number of reagents involved in the system and C be a $r \times 1$ vector storing the concentration of all reagents of the system, including biomasses, in mM at a given time. Let p be the number of processes affecting the concentrations of the reagents; the derivative of C over time is expressed from the balance equation of C as

$$\dot{C} = A * R \tag{1}$$

where A is a $r \times p$ matrix storing the (unitless) stoichiometric coefficients of every process for every reagent, and R is the $p \times 1$ vector of the rate (in time^{-1}) of every process. By convention, the stoichiometric coefficients are either positive or negative depending on the production or consumption of the corresponding chemical species, respectively. The processes described by the MTS model are metabolic reactions. Let g be the number of guilds, A_{met} the matrix of dimension $r \times g$ storing the stoichiometric coefficients of the metabolism of every guild, and R_{met} the $g \times 1$ vector of the rate of each guild reaction. Eventual stoichiometric matrices describing other processes are horizontally concatenated to A_{met}, and their rates are vertically concatenated to R_{met}. The A_{met} matrix is a linear combination of two matrices A_{an} and A_{cat}, both of dimensions $r \times g$, respectively storing the coefficients of the anabolic and catabolic reactions, and adjusted to close the elemental balance in each reaction separately. The stoichiometric coefficients of a catabolic reaction are set so that exactly one electron donor molecule is consumed (unitless stoichiometric coefficients are actually mol/molDonor ratios). The stoichiometric coefficients of an anabolic reaction are set so that exactly one biomass molecule is produced (unitless stoichiometric coefficients are actually mol/C-mol-Biomass ratios) as the production of one unit of biomass is considered as the elementary event in the MTS model.

For the anabolic reaction, we use the convention proposed by Heijnen: the C-source for anabolism is either the electron donor when it is organic or HCO_3^-. We posit here that the reduced and oxidized forms of the catabolism's electron donor should be included in the anabolism. Our working hypothesis is that the nitrogen source used for anabolism is ammonium (no organic sources of nitrogen are present in the simulated culture media), because it is available in every culture medium simulated in this article. Another choice would have been nitrate, but its possible role as an electron acceptor would have made the interactions between guilds more complex and complicated the message of the simulations.

The Gibbs free energy of formation of every chemical species used in the simulations is taken from Kleerebezem and collaborators (Kleerebezem and Van Loosdrecht 2010).

Linking growth stoichiometry to energy balance

The overall metabolism of the whole microbial community can thus be expressed as:

$$A_{met} = A_{an} + \lambda * A_{cat} \quad (2)$$

where λ is the number of times the catabolic reaction of a guild has to be performed for the total produced energy to equal the energy barrier of growth (therefore expressed as $mol_{Donor}.mol_{Biomass}^{-1}$). λ is then a diagonal matrix of guild specific scalar factors (denoted λ_g) that ensures the coupling of energy and stoichiometric balances, as explained by Kleerebezem (Kleerebezem and Van Loosdrecht 2010), a factor also sometimes denoted f_{cat} (Heijnen and Kleerebezem 2010):

$$\lambda_g = -\frac{\Delta G_{an} + \Delta G_{dis}}{\Delta G_{cat}} \quad (3)$$

where ΔG_{an} is the Gibbs free energy change for the anabolic reaction, ΔG_{cat} is the Gibbs free energy change for the catabolic reaction and ΔG_{dis} is the dissipated free energy of growth (Heijnen and Kleerebezem 2010). The dissipated free energy is the Gibbs free energy change in the overall growth reaction (anabolism and catabolism). This energy dissipation makes the overall Gibbs free energy change of growth negative, so the reaction is spontaneous. ΔG_{an} and ΔG_{cat} are computed from the Gibbs free energy change for anabolic and catabolic reactions, corrected for non-standard temperature and concentrations. We posit that only exergonic catabolic reactions can lead to growth. Therefore, if

ΔG_{cat}happens to be positive during the computations, it is set to zero, resulting in insignificant growth. The λ_g factors are computed at each time step of system integration.

Water and biomass activities are not included in the mass action ratio. ΔG_{an}, ΔG_{cat} and ΔG_{dis} values are expressed in kJ.C-molBiomass^{-1}. The value of each λ_g is such that the Gibbs energy variation of the metabolic reaction is equal to the dissipated energy, which in this article, is assumed to be identifiable with the variable $-Y_{GX}^{max}$, empirically defined by Heijnen as

$$\Delta G_{dis} \approx -Y_{GX}^{max} \tag{4}$$
$$= 200 + 18 \cdot \left(6 - NoC_{C_s}\right)^{1.8}$$
$$+ exp\left(\left(\left(3.8 - \gamma_{C_S}\right)^2\right)^{0.16} \cdot \left(3.6 + 0.4 \cdot NoC_{C_s}\right)\right)$$

where the chain length of the carbon source is denoted (NoC_{C_s}) and and (γ_{C_S}) is the degree of reduction of the carbon source required for heterotrophic growth. The degree of reduction of a carbon source is computed as in the original publication by Heijnen and collaborators (Heijnen and Dijken 1991). That is, by summing 4 electron.carbon^{-1}, 1 electron.hydrogen^{-1}, -2 electron.oxygen^{-1} -3 electron.nitrogen^{-1} and +/- 1 electron.charge^{-1}. For example, the reduction degree γ_{C_S} of the carbon source acetate ($C_2H_3O_2^-$) is $4 \times 2 + 3 - 2 \times 2 + 1 = 8$ electron.acetate^{-1}.

Coupling stoichiometry, energy balance and microbial dynamics

The growth rate function used in the simulations is the multi-substrate growth rate function described in the supplementary materials of Desmond-Le Quéméner and Bouchez (Desmond-Le Quéméner and Bouchez 2014). This formula arises from simple hypotheses concerning microbial growth at microscopic scale. These principles can be summarized as:

- a microbial cell needs to overcome a fixed energy barrier in order to divide

- this energy barrier can be broken down into anabolic energy ΔGan and dissipated energy ΔGdis
- the energy available to overcome the energy barrier is the catabolic energy ΔGcat obtained from the catabolism of substrate molecules
- substrate molecules are considered as particles randomly distributed around the cells
- if a fictional, fixed volume V_H ("harvest volume") around the cell contains enough substrate to overcome the energy barrier, the cell is said to be in an "activated" state
- only an activated cell is able to divide

Considering these hypotheses, the proportion of activated cells in the culture medium at a given time can be expressed using a probabilistic reasoning (detailed in (Desmond-Le Quéméner and Bouchez 2014)). Hence, for a given guild, the formula of the microbial growth rate is

$$\mu = \mu_{max} \prod_{i=0} e^{\frac{A_{met,i}}{V_h[S_i]}} \quad (5)$$

where μ is the growth rate (time^{-1}), $A_{met,i}$ the stoichiometric coefficient of substrate i (mol.C-mol-Biomass^{-1}) computed in equation 2, and $[S_i]$ the concentration of substrate i (mol.volume^{-1}). Although visually different, this formula is consistent with the one presented in the introductory article of the MTS model (Desmond-Le Quéméner and Bouchez 2014) (see supplementary materials 1). This formula encompasses two parameters: μ_{max}, which represents the maximum growth rate (time^{-1}) and V_h, which represents the harvest volume (m^3.C-mol-Biomass^{-1}). These parameters both aggregate generic physical phenomena and particular biological characteristics that would be very difficult to assess accurately for each specific guild considered. As the purpose of this article is to document generic growth patterns as derived from MTS theory, we intentionally made simple generic choices for the value of these parameters. The value of μ_{max} for every guild was set to $\frac{k_B T}{h}$ where k_B is the Boltzmann constant (m^2.kg.s^{-2}.K^{-1}), T the temperature of the system (K) and h Planck's constant (m^2.kg.s^{-1}); the result is close

to 2.23e16 hour^{-1} at 298.15 K. This term comes from Eyring's transition state theory (Eyring 1935), on which the MTS theory is based. The value of the V_h parameter was set to 1 m^3.C-mol-Biomass^{-1} for all the guilds, except in the final simulation where its value was set to 10 m^3.C-mol-Biomass^{-1} as a working hypothesis because, given the low yields of the autotrophs, a value of 1 m^3.C-mol-Biomass^{-1} proved to be insufficient (see supplementary materials 2). Considering the previously estimated ratio of 22.7e12 cell.C-mol-Biomass^{-1}, the individual cell harvest volume for a V_h of 1 m^3.C-mol^{-1} would be 4.40e-14 m^3.C-mol^{-1}. A sphere of this volume would have a radius of approximately 14 µm (see supplementary materials 3). The V_h parameter modulates the probability for a cell to be surrounded by sufficient substrate to be activated. From a biological standpoint, V_h results from all adaptations implemented by the cells to increase their ability to collect chemical resources in the culture medium (such as specific membrane transporters or chemotaxis), but also may vary according to the physical characteristics of the biotope such as the diffusivity of the substrates or agitation.

While pH can have many different impacts on the regulation of microbial growth, here only its influence on reaction equilibrium is taken into account through the mass action ratio. Other types of pH effects are beyond the scope of the current MTS model.

The $g \times 1$ R_{met} vector storing the rate of each metabolic reaction is

$$R_{met} = diag(\mathrm{M}) * [X] \tag{6}$$

where M is the $g \times 1$ vector of the microbial growth rate of each guild and $[X]$ the $g \times 1$ vector of the biomass concentration of each guild.

This ordinary differential equation system is implemented and solved using Matlab (MATLAB Release 2014a, The MathWorks, Inc., Natick, Massachusetts, United States.). Code files are available as supplementary materials 2.

Microbial guilds considered

- Glucosotroph:
    - Catabolism: $C_6H_{12}O_6 + 6\ O_2 \rightarrow 6\ HCO_3^- + 6\ H^+$ ($\Delta G^{0'}$ = -2841.3 kJ.molDonor^{-1})
    - Anabolism: $0.167\ C_6H_{12}O_6 + 0.158\ NH_4^+ \rightarrow 0.430\ H_2O + 0.164\ H^+ + 0.00625\ HCO_3^- + C_1H_{1.613}O_{0.557}N_{0.158}$ ($\Delta G^{0'}$ = -28.3 kJ.C-mol-Biomass^{-1})
    - Dissipated energy: 236.05 kJ.C-mol-Biomass^{-1}
- Aerobic acetotroph (Ordinary Heterotrophic Organisms):
    - Catabolism: $C_2H_3O_2^- + 2\ O_2 \rightarrow + 2\ HCO_3^- + 1\ H^+$ ($\Delta G^{0'}$ = -844.4 kJ.molDonor^{-1})
    - Anabolism: $0.503\ C_2H_3O_2^- + 0.158\ NH_4^+ + 0.338\ H^+ \rightarrow 0.4305\ H_2O + 0.0063\ HCO_3^- + C_1H_{1.613}O_{0.557}N_{0.158}$ ($\Delta G^{0'}$ = 23.9 kJ.C-mol-Biomass^{-1})
    - Dissipated energy: 432.12 kJ.C-mol-Biomass^{-1}
- Sulfate reducing acetotroph:
    - Catabolism: $1\ C_2H_3O_2^- + 1\ SO_4^{-2} \rightarrow 2\ HCO_3^- + 1\ HS^-$ ($\Delta G^{0'}$ = -47.7 kJ.molDonor^{-1})
    - Anabolism: $0.503\ C_2H_3O_2^- + 0.158\ NH_4^+ + 0.338\ H^+ \rightarrow 0.4305\ H_2O + 0.0063\ HCO_3^- + C_1H_{1.613}O_{0.557}N_{0.158}$ ($\Delta G^{0'}$ = 23.9 kJ.C-mol-Biomass^{-1})
    - Dissipated energy: 432.12 kJ.C-mol-Biomass^{-1}
- Denitrifying acetotroph:
    - Catabolism: $C_2H_3O_2^- + 1.6\ NO_3^- + 0.6\ H^+ \rightarrow 0.8\ N_2 + 2\ HCO_3^- + 0.8\ H_2O$ ($\Delta G^{0'}$ = -792.1 kJ.molDonor^{-1})
    - Anabolism: $0.503\ C_2H_3O_2^- + 0.158\ NH_4^+ + 0.338\ H^+ \rightarrow 0.4305\ H_2O + 0.0063\ HCO_3^- + C_1H_{1.613}O_{0.557}N_{0.158}$ ($\Delta G^{0'}$ = 23.9 kJ.C-mol-Biomass^{-1})
    - Dissipated energy: 432.12 kJ.C-mol-Biomass^{-1}

- Iron reducing acetotroph:
    - Catabolism: $C_2H_3O_2^- + 4\ H_2O + 8\ Fe^{+3} \rightarrow 9\ H^+ + 2\ HCO_3^- + 8\ Fe^{+2}$ ($\Delta G^{0'}$ = -809.6 kJ.molDonor^{-1})
    - Anabolism: $0.503\ C_2H_3O_2^- + 0.158\ NH_4^+ + 0.338\ H^+ \rightarrow 0.4305\ H_2O + 0.0063\ HCO_3^- + C_1H_{1.613}O_{0.557}N_{0.158}$ ($\Delta G^{0'}$ = 23.9 kJ.C-mol-Biomass^{-1})
    - Dissipated energy: 432.12 kJ.C-mol-Biomass^{-1}
- Ammonium oxidizing bacteria (AOB):
    - Catabolism: $NH_4^+ + 1.5\ O_2 \rightarrow 1\ NO_2^- + 1\ H_2O + 2\ H^+$ ($\Delta G^{0'}$ = -269.9 kJ.molDonor^{-1})
    - Anabolism: $HCO_3^- + 0.828\ NH_4^+ \rightarrow 1.101\ H_2O + 0.670\ NO_2^- + 0.499\ H^+ + C_1H_{1.613}O_{0.557}N_{0.158}$ ($\Delta G^{0'}$ = 267.7 kJ.C-mol-Biomass^{-1})
    - Dissipated energy: 3500 kJ.C-mol-Biomass^{-1}
- Nitrite oxidizing bacteria (NOB);
    - Catabolism: $NO2^- + 0.5\ O_2 \rightarrow 1\ NO_3^-$ ($\Delta G^{0'}$ = -79.1 kJ.molDonor^{-1})
    - Anabolism: $HCO_3^- + 2.64\ NO_2^- + 1.16\ H^+ \rightarrow 0.27\ H_2O + 2.49\ NO_3^- + C_1H_{1.613}O_{0.557}N_{0.158}$ ($\Delta G^{0'}$ = 241.3 kJ.C-mol-Biomass^{-1})
    - Dissipated energy: 3500 kJ.C-mol-Biomass^{-1}

The $\Delta G^{0'}$ values indicated here correspond to standard Gibbs free energy changes corrected for a realistic H^+ concentration of 1e-7 mol.L^{-1} at pH = 7. In the simulations, Gibbs free energy calculations are refined to account for concentrations of all chemical species. For each metabolism, the linear combination of the catabolic and the anabolic reaction gives what is assumed to be the overall growth reaction of the population considered.

Virtual culture conditions

The temperature is set to 298.15 K and the pH is assumed to be 7 in every system, unless specified otherwise.

Aerated batch cultures in minimal M9 medium

The culture medium used in the simulations is a minimal medium homologous to the M9 minimal medium. The concentrations of chemical species used for model initialization are glucose: 17.05 mM, proton: 3.98e-5 mM (pH=7.40), ammonium: 18.69 mM, bicarbonate: 0 mM. When using an alternative carbon source, the same quantity of mol_C is used to enable yield comparison. Thus the concentration of acetate is 51.15 mM. Initial microbial inoculation was set to 1 mM (22.7 $cell.mL^{-1}$).

The only abiotic process implemented in aerated batch systems is aeration. Its rate is $k_L a * ([O_2]_{sat} - [O_2])$ where $[O_2]$ is the current oxygen concentration, $[O_2]_{sat}$ the saturation concentration of oxygen in water (0.273 mM at the temperature of the system according to the Henry law) and $k_L a$ the oxygen transfer coefficient ($time^{-1}$), set to 100 day^{-1} for the simulations. The stoichiometric coefficients of the aeration process consist in a $r \times 1$ $A_{aeration}$ matrix. The coefficient of $A_{aeration}$ for O_2 is 1, and 0 for every other reagent.

Chemostat culture system

The system is submitted to chemostat dynamics; the rate of matter transfer in a chemostat system is $d * (C_{sat} - C)$ (Monod 1949) where C_{sat} is the $r \times 1$ matrix of the concentration of the concentration in the input of the chemostat and d is the dilution rate of the chemostat (input and output flow rate divided by the tank volume) in $time^{-1}$. The stoichiometric coefficient of matter transfer on the reagent concerned is 1. The chemostat culture system is modeled as a 1 m^3 perfectly mixed single compartment. It is subjected to chemostat dynamics with a flow rate of 1 $m^3.day^{-1}$. No aeration process is implemented in this system. Oxygen is fed to the culture medium through the chemostat inflow; its influent concentration is 2.73e-1 mM (8.73 $g.m^{-3}$, which is its saturation concentration). Input concentrations used for this system are meant to emulate a likely groundwater: ammonium 1 mM (18.04 $g.m^{-3}$), bicarbonate 1 mM (61.02 $g.m^{-3}$), nitrate 1.93e-1 mM (12 $g.m^{-3}$), sulfate 6.24e-1 mM (60 $g.m^{-3}$), Fe^{+3}

5.37e-1 mM (30 $g.m^{-3}$). Multiple simulations were ru using acetate concentrations ranging from 0 to 1.5 mM (88.56 $g.m^{-3}$). The initial microbial inoculation was set to 1 mM (22.7e6 $cell.mL^{-1}$) for each microbial guild. The simulation was run until biomass stabilized.

Activated sludge system

The activated sludge system is modeled as an aerated batch. Aeration conditions are identical to those used for aerated batch cultures in M9 medium, except that the oxygen saturation concentration is set to 0.2556 $mol.m^{-3}$, according to an empirical relationship used for wastewater modeling (Hiatt and Grady 2008). The composition of the medium is set to simulate a simplified filtered urban wastewater containing 1.76 mM acetate (103.9 $g.m^{-3}$) and 3.78 mM ammonium (68.19 $g.m^{-3}$). The initial microbial inoculation was set to 1 mM (22.7e6 $cell.mL^{-1}$) for each microbial guild. Values of V_h were common to all guilds and set to 10 m^3.C-mol-$Biomass^{-1}$. For the sake of simplicity, the influent was considered to be free of particles and only aerobic growth of microbial populations was modeled.

Results

Modeling a pure culture growing in a minimal medium containing multiple elemental resources

To show how energy balance, stoichiometry and microbial dynamics are inherently coupled in our modeling framework, here we present model implementation in the simplest case of a single population growing in a minimal medium. The results enable analysis of the model's dynamic properties.

We consider the growth of an axenic culture in a minimal M9 medium, in an aerated batch inoculated with 1 mM biomass (22.7e6 cell.mL^{-1}). A single glucosotroph guild (see Material and Methods) is simulated.

The metabolism of the population consists in the combination of an anabolic and a catabolic reaction and can be expressed as a function of λ;

$$-0.167\ C_6H_{12}O_6 - 0.158\ NH_4^+ + 0.430\ H_2O + 0.164\ H^+ + 0.00625\ HCO_3^- + 1\ C_1H_{1.613}O_{0.557}N_{0.158}$$

$$+ \lambda\ (-1\ C_6H_{12}O_6 - 6\ O_2 + 6\ HCO_3^- + 6\ H^+) = 0$$

where λ can be calculated at each time step according to anabolic, catabolic and dissipated Gibbs free energy variations is

$\lambda = -\frac{\Delta G_{an} + \Delta G_{dis}}{\Delta G_{cat}}$ (see equation 3 in material and methods for details)

The growth rate μ of the population is then expressed as a function of its substrates (namely chemical species having a negative stoichiometric coefficient): $\mu = \mu_{max} \prod_{i=0} e^{\frac{A_{met,i}}{V_h [S_i]}}$

where μ_{max} is a constant that is independent of the microbial population considered (see Material and Methods), $A_{met,i}$ is the stoichiometric coefficient of the substrate S_i, V_h is the harvest volume (independent of substrate and population, see Material and Methods) and $[S_i]$ is the concentration of

the substrate. In the following description of the MTS model predictions, each term $e^{\frac{A_{met,i}}{V_h[S_i]}}$ of equation 5 for each substrate will be called "tuning factor of substrate i". For each substrate, this tuning factor corresponds to the inhibition it exerts on the growth rate depending on the ratio of its demand to its supply in the culture medium. During the course of the simulation, the Gibbs free energy changes in the anabolic and catabolic reactions vary according to the changes in temperature and in the concentration of the reagents over time. The λ factor preserves the energy balance of the metabolism as a whole and is consequently dynamically adjusted. This process ensures the dynamic coupling of microbial growth rates, stoichiometry and energy balance at each time step of the simulation, according to changes in the concentration of the reagents over time.

Figure 1a tracks the concentrations in the system during the dynamic simulation. The simulated population grows linearly until it reaches a plateau at 74.47 mM (1.69e9 cell.mL^{-1}) in 30 hours (1.25 days). Figure 1a also shows the tuning factors of the population, which help understand population dynamics in the MTS model; in the current simulation, they indeed reveal a dynamic, two-step growth limitation phenomenon. In the beginning, as glucose is abundant in the culture medium, the growth rate is mostly limited by oxygen, the electron acceptor. As the aeration rate is constant, the population grows according to this linear aeration rate and depletes glucose and ammonium. At the 30th hour of the simulation, the concentration of glucose reaches a level at which it becomes significantly limiting; its tuning factor plummets from 0.98 to 8e-23 in less than 5 hours (0.2 day). As glucose becomes scarce, the growth rate of the population decreases exponentially as a result of the limitation of the electron donor, to reach 3.3e-6 day^{-1} at the end of the simulation. With such a low metabolic rate, aeration replenishes the oxygen concentration in the batch culture almost to saturation. Consequently, the tuning factor of oxygen (electron acceptor) increases during the glucose-limited phase of growth, and glucose (electron donor) becomes the main limiting substrate. Conversely, simulations were performed using exactly the

same model structure and only changing the initial nitrogen concentration to represent growth in a nitrogen-poor M9 medium. In that case, the growth rate appears to be limited by oxygen in the first phase, after which growth virtually ceases due to ammonia exhaustion (Fig 1c). These simulations reveal that the MTS model can jointly capture the influence of all resources on growth dynamics, whether these resources contribute to the energy supply of the cell (electron donors or acceptors) directly or not (nutrients such as ammonia).

Modeling two competing populations: emergence of an energy-driven ecological succession

One of the factors that influence community structure is the outcome of competition between populations. The MTS model's ability to predict the outcome of competition between multiple populations in the sense of functional guilds has been questioned.

To answer this question, we implemented the minimalist case of two heterotrophic populations growing together in an aerated batch, modeled as a glucose oxidizing guild and an aerobic acetate oxidizing guild. Both guilds have the same absolute μ_{max} value, and the same V_h value. The difference between the dynamics of the two guilds only resulting from thermodynamic constraints was then studied.

The two guilds differ in the energy they dissipate during their respective metabolic reaction according to Heijnen's formula (Heijnen and Kleerebezem 2010); -432.12 kJ.C-mol-Biomass^{-1} for acetate oxidizers vs. -236.05 kJ.C-mol-Biomass^{-1} for glucose oxidizers. However, the number of electrons donated per carbon atom of electron donor is the same (4 electrons per carbon atom of electron donor). The initial quantity of each electron donor is 17.05 mM of glucose and 51.15 mM of acetate. These quantities were adjusted in this simulation so that they represented the same quantity of carbon (and consequently the same quantity of electrons) in the raw element count.

Figure 1d tracks the system's variables over the course of the simulation. The model predicts that the glucose oxidizer guild grows first. Their population is predicted to stabilize at 74.47 mM (1.69e9 cell.mL^{-1})

after 31 hours (1.29 days). The population of acetate oxidizers is predicted to grow from the 31st to the 84th hour (1.3 – 3.5 days) and to stabilize at 50.19 mM (1.14e9 cell.mL^{-1}). The growth of the glucose oxidizer guild follows the same kinetics and the same sequential inhibition as in the previous, mono-guild simulation (Fig 1b). The growth of the acetate oxidizer guild is also sequentially inhibited by its electron acceptor (oxygen) then by its electron donor (acetate).

As the two guilds have a different electron donor (glucose and acetate) but the same electron acceptor (oxygen), they compete for the electron acceptor. The results shows that the guild of glucose oxidizers grows first and also develops a higher stabilized population density than the acetate oxidizers guild (Fig 1d). In presence of the glucose oxidizer guild, the growth of the acetate oxidizer guild is delayed compared to its growth in the absence of the glucose oxidizer guild (Fig 1a). This outcome arises from the thermodynamic properties of the metabolisms involved. The same amount of oxygen per carbon is needed to oxidize glucose or acetate, in order to preserve the elemental balance, and the batch simulation is initialized with the same quantity of carbon for both substrates. However, in the simulation conditions, glucose oxidation is more exergonic than acetate oxidation when normalized by carbon atom. The guild growing on glucose does not need to dissipate as much energy. These two properties are aggregated and related to the growth yields by the λ factor, more acetate than glucose must be oxidized in order to produce the energy required for growth (approximately 0.503 + 1 × 0.6 = 1.1 mol-Acetate.mol-Biomass^{-1} versus approximately 0.167 + 1 × 0.09 = 0.25 mol-Glucose.mol-Biomass^{-1}). Consequently, the tuning factor associated with oxygen is lower for acetate oxidizers than for glucose oxidizers (Fig 1d). Both guilds reduce the oxygen concentration while growing, but the thermodynamic considerations implemented by the MTS model give a better yield to the glucose oxidizers. Hence they can draw down O_2 to lower levels (compared with acetate oxidizers). Consequently the growth of glucose oxidizers brings the oxygen concentration to a level at which the growth rate of acetate oxidizers is insignificant and in practice prevents their growth during the first 30 hours (1.25 days) of the

simulation. Interestingly, the possibility to obtain such a pattern could be experimentally tested in a system where the oxygen supply rate would less than the oxygen uptake rate of the glucosotrophic guild.

It should be noted that the two populations have exactly the same growth parameter values implemented in the MTS model. Despite similar kinetic parameters for the two populations, a microbial succession emerges from the model. Therefore, simple mass and energy balance calculations coupled to the flux-force relationship between energy and rate determine the outcome of competition and an ecological succession emerges. Our model thus appears to inherently exhibit an original property: the ability to account for microbial successions as a result of competition for available resources and energy. In view of this result, the possibility to directly account for community assembly patterns observed in nature was further questioned.

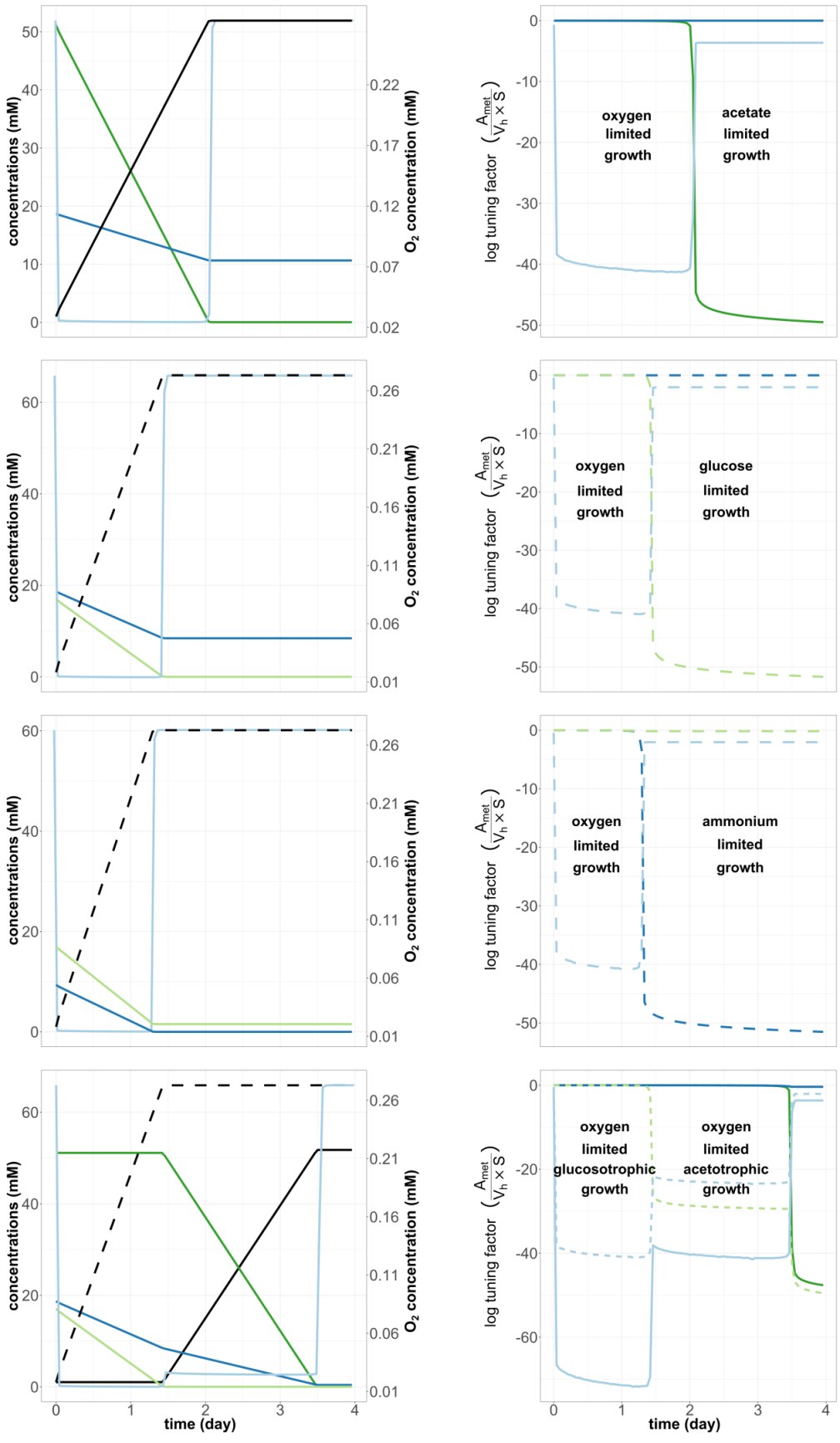

concentrations (mM)
O2 concentration (mM)
log tuning factor (Amet / Vh × S)
oxygen limited growth
acetate limited growth
glucose limited growth
ammonium limited growth
oxygen limited glucosotrophic growth
oxygen limited acetotrophic growth
time (day)

Figure 1 : Aerobic microbial cultures growing in different conditions. Graphs on left column are concentrations over time; acetate concentration in dark green, glucose concentration in light green, ammonium concentration in dark blue, oxygen concentration in light blue, acetotroph biomass concentration in plain black, glucosotroph biomass concentration in dashed black. Graphs on right column are natural logarithm of the tuning factors over time; tuning factors follow the same color scheme as in the left column; tuning factors for acetotroph guild are represented by plain lines, tuning factors for glucosotroph guild are represented by dashed lines. The lower a tuning factor is, the more significant is the limitation exerted by the substrate on the microbial population's growth rate. Time is in hour, concentration is in mM, secondary concentration axis (right side) on the left column is for oxygen. Biomass is quantified as mM of carbon-normalized biomass. (a) Acetotroph guild monoculture. (b) Glucosotroph guild monoculture. (c) ammonium-limited glucosotroph guild monoculture. (d) Acetotroph and glucosotroph co-culture.

Modeling competition between multiple populations: emergence of a community structured according to a "redox tower" of microbial metabolism

Natural systems in which electron donors are available in excess and electron acceptors are limited exhibit typical compartmentalization of microbial activities. This occurs particularly during hypolymnia of eutrophic freshwater lakes, when oxygen depletion results in anoxic zones (Müller et al 2012). This phenomenon has been studied for decades (Boehrer and Schultze 2008, Hutchinson 1957). It leads to sequential vertically stratified consumption of available electron acceptors according to the hierarchy of Gibbs free energy changes of half reduction reactions, leading to "redox towers of microbial metabolisms". To assess the ability of our approach to account for such compartmentalization patterns, we simulated competition between multiple populations for a single electron donor (acetate) and in the presence of various electron acceptors. The competition was analyzed in a chemostat. In these systems, microbial populations are subjected to a continuous flow of substrate and an output rate for biomass. Each population then has two possible stationary states: either its growth rate negates the dilution rate or the population is washed out. We implemented a microbial community channeling electrons from acetate to oxygen, nitrate, sulfate and iron, and we simulated its dynamics according to the MTS model.

The implemented guilds are aerobic acetotroph, denitrifying acetotroph, sulphate reducing and iron reducing guilds (see Material and Methods).

The culture medium is modeled as a perfectly mixed 1 m^3 single compartment subjected to chemostat dynamics with a flow rate of 1 $m^3.day^{-1}$, corresponding to a dilution rate of 1 day^{-1}.

Figure 2 shows the concentration of each guild during the stationary phase, as a function of the influent acetate concentration. The concentration of all populations as a function of time for an input acetate concentration of 1.76 mM is shown as supplementary material 4. As previously shown, the growth of the most thermodynamically advantaged guild delays the growth of competing guilds. The population able to adjust its growth rate to the dilution rate with the minimal amount of acetate is the aerobes guild. This is explained by the superior exergonicity of their metabolism as discussed in connection with the competition between two species above. With increasing acetate concentrations, aerobes coexist with nitrate reducers, then iron reducers, then sulfate reducers. Indeed, when the quantity of electrons provided in the form of acetate exceeds the quantity of electrons acceptable by a given acceptor (i.e. oxygen, nitrate, iron or sulfate), some unoxidized acetate molecules remain in the culture medium. These acetate molecules are available for a guild catalyzing a less exergonic reaction. Therefore the complexity of the community increases with an increase in the incoming electron donor molecules in the system, as the possible electron acceptors are saturated one by one. A community structure pattern compliant with the typical redox tower pattern is thus parsimoniously predicted by the MTS model.

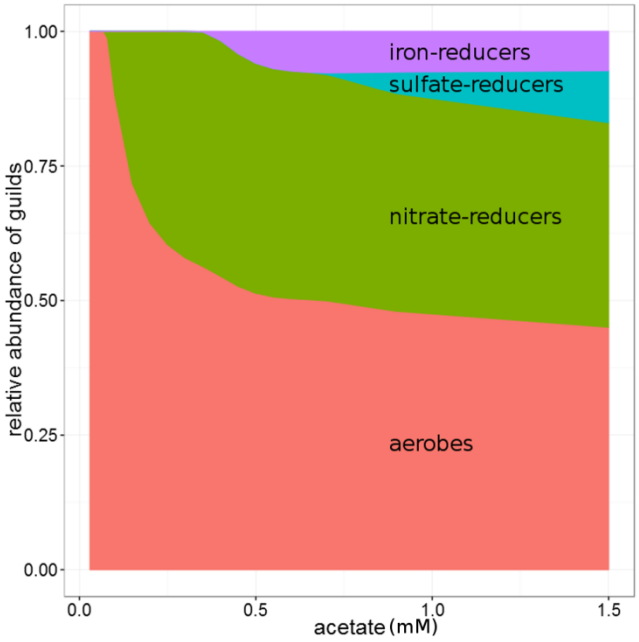


Figure 2 : Competition outcome in a chemostat fed with acetate in the presence of various electron acceptors as predicted by the MTS model. The proportion of each guild at stationary state is depicted as a function of the acetate input concentration (mM). Simulation results predict patterns ranging from competitive exclusion at low acetate concentration to full coexistence at non limiting acetate concentrations.

Towards more predictive models for environmental biotechnology applications: microbial dynamics in a simplified activated sludge community

To illustrate the use of the MTS model in the practical case of environmental bioprocess modeling, we implemented a simplified model of the community found in the aerated tanks of wastewater treatment plants (WWTPs). A set of engineering models called “Activated Sludge Model” (ASM) is commonly used for the operation, design and optimization of WWTPs (Hauduc et al 2009). These models focus on the accuracy of the system state variables. This accuracy depends on how closely the simulated system resembles the reference system from which *a priori* knowledge is inputted through calibration. The calibrated parameters are thus bound to a range of experimental conditions. On the contrary, the MTS-based approach we present here does not rely on calibration based on experimental data: we derived

kinetics and yields directly i.e.: without calibration, from fundamental generic assumptions, mass and energy balance calculations.

An aerated batch incubation containing wastewater and a simplified activated sludge inoculum was then simulated using the MTS and ASMN models. The ASMN model is a modified version of the ASM no. 1 model (Henze et al 1987), with nitrification split into two steps between two guilds(Hiatt and Grady 2008). This model thus splits the microbial community into three functional guilds:

- Ordinary heterotrophic organisms (OHO): heterotrophs consuming an unspecified carbonaceous substrate
- Ammonium oxidizing bacteria (AOB): autotrophs oxidizing ammonium to nitrite
- Nitrite oxidizing bacteria (NOB): autotrophs oxidizing nitrite to nitrate

The model was implemented with default parameters (for more details, see supplementary material 5). The MTS model was implemented with the same three guilds and the two generic parameters μ_{max} and V_h described in Material and Methods. As a working hypothesis, we assumed that the electron donor of the OHO guild is acetate, as it appears to be the most abundant volatile fatty acid in wastewater (Rössle and Pretorius 2001). Simulation results are shown in figure 3, where the predictions of the MTS model are put into perspective by comparing them with simulation results obtained with a default calibration of parameters for the ASMN model (see supplementary material 5). Despite overall similarity, the two simulations exhibit some differences. The main difference is that the three guilds grow simultaneously in the ASMN simulation and sequentially in the MTS model. This difference is linked to the structure of each model. In the ASMN model, growth depends on the combination of Monod-type affinity functions so that a population can grow as soon as its substrates are present in the medium. In the MTS model, the growth of heterotrophs leads to oxygen limitations that virtually prevent the growth of the two autotrophic guilds AOB and NOB in a similar way to that described in previous simulations. Moreover,

the low exergonicity of the nitrite oxidation reaction requires that a sufficient amount of nitrite accumulates in the batch before NOB can compete with AOB for oxygen, leading to a sequential growth pattern and transient nitrite accumulation. This pattern has already been documented in piggery wastewater treatment (Rajagopal et al 2011)), and also in soils, where a Monod-law based model was calibrated to reproduce it (Venterea and Rolston 2002). This sequential growth phenomenon is also apparent in the oxygen concentration profiles, where the three stage consumption pattern is most apparent in the MTS simulation. Despite these differences, the simulations exhibit remarkable similarities. At the end of the simulation, the yield of each guild is similar in the two models: the final quantities of biomass in the batch according to ASMN are 108.4 $gCOD.m^{-3}$ OHO, 50.63 gCOD AOB and 45.07 gCOD NOB, while the MTS model predicts 114.8 $gCOD.m^{-3}$ OHO, 52.4 $gCOD.m^{-3}$ AOB and 45.1 $gCOD.m^{-3}$ NOB (Fig 3). These results show that the abundances of each guild as predicted by thermodynamic rules are consistent with the growth yields measured in activated sludge and used for calibration in the ASMN model. The yields predicted by the MTS model depend on the energy dissipated by the metabolisms, which was computed using the empirical formula that Heijnen and collaborators (Heijnen and Kleerebezem 1999) calibrated on experimental culture data. This is why the yields predicted by the MTS model match those of the ASMN model. In particular, the calibration process used to produce the empirical yield parameter of the ASMN model's heterotrophic population captures the average yield of many thermodynamically constrained populations growing on diverse substrates. The average yield resulting from all these metabolisms is close to the yield of growth on acetate, which is one of the most abundant sources of carbon in such systems.

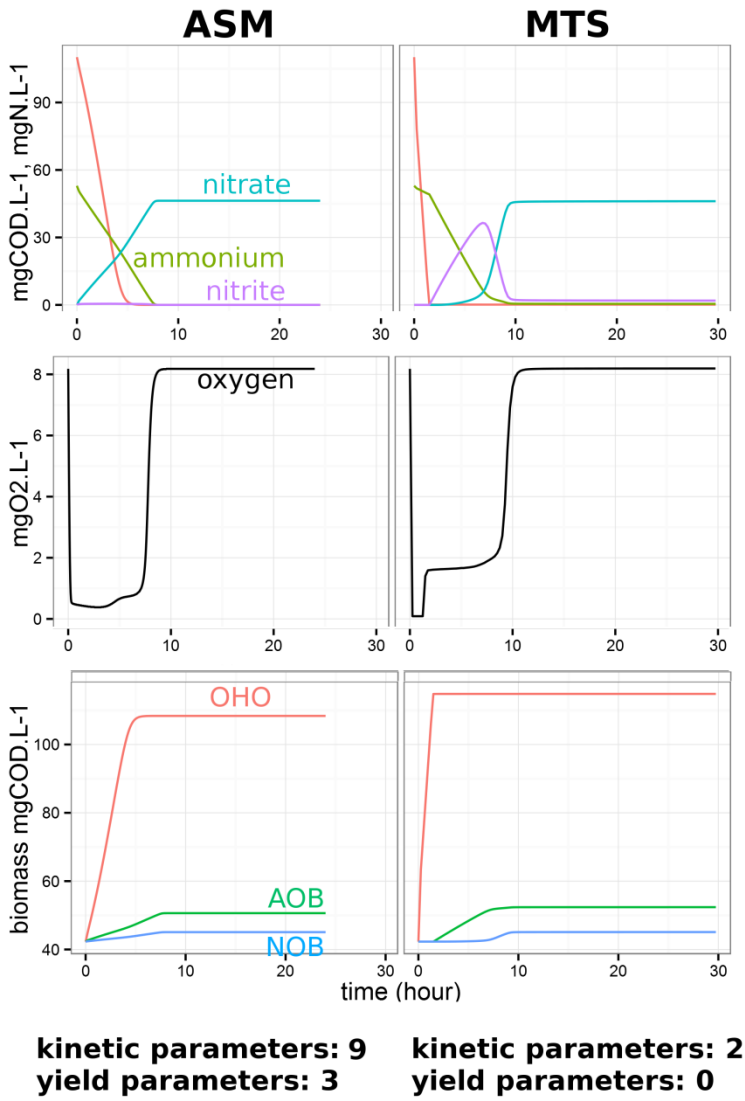


Figure 3 : Concentration of main chemical species and biomass over time for a simplified activated sludge community in an aerated batch reactor. Left column shows the prediction from the MTS model; right column shows the predictions from ASM. The units used to express the concentration of the chemical species are the same as in the ASM models.

Discussion

In this article, we present the dynamics of single and multiple microbial populations arising directly from the MTS theory (Desmond-Le Quéméner and Bouchez 2014) coupled to a thermodynamic and stoichiometric balance calculation framework developed in previous papers (reviewed in (Kleerebezem and Van Loosdrecht 2010)). While several approaches have been proposed in the past to link

thermodynamic balance calculations and kinetics (Heijnen and Kleerebezem 1999, Hoh and Cord-Ruwisch 1996, Jin and Bethke 2003, Noguera et al 1998, Roels 1980, Westerhoff et al 1983), the MTS theory differs fundamentally from all these previous approaches. What is completely novel is that it makes population dynamics emerge as the statistical outcome of all the individual division events described at the microscopic level, from fundamental and generic principles. Strikingly, despite the simplicity of the theory's core principles, consistent microbial dynamics, successions and functional community assembly patterns were simulated without population specific parameter calibration, for systems ranging from pure to mixed cultures. In this article, we therefore document that the kinetic equations arising from MTS theory intrinsically include many important generic properties to adequately model microbial population and community dynamics.

First, using a stoichiometric approach, we show how the model jointly captures the influence of all substrates simultaneously (see equation 5). However, the simulation results also illustrate how growth dynamics actually appear to be limited by one substrate at a given time. This emerging property of the model recalls Liebig's law of the minimum, which states that the growth rate of an organism requiring multiple different resources is controlled by the scarcest resource only. This situation indeed corresponds to the widely accepted intuitive understanding of the way microbial cultures behave in the laboratory, and has often been investigated, tested and confirmed (Kovarova-Kovar and Egli 1998). Liebig's hypothesis is thus widely used to model microbial dynamics, sometimes implicitly, as in the most simple, mono-substrate expression of Monod's equation, sometimes explicitly, as in multi-substrate implementations of the Monod model where the limiting substrate has to be selected at each time step by computing the minimum of all resource dependent growth factors to tune down the maximum growth rate (Droop 1974). Without introducing the Liebig hypothesis, considering the growth dependence of multiple substrates in Monod based models often leads to inconsistent growth patterns that require extensive adaptation of model structure (Bajpai-Dikshit et al 2003, Bapat et al 2006,

Nikolajsen et al 1991). In our case, the MTS model can produce sequential growth limitation either by electron donor, acceptor or by nutrient. The exponential nature of the relationship makes tuning factors "rise" from insignificant to significant in a narrow concentration range (as seen in fig 1), thus allowing pronounced limitation switches. Under the simple conditions simulated, virtual cultures thus exhibit "Liebig like behavior" that is obtained parsimoniously, as an emerging property of the model. The modeling framework we propose is therefore a simple and elegant way to jointly capture the effect of electron donor, acceptor and nutrient concentrations on microbial dynamics, without infringing Liebig's hypothesis.

Second, the simulations involving several populations in competition either for electron donor or acceptor result in community assembly patterns structured like microbial redox towers (see Fig 2), as observed in many natural habitats such as the anoxic hypolymnia of eutrophic lakes (Boehrer and Schultze 2008). These types of patterns have already been generated using Monod based kinetic models (Bethke et al 2011, Gonzalez-Cabaleiro et al 2015). However, our contribution differs significantly from previous works regarding the scientific conclusions that can be drawn from the simulations.

Gonzalez-Cabaleiro's model (González-Cabaleiro et al 2015b) uses a modified Herbert-Pirt equation in which the growth yield is a function of thermodynamic variables. The kinetic equation used to describe substrate consumption rate is a Monod-like function. This model is therefore a juxtaposition of a framework for the calculation of the thermodynamic balance and a phenomenological description of the dynamic growth process that results in ecological successions. Since kinetic parameters were not calibrated in Gonzalez-Cabaleiro's article, their simulations indeed show the phenomenological dependence of ecological successions on thermodynamic variables. However their simulations provide no support for using Monod's law itself as a way to reproduce this pattern. Indeed, since Monod's law is an empirical equation, it carries no hypotheses per se; any similarly shaped curve would have provided

the same simulation results. In contrast, in the case of the MTS model, the thermodynamic variables are used to compute the division probability at the level of each individual. The MTS theory then provides a mechanistic explanation of the influence of energy on microbial division. The sum of all individual division events then propagates at the scale of the population, resulting in the MTS growth equation. Therefore, both the role of thermodynamic variables and the hypotheses on how they are mechanistically linked to the growth rate receive support from the simulation. Our simulations therefore show that the mechanistic explanation of the influence of energy on microbial division at the heart of the MTS theory is sufficient to explain ecological successions.

An interesting parallel can then be drawn with Jin and Bethke's work (Jin and Bethke, 2007), since both their model and the MTS model propose a linkage between microbial kinetics and thermodynamic variables, based on two totally different theories (a probabilistic reasoning for the MTS model, non-equilibrium thermodynamics for Jin and Bethke's model). Both models are fundamentally different since Jin and Bethke's model expresses the rate of catabolism, while the MTS model expresses the rate of biomass synthesis. Jin and Bethke's model therefore cannot account for the limitation of growth by nutrients as the MTS model does (Fig 1c). However, the fact that two theory-based kinetic models provide alternative formulations of microbial growth is interesting, and the careful comparison of the models may provide fruitful insights into microbial growth kinetics.

Regarding the simulation of a simplified activated sludge ecosystem, the MTS model was shown to make predictions qualitatively similar to those of an engineering model (namely, implementation of the ASM). To implement the system whose simulation is depicted in Fig 3, the ASM model requires identification of nine kinetic parameters along with three growth yield parameters, whose values have been carefully adjusted during decades of experimentation (Gujer et al 2000). In contrast, the kinetics obtained in the MTS approach emerges from a theoretical construct in which the only two parameters (μ_{max} and V_h)

common to all guilds were considered. While allowing population specific values for V_h could indeed improve the quantitative accuracy of MTS simulations (for an illustration see Supplementary material 6), we considered default parameters values in the simulations displayed in Figure 3to show that our model is able to generate consistent dynamics with a remarkably small number of parameters, suggesting that the energy/rate dynamic equation resulting from the MTS theory already has interesting predictive abilities in the absence of parameter fitting.

Clearly, the community dynamics obtained by simulation corresponds to those of an idealized functional partitioning of the community resulting from a first layer of energetic and stoichiometric drivers. As implemented in this article, the MTS model obviously does not account for the whole range of phenomena that influence microbial community dynamics and community structure in real complex environmental or engineered settings like those encountered in activated sludge plants. Other processes such as inhibition due to compound toxicity, indirect pH and temperature adaptation effects (Shammas 1986), non-metabolic inter-species interactions, spatial organization (Meister et al 2017), etc. need to be added to obtain a more realistic picture of genuine complex microbial community dynamics. However, including such phenomena in the model was beyond the scope of the present study. The microbial dynamics simulated in this article were indeed not the result of an effort to reproduce specific experimental patterns through the calibration of empirical equations. Rather, the most salient message is to document and analyze the consistency of generic growth dynamics and community assembly patterns as emerging directly from a new kinetic theory of microbial growth relying on first principles coupled with thermodynamic and stoichiometric calculations.

To that extent, the simulations reported here illustrate a set of key properties of MTS dynamic models. To our knowledge, a microbial population dynamics model *per se* has never before exhibited all these properties, that is, without the need for additional hypotheses or specific parameter calibration. What is

of the utmost importance and constitutes the novelty of our contribution, is the fact that these properties are obtained parsimoniously through the combination of fundamental and generic principles translated into mathematical equations, and not from the calibration of population specific parameters. More generally, we advocate the need in microbial ecology to propose new theoretical abstractions to grasp a whole category of phenomena in an inclusive picture. We believe that such approaches pave the way for a new class of microbial ecology and engineering models, built on more robust theoretical foundations and exhibiting enhanced predictive abilities.

Acknowledgments

The authors thank Dr Bart Haegeman for the very useful comments he made on a preliminary version of this manuscript. The authors are also grateful to *Région Ile de France* for funding Hadrien Delattre's PhD in the framework of the DIM R2DS project. The authors thank the French "*Agence Nationale de la Recherche*" for its financial support through the "THERMOMIC" project ANR-16-CE04-0003-01.

Conflict of interest

The authors declare no competing financial interests.

Chapter 5:

Consistent dynamics of simplified activated sludge ecosystem obtained by the calibration of a parameter-parsimonious thermodynamic model

This chapter corresponds to an article draft not yet submitted.

While the previous article was aimed at studying the intrinsic properties of the MTS model's growth function, independently from its parameters' value, the current article investigates the effect of the model's parameters on its predictions. In order to do so, two questions need to be addressed.

Firstly, the model's parameters, namely, the maximum growth rate μmax and the harvest volume Vh, are calibrated in order to have an idea of their value according to experimental data. This is done using respirometry data, and considering a simplified version of the activated sludge microbial community striving in activated sludge. As a result, a preliminary estimation of the MTS model's parameters value is presented in this article. However, further calibration with more complete experimental data should be performed before to have a more accurate picture of the value of the MTS model's parameters.

Secondly, the effect of the value of those parameters on the MTS model's predictions of a simplified activated sludge community is evaluated. The predictions by the MTS model are then compared with those of a modified ASMN model in order to provide a reference.

This article reveals that the qualitative predictions of the activated sludge microbial community dynamics made by the MTS model heavily depends on the thermodynamic constraints that the model implements, and not on the precise value of the parameters. The results presented in this chapter then support the ability of the MTS model to account for dynamics of microbial communities in mixed culture bioprocesses while involving less parameter calibration than engineering models.

Consistent dynamics of simplified activated sludge ecosystem obtained by the calibration of a parameter-parsimonious thermodynamic model

Hadrien Delattre[1], Ahlem Filali[1], Elie Desmond-Le Quéméner[1,2], Etienne Paul[3], Théodore Bouchez[1]*.

[1]Irstea, UR HBAN, F-92761, Antony, France.

[2]LBE, Univ Montpellier, INRA, Narbonne, France

[3]LISBP, INSA Toulouse, France

Correspondence: T Bouchez, Irstea, UR HBAN, 1 rue Pierre-Gilles de Gennes CS 10030, F-92761 Antony, France

E-mail: theodore.bouchez@irstea.fr

Abstract

The development of predictive models plays a crucial role in bioprocess engineering. The most commonly used approach is to capture all the experimentally observed dynamics of the bioprocess into the calibrated parameters of an empirical engineering model. However, this solution is a double-edged sword since dynamics are reproduced while their causes are not elucidated. To address this issue, we developed the Microbial Transition State (MTS) theory, a theoretical model of microbial growth making the link between metabolic energy, growth yield and rate explicit. This model was shown in a previous study to make encouraging predictions of the dynamics of a simplified activated sludge community in a batch. In the present study we calibrate the kinetic parameters of this model on experimental data in order to evaluate the improvement it brings to the predictions. It is shown that the model then provide better predictions of the dynamics of the same microbial community in a batch and a continuous reactor, while keeping a low number of free parameters (4) and implementing a theoretical growth description only based on mass and energy balance constraints.

Introduction

The basic principle behind all mixed culture bioprocesses is that the upkeep of precise culture conditions, which can often be described in physical terms such as temperature and concentrations, will shape the microbial community so as to perform a set of targeted metabolic functions. Those culture conditions are known from empirical engineering rules. Models have been developed for the management and design of widely used mixed culture bioprocesses such as biological carbon and nitrogen removal (Gujer et al 2000) or anaerobic digestion (Batstone et al 2001), relying on an empirical expert knowledge of those processes. Although being able to quantitatively reproduce behaviors observed in the nominal conditions of the bioprocess, they are based on a partial understanding of the forces at play in the community's metabolic structuration. The adaptation of an existing mixed culture bioprocess model to different conditions (Ciggin et al 2013), or even the development of a model for a new bioprocess from metabolic specifications, require experimentations costly in time and money to calibrate the engineering model.

Generic models explaining microbial communities' metabolic structuration would be useful in those situations. Reciprocally, environmental bioprocesses, and especially activated sludge processes, constitute good model ecosystems for microbial ecology as their allow to study communities in well-defined and controllable culture conditions (Daims et al 2006). Attempts to derive generic relationships between physicochemical conditions (temperature and concentrations) and microbial growth yields have been proposed using thermodynamic calculations. Interestingly, the first thermodynamic model of microbial growth was developed in the context of activated sludge processes (McCarty 1965). This model and others since then have shown that there is a relationship between the thermodynamic properties of metabolisms and the abundance of microbes catalyzing it in a given culture medium (Heijnen and Dijken 1991, Roden and Jin 2011, Roels 1980, von Stockar et al 2008). While some methods implies to set some hypotheses about the electron transport chains of the modelled metabolisms to predict their yield and

rate (Jin and Bethke 2003, McCarty 1965, Noguera et al 1998), others are said to be “black box” and focus on the input and outputs of matter, entropy and energy (Heijnen and Dijken 1991, von Stockar and Liu 1999). Those methods have recently been applied to sets of metabolisms to predict the functional structuration of microbial communities in various chemical conditions (González-Cabaleiro et al 2015b, Van de Leemput et al 2011). However, while making the role of some physical factors explicit in the prediction of metabolic yields, for the modelling of growth rates, most of these approaches consist in binding energy-related terms (such as Gibbs energy differentials of reactions) into empirical equations. For example, the independent approaches developed by Heijnen and collaborators (Heijnen and Kleerebezem 1999), and Hoh and Cord-Ruwish (Hoh and Cord-Ruwisch 1996), to predict the growth rate associated to a metabolism in defined chemical conditions end up adding an energy-related term to a Monod growth rate equation. Thus, while empirical relationships between thermodynamic parameters and microbial growth are documented by the existing models, no underlying theory does satisfyingly explain this influence.

The Microbial Transition State (MTS) model is a microbial population dynamics model linking energetic gradients to population growth kinetics (Desmond-Le Quéméner and Bouchez 2014, Wade et al 2016). This model is aimed at describing generic mechanisms at play in the functional structuration of a microbial community so it can predict biomass and reagents concentration variation over time for a wide range of chemical settings. As the MTS model is built upon a statistical thermodynamic description of microbial growth, the comparison of its predictions with actual microbial growth outcomes provides feedback on the theory it implements. A previous article illustrated the ability of the MTS model to reproduce microbial community assembly patterns. For those predictions to be the result of the model’s construction, parsimony has been assumed as a working hypothesis as the number of population-specific parameters was kept to a minimum, and the model’s parameters were given default values. This first implementation of the MTS model was used to predict the growth dynamics of a microbial community

emulating this of activated sludge in a batch. The prediction was quite close to a reference simulation of the same system by the ASMN model, considering that this implementation of MTS had only 3 (uncalibrated) free parameters while the ASMN model had 12 parameters, each one calibrated on decades of measurements. However, predictions of the dynamics of an activated sludge community in a continuous reactor by this same implementation of MTS were nowhere close to those of ASMN. This is a concern for the use of the MTS model for bioprocess modelling as activated sludge is mainly cultivated in continuous culture. To be able to make accurate predictions of the dynamics of activated sludge in a continuous reactor is important for the use of the MTS model as it is the most common culture condition for this bioprocess. In order to make the MTS model predictions closer to the reference, knowledge about the value range of its kinetic parameters, and the sensibility of the model to them, needs to be gathered. In this article, we investigate the sensibility of the MTS model to its parameters value, and by extension, how the chosen level of parsimony (the number of different kinetic parameters) affects the predictive ability of the MTS model. This is done by calibrating the parameters of the model on experimental data, and then comparing results with the prediction of an uncalibrated MTS model (ref article 1), and with these of an engineering model (ASMN (Hiatt and Grady 2008)). The calibration process also greatly change the value of the MTS model's parameters from the values used in the previous article, so the results presented here also allow for an evaluation of the model's sensitivity to its parameters. The cases simulated are: (1) the competition for oxygen in a batch and (2) the washout by the reduction of sludge retention time and the competition for oxygen in a perfectly mixed continuous reactor with sludge recirculation.

Material and methods

Activated sludge metabolic model

The microbial community considered in this article is the main functional community of activated sludge, and is modelled as being divided into 3 microbial guilds: Ordinary Heterotrophic Organisms (OHO), Ammonium Oxidizing Bacteria (AOB) and Nitrite Oxidizing Bacteria (NOB), as in the ASMN model (Hiatt and Grady 2008). A guild is a microbial population characterized by an explicitly defined catabolic and anabolic reaction, a dissipated energy (Gibbs free energy differential of the overall metabolism, expressed in kJ per mol of biomass produced) and a biomass concentration. The dissipated energy of each guild's growth is a guild-specific value, computed according to the principles defined by Heijnen and collaborators (Heijnen and Dijken 1991). Biomass is modelled as a single, carbon-normalized molecule whose formula is as defined by Battley: $C_1H_{1.613}O_{0.557}N_{0.158}$, of molecular weight 24.76 $g.mol^{-1}$, of Chemical Oxygen Demand (COD) 1.708 $gDCO.g^{-1}$ and of Gibbs free energy differential of formation in standard conditions of -82.16 $kJ.mol^{-1}$ (Battley 1998). The anabolic and catabolic reactions defining each guild are;

- OHO:
    - Catabolism: 1 $C_2H_3O_2^-$ + 2 O_2 → + 2 HCO_3^- + 1 H^+ ($dG^{0'}$ = -844.4 $kJ.mol^{-1}$)
    - Anabolism: 0.503 $C_2H_3O_2^-$ + 0.158 NH_4^+ + 0.338 H^+ → 0.4305 H_2O + 0.0063 HCO_3^- + 1 $C_1H_{1.613}O_{0.557}N_{0.158}$ ($dG^{0'}$ = 23.9 $kJ.mol^{-1}$)
    - Dissipation: 432.12 $kJ.C\text{-}molBiomass^{-1}$
- AOB:
    - Catabolism: 1 NH_4^+ + 1.5 O_2 → 1 NO_2^- + 1 H_2O + 2 H^+ ($dG^{0'}$ = -269.9 $kJ.mol^{-1}$)
    - Anabolism: 1 HCO_3^- + 0.828 NH_4^+ → 1.101 H_2O + 0.670 NO_2^- + 0.499 H^+ + 1 $C_1H_{1.613}O_{0.557}N_{0.158}$ ($dG^{0'}$ = 267.7 $kJ.mol^{-1}$)
    - Dissipation: 3500 $kJ.C\text{-}molBiomass^{-1}$

- NOB:
    - Catabolism: 1 $NO2^-$ + 0.5 O_2 → 1 NO_3^- ($dG^{0'}$ = -79.1 kJ.mol^{-1})
    - Anabolism: 1 HCO_3^- + 2.64 NO_2^- + 1.16 H^+ → 0.27 H_2O + 2.49 NO_3^- + 1 $C_1H_{1.613}O_{0.557}N_{0.158}$ ($dG^{0'}$ = 241.3 kJ.mol^{-1})
    - Dissipation: 3500 kJ.C-$molBiomass^{-1}$

Where "$dG^{0'}$" denotes of the Gibbs energy differential of a reaction in standard conditions of temperature (298 K) and concentrations, corrected for a pH of 7 (the concentration of every chemical specie is assumed to be 1 M, except for H^+ which is at 1e-7 M).

MTS model

The MTS model is implemented as defined in chapter 4. Its growth rate formula is

$$\mu = \mu_m \cdot \prod_i e^{\frac{\upsilon_{an}+\lambda\cdot\upsilon_{cat}}{[S_i]\cdot V_h}} \quad (1)$$

Where μ is the growth rate of a guild (in day^{-1}), μ_m is the maximum growth rate of the guild (in day^{-1}), V_h is the harvest volume of the guild (the fictional volume into which substrate particles are accessible for a cell to metabolize them, in $m^3.mol_{Biomass}^{-1}$), $[S_i]$ is the concentration of a substrate S_i (in $mol.m^{-3}$) (a "substrate" being any chemical specie consumed according to the metabolism's stoichiometry), υ_{an} is the stoichiometric coefficient for the substrate S_i in the guild's anabolic reaction (in $mol_{Si}.mol_{Biomass}^{-1}$), υ_{cat} the stoichiometric coefficient for the substrate S_i in the guild's catabolic reaction (in $mol_{Si}.mol_{ElectronDonor}^{-1}$) and λ being the number of times the catabolic reaction is performed per anabolic reaction in order to close the energy balance of the metabolism, considering its Gibbs energy differential must equal the dissipated energy defined above for the guild (in $mol_{ElectronDonor}.mol_{Biomass}^{-1}$). V_h derives from the MTS theory. It represents the physical volume in which a cell can harvest molecular resources used for growth and might thus also be influenced by phenomena such as cell motility and substrate

diffusion. Though this article is the first attempt so far in the literature to calibrate the μ_m and V_h parameters of the MTS model, it is expected that those parameters are sensible to temperature. The λ factor is updated at each simulation time steps in order to account for the modification of the Gibbs energy differentials of anabolism and catabolism depending on the mass action ratio. The overall stoichiometric coefficient of each chemical specie S_i involved in the metabolism is then expressed as $\upsilon_{an} + \lambda \cdot \upsilon_{cat}$ (in $mol_{Si}.mol_{Biomass}^{-1}$), and the growth rate of each guild is then limited by a product of exponentials of the form $e^{\frac{\upsilon_{an}+\lambda\cdot\upsilon_{cat}}{[S_i]\cdot V_h}}$, taking the limitation by the concentration of each chemical specie required by the metabolism into account in the calculation of the guild's growth rate.

Activated sludge respirometric tests

Two respirometric tests conducted on activated sludge sampled from the aeration tank of a full-scale wastewater treatment plant (south of France), were used to calibrate the MTS model (μ_{max} and V_h). Before starting the tests, the activated sludge was maintained under endogenous conditions until the Oxygen Uptake Rate (OUR) reached a nearly stable signal. In the first experiment (further referred to as experiment 1), AllylThioUrea (ATU) was used (20 mg/L) to stop the nitrifying activity before injecting 25 mg/L of acetate. In the second experiment (further referred to as experiment 2), 5 $mgN.L^{-1}$ of ammonium was injected into the activated sludge mixed liquor. During both experiments, the pH was regulated at 7.3±0.3 and the temperature was maintained at 20 ±0.5°C.

Calibration process

MTS model parameters calibration for OHO guild is performed using the OUR curve of experiment 1 as a reference. AOB and NOB guilds are calibrated using the OUR curves of experiment 2. It is assumed that the observed oxygen consumption during the experiments happens because of two processes; exogenous and endogenous respiration. Exogenous respiration is the oxygen demand caused by the

catabolism of the added substrate and other processes such as its storage as polymers. It is considered that most of the exogenous oxygen demand is caused by catabolism, so the other processes are ignored.

To do so, the stoichiometry of the metabolisms is set so that their energy balance is closed, using the method defined by Heijnen and collaborators (Heijnen and Dijken 1991). It has been documented that it provides a reasonable estimation of the microbial growth yields without making assumptions based on experimental observations (von Stockar et al 2008). This method is also the one used by the MTS model to dynamically adjust stoichiometry. The total amount of oxygen needed for the catabolism of the introduced amounts of electron donors were deduced; the remaining amount of oxygen consumed in experiments was considered as endogenous respiration and removed from the curves on which the MTS model's parameters were calibrated. See sup mat for the experimental data, and its transformation as a R script (ref sup mat).

The calibration process consists in the heuristic search for the kinetic parameters of the MTS model (μ_{max} and V_h) for the simulation to fit the experimentally observed exogenous OUR curve. The parameters are calibrated using a MTS simulation program (ref article 1) written using the Matlab programming language (MATLAB Release 2014a, The MathWorks, Inc., Natick, Massachusetts, United States). The k_La of the calibration simulations has been set to 2000 day^{-1}, so oxygen is not limiting during the simulations as it is the case during the experiments. The distance between the MTS simulation and the reference is computed as the sum of the squares of the point-by-point differences between the two OUR curves (being expressed in $mgO_2.L^{-1}.min^{-1}$). The μ_{max} parameter's initial value for the OHO population was 33.1 day^{-1}, and its value is constrained to lie between 0.8 and 144 day^{-1}. 0.8 day^{-1} is the maximum growth rate usually considered for autotroph populations at 20°C in the ASM1 models (Henze et al 1987). While both the MTS and the Monod model do have a parameter named μ_m, it is not expected to have similar values since MTS's μ_{max} is subjected to much stronger limiting factors than this of Monod's model.

Consequently the calibrated value of μ_{max} should not be expected to go below this calibrated for the ASM models. 144 day^{-1} corresponds to the doubling time observed for vibrio natriegens, the fastest-replicating microbe according to current knowledge (Weinstock et al 2016). The initial V_h parameter's initial value for the OHO population was 2.5 $m^3.mol^{-1}$, Its value has not been constrained during the calibration as there exist no previous reference for the value of this parameter. The initial values have been chosen based on those assumed in the introductory paper of the model (Desmond-Le Quéméner and Bouchez 2014). The initial value of the μ_{max} and V_h parameters for the calibration of the AOB and NOB guild are the values calibrated for the OHO guild and the boundary values considered are the same. The initial OHO and total autotrophic biomass (AOB + NOB) concentrations were set to 1.89 mM (80 $gCOD.m^{-3}$) and 1.11 mM (47 $gCOD.m^{-3}$), respectively. The later were estimated based on the maximum value of the exogenous OUR recorded at the beginning of each test and considering μ_{max} and Y of each biomass (default values of ASMN were used). The respective proportion of AOB and NOB guilds biomass density is a supplementary calibrated parameter in the calibration of autotrophs parameters, constrained so that the AOB initial biomass cannot be higher than this of OHO and the NOB initial biomass cannot be higher than this of AOB. The initial assumed proportion is 50% / 50%. The ASM parameters for the estimation of initial biomass are taken from the ASMN model.

In order to provide a comparison with the previous article, simulations of the uncalibrated MTS model have been performed alongside with those of the calibrated model. The set of parameters of this model corresponds to this used in the previous article (cf chapter 4);

$$\begin{cases} \mu_{max}^{OHO} = \mu_{max}^{AOB} = \mu_{max}^{NOB} = 5.35e17\ day^{-1} \\ V_h^{OHO} = 1\ m^3.mol^{-1} \\ V_h^{AOB} = V_h^{NOB} = 10\ m^3.mol^{-1} \end{cases}$$

Simulation conditions

The effects of the calibration on the predictions by the MTS model are evaluated by performing a simulation of the dynamics of the calibrated guilds in two different culture conditions. One culture condition simulated in this article corresponds to an aerated batch, at pH 7, a temperature of 20°C, a volumetric mass transfer coefficient k_La of 100 day^{-1} (line with the order of magnitude of values of k_La experimentally estimated in wastewater treatment plants (Capela et al 2004)) and the following initial concentrations;

- Acetate: 1.76 $mol.m^{-3}$ (110 $gCOD.m^{-3}$)
- Ammonium : 3.78 $mol.m^{-3}$ (53 $gN.m^{-3}$)
- Bicarbonate: 4.91e-1 $mol.m^{-3}$ (30 $g.m^{-3}$)
- OHO biomass: 11 $mol.m^{-3}$ (42.29 $gCOD.m^{-3}$)
- AOB biomass: 1 $mol.m^{-3}$ (42.29 $gCOD.m^{-3}$)
- NOB biomass: 1 $mol.m^{-3}$ (42.29 $gCOD.m^{-3}$)

The second growth condition corresponds to a continuous reactor with a volume of 1000 m^{-3} connected to a point settler. Influent flow rate was set to 1000 $m^3.day^{-1}$; whereas, waste flow rate was varied from 10 to 999 $m^3.day^{-1}$ to simulate a sludge retention time ranging from 1 to 60 days. To assess the effect of oxygen concentration on microbial dynamics, two values of the volumetric mass transfer coefficient were used (k_La: 45 d^{-1} and 100 d^{-1}). Finally, simulations were performed considering a temperature of 20°C, a pH 7 and the following input concentrations;

- Acetate: 4.75 $mol.m^{-3}$ (300 $gCOD.m^{-3}$)
- Ammonium : 3.57 $mol.m^{-3}$ (50 $gN.m^{-3}$)
- Bicarbonate: 4.91e-1 $mol.m^{-3}$ (30 $g.m^{-3}$)
- OHO biomass: 3.90 $mol.m^{-3}$ (165 $gCOD.m^{-3}$)

- AOB biomass: 3.08 mol.m^{-3} (131 gCOD.m^{-3})
- NOB biomass: 0.99 mol.m^{-3} (42 gCOD.m^{-3})

The initial biomass concentrations are the stabilized biomass concentrations obtained with the maximum SRT. It has been done so in order to speed up time required by the system to stabilize.

ASMN model

A mathematical model modified from the well-established Activated Sludge Model no. 1 (Henze et al 1987) was used for modelling the batch and continuous simulation conditions with a simplified activated sludge community. The original model was extended with a two-step nitrification and simplified so as to describe only the aerobic growth of heterotrophs (X_{BH}), ammonia-oxidizing bacteria (X_{AOB}) and the nitrite-oxidizing bacteria (X_{NOB}). The parameters of the model are detailed as supplementary material 1. The simulation software used is WEST (MIKEbyDHI.com).

Results

The uncalibrated MTS model (UMM) simulated in this article have been used in a previous implementation of the model (cf chapter 4). A microbial community emulating this of an activated sludge in a batch reactor was simulated and it appeared that the yields from the UMM were close to those of a parallel implementation of the ASMN model. However, the UMM's prediction of the dynamics of the same microbial community in a continuous reactor were very different from the simulation by the ASMN model. The results below show the result of the calibration process of the MTS model, giving rise to the Calibrated MTS Model (CMM), and then compared simulations of the aforementioned activated sludge community by the CMM, UMM and ASMN, in a batch then in a continuous reactor, in order to document how the calibration of the kinetic parameters of the MTS model change its predictions of the yields and dynamics of a microbial community.

Calibration of the MTS model on an experimental record of activated sludge respiration of acetate

A respirometric experiment was performed using acetate as electron donor. Oxygen concentration was continuously monitored and corresponding data were used to calibrate a MTS model representing the growth of OHO populations. The calibration of the parameters of the OHO population on the OUR of experiment 1 yielded the following heuristically optimal parameter set;

$$\begin{cases} \mu_{max} = 3.75\ day^{-1} \\ V_h = 147\ m^3.mol^{-1} \end{cases}$$

The calibrated OHO parameter set was then used as a starting point to calibrate the AOB and NOB guilds. As a working hypothesis for the purpose of parsimony, it was assumed that the two autotrophic guilds AOB and NOB do have the same μ_{max} and V_h, while the initial biomass concentration was deemed to be guild-specific. The parameter set thus obtained by calibration on the OUR curve of respirometric experiment 2, performed using ammonia as electron donor:

$$\begin{cases} \mu_{max} = 2.08\ day^{-1} \\ V_h = 296\ m^3.mol^{-1} \\ X_0^{AOB} = 6.52e-1\ mol.m^{-3}(27.57\ gCOD.m^{-3}) \\ X_0^{NOB} = 4.59e-1\ mol.m^{-3}(19.41\ gCOD.m^{-3}) \end{cases}$$

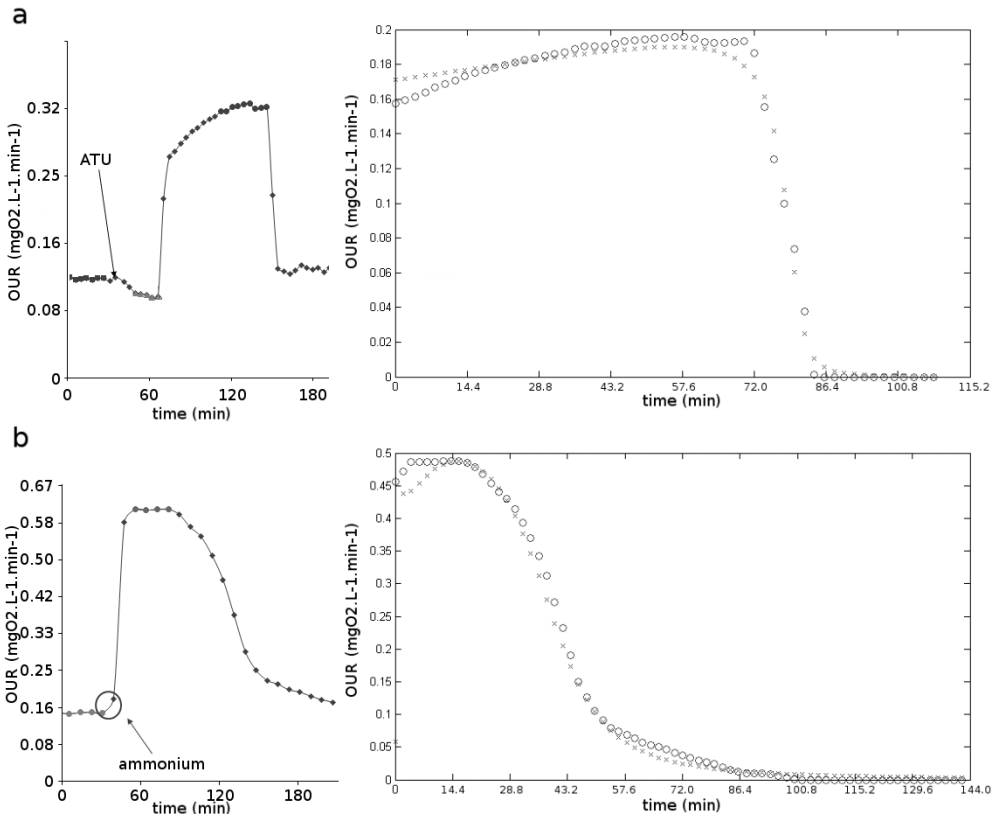


Figure 4 : Left: OUR curve of the raw experimental data, with injections in the culture medium specified. Right: Superimposition of the experimentally observed OUR curve (circles) with the prediction by the MTS model with parameters minimizing the summed square distance to the observed data (crosses). X axis is time in min, Y axis is oxygen consumption in $mgO_2.L^{-1}.min^{-1}$.a) experiment 1; respirogram following the injection of acetate, b) experiment 2; respirogram following injection of ammonium. Timescale in left and right column does not start at the same point.

Sensitivity of the simulations to the parameters

In order to evaluate the sensitivity of the MTS model's simulation to the value of its parameters μ_{max} and V_h, and also the identifiability of the parameters, the distance of simulations by the MTS model to the experimental references used to calibrate it have been computed using sets of value for μ_{max} and V_h varying around their optimized value. The k_La has been set to 2000 day^{-1} for those simulations, so the populations are not limited by oxygen. The results for the three references are displayed on Figure 2. It appears that there exists a single optimal point regarding each reference data in the $\mu_{max} \times V_h$ parameter space. Consequently the kinetic parameters of the MTS model can then be considered as formally identifiable. However, as the optimum (μ_{max}, V_h) couple lies in a shallow "valley" inside surface's topology, it should be outlined that the accuracy of the MTS model simulations is not very sensitive to

the value of the parameters (±25% variations of both parameters stays inside the valley, for both experiments).

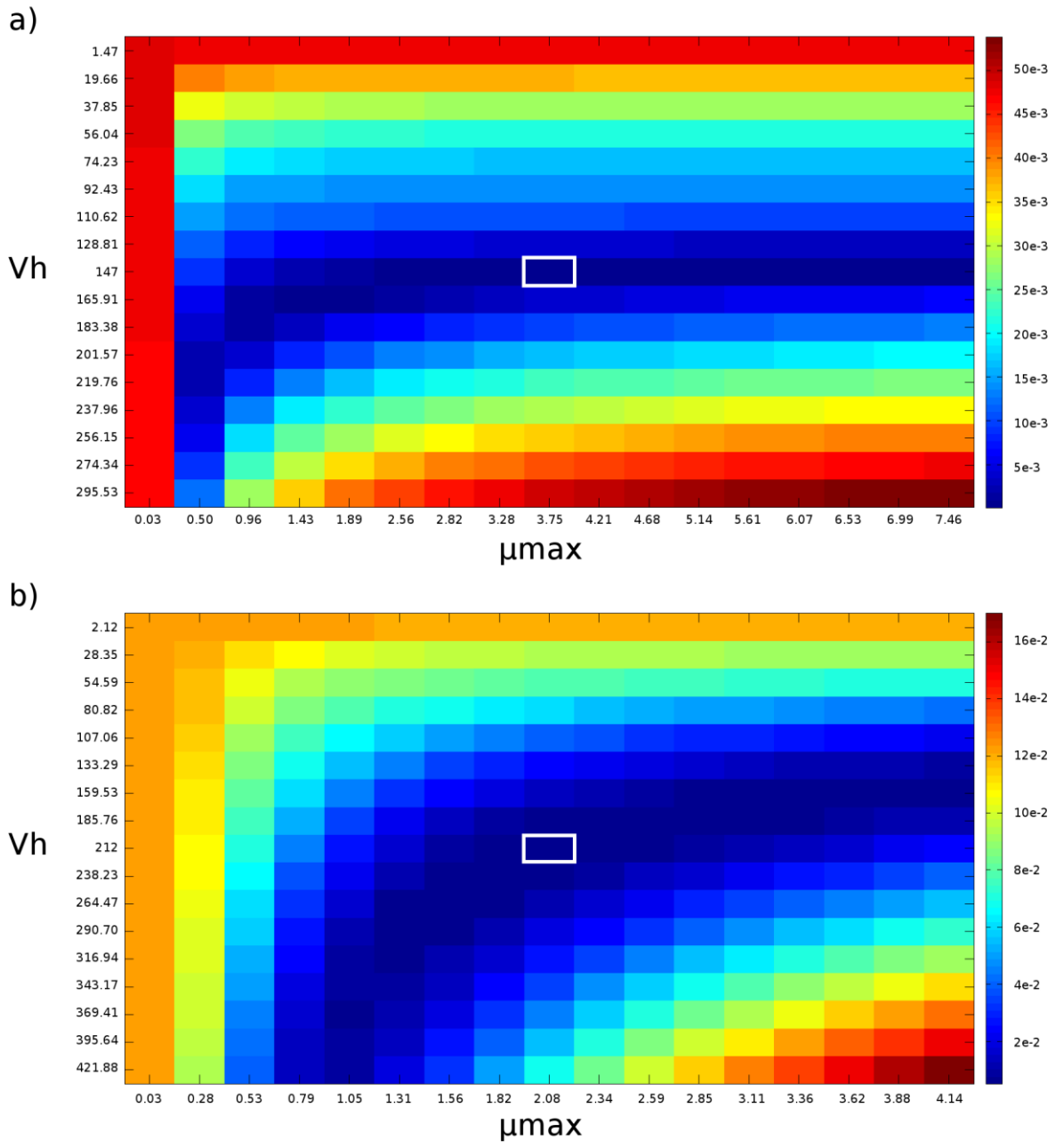


Figure 2 : Distance of the MTS OUR prediction from the experimental reference, depending on the couple of µmax and Vh values. The value of µmax varies along the X axis, the value of Vh varies along the Y axis. The optimum, calibrated (μ_m; V_h) couple lies in the center of each subfigure (white frame). The distance is expressed as the sum of squared difference between points, normalized by the number of points in the measurement. (a) distance to experiment 1 (b) distance to experiment 2

Simulation of the population dynamics of the guilds in a batch

The simulation of the OHO, AOB and NOB guilds is performed in batch culture conditions by the CMM, the UMM and an equivalent implementation of the ASMN model. The prediction by the CMM is compared to this of the UMM to evaluate the effect of the calibration on the MTS model, to this of the ASMN simulation which is here considered as a reference for the dynamics of this system (Figure 3).

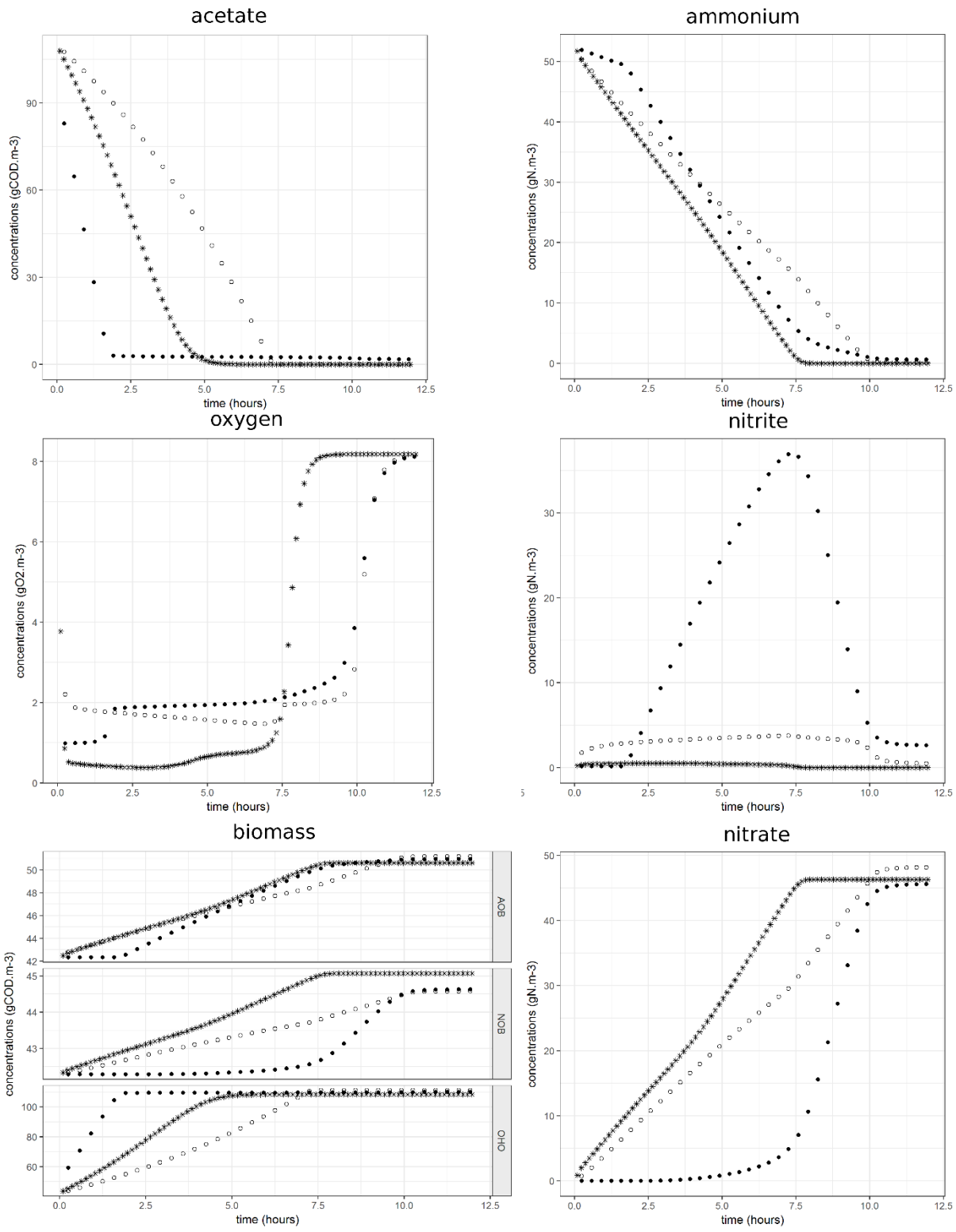


Figure 3: Concentration of the metabolites and biomasses over time in a batch setting, as predicted by an uncalibrated MTS model (●), a calibrated MTS model (○) and a modified ASMN model (*). Time is in hour, concentrations are expressed in ASM units (gCOD, gN, gO_2 per m^3).

The UMM predicts that the OHO population grows immediately, and the AOB and NOB populations start to grow significantly later at respectively 1.83 h and 7.2 h. This growth lag is linked with a predicted nitrite accumulation (up to 2.66 $mol.m^{-3}$, i.e. 37.26 $gN.m^{-3}$) before the NOB population growth. In contrast, the CMM predicts the immediate growth of the AOB and NOB populations. The OHO guild reaches its stationary state 3 hours later than the UMM, and the final OHO biomass density predicted by the CMM is also 2.60 $mol.m^{-3}$ (1.10 $gCOD.m^{-3}$) higher than this predicted by the UMM. Minor differences are also observed in the dynamics of AOB and NOB guilds. Reciprocally to the different growth dynamics of AOB and NOB, nitrite accumulation in the culture medium is far less important according to the CMM model than as by the UMM (up to 2.5e-1 $mol.m^{-3}$, i.e. 3.58 $gN.m^{-3}$), so the CMM predicts nitrite dynamics far closer to those obtained with the ASMN model than to those of the UMM.

Simulation of the population dynamics of the guilds in a continuous reactor

In the context of wastewater treatment, activated sludge develops in continuous reactors. The dynamics of chemical species in this environment are mediated by microbial guilds, whose competition give rise to dynamic patterns known as for now as engineering rules. In order to evaluate whether the calibration of the MTS kinetic parameters allows for more “realistic” predictions of population dynamics in a continuous reactor, simulations by the CMM, UMM and the ASMN model have been performed, with different values of SRT and k_La (Figure 4). Again, the ASMN model’s simulation has been considered as a reference for the dynamics in such system. The data used to generate the figure consists in the stable state of systems characterized by different conditions of k_La and SRT. Two values of k_La have been tested; 45 day^{-1} and 100 day^{-1}, and values of SRT ranging from 60 to 1 day have been tested on each. An additional set of simulations has been implemented at k_La 45 day^{-1}, corresponding to the simulation of

the CMM except that the AOB and NOB are given the same parameters as the OHO guild (visible on Figure 4). The data is available as supplementary material.

While the growth yield on the electron donor (acetate, ammonium or nitrite) is similar in both models, the rest of the stoichiometry is different and this causes some discrepancies between the stationary states predicted by the two models. Indeed, the effluent total free nitrogen mass (ammonium, nitrite and nitrate altogether) is higher according to the MTS model (40 $mgN.L^{-1}$) than according to the ASMN model (approximately 35 $mgN.L^{-1}$), and the OUR predicted by the MTS model (339 $mgO_2.L^{-1}$ at 60 days SRT, 45 day^{-1} k_La) is higher than this predicted by the ASMN model (266 $mgO_2.L^{-1}$ at 60 days SRT, 45 day^{-1} k_La). Moreover, the OUR normalized by the concentration of stabilized biomass is higher for every guilds at every SRT according to MTS prediction when compared with ASMN simulation.

The loss of the nitrifying activity along with the diminution of the SRT is visible, in all simulation records of the ASMN and CMM, as the change of the concentrations of the free nitrogen forms (ammonium, nitrite and nitrate). Indeed, when the nitrifying activity is fully catalyzed (both AOB and NOB guilds being active) the effluent nitrogen is mostly in the nitrate form, while when the NOB guild get washed out, the effluent nitrate concentration decreases and gives way to nitrite. In turn, when the AOB guild's activity decreases, ammonium is not metabolized and becomes the main free nitrogen form in the effluent. The loss of the nitrifying activity occurs in two stages (loss of NOB activity then loss of AOB activity) according to both ASMN's and CMM's simulations. Interestingly, it is predicted by the CMM whether heterotroph and autotrophs guilds are given specific parameters or not. In the ASMN simulation for k_La 45 day^{-1} and k_La 100 day^{-1}, and in the MTS simulation at k_La 45 day^{-1}, the decrease of the NOB guild's activity happens at a SRT of approximately 6 days. The decrease of the NOB guild's activity is followed by the decrease of AOB guild's activity. While the ASMN model predicts the AOB activity to totally disappear for a SRT of 1.3 day, the MTS model also predicts that the nitrifying activity decreases with SRT, but far more gradually,

to the point the nitrifying guilds are still not washed out when the minimum SRT (1.0 day) is reached. The loss of the nitrifying activity only starts and is too progressive to actually happen before the minimum SRT is reached in the simulation of the MTS model at k_La 100 day^{-1}. The loss of the nitrifying activity along with the decrease of the SRT is predicted by the CMM, but is not by its uncalibrated counterpart in the tested SRT range (Figure 4).

Discussion

Sensitivity analysis

The evaluation of the sensitivity of the kinetic parameters of the MTS model displayed on Figure 2 shows that the calibrated V_h and μ_{max} parameters are formally identifiable, that is, for a given reference dataset there exist a single (μ_{max}, V_h) couple minimizing the distance between the model's prediction and the reference data.

The approach of the ASM models toward real systems is to indiscriminately capture all the phenomena influencing the microbial growth yield and dynamics in their parameters. The characteristic of this approach is the quantitative accuracy of the predictions without the need to characterize those phenomena individually, which is a double-edged sword. The MTS approach is to have the influence of stoichiometry and thermodynamics directly accounted for by the construction of the model. The kinetic parameters, once calibrated, represent a residual of yet uncharacterized phenomena influencing growth dynamics. This residual is shown in the simulations results to have less influence on the dynamics than stoichiometry and thermodynamics, since the predictions made by the MTS model are not very sensitive to the value of those parameters, as pictured by the sensitivity graphs (Figure 2). This low sensitivity of the MTS model's predictions to its parameters' value is also explained by the fact that since the populations' growth is not limited by oxygen, the V_h parameter has not much influence on the dynamics.

It should be noted that experimental respirometry measurement, on which calibration is based, is performed in a culture medium where oxygen is not a limiting resource.

Moreover, as shown by the batch predictions (Figure 2), if a substrate is limiting for growth, the influence exerted by the concentration of this substrate on growth dynamics cannot be cancelled by kinetic parameters as the limiting factor of a substrate is a negative exponential function of its concentration (cf equation 1).

Batch results

The calibrated and uncalibrated MTS models are compared in this section. A simulation by the ASMN model is also juxtaposed to the data as a reference. However, comparison with it should be qualitative and not quantitative, since the CMM has been calibrated in conditions which may have implied different parameters for the ASMN model (as respirometry experiments constitute growth conditions different from the nominal conditions of municipal wastewater).

The calibration of the kinetic parameters does not affect heterotroph and autotroph guilds' dynamics the same way. Indeed, it can be seen on Figure 2 that the calibration makes the OHO guild's growth slower while the growth of the autotroph guilds is quicker and starts earlier (UMM predicts a lag phase before AOB and NOB growth, while this lag phase is absent in CMM's prediction).

To correlate those observations to the growth formula of the MTS model, let's consider a single limiting substrate S and neglect its contribution to anabolism for the sake of simplicity. Its limiting factor applied to μ_m is then $e^{-\frac{\lambda}{V_h \cdot [S]}}$, that is $e^{-\frac{dG_{dis}+dG_{an}}{V_h \cdot dG_{cat} \cdot [S]}}$ (Desmond-Le Quéméner and Bouchez 2014). The $dG_{dis} + dG_{an}$ term is the energy barrier of microbial growth (kJ.C-$mol_{Biomass}^{-1}$) and $V_h \cdot dG_{cat} \cdot [S]$ the catabolic energy inside the harvest volume (kJ.C$mol_{Biomass}^{-1}$). Generally speaking, the higher the catabolic energy is compared to the energy barrier, the closer the limitation factor applied to μ_{max} is to 1. When the

limitation factor is close to 1, the growth rate μ is close to its maximum μ_{max}, so the μ_{max} parameter has a great influence on the simulation of the population's dynamics. On the contrary, when the available catabolic energy $dG_{cat} \cdot [S]$ is low, the influence of μ_{max} on the dynamics fades and V_h becomes the most influent parameter, simply because μ_{max} has a linear influence on μ while V_h is exponentiated.

As the OHO guild catalyzes the most exergonic electron transfer to oxygen, the limitation exerted by its substrates at the beginning of the simulation is not very important, consequently its growth rate is conditioned by the value of its μ_{max}, this is why its growth is slower according to CMM. On the other hand, the electron transfer catalyzed by the autotroph guilds AOB and NOB is less exergonic than this of OHO; the limiting substrate's concentration becomes the main factor influencing the population's growth dynamics. In this situation, the V_h parameter becomes of importance as it modulates the sensitivity of the growth rate function to the substrate's concentration.

While the increase of V_h through calibration was of little effect on the OHO guild's dynamics, it has a visible effect on the autotroph guilds' growth dynamics as they are limited by their substrate (may it be oxygen for AOB or both oxygen and nitrite for NOB). The most visible effect of the increase of V_h for the autotroph populations is the disappearance of the nitrite peak in CMM's prediction. The low value of V_h for autotroph populations in UMM makes the NOB population very sensitive to nitrite concentrations so they require nitrite to accumulate up to a high value (33 $gN.m^{-3}$) before they start to significantly grow. This nitrite accumulation phenomenon has been reported in some instances (Rajagopal et al 2011); however, it contradicts the simulation by ASMN for this setting (as the nitrite half saturation coefficient is low; K_{NO2} = 0.5 $gN.m^{-3}$). The increase of V_h value for autotroph populations in CMM allows the NOB population to significantly grow at a lower concentration of nitrite, thus making CMM's prediction qualitatively closer to this of ASMN. Considering all substrates, an increase of the value of V_h makes a population able to grow significantly with a lower concentration of substrate. Reflecting this, it is visible

in the results shown on Figure 2 that the final concentration of acetate, ammonium and nitrite is lower as predicted by CMM (where V_h is higher for all guilds), and the total amount of produced biomass is respectively higher. It can then be interpreted that the V_h parameter, although being a kinetic parameter, have an influence on growth activation, as it is involved in the ratio of energy barrier over available catabolic energy ($\frac{dG_{dis}+dG_{an}}{V_h \cdot dG_{cat} \cdot [S]}$) for biomass production. However this effect is rather small considering that multiplying V_h by more than 100 increases the total amount of biomass produced from substrate by less than 1% in the timeframe of the simulation.

From these observations it can be said that the μ_{max} parameter has more influence on populations not very limited by their substrate while the V_h parameter has more influence on populations facing important limitation by their substrate.

Continuous reactor results

Before discussing the characteristics of the models in terms of dynamics, we must address a number of quantitative differences between the stationary states predicted by the MTS and ASMN models, as put into light by the results (Figure 4 and 5). Those differences result from stoichiometric differences between the models.

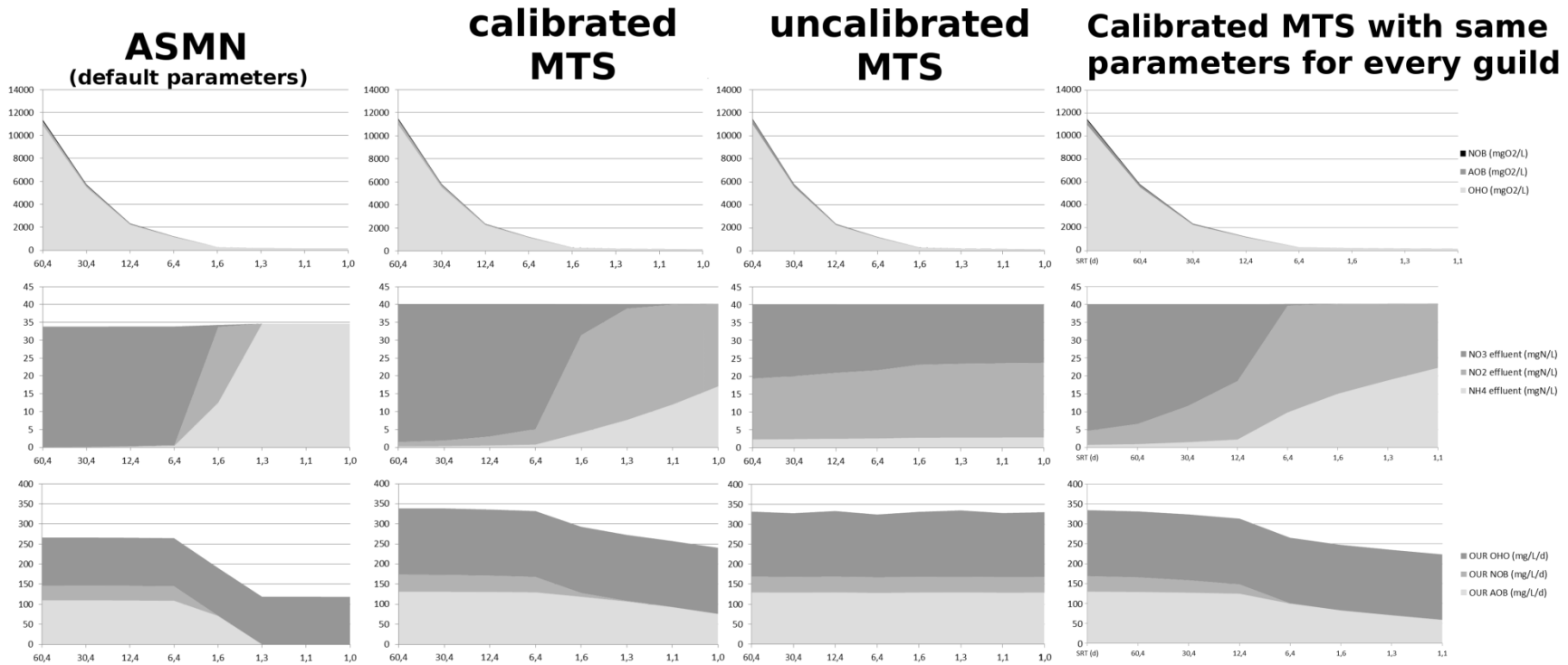

ASMN
(default parameters)
calibrated MTS
uncalibrated MTS
Calibrated MTS with same parameters for every guild
NOB (mgO2/L)
AOB (mgO2/L)
OHO (mgO2/L)
NO3 effluent (mgN/L)
NO2 effluent (mgN/L)
NH4 effluent (mgN/L)
OUR OHO (mg/L/d)
OUR NOB (mg/L/d)
OUR AOB (mg/L/d)
SRT (d)
60,4
30,4
12,4
6,4
1,6
1,3
1,1
1,0

Figure 4: Concentrations of biomass of OHO, AOB and NOB guild (first row), ammonium, nitrite and nitrate (second row) and Oxygen Uptake Rate (third row) as predicted by the ASM (first column), calibrated MTS (second row), uncalibrated MTS model (third row) and a calibrated MTS model where all guilds have the same parameters, which are the parameters calibrated for the OHO guild in this contribution (μmax = 3.75 day 1 and Vh = 147 m3.mol 1). The culture medium in each case is a continuous reactor with variable Sludge Retention Time (abcisses) and kLa = 45 day 1

ASMN
(default parameters)

calibrated MTS

uncalibrated MTS

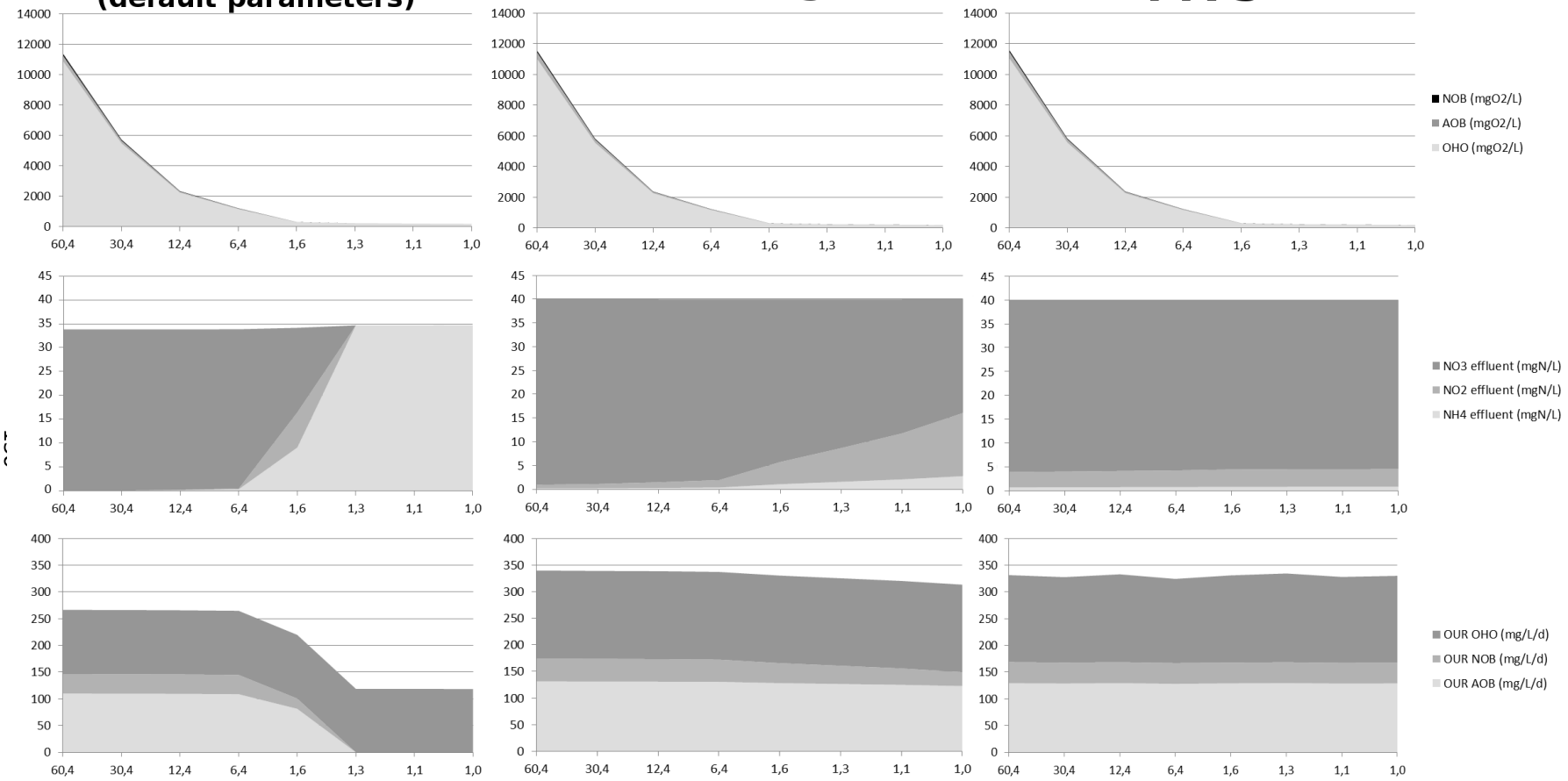


Figure 5: Concentrations of biomass of OHO, AOB and NOB guild (first row), ammonium, nitrite and nitrate (second row) and Oxygen Uptake Rate (third row) as predicted by the ASM (first column), calibrated MTS (second row) and uncalibrated MTS (third row) in a continuous reactor with variable Sludge Retention Time (abscises) and k_La = 100 day^{-1}

The difference between the amounts of effluent free nitrogen in both models can be explained by the fact that they do not rely on the same biomass composition. Indeed the MTS model explicitly considers Battley's biomass, of formula $C_1H_{1.613}O_{0.557}N_{0.158}$ measured from a Saccharomyces cerevisiae pure culture (Battley 1998) implying a nitrogen requirement of 5.23e-2 gN.gCOD-$biomass^{-1}$, while the ASM model implicitly considers Hoover and Porges' biomass $C_5H_7O_2N$ ($C_1H_{1.4}O_{0.4}N_{0.2}$) measured from a mixed culture growing on dairy waste (Hoover et al 1953), implying a nitrogen requirement of 8.76e-2 gN.gCOD-$biomass^{-1}$. The higher quantity of fixed nitrogen in Hoover and Porges' biomass explains the lower effluent concentration of free nitrogen in the ASMN simulations.

On the other hand the higher OUR per unit biomass predicted by the MTS model can be attributed to neither anabolism nor catabolism as the growth yield on the electron donor is close in both models, and it is assumed that anabolism does not involve oxygen. However, the values used by ASMN should not be considered as an absolute reference to this regard as the parameters of the MTS model have been calibrated in conditions into which the parameters of the ASMN model would have been different, as respirometry experiments are performed in culture conditions different from those of municipal wastewater in an aeration tank. Nevertheless, oxygen is predicted in the results to be a far more limiting resource according to the MTS model than to the ASMN model since the stabilized oxygen concentrations are always higher when predicted by the ASMN model than by the MTS model. The crossed effect of k_La and SRT on nitrification loss is much more important according to MTS than to ASMN.

Now regarding the dynamics of the system, the constraint on growth is twofold in this simulated environment; first the SRT sets the growth rate the guilds must reach in order to stabilize, second the oxygen concentration limits the growth rate of each implemented guild. Schematically speaking, the

equilibrium state in this system depends on three interacting variable; the substrate concentration, the growth rate of the considered population and the biomass concentration of the population. The substrate concentration increases the growth rate; the growth rate increases the biomass concentration and the biomass concentration decreases the substrate concentration. There is then a negative feedback loop between the biomass and the substrate concentration. Specifically considering dissolved oxygen as the substrate, two parameters come into play; the k_La increases the dissolved oxygen concentration, and the SRT increases the biomass concentration and decreases the oxygen concentration. This scheme results from the system's structure and does not depends on the growth model considered, so it is true for both MTS and ASMN.

A consequence of this scheme is that the SRT defines the growth rate μ^* that a guild has to reach in order for its biomass concentration to be stabilized in the reactor. To reach a higher μ^*, the stabilized biomass concentrations X^* of each guild decreases so as to reduce their resource consumption and then increase the concentration of available resources in the culture medium. The more a population's growth is limited by its resources in a given environment, the lower its X^* (and consequently its catalytic activity) has to be low in order to be stabilized at a given SRT.

Both ASMN and MTS models predict that stabilized biomass decreases with the SRT. Both models predict a two-stage loss of the nitrifying activity, as the concentration of the free forms of nitrogen in the culture medium depends on the SRT, from almost only nitrate to a mix between ammonium, and nitrite to almost only ammonium at low SRT. This pattern, which is critical for the management of wastewater (Van Dongen et al 2001), was captured as an empirical engineering rule by the calibration of the ASMN model, notably by attributing different affinity for oxygen to the AOB and NOB guilds. On its part, the MTS model reproduces this pattern as a consequence of the difference of exergy (and thus resources requirement) between the two metabolisms. Indeed, the phenomenon occurs in the MTS simulations

while both the AOB and the NOB guilds are given the same kinetic parameters. Moreover, it is also reproduced by the MTS model when all guilds have the same parameters (Figure 4). This kinetic difference between the AOB and NOB guild, which is enforced in the ASMN model through the calibration of the kinetic parameters on experimental data, then appears to arise from the physicochemical variables accounted for by the MTS model. Though the two-stage loss of nitrification prediction by the UMM is hardly observable from the graphical results, the evolution of concentrations and limitations along the SRT suggests that it happens at an unrealistically low range of SRT (below 1 day, which cannot be achieved in this setting). Indeed, the prediction of the two stage loss of nitrification arises from the MTS model's non calibrated parameters, such as the yields. The effect of the calibration of μ_m and V_h on respirometric data only change the range of SRT at which the phenomenon occurs to a range close to the one practically observed.

As the limiting factor applied to the μ_m parameter is inversely correlated with the value of the SRT in both model, the higher the SRT the more the guilds are limited by their substrate (limitation data are provided as supplementary material). According to the role of the μ_m and V_h parameters as it has been described in the batch section, the V_h parameter has more influence at high SRT while the μ_m value has more influence at low SRT. Also, the observation of the limitations shows that the most limiting substrate of the AOB and NOB guilds as simulated by the MTS model switches from their electron donor (respectively ammonium and nitrite) as high SRT to their electron acceptor (oxygen) at low SRT.

To summarize, both the ASMN and the MTS model predict the two stage loss of nitrification, however the MTS model predicts it as arising from physicochemical parameters, that which are not calibrated but inferred from an accurate description of the growth environment. The differences between the two models lay in the sensibility of the response of the variables, thus influencing the range of SRT at which

the phenomenon occurs. Those differences depend on the kinetic parameters, and stoichiometric differences between the two models.

Conclusion

The result we present indicates the MTS model is able to reproduce behaviors (such as the nitrate shunt) previously modelled as empirical engineering rules solely based on a microscopic description of microbial growth and the thermodynamic properties of the involved chemical species. The fact that the MTS model's predictions are not very sensitive to the value of its kinetic parameters indicates that the model is able to produce consistent predictions from matter and energy conservation constraints alone in this kind of system. The calibration of the kinetic parameters of the model on respirogram experiments improved the MTS model's prediction of phenomena different from respiration, as seen in the sections on batch and continuous reactor simulation. Moreover, the results presented in this contribution provide an evaluation of the ability of the MTS to predict chemical dynamics in an environmental bioprocess, how it is affected by the level of parsimony considered and gives a first empirical estimate of the values of the parameters. Notably, the weak sensibility of the MTS model's predictions to its kinetic parameters suggests that energy and matter flows as described by the model are a major factor driving microbial community dynamics, so the MTS model could be applied to other bioprocesses as a generic microbial community structuration model.

Chapter 6:
Predicting microbial growth yield from the nature of metabolic reactions

This chapter corresponds to an article draft not submitted yet. This article investigates the relationship between physicochemical characteristics of metabolisms and the associated growth yield.

Investigating this relationship equates to investigating the relationship between the physicochemical characteristics of metabolisms and the energy balance of microbial growth. While the two previous chapters investigated the kinetic aspects of the MTS model, the dynamics predicted by the model depends on the energy balance of the modelled metabolisms. However the MTS model itself does not propose a theory to equilibrate the energy balance of microbial growth. Instead, it has to borrow an empirical relationship from the literature. In the two previous chapters, the assumed relationship was the one defined by Heijnen and collaborators and described in chapter 3 of this memoir. However, more recent attempts have been made to calibrate more accurate relationships, using wider datasets.

The current chapter then presents an effort to qualify more robustly the relationship between metabolisms' physicochemical parameters (catabolic energy, number of carbons of the carbon source...) and the metabolisms' growth yield. This is then an attempt to improve the current knowledge on metabolism energy balance, which is the "static" part of microbial thermodynamics, and on whose dynamics predictions relies.

Predicting microbial growth yield from the nature of metabolic reactions

Hadrien Delattre[1], Elie Desmond-Le Quéméner[1,2], Robbert Kleerebezem[3], Théodore Bouchez[1]*

[1] Irstea, UR HBAN, F-92761, Antony, France

[2] LBE, Univ Montpellier, INRA, Narbonne, France

[3] Department of Biotechnology, Delft University of Technology, Julianalaan 67, Delft 2628 BC, The Netherlands

Correspondence: T Bouchez, Irstea, UR HBAN, 1 rue Pierre-Gilles de Gennes CS 10030, F-92761 Antony, France

E-mail: theodore.bouchez@irstea.fr

Introduction

An accurate method to estimate the growth yield of a microbial population is critical for the modelling of equilibria and dynamics of microbial populations and communities. Both the modelling of natural ecosystems and the engineering of bioprocesses benefit from the improvement of such methods.

Since the 1960s, many attempts have been made to correlate physicochemical properties of the metabolisms to their growth yield (Bauchop and Elsden, 1960; Mayberry et al., 1968; Minkevich and Eroshin, 1973; Roels, 1983). However most of these attempts produced yield predictors plagued by interal inconsistencies and/or valid only for a certain category of metabolisms (Heijnen and Dijken, 1991).

The most accurate growth yield predictions methods nowaday are based on expert knowledge about microbial thermodynamics (McCarty, 2007; Heijnen et al., 1992; Liu et al., 2007). Those models boil down to considering microbial growth as an energy transfer phenomenon between catabolism and anabolism. Some of the Gibbs energy generated by catabolism is conserved in anabolism, while another part is "dissipated". Dissipated energy is the part of catabolic energy which is not converted into biomass. It may be either invested in maintenance or expelled out of the cell as heat or molar entropy ((von Stockar and Liu, 1999) for a detailed description of this phenomenon). For instance, Heijnen and collaborators developed in the 1990s a formula linking the dissipated energy per amount of biomass synthetized to the number of carbons and reduction degree of the carbon source of the metabolism (Heijnen et al., 1992). Another more accurate formula based solely on the electron donor's reduction degree was later proposed by Liu and collaborators (Liu et al., 2007). This suggests that the prediction of microbial growth yield based on physicochemical parameters can still be improved by the enlargement of data sets. Another example of growth yield predictor from thermodynamic variables is the model developped by McCarty and collaborators since the 1960s estimates by assuming an "energy transfer efficiency" between catabolism and anabolism (McCarty, 1965). While this parameter was initially assumed to be the same for all metabolism, the authors later noted that its value is likely to depend on the metabolism (McCarty, 2007). Finally, as there is no widely accepted theory explaining the values taken by metabolic efficiency up to now, the expert models predicting it are empirical models calibrated on experimental data, and are likely to be improved.

A relatively recent attempt to directly predict the microbial growth yield from the Gibbs energy change of catabolism led to encouraging growth yields predictions (Roden and Jin, 2011). This result opens the door to

yield prediction models requiring less theoretical knowledge to be used.

This contribution examines new growth yields prediction formulas based on a more comprehensive set of thermodynamic variables and more robust than those previously proposed in the literature. For that, the largest literature collection to date was gathered in order to infer and calibrate the statistical models. This dataset, available as a supplementary material, connects experimental microbial growth yield data to thermodynamic information about their metabolisms. It leverages on multiple previous data collections ((Roden and Jin, 2011; Liu et al., 2007; Batstone, 2001)). While two of those data collections allowed to propose new growth yield correlation formulas or to validate previous estimations ((Roden and Jin, 2011; Liu et al., 2007)), the other was purposed at the calibration of an engineering model (ADM1). At the end, the exploration of this substantial dataset not only allows for the calibration of robust microbial growth yield predictors, but also gives pointers to efforts to be made for the improvement of such predictors.

Material and methods

Collection and computation of experimental data

The data consist in a set of experimental yield measurements, to which physicochemical characteristics (catabolic and anabolic stoichiometric formulas etc), as well as metadata (organism, bibliographical reference etc) are associated. It is available as a supplementary material in the form of a table (ref sup mat 1).

The data were collected from three previous data collections, respectively indicated by the "collector" column of the table as "Liu" (Liu et al., 2007), "Roden" (Roden and Jin, 2011) and "Batstone" (Batstone, 2001), plus a few other references indicated as "Delattre". Furthermore, the literature source of the collected data is indicated by the "reference" column. Duplicates between the collections were manually checked and deleted. The original yield value and its units as reported in the literature source are present in the table as the "Y" and "Units" columns. The "observed yield" column of the table contains the yields converted to gram of biomass produced per mol of electron donor consumed. The conversion is done assuming the biomass formula is $C_5H_7O_2N$ ($113.11\,\mathrm{g \cdot mol^{-1}}$, $18.8\,\mathrm{kJ \cdot mol^{-1}}$ (Roden and Jin, 2011)) which was obtained experimentally from an activated sludge culture (Hoover and Porges, 1952), and the yields are normalized per C-mol of biomass. In order to have

more easily interpretable variables, the yield in data analysis is converted to $mol_X \cdot mol_D^{-1}$. For every experimental yield record, a catabolic and anabolic reaction has been associated, and the respective Gibbs energy differential have been computed considering standard temperature (298 K) and a set of non-standard concentrations. The set of non-standard concentrations has been determined individually for each metabolism according to a set of rules based on those used by Roden (Roden and Jin, 2011). They can be summed up as follow;

- The activity of H_2O and chemical species in solid state is 1 M
- The concentration of HCO_3^- is 1×10^{-2} M
- The concentration of H^+ is 1×10^{-7} M (pH 7)
- The concentration of O_2 is 2.5×10^{-4} M (which is close to saturation at 298 K)
- The concentration of H_2 is 6.4×10^{-4} M if consumed, or 1×10^{-8} M if produced
- The concentration of any other consumed chemical specie is 1×10^{-2} M
- The concentration of any other produced chemical specie is 1×10^{-3} M

The O_2, H_2 and N_2 chemical species are considered to be in aqueous state in the metabolic reactions (which affects their Gibbs energy differential of formation). The stoichiometry of the catabolic and anabolic reactions, as well as the specific set of concentration and their Gibbs energy differential, has been automatically computed using specifically developed python code available as supplementary material (ref sup mat 2). The catabolism associated to very few (2) data records were endergonic (aerobic metabolism on gluconate, from (Liu et al., 2007)) and were deleted from the table presented in the supplementary material.

Analysis of experimental data

The statistical analysis of the data set was performed with R (R Development Core Team, 2008) using a R-markdown file. The report file of the analysis, from which all the figures of this article are generated, is available as a supplementary material (ref sup mat).

Result

Data distribution

The data consists in 341 experimental measurement of the yield of 101 different metabolisms, from 89 different bibliography references. Those measurements have been done on cultures of approximately 100 different organisms.

The number of records per metabolism has been represented for each acceptor/donor couple in the data (figure 1). For aerobic metabolisms, which use dioxygen as their electron acceptors, yields have been measured for numerous electron donors, and eventually multiple times each. For other inorganic electron acceptors such as nitrate or sulfate, the diversity of tested electron donor is lower by far, and so tends to be the number of observations per individual metabolisms. An exception is the methanogenic metabolism, which is deemed to use bicarbonate as acceptor and dihydrogen as donor. Quantitatively speaking, the maximum number of literature report per acceptor/donor couple is 34 (dihydrogen with bicarbonate) while the median is 2. As multiple different catabolisms can correspond to a single acceptor/donor couple (for example, hydrogenotrophic methanogenesis and homoacetogenesis), the distribution of the reports per type of catabolic reaction is also of interest. Indeed, considering specific catabolisms, the maximum number of observations is 31 (hydrogenotrophic methanogenesis) while the median is 1.

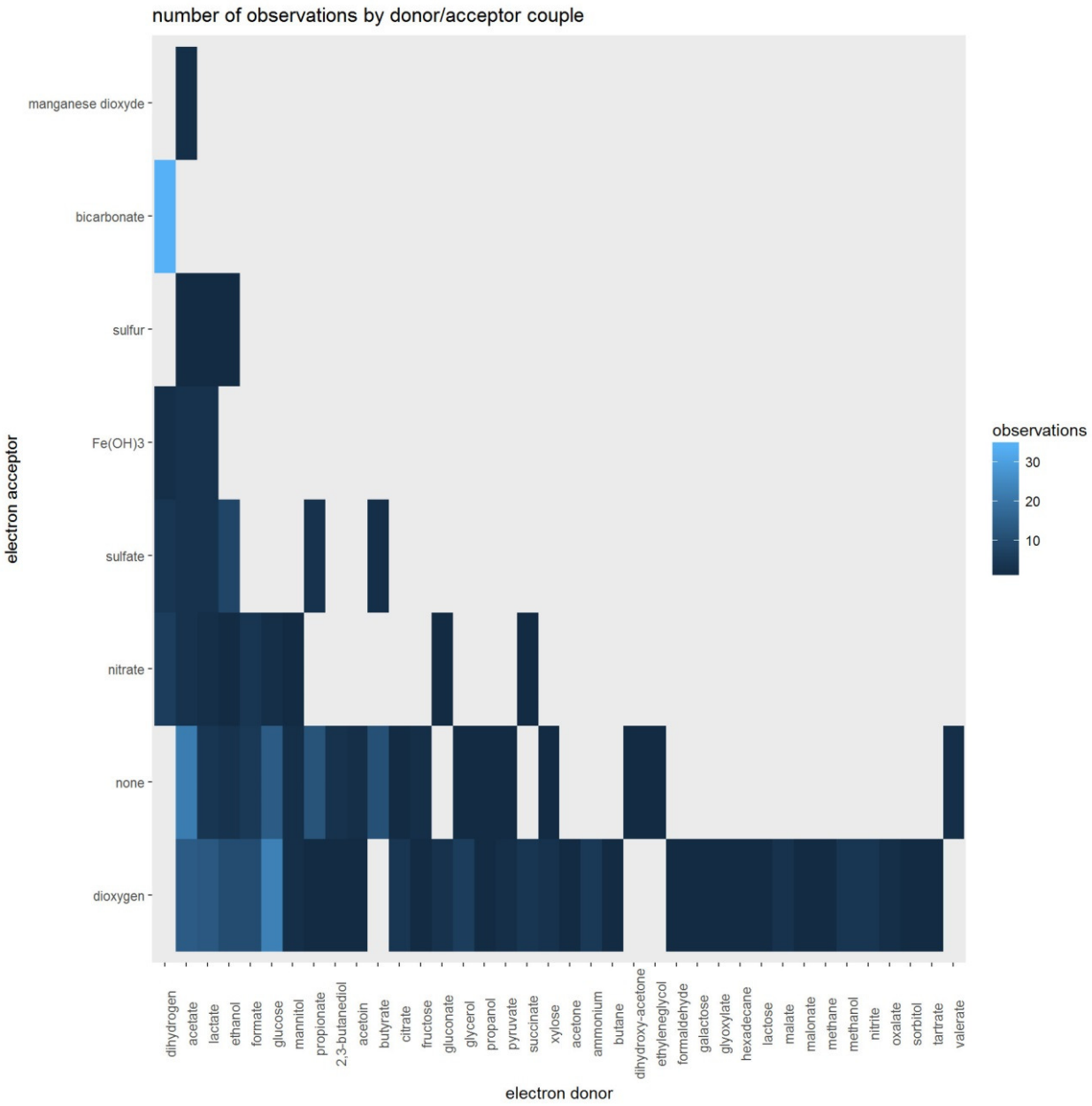


Figure 1: *number of growth yield measurements found in the literature, depending on the electron donor and acceptor. Bicarbonate as electron acceptor corresponds to hydrogenotrophic methanogenesis, while "none" as electron acceptor corresponds to fermentation metabolisms.*

The distribution of the reported growth yields depending on the metabolism is displayed on a box plot (Figure 2). The collected data displays measures of observed growth yields continuously spreading between 0 and approx-

imately 9 $mol_X \cdot molDonor^{-1}$ (80 $gBiomass \cdot molDonor^{-1}$). There are also two outliers; respectively the yield of the metabolism of hexadecane and lactose, whose yield is very high compared to the rest of the dataset. This is no surprise since their number of carbon is far higher than those of the other donors of the dataset (respectively 16 and 12). Interestingly, the variance of the measured yield value of a metabolism seems not to be correlated to the number of observations collected for this metabolism.

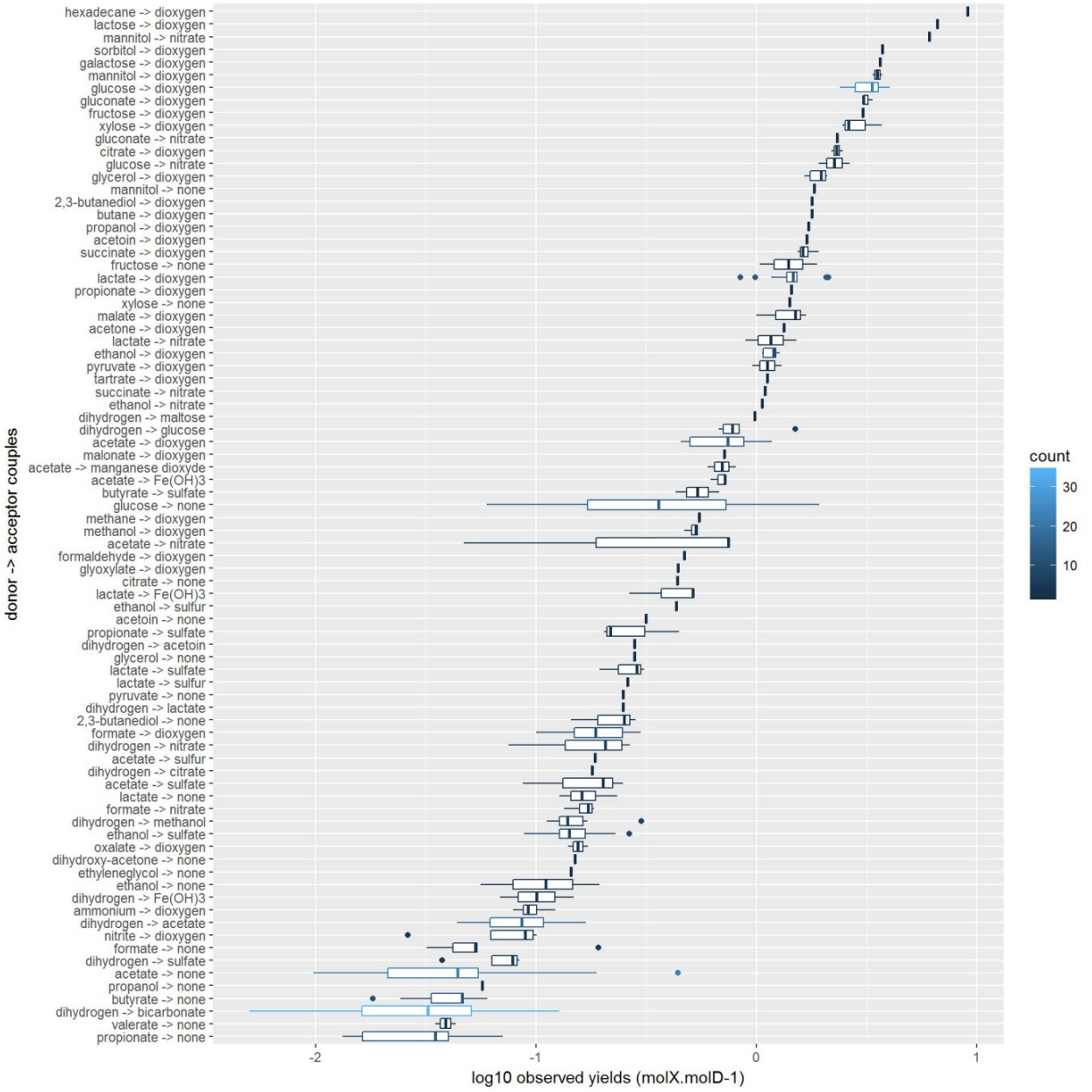


Figure 2: *dispersion of the microbial growth yield per metabolism (log10 scale), sorted by average yield. Color indicates the number of experimental observations collected for each yields*

Assessment of the goodness of fit of previous models on the dataset

Three previously proposed growth yield prediction models are implemented; two expert models by Heijnen (Heijnen et al., 1992), Liu and collaborators (Liu et al., 2007) and one direct yield predictor by Roden and collaborators (Roden and Jin, 2011). The TEEM model developed by McCarty and collaborators (McCarty, 1965) is not included in this comparison because its modern implementation implies to make hypotheses on the metabolisms which would have greatly complexified the comparison.

The goodness of fit of those models is assessed by measuring their ability to predict the maximum observed yield of each metabolisms, so as to provide a reference to the new models that this contribution intends to propose.

Indeed, instead of considering all the yields of the dataset, we make the choice to consider only the maximum yield for each metabolism as the response variable to predict. This reduces the number of observations from 341 to 101. The reason behind this choice is that according to the energy balance model proposed by Heijnen and collaborators (Heijnen et al., 1992), the dissipated energy can be decomposed into two parts. The first part is the energy that is actually dissipated as a driven force for growth, the second part is the energy dissipated in maintenance. Heijnen and collaborators consider that the first part is predicted from thermodynamic constraints, while the second part depends on the growth rate of the population. The maintenance then lessens the observed growth yield and depends on the culture conditions. As the data aggregated in the current dataset contains measurements from cells in various and unknown culture conditions, only the maximum yield is likely to correspond to the part of dissipated energy which can be estimated from physicochemical characteristics of the metabolisms. This is why the maximum yield is considered as the response variable.

The performance of each model is measured as Mean Squared Error ($mean(observation - prediction)^2$). This measurement for all models is shown in the recapitulative MSE table (Table 1).

model	formula	MSE
(Heijnen et al., 1992)	$\Delta G_{dis} = -(200 + 18 \cdot (6 - NoC_{Cs})^{1.8} + exp(((3.8 - \gamma_{Cs})^2)^{0.16} \cdot (3.6 + 0.4 \cdot NoC_{Cs})))$	0.18
(Liu et al., 2007)	$\Delta G_{dis} = \begin{cases} -\frac{666.2}{\gamma_D} - 243.1 & \gamma_D \leq 4.67 \\ -157 \cdot \gamma_D + 339 & \gamma_D > 4.67 \end{cases}$	0.15
(Roden and Jin, 2011)	$Y_{X/S} = 2.08 + 0.0211 \cdot \Delta G_{cat}$	0.41

Table 1: *Performance of multiple previous growth yield prediction models, evaluated by predicting the maximum observed yield of each metabolisms on the current dataset. The performance is expressed as Mean Squared Error.* NoC_{Cs} *is the number of carbon per molecule of carbon source of the anabolism and* γ_{Cs} *is the reduction degree of the carbon source of the anabolism (*$\text{mol}_\text{e} \cdot \text{C} - \text{mol}_\text{Cs}^{-1}$*). The yield* $Y_{X/S}$ *(*$\text{C} - \text{mol}_\text{X} \cdot \text{mol}_\text{D}^{-1}$*) can be computed from* ΔG_{dis} *as* $Y_{X/D} = \frac{1}{\frac{\Delta G_{dis} - \Delta G_{an}}{\Delta G_{cat}} - \nu_D^{an}}$*, with* ΔG_{cat} *and* ΔG_{an} *respectively the catabolic and anabolic energy of growth, in* $\text{kJ} \cdot \text{mol}^{-1}$*,* ν_D^{an} *the stoichiometric coefficient for the electron donor in the anabolic reaction, and assuming that the catabolic equation is normalized per electron donor and the anabolic reaction is normalized per C-mol of biomss*

Selection of the variables to build a new growth yield predictor

In order to build a new growth yield prediction model, two aspects have to be considered; the response variable and the explanatory variable.

The choice of the response variable, the maximum yield per metabolism, has been addressed in the previous section. Regarding the explanatory variables, the following variables are known for each metabolism in the dataset;

- ne: the number of electrons transfered between the electron donor D and the electron acceptor A during the catabolic reaction ($\text{mol}_\text{e} \cdot \text{mol}_\text{D}^{-1}$)
- ΔG_{cat}: the Gibbs energy change of the catabolic reaction ($\text{kJ} \cdot \text{mol}_\text{D}^{-1}$)

- NoC_D: the number of carbon in the electron donor ($C - mol_D \cdot mol_D^{-1}$)
- γ_D: the reduction degree of the electron donor ($mol_e \cdot mol_D^{-1}$)
- NoC_{Cs}: the number of carbon in the carbon source of the anabolism ($C - mol_{Cs} \cdot mol_{Cs}^{-1}$)
- γ_{Cs}: the reduction degree of the carbon source ($mol_e \cdot mol_{Cs}^{-1}$)

Informations about the anabolism cannot be used since modelling the anabolic reaction of a population imply to make many simplifying hypotheses, which would make a yield prediction model less robust.

A Principal Component Analysis (PCA) is performed on the candidate explanatory variables in order to see the correlations between them (3)

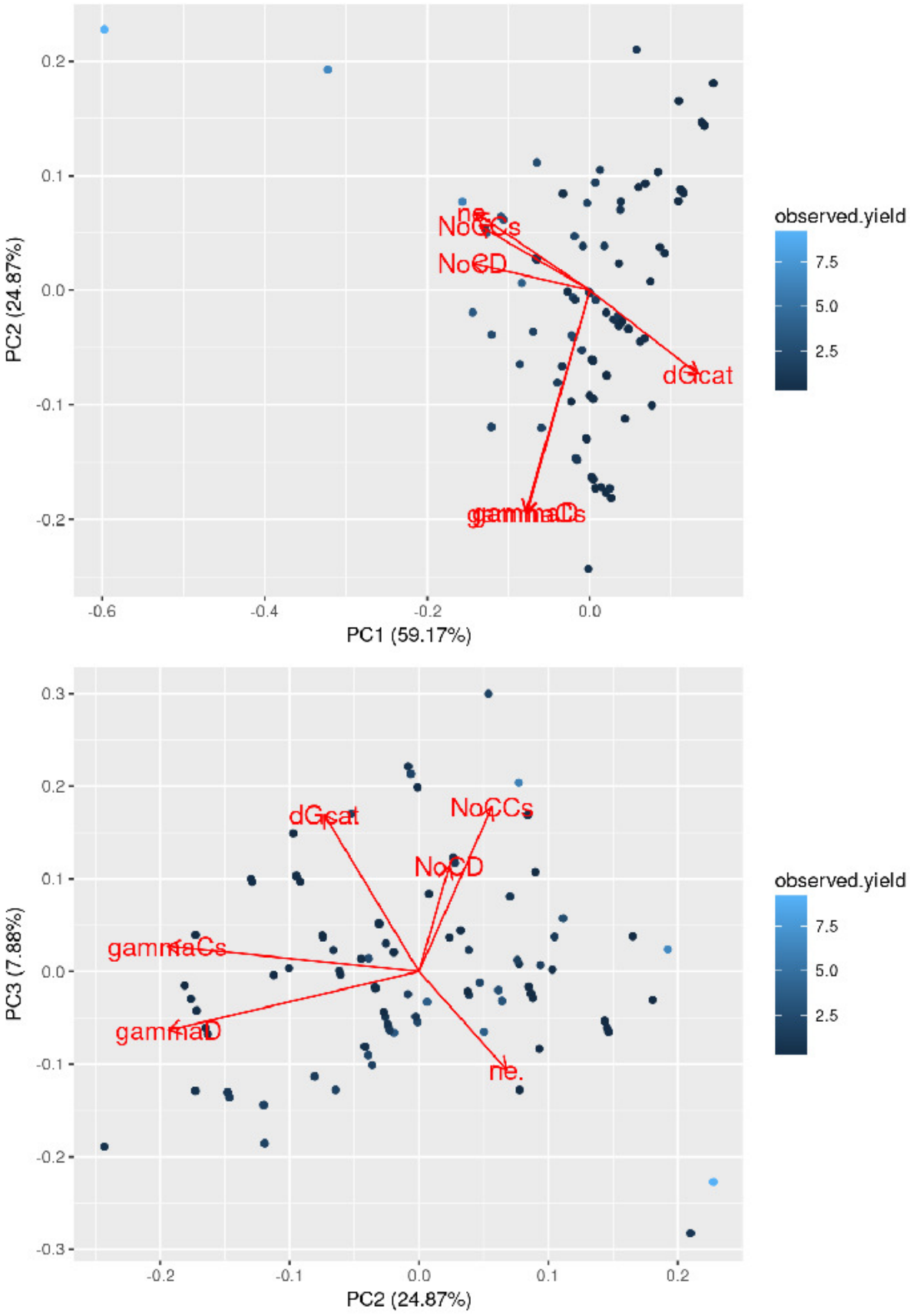


Figure 3: *PCA plot of the candidate explanatory variables* ΔG_{cat}, $ne.$, γ_D, $gamma_{Cs}$, NoC_D *and* NoC_{Cs}. *The upper panel shows PCA axis one (59% of the variance) vs PCA axis two (25% of the variance), and the lower*

It can be observed from this PCA plot that the candidate explanatory variables are highly correlated in the current dataset. Indeed, ΔG_{cat} and ne. are colinear on the three first components. This means that both variables explains the same part of the variance of the explanatory variables. Consequently it is not useful to include both of them in a regression model.

Measuring the correlation coefficient between each candidate explanatory variable and the maximum yields shows that numbers of carbon NoC_D and NoC_{Cs} are sensibly more correlated to maximum yields than the reduction degrees γ_D and γ_{Cs}, as seen in the following table;

variable	**coefficient of correlation to the maximum growth yield**
ΔG_{cat}	-0.91
ne	0.84
γ_D	0.19
NoC_D	0.76
γ_{Cs}	0.20
NoC_{Cs}	0.78

The variables from which a new regression model shall be built are then ΔG_{cat}, NoC_D, NoC_{Cs}. While Liu and collaborators ditched the NoC_{Cs} used by Heijnen and collaborators in their own predictor (Liu et al., 2007), the numbers of carbon seems to be more predictive the reduction degrees in the light of the current analysis. Alternative models, using different combinations of NoC_D, NoC_{Cs}, γ_D and γ_{Cs} have been tested and indeed exhibit high MSE (cf supplementary material).

Definition of new regression models

Generic predictors of maximum growth yield

A new generic regression model (lets call it "model A") is calibrated, considering the variables ΔG_{cat}, NoC_D, NoC_{Cs} and their interactions up to order two (one variable multiplied by another). The MSE of this model is 0.15, which means its goodness of fit on the current dataset are equivalent to this of Liu's model (cf table 1). Figure 4 shows a Bland-Altman plot of model A.

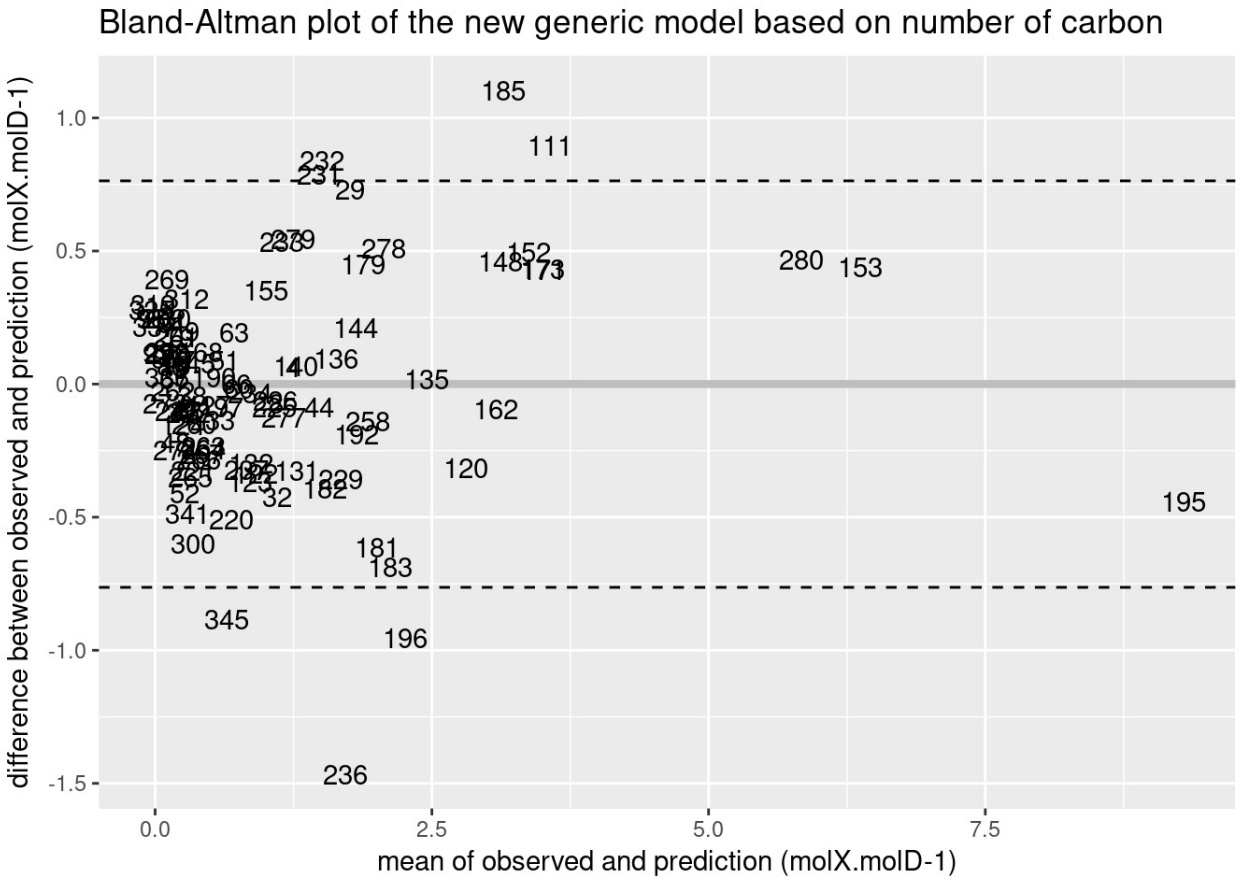


Figure 4: *Bland-Altman plot of the new maximum yield prediction model A. Numbers displayed on the plot are the index of the experimental observations collected in the spreadsheet given as supplementary material. This predictor relies on the Gibbs energy change of the catabolism and the number of carbons of the carbon source and the electron donor of the metabolism*

As this model takes variables interactions into account, its formulation is rather lengthy (it can be seen in the supplementary material), which defeats the purpose of practicality behind the direct prediction of the maximum yield.

The model A is then simplified by keeping only the variables of its most significant terms, that is ΔG_{cat}, $\Delta G_{cat} \cdot NoC_D$ and $\Delta G_{cat} \cdot NoC_{Cs}$.

The MSE of the resulting model (lets call it "model B") is 0.23, which means its goodness of fit on the current dataset is not as good as those of the expert models, but better than this of Roden's model. Figure 5 shows a

Bland-Altman plot of model B.

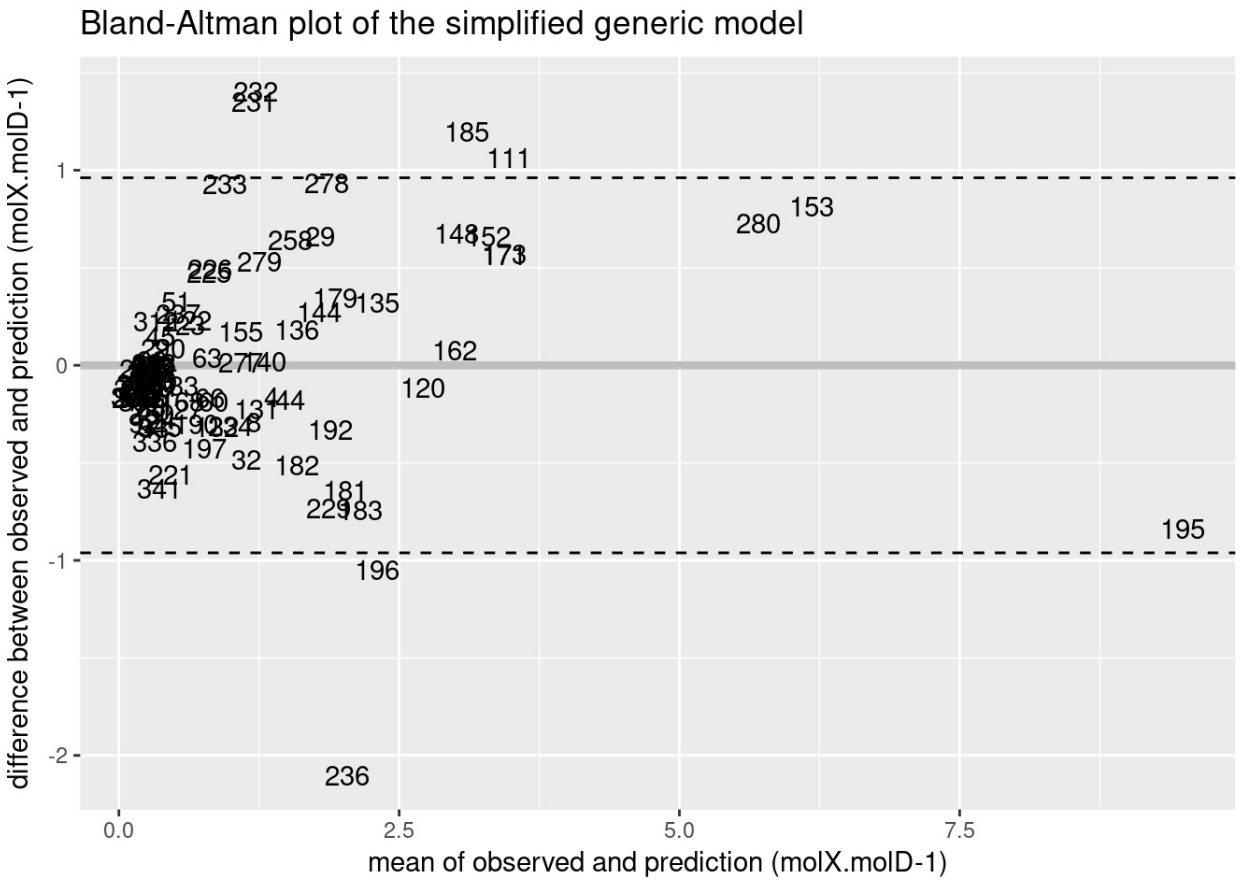


Figure 5: *Bland-Altman plot of the new maximum yield prediction model B. Numbers displayed on the plot are the index of the experimental observations collected in the spreadsheet given as supplementary material. This predictor consists in a simplified version of model A*

Partitioned predictors of maximum growth yields

According to the results of the previous subsection, it seems it is not possible to calibrate a yield predictor better than those already existing in the literature using the selected variables. Liu's model partition the metabolisms into two categories, depending on the value of γ_D, and calibrate a specific model for each partition. Partitioning the dataset is a way to develop a better regression model to predict maximum growth yields. However, the number of carbons NoC_{Cs} and NoC_D were observed in the previous subsection to be better predictors of the maximum growth yields than the reduction degrees γ_D and γ_{Cs} for the overall dataset.

number of observations	**fermentation**	**respiration**	**total**
$\gamma_D \leq 4.67$	27	47	74
$\gamma_D > 4.67$	11	16	27
total	38	63	101

Table 2: *Number of metabolisms recorded in the dataset depending on the fermentation vs respiration and* $\gamma_D \leq 4.67$ *vs* $\gamma_D > 4.67$ *partitions*

On the other hand, partitioning metabolisms between respirations and fermentations seems a sensible choice *a prior* for two reasons;

- Respiration and fermentation are two fundamentally different metabolic mechanisms. By definition, respirations are metabolisms upkeeping a membrane gradient as a mean to store energy, while fermentations do not.
- Respirating and fermentating microbes are usually studied by different communities of microbiologists. Calibrating specific maximum yield prediction models for both partitions would then increase the practicality of the outcome of this contribution

Partitioning the metabolisms using the two proposed partitions (fermentation/respiration and $\gamma_D \leq 4.67$ vs $\gamma_D > 4.67$) at the same time would lead to partitions too small to consider the regression model of some partitions as robust (cf table 2). Consequently only one partition will be used.

In order to choose the right partition, and the right explanatory variables, the correlation between all explanatory variables and all partitions is measured in the current dataset (Table 3). Three important observations can be drawn from these measurements.

Firstly, the reduction degrees γ_D and γ_{Cs} are poorly correlated to the maximum yields of all partitions.

Secondly, ne, NoC_{Cs} and NoC_D are highly correlated to the maximum yields of respiration, but poorly correlated to those of fermentation.

Thirdly, none of the explanatory variables is highly correlated to the maximum yields of fermentation metabolisms whose γ_D is below 4.67. This set of 27 observations designates fermentations on acetate, citrate, dihydrogen,

variable	metabolism	$\gamma_D \leq 4.67$?	correlation
ΔG_{cat}	fermentation	False	-0.94
ΔG_{cat}	fermentation	True	-0.37
ΔG_{cat}	respiration	False	-0.97
ΔG_{cat}	respiration	True	-0.92
ne	fermentation	False	-0.12
ne	fermentation	True	-0.28
ne	respiration	False	-0.98
ne	respiration	True	-0.93
γ_{Cs}	fermentation	False	-0.31
γ_{Cs}	fermentation	True	0.13
γ_{Cs}	respiration	False	-0.07
γ_{Cs}	respiration	True	-0.46
γ_D	fermentation	False	-0.31
γ_D	fermentation	True	0.26
γ_D	respiration	False	-0.15
γ_D	respiration	True	0.43
NoC_{Cs}	fermentation	False	0.19
NoC_{Cs}	fermentation	True	0.64
NoC_{Cs}	respiration	False	0.97
NoC_{Cs}	respiration	True	0.94
NoC_D	fermentation	False	0.19
NoC_D	fermentation	True	0.45
NoC_D	respiration	False	0.97
NoC_D	respiration	True	0.88

Table 3: *Pearson correlation coefficient measured between each explanatory variable and the maximum growth yields of the dataset, depending on the fermentation vs respiration and $\gamma_D \leq 4.67$ vs $\gamma_D > 4.67$ partitions*

dihydroxy-acetone, formate, fructose, glucose, glycerol, lactate, mannitol, propionate, pyruvate and xylose. Whatever, NoC_D and NoC_{Cs} still are the variable being the highly correlated to their maximum yields.

The choice is then made to partition the model between respiration and fermentation, because the maximum yields exhibit better correlations to the explanatory variables this way, because the two partitions are more equitably populated (cf table 2) so the individual regression models will be more robust, and because of the *a priori* reasons mentioned earlier. Moreover, it seems that ΔG_{cat}, NoC_{Cs} a,d NoC_D are the best variable choice for both respiration and fermentation.

Specific models are then calibrated for respiration and fermentation maximum growth yields. Lets call the resulting conditional model "model C". Its formula is

$$Y_{X/S}^{max} = \begin{cases} 8.39e-2 - \Delta G_{cat} \cdot (1.14e-4 + 1.91e-4 \cdot NoC_D + 2.73e-4 \cdot NoC_{Cs}) & if fermentation \\ 2.96e-2 - \Delta G_{cat} \cdot (1.15e-3 - 3.36e-4 \cdot NoC_D + 3.24e-4 \cdot NoC_{Cs}) & if respiration \end{cases} \quad (1)$$

With $Y_{X/S}^{max}$ the maximum yield of the metabolism ($C - mol_X \cdot mol_D^{-1}$), ΔG_{cat} the Gibbs energy change of catabolism ($kJ \cdot mol_D^{-1}$), NoC_{Cs} the number of carbon per carbon source molecule and NoC_D the number of carbon per electron donor molecule.

A Bland-Altman of the model C is shown as Figure 6. The MSE of this model is 0.14 on the current dataset, which means it is slightly better than this of the previous approaches.

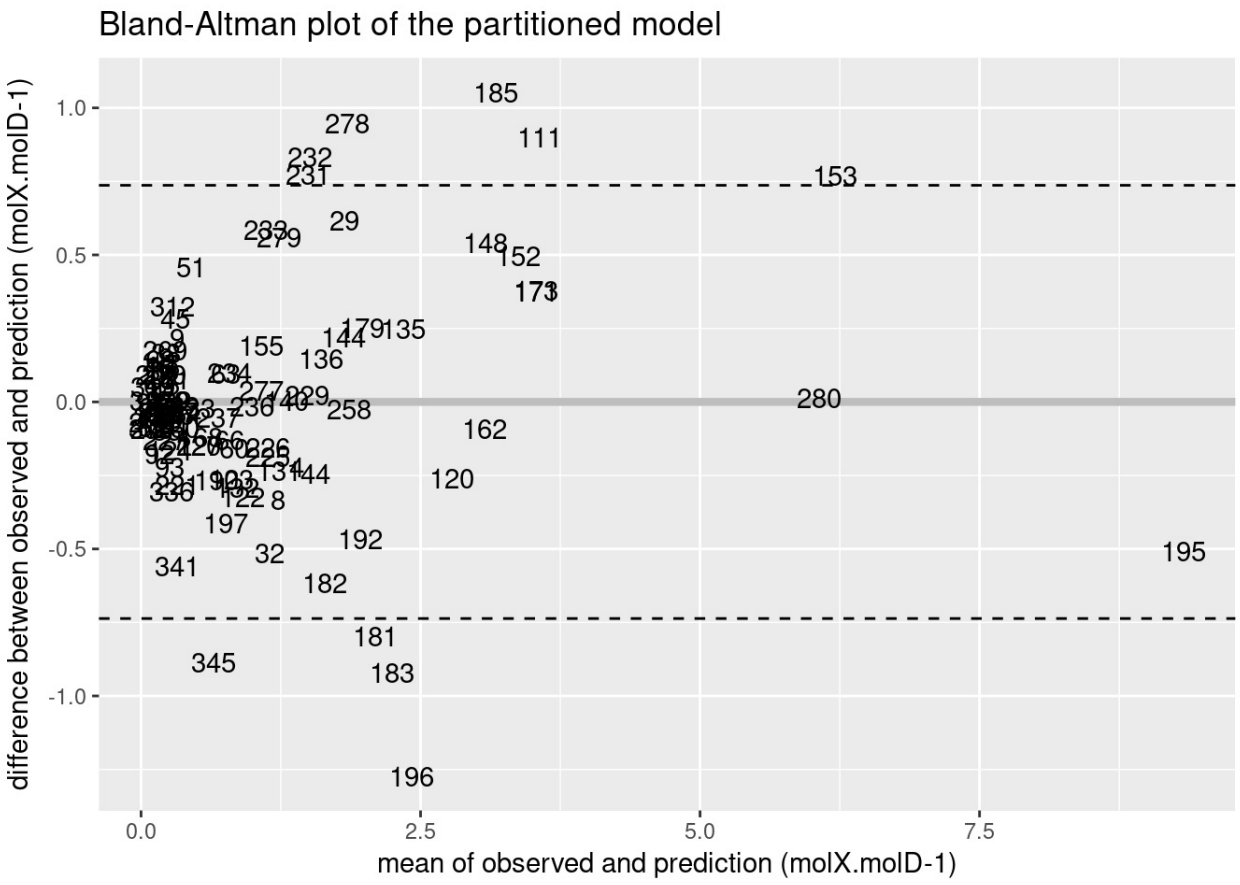


Figure 6: *Bland-Altman plot of the new maximum yield prediction model C. Numbers displayed on the plot are the index of the experimental observations collected in the spreadsheet given as supplementary material. This predictor consists in a conditional regression model with different coefficients depending on whether the metabolism is a respiration or a fermentation*

Discussion

Biases in the current knowledge about metabolisms' yields

Since the dataset collected for the need of this contribution is the largest of its kind, it can be thought of as an overview of the state of experimental knowledge regarding microbial growth yield. The display of its distribution (subsection) revealed two important biases in the reported data. One bias affect the response variable (the maximum growth yield) and the other affects the explanatory variables.

The first bias is that the dataset contains only one yield measurement for most of the metabolisms, while some other metabolisms are documented with up to 31 yield reports. In the numerous cases where only one yield was reported for a metabolism, this yield measurement was assumed to be the maximum yield of the metabolism. This leads to the underestimation of the maximum growth yield in most cases. Hopefully, this bias is not sensible on the Bland-Altman plots of the models proposed in this contribution (no visible bias on the residuals distribution along the vertical axis of the plots).

The second bias in the dataset is that some metabolisms are well studied (for example, aerobic metabolisms and methanogenesis) while others are sparsely documented. Indeed, if an empirical relationship linking physicochemical variables to maximum growth yield is to be calibrated from experimental data, physicochemical variable should ideally be uniformly distributed in experimental data. However, this is not the case in the data collected from the literature.

Those biases may be detrimental to any attempt to build a empirical growth yield predictor from literature data. A possible reason for those biases are that the measurement of the yield of some metabolisms requires more complex experimental settings than others. Another reason is that experiments on specific metabolisms are often engaged because of industrial interest. Bridging the gaps of experimental data would then likely need dedicated experimental efforts. To summarise, this study revealed biases in microbial growth yield reports in the literature which may lessen the robustness of any growth yield predictor calibrated on physicochemical variables. Hopefully, this claim can be mitigated by the fact that the existing data reveals strong correlations between the maximum yield of most metabolisms and some physicochemical variables (as observed in table 3).

Explanations to the poor performances of Roden's model on the current dataset

As the performance of previously published yield prediction models has been tested on the current dataset, the simple regression model proposed by Roden and collaborators performs rather poorly when compared to the two expert models (Table 1). The data collected by Roden and collaborators to calibrate their model is included as an important part of the current dataset (123 observations over 341), so this lack of fit is surprizing at first sight. However, two possible explanations can be raised. Firstly, Roden's model was calibrated in order to predict all yields, and not only maximum yields, so it may put a first bias on the evaluation of its goodness of fit on the current

dataset. In addition, this model is based on the Gibbs energy change of catabolism ΔG_{cat}. This quantity was adjusted to non-standard concentrations by Roden and collaborators. The authors gave some rules they used to decide which concentrations to assume for each yield observation; those rules were applied to the current dataset however the Gibbs energies are sometimes different from those computed by Roden (the difference is displayed on the "catabolisms" sheet of the data's spreadsheet, in supplementary material). While Roden and collaborators only gave a set of rules and the final ΔG_{cat} they associate to each metabolisms, the ΔG_{cat} calculations for the current dataset have been performed in an explicit and reproducible way (cf python scripts provided as supplementary material). Those two reasons are then possible explanations as to why the goodness of fit of Roden's model is less good than this of the expert models by Heijnen and Liu.

Delineation of a subset of poorly predicted maximum yields

The measurement of the correlation coefficient between all explanatory variables and all subparts of the dataset (considering respiration/fermentation and $\gamma_D \leq 6.67/\gamma_D > 4.67$ as the two partitions) in table 3, reveals that one precise subpart of the dataset is notably less correlated to the explanatory variables than the others. This subset is fermentation the metabolisms whose electron donor's reduction degree is below $4.67\,\mathrm{mol_e \cdot C - mol_D^{-1}}$. This is the only segment of the dataset for which some explanatory variable seems to be lacking. Since many of the carbonaceous substrates in this subset can give multiple different degradation products, it is possible that the accumulation of some products in the culture medium may influence the yield of the reaction. Moreover, other metabolisms are likely to be catalyzed in parallel (degradation of the same carbonaceous substrate into a different set of product, or degradation of the products), and this possibly interfere with the estimation of the maximum yield of a precise metabolic pathway.

Another explanation is that it is sometimes hard to determine the "electron donor" and the "electron acceptor" in fermentation metabolisms, as different parts of a single molecule can act as donor and acceptor. For example in acetoclastic methanogenesis, the "COO^-" part of acetate (CH_3COO^-) gives an electron to the CH_3 part. Both parts respectively become CO_2 and CH_4, thus this reaction consists in an electron transfer inside a single molecule. It is possible that a new convention to determine variables such as γ_D or NoC_D in such conditions is needed to draw better correlations between them and the growth yield. Those explanations, through being mere

suggestions, are the most likely to the authors of this contribution at this stage.

Conclusion

The accurate prediction of microbial yields is of paramount importance for the modelling of microbial communities, both for mixed culture bioprocesses and environmental studies.

Two kinds of approaches have been developed to address this issue. The first consists in expert models relying on a semi-empirical description of the energy balance of microbial growth. The second consists in a more purely statistical approach, proposing a direct expression of the growth yield as a function of physicochemical parameters. The current contribution chose the practicality of the second approach. Nevertheless, it also pays heed to theoretical notions introduced by the expert models in order to provide a growth yield predictor as robust as possible.

First and foremost, the main result of the contribution is its dataset. It is indeed the largest collection of experimental yield measurements from the literature up to now, and it is provided as a supplementary material along with this article, in order to make further statistical investigations possible. The second result of this contribution is a new maximum growth yield predictor (equation 1), which is simpler to use and slightly more accurate than the previously proposed expert models.

Last but not least, this study allows to point at the limits of growth yield prediction by calibration of an empirical model from experimental data. Particularly, it reveal a subset of metabolisms (fermentation metabolisms whose electron donor's reduction degree is below $4.67\,\mathrm{mol_e \cdot C - mol_D^{-1}}$) whose maximum yield is less correlated to the explanatory physicochemical variables than the other metabolisms. Further investigations are needed in order to find more relevant variables to predict those metabolisms' yields.

Chapter 7:

Discussion and conclusion

7.1 Discussion

7.1.1 Prediction of the "redox tower" by the MTS model

One of the most important results of the study of the MTS model is the demonstration of its ability to predict a growth pattern known as "redox tower" (cf chapter 4 for demonstration by simulation).

The redox tower is the accepted name given to an ecological succession reflecting the thermodynamics potential of the substrates in the culture medium. A "redox potential" can be associated to all catabolic reaction, and thus ordered on this single axis. The redox tower is then said to occur when the ecological succession observed in a microbial community reflects this deterministic axis. A *sine qua non* condition for a redox tower to occur is that the affected guilds must be in competition for a resource. Indeed, the ecological succession happens because the guild catalyzing the most exergonic catabolism prevents the implantation of the other guilds through competitive exclusion (they have a better yield so they develop a more important population so they consume more substrate and finally take a kind of monopoly on the resource).

While the concept of redox potential was not used in this memoir, it is a transformation of the Gibbs energy change of reaction; this phenomenon can then also be formulated in terms of Gibbs energy.

For example, let us consider the aquifer simulation of chapter 4 (third part of the results). Here, the culture medium contains acetate as only electron donor, and oxygen, nitrate, sulfate and iron as electron acceptors. In terms of Gibbs energy change of catabolism; oxygen comes first, then nitrate, then sulfate, then iron. The succession observed in the simulations complies to this ordering. Indeed, the simulated culture medium has the dynamics of a chemostat, so a constant amount of acetate (i.e. a constant amount of electrons) is inputed into the culture medium. Multiple simulations have been performed to test the effect of different concentrations of acetate on the communities structure (once stabilization is reached). The concentration of electron acceptors remained constant. It appears that the electrons available in the culture medium as acetate are transfered to oxygen first, so the aerobic guild is the only one to stabilize in the chemostat at low acetate concentrations. Once the electron well that oxygen represents is saturated (all oxygen is reduced) the remaining acetate reacts with nitrate. As acetate starts to be available for the denitrifier guild, this guild is then able to remain in the chemostat at steady state, along with the aerobic guild. Again, once all nitrate is reduced, the remaining acetate is consumed

by sulfate-reducers etc.

The redox tower is also experimentally shown to occur in environmental communities. It gives rise to a structuration of the microbial community. Indeed it is observed in ecosystems such as aquifers (Chapelle and Lovley, 1992) and lake hypolimnia (Müller et al., 2012). Moreover, the spatial structuration induced in the classical Winogradsky column experiment (Zavarzin, 2006) is notably explained by the occurence of a redox tower.

Since the redox tower has been known for a long time, and since other microbial thermodynamics models have been demonstrated to be able to simulate this phenomenon (González-cabaleiro et al., 2015; Bethke et al., 2011), some could argue that its prediction by the MTS model is trivial and brings no novelty. However this is not true since the MTS model makes this prediction without the need to calibrate parameters or to modulate empirical kinetic equations with additional thermodynamic factors. This model implements, by construction, a mechanism giving rise to the redox tower. On the other hand, the two other models cited earlier do not have the same approach.

The approach behind the model designed by Jin and Bethke (reviewed in the bibliography of this memoir, chapter 2) can be said to be close to this of the MTS model. Indeed, Jin and Bethke derive their growth rate function from theoretical considerations applied to microbial catabolism's electron transfer chain (Jin and Bethke, 2002). The approach of Jin and Bethke and this of the MTS model are both based on theory (two totally different theories) and both of them provide a new growth rate expression which allows to simulate redox tower phenomena. Nevertheless, the MTS model has the advantage to be able to simulate growth limitation by nutrient, that which is not included in Jin and Bethke's theory. More details about this feature are provided in the next section (subsection 7.1.2.2).

Regarding the model used by González-Cabaleiro and collaborators, it is based on the one defined by Heijnen and collaborators ((Heijnen and Kleerebezem, 1999)), itself based on an empirical model (the Herbert-Pirt model, described in chapter 2). The growth dynamics of this model relies on Monod's law; such dynamics does not induce redox towers by construction. The approach behind this model is to introduce energy constraint by calibration of the dissipated energy and of the maintenance energy cost. On the other hand, the MTS model relies on the same calibration-based method to estimate dissipated energy, but it includes it into a novel growth kinetics formula which comes from a theory explaining the redox tower.

In other words, the important difference between predicting a redox tower pattern using Monod's kinetics and using MTS's kinetics lays in the conclusions that can be drawn from the simulation results. Predicting a redox tower using Monod's law calibrated according to thermodynamic variables evidences that the used thermodynamic variables are sufficient to explain the redox tower pattern. However, any other curve whose shape is similar to this of Monod's law could have been used, so this result does not evidence anything regarding Monod's law in itself. Contrarily, reproducing a redox tower pattern using the MTS model not only evidence that thermodynamic variables are sufficient to explain the phenomenon, but this result also provides support for the theoretical explanation proposed by the MTS model on how these thermodynamic variables influences growth dynamics.

Consequently, the MTS model provides a more parcimonious explanation to the redox tower phenomenon than González-Cabaleiro's model, since it does not need to make the hypothesis of an external kinetic model as its growth kinetics ensues from its theory. While it is certain that biological systems cannot be reduced to simple hypotheses, the theory behind the MTS model constitutes an attempt to understand the basal drivers on top of which the complexity of biology adds up. On the other hand, González-Cabaleiro's model do not offers this possibility.

7.1.2 Consequences of the growth rate function of the model

The growth rate function of the MTS model arises from simple hypotheses about microbial growth at microscopic scale. These hypotheses can be summarized as;

- a microbial cell needs to overcome a fixed energy barrier in order to divide
- this energy barrier can be decomposed into the anabolic energy ΔG_{an} and the dissipated energy ΔG_{dis}
- the energy to overcome the energy barrier is the catabolic energy ΔG_{cat}, which is obtained by catalyzing substrate oxidation
- substrate exists in the form of particles uniformly distributed around the cells
- if a fictional, fixed volume V_H ("harvest volume") around the cell contains enough substrate to overcome the energy barrier, the cell is said to be in "activated" state
- only an activated cell is able to divide

Considering those hypothesis, statistical physics allows to formulate the proportion of the $N^{\ddagger}$ activated cells over the N total number of cells;

$$\frac{N^{\ddagger}}{N} = \prod_{i=0} exp(\frac{Y_{S_i}^{met}}{V_H \cdot [S_i]}) \tag{7.1}$$

with S_i the i^{th} consumed chemical specie, $Y_{S_i}^{met}$ its stoichiometric coefficient ($\text{mol}_{\text{S}_\text{i}} \cdot \text{mol}_{\text{X}}^{-1}$, negative), and $[S_i]$ its concentration (in $\text{mol} \cdot \text{volume}^{-1}$). V_H, the harvest volume, is then expressed in $\text{volume} \cdot \text{mol}_{\text{X}}^{-1}$.

The value of $Y_{S_i}^{met}$ is dynamically computed as;

$$Y_{S_i}^{met} = Y_{S_i}^{an} + Y_{S_i}^{cat} \cdot \frac{-\Delta G_{an} + \Delta G_{dis}}{\Delta G_{cat}} \tag{7.2}$$

So the stoichiometric coefficients of metabolism are adjusted such that catabolism is run enough times to allow to overcome the energy barrier.

The ratio $\frac{N^{\ddagger}}{N}$ of "activated" cells can translate into an expression of the growth rate μ of a population;

$$\mu = \mu_{max} \cdot \prod_{i=0} exp(\frac{Y_{S_i}^{met}}{V_H \cdot [S_i]}) \tag{7.3}$$

with μ_{max} the maximum growth rate (time^{-1}).

This growth rate function is specific to the MTS model as it is a direct consequence of its hypotheses on microbial growth. The specificities of this function, and their consequences in microbial populations' dynamics modelling, have been highlighted during this thesis. They are discussed in this section.

7.1.2.1 MTS's growth function is natively multi-substrate and provides a possible explanation for Liebig's law

The MTS model expresses the growth rate of a population as a function of the concentrations of all the chemical species required by the growth reaction. Comparatively, most of the other microbial growth models are based on empirical growth laws (such as Monod's), which assume the growth rate depends on the concentration of a single limiting substrate. Many different propositions have been made to modify empirical growth rate functions so they can take multiple substrates into account. However, none of those propositions has reached consensus (Bungay, 1994).

Indeed, Monod's growth law (and others such as Contois') implicitely assumes Liebig's law of the minimum. This empirical law, which has initially been formulated for the growth of plants, states that the growth of organisms is limited only by the scarcest resource at a given time. This principle is widely used in ecology and has been applied at the scale of organisms, populations and communities.

As an example, this law has been implicitely assumed by Gonzalez-Cabaleiro and collaborators when simulating the dynamics of a microbial community (González-cabaleiro et al., 2015) using the energy-based model defined by Heijnen and collaborators (Heijnen and Kleerebezem, 1999) (presented in the bibliography review of this memoir). The dynamic model described by Heijnen and collaborators do not provide a novel expression for the relationship between substate concentration and substrate consumption rate. Instead, they consider it follows Monod's law. This relationship then accounts for the limitation of growth by a single substrate. However this is not satisfying when attempting to describe the dynamics of a community, where populations can possibly be limited by multiple substrates. The solution considered by Gonzalez-Cabaleiro and collaborators was then to compute the value the growth rate would take for each substrate separately, and then to keep the minimum among those values. This is then an example of how Liebig's law is sometimes implicitely assumed in a microbial growth model.

However, while Liebig's law is often assumed while modelling populations' and communities' dynamics (sometimes implicitely as show by the previous example), it does not always reflects the observed behavior of microbial populations and communities (Danger et al., 2008).

The growth rate function of the MTS model provides an interesting contribution on this issue. Indeed, Liebig's law is not enforced as an hypothesis in the MTS model. According to it's growth function, it is a priori possible that a population can be limited by two substrate at a given time. Indeed, supplementary material 9.4 illustrates that it is possible for a population simulated by the MTS model to be limited by both its electron donor and acceptor at the same time. As shown in the supplementary material, this situation depends on the stoichiometric demand and offer for both chemical species. Since the MTS model adjusts stoichiometry (and therefore the "demand" for a given chemical specie) depending on energy balance, the concentration of a resource is then not the only factor defining its "scarcity". Such "colimitation" situation cannot be simulated by a model enforcing Liebig's law. Conservatively, other populations simulated using the MTS model have been observed to comply to Liebig's law (cf chapter 4).

Liebig's law is not enforced as an hypothesis in the MTS model. According to it's growth function, it is a priori possible that a population can be limited by two substrate at a given time. However, in practice, it has been observed that the populations simulated using the MTS model complied to Liebig's law in some situations (cf chapter 4). Indeed, the growth rate of a population simulated using the MTS model is multiplied by a specific factor (between 0 and 1) for each of its substrates, called "tuning factor", hence a possible limitation by multiple substrates at once. However it has been observed in multiple situations that for a simulated population, one tuning factor was significantly lower than all the others, to the point that all tuning factors could have been neglected except the one of the most limiting substrate. This situation makes the population apparently comply to Liebig's law. It happens because the tuning factors are expressed as exponentials of ratios of a stoichiometric coefficient over a substrate concentration (which can be pictured as a "demand over offer" ratio). Therefore, tuning factors are very sensitive to the changes in substrate concentrations. To illustrate this statement, supplementary material 9.3 (section 9.3) shows the effect of the same concentration change on $\frac{[S]}{K_S+[S]}$ and $exp^{\frac{Y_S^{met}}{V_H \cdot [S]}}$. The sensibility of the tuning factors of the MTS model induce "toppling" behaviors; the influence of a substrate on growth can switch from "negligible" to "limiting" over a very narrow range of concentration (as observed in chapter 4). When this phenomenon occurs while simulating growth kinetics, it makes the simulated populations apparently comply to Liebig's principle as the limiting resource quickly "switches". However, as can be deduced from equation 7.3, a low value of V_H can alleviate the steepness of the switch between two limiting substrates.

As a conclusion, the MTS model then provides a possible, theoretical explanation to the apparent compliance of some experimental growth observations to Liebig's law. At the same time, the MTS model is also able to simulate situations into which multiple substrates exerts a significant limitation on growth at the same time. It is then intrinsically more flexible than a model enforcing Liebig's law as an hypothese.

7.1.2.2 MTS's growth function accounts for growth limitation by both energy and nutrients

According to the MTS model's energy balance, the growth reaction is a combination of the catabolic reaction and the anabolic reaction. Consequently a chemical specie is consummed for growth because it is required either by the catabolism for energy-generation purpose or by anabolism for biomass synthesis, or both. Chemical species required by anabolism only are called "nutrients". The MTS model translates these two different needs into stoichiometric coefficients, which in turn become tuning factors (cf previous subsection). As an illustration, the

growth limitation exerted by an extended set of nutrients in a M9 medium according to the MTS model has been computed (cf supplementary material 9.1). In a nutshell, the MTS model's growth rate function formulates the limitation of growth by energy source and nutrients into a unified framework.

This feature is an advantage when compared to other models, particularly the one proposed by Jin and Bethke (introduced in the bibliography review, chapter 2) (Jin and Bethke, 2002). This model derives an expression of the rate of the catabolic reaction based on nonlinear nonequilibrium thermodynamics. The authors state that the growth rate of a population simulated using their model can be expressed as a function of the rate of the catabolism in cases when the growth is "energy-limited". An energy-limited growth means that the consummed chemical specie being the most limiting for growth is required by the catabolism (likely an electron acceptor or donor). Consequently, the theory behind Jin and Bethke's model cannot represent the limitation exerted by a nutrient on the growth of a microbial population. On the other hand, the MTS model is able to model this limitation indistinctly from the energy limitations.

7.1.2.3 MTS's functional response is more compliant to some experimental observations than Monod's

The relationship between the growth rate and the concentration of a substrate has a different "shape" depending on whether it is simulated by the MTS model or a Monod law. Such shape, as visible on the graphs of supplementary material 9.3 (section 9.3), is called the "functional response" of the model.

The notion of functional response was proposed in ecology in the 1950s; the idea was to study the consequences of the "shape" of a growth rate function regarding population dynamics, independently from its formula (Denny, 2014). For instance, the shape of the functional response of the MTS model corresponds to a "type III" functional response, while Monod's law induce a "type II" functional response.

The type III functional response is characterized by a sigmoid curve. In practice, it means that growth is unsignificant if substrate concentration is too low, until a threshold susbtrate concentration is reached. It then increase exponentially and stabilizes to a plateau (defined by the model's maximum growth rate μ_{max}). The fact that the growth rate does not increase with substrate concentration until a threshold value is reached has been observed in some experiments (Kovarova-Kovar and Egli, 1998).

A consequence of the MTS model inducing a type III response was commented in the results from chapter 4. In one of the simulations of this publication, a simplified activated sludge microbial community has been simulated.

In particular, a nitrite-oxidizing autotroph population ("NOB") is in competition with others. At the beginning of the simulation, there is no nitrite in the culture medium, as it is being produced by another population. In a first time, the NOB population does not grow as the concentration of nitrite, its electron-donor, is too low. Once the nitrite concentration accumulates enough to pass a certain threshold value, the NOB population starts to grow and to consume nitrite. As a consequence the nitrite concentration in the culture medium is predicted to rise then fall. As shown in the article (chapter 4), this "nitrite peak" is not predicted using Monod law as its type II functional response implies that populations starts to grow once the first molecule of substrate enters the culture medium. While this prediction is not in adequation with what is usually observed in the modeled system, experimental observations of such nitrite peaks have been reported (Rajagopal et al., 2011).

Some models using Monod's law perform transformations on the equations making the functional response of the model closer to this of a type III functional response. For example, the microbial population dynamics model described by Heijnen and collaborators (Heijnen and Kleerebezem, 1999) is an energy-based growth model which consider that if the substrate is not concentrated enough to meet the maintenance's energy requirement, the growth rate becomes negative. This consideration induces the existence of a minimal substrate concentration allowing growth, hence a functional response resembling a type III. This model is described in more details in the bibliography chapter of this memoir (chapter 2).

The MTS model also has a type III functional response, however it derives it from totally different considerations (summarized earlier in this chapter in section 7.1.2). It can then be considered as an alternative explanation to the existence of a threshold substrate concentration for microbial growth, which is sometimes observed.

While the relationship between microbial growth rate is sometimes observed to correspond to a type III functional response, Monod's growth law (a type II functional response) is massively used in microbiology, which suggests that this functional response is more often encountered experimentally than the type III response. It can be seen on supplementary material 9.3 (section 9.3) that the functional response of the MTS model tends toward a type II response when its V_H parameter is sufficiently high.

7.1.3 Sensitivity of the MTS model's predictions to the value of its parameters

The sensitivity of the MTS model's prediction to the value of the μ_{max} and V_H parameters, as well as the identifiability of those two parameters, have been studied during this thesis (cf chapter 5). To do so, the parameters of microbial populations simulated by the MTS model were calibrated in order to fit experimental Oxygen Uptake Rate (OUR) curves. The populations were an aerobic acetrotroph guild (OHO, heterotroph), an ammonium oxidizer guild (AOB, autotroph) and a nitrite oxidizer guild (NOB, autotroph). It must be emphasized that those results, and in particular the values obtained through calibration, are only preliminary explorations. Indeed, obtaining robust estimation of the the value of the kinetic parameters μ_{max} and V_H of the MTS model would require much more experimental data and repetition that was was used in this article. This specific issue is discussed in subsection 7.1.3.2.

7.1.3.1 Predictions of the MTS model are formostly influenced by energy and stoichiometry constraints

Independently from the correctness of the estimation of the MTS model's parameters, an interesting conclusion can be drawn from this study. Indeed, the predictions of the MTS model have been found not to be very sensitive to the value of its kinetic parameters. The fact that the kinetic parameters have few influence on the behavior of the MTS model's simulations. Indeed, the kinetic parameters, once calibrated, represent a residual of yet uncharacterized phenomena influencing growth dynamics. This residual is then shown in the simulations results to have less influence on the dynamics than stoichiometry and thermodynamics. Consequently, the most important part of its explanatory power comes from the energy and matter balance constraints it implements, which are not calibrated but knowable *a priori*.

Two evidences of the weakness of the sensibility of the MTS to its parameters can be found in the article corresponding to the chapter 5 of this memoir.

Firstly, a sensitivity study has been made, measuring the distance (as sum of squared difference) between the experimental OUR curves and their prediction by simulation of the MTS model as a function of the value of the two kinetic parameters V_H and μ_{max}. When the distance between the MTS model prediction and the experimental reference OUR is drawn as a surface with μ_{max} and V_H as coordinates, it reveals a shallow "basin" encompassing the optimal (μ_{max}, V_H) couple. The same topology was observed for the two tested guilds (a hetetrotroph guild

and two autotroph guilds). The existence of a single (μ_{max}, V_H) couple minimizing the distance with the observed data also indicates that those parameters are identifiable from experimental data.

Another evidence of the weakness of the sensibility of the MTS model to the value of these parameters is that two sets of drastically different kinetic parameters values led to comparatively small differences in the dynamics of the MTS model simulation.

Indeed in chapter 5, the growth of the three guilds (OHO, AOB, NOB) in a batch is simulated using two different sets of kinetic parameters. In one set (called the "uncalibrated" set), the value of μ_{max} is $5.35 \times 10^{17}\,\mathrm{day}^{-1}$ and the value of V_H is $1\,\mathrm{m}^3 \cdot \mathrm{mol}^{-1}$ for every guild, while in the second set (called the "calibrated" set), the value of μ_{max} is around $3\,\mathrm{day}^{-1}$ and V_H is around $200\,\mathrm{m}^3 \cdot \mathrm{mol}^{-1}$ (approximations are given here because the exact value depends on the guild). While the kinetic parameters are totally different between the two simulations, the ensuing differences in between the simulated dynamics were relatively small. Indeed, in the simulation with the "uncalibrated" parameters, the growth of the OHO population (the only heterotroph population) takes 2 hours, while in the simulation using the "calibrated" parameters, it takes 7.5 hours. As a whole, the other dynamic phenomena of the simulation were also elongated by 5 hours at most. This difference is significant when trying to build a predictive model, however they are remarkably small considering that the maximum growth rate differs by a factor 1e17 between the two simulations.

Those two observations evidence that the mechanics of matter and energy balance implemented in the MTS model by construction contains the major part of its explanatory power.

7.1.3.2 Call for further experimentations to estimate the value of the kinetic parameters of the MTS model

The attempt made in the chapter 5 to calibrate the kinetic parameters of the MTS model were made using only two OUR curves. This work was preliminary and it is certain that many additional experimental data would be required to provide a robust estimation of the kinetic parameters of a guild.

Moreover, the experimental data used to calibrate the MTS model were oxygen uptake rate, in other words; the part of the derivative of oxygen concentration caused by the growth of biomass. This data shows a partial view of the dynamics of the observed experimental system. Concretely, if it had been possible to calibrate the MTS model on another type of data from the same experiment, for example, substrate concentrations through time, the value

of the calibrated parameters may have been different. Indeed, by calibrating the MTS model on OUR curves, a set of parameters making the model reproduce quite well the experimental OUR curve was obtained. However, this does not provide any garantee that other dynamics of the same experiment (such as substrate consumption) are accurately reproduced by the calibrated MTS model. Unfortunately such data was not available for this experiment to directly challenge this claim. However, two arguments can be raised to support it.

Firstly, the comparison between simulations made using the ASMN model ((Hiatt and Grady, 2008)) and the MTS model in chapter 5 strongly suggested stoichiometric differences between the two models. The ASM models family are calibrated engineering models used to simulate activated sludge systems. Their focus is on the accuracy and the practical usability of their predictions. The concentrations and dynamics they predict can then be considered to be close to what can be found in a real activated sludge system. They were therefore considered as a reference to compare with predictions by the MTS model. Both ASMN and MTS models predicted similar yields for guilds OHO, AOB and NOB on their respective electron donor (acetate, ammonium, nitrite, respectively); this result has been observed in multiple compared simulations undertaken in chapter 4 and 5. However, other yields, such as the yield on ammonium ($\mathrm{mol}_{\mathrm{biomass}} \cdot \mathrm{mol}^{-1}_{\mathrm{ammonium}}$) (for OHO guild) or the yield on oxygen ($\mathrm{mol}_{\mathrm{biomass}} \cdot \mathrm{mol}^{-1}_{\mathrm{oxygen}}$), were different between ASMN and MTS. The difference between the yields on ammonium can be explained by the fact that both models use a different biomass formula, with a different proportion of nitrogen. No definitive explanation was proposed for the difference between the yields on oxygen. However it was suggested that the difference between the conditions into which the ASMN model and the MTS model were calibrated could explain this stoichiometry difference. Indeed, while the MTS model assume that the stoichiometry of growth is adjusted to respect the balance of every element, the ASM models' stoichiometry does not (balance is closed on Chemical Oxygen Demand instead), and may be affected by its calibration conditions.

Secondly, the V_H parameter calibrated on the OUR curves in chapter 5 makes the functional response of the MTS model very different from this of the ASMN model. This result is shown in supplementary material 9.3 section 9.3. Those results reveal it is true at least for the functional response of the MTS model on acetate and on oxygen. In this supplementary material section, the functional response of the calibrated MTS model is superimposed to this of the ASMN model (to be more precise; the implementation of the ASMN model used in chapter 5). It is also shown that the value range of V_H which makes the functional responses of MTS closer to

this of the ASMN model are orders of magnitude below the one calibrated on the OUR curve. This suggests that if the parameters of the MTS had been calibrated on other data, such as the electron donor absorption curves instead of oxygen uptake curve, the calibrated parameters could have been different.

These two arguments give strong evidence that calibrating the parameters of the MTS model on a single experimental variable cannot lead to a robust estimation of those parameters. Neither does the comparison of simulations by the MTS model with calibrated engineering models. On one hand this comparison allows to highlight some variables as "well explained". On the other hand the observed discrepancies between the two models are hard to explain.

This reflection allows to draw some requirements for experimental data results usable to calibrate the value of the parameters of the MTS model. Indeed, such experimental results would necessarily include the dynamic tracking of the extracellular concentration of all chemical species involved with the growth process, including biomass. It would also be necessary to have a good estimation of the mean elemental composition of biomass, in order to limit the stoichiometric bias caused by the use of a generic biomass molecule by the MTS model. Moreover, it would also require to make sure that the metabolic reactions simulated by the MTS model accurately represents the reactions performed in the physical experiment. As an exemple, if a population produces a significant amount of extracellular polymers, this reaction, which is not represented by the MTS model, consummes substrate and then may skew the estimation of the parameters. Obtaining experimental data compliant to these requirements from the literature is very hard. Therefore, the robust estimation of the value of the parameters of the MTS model will probably imply careful, dedicated experiments.

As a conclusion, the calibration of the MTS model using experimental data requires to dynamically monitor a lot of parameters which are usually hard to track in an experimental setting (such as biomass). These data will be required in order to provide robust estimations of the kinetic parameters of the MTS model. However, a side effect to the preliminary attempts to estimate those parameters revealed that whatever the value of the kinetic parameters, the most important part of the dynamics simulated by the MTS parameters depends on stoichiometric and energetic factors which are not calibrated because they depends on physicochemical variables.

7.1.4 Use of the MTS model for simulating mixed culture bioprocesses

During this thesis, the MTS model has been used multiple times (cf chapters 4 and 5) to model the dynamics of microbial communities involved in a bioprocess (wastewater treatment through activated sludge). The modelling of mixed culture bioprocesses is largely dominated by engineering models calibrated on experimental data. Those models are based on expert knowledge of their system. They are not intended to provide a representation of the actual drivers behind the system's dynamics. Instead, they are designed to give accurate predictions of the variables of interest (Chemical Oxygen Demand, nitrogen effluent concentration etc), as long as the system stays in its nominal conditions. On the other hand, MTS model expresses a totally different take on microbial communities modelling, not focusing on the quantitative accuracy of its predictions. How can this model contribute to bioprocesses modelling?

The purpose of the MTS model is not to supersede the engineering models. Rather, it can give a valuable contribution to bioprocess modelling by allowing to model bioprocesses for which the amount of experimental data is insufficient to build an engineering model. Indeed, while community dynamics phenomena are predicted by engineering models because of their calibration, the MTS model is able to make some of this dynamics arise from the physicochemical parameters of the system. The second article (chapter 5) contains a result supporting this claim. Indeed, the MTS model was shown to be able to reproduce the sequencial loss of the nitrifying activity of activated sludge in a continuous reactor with the decrease of the sludge retention time. In a first time, the conversion of nitrite to nitrate is lost, then the conversion of ammonium to nitrite is lost for lower sludge retention times. While this phenomenon is reproduced by engineering models of this kind of system (ASMN (Hiatt and Grady, 2008)), it is done through parameter calibration. Indeed, the two nitrifying guilds of the community are given different values for their affinity for oxygen. The fact that both guild are given different kinetic parameters is necessary for the two-stage loss of nitrification to be reproduced by an engineering model. This difference of affinity for oxygen is found through calibration on experimental data. In the MTS model, this behavior is reproduced because of the difference of yield of the two guilds. While affinity for substrate in Monod's law is a calibrated parameter with no fully accepted fundamental meaning (Liu, 2007), the yields of the MTS models are computed from the exergy of their metabolic reactions. The yield difference between two guilds in MTS are thus the result of the matter and energy balance constraints of the two metabolisms. The calibration of the kinetic

parameters of the model in the article does not affect the behavior itself but rather the sludge retention time at which it happens. Indeed, this behavior is predicted by the MTS model while both the ammonium-oxidizing and the nitrite-oxidizing guild are given the same kinetic parameters.

This result suggests that the MTS model has a potential use in the modelling of novel bioprocesses for which calibrated models do not yet exist since it features enhanced abilities in predicting behaviors not known from experimental data. The MTS model, could then be envisioned as a possible candidate to perform exploratory simulations of mixed culture bioprocess settings not existing yet.

Moreover, the MTS model can also contribute to the development of mixed culture bioprocesses for which an engineering model already exist. The MTS model can help improving of the understanding of the physicochemical drivers behind the community dynamics simulated by engineering models. Indeed, the MTS model predicts populations dynamics based on a set of hypotheses on microbial growth. The fact that a specific pattern is predicted both by the MTS and an engineering model (as it can be seen in chapter 5) gives support to the hypotheses on which the MTS model is based.

Expliciting the physicochemical drivers behind observed population dynamics patterns is beneficial to the engineering models of microbial communities. Indeed, departure of the simulated system from the conditions into which the engineering model was calibrated require extensive measurements in order to adapt the engineering model. On the other hand, the explicitation generic physical principles behind some dynamics lessen this work of adaptation of the models.

7.1.5 The estimation of dissipated energy to predict microbial growth yield

7.1.5.1 The method used to estimate dissipated energy in the MTS model's simulation

The MTS model's innovation lays in its formula expliciting the relationship between Gibbs energies and the biomass synthesis rate for a given metabolism. This model also borrows prior concepts from Heijnen and collaborators about metabolism energy balance. According to those concepts, the metabolism's energy balance is solved by setting the proportion of dissipated energy per quantity of biomass produced. This value has been shown to be correlated to metabolisms' characteristics such as their carbon source (Heijnen et al., 1992). However, to this day, there exist no theory explaining the observed correlations. Consequently the MTS model has to rely

on an empirical relationship to set this value and solve the balance of the simulated metabolisms. The choice has been made to stick with the formula originally proposed by Heijnen and collaborators for dissipated energy. Other attempts to calibrate empirical expressions of the dissipated energy have been published since and provided some improvement in the accuracy of the yield predictions ((Liu et al., 2007)). This dissipated energy calculation method was not implemented in the MTS simulations done during this thesis as it was acknowledged while many results based on Heijnen's method were already produced and exploited. Moreover, the yields predicted using Heijnen's relationship were shown to be very close to those predicted using the ASMN model (cf chapter 4, last result section). Therefore, yields predictions using Heijnen's method were already deemed as satisfying. To sum up, the use of Heijnen's dissipated energy formula is not constitutive of the MTS model. It could thus be changed if a more satisfying model of the energy balance of a metabolism could be proposed.

It is worth mentioning that the MTS model omits maintenance in its implementation of Heijnen's energy balance model. In the material and methods (chapter 3), the λ factor, expressing the number of catabolic reaction per anabolic reaction ($\text{mol}_{\text{electron}-\text{donor}} \cdot \text{C} - \text{mol}^{-1}_{\text{biomass}}$) is formulated as

$$\lambda = \frac{-\Delta G_{an} + \Delta G_{dis}}{\Delta G_{cat}} \tag{7.4}$$

(cf equation 3.3). It implies that the overall Gibbs energy differential of microbial growth is ΔG_{dis}; that is, the dissipated energy, in joule per C-mole of biomass produced. However, Heijnen and collaborators consider that the energy dissipated for maintenance also adds up to the overall Gibbs energy differential of growth, as $\frac{m_G}{\mu}$ (Heijnen and Kleerebezem, 1999), where m_G is an energy flux dedicated to maintenance ($\text{joule} \cdot \text{mol}^{-1}_{\text{biomass}} \cdot \text{time}^{-1}$) and μ the population's growth rate.

This specificity of implementation in the MTS model is justified. Indeed, this concept of maintenance is part of the classical Herbert-Pirt growth model (described in the bibliography review of this memoir, chapter 2). Including this maintenance term in the MTS model would had implied to include hypotheses of the Herbert-Pirt model into the MTS model. For example, it would have implied to consider that the maintenance energy cost is linked to the growth rate of the population, and that biomass decay happens when the concentration of substrate is insufficient to supply the maintenance requirements (cf bibliography review, chapter 2, for a description of Heijnen's dynamic model based on Herbert-Pirt's model). However, the purpose of this thesis is to test the effect of the hypotheses

specific to the MTS model on simulations. Not implementing it was a better choice in the process of investigating the MTS model's intrinsic properties, as it would have meddled with multiple features of the model

7.1.5.2 Toward a better knowledge of dissipated energy per unit of biomass synthetized

Predicting the dissipated energy of a metabolism is a method to predict the growth yield (whose specific advantages are discussed in (Heijnen and Dijken, 1991)). An example of an other method to predict microbial growth yield is the TEEM model, described in the bibliography of this memoir (chapter 2).

While models such as the MTS model or Jin and Bethke's model (Jin and Bethke, 2002) provide a theory for the link between physicochemical variables (concentrations, temperature...) and the microbial growth rate, there is no theory expliciting the link between physicochemical variables and microbial growth yield up to now. The only way to propose a new formula to predict microbial growth yield more accurately is to calibrate an empirical model based on experimental observations.

This have been done through multiple studies in the past. An example is Heijnen's dissipated energy method, based on the number of carbons and the reduction degree per carbon of the carbon source of the metabolism. Another example is the TEEM model, whose energy transfer coefficient ϵ have been calibrated on experimental yield measurements too (McCarty, 2007).

As discussed in the previous subsection, the microbial growth yield estimation method is not part of the novel aspects of the MTS model, and it relies on previous methods to estimate dissipated energy. Since the stoichiometric coefficients plays an important role in the formula of the growth rate (cf equation 7.3), the accuracy of the empiricial estimation of microbial growth yield is of paramount importance for the prediction of microbial community dynamics by the MTS model. New empirical relationships to predict microbial growth yield have then been investigated. To do so, databases of experimental growth yields measurements collected by previous studies ((Batstone, 2001; Liu et al., 2007; Roden and Jin, 2011)) were combined. This resulted in the largest, most comprehensive yield database to date. The stoichiometry and Gibbs energy change of the anabolism and catabolism attached to each yield measurement was then estimated. This dataset was then exploited using data analysis methods. The article describing this endeavor and its results is the chapter 6 of this memoir.

The first result revealed by the analysis of the dataset is that while this dataset is currently the most comprehensive, the knowledge it contains on experimental yields measurements is biased. Indeed, some metabolisms

have been the subject of a lot of studies (for example, hydrogenotrophic methanogenesis) while only one or two yield measurements were found for most of the other metabolisms taken individually. As an example, while the yield of aerobic guilds on many different electron donors have been recorded, the yield of denitrifier guilds have been tested with far less different electron donor. This dataset being larger than the one used in previous studies, all the previous studies also suffer from this lack of experimental data. This bias is an important issue since yield prediction models are calibrated on this kind of data. If the distribution of the physicochemical parameters of the metabolisms is too skewed, no robust relationship between those parameters and growth yield can be drawn from the data. Consequently, no totally satisfying empirical predictor of microbial growth yield can be devised as long as this bias is not addressed. Doing so would probably involve performing specific experiments, to ensure an equal representation for a large panel of metabolisms and to ensure that the culture conditions are the comparable for each yield measurement.

Maximum growth yield per metabolism has been chosen as the variable to predict, instead of simply any observed growth yield. Indeed, the yield measurements were collected from various experimental sources, implying diverse growth conditions. In those conditions, the measured experimental growth yield is expected to be mitigated by the maintenance cost, which is unknown and not expected to depend on the intrinsic characteristics of the metabolisms. This motivated the choice of maximum growth yield as the variable to predict

However, despite this bias on the initial data, yield predictions methods based on physicochemical parameters have already provided estimations sufficiently consistent to be used to model microbial communities (González-cabaleiro et al., 2015). The yields estimated using Heijnen's method also have been found to be conform to expectations in the simulations of the MTS model done during this thesis (cf last result section of chapter 4). This justifies the relevance of the attempt of calibrating a yield prediction model from this dataset.

Using linear regression on the dataset to determine the best predictor led to the calibration of new regression models. Those regression models proposed in the article chapter 6, have had their goodness of fit computed on the current dataset and compared with the goodness of fit of models previously proposed in the literature (Heijnen's (Heijnen et al., 1992), Liu's (Liu et al., 2007) and Roden's (Roden and Jin, 2011)). It appeared that the number of carbon of the electron donor and carbon source were better correlated to the yield of the metabolisms in the dataset than their degrees of reduction, contrarily to what was assumed by Liu and collaborators (among others).

A regression model based on catabolic Gibbs energy change and the number of carbons of the carbon source and electron donor of the metabolism was built to predict the maximum yield of the metabolisms of the dataset. While has not been possible to make it more performant (in terms of Mean Squared Error) than the previously proposed models, the choice of the variables used in the predictor (number of carbons) can be thought to be more judicious than this made by Liu and collaborators, when considering the dataset.

The model proposed by Liu and collaborators is the previously proposed model whose Mean Squared Error was found to be the lowest on the dataset. This model separates metabolisms into two categories depending on the degree of reduction of their electron donor, and then consists in two distinct calibrated formulas. In order to provide a yield prediction method with a lower Mean Squared Error than this of Liu and collaborator on the dataset, calibrating a conditional model was attempted. It appeared that partitioning the metabolisms between respirations and fermentations led to a Mean Squared Error slightly lower than this of Liu and collaborators' model on the dataset.

At first sight, the approach used by this article is not novel when compared to previously proposed yield prediction methods based on physicochemical parameters. However, this article has the merit of justify the formula of the new models it proposes by measuring the correlation between observed yields and multiple physicochemical variables. To this regard, it intends to propose a selection of variables and a partitioning of the metabolisms more robust than this of previous models. Moreover, it precisely points at the limitations of such approach by exposing the bias in the current knowledge of microbial growth yield in the literature.

7.2 Perspectives and conclusion

The fundamental properties and some applications of the MTS model were explored during this thesis. However the version of the MTS model considered during this thesis should not be considered as definitive. Indeed, the model, in its current implementation, contains some inconsistencies and limitations that are still to be overcome. These aspects are brought to light in this section in the form of development perspectives for the model. All of these perspectives point at issues in need for investigation in order to increase the ability of the MTS model to capture the drivers of microbial communities. This section ends with the general conclusion of this memoir.

Three different perspectives are discussed in this section; (1) the spatialization of the model, (2) the accounting

for the metabolic versatility of the populations and (3) the clarification of the link between growth yield and growth rate. Those three perspective address rather independent issues, respectively concerning (1) the relationship between a community, and its environment, (2) the plasticity of the notion of population in a community, and (3) individual populations and their relationship between eachother.

7.3 Spatialization of the MTS model

All the simulations performed during this thesis assumed that microbial growth took place in a perfectly mixed culture medium. Since the aim of this thesis was to investigate the fundamental properties of the MTS model, considering culture media of more than zero dimensions would have blured the intended message. However, the spatialization of the simulations of the MTS model constitutes a possible way to enhance its application scope.

Indeed, aside from lab reactors, very few environments can be reasonably considered as "perfectly mixed". Indeed it omits the ability of microbes to structurate their environment and create microenvironments (biofilms and bioflocs are illustrations of this phenomenon). Moreover, the structuration of the environment through the action of microbes is thought to play an important role in the development of their social interactions (Velicer et al., 1998). Considering for example the context of fermented food bioprocesses, microbial communities can be spatially structured as biofilms associated with surfaces, suspended biofilm, dispersed growth in semi-solid substrate; all of which eludes the conception of freely swimming microbes in a perfectly mixed culture medium. As said by Dykhuizen, "Originally an environment is undivided, but as species develop, they divide the environment into separate niches. Species create niches. Different niches have different amounts of resource and are differentially stable" (Dykhuizen, 1998).

In itself, the spatialization of the MTS model will probably be non-trivial and require some work over its underlying theory. Indeed, the MTS model is based on hypotheses about the microscopic environment of the cells. For example, the growth rate formula (equation 7.3) holds true as long as the substrate is uniformly distributed in the cells' microscopic environment, and as long the harvest volumes around the cells do not significantly overlap with eachother. In the case of spatialization through finite element analysis, the meshing will likely be large enough for those considerations not to be raised. However it may be an issue in the case of individual-centric simulations.

7.4 Accounting for the metabolic versatility of microbial populations

Another possible perspective for the improvement of the MTS model is to account for the metabolic versatility of the microbial populations. Indeed, the MTS model, in its current implementation, assigns a "population" to each simulated metabolism. A microbial population is characterized by a biomass concentration, indicating its population density in the culture medium. A population of microbes always catalyzes the same metabolism, and its biomass concentration depends on its anabolic reaction.

This way to represent a microbial community provides coherent results when simulating communities relying on far-from-equilibrium catabolisms. However it is a far less good approximation of the reality of communities catalyzing close-to-equilibrium catabolisms (Rodríguez et al., 2008). The explanation to this statement by Rodriguez an collaborators is detailed in the next paragraph.

Changing its catalyzed metabolism is presumably costly for a cell. Indeed it has been observed in some experimental settings that the redox tower did not apply because some microbial populations favored keeping catalyzing suboptimal metabolisms over switching it (Venturelli et al., 2015; Chen et al., 2017). Contrarily, in culture mediums such as non-methanogenic anaerobic mixed cultures, all metabolisms are close-to-equilibrium, and a slight change in the physicochemical conditions of the culture medium can suddenly turn an exergonic catabolism endergonic, and vice versa. In this kind of setting, the cells are more keen to change their metabolic pathways in order to adapt to their environment. This metabolic versatility is proposed as a reason why the empirical modelling of the population dynamics in anaerobic digesters is less successful than this of activated sludge (Rodríguez et al., 2008).

Moreover, this metabolic versatility phenomenon is not restricted to some specific anaerobic communities. Indeed, it has been experimentally demonstrated that a clonal E. *coli* population is able to differentiate into two functional guild in the presence of two electron donors (Friesen, 2004).

The ability to account for metabolic versatility is then an interesting development perspective to improve the ability of the MTS model to simulate the dynamics of microbial communities. Interestingly, an adaptation of Jin and Bethke's model accounting for the dynamics of pools of enzymes instead of biomasses have already been proposed in the literature (Li et al., 2017). The reasoning of Li and collaborators may be easier to apply to Jin and

Bethke model than to the MTS model since what Jin and Bethke's model fundamentally describe is the rate of ATP production by catabolism, while the MTS model describes the rate of cell division. However this contribution could constitute a possible starting point in the accounting for metabolic versatility in the MTS model.

7.4.1 Investigation the link between dissipated energy and growth rate

A significant issue regarding the MTS model is that it implies a positive relationship between the microbial growth yield and rate. The justification of this relationship, as well as the highlight of the position of other microbial thermodynamic models on this subject, is detailed in supplementary material 9.5. This positive relationship clashes with the common assumption in microbial population dynamics that the relationship between yield and rate is negative (Lele and G watve, 2014). This subsection discusses the implications of this clash and what it implies on the future development of the MTS model.

The MTS model is not the only microbial thermodynamics model to imply a positive relationship between yield and rate. For example, Heijnen's dynamic model also implies the positivity of this relationship, although with a different justification. Nevertheless, the fact that a thermodynamic model implies a positive relationship between yield and rate is not a sufficient argument to ditch it. Indeed, though most experimental observations support a negative relationship between yield and rate, counterexamples exist in the literature, so it is still not known whether this relationship is negative under any circumstance or not. Moreover, a lot of different theories have been proposed to explain the negativity of this relationship; thermodynamics is just one source of explanations among other (the reader is advised to refer to (Lele and G watve, 2014) for a thorough review covering the arguments put forward in the last two sentences).

Nevertheless, a negative relationship between growth yield and rate could possibly capture more of the mechanisms behind microbial communities structure than a positive relationship. Indeed, in an evolutionary perspective, if the relationship between yield and rate is positive, all the microbial populations of a community have one objective under any circumstance; maximizing those variables. The competition between different populations then settles on an unidimensional axis (yield and rate being the same objective). Consequently, this hypothesis implies that the population catalyzing the most thermodynamically favorable metabolism (with minimum energy dissipation) inevitably dominates the ecological succession, as the competitive exclusion principle applies. This

principle corresponds to the commonly accepted "redox tower" phenomenon, that can be simulated by models implying a positive relationship between yield and rate (cf subsection 7.1.1).

However, this pattern is not observed in every microbial communities; far from it. Mutualistic relationships, such as syntrophy, are commonly experimentally observed in microbial communities while the competition between all populations on a single objective would preclude this. In fact, mutualistic relationships contradicting the redox tower principle are observed to happen even in systems were the redox tower is known to be observable, such as in sediments (Chen et al., 2017) or aquifers (Bethke et al., 2011).

The ecological consequences of the negativity of the relationship between yield and rate have already been explored, notably by Pfeiffer and collaborators (Pfeiffer and Bonhoeffer, 2002; Kreft and Bonhoeffer, 2005). According to this hypothesis, equilibria between multiple guilds can be expressed in terms of game theory, which brings subtleties to community structuration. For example, Pfeiffer and collaborators considered a metabolism with a high yield and a low growth rate they called "respiration" and a metabolism with a low yield and a high growth rate they called "fermentation". They demonstrated that the "respiration" metabolism allows for the coexistence of multiple populations exploiting the same substrate, and constitutes a preferable strategy in a structured growth medium, were the dilution rate is low, while the "fermentation" strategy induces a tough competition between the populations and is an advantageous strategy when the dilution rate is high.

Interestingly, while the model of Jin and Bethke implies a negative relationship between yield and rate, they implemented a version of it to reproduce a redox tower. They succesfully simulated this phenomenon, however they failed to reproduce contradictions with the redox tower principle they observed in the culture experiments they conducted in parallel of their simulations. The probable cause to this prediction failure is that then assigned fixed positions on the yield/rate axis of each implemented metabolisms in their simulation (Bethke et al., 2011).

As a conclusion, while it is not known for certain whether thermodynamic microbial kinetics models should assume a positive or negative relationship between the growth yield and the growth rate, a negative relationship could help to capture more of the complexity of microbial communities structuration. By assuming a positive relationship between the growth yield and rate, the MTS model clashes with the current microbial growth kinetics model based on nonequilibrium thermodynamics (Jin and Bethke's model). The fact that two different theoretical approaches on microbial growth can lead to a totally opposed conclusion on a matter as fundamental as the

relationship between yield and rate is a rather remarkable paradox. As explained in the previous paragraph, the position of the MTS model on this paradox precludes it from capturing some mutualistic relationships. Consequently, the theory of the MTS model will have to be reexamined in the future in the light of this paradox before it is possible to go further with the simulation of community structuration from first principles.

7.5 Conclusion

Experimental evidences suggest that the functional organization of a microbial community is structured by deterministic processes, and is then predictable (up to some point) by a theoretical model. Energy flows have been observed to play an important role in shaping the functional organization of microbial communities and many population models have included it in various ways in order to predict microbial populations growth yields and rates.

Those different models convey different approaches about the study of microbial populations dynamics. In all cases, the design of such model starts with the identification of invariants in experimental observations. Some then calibrate empirical models on experimental data, in order to characterize the link between physicochemical parameters and variables. Most yield prediction models ((Heijnen et al., 1992; Liu et al., 2007; Roden and Jin, 2011)) are an example of this approach. Those models can make accurate quantitative predictions and consequently give direct benefits to the study of microbial communities. Some models also propose a theoretical explanation to the observed invariants. Those models are less predictive, since they are purposed to describe a process happening simultaneously with many non-modelled processes. Examples of models conveying this approach are the one by Jin and Bethke, Hoh and Cord-Ruwisch's "equilibrium-based model", and the MTS model (all introduced in chapter 2). The investigation of the hypotheses behind those models are hard, because of the multitude of unknown processes occuring at the same time in microbial growth. However, earned support for a theoretical model is rewarding, as it allows to make predictions about microbial communities never observed before. Ultimately, both approaches are legitimate and complementary. In fact, most models pertain to both, to some extent, for example, the TEEM model (introduced in chapter 2) predicts growth yield based on parameters calibrated on experimental data, however it also involve theoretical hypotheses about the intracellular details of metabolisms. As regards the MTS model, though it relies of empirical prediction of dissipated energy and has some kinetic parameters, it is

rather strongly inclined toward the second approach.

The MTS model was shown during this thesis to be able to reproduce microbial growth patterns, both at the scale of individual populations (Liebig's rule, minimal substrate requirement for growth...) and communities (ecological successions along time and concentration gradients). Its specificity, and novelty, is to make those results arise from a theoretical representation of microbial growth, a microscopic model pertaining to statistical physics. The parcimony in the implementation of the model allows to conclude that the patterns and behaviors predicted by the MTS model are the result of the theoretical hypotheses behind it. Altogether, very few models have attempted to propose a microscopic-scale energy-based theory to explain patterns in microbial growth dynamics. Jin and Bethke's model is a possible example, however its application-scope is more limited than this of the MTS model (as discussed in section 7.1.2.2 ealier in this chapter). Therefore, no model have conveyed this approach this far before the MTS model.

While energy-based models (let alone theory-inclined ones) have not been the most favored approach toward microbial community modelling up to now, the abilities of the MTS model, as revealed during this thesis, constitutes an positive signal toward this direction. This contribution advocates for the development of a new class of microbial ecology and engineering models, built on more robust theoretical grounds. Such models are indeed needed to predict the behavior of microbial communities in situations not encountered yet, and thus address biotechnological and environmental challenges in the near future.

Chapter 8:

Bibliography

Andersen KB, Von Meyenburg K (1980). Are growth rates of Escherichia coli in batch cultures limited by respiration? Journal of Bacteriology 144: 114-123.

Andrews JF (1968). A mathematical model for the continuous culture of microorganisms utilizing inhibitory substrates. Biotechnology and Bioengineering 10: 707-723.

Araujo Granda P, Gras A, Ginovart M, Moulton V (2016). INDISIM-Paracoccus, an individual-based and thermodynamic model for a denitrifying bacterium. Journal of Theoretical Biology 403: 45-58.

Araujo P, Gras A, Ginovart M (2015). Thermodynamic behavior-rules for a bacterial individual-based model to study the denitrification process. IFAC-PapersOnLine 28: 743-748.

Araujo PG, Gras A, Ginovart M (2016). MbT-Tool: An open-access tool based on Thermodynamic Electron Equivalents Model to obtain microbial-metabolic reactions to be used in biotechnological process. Computational and Structural Biotechnology Journal 14: 325-332.

Bajpai-Dikshit J, Suresh AK, Venkatesh KV (2003). An Optimal Model for Representing the Kinetics of Growth and Product Formation by Lactobacillus rhamnosus on Multiple Substrates. Journal of Bioscience and Bioengineering 96: 481-486.

Bapat PM, Bhartiya S, Venkatesh KV, Wangikar PP (2006). Structured kinetic model to represent the utilization of multiple substrates in complex media during rifamycin B fermentation. Biotechnology and Bioengineering 93: 779-790.

Batstone D, Angelidaki I, Pavlostathis SG, Sanders WTM, Vavilin V, Keller J et al (2001). Anaerobic digestion model No 1 (ADM1).

Battley E (1998). thermochimica acta The development of direct and indirect methods for the study of the thermodynamics of microbial growth 1. Thermochimica Acta 309: 17-37.

Battley EH, Ptnam RL, Boerio-goates J (1997). 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. Termochimica Acta 298: 37-46.

Bauchop T, Elsden SR (1960). The growth of microorganisms in relation to their metabolism and energy supply. Acta microbiologica Bulgarica 23: 457-469.

Bethke CM, Sanford Ra, Kirk MF, Jin Q, Flynn TM (2011). The thermodynamic ladder in geomicrobiology. American Journal of Science 311: 183-210.

Boehrer B, Schultze M (2008). Stratification Lakes. Reviews of Geophysics 46: 1-27.

Bungay HR (1994). Growth rate expressions for two substrates one of which is inhibitory. Journal of Biotechnology 34: 97-100.

Burke C, Steinberg P, Rusch DB, Kjelleberg S, Thomas T (2011). Bacterial community assembly based on functional genes rather than species. Proceedings of the National Academy of Sciences of the USA 108: 14288-14293.

Button DK (1998). Nutrient uptake by microorganisms according to kinetic parameters from theory as related to cytoarchitecture. Microbiology and molecular biology reviews : MMBR 62: 636-645.

Capela S, Gillot S, Héduit A (2004). Comparison of Oxygen-Transfer Measurement Methods Under Process Conditions. Water Environment Research 76.

Chapelle F, Lovley D (1992). Competitive exclusion of sulfate reduction by Fe (lll) reducing bacteria. pp 29-36.

Chen J, Hanke A, Tegetmeyer HE, Kattelmann I, Sharma R, Hamann E et al (2017). Impacts of chemical gradients on microbial community structure 11: 1-12.

Ciggin aS, Insel G, Majone M, Orhon D (2013). Model evaluation of starch utilization by acclimated biomass with different culture history under pulse and continuous feeding. Bioresource Technology 138: 163-171.

Contois DE (1959). Kinetics of bacterial growth: relationship between population density and specific growth rate of continuous cultures. Journal of general microbiology 21: 40-50.

Corkrey R, Olley J, Ratkowsky D, McMeekin T, Ross T (2012). Universality of thermodynamic constants governing biological growth rates. PLoS One 7: e32003.

Corkrey R, McMeekin TA, Bowman JP, Ratkowsky DA, Olley J, Ross T (2016). The Biokinetic Spectrum for Temperature. PLoS One 11: e0153343.

Crawford R, Crawford D (1996). Bioremediation: Principles and Applications. Cambridge University Press.

Daims H, Taylor MW, Wagner M (2006). Wastewater treatment: a model system for microbial ecology. Trends Biotechnol 24: 483-489.

Danger M, Daufresne T, Lucas F, Pissard S, Lacroix G (2008). Does Liebig's law of the minimum scale up from species to communities? Oikos 117: 1741-1751.

De Filippis FAGPFJAGDE (2016). Metatranscriptomics reveals temperature-driven functional changes in microbiome impacting cheese maturation rate. Scientific Reports 6: 1-12.

De Wit R, Bouvier T (2006). 'Everything is everywhere, but, the environment selects'; what did Baas Becking and Beijerinck really say? Environmental Microbiology 8: 755-758.

Denny M (2014). Buzz Holling and the Functional Response. Bulletin of the Ecological Society of America 95: 200-203.

Desmond-Le Quéméner E, Bouchez T (2014). A thermodynamic theory of microbial growth. The ISME journal: 1-5.

Droop MR (1974). The nutrient status of algal cells in continuous culture. Journal of the Marine Biological Association of the United Kingdom 54: 825.

Dykhuizen DE (1998). Santa Rosalia revisited: Why are there so many species of bacteria? Antonie van Leeuwenhoek 73: 25-33.

Eyring H (1935). The Activated Complex in Chemical Reactions. The Journal of Chemical Physics 3: 107-115.

Falkowski PG, Fenchel T, Delong EF (2008). The Microbial Engines That Drive Earth 's Biogeochemical Cycles. Science 320: 1034-1039.

Fernández A, Huang S, Seston S, Xing J, Hickey R, Criddle C et al (1999). How stable is stable? Function versus community composition. Applied and Environmental Microbiology 65: 3697-3704.

Fernandez aS, Hashsham Sa, Dollhopf SL, Raskin L, Glagoleva O, Dazzo FB et al (2000). Flexible community structure correlates with stable community function in methanogenic bioreactor communities perturbed by glucose. Applied and Environmental Microbiology 66: 4058-4067.

Friesen ML (2004). Experimental Evidence for Sympatric Ecological Diversification due to Frequency-Dependent Competition in Escherichia coli. Evolution 58: 245-260.

Gause F (1934). The Struggle for Existence. The Yale journal of biology and medicine 7: 609-609.

Gloag ES, Turnbull L, Whitchurch CB (2015). Bacterial stigmergy: an organising principle of multicellular collective behaviours of bacteria. Scientifica (Cairo) 2015: 387342.

González-Cabaleiro R, Lema JM, Rodríguez J, Kleerebezem R (2013). Linking thermodynamics and kinetics to assess pathway reversibility in anaerobic bioprocesses. Energy & Environmental Science 6: 3780-3780.

González-Cabaleiro R, Lema JM, Rodríguez J (2015a). Metabolic energy-based modelling explains product yielding in anaerobic mixed culture fermentations. PLoS ONE 10: 1-17.

González-Cabaleiro R, Ofiţeru ID, Lema JM, Rodríguez J (2015b). Microbial catabolic activities are naturally selected by metabolic energy harvest rate. The ISME Journal 9: 2630-2641.

Griggs D, Stafford-Smith M, Gaffney O, Rockström J, Öhman MC, Shyamsundar P et al (2013). Policy: Sustainable development goals for people and planet. Nature 495: 305-307.

Großkopf T, Soyer OS (2016). Microbial diversity arising from thermodynamic constraints. The ISME Journal: 1-9.

Großkopf T, Zenobi S, Alston M, Folkes L, Swarbreck D, Soyer OS (2016). A stable genetic polymorphism underpinning microbial syntrophy. The ISME Journal: 1-10.

Gujer W, Morgens I, Mino T, Van Loosdrecht MCM (2000). activated sludge models asm1, asm2, asm2d and asm3.

Haldane JBS (1930). Enzymes.

Hanselman KW (1991). microbial energetics applied to waste repositories. Experientia 47.

Hansson L (2003). Why ecology fails at application: should we consider variability more than regularity? Oikos2 100: 624-627.

Hardin G (1960). The competitive exclusion principle. pp 1292-1297.

Harwood CR, Cutting SM (1990). Molecular biological methods for Bacillus: Chichester, United Kingdom.

Hauduc H, Gillot S, Rieger L, Ohtsuki T, Shaw a, Tak??cs I et al (2009). Activated sludge modelling in practice: an international survey. Water science and technology : a journal of the International Association on Water Pollution Research 60: 1943-1951.

Hauduc H, Rieger L, Takacs I, Heduit A, Vanrolleghem PA, Gillot S (2010). A systematic approach for model verification: application on seven published activated sludge models. Water Sci Technol 61: 825-839.

Heijnen JJ, Dijken JPV (1991). In Search of a Thermodynamic Description of Biomass Yields for the Chemotrophic Growth of Microorganisms. biotechnology and bioengineering 39: 833-858.

Heijnen JJ, Van Loosdrecht MCM, Tijhuis L (1992). A black box mathematical model to calculate auto- and heterotrophic biomass yields based on gibbs energy dissipation. Biotechnology and Bioengineering 40: 1139-1154.

Heijnen JJ, Kleerebezem R (2010). Bioenergetics of microbial growth. Encyclopedia of Industrial Biotechnology: Bioprocess, Bioseparation and Cell Technology: 1-24.

Heijnen JJK, R. (2010). Bioenergetics of microbial growth. Encyclopedia of Industrial Biotechnology: Bioprocess, Bioseparation and Cell Technology: 1-24.

Henze M, Grady CPL, Gujer W, Marais GVR, Matuso T (1987). Activated Sludge Model No. 1. IAWQ Scientific and Technical Report No 1 1.

Henze M, Van Loosdrecht MCM, Mino T (2000). Activated Sludge Models ASM1.

Herbert BYD, Elsworth R (1956). The Continuous Culture of Bacteria ; a Theoretical and Experimental Study. Journal of general microbiology 14: 601-622.

Hiatt WC, Grady CPL (2008). An Updated Process Model for Carbon Oxidation, Nitrification, and Denitrification. Water Environment Research 80: 2145-2156.

Hoh CY, Cord-Ruwisch R (1996). A practical kinetic model that considers endproduct inhibition in anaerobic digestion processes by including the equilibrium constant. Biotechnology and Bioengineering 51: 597-604.

Hoover SR, Jasewicz L, Porges N (1953). Aerobic Treatment of dairy wastes. applied microbiology 1: 262–270.

Hutchinson GE (1957). A Treatise on Limnology: Geography, physics, and chemistry. pt. 1. Geography and physics of lakes: 1016.

Huttenhower C, Gevers D, Knight R, Abubucker S, Badger JH, Chinwalla AT et al (2012). Structure, function and diversity of the healthy human microbiome. Nature 486: 207-214.

Istok JD, Park M, Michalsen M, Spain aM, Krumholz LR, Liu C et al (2010). A thermodynamically-based model for predicting microbial growth and community composition coupled to system geochemistry: Application to uranium bioreduction. Journal of contaminant hydrology 112: 1-14.

Jacob EB, Becker I, Shapira Y, Levine H (2004). Bacterial linguistic communication and social intelligence. Trends in Microbiology 12: 366-372.

Jin Q, Bethke CM (2002). Kinetics of Electron Transfer through the Respiratory Chain. Biophysical Journal 83: 1797-1808.

Jin Q, Bethke CM (2003). A New Rate Law Describing Microbial Respiration 69: 2340-2348.

Jin Q, Bethke CM (2005). Predicting the rate of microbial respiration in geochemical environments. Geochimica et Cosmochimica Acta 69: 1133-1143.

Jin Q, Bethke CM (2007). The thermodynamics and kinetics of microbial metabolism. American Journal of Science 307: 643-677.

Jin Q, Bethke CM (2008). Reply to comment: The thermodynamics and kinetics of microbial metabolism by John Walther. American Journal of Science 308: 1117-1118.

Ju F, Guo F, Ye L, Xia Y, Zhang T (2014). Metagenomic analysis on seasonal microbial variations of activated sludge from a full-scale wastewater treatment plant over 4 years. Environmental Microbiology Reports 6: 80-89.

Kaiser K, Wemheuer B, Korolkow V, Wemheuer F, Nacke H, Schöning I et al (2016). Driving forces of soil bacterial community structure, diversity, and function in temperate grasslands and forests. Scientific Reports 6: 33696-33696.

Kleerebezem R, Van Loosdrecht MCM (2010). A Generalized Method for Thermodynamic State Analysis of Environmental Systems. Critical Reviews in Environmental Science and Technology 40: 1-54.

Kovarova-Kovar K, Egli T (1998). Growth Kinetics of Suspended Microbial Cells: From Single-Substrate-Controlled Growth to Mixed-Substrate Kinetics. Microbiol Mol Biol Rev 62: 646-666.

Kreft JU, Bonhoeffer S (2005). The evolution of groups of cooperating bacteria and the growth rate versus yield trade-off. Microbiology 151: 637-641.

LaRowe DE, Amend JP (2016). The energetics of anabolism in natural settings. The ISME Journal.

Lawton JH (1999). Are There General Laws in Ecology ? oikos 84: 177-192.

Lele UN, G watve M (2014). Bacterial Growth Rate and Growth Yield: Is There A Relationship? Proceedings of the Indian National Science Academy 80: 537-537.

Lemanceau P, Blouin M, Muller D, Moenne-Loccoz Y (2017). Let the Core Microbiota Be Functional. Trends Plant Sci 22: 583-595.

Li M, Qian WJ, Gao Y, Shi L, Liu C (2017). Functional Enzyme-Based Approach for Linking Microbial Community Functions with Biogeochemical Process Kinetics. Environ Sci Technol 51: 11848-11857.

Linton JD, Stephenson RJ (1978). A preliminary study on growth yields in relation to the carbon and energy content of various organic growth substrates. FEMS Microbiology Letters 3: 95-98.

Little AEF, Robinson CJ, Peterson SB, Raffa KF, Handelsman J (2008). Rules of engagement: interspecies interactions that regulate microbial communities. Annual review of microbiology 62: 375-401.

Liu J-S, Vojinović V, Patiño R, Maskow T, von Stockar U (2007). A comparison of various Gibbs energy dissipation correlations for predicting microbial growth yields. Thermochimica Acta 458: 38-46.

Liu Y (2006). A simple thermodynamic approach for derivation of a general Monod equation for microbial growth. Biochemical Engineering Journal 31: 102-105.

Liu Y (2007). Overview of some theoretical approaches for derivation of the Monod equation. Applied Microbiology and Biotechnology 73: 1241-1250.

Louca S, Jacques SMS, Pires APF, Leal JS, Srivastava DS, Parfrey LW et al (2016a). High taxonomic variability despite stable functional structure across microbial communities. Nature Ecology & Evolution 1: 0015-0015.

Louca S, Parfrey LW, Doebeli M (2016b). Decoupling function and taxonomy in the global ocean microbiome. Science 353: 1272-1277.

Louca S, Doebeli M (2017). Taxonomic variability and functional stability in microbial communities infected by phages. Environ Microbiol 19: 3863-3878.

Ludovisi a (2009). Exergy vs information in ecological successions: Interpreting community changes by a classical thermodynamic approach. Ecological Modelling 220: 1566-1577.

Lusk JE, Williams RJP, Kennedy EP (1968). Magnesium and the growth of Escherchia coli. Journal of biological chemistry 243: 2618-2624.

Mayberry WR, Prochazka GJ, Payne WJ (1967). Growth Yields of Bacteria on Selected Organic Compounds. Appl Envir Microbiol 15: 1332-1338.

Mayberry WR, Prochazka GJ, Payne WJ (1968). Factors derived from studies of aerobic growth in minimal media. Journal of Bacteriology 96: 1424-1426.

McCarty PL (1965). Thermodynamics of biological synthesis and growth. Air Water Pollut 9: 621-639.

McCarty PL (2007). Thermodynamic Electron Equivalents Model for Bacterial Yield Prediction: Modifications and Comparative Evaluations. Biotechnology and Bioengineering 97: 377-388.

McGill BJ, Enquist BJ, Weiher E, Westoby M (2006). Rebuilding community ecology from functional traits. Trends in Ecology and Evolution 21: 178-185.

McGovern PE, Zhang J, Tang J, Zhang Z, Hall GR, Moreau Ra et al (2004). Fermented beverages of pre- and proto-historic China. Proceedings of the National Academy of Sciences 101: 17593-17598.

Meister M, Winkler D, Rezavand M, Rauch W (2017). Integrating hydrodynamics and biokinetics in wastewater treatment modelling by using smoothed particle hydrodynamics. Computers & Chemical Engineering 99: 1-12.

Milo R, Phillips R (2015). Cell biology by the number.

Minkevich IG, Eroshin VK (1973). Productivity and heat generation of fermentation under oxygen limitation. Folia Microbiologica 18: 376-385.

Monod J (1949). The Growth of Bacterial Cultures. Annual Review of Microbiology 3: 371-394.

Moon HG, Jang YS, Cho C, Lee J, Binkley R, Lee SY (2016). One hundred years of clostridial butanol fermentation. FEMS Microbiol Lett 363.

Müller B, Bryant LD, Matzinger A, Wüest A (2012). Hypolimnetic oxygen depletion in eutrophic lakes. Environmental Science and Technology 46: 9964-9971.

Nature Methods Editorial (2009). Metagenomics versus Moore's law. Nature Methods 6: 623.

Nelson MB, Martiny AC, Martiny JBH (2016). Global biogeography of microbial nitrogen-cycling traits in soil. Proceedings of the National Academy of Sciences 113: 8033-8040.

Nemergut DR, Schmidt SK, Fukami T, O'Neill SP, Bilinski TM, Stanish LF et al (2013). Patterns and processes of microbial community assembly. Microbiol Mol Biol Rev 77: 342-356.

Nikolajsen K, Nielsen J, Villadsen J (1991). Structured modeling of a microbial system: III. Growth on mixed substrates. Biotechnology and bioengineering 38: 24-29.

Noguera DR, Brusseau Ga, Rittmann BE, Stahl Da (1998). A Unified Model Describing the Role of Hydrogen in the Growth of Desulfovibrio vulgaris under Different Environmental Conditions.

Odum EP (1969). Strategy of Ecosystem Development. Science 164: 8.

Pfeiffer T, Schuster S, Bonhoeffer S (2001). Cooperation and competition in the evolution of ATP-producing pathways. Science (New York, NY) 292: 504-507.

Pfeiffer T, Bonhoeffer S (2002). Evolutionary Consequences of Tradeoffs between Yield and Rate of ATP Production. Zeitschrift für Physikalische Chemie 216: 51-51.

Pidello A (2014). Principes de chimie redox en écologie microbienne: 143.

Pirt SJ (1965). The Maintenance Energy of Bacteria in Growing Cultures. Proceedings of the Royal Society of London Series B, Biological Sciences 163: 224-231.

Pirt SJ (1982). Maintenance energy: a general model for energy-limited and energy-sufficient growth. Archives of microbiology 133: 300-302.

Prochazka GJ, Payne WJ, Mayberry WR (1970). Calorific content of certain bacteria and fungi. Journal of bacteriology 104: 646-649.

R Development Core Team (2008). R: A Language and Environment for Statistical Computing: Vienna, Austria

Raes J, Letunic I, Yamada T, Jensen LJ, Bork P (2011). Toward molecular trait-based ecology through integration of biogeochemical, geographical and metagenomic data. Molecular systems biology 7: 473-473.

Rajagopal R, Rousseau P, Bernet N, Beline F (2011). Combined anaerobic and activated sludge anoxic/oxic treatment for piggery wastewater. Bioresour Technol 102: 2185-2192.

Roden EE, Jin Q (2011). Thermodynamics of microbial growth coupled to metabolism of glucose, ethanol, short-chain organic acids, and hydrogen. Applied and Environmental Microbiology 77: 1907-1909.

Rodríguez J, Lema JM, Kleerebezem R (2008). Energy-based models for environmental biotechnology. Trends in Biotechnology 26: 366-374.

Roels J (1983). Energetics and kinetics in biotechnology. Elsevier Biomedical Press.

Roels Ja (1980). Application of macroscopic principles to microbial metabolism. Biotechnology and bioengineering 103: 2-59; discussion 51.

Rössle WH, Pretorius Wa (2001). A review of characterisation requirements for in-line prefermenters paper 2: Process characterisation. Water SA 27: 413-422.

Rousk J, Bengtson P (2014). Microbial regulation of global biogeochemical cycles. Front Microbiol 5: 103.

Shammas N (1986). Interactions of Temperature, pH, and Biomass on the Nitrification Process. Water Pollution Control Federation 58: 52-59.

Shen TCD, Albenberg L, Bittinger K, Chehoud C, Chen YY, Judge Ca et al (2015). Engineering the gut microbiota to treat hyperammonemia. Journal of Clinical Investigation 125: 2841-2850.

Simberloff D (2004). Community Ecology: Is It Time to Move On? (An American Society of Naturalists Presidential Address). The American Naturalist 163: 787-799.

Stouthamer aH (1973). A theoretical study on the amount of ATP required for synthesis of microbial cell material. Antonie van Leeuwenhoek 39: 545-565.

Strycharz-Glaven SM, Snider RM, Guiseppi-Elie A, Tender LM (2011). On the electrical conductivity of microbial nanowires and biofilms. Energy & Environmental Science 4: 4366.

Svirezhev SEJYM (2004). Towards a Thermodynamic Theory for Ecological Systems.

Treseder KK, Balser TC, Bradford Ma, Brodie EL, Dubinsky Ea, Eviner VT et al (2012). Integrating microbial ecology into ecosystem models: Challenges and priorities. Biogeochemistry 109: 7-18.

Van de Leemput Ia, Veraart AJ, Dakos V, De Klein JJM, Strous M, Scheffer M (2011). Predicting microbial nitrogen pathways from basic principles. Environmental Microbiology 13: 1477-1487.

Van Dongen U, Jetten MSM, Van Loosdrecht MCM (2001). The SHARON®-Anammox® process for treatment of ammonium rich wastewater. Water Science and Technology 44: 153-160.

VanBriesen JM, Rittmann BE (2000). Mathematical description of microbiological reactions involving intermediates. Biotechnology and bioengineering 67: 35-52.

Velicer GJ, Kroos L, Lenski RE (1998). Loss of social behaviors by myxococcus xanthus during evolution in an unstructured habitat. Proceedings of the National Academy of Sciences of the United States of America 95: 12376-12380.

Velicer GJ (2003). Social strife in the microbial world. Trends in Microbiology 11: 330-337.

Venturelli OS, Zuleta I, Murray RM, El-Samad H (2015). Population Diversification in a Yeast Metabolic Program Promotes Anticipation of Environmental Shifts. PLoS Biology 13: 1-24.

Verstraete W (2007). Microbial ecology and environmental biotechnology. ISME J 1: 4-8.

von Stockar U, Liu J (1999). Does microbial life always feed on negative entropy ? Thermodynamic analysis of microbial growth 1412.

von Stockar U, Vojinović V, Maskow T, Liu J (2008). Can microbial growth yield be estimated using simple thermodynamic analogies to technical processes? Chemical Engineering and Processing: Process Intensification 47: 980-990.

Wade MJ, Harmand J, Benyahia B, Bouchez T, Chaillou S, Cloez B et al (2016). Perspectives in mathematical modelling for microbial ecology. Ecological Modelling 321: 64-74.

Weinstock MT, Hesek ED, Wilson CM, Gibson DG (2016). Vibrio natriegens as a fast-growing host for molecular biology. Nat Methods 13: 849-851.

Westerhoff HV, Hellingwerf KJ, Van Dam K (1983). Thermodynamic efficiency of microbial growth is low but optimal for maximal growth rate. Proceedings of the National Academy of Sciences of the United States of America 80: 305-309.

Whitman WB, Coleman DC, Wiebe WJ (1998). Prokaryotes: the unseen majority. Proc Natl Acad Sci U S A 95: 6578-6583.

Wickham H (2014). Tidy Data. Journal of Statistical Software 59: 1-23.

Widder S, Allen RJ, Pfeiffer T, Curtis TP, Wiuf C, Sloan WT et al (2016). Challenges in microbial ecology: building predictive understanding of community function and dynamics. The ISME journal: 1-12.

Zavarzin GA (2006). Winogradsky and modern microbiology. Microbiology 75: 501-511.

Chapter 9:

Appendix

9.1 Growth limitations by nutrients predicted by the MTS model

The MTS model considers that every chemical species required by the growth process of a microbial guild exerts a limitation on its growth rate. This limitation takes the form of what is called a "tuning factor" in the model's semantics. This factor taking value between 0 and 1 can be high or low depending on the demand over availability ratio of the chemical species in the culture medium. The MTS model also implies to consider a model, generic biomass formula. While many biomass formulas have been defined and can be used interchangeably without affecting the construction of the model itself, the biomass formula proposed by Battley (Battley et al., 1997) have been used during this thesis. This biomass formula have been obtained from experimental measurements on a culture of Saccharomyces *cerervisiae*; its full formula is $\mathrm{C_1H_{1.613}O_{0.557}N_{0.158}P_{0.012}S_{0.003}K_{0.022}Mg_{0.003}Ca_{0.001}}$, however it was generally assumed that the simpler formula $\mathrm{C_1H_{1.613}O_{0.557}N_{0.158}}$ can be used interchangeably for the sake of simplicity. This chapter raises the question of the impact of this simplifying assumption.

The more complex formula involves some supplementary elements (P, K, Mg, Ca) in biomass composition, which are thought to be provided to the cell by respectively $\mathrm{PO_4^{-2}}$, $\mathrm{SO_4^{-2}}$, $\mathrm{K^+}$, $\mathrm{Mg^{+2}}$ and $\mathrm{Ca^{+2}}$. As those nutrients are unlikely to be required by the catabolic reaction, it can be assumed that their corresponding stoichiometric coefficient in the energy-balanced metabolic reaction of any microbial guild equals their stoichiometric number in biomass composition, that is, $0.012\mathrm{PO_4^{-2}}$, $0.003\mathrm{SO_4^{-2}}$, $0.022\mathrm{K^+}$, $0.003\mathrm{Mg^{+2}}$ and $0.001\mathrm{Ca^{+2}}$.

The tuning factor of each of those chemical species on the growth rate of any guild can then be computed while assuming their concentration from this of a minimal M9 medium (Harwood and Cutting, 1990) and that the V_h parameter is $1\,\mathrm{m^3 \cdot mol_X^{-1}}$;

chemical species	concentration in M9 medium (M)	tuning factor
$\mathrm{PO_4^{-2}}$	3.48e-1	0.96
$\mathrm{SO_4^{-2}}$	9.97e-2	0.97
$\mathrm{K^+}$	1.10e-1	0.81
$\mathrm{Mg^{+2}}$	7.8e-4	0.02
$\mathrm{Ca^{+2}}$	9.99e-3	0.90

The observation which can be made from table 9.1 is that considering the composition of a M9 medium, the limitation exerted by a nutrient on growth greatly depends on the nutrient (the overall limitation for all chemical species in 9.1) is 1.36e-2). However, the importance of this limitation also greatly depends on the values of the kinetic parameters of the MTS model. In the first article (chapter 4), the V_h parameter is $1\,m^3 \cdot mol^{-1}$ for most simulated guilds, however the μ_{max} parameter is so high (of order of magnitude 1e17) that even the tuning factor of magnesium has no significant impact on the simulation kinetics. In the second article, a μ_{max} of order of magnitude 1 was set, but the value of the V_h parameter was of order of magnitude 100, which would have led to different tuning factor values (for example, with a V_h of $100\,m^3 \cdot mol_X^{-1}$) the tuning factor of Mg^{+2} is 0.99.

As a conclusion, the tuning factors of nutrients ordinarily considered as neglectible in mass balance (K, S, P, Na...) can possibly be significant in some cases, but the two different takes on the MTS model parametrization tested during this thesis implied kinetic parameters values which made those tuning factors actually neglectible. Nevertheless, limitations of microbial growth by nutrients such as magnesium do exist (Lusk et al., 1968), and the possibility offered by the MTS model to account for the impact of any nutrient on the growth rate of a microbial population is an asset when compared to growth models not based on an explicit description of growth stoichiometry.

9.2 Example of getExperimentalDesign files

The `getExperimentalDesign.m` function is expected by `simulate.m` to return either a single simulation structure or a cell array of simulation structures. If a cell array of simulation structures is returned by the `getExperimentalDesign.m` function, each thus-defined system s integrated one after another by the program. This functionality is intended to allow for successive simulations with automatically generated parameters values, such as in parametric sweeps.

In order to help with the exploitation of the results, the program creates a file called "experimetalDesignLog.csv" in the simulation directory. This file is a csv file linking the simulations' indexes and names to the path to their result file. This file is created whatever the number of simulations performed.

An example of the definition of the structure for a single simulation is the following;

```
experiment = struct('name', 'experiment 1', ...
'chunkSize', 1, ...
'totalTime', 5, ...
'resultRoot', '../out/results/', ...
'errorDir', '../out/errors/', ...
'solver', @ode113, ...
'chunkSizeEasingCoeffient', 1, ...
'solverOptions', odeset('NonNegative', 1:p.C, 'AbsTol', 1e-9), ...
'fixedParameters', fixedParameter);
experimentalDesign = experiment;
```

An example of the definition of the structure for a list of simulations is the following;

```
mumaxSize = 3;
VhSize = 5;
mumaxValues = linspace(5, 12, mumaxSize);
VhValues = linspace(1, 50, VhSize);
```

```
function experiment = getExperiment(x)
mumax = mumaxValues(mod(x, mumaxSize) + 1);
vh = ceil(x / VhSize);
experiment = struct('name', sprintf('experiment %d', x), ...
'chunkSize', 1, ...
'totalTime', 5, ...
'resultRoot', '../out/results/', ...
'errorDir', '../out/errors/', ...
'solver', @ode113, ...
'chunkSizeEasingCoeffient', 1, ...
'solverOptions', odeset('NonNegative', 1:p.C, 'AbsTol', 1e-9), ...
'fixedParameters', struct('initialStateDictionary', initialStateDictionary, ...
'vh', ones(p.C, p.G) * vh, ...
'mumaxharv', ones(1, p.G) * mumax, ...
'hooks', {hooks}, ...
'guilds', {guilds}));
end
experimentalDesign = arrayfun(@(x) getExperiment(x), 1:(mumaxSize * VhSize));
```

In this example, a parametric sweep is performed through the multiple simulations. 3 values of the `mumaxharv` parameter are tested in combination with 5 values of the `vh` parameter. The simulation structures are produced as a list using Matlab's `arrayfun` function along a simple integer sequence.

9.3 Specificities of the MTS model's growth function

In this supplementary material, the growth function of the MTS model is compared to this of Monod's law.

Monod's law can natively account for the influence of only one limiting substrate, so it will be compared to a single-substrate form of the MTS growth function.

Moreover, both growth functions express growth limitation as a factor (ranging from 0 to 1) applied to a "maximum growth rate" constant μ_{max}. This constant has a theoretical meaning in the context of the MTS construction (cf first section of chapter 7), while many attempts have been made to attach a theoretical meaning to Monod's law μ_{max} and none of them reached consensus (Liu, 2007). While the maximum growth rate constant of the MTS model and Monod's law usually take value in different ranges, the purpose of this appendix is to compare the shape of the growth functions qualitatively, so those constants will be excluded from the comparison. The functions being compared in this document are then;

$$\mu([S]) = e^{-\frac{1}{V_H \cdot [S]}} \tag{9.1}$$

for the MTS model, and

$$\mu([S]) = \frac{[S]}{K_S + [S]} \tag{9.2}$$

for Monod's law.

The only parameters in those functions is then the Harvest Volume V_H for the MTS model growth function and the substrate affinity coefficient K_S for Monod's law. As the issue in this document is to compare the response curve of the two aforementioned simplified functions (equation 9.1 and 9.2), a value of 1 will be considered for both parameters.

Superimposition of Monod's and MTS's growth curves

The superimposition of both functions and reveal that their characteristic shapes are fundamentally different. Indeed, using the MTS growth mode, growth does not start until a critical substrate concentration is reached.

The consequences of this are evoked in the discussion of this memoir (chapter 7).

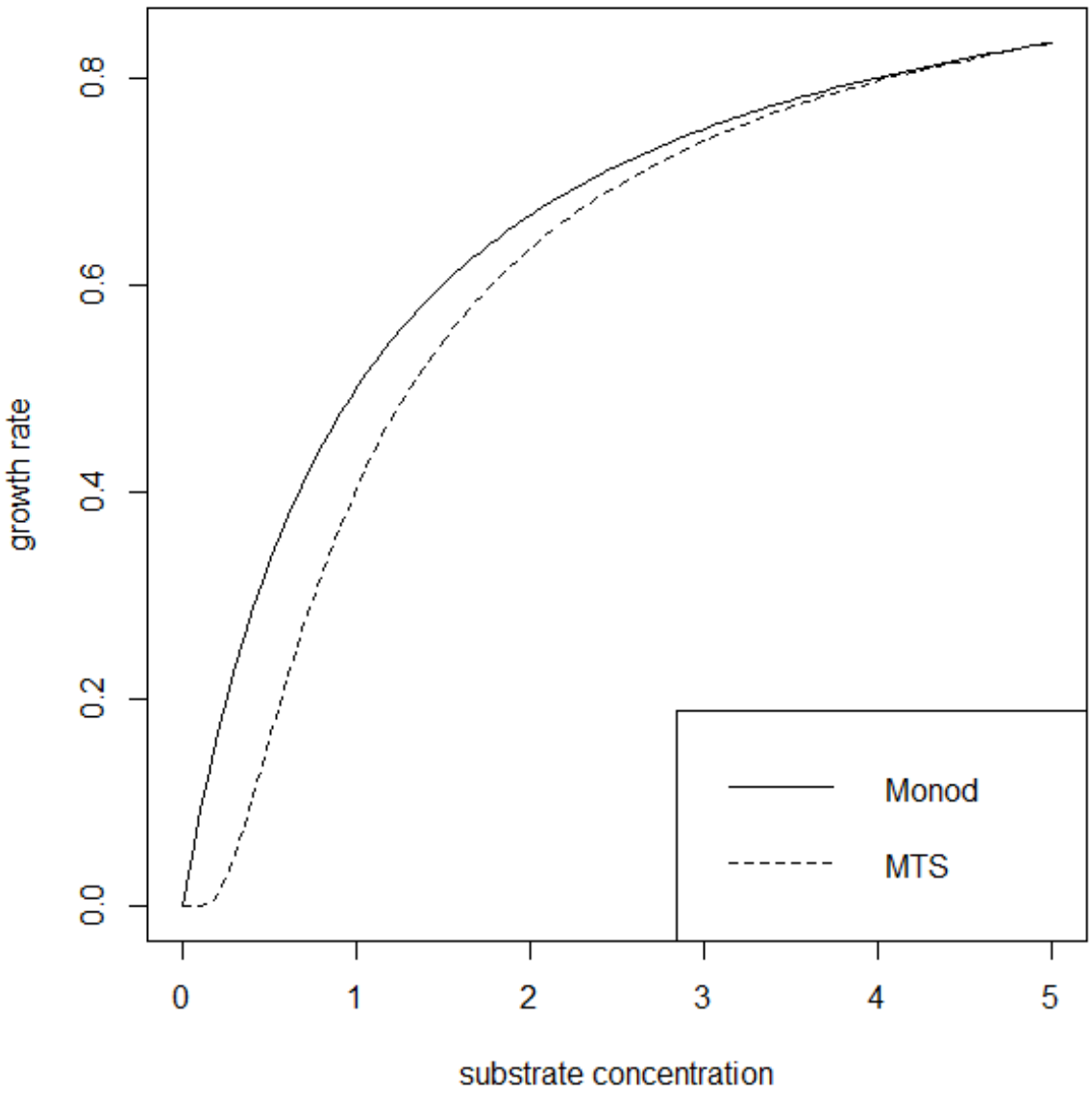
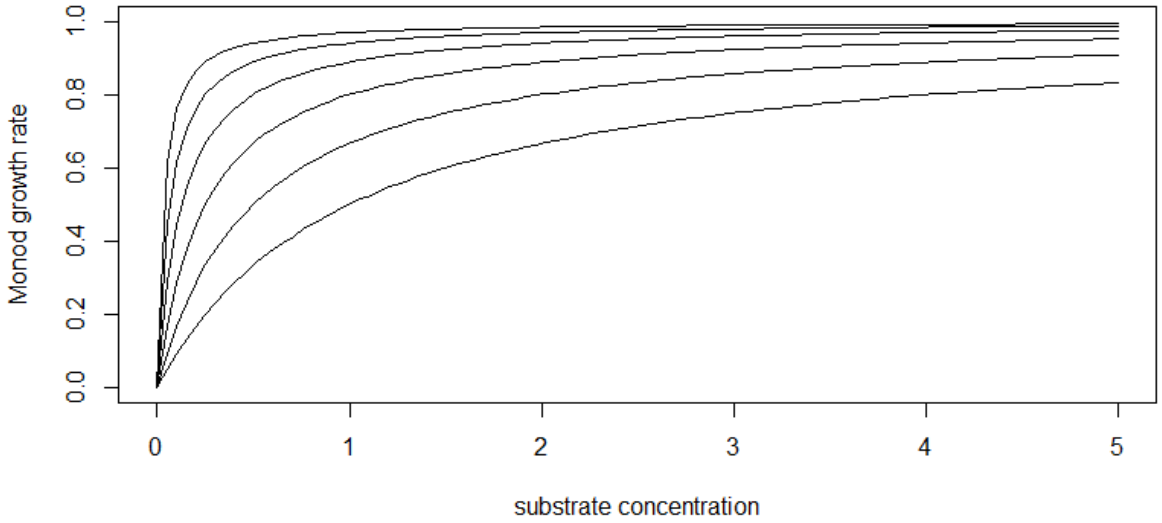


Figure 9.1: comparison between the growth rate function of the MTS model and of Monod's law. The two plotted functions are actually $\frac{[S]}{1+[S]}$ for Monod's law and $exp^{-\frac{1}{1\cdot[S]}}$ for the MTS model

Effect of the value of K_S and V_H on the shape of Monod's and MTS's growth curve respectively

Monod's law and MTS model's growth function are plotted for various values of V_H and K_S. It can be seen that while the MTS growth function's shape differs from this of Monod's law, it's shape changes and it seems to converge toward this of Monod's law as V_H gets higher.



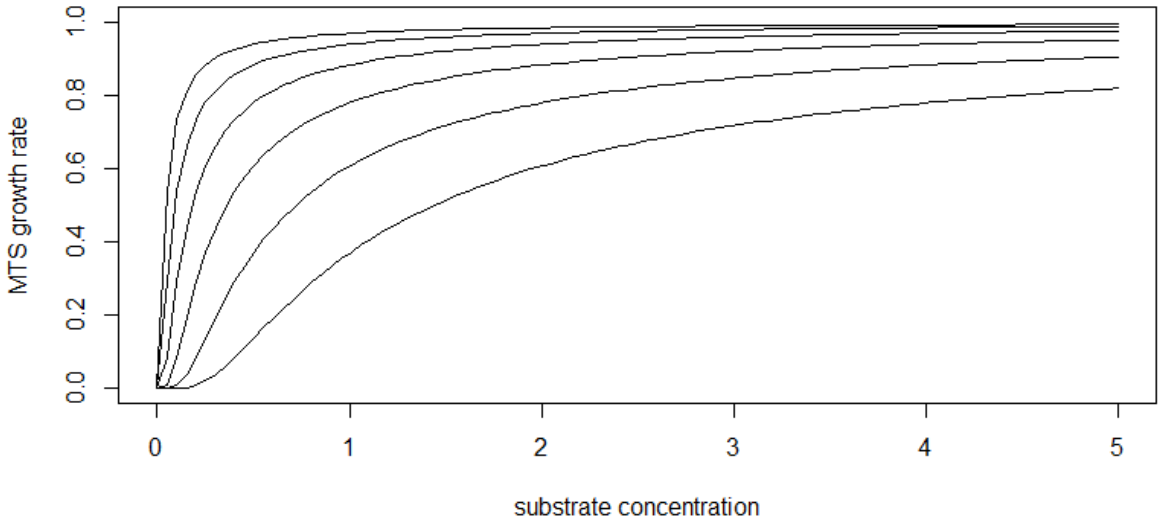


Figure 9.2: Influence of the K_S and V_H parameters, respectively on Monod's growth law and the MTS growth function. The two plotted functions are actually $\frac{[S]}{1/p+[S]}$ for Monod's law and $exp^{-\frac{1}{p \cdot [S]}}$ for the MTS model, with p taking the following values: 1, 2, 4, 8, 16, 32

Difference between Monod's and MTS's growth curves

The substraction of Monod's growth function and the MTS model's growth function shows that both functions converge toward the same value (1) when $[S]$ gets large enough. The difference between the two functions lies at low $[S]$ values.

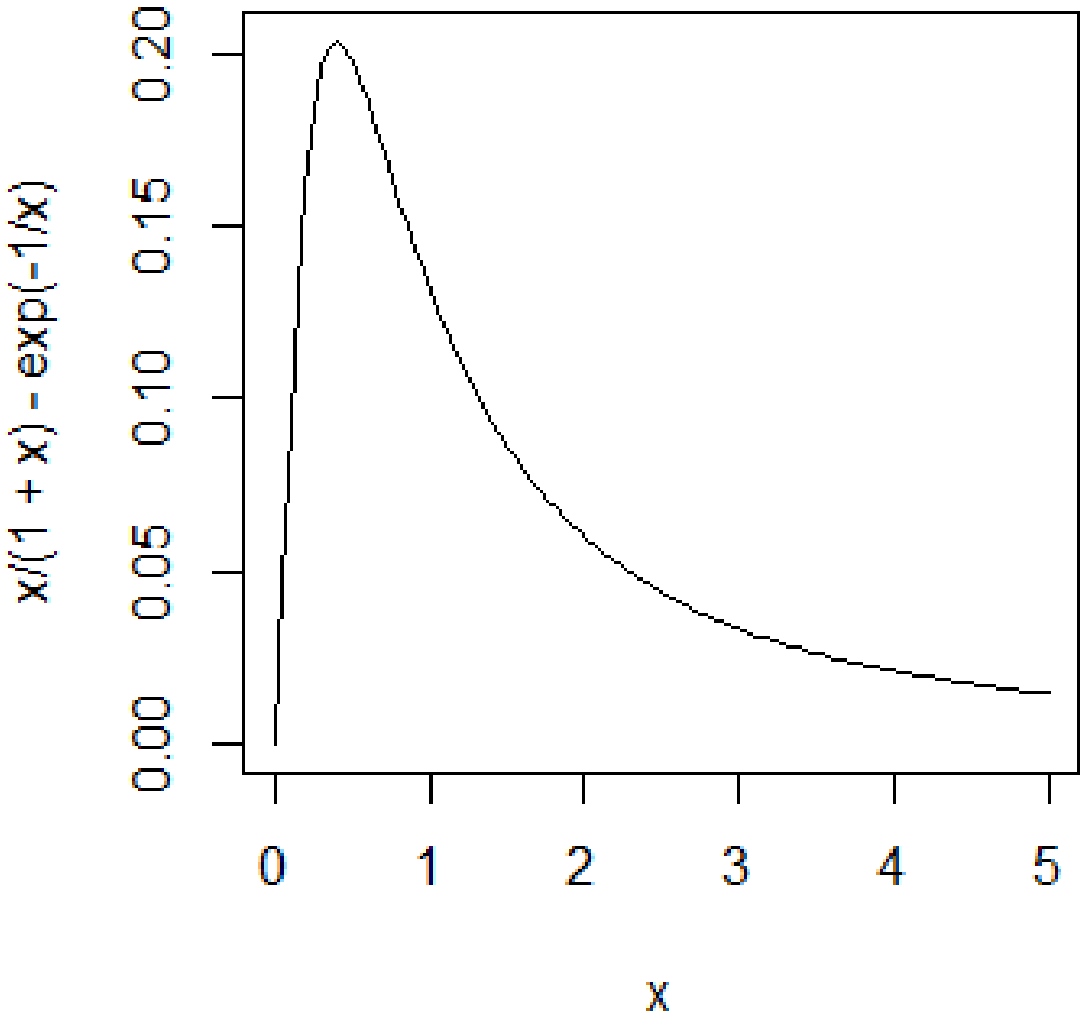


Figure 9.3: difference between the growth function of Monod's growth law and the MTS growth function, depending on substrate concentration. The plotted function is actually $\frac{[S]}{1+[S]} - exp^{-\frac{1}{1\cdot[S]}}$

Comparison between Monod's and MTS's growth curves considering their usual parameters value range

It should be noted that the two parameters K_S and V_H usually do not have the same units, and then not the same values in the respective implementations of the growth law. Indeed, the concentration unit is usually $\text{mol} \cdot \text{L}^{-1}$ when using Monod's law, while it is $\text{mol} \cdot \text{m}^{-3}$ ($1000\,\text{mol} \cdot \text{L}^{-1}$) when using the MTS model in this memoir.

The comparison of Monod's and MTS's growth function in the previous section were performed without μ_{max} and without considering the usual value range of K_S and V_H. In this section, the usual value ranges of the parameters in the comparison between the two growth functions is taken into consideration. To do so, the growth functions of an aerobic acetotroph population simulated using Monod's law and the MTS model are compared.

The specifications of the ASM model (Henze et al., 1987) gives different values for the affinity coefficient on oxygen (K_{OH}) and carbonaceous substrate (K_S). it is said that the value of K_{OH} is $0.1\,\text{gCOD} \cdot \text{m}^{-3}$ and the value of K_S is $20\,\text{gCOD} \cdot \text{m}^{-2}$ for heterotroph populations striving in activated sludge. Considering that acetate is the most abundant volatile fatty acid in activated sludge (Rössle and Pretorius, 2001), the K_S value is assumed to corresponds to $0.31\,\text{mol}_{\text{acetate}} \cdot \text{m}^{-3}$. It is also said that the value of μ_{max} for this population is $6\,\text{day}^{-1}$.

The calibration of the MTS model's parameters V_H and μ_{max} on Oxygen Uptake Rate (OUR) curves from a sample of activated sludge (done in chapter 5 of this memoir) gave the value $147\,\text{m}^3 \cdot \text{mol}^{-1}$ for V_H and $3.75\,\text{day}^{-1}$ for μ_{max}.

In order to properly reproduce the growth function of this population according to the MTS model, it is also necessary to account for the effect of the variation of the concentration of acetate and oxygen on the MTS model's stoichiometry. As the stoichiometry of the growth reaction depends on the Gibbs energy change of the anabolic reaction, the catabolic reaction and dissipation, the following statements are assumed;

- the conditions are standard (temperature is 298 K and the concentrations are 1 M), with exceptions
- the concentration of protons is $1 \times 10^{-4}\,\text{mol} \cdot \text{m}^{-3}$ (pH 7)
- $\Delta G_{an}^{0'} = 23.9e3J.mol_{biomass}^{-1}$

- $\Delta G^{0'}_{cat} = -844.4e3 J.mol^{-1}_{acetate}$
- the dissipated energy is $-432.12 \times 10^3\ \mathrm{J \cdot mol^{-1}_{biomass}}$

The λ factor of the metabolism can then be expressed as

$$\lambda = \frac{-432.12e3 - (23.9e3 - 0.5 \cdot RTln[acetate])}{-844.4e3 - RT(ln[acetate] + 2ln[O_2])} \tag{9.3}$$

When the functional response on acetate is tested, the oxygen concentration is assumed to be the saturation concentration at 298 K ($0.275\ \mathrm{mol \cdot m^{-3}}$).

The stoichiometric coefficienf for acetate is -1 in the catabolic reaction and -0.5 in the anabolic reaction.The relationship between the growth rate predicted by the MTS model and the acetate concentration then becomes;

$$\mu([acetate]) = \mu_{max} \cdot exp^{-\frac{-1 \cdot \lambda - 0.5}{V_H \cdot [acetate]}} \tag{9.4}$$

When the functional response on oxygen is tested, the acetate concentration is assumed to be $1\ \mathrm{mol \cdot m^{-3}}$.

The stoichiometric coefficienf for oxygen is -2 in the catabolic reaction and 0 in the anabolic reaction. The relationship between the growth rate predicted by the MTS model and the oxygen concentration is then;

$$\mu([O_2]) = \mu_{max} \cdot exp^{-\frac{-2 \cdot \lambda}{V_H \cdot [O_2]}} \tag{9.5}$$

Using the values of V_H and μ_{max} calibrated on OUR curves in chapter 5, the comparison of MTS's and Monod's growth curves gives the following graph for the functional responses of MTS and ASM on acetate;

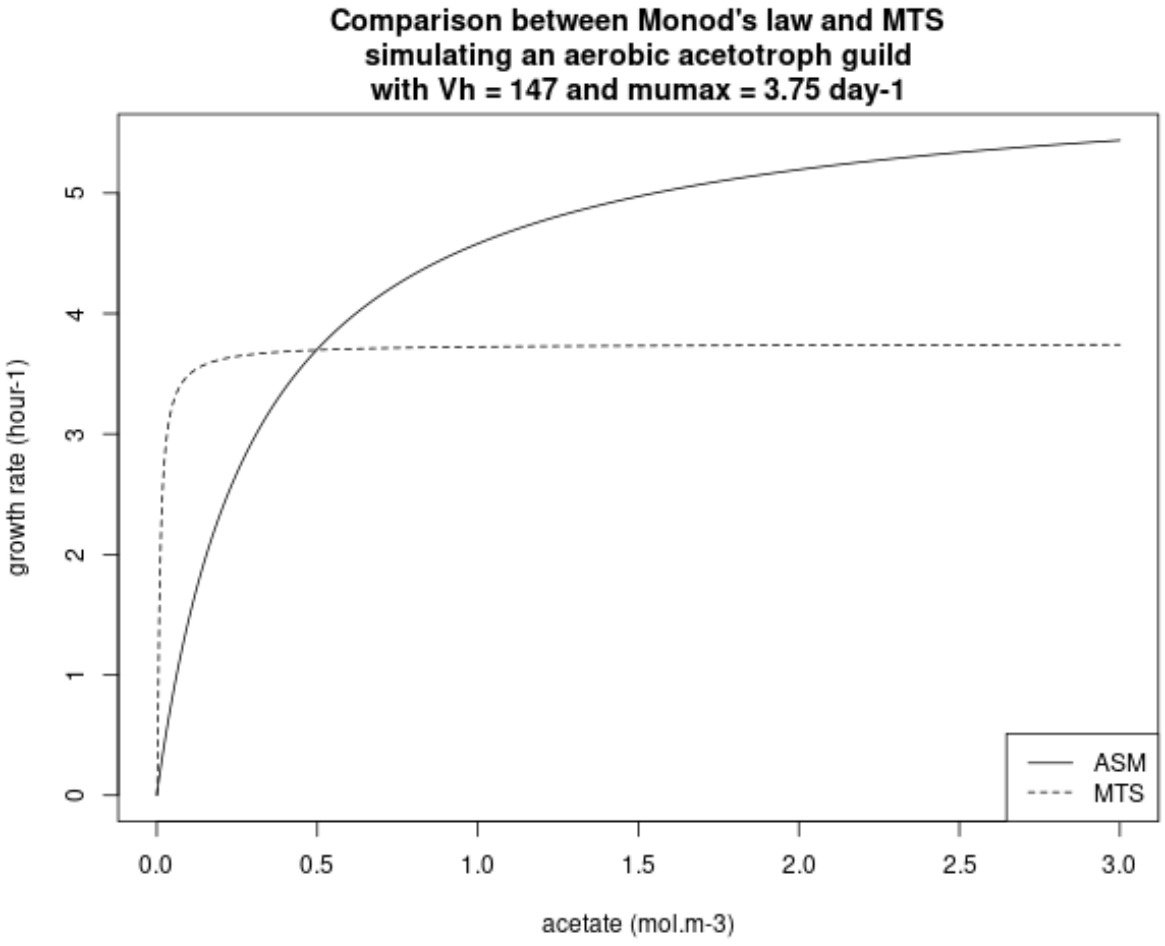


Figure 9.4: superimposition of the Monod's law growth function of the ASM1 model and the calibrated MTS's growth function for an aerobic acetotroph guild, depending on the concentration of acetate

The functional responses between the two models are totally different. A couple of μ_{max} and V_H values making the functional response of the MTS model closer to this of the calibrated Monod's law is $\mu_{max} = 6day^{-1}$ and $V_H = 4m^3.mol^{-1}$;

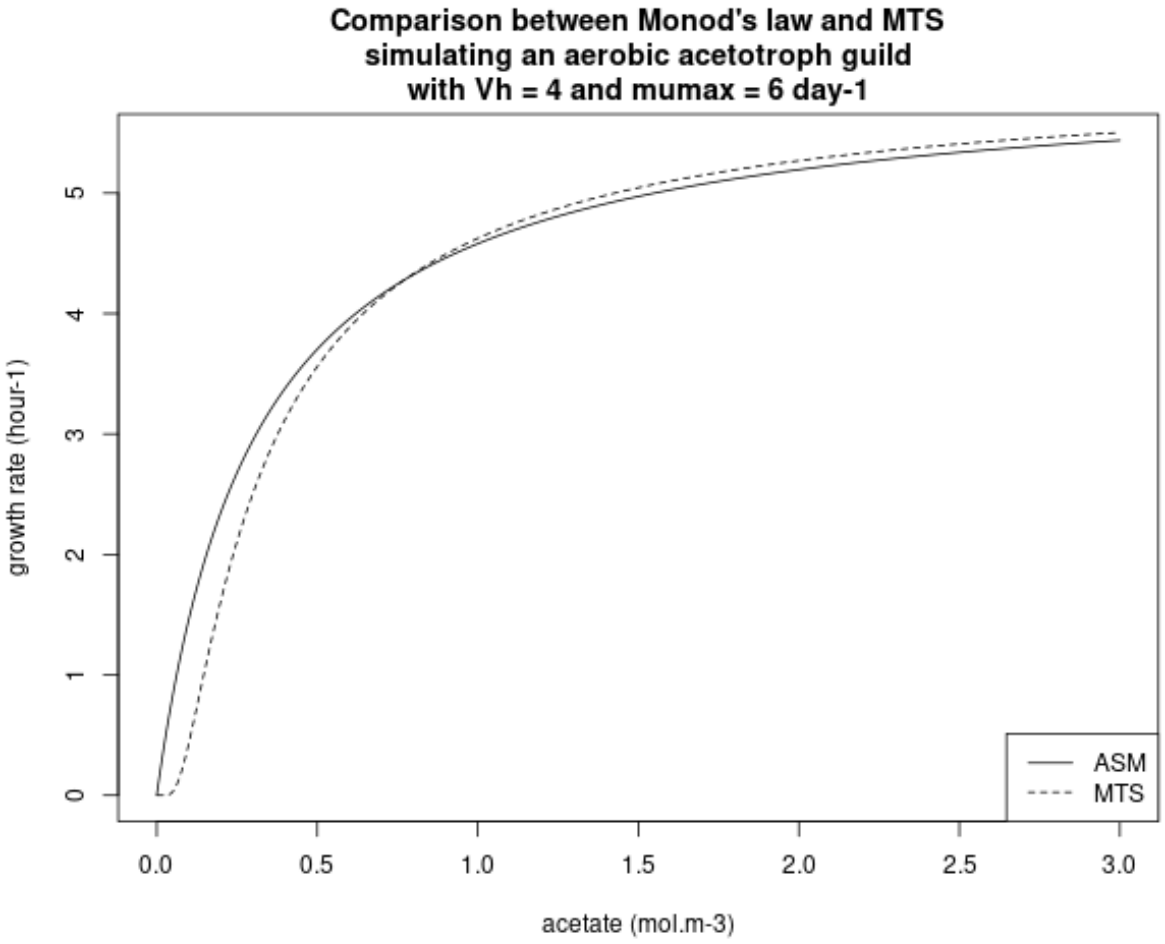


Figure 9.5: superimposition of the Monod's law growth function of the ASM1 model and the MTS's growth function with alternative parameters for an aerobic acetotroph guild, depending on the concentration of acetate

The same way, using the values of V_H and μ_{max} calibrated on OUR curves in chapter 5, the comparison of MTS's and Monod's growth curves gives the following graph for the functional responses of MTS and ASM on oxygen;

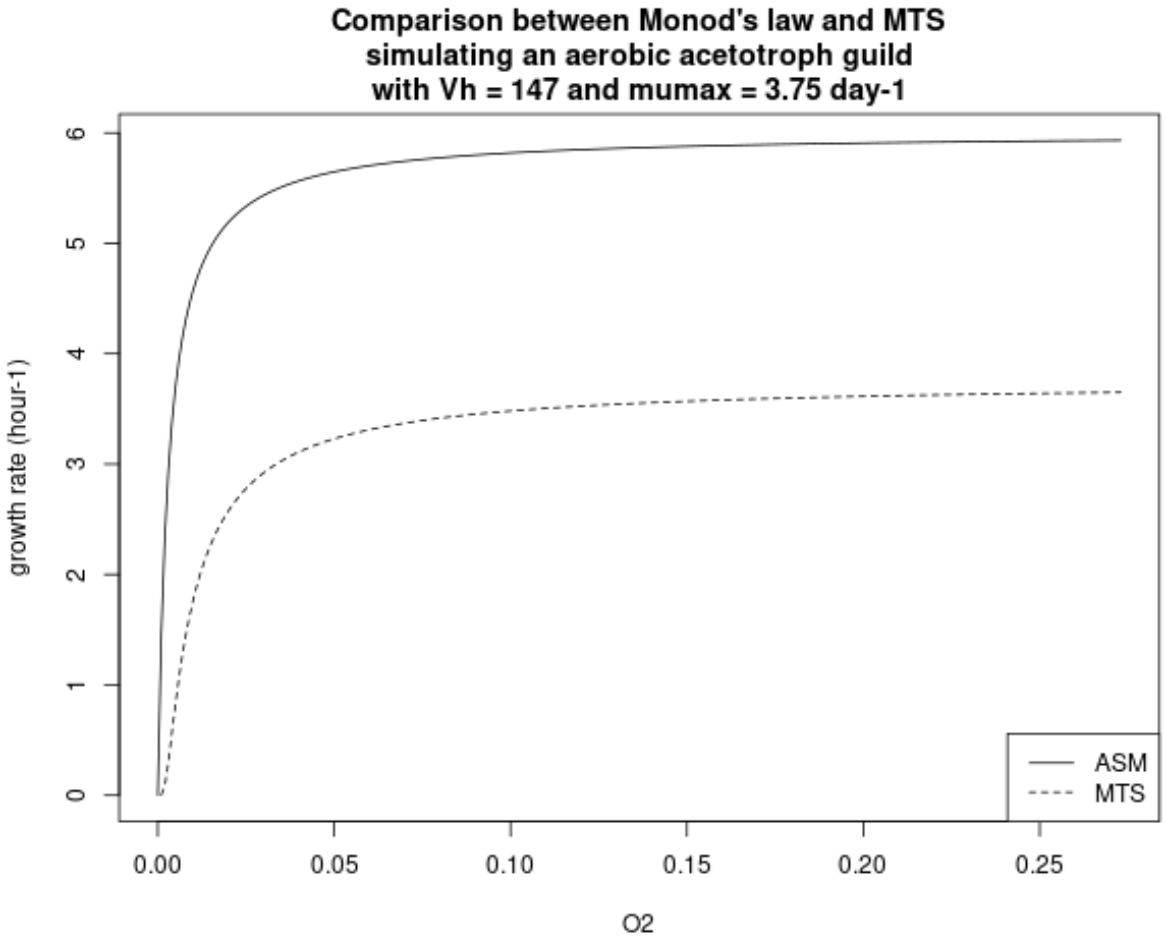


Figure 9.6: superimposition of the Monod's law growth function of the ASM1 model and the calibrated MTS's growth function for an aerobic acetotroph guild, depending on the concentration of oxygen

1505

A couple of μ_{max} and V_H values making the functional response of the MTS model closer to this of the calibrated Monod's law is $\mu_{max} = 6day^{-1}$ and $V_H = 400m^3.mol^{-1}$;

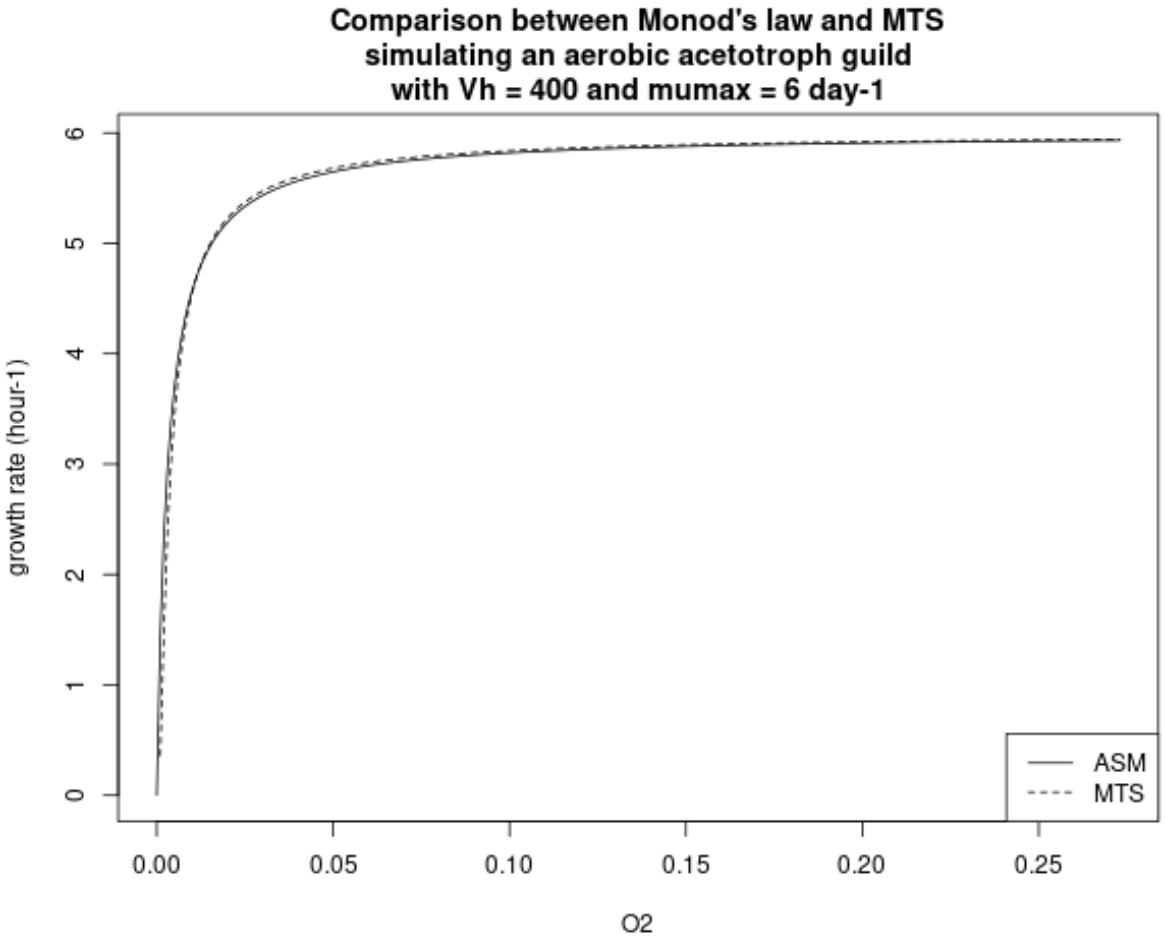


Figure 9.7: superimposition of the Monod's law growth function of the ASM1 model and the MTS's growth function with alternative parameters for an aerobic acetotroph guild, depending on the concentration of oxygen

9.4 Compliance of a population simulated by MTS to Liebig's law

The growth of a denitrifying acetotroph population in a batch has been simulated using the MTS model. Different initial nitrate concentrations have been considered. Depending on the nitrate concentration, the population's growth stops either because of lack of acetate or nitrate. At a narrow range of initial nitrate concentrations, the limitation of growth by both acetate and nitrate has a simular value. In this specific case, the MTS model's predictions does not comply to Liebig's law.

The tuning factors are displayed in each situations. The tuning factor of a chemcial specie is the natural logarithm of the factor applied to the growth rate because of the limitation exerted by the chemical specie on growth.

The initial concentrations are;

- acetate: 1 mM
- biomass: 1 mM

When the initial concentration of nitrate is 0.85 mM the growth is nitrate-limited. When the initial concentration of nitrate is 0.862 mM the growth is limited by both acetate and nitrate. When the initial concentration of nitrate is 0.88 mM the growth is limited by acetate.

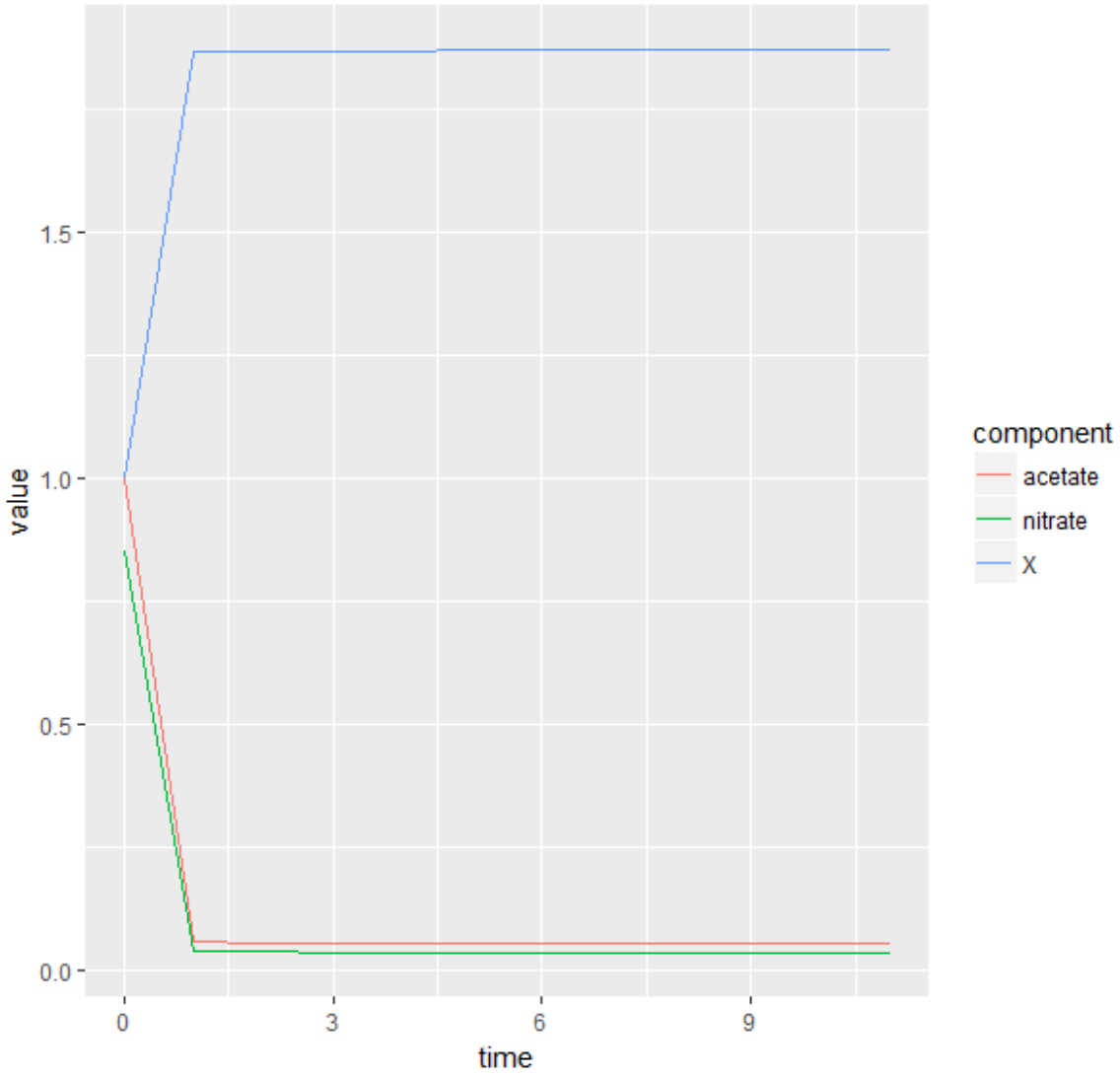


Figure 9.8: Concentrations of acetate, nitrate and biomass (X) (mM) along simulation time (hour) when initial nitrate is 0.85 mM

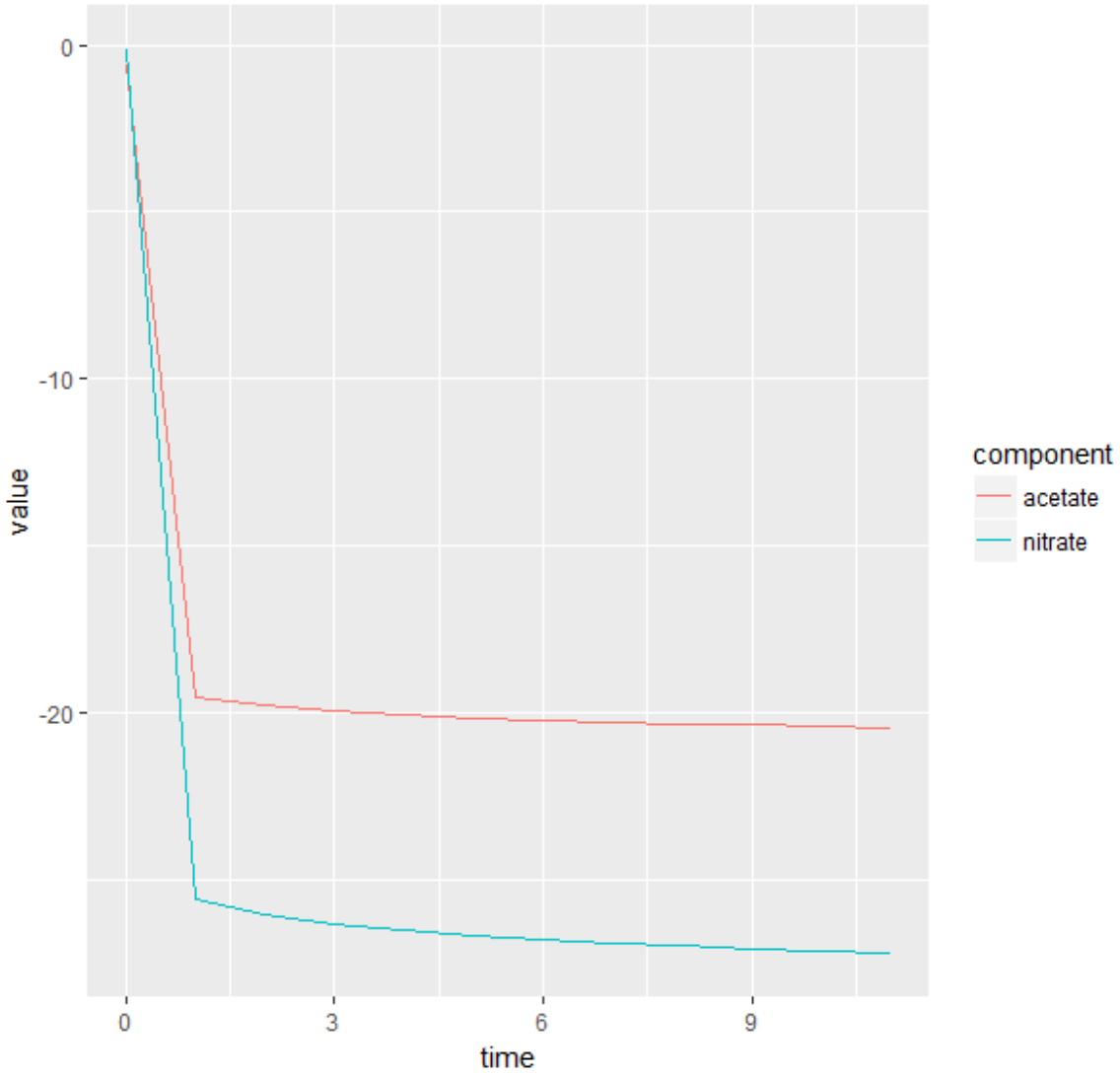


Figure 9.9: tuning factors of acetate and nitrate (unitless) along simulation time (hour) when initial nitrate is 0.85 mM

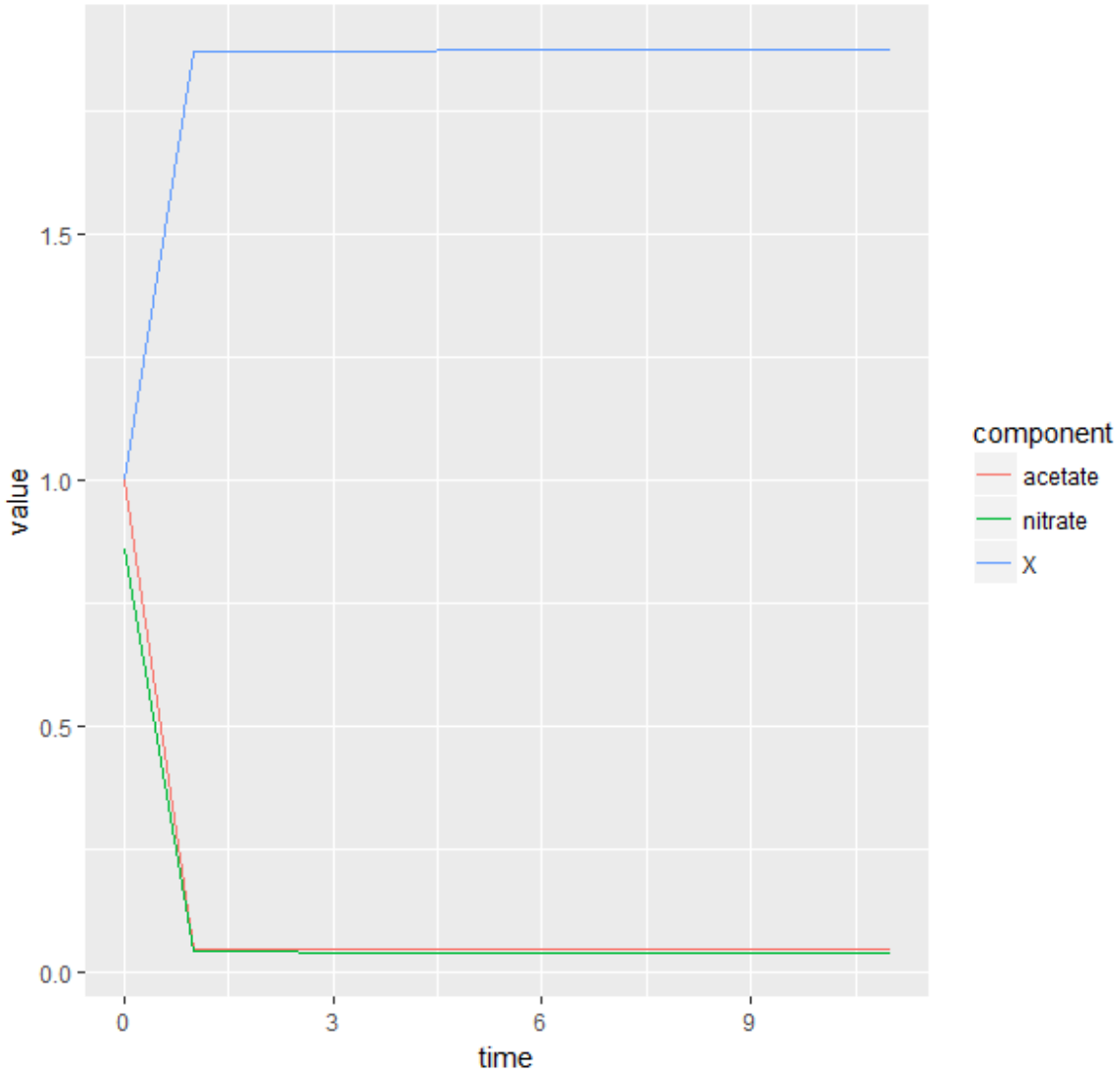


Figure 9.10: Concentrations of acetate, nitrate and biomass (X) (mM) along simulation time (hour) when initial nitrate is 0.862 mM

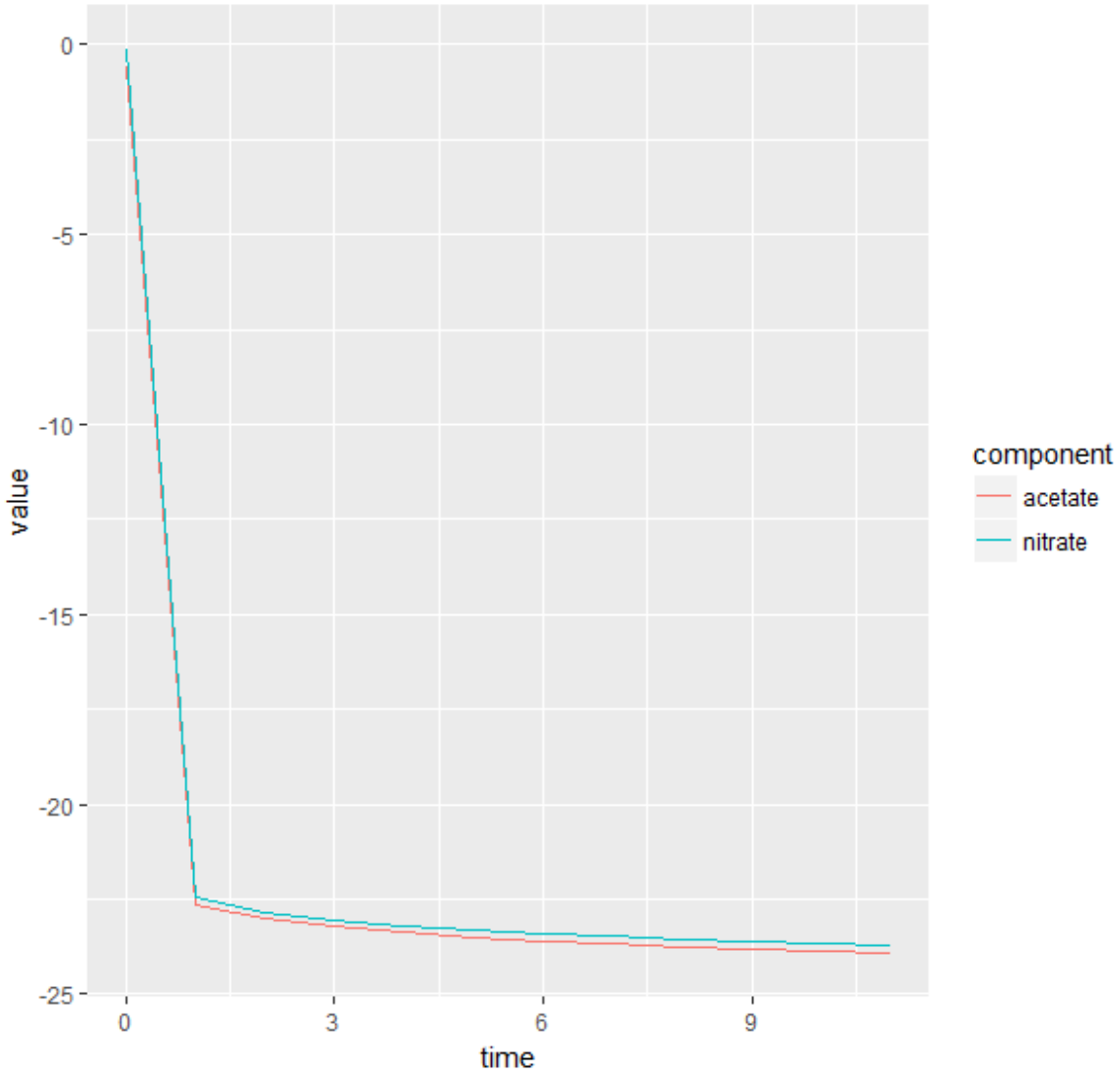


Figure 9.11: tuning factors of acetate and nitrate (unitless) along simulation time (hour) when initial nitrate is 0.862 mM

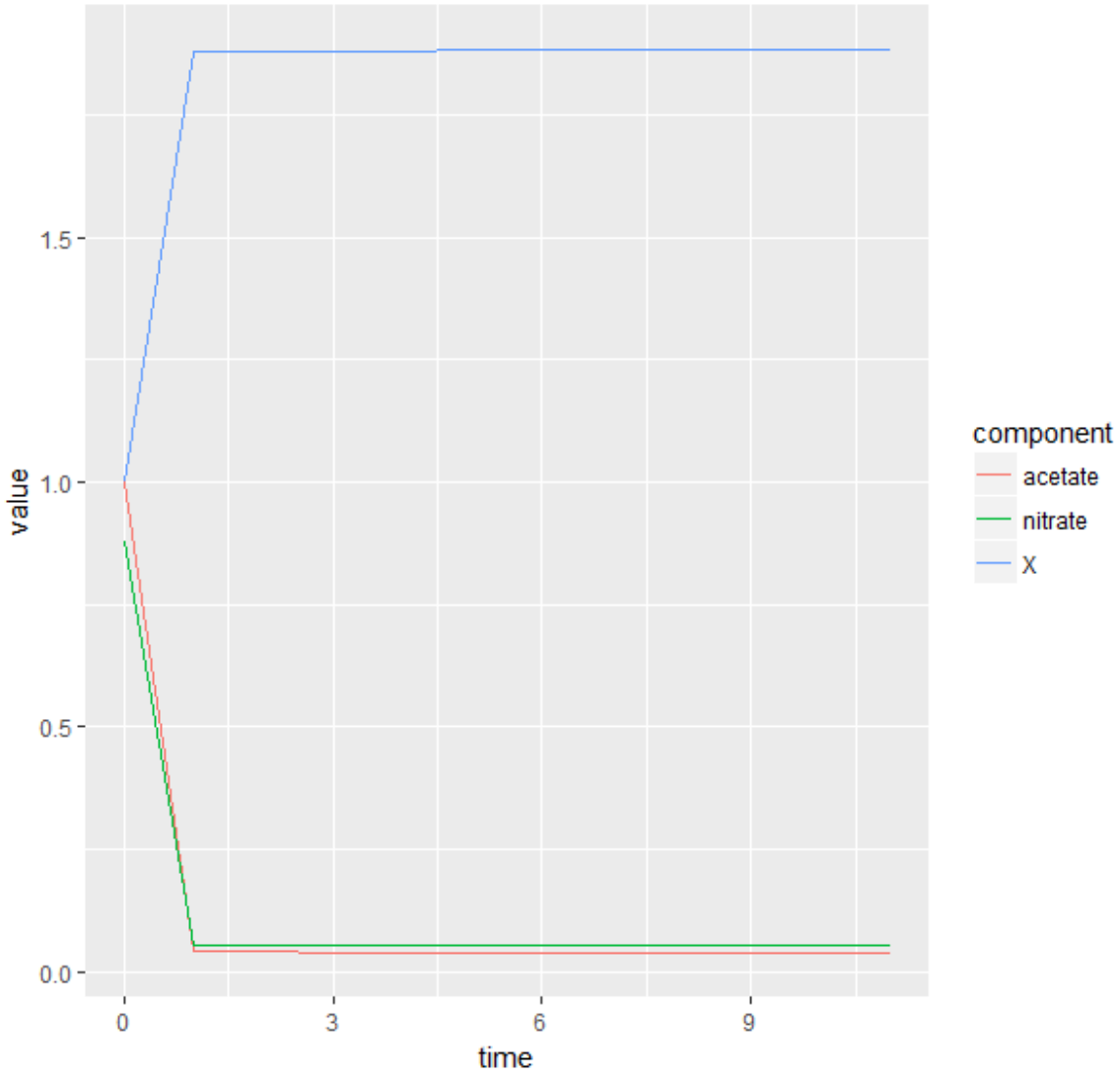


Figure 9.12: Concentrations of acetate, nitrate and biomass (X) (mM) along simulation time (hour) when initial nitrate is 0.88 mM

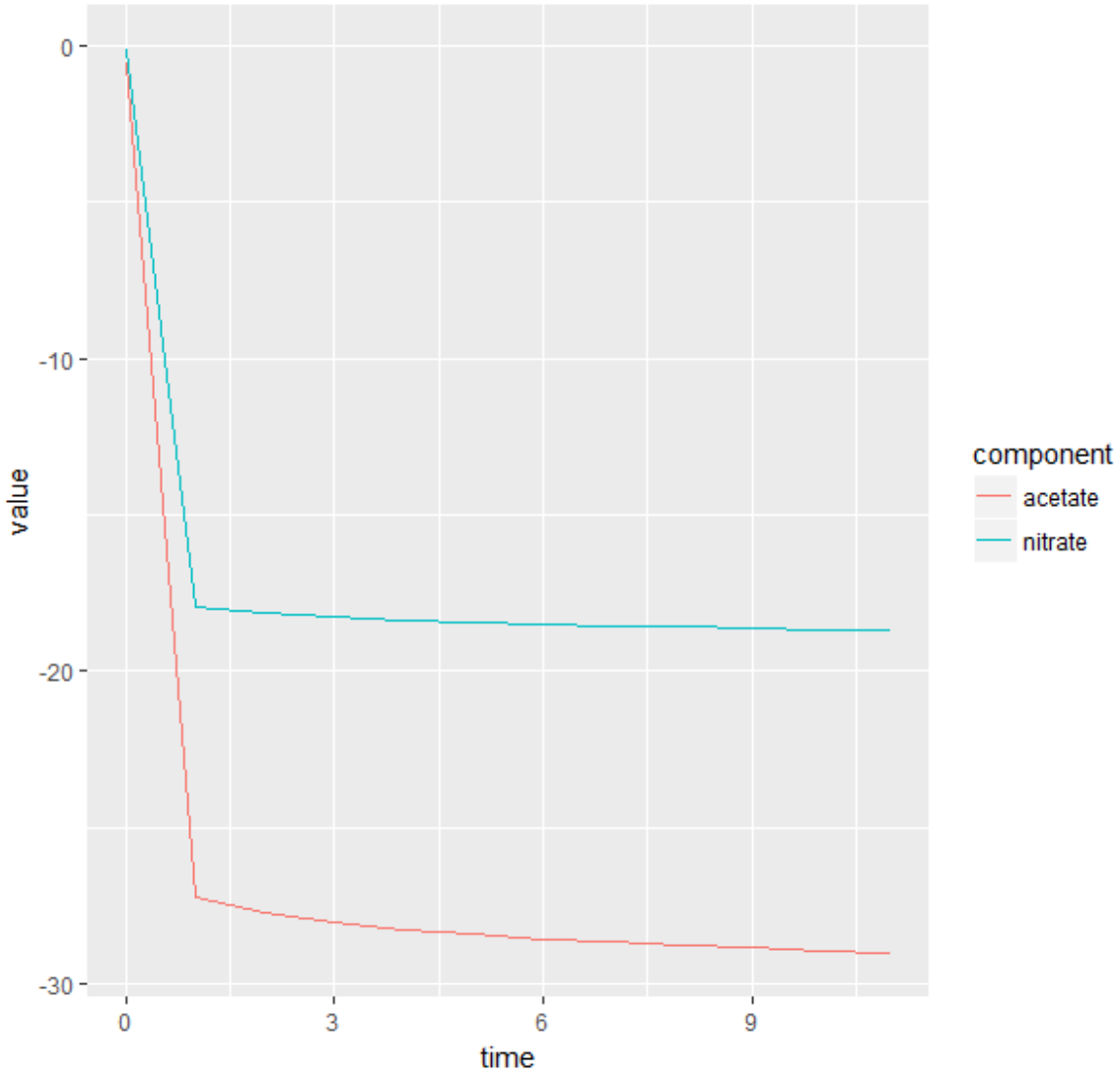


Figure 9.13: tuning factors of acetate and nitrate (unitless) along simulation time (hour) when initial nitrate is 0.88 mM

9.5 Relationship between growth yield and rate in some microbial thermodynamics models

Different microbial thermodynamics models have different, contradictory consequences on the relationship between the microbial growth yield and rate. While all the considered models are reviewed in the bibliography of this memoir (chapter 3), the purpose of this appendix is to precisely point at the relationship between yield and rate implied by each model.

Heijnen's dynamic model

This model was proposed by Heijnen and collaborators as a population growth dynamics model associated to their previously defined metabolism energy balance model (Heijnen et al., 1992) (Heijnen and Kleerebezem, 1999).

The relationship between the growth yield and the growth rate implied by this kinetic model is highlighted by this expression from the model;

$$\mu = Y_{X/S}^{max} \cdot (q_S - m_S) \tag{9.6}$$

where μ is the growth rate of the population ($hour^{-1}$), $Y_{X/S}^{max}$ is the maximum growth yield on the substrate S (the yield in the absence of maintenance) ($mol_X \cdot mol_S^{-1}$), q_S is the consumption rate of the substrate S ($mol_S \cdot hour^{-1}$) and m_S the substrate S absorption specifically required by maintenance ($mol_S \cdot hour^{-1}$).

From 9.6, the growth rate μ is positively correlated to the maximum growth rate $Y_{X/S}^{max}$.

The maximum growth rate is computed as a function of the Gibbs energy change of catabolism ΔG_{cat}, anabolism ΔG_{an} and dissipation ΔG_{dis};

$$Y_{X/S}^{max} = -\frac{1}{\frac{\Delta G_{dis} - \Delta G_{an}}{\Delta G_{cat}} \cdot Y_S^{cat} - Y_S^{an}} \tag{9.7}$$

where Y_S^{cat} is the stoichiometric coefficient of S in the catabolic reaction and Y_S^{an} the stoichiometric coefficient of S in the anabolic reaction. Both coefficients are positive if S is a reagent and negative if S is a product.

From 9.7 it can be seen that $Y_{X/S}^{max}$ is negatively correlated to ΔG_{dis}. Consequently, according to this model, μ is positively correlated to $Y_{X/S}^{max}$ and negatively correlated to ΔG_{dis}.

MTS model

The MTS model (Desmond-Le Quéméner and Bouchez, 2014) is built on top of the metabolic energy balance model defined by Heijnen and collaborators (Heijnen et al., 1992). Consequently, it assumes the negative relationship between $Y_{X/S}^{max}$ and ΔG_{dis} shown by equation 9.7.

However, the MTS model does not take up the kinetic model proposed by Heijnen and collaborators to complement their energy balance model. Instead it proposes its own growth kinetics formula, derived from a theoretical reasoning. A formulation of this growth kinetics formula expliciting the link between $Y_{X/S}^{max}$ and μ is:

$$\mu = \mu_{max} \cdot e^{-\frac{1}{Y_{X/S}^{max} \cdot V_H \cdot [S]}} \tag{9.8}$$

where μ_{max} is the maximum growth rate of the population (time^{-1}), V_H a kinetic parameter specific to the MTS model and $[S]$ the concentration of the substrate S. This formulation of μ assume that S is the only substrate accounted for, that its stoichiometric coefficients are -1 in the catabolic reaction and 0 in the anabolic reaction. Though other cases can be handled by the MTS model without modifying the relationship between $Y_{X/S}^{max}$ and μ, those hypotheses have been followed in order to simplify the expression.

The relationship described by equation 9.8 implies a positive relationship between $Y_{X/S}^{max}$ and μ, like equation 9.6 but with totally different justifications.

Consequently it can be said that according to the MTS model, μ is positively correlated to $Y_{X/S}^{max}$ and negatively correlated to ΔG_{dis}, like in Heijnen's kinetic model.

Jin and Bethke's model

Jin and Bethke's model does not consider the energy balance of microbial growth as defined by Heijnen and collaborators. It does not give a description of anabolism. Instead, it focus on the rate of the catabolic reaction, and gives it a theoretical expression relying on nonlinear nonequilibrium thermodynamics (Jin and Bethke, 2002). The authors assume that some of the energy ΔG_{cat} (kJ $\cdot$ mol$_e^{-1}$) produced by the catabolic reaction is stored as

ATP. The remaining part then constitutes the "drive" f of the catabolic reaction:

$$f = -\Delta G_{cat} - m \cdot \Delta G_P \tag{9.9}$$

where ΔG_P ($\text{kJ} \cdot \text{mol}_{\text{ATP}}^{-1}$) is the phosphorylation energy of ATP and m is the ratio of the number of ATP molecules produced per electrons transfered in the catabolic reaction ($\text{mol}_{\text{ATP}} \cdot \text{mol}_{\text{e}}^{-1}$).

Assuming that ATP is expended to cover the energy requirements of anabolism and maintenance, the drive f can be seen as the value $-\Delta G_{dis}$ from previously introduced models, normalized per electron transfered.

To make a long story short, nonequilibrium thermodynamics implies that the rate of the catabolic reaction is positively correlated to its overall Gibbs energy change, which can be expressed as ΔG_{dis}.

From equation 9.9, it can then be said that the relationship between $Y_{X/S}^{max}$ and μ is negative (assuming that S is the electron donor of the catabolism) since the energy of catabolism is either invested in ATP (then anabolism) or in f.

To summarise, while the model defined by Jin and Bethke does not address yield and dissipated energy as Heijnen's model does, it can be said that it implies a negative relationship between μ and $Y_{X/S}^{max}$, and a positive relationship between μ and ΔG_{dis}, which is the opposite of what Heijnen's model and the MTS model assume.

Supplementary materials of chapter 4

Supplementary material 1: coherence between the notation of MTS's introductory article and the current article

The introductory article of the MTS model (Desmond-Le Quéméner and Bouchez 2014) makes use of units to describe the parameters of the MTS model which differs from those used in the current article. This leads to two apparently different formulations of the growth rate function of the model. This supplementary material explains why those two formulations are consistent.

In the introductory article, the theory was implemented for the analysis of an elementary microbial division event in the simplest case, namely in the case for which the stoichiometric growth equation corresponds to:

M+ S -> 2M + P

M being the microbe, S the substrate, P the metabolic product

Using the multi-substrate formalism detailed in the current article, the catabolic stoichiometric coefficient A_{cat} for S is -1 while the anabolic stoichiometric coefficient A_{an} for S is 0. This lead to the following growth formula;

$$\mu = \mu_{max} e^{-\frac{\lambda}{[S] V_h}}$$

which is the growth formula presented in the introductory article.

Supplementary material 2: effect of the value of Vh on the growth of the AOB guild

The Ammonium Oxydizer Bacteria (AOB) guild and the Nitrite Oxidizer Bacteria (NOB) guild are very sensitive to the oxygen concentration. For this reason, the value of their Vh parameter was set to 10 $m^3.mol_{biomass}^{-1}$ in the article instead of 1 $m^3.mol_{biomass}^{-1}$ as with other guilds simulated in the article. This change modifies the point at which the growth rate of the guild, as a function of oxygen concentration, starts. This is illustration by the following figure;

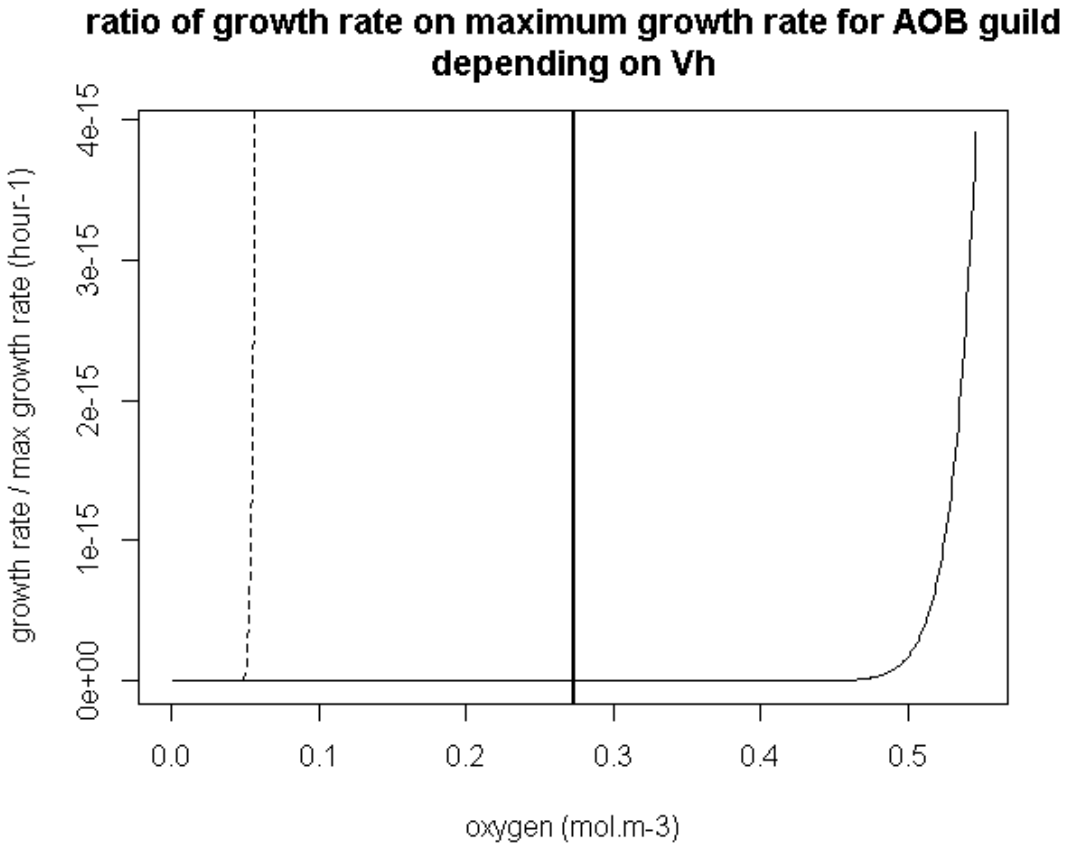


Figure 1 : ratio of growth rate over maximum growth rate for the AOB guild, as a function of oxygen concentration in the culture medium. The plain line represents the case where Vh = 10 m3.mol-1, the dashed line represents the case where Vh = 1 m3.mol-1. The solid vertical line represents oxygen's saturation concentration at 298 K.

Supplementary material 3: estimation of the harvest volume as a sphere around microbial cells

The Harvest Volume (V_h) parameter of the MTS model is expressed in $m^3.mol_{Biomass}^{-1}$, however in the MTS theory it corresponds to a fictional volume surrounding each microbial cells. This supplementary material shows estimation of the radius of the harvest volume around each cell from its value in $m^3.mol_{Biomass}^{-1}$.

This conversion compels us to make some hypotheses;

- All the cells of the population have a volume of 1 μm^3 (1e-18 m^3)
- This volume contains 141e-15 g carbon (geometric mean of the interval 100-200 fg given by Milo and Philips for a cell volume of 1 μm^3 (Milo and Phillips 2015)
- The cells are exclusively made of Battley's biomass molecules, whose molecular weight is 24.68 $g.mol^{-1}$ and whose carbon proportion is 0.48 $gC.gBiomass^{-1}$
- Consequently there is 141e-15 / 0.48 / 24.68 = 1.19e-14 mol of biomass per cell
- V_h * 1.19e-14 then is the harvest volume in m^3 per cell.

The value of 1 $m^3.mol_{biomass}^{-1}$ used in the article then corresponds to 1.19e-14 $m^3.cell^{-1}$ = 1.19e4 $\mu m^3.cell^{-1}$. Considering that the harvest volume is a sphere, its radius is then $r = \sqrt[3]{\frac{3\,V_h}{4\pi}} =$

$14.16\ \mu m.cell^{-1}$. Using the same calculation, a V_h of 10 $m^3.mol_{biomass}^{-1}$ corresponds to a radius of approximately 30 $\mu m^3.cell^{-1}$.

Supplementary material 4: growth kinetics of competing guilds in a chemostat

The figure below represents the growth kinetics of four competing guilds in a chemostat. The setting is the same as for the figure 2 of the article. The four guilds use acetate as electron donor; the input concentration of acetate is 1.76 mM.

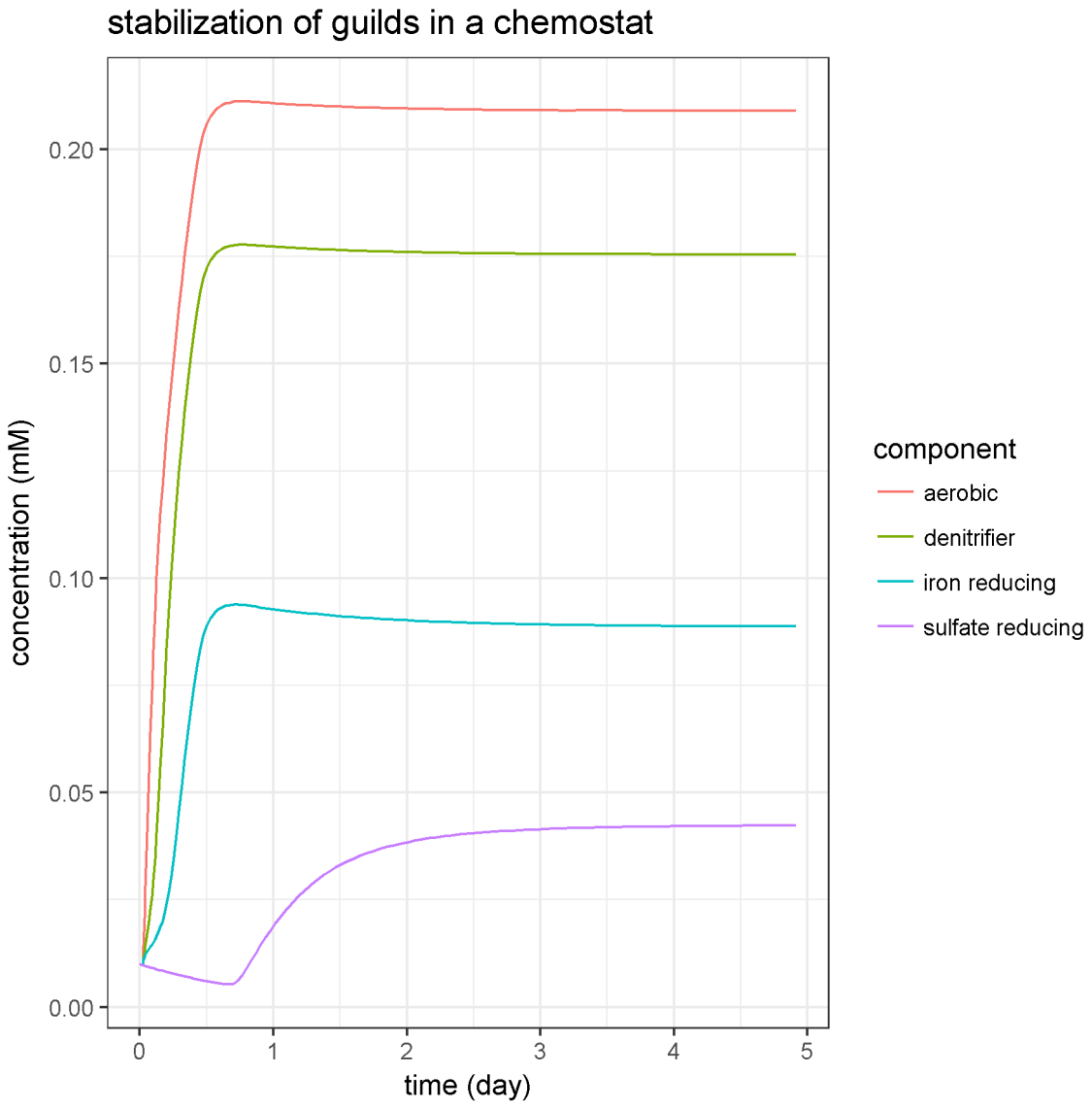


Figure 2: growth kinetics of four guilds in a chemostat, in the simulation setting defined for the figure 2 of the article, with an input acetate concentration of 1.76 mM.

Supplementary material 5: comparison of results obtained with the Activated Sludge Model and the MTS Model

		X_{bh}	X_{aob}	X_{nob}
ASM	Time before stationary state is reached (h)	0.2	0.2	0.2
	Biomass concentration at stationary state (mM)	2.67	1.19	1.06

Harvesting	Time before stationary state is reached (h)	2	15	24
	Biomass concentration at stationary state (mM)	3.57	1.64	1.18

A modified version of the Activated Sludge Model 1 (ASM1) has been implemented (see below) and its predictions in terms of biomass yields and kinetics has been compared with those of the MTS model presented in the fourth part of the results. Hydrolysis and decay reactions normally included in ASM1 have been removed in this implementation as they represent phenomena different from growth and modify the nature of the stationary state (the existence of a decay reaction implies that there is no more biomass when the system reaches its stationary state).

This comparison shows that the biomass yields predicted by the MTS model without parameter calibration are close to those predicted by the ASM, which are the result of years of measurements on real life systems. The difference between the two predictions lays in the kinetics; while every guild in the ASM grows simultaneously, the growth predicted by the Harvesting model is slower and sequential, showing a temporal ecological succession which is not predicted by the ASM.

Description of the modified ASM1 model

A mathematical model modified from the well-established Activated Sludge Model no. 1 (Henze et al 2000) was used for modelling the batch experiment with a simplified activated sludge community.

The original model was extended with a two-step nitrification and simplified so as to describe only the aerobic growth of heterotrophs (X_{BH}), ammonia-oxidizing bacteria (X_{AOB}) and the nitrite-oxidizing bacteria (X_{NOB}).

The description of processes and the basic stoichiometry between model components of the modified ASM 1 model are presented in Table S1 in the usual matrix format. Tables S2 and S3 present the kinetic rate expression and the parameters value of the model, respectively. The simulation software used is WEST (MIKEbyDHI.com).

Table S1. Matrix representation of the modified ASM1 model

Process		S_S	X_S	X_{BH}	X_{AOB}	X_{NOB}	X_P	S_O	S_{NO_2}	S_{NO_3}	S_{NH}	S_{ND}	X_{ND}	S_{ALK}
1	Aerobic growth of heterotrophs	$-\frac{1}{Y_H}$		1				$-\left(\frac{1-Y_H}{Y_H}\right)$			-i_{XB}			$-\frac{i_{XB}}{14}$
2	Aerobic growth of AOB				1			$-\left(\frac{3.43-Y_{AOB}}{Y_{AOB}}\right)$	$\frac{1}{Y_{AOB}}$		$-i_{XB}-\frac{1}{Y_{AOB}}$			$-\frac{i_{XB}}{14}-\frac{2}{14\times Y_{AOB}}$
3	Aerobic growth of NOB					1		$-\left(\frac{1.14-Y_{NOB}}{Y_{NOB}}\right)$	$-\frac{1}{Y_{NOB}}$	$\frac{1}{Y_{NOB}}$	-i_{XB}			$-\frac{i_{XB}}{14}$

Table S2. Kinetics of the modified ASM1 model

Process	Kinetic rate expression
1	$\hat{\mu}_{\mathrm{H}} \times \frac{S_S}{K_S+S_S} \times \frac{S_O}{K_{O,H}+S_O} \times \frac{S_{NH}}{K_{NH,H}+S_{NH}} \times X_{BH}$
2	$\hat{\mu}_{\mathrm{AOB}} \times \frac{S_{NH}}{K_{NH}+S_{NH}} \times \frac{S_O}{K_{O,AOB}+S_O} \times X_{AOB}$
3	$\hat{\mu}_{\mathrm{NOB}} \times \frac{S_{NO_2}}{K_{NO_2}+S_{NO_2}} \times \frac{S_O}{K_{O,NOB}+S_O} \times X_{NOB}$

Table S3. Parameters value of the modified ASM1

Name	Description	Unit	Value	Source
$\hat{\mu}_{\mathrm{H}}$	Maximum specific growth rate, heterotrophs	day^{-1}	6.25^{*} ($\theta = 1.072$)	ASMN (Hiatt and Grady 2008)
Y_H	Heterotrophic yield	gCOD g^{-1} COD	0.6	ASMN
K_S	Half-saturation coefficient for S_S, heterotrophs	gCOD m^{-3}	20	ASMN
$K_{O,H}$	Half-saturation coefficient for S_O, heterotrophs	gCOD m^{-3}	0.1	ASMN
$K_{NH,H}$	Half-saturation coefficient for S_{NH}, heterotrophs	gN m^{-3}	0.05	(Hauduc et al 2010)
$\hat{\mu}_{\mathrm{AOB}}$	Maximum specific growth rate, AOB	day^{-1}	0.78^{*} ($\theta = 1.103$)	ASMN
Y_{AOB}	Autotrophic yield, AOB	gCOD g^{-1} COD	0.18	ASMN
K_{NH}	Half-saturation coefficient for S_{NH}, AOB	gCOD m^{-3}	1.31^{**}	ASMN
$K_{O,AOB}$	Half-saturation coefficient for S_O, AOB	gCOD m^{-3}	0.6	ASMN
$\hat{\mu}_{\mathrm{NOB}}$	Maximum specific growth rate, NOB	day^{-1}	0.78^{*} ($\theta = 1.103$)	ASMN
Y_{NOB}	Autotrophic yield, NOB	gCOD g^{-1} COD	0.06	ASMN
K_{NO_2}	Half-saturation coefficient for S_{NO_2}, NOB	gCOD m^{-3}	0.45^{***}	ASMN
$K_{O,NOB}$	Half saturation parameter for S_O, NOB	gCOD m^{-3}	1.2	ASMN

* Value at 20°C, considering Arrhenius equation with the corresponding **θ** value
** Estimated based on half-saturation coefficient for free ammonia considering a temperature of 25°C and a pH of 7
*** Estimated based on half-saturation coefficient for free nitrous acid considering a temperature of 25°C and a pH of 7

Supplementary material 6: Activated sludge simulation with alternative parameters

The MTS model has been used in this article to simulate a simplified activated sludge microbial community. The following values have been used for the parameters of the guilds implemented in the MTS model;

- the μ_{max} of every guild was set to 2.23e16 $hour^{-1}$

- the V_h of every guild was set to 10 $m^3.C\text{-}mol_{biomass}^{-1}$

The simulation's results showed qualitative similarities with a simulation of the same microbial guilds using calibrated growth equations from the ASMN model. This raises questions about the possibility to adjust the kinetic parameters of the MTS model in order to make it more quantitatively predictive.

While the calibration of the MTS model's kinetic parameters on experimental data is obviously out of the scope of this article, this supplementary material illustrates the possibility to make the model's prediction closer to observations by adjusting the value of its kinetic parameters.

In the example presented here, another MTS simulation of the simplified activated sludge system is presented with alternative parameters;

- the μ_{max} of every guild was kept 2.23e16 $hour^{-1}$

- the V_h of OHO guild was set to 1 $m^3.C\text{-}mol_{biomass}^{-1}$

- the V_h of AOB guild was set to 10 $m^3.C\text{-}mol_{biomass}^{-1}$

- the V_h of NOB guild was set to 50 $m^3.C\text{-}mol_{biomass}^{-1}$

The results of the simulation are displayed on figure 3 and 4.

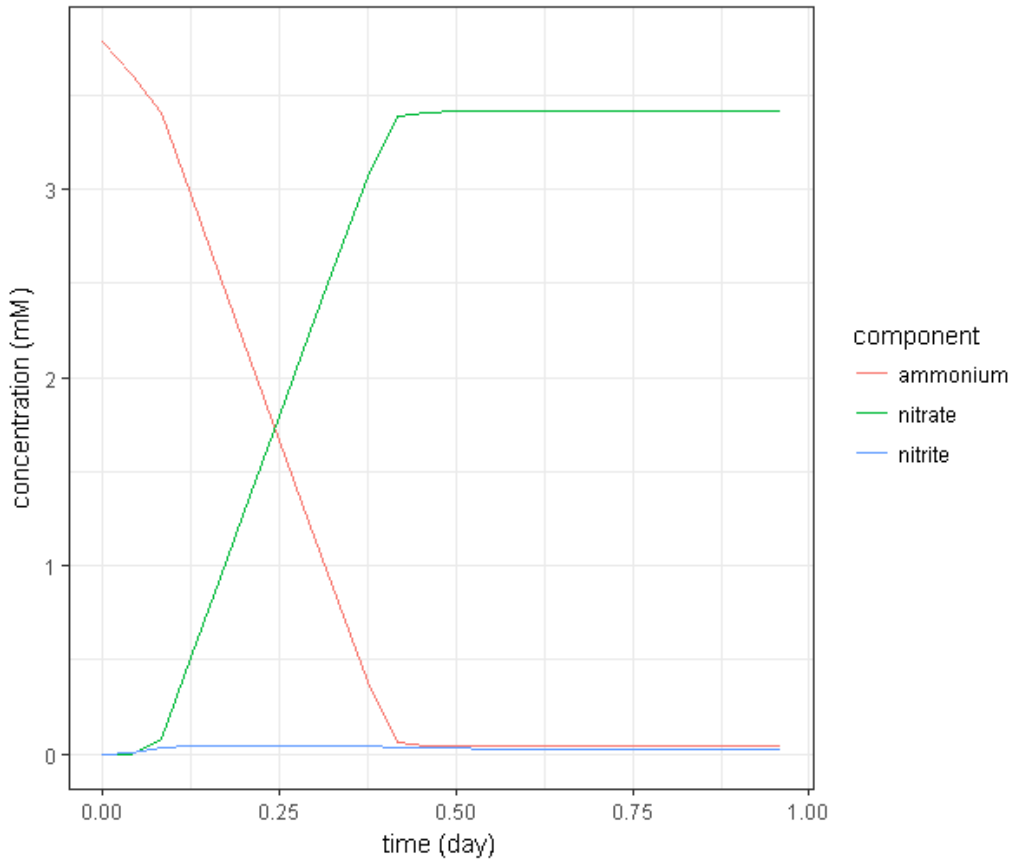


Figure 3: concentration (mM) of different forms of nitrogen along simulation time (day)

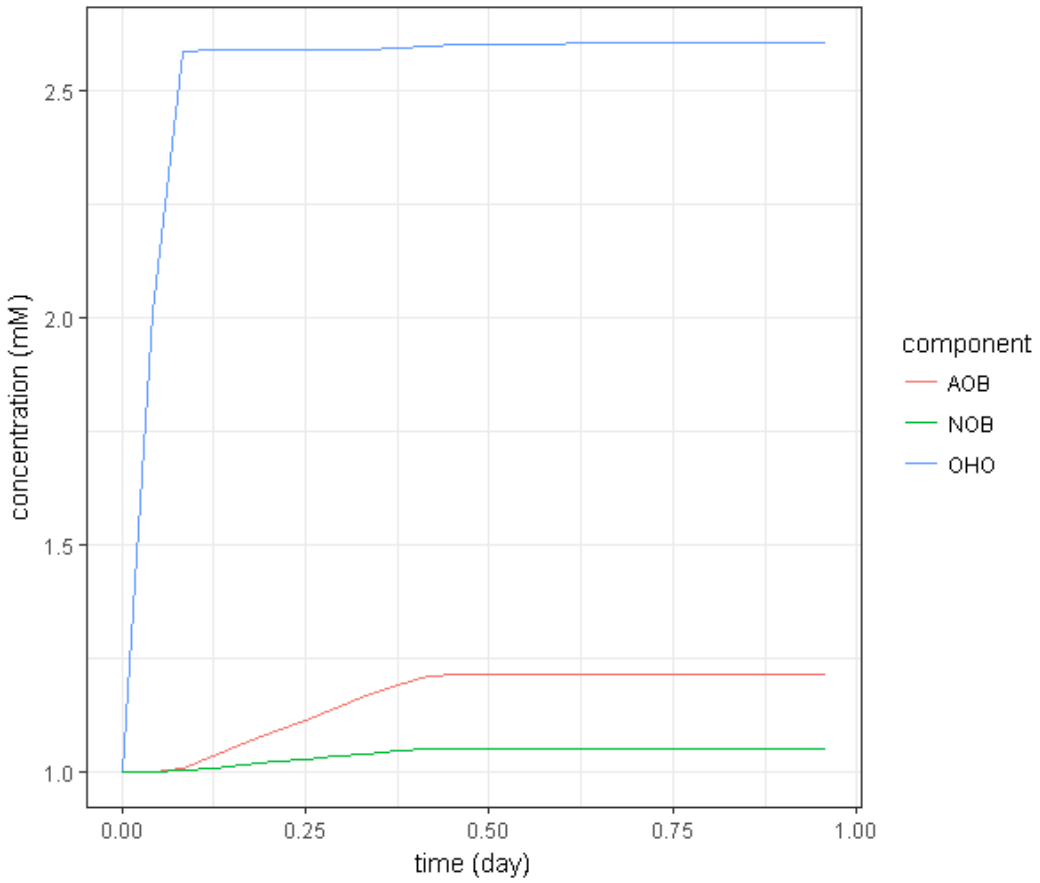


Figure 4: concentration (mM) of the simulated guilds along simulation time (day)

An important observation that can be drawn from figure 3 is that the transient high nitrite peak observed in the simulation performed in the main article has such low amplitude here that it can be considered as not being reproduced in the current simulation. This difference is caused by the larger harvest volume of the NOB guild. While the NOB guild suffers the most severe energy and substrate limitation (compared to the OHO and AOB guilds), here their large V_h partly compensate for this disadvantage. Consequently, the NOB guild does not have to wait nitrite to accumulate in the culture medium before it starts to significantly grow. Instead, the NOB guild's growth follows the AOB guild's growth, by consuming the nitrite once it is produced. Consequently, as seen on figure 4, the growth of the AOB and NOB guilds is almost simultaneous.

As argued in the body of the article, while experimental reports of transient nitrite accumulation in activated sludge systems exist (Rajagopal et al 2011), it does not correspond to the expected behavior of a nominal activated sludge system. The alternative parameters proposed in this supplementary material section make this behavior disappear; therefore it suggests that the predictive accuracy of the MTS model can be improved by adjusting the values of its kinetic parameters.

Supplementary materials of chapter 5

Parameters of the ASMN model

The description of processes and the basic stoichiometry between model components of the modified ASM 1 model are presented in Table S1 in the usual matrix format. Tables S2 and S3 present the kinetic rate expression and the parameters value of the model, respectively.

Table S1. Matrix representation of the modified ASM1 model

Process		S_S	X_S	X_{BH}	X_{AOB}	X_{NOB}	X_P	S_O	S_{NO_2}	S_{NO_3}	S_{NH}	S_{ND}	X_{ND}	S_{ALK}
1	Aerobic growth of heterotrophs	$-\frac{1}{Y_H}$		1				$-\left(\frac{1-Y_H}{Y_H}\right)$			$-i_{XB}$			$-\frac{i_{XB}}{14}$
2	Aerobic growth of AOB				1			$-\left(\frac{3.43-Y_{AOB}}{Y_{AOB}}\right)$	$\frac{1}{Y_{AOB}}$		$-i_{XB}-\frac{1}{Y_{AOB}}$			$-\frac{i_{XB}}{14}-\frac{2}{14\times Y_{AOB}}$
3	Aerobic growth of NOB					1		$-\left(\frac{1.14-Y_{NOB}}{Y_{NOB}}\right)$	$-\frac{1}{Y_{NOB}}$	$\frac{1}{Y_{NOB}}$	$-i_{XB}$			$-\frac{i_{XB}}{14}$

Table S2. Kinetics of the modified ASM1 model

Process	Kinetic rate expression
1	$\hat{\mu}_{\mathrm{H}} \times \frac{S_S}{K_S+S_S} \times \frac{S_O}{K_{O,H}+S_O} \times \frac{S_{NH}}{K_{NH,H}+S_{NH}} \times X_{BH}$
2	$\hat{\mu}_{\mathrm{AOB}} \times \frac{S_{NH}}{K_{NH}+S_{NH}} \times \frac{S_O}{K_{O,AOB}+S_O} \times X_{AOB}$
3	$\hat{\mu}_{\mathrm{NOB}} \times \frac{S_{NO_2}}{K_{NO_2}+S_{NO_2}} \times \frac{S_O}{K_{O,NOB}+S_O} \times X_{NOB}$

Table S3. Parameters value of the modified ASM1

Name	Description	Unit	Value	Source
$\hat{\mu}_{\mathrm{H}}$	Maximum specific growth rate, heterotrophs	day^{-1}	6.25* ($\theta = 1.072$)	ASMN (Hiatt and Grady 2008)
Y_H	Heterotrophic yield	gCOD g^{-1} COD	0.6	ASMN
K_S	Half-saturation coefficient for S_S, heterotrophs	gCOD m^{-3}	20	ASMN
$K_{O,H}$	Half-saturation coefficient for S_O, heterotrophs	gCOD m^{-3}	0.1	ASMN
$K_{NH,H}$	Half-saturation coefficient for S_{NH}, heterotrophs	gN m^{-3}	0.05	(Hauduc et al. 2010)
$\hat{\mu}_{\mathrm{AOB}}$	Maximum specific growth rate, AOB	day^{-1}	0.78* ($\theta = 1.103$)	ASMN
Y_{AOB}	Autotrophic yield, AOB	gCOD g^{-1} COD	0.18	ASMN
K_{NH}	Half-saturation coefficient for S_{NH}, AOB	gCOD m^{-3}	1.31**	ASMN
$K_{O,AOB}$	Half-saturation coefficient for S_O, AOB	gCOD m^{-3}	0.6	ASMN
$\hat{\mu}_{\mathrm{NOB}}$	Maximum specific growth rate, NOB	day^{-1}	0.78* ($\theta = 1.103$)	ASMN
Y_{NOB}	Autotrophic yield, NOB	gCOD g^{-1} COD	0.06	ASMN
K_{NO_2}	Half-saturation coefficient for S_{NO_2}, NOB	gCOD m^{-3}	0.45***	ASMN
$K_{O,NOB}$	Half saturation parameter for S_O, NOB	gCOD m^{-3}	1.2	ASMN

* Value at 20°C, considering Arrhenius equation with the corresponding **θ** value
** Estimated based on half-saturation coefficient for free ammonia considering a temperature of 25°C and a pH of 7
*** Estimated based on half-saturation coefficient for free nitrous acid considering a temperature of 25°C and a pH of 7

Constancy of the yield on oxygen

The kinetic parameters of the MTS model are calibrated using data assuming a constant yield. However the yield is dynamic in the MTS model. Does this difference disqualifies the reference data?

First, it does not disqualifies the reference data by lack of realism (argument A1), in the sense that the reference data comes from experimental data, transformed according to the hypothesis of a constant growth yield which is a commonly used simplifying hypothesis in population microbiology, and the method used to estimate this constant yield has been itself obtained by calibration on experimental data and has been shown by separate authors to provide satisfying predictive results on aerobic metabolisms.

Another argument (argument A2) against the use of this reference data would be that the assumption of a constant yield is a simplifying hypothesis anyway, and it may consequently put a bias on the estimation of the kinetic parameters of a model with dynamic yield. To state the problem more clearly; static and dynamic yields represent two different ideas of the microbial growth yield. A static yield is derived from considering the initial and final state of a growth event, while a dynamic yield has a meaning also between those two states. If an "overall yield" was to be computed from the initial and final state of the simulation of a model considering dynamic yields, maybe this overall yield would depend on the kinetic parameters of the model. In the context of a parameter calibration, multiple values of parameters are tested in order to find an optimum. The parameters values giving rise to the "overall yield" corresponding to the constant yield used to transform the reference data may be favored as optimum in the calibration process, above other criteria. Indeed, the constant yield, and its dynamic counterpart the overall yield, affects the area under the fitted curve, however, the "shape" of the curve is another fitting criterion.

In order to address argument A2 with tangible counter-arguments, some additional information is given on the calibration simulations.

The tracking of the lambda factor of the OHO population in the OHO population's calibration simulation with optimum parameters is given as fig a. It shows that the lambda factor's variation along simulation time is very small (approximately 3% of its mean value). Consequently the lambda factor can be considered as a constant in the case of the OHO population. The yield in the MTS model being a linear transformation of the lambda factor, this observation effectively counters argument A2 in the case of the OHO population. Moreover, the tracking of the value of the overall OUR (indicative of the overall yield) predicted by the MTS model for the OHO population for various values of the kinetic parameters shows that the yield of the OHO population is independent of the value of the kinetic parameters (different values of OUR recorded for very low parameters are caused by the truncation of the OUR record).

On the other hand, the tracking of the lambda factor for the AOB and NOB populations (fig b) shows that it varies greatly along simulation time. Unlike with the OHO population, the yield of the AOB and NOB populations cannot be considered as constant during the calibration simulation. However the calibration process of AOB and NOB populations is not disqualified by argument A2. Indeed, the measurement of the overall OUR of the AOB and NOB populations for various kinetic parameters values shows that the yield of the AOB and NOB populations depends on the value of the kinetic parameters. The OUR of the reference data is 0.67 mM (21.64 $mgO_2.L^{-1}$); this value lays at the far right of the graph. The parameters values selected as optimum by the calibration process are at the center of the graph. Consequently it appears that the overall yield displayed by the reference data did not act as an objective for the calibration process, which counters argument A2 in the case of the AOB and NOB populations.

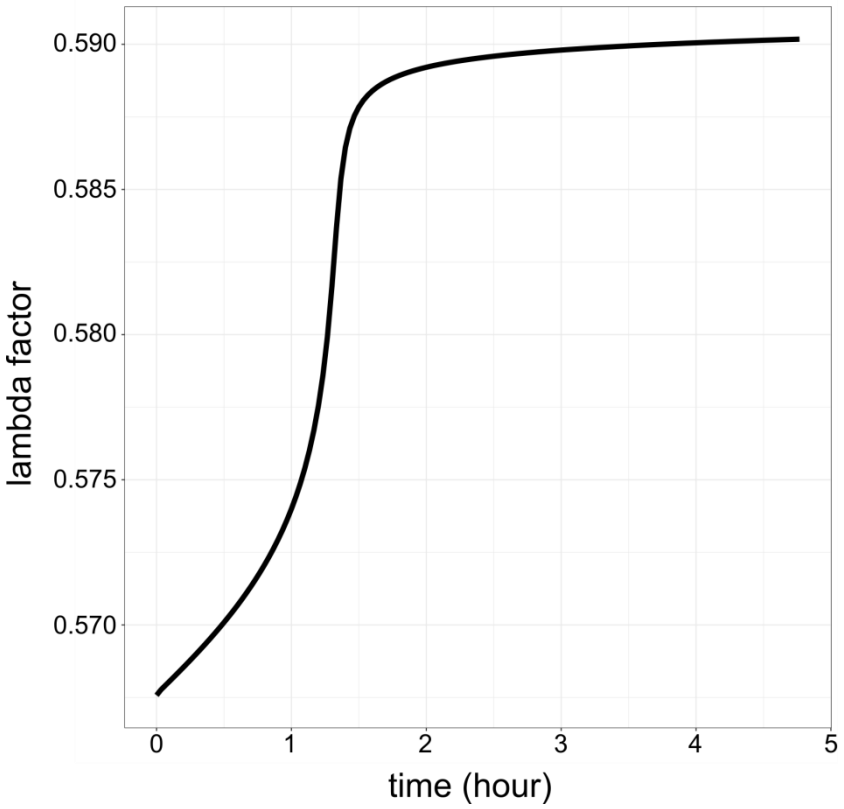


Figure a : lambda factor of the OHO population along simulation time, during the calibration simulation of the parameters of the OHO population, with optimum parameters (μ_{max} = 3.75 day^{-1}, V_h = 147 $m^3.mol^{-1}$)

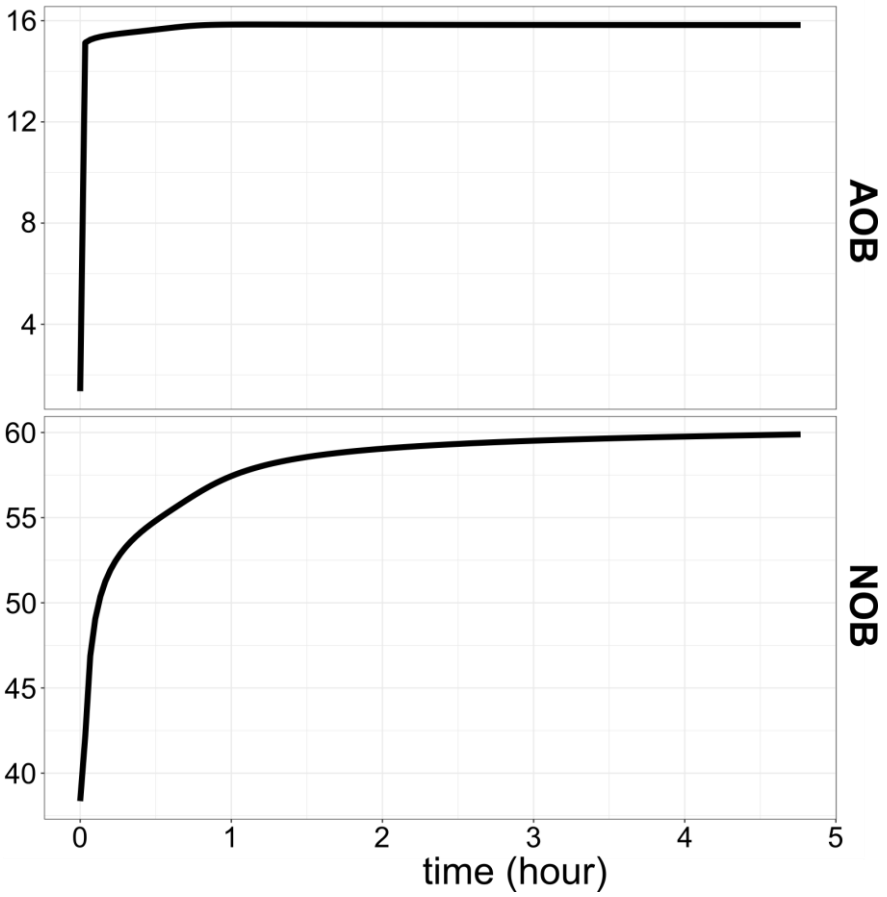


Figure b : lambda factor of the AOB and NOB populations along simulation time, during the calibration simulation of the parameters of the AOB and NOB populations, with optimum parameters (μ_{max} = 2.08 day^{-1}, V_h = 296 $m^3.mol^{-1}$, initial AOB population = 6.52e-1 mM, initial NOB population = 4.59e-1 mM)

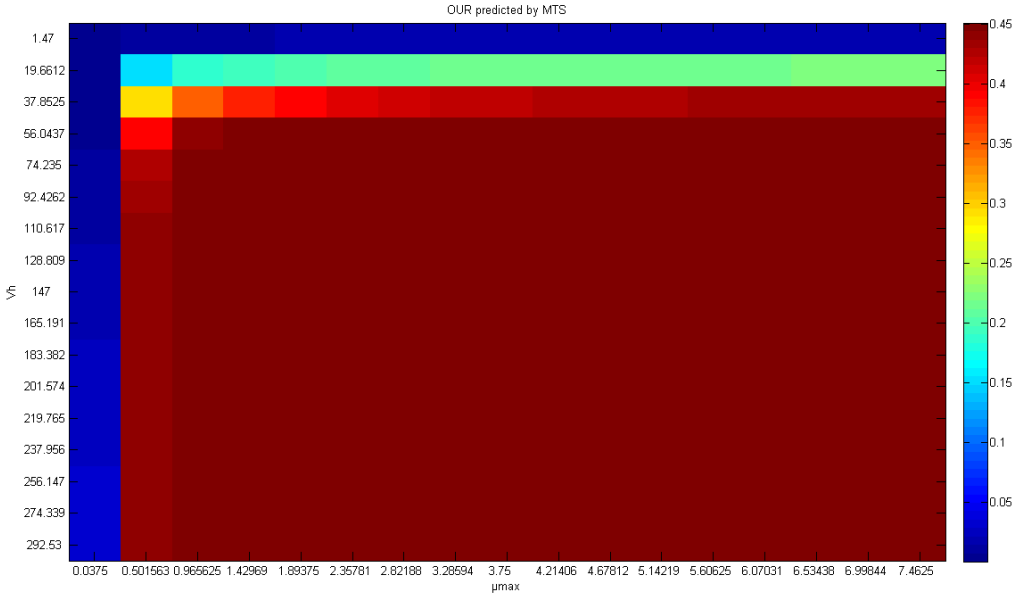


Figure c: total amount of oxygen (in mol) consumed by the OHO population as predicted by the MTS model for the calibration simulation of the OHO population

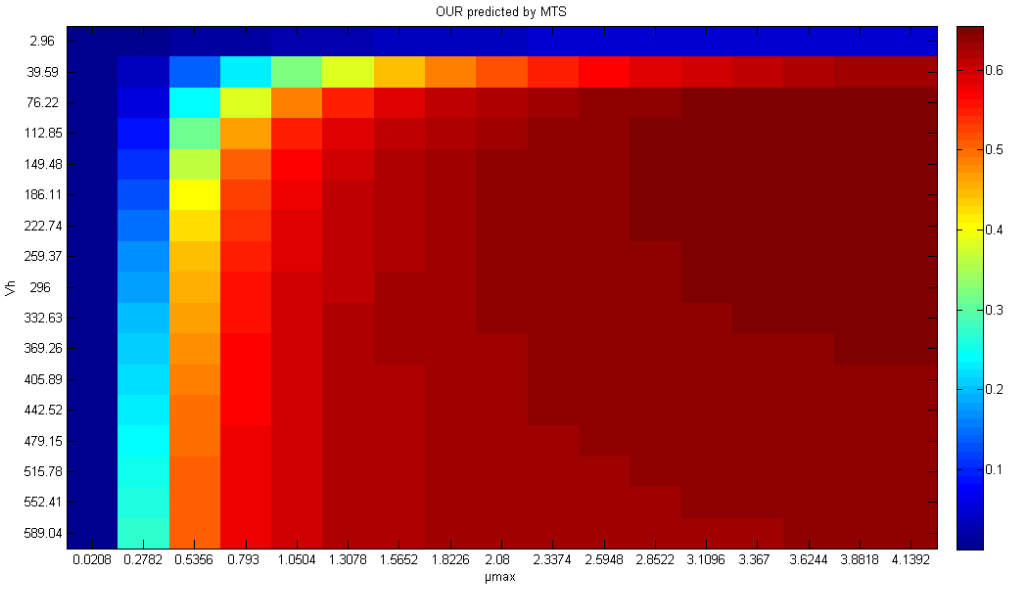


Figure d : total amount of oxygen (in mol) consumed by the AOB and NOB populations as predicted by the MTS model for the calibration simulation of the AOB and NOB populations

Supplementary material of chapter 6

Data collected from the literature and used in chapter 6 to investigate the link between growth yield and metabolic properties. In the "growth" column, "A" means "aerobic" and "N" means "anaerobic". In the "trophism" column, "H" means "heterotroph" and "A" means "autotroph". In the "metabolism type" column, "R" means "respiration" and "F" means "fermentation".

index	donor	acceptor	ne-	dGcat (kJ. mol_D^{-1})	observed yield (gX.mol_D^{-1})	growth	gammaCs (electrons per carbon source molecule)	NoCCs (number of carbon per carbon source molecule)	gammaD (number of electron per electron donr molecule)	NoCD (number of carb on per electron donor molecule)	trophism	metabolism type	dGan (kJ. $mol X^{-1}$)
1	ethanol	O2	12	-1308	25.22	A	12	2	12	2	H	R	69.50992
2	ethanol	O2	12	-1308	27.98	A	12	2	12	2	H	R	69.50992
3	ethanol	O2	12	-1308	23.96	A	12	2	12	2	H	R	69.50992
4	ethanol	O2	12	-1308	28.93	A	12	2	12	2	H	R	69.50992
5	ethanol	O2	12	-1308	24.05	A	12	2	12	2	H	R	69.50992
6	ethanol	O2	12	-1308	27.1	A	12	2	12	2	H	R	69.50992
7	ethanol	O2	12	-1308	26.98	A	12	2	12	2	H	R	69.50992
8	ethanol	NO3-	12	-1212	24	N	12	2	12	2	H	R	69.50992
9	ethanol	S	12	-167	9.8	N	12	2	12	2	H	R	69.50992
10	ethanol	SO4-2	4	-74	5.2	N	12	2	12	2	H	R	69.50992
11	ethanol	SO4-2	4	-74	3	N	12	2	12	2	H	R	69.50992
12	ethanol	SO4-2	4	-74	2.6	N	12	2	12	2	H	R	69.50992
13	ethanol	SO4-2	4	-74	3.4	N	12	2	12	2	H	R	69.50992
14	ethanol	SO4-2	4	-74	3.4	N	12	2	12	2	H	R	69.50992
15	ethanol	SO4-2	4	-74	2	N	12	2	12	2	H	R	69.50992
16	ethanol	SO4-2	4	-74	3	N	12	2	12	2	H	R	69.50992
17	ethanol	SO4-2	6.8	-95	6	N	12	2	12	2	H	R	69.50992
18	ethanol	none	12	-52	2.5	N	12	2	12	2	H	F	69.50992
19	ethanol	none	12	-52	4.4	N	12	2	12	2	H	F	69.50992
20	lactate	O2	12	1333	30.96	A	12	3	12	3	H	R	60.91319
21	lactate	O2	12	-1333	19.1	A	12	3	12	3	H	R	60.91319
22	lactate	O2	12	-1333	34.58	A	12	3	12	3	H	R	60.91319
23	lactate	O2	12	-1333	32.54	A	12	3	12	3	H	R	60.91319
24	lactate	O2	12	-1333	33.36	A	12	3	12	3	H	R	60.91319
25	lactate	O2	12	-1333	22.37	A	12	3	12	3	H	R	60.91319
26	lactate	O2	12	-1333	32.54	A	12	3	12	3	H	R	60.91319
27	lactate	O2	12	-1333	33.22	A	12	3	12	3	H	R	60.91319

28	lactate	O2	12	-1333	46.78	A	12	3	12	3	H	R	60.91319
29	lactate	O2	12	-1333	48.14	A	12	3	12	3	H	R	60.91319
30	lactate	O2	12	-1333	26.44	A	12	3	12	3	H	R	60.91319
31	lactate	O2	12	-1333	33.9	A	12	3	12	3	H	R	60.91319
32	lactate	NO3-	12	-1238	20.2	N	12	3	12	3	H	R	60.91319
33	lactate	Fe(OH)3	4	-462	11.83	N	12	3	12	3	H	R	60.91319
34	lactate	Fe(OH)3	4	-462	6	N	12	3	12	3	H	R	60.91319
35	lactate	Fe(OH)3	4	-462	11.66	N	12	3	12	3	H	R	60.91319
36	lactate	S	4	-111	5.9	N	12	3	12	3	H	R	60.91319
37	lactate	SO4-2	4	-100	6.5	N	12	3	12	3	H	R	60.91319
38	lactate	SO4-2	4	-100	7	N	12	3	12	3	H	R	60.91319
39	lactate	SO4-2	4	-100	4.4	N	12	3	12	3	H	R	60.91319
40	lactate	none	12	-78	3.4	N	12	3	12	3	H	F	60.91319
41	lactate	none	12	-78	2.9	N	12	3	12	3	H	F	60.91319
42	lactate	none	12	-78	5.3	N	12	3	12	3	H	F	60.91319
43	lactate	none	12	-78	3.9	N	12	3	12	3	H	F	60.91319
44	propionate	O2	14	-1487	32.54	A	14	3	14	3	H	R	80.85191
45	propionate	SO4-2	14	-118	10.1	N	14	3	14	3	H	R	80.85191
46	propionate	SO4-2	6	-59	4.93	N	14	3	14	3	H	R	80.85191
47	propionate	SO4-2	6	-59	4.6	N	14	3	14	3	H	R	80.85191
48	propionate	none	6	-24	1.6	N	14	3	14	3	H	F	80.85191
49	propionate	none	6	-24	1.02	N	14	3	14	3	H	F	80.85191
50	butyrate	SO4-2	20	-171	9.76	N	20	4	20	4	H	R	80.06082
51	butyrate	SO4-2	20	-171	15.3	N	20	4	20	4	H	R	80.06082
52	butyrate	none	4	-30	1.36	N	20	4	20	4	H	F	80.06082
53	butyrate	none	4	-30	0.55	N	20	4	20	4	H	F	80.06082
54	butyrate	none	4	-30	1.08	N	20	4	20	4	H	F	80.06082
55	acetate	O2	8	-847	18.35	A	8	2	8	2	H	R	81.99803
56	acetate	O2	8	-847	22.8	A	8	2	8	2	H	R	81.99803
57	acetate	O2	8	-847	16.8	A	8	2	8	2	H	R	81.99803
58	acetate	O2	8	-847	20.3	A	8	2	8	2	H	R	81.99803
59	acetate	O2	8	-847	15.37	A	8	2	8	2	H	R	81.99803

60	acetate	NO_3^-	8	-784	17	N	8	2	8	2	H	R	81.99803
61	acetate	NO_3^-	8	-784	16.9	N	8	2	8	2	H	R	81.99803
62	acetate	MnO_2	8	-605	13.62	N	8	2	8	2	H	R	81.99803
63	acetate	MnO_2	8	-605	18.24	N	8	2	8	2	H	R	81.99803
64	acetate	$Fe(OH)_3$	8	-736	14.05	N	8	2	8	2	H	R	81.99803
65	acetate	$Fe(OH)_3$	8	-736	16.32	N	8	2	8	2	H	R	81.99803
66	acetate	$Fe(OH)_3$	8	-736	16.45	N	8	2	8	2	H	R	81.99803
67	acetate	S	8	-86	4.2	N	8	2	8	2	H	R	81.99803
68	acetate	SO_4^{-2}	8	-64	4.55	N	8	2	8	2	H	R	81.99803
69	acetate	SO_4^{-2}	8	-64	5.6	N	8	2	8	2	H	R	81.99803
70	acetate	SO_4^{-2}	8	-64	1.97	N	8	2	8	2	H	R	81.99803
71	acetate	none	1	-31	2.7	N	8	2	8	2	H	F	81.99803
72	acetate	none	1	-31	1.6	N	8	2	8	2	H	F	81.99803
73	acetate	none	1	-31	1.94	N	8	2	8	2	H	F	81.99803
74	acetate	none	1	-31	1.25	N	8	2	8	2	H	F	81.99803
75	acetate	none	1	-31	1.15	N	8	2	8	2	H	F	81.99803
76	formate	O2	2	-233	3.66	A	2	1	2	1	H	R	38.08406
77	formate	O2	2	-233	5.9	A	2	1	2	1	H	R	38.08406
78	formate	O2	2	-233	6.3	A	2	1	2	1	H	R	38.08406
79	formate	O2	2	-233	6.75	A	2	1	2	1	H	R	38.08406
80	formate	O2	2	-233	4.68	A	2	1	2	1	H	R	38.08406
81	formate	O2	2	-233	4.37	A	2	1	2	1	H	R	38.08406
82	formate	O2	2	-233	3.29	A	2	1	2	1	H	R	38.08406
83	formate	NO_3^-	2	-217	4.05	N	2	1	2	1	H	R	38.08406
84	formate	NO_3^-	2	-217	3.03	N	2	1	2	1	H	R	38.08406
85	formate	NO_3^-	2	-150	4.15	N	2	1	2	1	H	R	38.08406
86	formate	none	2	-26	1.2	N	2	1	2	1	H	F	38.08406
87	H2	NO_3^-	2	-218	5.6	N	0	1	2	0	A	R	343.3657
88	H2	NO_3^-	2	-218	2.8	N	0	1	2	0	A	R	343.3657
89	H2	NO_3^-	2	-218	5.42	N	0	1	2	0	A	R	343.3657

90	H2	NO3-	2	-218	4.03	N	0	1	2	0	A	R	343.3657
91	H2	NO3-	2	-218	6.05	N	0	1	2	0	A	R	343.3657
92	H2	NO3-	2	-150	1.7	N	0	1	2	0	A	R	343.3657
93	H2	Fe(OH)3	2	-230	3.35	N	0	1	2	0	A	R	36.74476
94	H2	Fe(OH)3	2	-230	1.55	N	0	1	2	0	A	R	36.74476
95	H2	SO4-2	2	-38	1.7	N	0	1	2	0	A	R	36.74476
96	H2	SO4-2	2	-38	1.85	N	0	1	2	0	A	R	36.74476
97	H2	SO4-2	2	-38	0.85	N	0	1	2	0	A	R	36.74476
98	H2	SO4-2	2	-38	1.9	N	0	1	2	0	A	R	36.74476
99	H2	HCO3-	2	-30	0.4	N	0	1	2	0	A	R	36.74476
100	H2	HCO3-	2	-30	1.1	N	0	1	2	0	A	R	36.74476
101	H2	HCO3-	2	-30	0.9	N	0	1	2	0	A	R	36.74476
102	H2	HCO3-	2	-30	0.6	N	0	1	2	0	A	R	36.74476
103	H2	HCO3-	2	-30	0.65	N	0	1	2	0	A	R	36.74476
104	H2	HCO3-	2	-30	1.15	N	0	1	2	0	A	R	36.74476
105	H2	HCO3-	2	-30	1.6	N	0	1	2	0	A	R	36.74476
106	H2	HCO3-	2	-30	2.2	N	0	1	2	0	A	R	36.74476
107	glucose	O2	24	-2873	68.48	A	24	6	24	6	H	R	27.12507
108	glucose	O2	24	-2873	75.6	A	24	6	24	6	H	R	27.12507
109	glucose	O2	24	-2873	59.7	A	24	6	24	6	H	R	27.12507
110	glucose	O2	24	-2873	65.36	A	24	6	24	6	H	R	27.12507
111	glucose	O2	24	-2873	90.85	A	24	6	24	6	H	R	27.12507
112	glucose	O2	24	-2873	80.68	A	24	6	24	6	H	R	27.12507
113	glucose	O2	24	-2873	63.73	A	24	6	24	6	H	R	27.12507
114	glucose	O2	24	-2873	81.36	A	24	6	24	6	H	R	27.12507
115	glucose	NO3-	24	-2683	43.39	N	24	6	24	6	H	R	27.12507

116	glucose	O2	24	-2873	80	A	24	6	24	6	H	R	27.12507
117	glucose	O2	24	-2873	62.38	A	24	6	24	6	H	R	27.12507
118	glucose	O2	24	-2873	81.7	A	24	6	24	6	H	R	27.12507
119	glucose	O2	24	-2873	63.73	A	24	6	24	6	H	R	27.12507
120	glucose	NO3-	24	-2683	60	N	24	6	24	6	H	R	27.12507
121	glucose	none	24	-338	14.32	N	24	6	24	6	H	F	27.12507
122	glucose	none	24	-338	16.83	N	24	6	24	6	H	F	27.12507
123	glucose	none	24	-294	15.28	N	24	6	24	6	H	F	27.12507
124	oxalate	O2	2	-271	3.890984	A	2	2	2	2	H	R	-38.148
125	oxalate	O2	2	-271	3.16708	A	2	2	2	2	H	R	-38.148
127	glyoxylate	O2	4	-497	9.95368	A	4	2	4	2	H	R	8.511992
128	formate	O2	2	-233	2.71464	A	2	1	2	1	H	R	38.08406
129	formate	O2	2	-233	2.2622	A	2	1	2	1	H	R	38.08406
130	formate	O2	2	-233	4.07196	A	2	1	2	1	H	R	38.08406
131	tartrate	O2	10	-1204	25.33664	A	10	4	10	4	H	R	24.06
132	malonate	O2	8	-877	16.15211	A	8	3	8	3	H	R	66.74002
133	citrate	O2	18	-2036	51.3067	A	18	6	18	6	H	R	52.89188
134	citrate	O2	18	-2036	52.93548	A	18	6	18	6	H	R	52.89188
135	citrate	O2	18	-2036	55.78585	A	18	6	18	6	H	R	52.89188
136	malate	O2	12	-1367	38.00496	A	12	4	12	4	H	R	49.85226
137	malate	O2	12	-1367	22.622	A	12	4	12	4	H	R	49.85226
138	malate	O2	12	-1367	33.933	A	12	4	12	4	H	R	49.85226
139	citrate	O2	18	-2036	49.54218	A	18	6	18	6	H	R	52.89188
140	pyruvate	O2	10	-1141	29.45384	A	10	3	10	3	H	R	49.22641
141	pyruvate	O2	10	-1141	21.71712	A	10	3	10	3	H	R	49.22641
142	succinate	O2	14	-1519	37.37154	A	14	4	14	4	H	R	71.69533
14	succinate	O2	14	1519	34.83788	A	14	4	14	4	H	R	71.69533

3													
144	succinate	O2	14	-1519	43.43424	A	14	4	14	4	H	R	71.69533
145	succinate	O2	14	-1519	36.1952	A	14	4	14	4	H	R	71.69533
146	gluconate	O2	32	-2590	69.22332	A	32	6	32	6	H	R	181.8515
147	gluconate	O2	32	-2590	69.22332	A	32	6	32	6	H	R	181.8515
148	gluconate	O2	32	-2590	75.87419	A	32	6	32	6	H	R	181.8515
149	glucose	O2	24	-2873	78.4531	A	24	6	24	6	H	R	27.12507
150	glucose	O2	24	-2873	73.29528	A	24	6	24	6	H	R	27.12507
151	glucose	O2	24	-2873	61.21513	A	24	6	24	6	H	R	27.12507
152	galactose	O2	24	-2884	81.98213	A	24	6	24	6	H	R	25.30477
153	lactose	O2	48	-5793	149.0337	A	48	12	48	12	H	R	22.33302
154	glucose	O2	24	-2873	79.67468	A	24	6	24	6	H	R	27.12507
155	acetate	O2	8	-847	26.55823	A	8	2	8	2	H	R	81.99803
156	glucose	O2	24	-2873	75.60272	A	24	6	24	6	H	R	27.12507
157	glucose	O2	24	-2873	80.76054	A	24	6	24	6	H	R	27.12507
158	acetate	O2	8	-847	16.64979	A	8	2	8	2	H	R	81.99803
159	glucose	O2	24	-2873	65.42282	A	24	6	24	6	H	R	27.12507
160	glucose	O2	24	-2873	54.15707	A	24	6	24	6	H	R	27.12507
162	fructose	O2	24	-2877	68.54466	A	24	6	24	6	H	R	26.40697
163	glucose	O2	24	-2873	80.76054	A	24	6	24	6	H	R	27.12507
164	xylose	O2	20	-2408	55.4239	A	20	5	20	5	H	R	23.9952
165	acetate	O2	8	-847	20.58602	A	8	2	8	2	H	R	81.99803
166	glucose	O2	24	-2873	54.2928	A	24	6	24	6	H	R	27.12507
167	acetate	O2	8	-847	18.55004	A	8	2	8	2	H	R	81.99803
168	formaldehyde	O2	4	-492	10.63234	A	4	1	4	1	H	R	13.52801
169	glucose	O2	24	-2873	82.79652	A	24	6	24	6	H	R	27.12507
170	lactate	O2	12	-1333	34.61166	A	12	3	12	3	H	R	60.91319

171	mannitol	O2	28	-3085	84.15384	A	26	6	26	6	H	R	31.14861
172	mannitol	O2	28	-3085	76.00992	A	26	6	26	6	H	R	31.14861
173	sorbitol	O2	28	-3081	84.15384	A	26	6	26	6	H	R	31.73381
174	glycerol	O2	14	-1637	38.61575	A	14	3	14	3	H	R	38.02168
175	glycerol	O2	14	-1637	46.96327	A	14	3	14	3	H	R	38.02168
176	glycerol	O2	14	-1637	37.3263	A	14	3	14	3	H	R	38.02168
178	glycerol	O2	14	-1637	45.47022	A	14	3	14	3	H	R	38.02168
179	glycerol	O2	14	-1637	47.57407	A	14	3	14	3	H	R	38.02168
180	glycerol	O2	14	-1637	43.43424	A	14	3	14	3	H	R	38.02168
181	acetoin	O2	20	-2240	38.36691	A	20	4	20	4	H	R	57.54961
182	acetone	O2	16	-1726	30.20037	A	16	3	16	3	H	R	73.98401
183	2.3-butanediol	O2	22	-2433	40.35765	A	22	4	22	4	H	R	63.37158
184	xylose	O2	20	-2408	59.04342	A	26	5	20	5	H	R	79.98634
185	xylose	O2	20	-2408	83.7014	A	26	5	20	5	H	R	79.98634
186	methanol	O2	6	-682	11.80868	A	6	1	6	1	H	R	50.55047
187	ethanol	O2	12	-1308	27.01067	A	12	2	12	2	H	R	69.50992
189	ethanol	O2	12	-1308	27.91555	A	12	2	12	2	H	R	69.50992
190	methanol	O2	6	-682	12.21588	A	6	1	6	1	H	R	50.55047
191	ethanol	O2	12	-1308	23.97932	A	12	2	12	2	H	R	69.50992
192	propanol	O2	18	-1946	39.02295	A	18	3	18	3	H	R	73.06015
193	methanol	O2	6	-682	12.21588	A	6	1	6	1	H	R	50.55047
194	methanol	O2	6	-682	10.63234	A	6	1	6	1	H	R	50.55047
195	hexadecane	O2	94	-10060	205.2268	A	94	16	94	16	H	R	77.53487
196	butane	O2	28	-2751	40.26716	A	26	4	26	4	H	R	82.60681
197	methane	O2	8	-821	12.4421	A	6	1	8	1	H	R	50.55047
198	H2	HCO3-	2	-30	0.429818	N	0	1	2	0	A	R	36.74476

199	formate	none	2	-26	0.950124	N	2	1	2	1	H	F	38.08406
200	formate	none	2	-26	1.221588	N	2	1	2	1	H	F	38.08406
201	formate	none	2	-26	0.723904	N	2	1	2	1	H	F	38.08406
202	formate	none	2	-26	4.343424	N	2	1	2	1	H	F	38.08406
204	H2	HCO3-	2	-30	0.361952	N	0	1	2	0	A	R	36.74476
205	H2	HCO3-	2	-30	1.266832	N	0	1	2	0	A	R	36.74476
206	H2	HCO3-	2	-30	2.239578	N	0	1	2	0	A	R	36.74476
207	H2	HCO3-	2	-30	0.294086	N	0	1	2	0	A	R	36.74476
208	H2	HCO3-	2	-30	0.316708	N	0	1	2	0	A	R	36.74476
209	H2	HCO3-	2	-30	0.33933	N	0	1	2	0	A	R	36.74476
210	H2	HCO3-	2	-30	0.361952	N	0	1	2	0	A	R	36.74476
211	H2	HCO3-	2	-30	0.384574	N	0	1	2	0	A	R	36.74476
212	H2	HCO3-	2	-30	0.316708	N	0	1	2	0	A	R	36.74476
213	H2	HCO3-	2	-30	0.11311	N	0	1	2	0	A	R	36.74476
214	H2	HCO3-	2	-30	0.316708	N	0	1	2	0	A	R	36.74476
215	H2	HCO3-	2	-30	0.520306	N	0	1	2	0	A	R	36.74476
216	H2	HCO3-	2	-30	0.475062	N	0	1	2	0	A	R	36.74476
217	H2	HCO3-	2	-30	0.67866	N	0	1	2	0	A	R	36.74476
218	H2	HCO3-	2	-30	1.153722	N	0	1	2	0	A	R	36.74476
219	H2	HCO3-	2	-30	1.515674	N	0	1	2	0	A	R	36.74476
220	citrate	none	18	-173	9.908436	N	18	6	18	6	H	F	52.89188
221	H2	citrate	6	-215	4.07196	N	18	6	2	0	H	F	52.89188
222	pyruvate	none	10	-105	5.632878	N	10	3	10	3	H	F	49.22641
225	fructose	none	24	-393	23.48164	N	24	6	24	6	H	F	26.40697
226	glucose	none	24	-388	23.88883	N	24	6	24	6	H	F	27.12507
227	dihydroxy-acetone	none	12	-203	3.3933	N	0	1	12	3	A	F	36.74476

228	glucose	none	24	-388	23.88883	N	24	6	24	6	H	F	27.12507
229	H2	glucose	24	-793	33.933	N	24	6	2	0	H	F	27.12507
230	H2	lactate	1	-78	5.632878	N	12	3	2	0	H	F	60.91319
231	fructose	none	1	-352	42.48412	N	24	6	24	6	H	F	26.40697
232	glucose	none	1	-348	43.56997	N	24	6	24	6	H	F	27.12507
233	xylose	none	1	-304	31.89702	N	26	5	20	5	H	F	79.98634
234	H2	glucose	12	-373	19.00248	N	24	6	2	0	H	F	27.12507
235	H2	glucose	12	-373	16.5593	N	24	6	2	0	H	F	27.12507
236	H2	maltose	24	-2358	22.26005	N	0	1	2	0	H	F	-82.7085
237	glucose	none	1	-217	14.93052	N	24	6	24	6	H	F	27.12507
238	H2	acetate	8	-154	1.085856	N	8	2	2	0	H	F	81.99803
239	H2	acetate	8	-154	2.443176	N	8	2	2	0	H	F	81.99803
240	H2	acetate	8	-154	2.171712	N	8	2	2	0	H	F	81.99803
241	H2	acetate	8	-154	1.1311	N	8	2	2	0	H	F	81.99803
242	H2	acetate	8	-154	1.040612	N	8	2	2	0	H	F	81.99803
243	H2	acetate	8	-154	0.995368	N	8	2	2	0	H	F	81.99803
244	H2	acetate	8	-154	1.945492	N	8	2	2	0	H	F	81.99803
245	H2	acetate	8	-154	1.855004	N	8	2	2	0	H	F	81.99803
246	H2	acetate	8	-154	2.03598	N	8	2	2	0	H	F	81.99803
247	H2	acetate	8	-154	1.945492	N	8	2	2	0	H	F	81.99803
248	H2	acetate	8	-154	2.03598	N	8	2	2	0	H	F	81.99803
24	H2	acet	8	-154	2.533664	N	8	2	2	0	H	F	81.99803

9		ate											
250	H2	acetate	8	-154	3.800496	N	8	2	2	0	H	F	81.99803
251	H2	acetate	8	-154	1.628784	N	8	2	2	0	H	F	81.99803
252	H2	acetate	8	-154	2.443176	N	8	2	2	0	H	F	81.99803
253	H2	acetate	8	-154	2.624152	N	8	2	2	0	H	F	81.99803
254	H2	acetate	8	-154	1.402564	N	8	2	2	0	H	F	81.99803
255	H2	glucose	12	-373	18.73102	N	24	6	2	0	H	F	27.12507
256	H2	glucose	12	-373	15.74491	N	24	6	2	0	H	F	27.12507
257	H2	glucose	12	-373	15.20198	N	24	6	2	0	H	F	27.12507
258	mannitol	none	26	-394	41.80546	N	48	12	26	6	H	F	-107.622
259	mannitol	none	26	-394	40.99106	N	48	12	26	6	H	F	-107.622
260	glycerol	none	14	-188	6.311538	N	14	3	14	3	H	F	38.02168
261	H2	acetoin	1	-98	6.33416	N	20	4	2	0	H	F	57.54961
262	ethylenegl ycol	none	1	-89	3.257568	N	10	2	10	2	H	F	41.15281
263	acetoin	none	1	-125	7.148552	N	20	4	20	4	H	F	57.54961
264	acetoin	none	1	-137	7.148552	N	20	4	20	4	H	F	57.54961
265	2.3-butanediol	none	1	-59	3.257568	N	22	4	22	4	H	F	63.37158
266	2.3-butanediol	none	1	-107	5.700744	N	22	4	22	4	H	F	63.37158
267	2.3-butanediol	none	1	-119	6.424648	N	22	4	22	4	H	F	63.37158
268	H2	methanol	2	-101	3.031348	N	8	1	2	0	H	F	95.05603
269	H2	methanol	2	-101	6.7866	N	8	1	2	0	H	F	95.05603
27	H2	met	2	-101	3.868362	N	8	1	2	0	H	F	95.05603

0		hanol											
271	H2	methanol	2	-101	2.850372	N	8	1	2	0	H	F	95.05603
272	H2	methanol	2	-101	2.533664	N	8	1	2	0	H	F	95.05603
273	ethanol	none	1	-38	1.266832	N	12	2	12	2	H	F	69.50992
274	propanol	none	1	-42	1.289454	N	18	3	18	3	H	F	73.06015
275	H2	methanol	2	-101	3.257568	N	8	1	2	0	H	F	95.05603
276	formate	NO3-	2	-150	3.755252	N	2	1	2	1	H	R	38.08406
277	succinate	NO3-	14	-933	24.79371	N	14	4	14	4	H	R	71.69533
278	gluconate	NO3-	32	-1251	52.52828	N	32	6	32	6	H	R	181.8515
279	lactate	NO3-	12	-831	34.27233	N	12	3	12	3	H	R	60.91319
280	mannitol	NO3-	26	-1997	137.3608	N	48	12	26	6	H	R	-107.622
281	propionate	none	6	-24	1.409351	N	14	3	14	3	H	F	80.85191
282	propionate	none	6	-24	0.79177	N	14	3	14	3	H	F	80.85191
283	propionate	none	6	-24	0.395885	N	14	3	14	3	H	F	80.85191
284	propionate	none	6	-24	0.870947	N	14	3	14	3	H	F	80.85191
285	propionate	none	6	-24	0.870947	N	14	3	14	3	H	F	80.85191
286	propionate	none	6	-24	0.79177	N	14	3	14	3	H	F	80.85191
287	propionate	none	6	-24	0.79177	N	14	3	14	3	H	F	80.85191
288	propionate	none	6	-24	0.300873	N	14	3	14	3	H	F	80.85191
289	propionate	none	6	-24	0.300873	N	14	3	14	3	H	F	80.85191
290	propionate	none	6	-24	0.300873	N	14	3	14	3	H	F	80.85191
291	butyrate	none	4	-30	1.250997	N	20	4	20	4	H	F	80.06082
292	butyrate	none	4	-30	1.045136	N	20	4	20	4	H	F	80.06082
293	butyrate	none	4	-30	1.045136	N	20	4	20	4	H	F	80.06082
29	butyrate	none	4	-30	1.045136	N	20	4	20	4	H	F	80.06082

4													
295	butyrate	none	4	-30	1.045136	N	20	4	20	4	H	F	80.06082
296	butyrate	none	4	-30	1.045136	N	20	4	20	4	H	F	80.06082
297	butyrate	none	4	-30	0.41172	N	20	4	20	4	H	F	80.06082
298	butyrate	none	4	-30	0.41172	N	20	4	20	4	H	F	80.06082
299	valerate	none	4	-20	0.79177	N	26	5	26	5	H	F	79.98634
300	valerate	none	4	-20	0.981795	N	26	5	26	5	H	F	79.98634
301	acetate	none	1	-31	0.870947	N	8	2	8	2	H	F	81.99803
302	acetate	none	1	-31	1.20349	N	8	2	8	2	H	F	81.99803
303	acetate	none	1	-31	0.506733	N	8	2	8	2	H	F	81.99803
304	acetate	none	1	-31	1.187655	N	8	2	8	2	H	F	81.99803
305	acetate	none	1	-31	1.187655	N	8	2	8	2	H	F	81.99803
306	acetate	none	1	-31	1.20349	N	8	2	8	2	H	F	81.99803
307	acetate	none	1	-31	0.221696	N	8	2	8	2	H	F	81.99803
308	acetate	none	1	-31	0.760099	N	8	2	8	2	H	F	81.99803
309	acetate	none	1	-31	0.760099	N	8	2	8	2	H	F	81.99803
310	acetate	none	1	-31	0.475062	N	8	2	8	2	H	F	81.99803
311	acetate	none	1	-31	0.475062	N	8	2	8	2	H	F	81.99803
312	acetate	none	1	-31	9.976302	N	8	2	8	2	H	F	81.99803
313	acetate	none	1	-31	4.275558	N	8	2	8	2	H	F	81.99803
314	acetate	none	1	-31	0.443391	N	8	2	8	2	H	F	81.99803
315	acetate	none	1	-31	0.443391	N	8	2	8	2	H	F	81.99803
316	acetate	none	1	-31	0.601745	N	8	2	8	2	H	F	81.99803
317	acetate	none	1	-31	0.395885	N	8	2	8	2	H	F	81.99803
318	H2	HCO3-	2	-30	0.886782	N	0	1	2	0	A	R	36.74476
319	H2	HCO3-	2	-30	2.897878	N	0	1	2	0	A	R	36.74476
320	H2	HCO3-	2	-30	0.221696	N	0	1	2	0	A	R	36.74476

321	H2	HCO3-	2	-30	0.950124	N	0	1	2	0	A	R	36.74476
322	H2	HCO3-	2	-30	0.79177	N	0	1	2	0	A	R	36.74476
323	H2	HCO3-	2	-30	0.79177	N	0	1	2	0	A	R	36.74476
324	H2	HCO3-	2	-24	1.108478	N	0	1	2	0	A	F	36.74476
325	H2	HCO3-	2	-24	2.228141	N	0	1	2	0	A	F	36.74476
326	H2	HCO3-	2	-24	2.160621	N	0	1	2	0	A	F	36.74476
327	acetate	O2	8	-847	10.60972	A	8	2	8	2	H	R	81.99803
328	acetate	O2	8	-847	10.29301	A	8	2	8	2	H	R	81.99803
329	acetate	O2	8	-847	10.92643	A	8	2	8	2	H	R	81.99803
330	acetate	O2	8	-847	11.24313	A	8	2	8	2	H	R	81.99803
331	acetate	O2	8	-847	11.55984	A	8	2	8	2	H	R	81.99803
332	ammonium	O2	6	-269	1.783312	A	0	1	4	0	A	R	326.0854
333	ammonium	O2	6	-269	2.080531	A	0	1	4	0	A	R	326.0854
334	ammonium	O2	6	-269	1.981458	A	0	1	4	0	A	R	326.0854
335	ammonium	O2	6	-269	2.278677	A	0	1	4	0	A	R	326.0854
336	ammonium	O2	6	-269	2.774041	A	0	1	4	0	A	R	326.0854
337	nitrite	O2	1	-82	2.278677	A	0	1	1	0	A	R	343.3657
338	nitrite	O2	1	-82	1.882385	A	0	1	1	0	A	R	343.3657
339	nitrite	O2	1	-82	2.179604	A	0	1	1	0	A	R	343.3657
340	nitrite	O2	1	-82	0.594437	A	0	1	1	0	A	R	343.3657
341	acetate	NO3-	8	-512	1.066466	N	8	2	8	2	H	R	81.99803
343	glucose	none	4.6	-347	2.71464	N	24	6	24	6	H	F	27.12507
344	glucose	none	4.6	-347	3.800496	N	24	6	24	6	H	F	27.12507
345	glucose	none	4.6	-347	4.614888	N	24	6	24	6	H	F	27.12507
346	glucose	none	4.6	-347	1.35732	N	24	6	24	6	H	F	27.12507
347	glucose	none	4.6	-347	1.900248	N	24	6	24	6	H	F	27.12507
34	glucose	none	4.	-347	4.07196	N	24	6	24	6	H	F	27.12507

8			6										
349	glucose	none	4.6	-347	4.07196	N	24	6	24	6	H	F	27.12507

```
suppressMessages(library(dplyr))
suppressMessages(library(magrittr))
suppressMessages(library(tidyr))
suppressWarnings(library(xlsx))
```

```
## Loading required package: rJava
```

```
## Loading required package: methods
```

```
## Loading required package: xlsxjars
```

```
suppressWarnings(library(knitr))
library(ggplot2)
library(ggfortify)

"yields_v7.xlsx" %>%
 read.xlsx("data") %>%
 na.omit() %>%
 tbl_df() -> mt.data

# biomass molecular weight in g.C-mol-1
# assuming C5H7O2N biomass (Hoover and Porges, 1952)
carbon.per.biomass <- 5
mwx <- 113.11 / carbon.per.biomass

# convert observed yield from gX.molS-1 to molX.molS-1
mt.data %<>% mutate(observed.yield = observed.yield / mwx)

# convert reduction degrees in mole.C-mol-1
mt.data %<>% mutate(gammaCs = gammaCs / ifelse(NoCCs == 0, 1, NoCCs))
mt.data %<>% mutate(gammaD = gammaD / ifelse(NoCD == 0, 1, NoCD))

# compute the dataset of the maximum yield per metabolism
mt.data %>%
 na.omit() %>%
 group_by(catabolism.index, e.donor, e.acceptor) %>%
 mutate(observations = n()) %>%
 filter(observed.yield == max(observed.yield)) %>%
 filter(row_number() == 1) %>%
 mutate(donor.reduction = ifelse(gammaD <= 4.67, "low", "high")) %>%
 ungroup() -> mt.data.max
```

data distribution

The data collected from the litterature encompasses a wide variety of metabolisms defined by their electron acceptor/donor couple. However, not all couples are documented with the same number of observations; some metabolisms have their growth yield well documented by a great number of experimental observations while other metabolisms have been seldomly documented. Lets plot a matrix of the number of observations we have per acceptor/donor couple to illustrate those discrepancies.

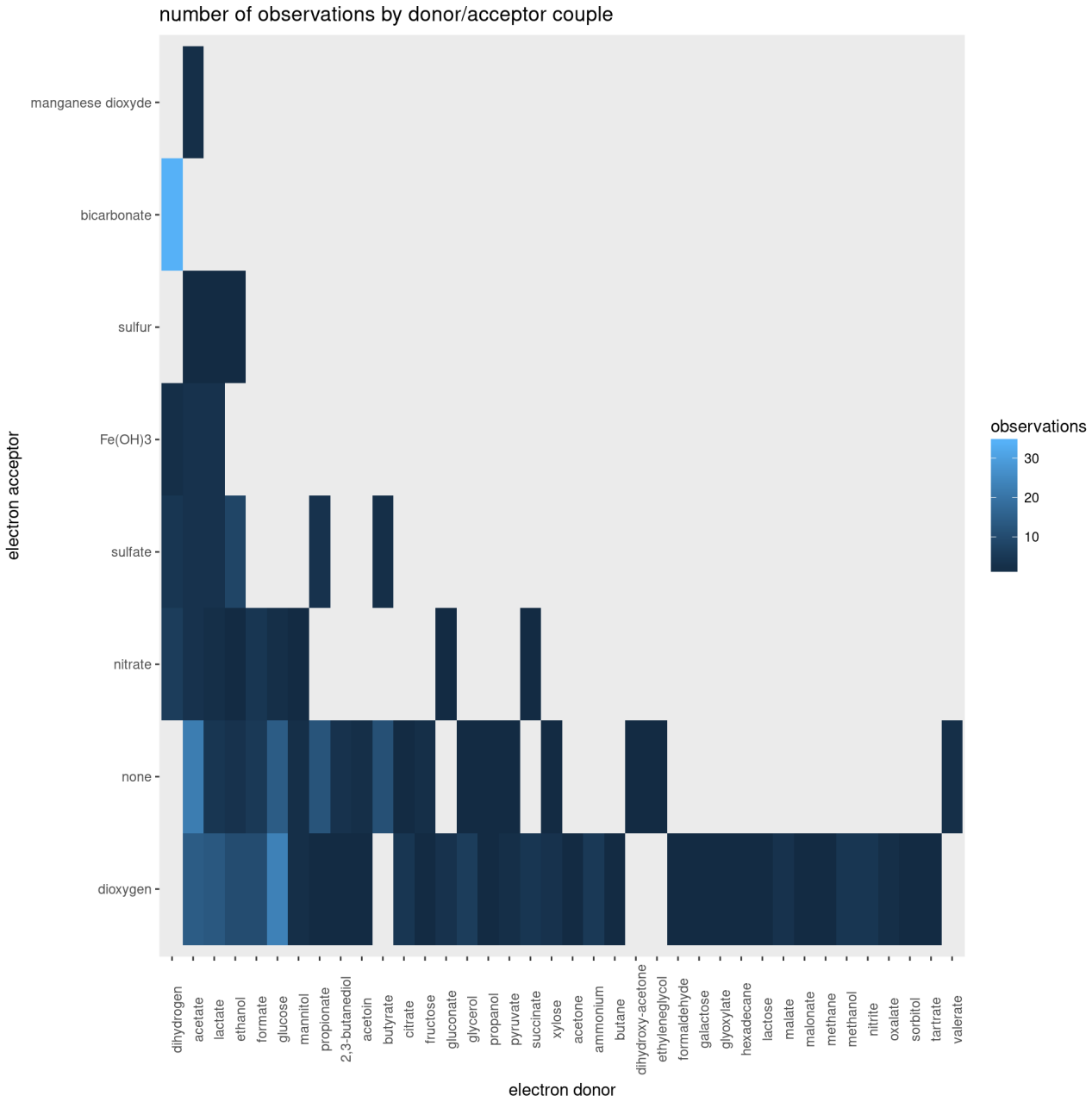
```
# count the number of literature report per catabolism
mt.data %>%
    group_by(catabolism.index) %>%
    summarise(observations = n()) %$%
    summary(observations)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   1.000   1.000   3.376   3.000  31.000
```

```
# count the number of literature report of growth yield per acceptor/donor pair
mt.data %>%
 group_by(e.donor, e.acceptor) %>%
 summarise(observations = n()) -> observations.per.metabolism

# display the distribution of the number of observations per metabolism
observations.per.metabolism %$%
    observations %>%
    summary() -> metabolism.count.summary

print(metabolism.count.summary)
```

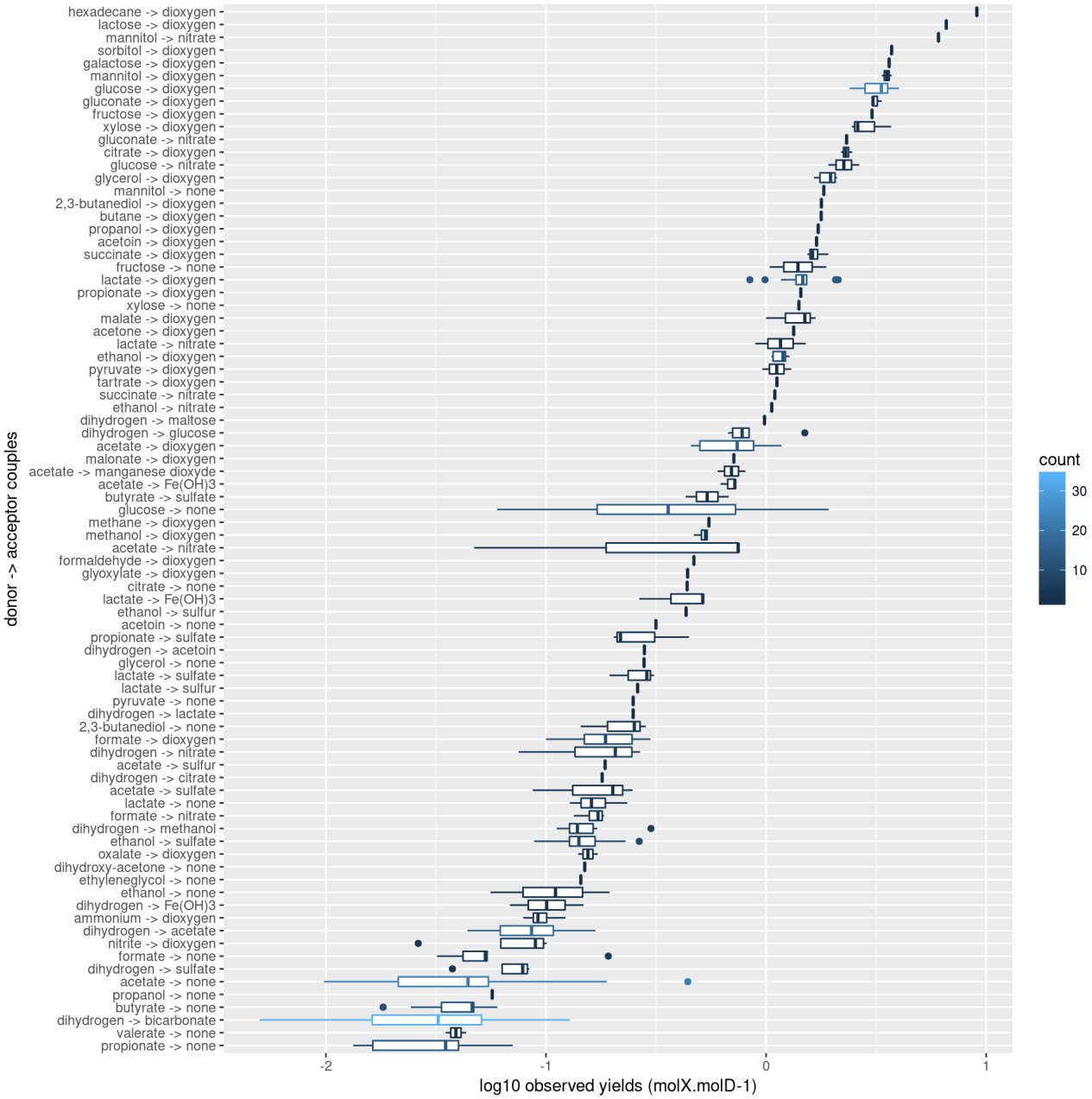
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   1.000   2.000   4.159   4.000  34.000
```

```
# number of observations per metabolism as a matrix
(observations.per.metabolism %>%
 group_by(e.donor) %>%
 mutate(observation.per.donor = n()) %>%
 group_by(e.acceptor) %>%
 # keep only inorganic electron acceptors as organic electron acceptors are
 # defined to include fermentation reactions in a consistent notation with othe
r
 # metabolisms while their electron donor is always dihydrogen
 filter(e.acceptor %in% c("dioxygen", "nitrate", "sulfate", "sulfur", "Fe(OH)3"
, "bicarbonate", "manganese dioxyde", "none")) %>%
 mutate(observation.per.acceptor = n()) %>%
 {(ggplot(., aes(x=reorder(e.donor, -observation.per.donor), y=reorder(e.accept
or, -observation.per.acceptor)))
       + geom_raster(aes(fill=observations))
       + ggtitle("number of observations by donor/acceptor couple")
       + xlab("electron donor")
       + ylab("electron acceptor")
       + theme(panel.grid.major = element_blank(),
               axis.text.x=element_text(angle = 90))
 )}
)
```



The experimental observations collected from the litterature cover a wide range of values of growth yield. The following figure illustrates the values of observed yields and their dispersion depending on the considered metabolism.

```
# dispersion of observed yields per metabolism
mt.data %>%
 unite("e.donor", "e.acceptor", col="metabolism", sep=" -> ") %>%
 group_by(metabolism) %>%
 mutate(avg.yield = mean(observed.yield)) %>%
 mutate(count = n()) %>%
 {(ggplot(., aes(reorder(metabolism, avg.yield), log10(observed.yield)))
       + geom_boxplot(aes(color=count))
       + coord_flip()
       + xlab("donor -> acceptor couples")
       + ylab("log10 observed yields (molX.molD-1)")
)}
```



Here are the data classified into the autotroph/heterotroph and aerobic/anaerobic categories;

```
mt.data %>%
    group_by(growth, trophism) %>%
    summarise(count = n(), avg.yield = mean(observed.yield)) %T>%
    # number of observations in each categories
    { xtabs(count ~ growth + trophism, data=.) %>% print() } %>%
    # average observed yield for each categories
    xtabs(avg.yield ~ growth + trophism, data=.)
```

```
##            trophism
## growth      autotroph heterotroph
##   aerobic           9         119
##   anaerobic        47         166
```

```
##            trophism
## growth       autotroph heterotroph
##   aerobic   0.08758987  1.82152737
##   anaerobic 0.06763830  0.37015275
```

previous prediction models

```
# compute goodness of fit criteria on a prediction
# SSR: Sum of Squared Residuals
# MSE: Mean Squared Error

# arguments;
# data: a data frame like mt.data or mt.data.max, with also a `predicted.yield`
column
# partition: a string designating the column to use as a partition
# if partition is null, goodness of fit criteria wont be applied on any partiti
on
goodness.of.fit <- function(data, partition=NULL) {
    data %<>% mutate(residuals = observed.yield - predicted.yield)

    # overall goodness of fit
    data %>%
    group_by("partition" = "overall") %>%
    summarise(SSR = sum(residuals ** 2),
              MSE = SSR / n()) -> overall.goodness.of.fit

    # goodness of fit per partition
    if (is.null(partition)) {
        return (overall.goodness.of.fit)
    } else {
        data %>%
        group_by("partition" = get(partition)) %>%
        summarise(SSR = sum(residuals ** 2),
                  MSE = SSR / n()) -> partitioned.goodness.of.fit
        return (rbind(partitioned.goodness.of.fit, overall.goodness.of.fit))
    }
}

bland.altman.plot <- function(data) {
    data %<>%
        ungroup() %>%
        mutate(difference = observed.yield - predicted.yield) %>%
        mutate(mean = (observed.yield + predicted.yield) / 2)

    mean.difference <- mean(data$difference)
    sd.difference <- sd(data$difference)

    data %>%
        {( ggplot(., aes(x=mean, y=difference))
           + geom_abline(intercept = mean.difference, slope = 0, size = 1.5, co
lor = "gray")
           + geom_abline(intercept = mean.difference + 1.96 * sd.difference, sl
ope = 0, linetype = 2)
```

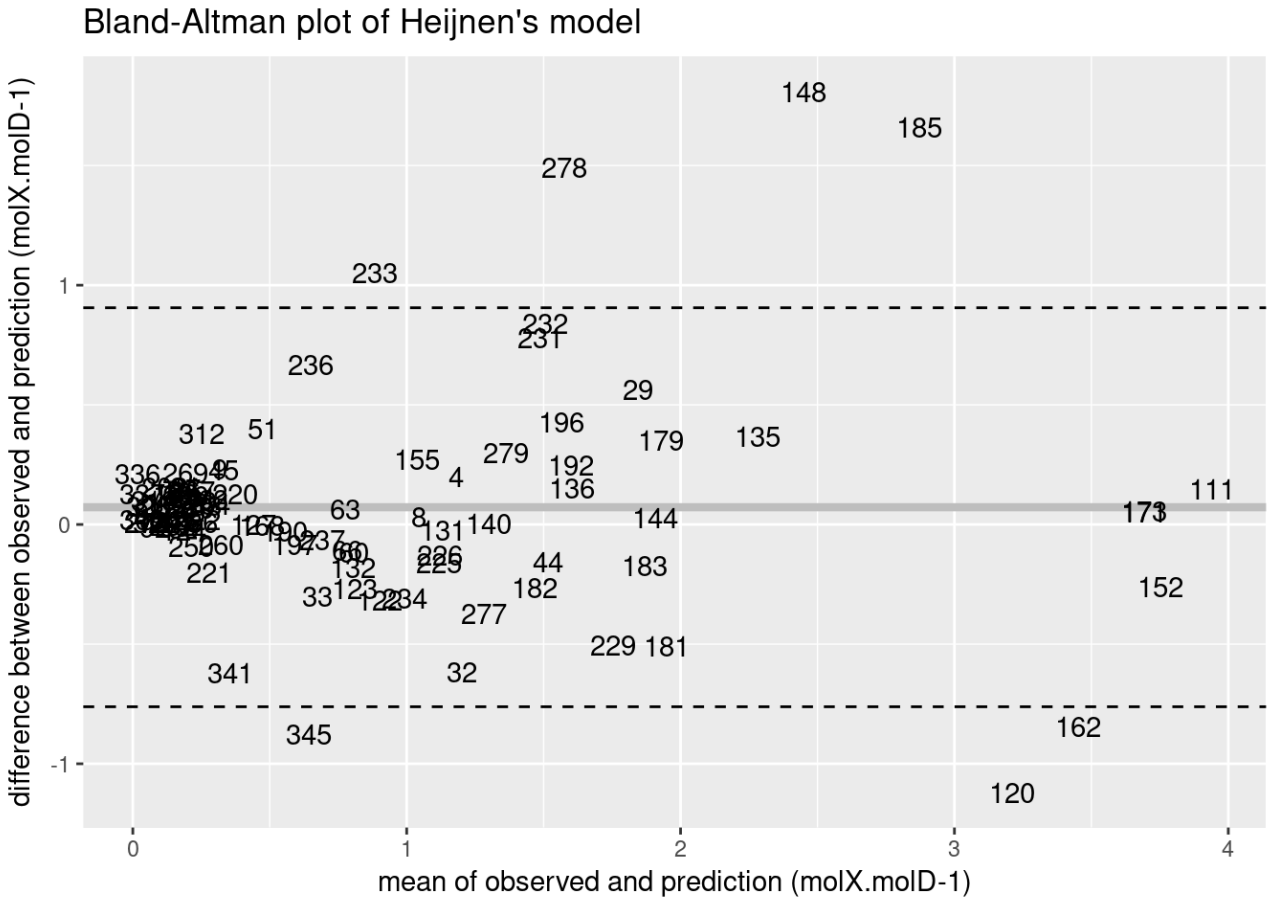
```
           + geom_abline(intercept = mean.difference - 1.96 * sd.difference, sl
ope = 0, linetype = 2)
           #+ geom_point(shape=1)
           + geom_text(aes(label=index))
           + xlab("mean of observed and prediction (molX.molD-1)")
           + ylab("difference between observed and prediction (molX.molD-1)")
        )}
}
```

Heijnen's model

```
# dissipated energy computed from Heijnen's predictor
YGmax <- function(NoCCs, gammaCs) { -(200 + 18 * (6 - NoCCs) ** 1.8 + exp(((3.8
- gammaCs)**2)**0.16 * (3.6 + 0.4 * NoCCs))) }

mt.data.max %>%
    mutate(dGdis = YGmax(NoCCs, gammaCs)) %>%
    mutate(dGdis = ifelse(e.donor %in% c("ammonium", "nitrite"), 3500, dGdis))
%>% # reverse electron transfer
    mutate(lambda = (dGdis - dGan) / dGcat) %>%
    mutate(predicted.yield = 1 / (lambda - nuDan)) %>%
    na.omit() -> mt.data.max.heijnen

# goodness of fit of the model
mt.data.max.heijnen %>%
    goodness.of.fit() %>%
    print()
```

```
## # A tibble: 1 x 3
##   partition      SSR       MSE
##       <chr>    <dbl>     <dbl>
## 1   overall 17.87086 0.1842357
```

```
mt.data.max.heijnen %>%
    goodness.of.fit() %>%
    select(MSE) %>%
    as.numeric() -> MSE.heijnen

 # draw a plot of predicted vs observed yields
mt.data.max.heijnen %>%
    bland.altman.plot() + ggtitle("Bland-Altman plot of Heijnen's model")
```

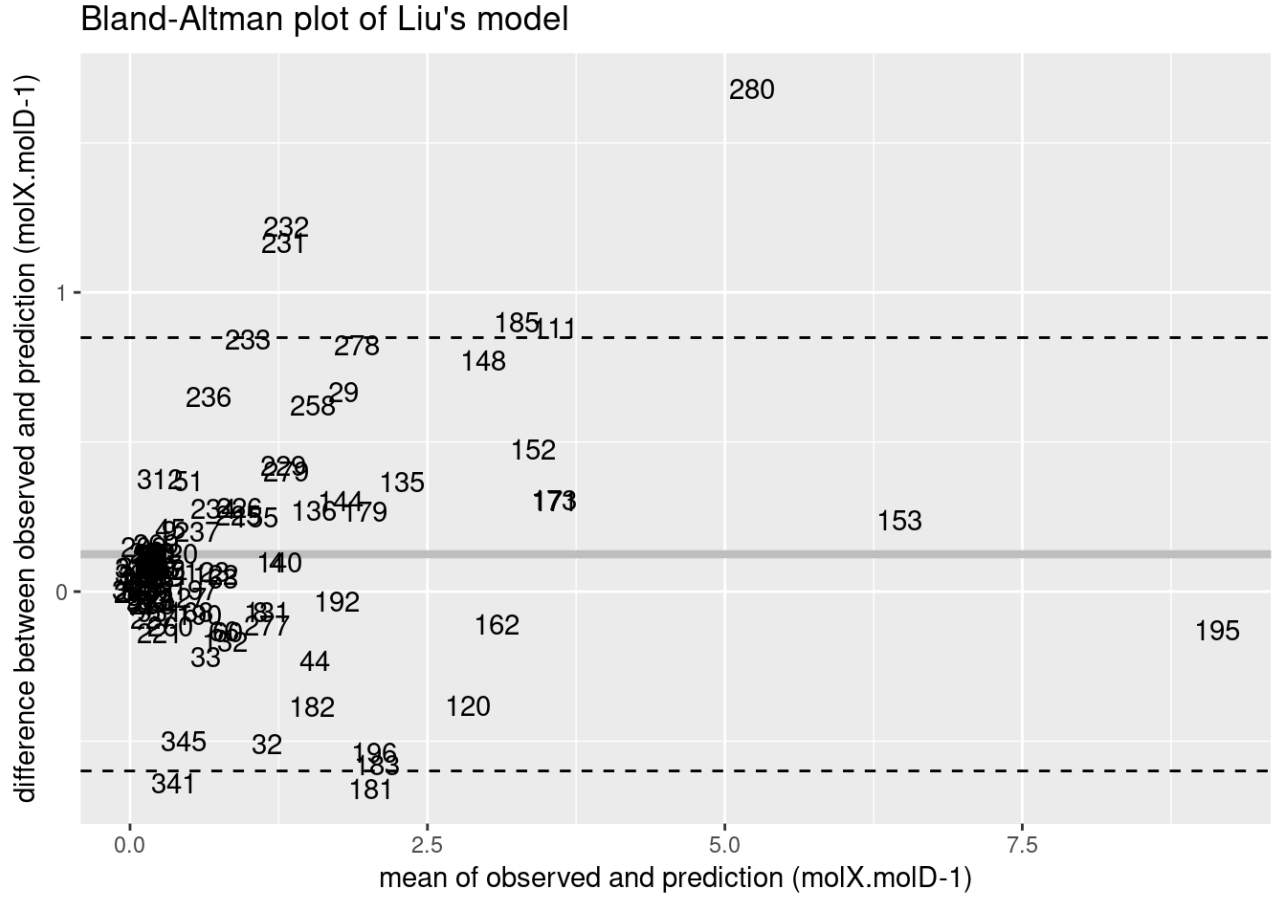


Liu's model

```
mt.data.max %>%
    mutate(dGdis = ifelse(gammaD <= 4.67,
                          -666.2 / (gammaD) - 243.1,
                          -157 * gammaD + 339)) %>%
    mutate(dGdis = ifelse(e.donor %in% c("ammonium", "nitrite"), -3500, dGdis))
%>% # reverse electron transfer
    mutate(lambda = (-dGdis + dGan) / -dGcat) %>%
    mutate(predicted.yield = 1 / (lambda - nuDan)) -> mt.data.max.liu

# goodness of fit of the model
mt.data.max.liu %>%
    goodness.of.fit() %>%
    print()
```

```
## # A tibble: 1 x 3
##   partition      SSR       MSE
##       <chr>    <dbl>     <dbl>
## 1   overall 15.22632 0.1507557
```

```
mt.data.max.liu %>%
    goodness.of.fit() %>%
    select(MSE) %>%
    as.numeric() -> MSE.liu

# draw a plot of predicted vs observed yields
mt.data.max.liu %>%
    bland.altman.plot() + ggtitle("Bland-Altman plot of Liu's model")
```

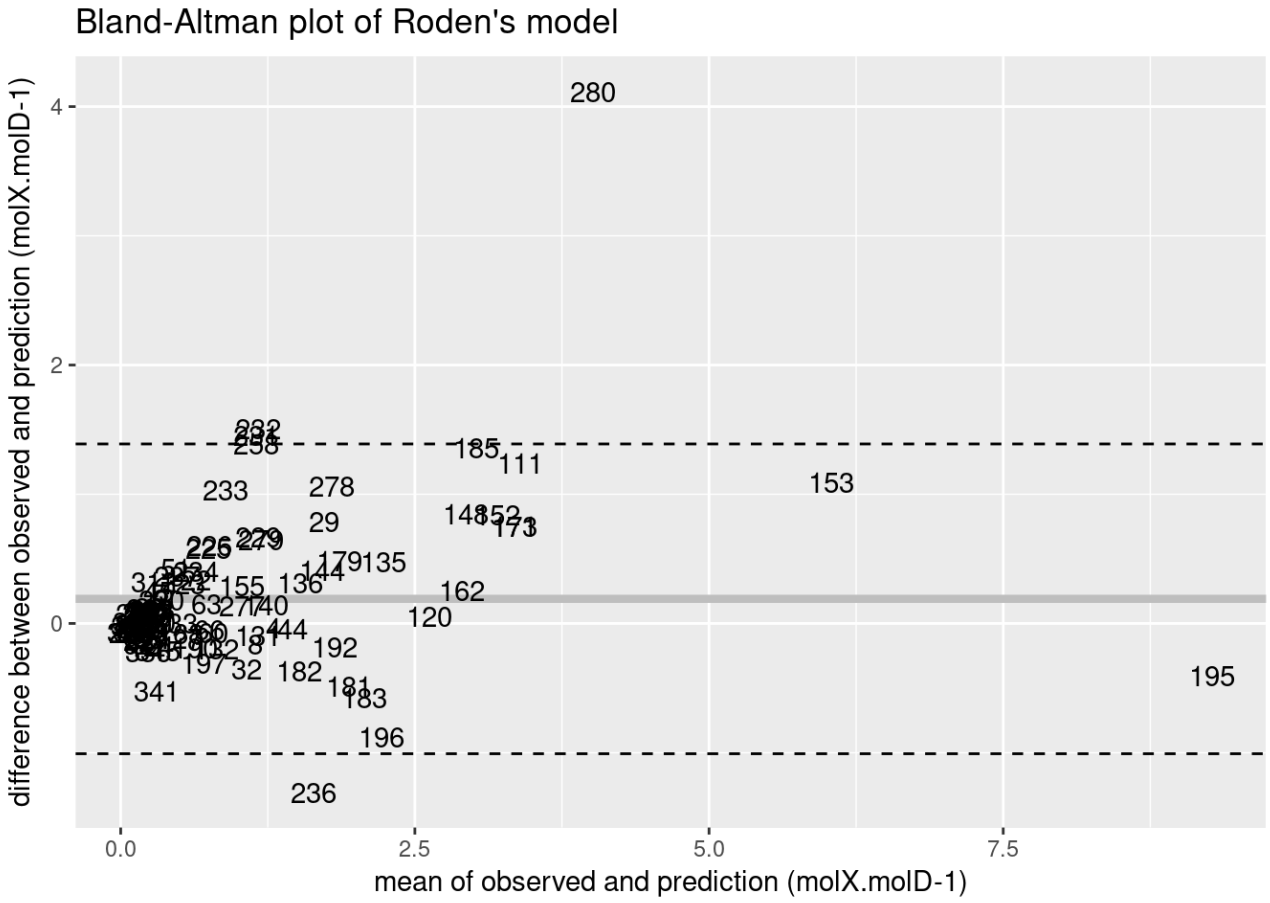


Roden's model

```
mt.data.max %>%
    mutate(predicted.yield = (2.08 + 0.0211 * -dGcat) / mwx) -> mt.data.max.rod
en

# goodness of fit of the model
mt.data.max.roden %>%
    goodness.of.fit() %>%
    print()
```

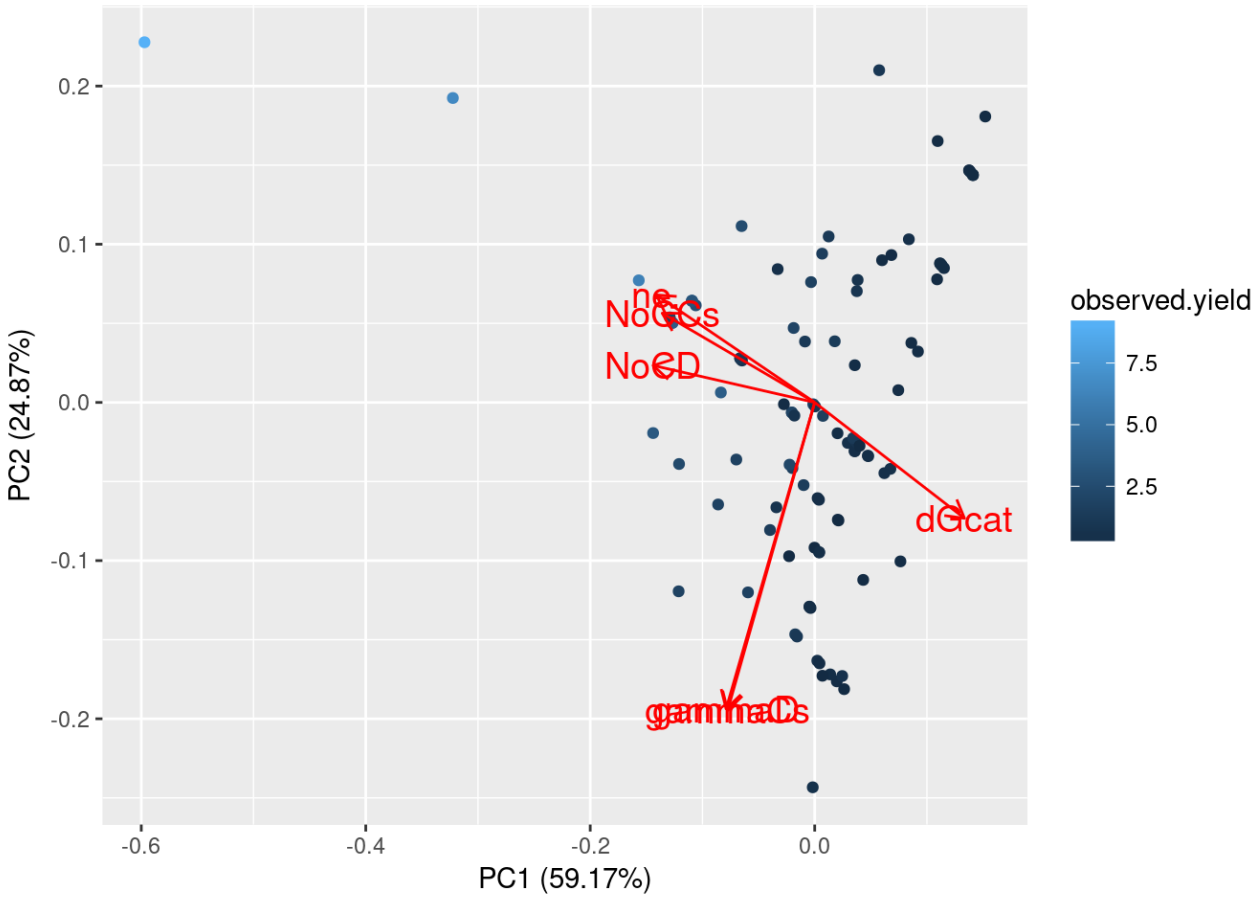
```
## # A tibble: 1 x 3
##   partition      SSR       MSE
##       <chr>    <dbl>     <dbl>
## 1   overall 41.07026 0.4066362
```

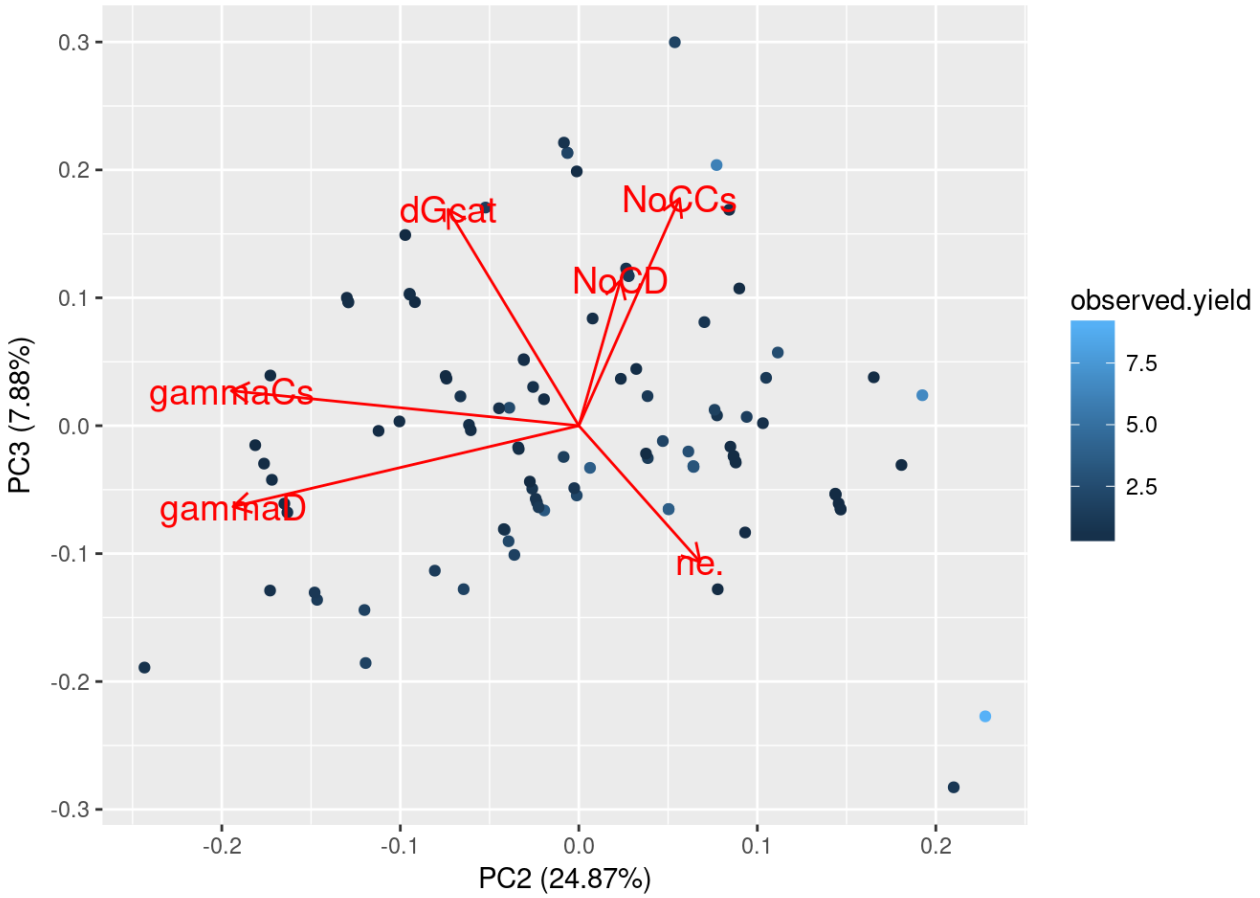
```
mt.data.max.roden %>%
    goodness.of.fit() %>%
    select(MSE) %>%
    as.numeric() -> MSE.roden

# draw a plot of predicted vs observed yields
mt.data.max.roden %>%
    bland.altman.plot() + ggtitle("Bland-Altman plot of Roden's model")
```



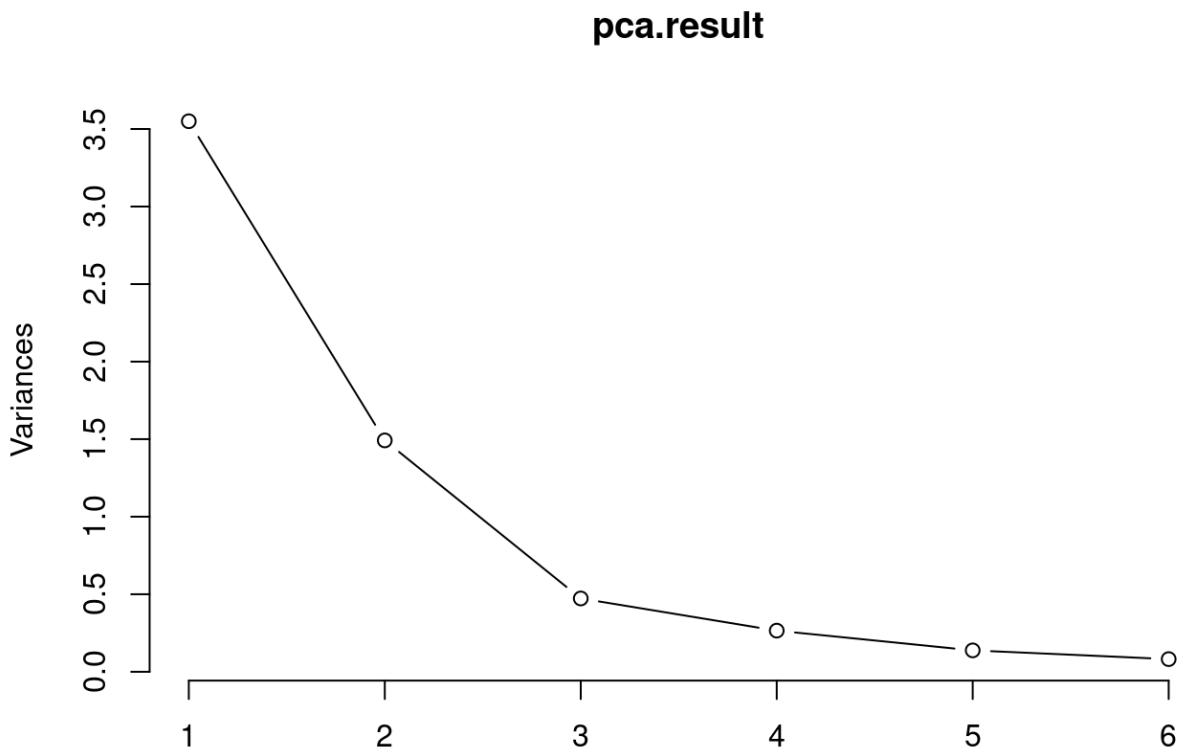
correlations between the explanatory variables

```
mt.data.max %>%
 select(ne., dGcat, gammaCs, NoCCs, gammaD, NoCD) %>%
 prcomp(scale.=T, center=T) %T>%
 { autoplot(., data = ungroup(mt.data.max),
            colour = "observed.yield",
            loadings = TRUE,
            loadings.label = TRUE,
            loadings.label.size = 5) %>% print() } %T>%
 { autoplot(., data = ungroup(mt.data.max),
            colour = "observed.yield",
            x = 2,
            y = 3,
            loadings = TRUE,
            loadings.label = TRUE,
            loadings.label.size = 5) %>% print() } -> pca.result
```


0.2
0.1
0.0
-0.1
-0.2
PC2 (24.87%)
-0.6
-0.4
-0.2
0.0
PC1 (59.17%)
NoCCs
NoCD
dCcat
observed.yield
7.5
5.0
2.5




```
# plot the explained variance per principal component
plot(pca.result, type="l")
```




```
# print the principal components
print(pca.result)
```

```
## Standard deviations (1, .., p=6):
## [1] 1.8841418 1.2214990 0.6876278 0.5150678 0.3715558 0.2859533
##
## Rotation (n x k) = (6 x 6):
##                PC1         PC2        PC3        PC4        PC5        PC6
## ne.     -0.4725236  0.22763845 -0.3567543  0.1173889 -0.5728495  0.5056598
## dGcat    0.4469454 -0.24533135  0.5664200 -0.1300413 -0.4611115  0.4355297
## gammaCs -0.2597615 -0.65203598  0.0914658  0.6575424  0.1885743  0.1763089
## NoCCs   -0.4550714  0.18899482  0.5950297  0.1595864 -0.3395378 -0.5122274
## gammaD  -0.2644323 -0.64884515 -0.2120819 -0.5751945 -0.2555727 -0.2606352
## NoCD    -0.4821522  0.07825131  0.3801179 -0.4250066  0.4930072  0.4395789
```

Correlation of the explanatory variables between eachother

```
mt.data.max %>%
    ungroup() %>%
    select(dGcat, gammaCs, NoCCs, ne., gammaD, NoCD) %>%
    # spearman correlation coefficient is used because we do not assume the
    # relation between the variables to be linear
    cor(method="spearman")
```

```
##               dGcat     gammaCs      NoCCs        ne.     gammaD
## dGcat    1.00000000 -0.04340505 -0.4208124 -0.7132874 -0.1000734
## gammaCs -0.04340505  1.00000000  0.2359981  0.1164430  0.8565269
## NoCCs   -0.42081244  0.23599812  1.0000000  0.4877175  0.2143523
## ne.     -0.71328735  0.11644304  0.4877175  1.0000000  0.2158116
## gammaD  -0.10007343  0.85652687  0.2143523  0.2158116  1.0000000
## NoCD    -0.43796939  0.25954054  0.8336416  0.4836768  0.3967044
##               NoCD
## dGcat   -0.4379694
## gammaCs  0.2595405
## NoCCs    0.8336416
## ne.      0.4836768
## gammaD   0.3967044
## NoCD     1.0000000
```

Correlation of the explanatory variable with the maximum yield, for the overall dataset and for specific partitions

```
grouped.mt.data.max <- group_by(mt.data.max, metabolism.type, donor.reduction)
for (variable in c("dGcat", "ne.", "gammaCs", "gammaD", "NoCCs", "NoCD")) {
    paste("correlation between subpartitions and", variable) %>%
        print()
    grouped.mt.data.max %>%
        summarise(cor(get(variable), observed.yield)) %>%
        print()
}
```

```
## [1] "correlation between subpartitions and dGcat"
## # A tibble: 4 x 3
## # Groups:   metabolism.type [?]
##   metabolism.type donor.reduction `cor(get(variable), observed.yield)`
##             <fctr>           <chr>                                <dbl>
## 1    fermentation            high                           -0.9449024
## 2    fermentation             low                           -0.3723215
## 3     respiration            high                           -0.9736257
## 4     respiration             low                           -0.9242158
## [1] "correlation between subpartitions and ne."
## # A tibble: 4 x 3
## # Groups:   metabolism.type [?]
##   metabolism.type donor.reduction `cor(get(variable), observed.yield)`
##             <fctr>           <chr>                                <dbl>
## 1    fermentation            high                           -0.1192747
## 2    fermentation             low                            0.2842146
## 3     respiration            high                            0.9813195
## 4     respiration             low                            0.9357717
## [1] "correlation between subpartitions and gammaCs"
## # A tibble: 4 x 3
## # Groups:   metabolism.type [?]
##   metabolism.type donor.reduction `cor(get(variable), observed.yield)`
##             <fctr>           <chr>                                <dbl>
## 1    fermentation            high                           -0.3084685
## 2    fermentation             low                            0.1354055
## 3     respiration            high                           -0.0682507
## 4     respiration             low                            0.4591582
## [1] "correlation between subpartitions and gammaD"
## # A tibble: 4 x 3
## # Groups:   metabolism.type [?]
##   metabolism.type donor.reduction `cor(get(variable), observed.yield)`
##             <fctr>           <chr>                                <dbl>
## 1    fermentation            high                           -0.3084685
## 2    fermentation             low                            0.2568150
## 3     respiration            high                           -0.1480200
## 4     respiration             low                            0.4318263
## [1] "correlation between subpartitions and NoCCs"
## # A tibble: 4 x 3
## # Groups:   metabolism.type [?]
##   metabolism.type donor.reduction `cor(get(variable), observed.yield)`
##             <fctr>           <chr>                                <dbl>
## 1    fermentation            high                            0.1916450
## 2    fermentation             low                            0.6398154
## 3     respiration            high                            0.9730822
## 4     respiration             low                            0.9403043
## [1] "correlation between subpartitions and NoCD"
```

```
## # A tibble: 4 x 3
## # Groups:   metabolism.type [?]
##   metabolism.type donor.reduction `cor(get(variable), observed.yield)`
##             <fctr>           <chr>                                <dbl>
## 1    fermentation            high                            0.1916450
## 2    fermentation             low                            0.4521973
## 3     respiration            high                            0.9730822
## 4     respiration             low                            0.8866956
```

new generic predictor

selection of the variables

Generic predictor based on catabolic energy and reduction degrees

```
generic.model.gamma <- lm(observed.yield ~ (dGcat + gammaD + gammaCs)^2, data=m
t.data.max)
generic.model.gamma %>%
    summary() %>%
    print()
```

```
##
## Call:
## lm(formula = observed.yield ~ (dGcat + gammaD + gammaCs)^2, data = mt.data.m
ax)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.8048 -0.2066 -0.0682  0.1172  3.8422
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -0.7556569  0.3471233  -2.177  0.03199 *
## dGcat          -0.0014835  0.0002030  -7.308  8.8e-11 ***
## gammaD          0.3489203  0.1368328   2.550  0.01239 *
## gammaCs         0.1423772  0.0888758   1.602  0.11252
## dGcat:gammaD    0.0004081  0.0001222   3.339  0.00121 **
## dGcat:gammaCs  -0.0003082  0.0001063  -2.900  0.00464 **
## gammaD:gammaCs -0.0570701  0.0249965  -2.283  0.02468 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5529 on 94 degrees of freedom
## Multiple R-squared:  0.8678, Adjusted R-squared:  0.8594
## F-statistic: 102.9 on 6 and 94 DF,  p-value: < 2.2e-16
```

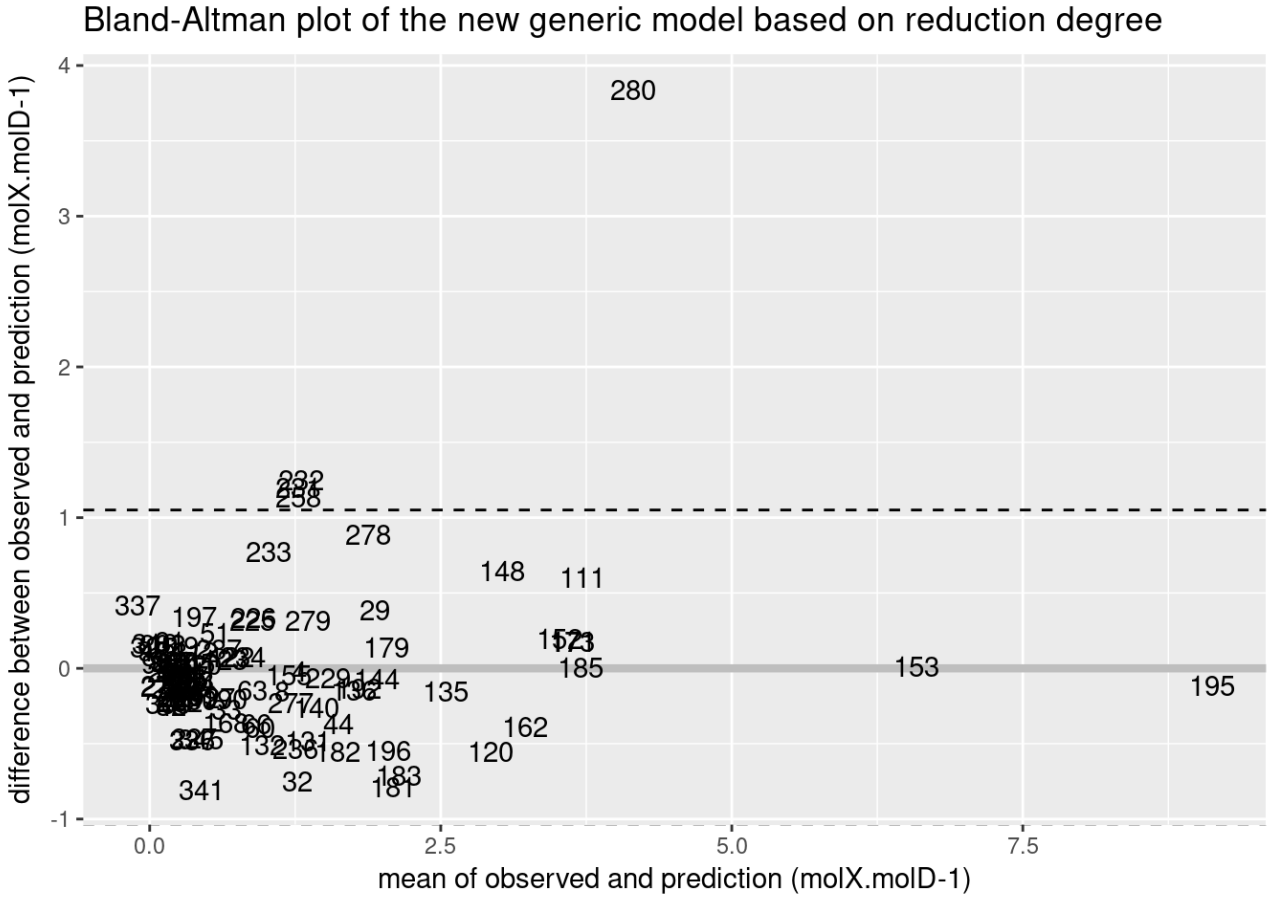
```
mt.data.max %>%
    mutate(predicted.yield = predict(generic.model.gamma)) -> mt.data.max.gener
ic.gamma

# goodness of fit of the model
mt.data.max.generic.gamma %>%
    goodness.of.fit("metabolism.type") %>%
    print()
```

```
## # A tibble: 3 x 3
##      partition       SSR       MSE
##         <fctr>     <dbl>     <dbl>
## 1 fermentation  6.397067 0.1683439
## 2  respiration 22.338703 0.3545826
## 3      overall 28.735771 0.2845126
```

```
mt.data.max.generic.gamma %>%
    goodness.of.fit("donor.reduction") %>%
    print()
```

```
## # A tibble: 3 x 3
##   partition       SSR       MSE
##       <chr>     <dbl>     <dbl>
## 1      high  3.446129 0.1276344
## 2       low 25.289642 0.3417519
## 3   overall 28.735771 0.2845126
```

```
mt.data.max.generic.gamma %>%
    goodness.of.fit() %>%
    select(MSE) %>%
    as.numeric() -> MSE.generic.gamma

# draw a plot of predicted vs observed yields
mt.data.max.generic.gamma %>%
    bland.altman.plot() + ggtitle("Bland-Altman plot of the new generic model b
ased on reduction degree")
```



Generic predictor based on catabolic energy and numbers of carbons

```
generic.carbon.model <- lm(observed.yield ~ (dGcat + NoCD + NoCCs)^2, data=mt.d
ata.max)
generic.carbon.model %>%
    summary() %>%
    print()
```

```
##
## Call:
## lm(formula = observed.yield ~ (dGcat + NoCD + NoCCs)^2, data = mt.data.max)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.46578 -0.23370  0.00158  0.21104  1.10453
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.495e-01  1.088e-01  -2.293   0.0241 *
## dGcat       -8.846e-04  8.878e-05  -9.964  < 2e-16 ***
## NoCD         1.271e-01  5.601e-02   2.270   0.0255 *
## NoCCs        4.200e-02  4.869e-02   0.863   0.3906
## dGcat:NoCD   2.551e-04  4.327e-05   5.897 5.79e-08 ***
## dGcat:NoCCs -2.419e-04  4.708e-05  -5.138 1.50e-06 ***
## NoCD:NoCCs   1.134e-03  9.429e-03   0.120   0.9045
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4018 on 94 degrees of freedom
## Multiple R-squared:  0.9302, Adjusted R-squared:  0.9257
## F-statistic: 208.8 on 6 and 94 DF,  p-value: < 2.2e-16
```

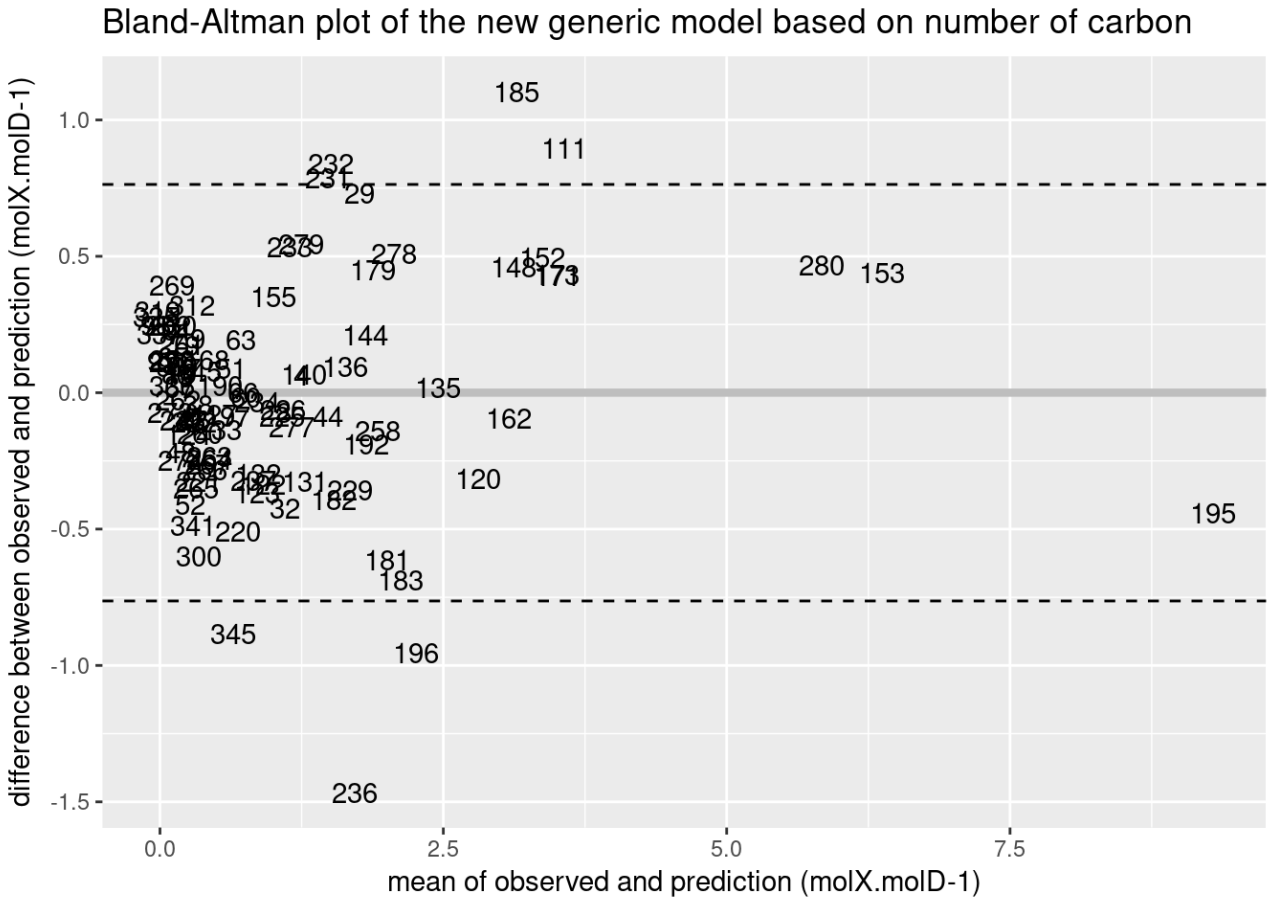
```
mt.data.max %>%
    mutate(predicted.yield = predict(generic.carbon.model)) -> mt.data.max.gene
ric.carbon

# goodness of fit of the model
mt.data.max.generic.carbon %>%
    goodness.of.fit("metabolism.type") %>%
    print()
```

```
## # A tibble: 3 x 3
##      partition       SSR       MSE
##         <fctr>     <dbl>     <dbl>
## 1 fermentation  6.998443 0.1841696
## 2  respiration  8.178914 0.1298240
## 3      overall 15.177357 0.1502709
```

```
mt.data.max.generic.carbon %>%
    goodness.of.fit("donor.reduction") %>%
    print()
```

```
## # A tibble: 3 x 3
##   partition       SSR       MSE
##       <chr>     <dbl>     <dbl>
## 1      high  3.665413 0.1357560
## 2       low 11.511944 0.1555668
## 3   overall 15.177357 0.1502709
```

```
mt.data.max.generic.carbon %>%
    goodness.of.fit() %>%
    select(MSE) %>%
    as.numeric() -> MSE.generic.carbon

# draw a plot of predicted vs observed yields
mt.data.max.generic.carbon %>%
    bland.altman.plot() + ggtitle("Bland-Altman plot of the new generic model b
ased on number of carbon")
```



generic model based on the reduction degree of the electron donor and the number of carbons of the carbon source

```
generic.gammaDNoCCs.model <- lm(observed.yield ~ (dGcat + gammaD + NoCCs)^2, da
ta=mt.data.max)
generic.gammaDNoCCs.model %>%
    summary() %>%
    print()
```

```
##
## Call:
## lm(formula = observed.yield ~ (dGcat + gammaD + NoCCs)^2, data = mt.data.max
)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.3881 -0.2761  0.0155  0.2003  2.4028
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -2.617e-01  2.265e-01  -1.155   0.2509
## dGcat        -1.247e-03  1.835e-04  -6.799 9.61e-10 ***
## gammaD       -2.698e-02  6.075e-02  -0.444   0.6580
## NoCCs         1.242e-01  8.534e-02   1.455   0.1491
## dGcat:gammaD  8.385e-05  4.291e-05   1.954   0.0536 .
## dGcat:NoCCs   9.618e-06  8.752e-06   1.099   0.2746
## gammaD:NoCCs  1.967e-02  2.161e-02   0.910   0.3650
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4604 on 94 degrees of freedom
## Multiple R-squared:  0.9083, Adjusted R-squared:  0.9025
## F-statistic: 155.3 on 6 and 94 DF,  p-value: < 2.2e-16
```

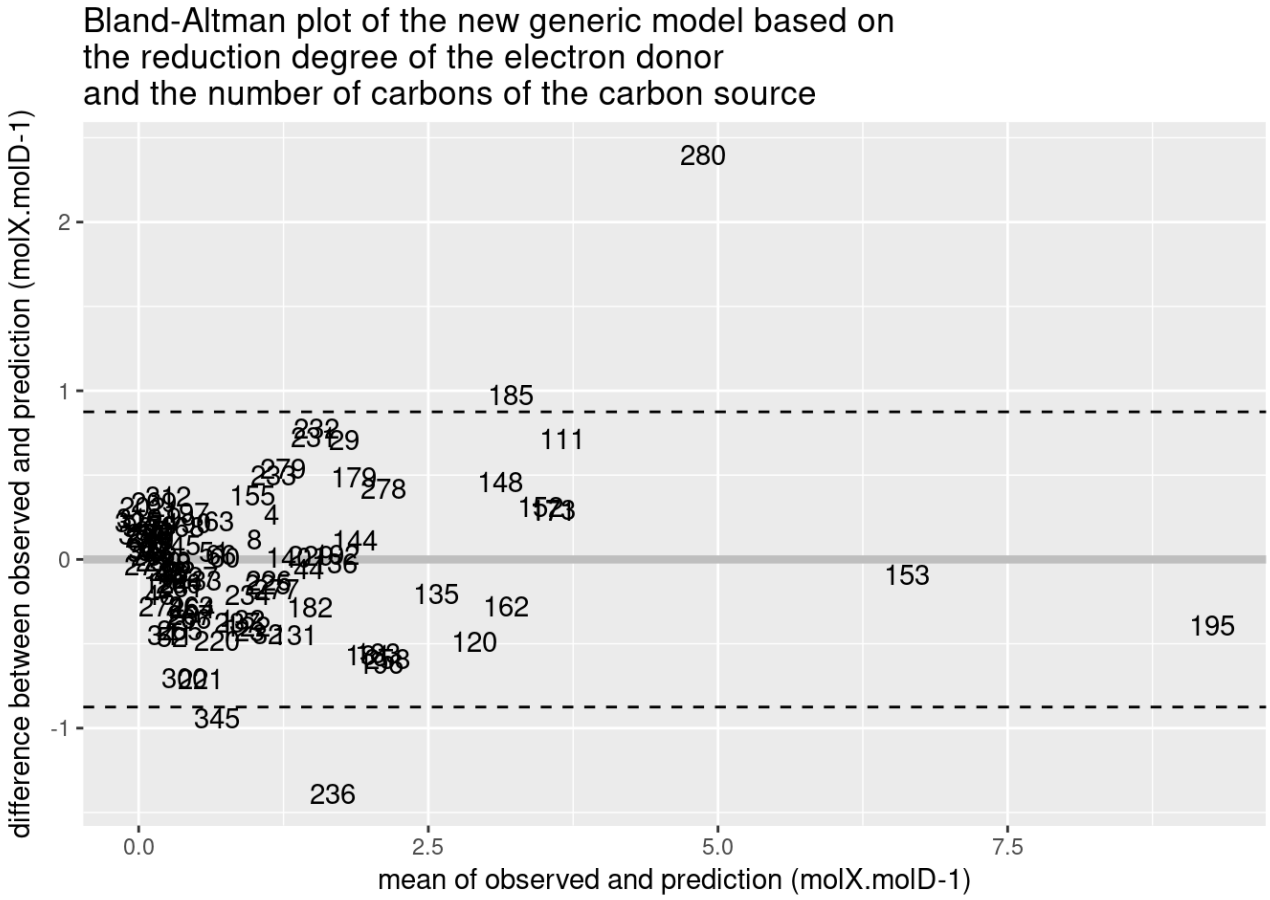
```
mt.data.max %>%
    mutate(predicted.yield = predict(generic.gammaDNoCCs.model)) -> mt.data.max
.generic.gammaDNoCCs

# goodness of fit of the model
mt.data.max.generic.gammaDNoCCs %>%
    goodness.of.fit("metabolism.type") %>%
    print()
```

```
## # A tibble: 3 x 3
##      partition       SSR       MSE
##         <fctr>     <dbl>     <dbl>
## 1 fermentation  7.711462 0.2029332
## 2  respiration 12.217329 0.1939259
## 3      overall 19.928792 0.1973148
```

```
mt.data.max.generic.gammaDNoCCs %>%
    goodness.of.fit("donor.reduction") %>%
    print()
```

```
## # A tibble: 3 x 3
##   partition       SSR       MSE
##       <chr>     <dbl>     <dbl>
## 1      high  3.295153 0.1220427
## 2       low 16.633639 0.2247789
## 3   overall 19.928792 0.1973148
```

```
mt.data.max.generic.gammaDNoCCs %>%
    goodness.of.fit() %>%
    select(MSE) %>%
    as.numeric() -> MSE.generic.gammaDNoCCs

# draw a plot of predicted vs observed yields
mt.data.max.generic.gammaDNoCCs %>%
    bland.altman.plot() + ggtitle("Bland-Altman plot of the new generic model b
ased on\nthe reduction degree of the electron donor\nand the number of carbons
of the carbon source")
```



generic model based on the number of carbons of the electron donor and the reduction degree of the carbon source

```
generic.gammaCsNoCD.model <- lm(observed.yield ~ (dGcat + gammaCs + NoCD)^2, da
ta=mt.data.max)
generic.gammaCsNoCD.model %>%
    summary() %>%
    print()
```

```
##
## Call:
## lm(formula = observed.yield ~ (dGcat + gammaCs + NoCD)^2, data = mt.data.max
)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.0896 -0.2438 -0.0794  0.2212  3.5210
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -1.988e-01  1.653e-01  -1.203   0.2321
## dGcat         -9.633e-04  1.881e-04  -5.120 1.62e-06 ***
## gammaCs        1.521e-02  4.562e-02   0.333   0.7396
## NoCD           2.486e-01  8.912e-02   2.789   0.0064 **
## dGcat:gammaCs  5.678e-06  4.394e-05   0.129   0.8975
## dGcat:NoCD     1.021e-05  9.546e-06   1.069   0.2877
## gammaCs:NoCD  -2.327e-02  2.227e-02  -1.045   0.2986
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5479 on 94 degrees of freedom
## Multiple R-squared:  0.8702, Adjusted R-squared:  0.8619
## F-statistic: 105.1 on 6 and 94 DF,  p-value: < 2.2e-16
```

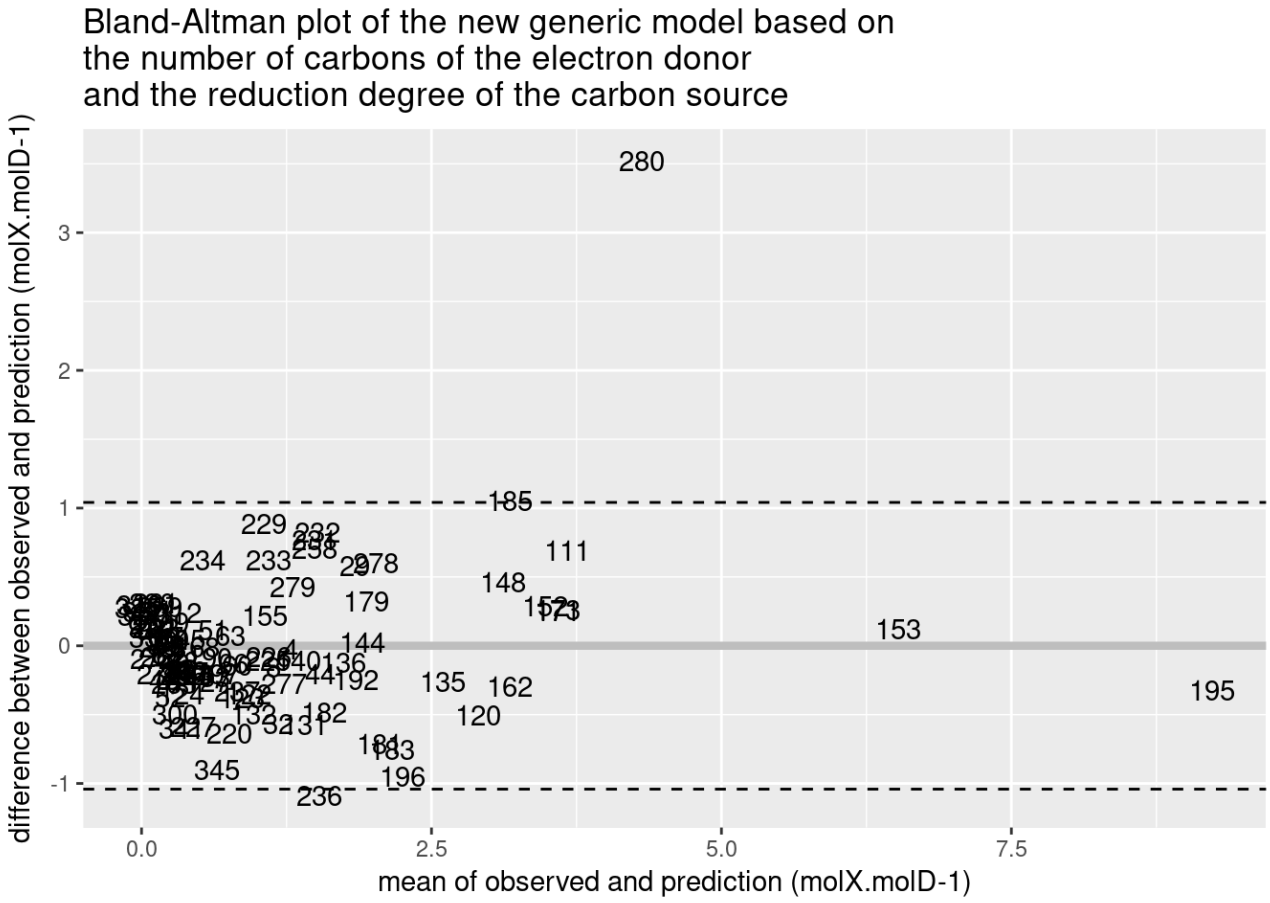
```
mt.data.max %>%
    mutate(predicted.yield = predict(generic.gammaCsNoCD.model)) -> mt.data.max
.generic.gammaCsNoCD

# goodness of fit of the model
mt.data.max.generic.gammaCsNoCD %>%
    goodness.of.fit("metabolism.type") %>%
    print()
```

```
## # A tibble: 3 x 3
##      partition       SSR       MSE
##         <fctr>     <dbl>     <dbl>
## 1 fermentation  7.793888 0.2051023
## 2  respiration 20.421295 0.3241475
## 3      overall 28.215184 0.2793583
```

```
mt.data.max.generic.gammaCsNoCD %>%
    goodness.of.fit("donor.reduction") %>%
    print()
```

```
## # A tibble: 3 x 3
##   partition      SSR       MSE
##       <chr>    <dbl>     <dbl>
## 1      high  3.74592 0.1387378
## 2       low 24.46926 0.3306657
## 3   overall 28.21518 0.2793583
```

```
mt.data.max.generic.gammaCsNoCD %>%
    goodness.of.fit() %>%
    select(MSE) %>%
    as.numeric() -> MSE.generic.gammaCsNoCD

# draw a plot of predicted vs observed yields
mt.data.max.generic.gammaCsNoCD %>%
    bland.altman.plot() + ggtitle("Bland-Altman plot of the new generic model b
ased on\nthe number of carbons of the electron donor\nand the reduction degree
of the carbon source")
```



simplification of the best generic model

simplify the generic model based on number of carbons by considering only the most significant terms

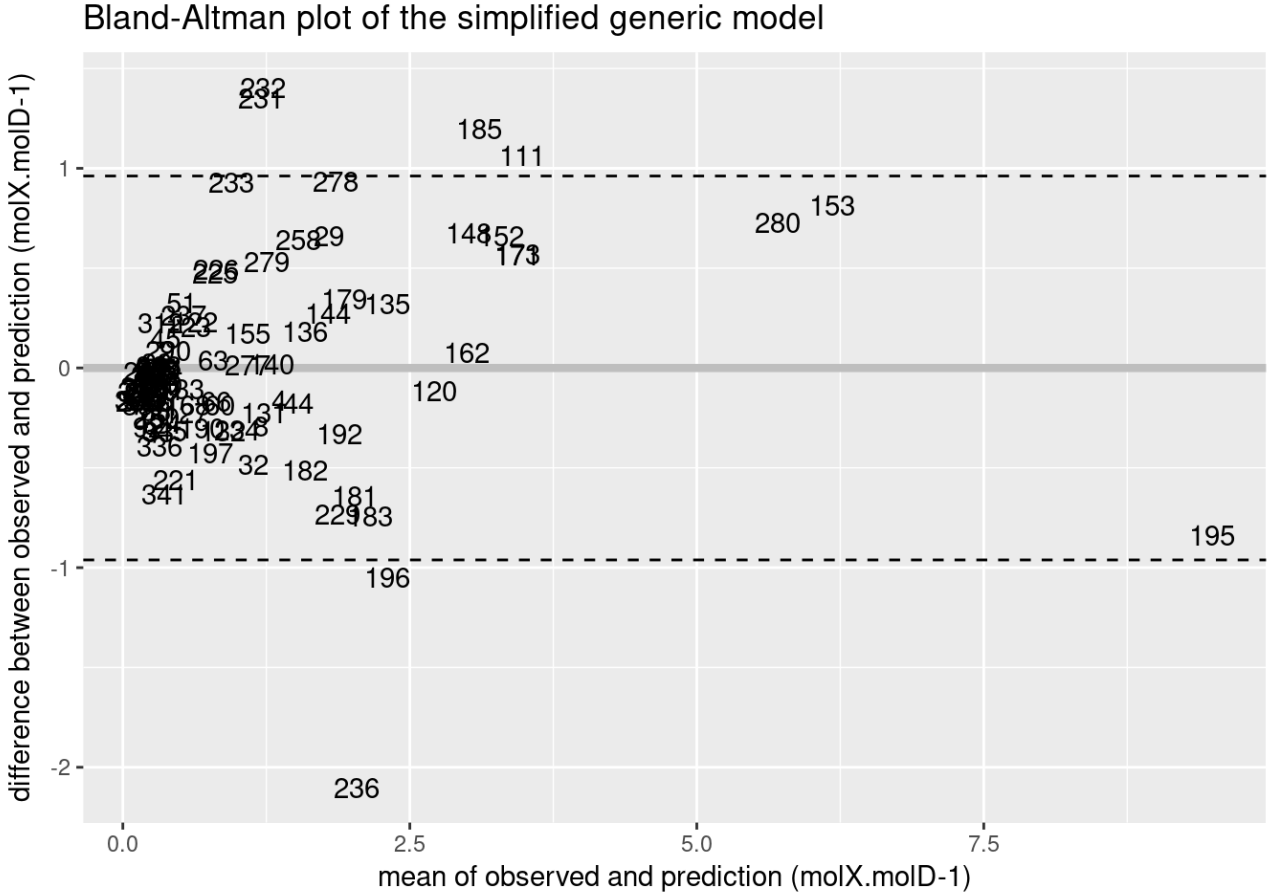
```
simplified.generic.model <- lm(observed.yield ~ dGcat + dGcat:NoCD + dGcat:NoCC
s, data=mt.data.max)
simplified.generic.model %>%
    summary() %>%
    print()
```

```
##
## Call:
## lm(formula = observed.yield ~ dGcat + dGcat:NoCD + dGcat:NoCCs,
##     data = mt.data.max)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.1017 -0.1875 -0.0829  0.1809  1.4059
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.851e-01  6.727e-02   2.752  0.00707 **
## dGcat       -9.594e-04  8.248e-05 -11.632  < 2e-16 ***
## dGcat:NoCD   2.698e-04  3.748e-05   7.198 1.30e-10 ***
## dGcat:NoCCs -2.702e-04  3.772e-05  -7.164 1.53e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.498 on 97 degrees of freedom
## Multiple R-squared:  0.8893, Adjusted R-squared:  0.8859
## F-statistic: 259.8 on 3 and 97 DF,  p-value: < 2.2e-16
```

```
mt.data.max %>%
    mutate(predicted.yield = predict(simplified.generic.model)) -> mt.data.max.
simplified.generic

# goodness of fit of the model
mt.data.max.simplified.generic %>%
    goodness.of.fit() %>%
    print()
```

```
## # A tibble: 1 x 3
##   partition      SSR       MSE
##       <chr>    <dbl>     <dbl>
## 1   overall 24.06099 0.2382276
```

```
mt.data.max.simplified.generic %>%
    goodness.of.fit() %>%
    select(MSE) %>%
    as.numeric() -> MSE.simplified.generic

# draw a plot of predicted vs observed yields
mt.data.max.simplified.generic %>%
    bland.altman.plot() + ggtitle("Bland-Altman plot of the simplified generic
model")
```



respiration/fermentation partitioned predictor

```
# respiration-specific model
mt.data.max.respiration <- filter(mt.data.max, metabolism.type == "respiration"
)
respiration.model <- lm(observed.yield ~ dGcat + dGcat:NoCD + dGcat:NoCCs, data
=mt.data.max.respiration)
respiration.model %>%
    summary() %>%
    print()
```

```
##
## Call:
## lm(formula = observed.yield ~ dGcat + dGcat:NoCD + dGcat:NoCCs,
##     data = mt.data.max.respiration)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.27317 -0.22724 -0.00283  0.20386  1.05373
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.964e-02  8.743e-02   0.339    0.736
## dGcat       -1.149e-03  8.750e-05 -13.127  < 2e-16 ***
## dGcat:NoCD   3.365e-04  3.681e-05   9.140 6.68e-13 ***
## dGcat:NoCCs -3.240e-04  3.670e-05  -8.829 2.21e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4344 on 59 degrees of freedom
## Multiple R-squared:   0.94,  Adjusted R-squared:  0.9369
## F-statistic:   308 on 3 and 59 DF,  p-value: < 2.2e-16
```

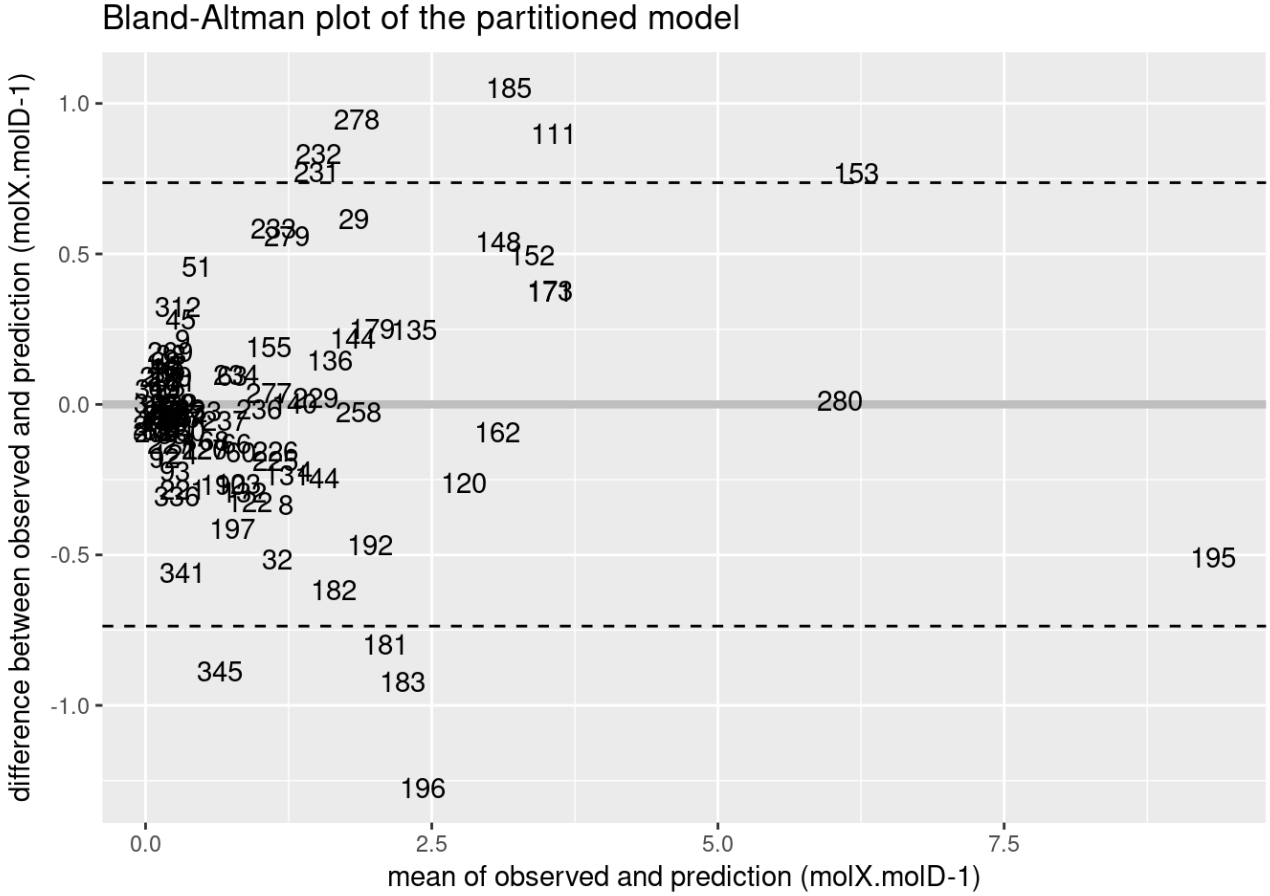
```
mt.data.max.respiration %>%
    mutate(predicted.yield = predict(respiration.model)) -> mt.data.max.respira
tion.prediction

# fermentation-specific model
mt.data.max.fermentation <- filter(mt.data.max, metabolism.type == "fermentatio
n")
fermentation.model <- lm(observed.yield ~ dGcat + dGcat:NoCD + dGcat:NoCCs, dat
a=mt.data.max.fermentation)
fermentation.model %>%
    summary() %>%
    print()
```

```
##
## Call:
## lm(formula = observed.yield ~ dGcat + dGcat:NoCD + dGcat:NoCCs,
##     data = mt.data.max.fermentation)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.88683 -0.08938 -0.03398  0.05031  0.83271
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.397e-02  6.794e-02   1.236   0.2250
## dGcat       -1.139e-04  1.577e-04  -0.722   0.4751
## dGcat:NoCD  -1.906e-04  7.514e-05  -2.537   0.0159 *
## dGcat:NoCCs -2.735e-04  6.200e-05  -4.411 9.82e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2967 on 34 degrees of freedom
## Multiple R-squared:  0.7313, Adjusted R-squared:  0.7076
## F-statistic: 30.84 on 3 and 34 DF,  p-value: 8.166e-10
```

```
mt.data.max.fermentation %>%
    mutate(predicted.yield = predict(fermentation.model)) -> mt.data.max.fermen
tation.prediction

# combine both predicted partitions
mt.data.max.partitioned.prediction <- rbind(mt.data.max.respiration.prediction,
mt.data.max.fermentation.prediction)

# goodness of fit of the model
mt.data.max.partitioned.prediction %>%
    goodness.of.fit("metabolism.type") %>%
    print()
```

```
## # A tibble: 3 x 3
##      partition       SSR        MSE
##         <fctr>     <dbl>      <dbl>
## 1 fermentation  2.993704 0.07878167
## 2  respiration 11.132951 0.17671351
## 3      overall 14.126655 0.13986787
```

```
mt.data.max.partitioned.prediction %>%
    goodness.of.fit() %>%
    select(MSE) %>%
    as.numeric() -> MSE.partitioned


# draw a plot of predicted vs observed yields
mt.data.max.partitioned.prediction %>%
    bland.altman.plot() + ggtitle("Bland-Altman plot of the partitioned model")
```




```
# oxidized-specific model
mt.data.max.oxidized <- filter(mt.data.max, donor.reduction == "low")
oxidized.model <- lm(observed.yield ~ dGcat + dGcat:NoCD + dGcat:NoCCs, data=mt
.data.max.oxidized)
oxidized.model %>%
    summary() %>%
    print()
```

```
##
## Call:
## lm(formula = observed.yield ~ dGcat + dGcat:NoCD + dGcat:NoCCs,
##     data = mt.data.max.oxidized)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.7964 -0.2010 -0.1011  0.1912  1.3491
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.166e-01  7.567e-02   2.862  0.00555 **
## dGcat       -8.230e-04  1.167e-04  -7.052 1.01e-09 ***
## dGcat:NoCD   2.285e-04  3.683e-05   6.203 3.42e-08 ***
## dGcat:NoCCs -2.638e-04  3.632e-05  -7.264 4.11e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4724 on 70 degrees of freedom
## Multiple R-squared:  0.8827, Adjusted R-squared:  0.8777
## F-statistic: 175.7 on 3 and 70 DF,  p-value: < 2.2e-16
```

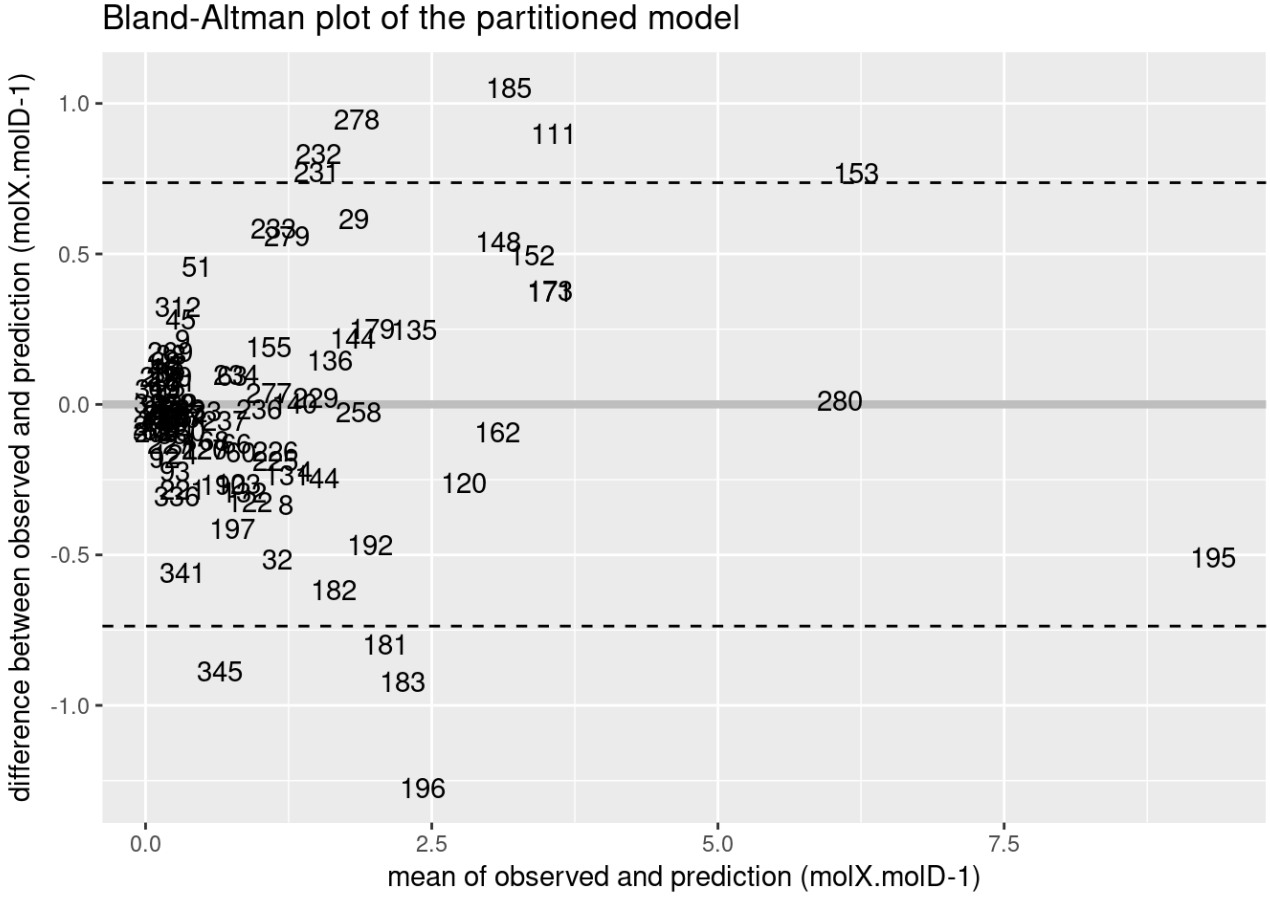
```
mt.data.max.oxidized %>%
    mutate(predicted.yield = predict(oxidized.model)) -> mt.data.max.oxidized.p
rediction

# reduced-specific model
mt.data.max.reduced <- filter(mt.data.max, donor.reduction == "high")
reduced.model <- lm(observed.yield ~ dGcat + dGcat:NoCD + dGcat:NoCCs, data=mt.
data.max.reduced)
reduced.model %>%
    summary() %>%
    print()
```

```
##
## Call:
## lm(formula = observed.yield ~ dGcat + dGcat:NoCD + dGcat:NoCCs,
##     data = mt.data.max.reduced)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.56744 -0.16578 -0.05904  0.02161  1.12817
##
## Coefficients: (1 not defined because of singularities)
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.911e-01  9.730e-02   1.964   0.0612 .
## dGcat       -7.480e-04  1.055e-04  -7.091 2.49e-07 ***
## dGcat:NoCD  -8.919e-06  6.905e-06  -1.292   0.2088
## dGcat:NoCCs         NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.374 on 24 degrees of freedom
## Multiple R-squared:  0.9601, Adjusted R-squared:  0.9568
## F-statistic: 289.1 on 2 and 24 DF,  p-value: < 2.2e-16
```

```
mt.data.max.reduced %>%
    mutate(predicted.yield = predict(reduced.model)) -> mt.data.max.reduced.pre
diction

# combine both predicted partitions
mt.data.max.partitioned.prediction.2 <- rbind(mt.data.max.oxidized.prediction,
mt.data.max.reduced.prediction)

# goodness of fit of the model
mt.data.max.partitioned.prediction.2 %>%
    goodness.of.fit("donor.reduction") %>%
    print()
```

```
## # A tibble: 3 x 3
##   partition       SSR       MSE
##       <chr>     <dbl>     <dbl>
## 1      high  3.356612 0.1243190
## 2       low 15.619812 0.2110785
## 3   overall 18.976424 0.1878854
```

```
mt.data.max.partitioned.prediction.2 %>%
    goodness.of.fit() %>%
    select(MSE) %>%
    as.numeric() -> MSE.partitioned.2


# draw a plot of predicted vs observed yields
mt.data.max.partitioned.prediction %>%
    bland.altman.plot() + ggtitle("Bland-Altman plot of the partitioned model")
```



models performances summary

```
data.frame(model=c("Heijnen",
                   "Liu",
                   "Roden",
                   "generic using reduction degrees",
                   "generic using number of carbons",
                   "simplified generic",
                   "respiration/fermentation partitioned model",
                   "donor reduction partitioned model"
                   ),
           MSE=c(MSE.heijnen,
                 MSE.liu,
                 MSE.roden,
                 MSE.generic.gamma,
                 MSE.generic.carbon,
                 MSE.simplified.generic,
                 MSE.partitioned,
                 MSE.partitioned.2
                 )) %>% kable()
```

model	MSE
Heijnen	0.1842357
Liu	0.1507557
Roden	0.4066362
generic using reduction degrees	0.2845126
generic using number of carbons	0.1502709
simplified generic	0.2382276
respiration/fermentation partitioned model	0.1398679
donor reduction partitioned model	0.1878854

Résumé substantiel de la thèse

Introduction

L'objectif de cette thèse est d'étudier le modèle MTS (*Microbial Transition State*), un modèle destiné à prédire les dynamiques de croissance de populations microbiennes à partir de principes premiers. Dans un premier temps, les caractéristiques découlant de l'équation du modèle furent étudiées. Dans un deuxième temps, des motifs de croissance et de structuration d'écosystèmes microbiens pour lesquels le modèle MTS constitue une explication suffisante ont été recherchés.

Ce modèle est pensé comme un outil permettant l'étude du lien théorique existant entre la cinétique de la production de biomasse par une population de microbes et les gradients d'énergie présents dans le milieu.

Le modèle MTS est basé sur une description microscopique de la croissance microbienne. Il suppose que les cellules d'une population microbienne, ainsi que les réactifs impliqués dans leur réaction de croissance, peuvent être pensées comme des points uniformément répartis dans leur milieu de culture. L'utilisation d'un modèle de bilan d'énergie, tel que celui développé par Heijnen et collaborateurs (Heijnen et al., 1991) permet d'estimer combien de particules de substrat sont nécessaires pour subvenir aux besoins en énergie et en nutriments pour répliquer la biomasse. Il est considéré qu'une population de microbe catalyse deux réactions ; une réaction catabolique et une réaction anabolique. La réaction catabolique est un transfert d'électron entre un donneur d'électrons D et un accepteur d'électrons A. La réaction anabolique consiste en la synthèse d'une molécule de biomasse générique. On considère en effet une molécule de biomasse fictive, normalisée par atome de carbone et reflétant la composition moyenne d'une cellule, afin de pouvoir écrire la réaction anabolique de façon stoechiométriquement explicite, et pouvoir exprimer le taux de croissance de la population en carbone de biomasse produite par unité de temps. Le modèle de bilan d'énergie de la croissance introduit par Heijnen et collaborateurs suppose que la réaction catabolique produit de l'énergie, dont une partie est utilisée pour rendre la réaction anabolique thermodynamiquement spontanée, et une autre partie est dissipée. La dissipation désigne l'énergie capturée par la cellule (par le catabolisme) qui n'est pas stockée sous forme de biomasse. Heijnen et collaborateurs proposent une formule empirique pour estimer la quantité d'énergie dissipée par unité de biomasse produite. Le bilan d'énergie de la croissance cellulaire est alors fermé en calculant le nombre de fois λ (en $mol_D.mol_X^{-1}$) que la réaction catabolique se produit par réaction anabolique. Ce nombre de fois s'exprime

$$\lambda = \frac{\Delta G_{dis} - \Delta G_{an}}{\Delta G_{cat}}$$	1

avec ΔG_{cat} et ΔG_{an} les différentiels d'énergies de Gibbs de la réaction catabolique et anabolique en $J.mol_D^{-1}$ et $J.mol_X^{-1}$ respectivement, et ΔG_{dis} l'énergie dissipée en $J.mol_X^{-1}$ telle que prédite par la formule proposée par Heijnen et collaborateurs.

Le coefficient stœchiométrique de chaque espèce chimique impliquée dans la réaction de croissance devient alors

$Y_{Si}^{met} = Y_{Si}^{an} + \lambda \cdot Y_{Si}^{cat}$	2

Avec Y_{Si}^{an} le coefficient stœchiométrique de l'espèce chimique S_i dans l'anabolisme (en $mol.mol_X^{-1}$), Y_{Si}^{cat} le coefficient stœchiométrique de l'espèce chimique S_i dans le catabolisme (en $mol.mol_D^{-1}$) et Y_{Si}^{met} le coefficient stœchiométrique de l'espece chimique Si dans la réaction de croissance (en $mol.mol_X^{-1}$) équilibrée de sorte à ce que son différentiel d'énergie de Gibbs soit égal à la valeur ΔG_{dis} issue de la formule de Heijnen de prédiction de l'énergie dissipée.

Le fondement de l'approche du modèle MTS consiste à considérer que les cellules sont entourées d'un volume de récolte fictif appelé V_h (*Harvest Volume*) dans lequel les particules de substrat sont accessibles à la cellule. Un raisonnement purement probabiliste permet alors d'exprimer mathématiquement la proportion de cellules ayant suffisamment de substrat dans leur volume de récolte pour pouvoir se répliquer. De cette formule peut être dérivé le taux de croissance de la population de microbes.

$\mu = \mu_{max} \cdot \prod_i e^{\frac{Y_{Si}^{met}}{V_h \cdot [S_i]}}$	3

avec μ le taux de croissance en $jour^{-1}$, μ_{max} le taux de croissance maximal en $jour^{-1}$, V_h le volume de récolte en $m^3.molX^{-1}$, Y^{met}_{Si} le coefficient stœchiométrique de la ressource Si en $mol_S.mol_X^{-1}$ (négatif) et [Si] la concentration de la ressource dans le milieu de culture en $mol_S.m^{-3}$.

Cette fonction de croissance donne lieu à une courbe de croissance en fonction du substrat différente de celle qui peut être obtenue avec la fonction de croissance de Monod (figure 1). On observe en effet sur la figure 1 que la fonction de croissance du modèle MTS a une forme sigmoïde, ce qui n'est pas le cas de celle du modèle de Monod. On note également que le « pied » de la sigmoïde du modèle MTS tend à devenir négligeable lorsque la valeur du paramètre V_h est grande (autour de 10 $m^3.mol^{-1}$).

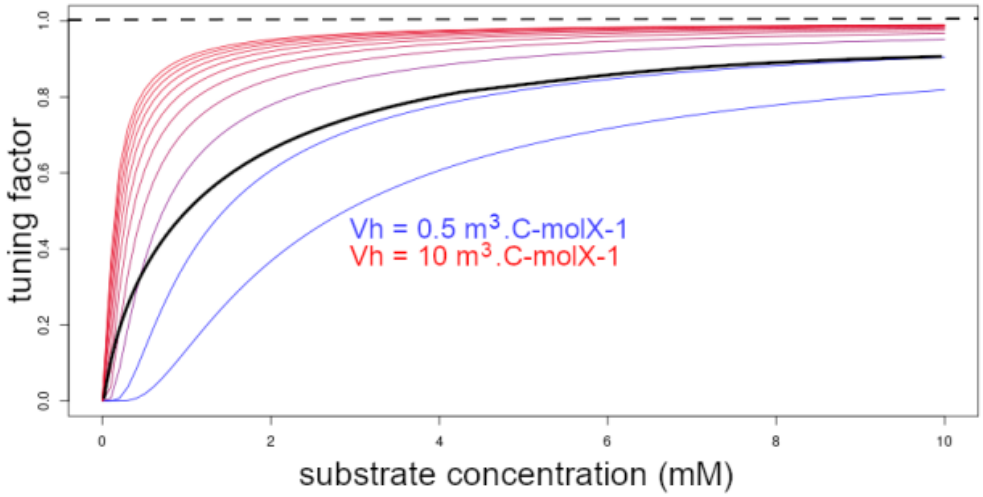


Figure 1 : relation entre la concentration de substrat [S] et la valeur de l'exponentielle négative associée dans le modèle MTS ($e^{-\frac{1}{[S]}}$) (courbes bleues et rouges), et comparaison avec la relation équivalente dans la fonction de croissance de Monod ($\frac{[S]}{K_S+[S]}$) (courbe noire). L'axe Y peut être interprété comme μ/μmax pour une fonction de croissance monosubstrat.

Résultats et discussion

Reproduction de la loi de Liebig

Le modèle MTS exprime le taux de croissance d'une population microbienne comme étant fonction de la concentration de chaque molécule consommée dans la réaction de croissance. Au contraire, la plupart des autres modèles de croissance microbienne basés sur des formules de taux de croissance empiriques (telle que celle de Monod) considère que la croissance n'est limitée que par une seule molécule (le substrat limitant). De nombreuses propositions ont été faites pour modéliser la limitation de la croissance par plusieurs substrats, néanmoins aucune d'entre elles n'a atteint le consensus (Bungay et al., 1994).

La fonction de croissance de Monod (de même que d'autres telles que celle de Contois (Contois, 1959)) supposent implicitement que la loi de Liebig s'applique. Cette loi empirique, initialement formulée pour la croissance des plantes, suppose que la croissance d'un organisme n'est limitée, à un instant donné, que par la ressource la plus rare. Ce principe est très souvent considéré en écologie, à la fois à l'échelle de l'organisme, de la population ou de la communauté. Il a néanmoins récemment été montré que cette loi ne s'applique pas toujours à l'échelle des populations et des communautés (Danger et al., 2008).

La fonction de croissance du modèle MTS apporte une contribution intéressante sur ce sujet. En effet, la loi de Liebig n'est pas présupposée par le modèle MTS. D'après la fonction de croissance propre au modèle MTS, il est mathématiquement possible que deux ressources exercent, à un instant donné, le même niveau de limitation sur la croissance d'une population.

Cependant il est observé dans les simulations du modèle que les niveaux de limitation exercés par les ressources sur le taux de croissance d'une population sont très différents. À un instant donné, une

ressource est beaucoup plus limitante que les autres. Cela est dû au fait que la limitation exercée par une ressource est de nature exponentielle ; ainsi un faible changement de concentration peut donner lieu à un changement important de la limitation exercé par la ressource.

Un tel phénomène est illustré par une simulation au cours de laquelle une population glucosotrophe aérobie est simulée dans un batch aéré. Les équations de croissance de la population sont;

- Catabolisme: $C_6H_{12}O_6 + 6\ O_2 \rightarrow 6\ HCO_3^- + 6\ H^+$ ($\Delta G^{0'} = -2841.3$ kJ.mol$_D^{-1}$)
- Anabolisme: $0.167\ C_6H_{12}O_6 + 0.158\ NH_4^+ \rightarrow 0.430\ H_2O + 0.164\ H^+ + 0.00625\ HCO_3^- + C_1H_{1.613}O_{0.557}N_{0.158}$ ($\Delta G^{0'} = -28.3$ kJ.C-mol-X^{-1})
- Energie dissipée: 236.05 kJ.C-mol-X^{-1}

Les dynamiques des concentrations sont montrés par les figures 2 et 3.

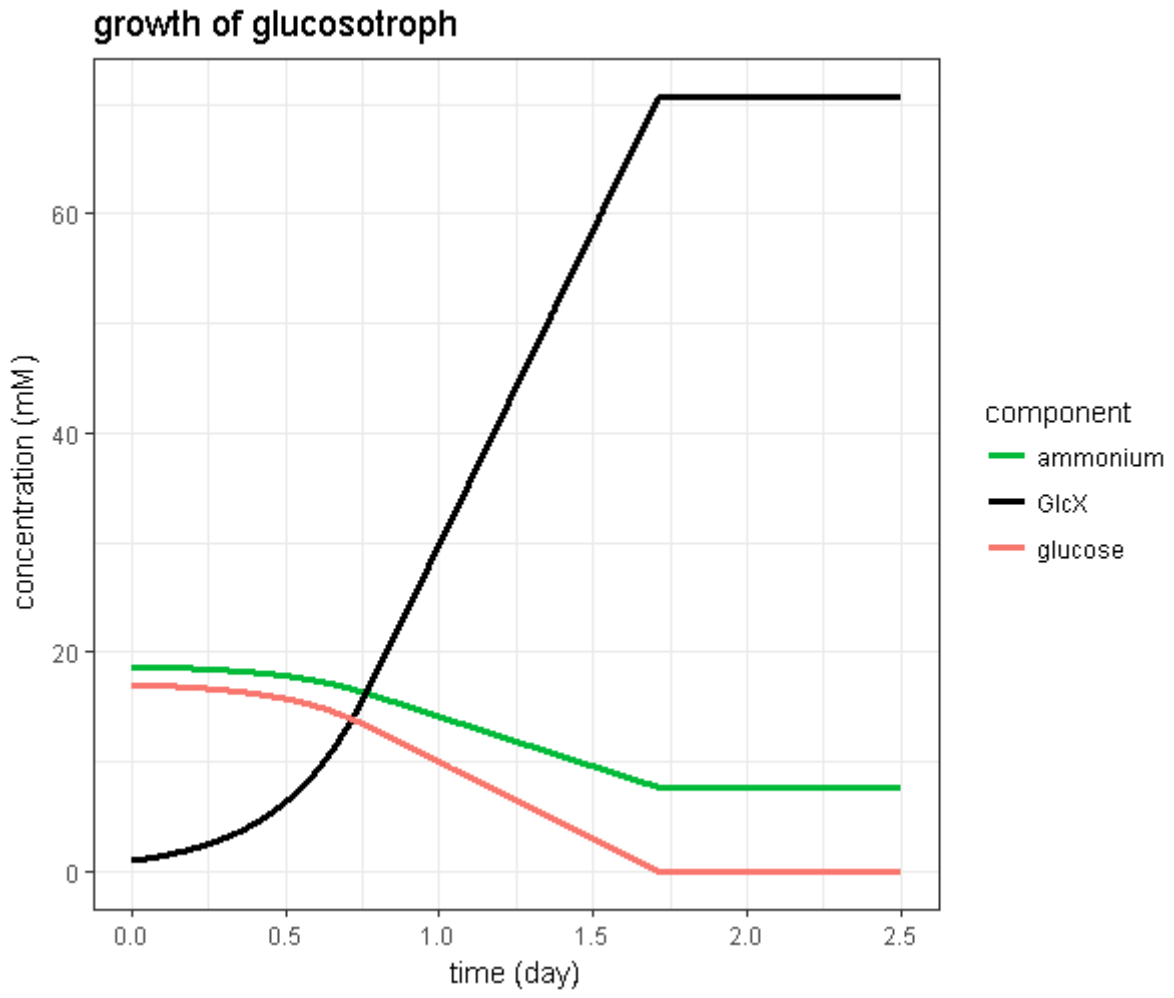


Figure 2 : Concentration d'espèces chimiques au cours de la simulation de la croissance d'une population glucosotrophe par le modèle MTS. La courbe rose représente la concentration en glucose, la courbe verte représente la concentration en ammonium et la courbe noire représente la concentration en biomasse de la population. Les concentrations sont exprimées en mM et le temps en jour.

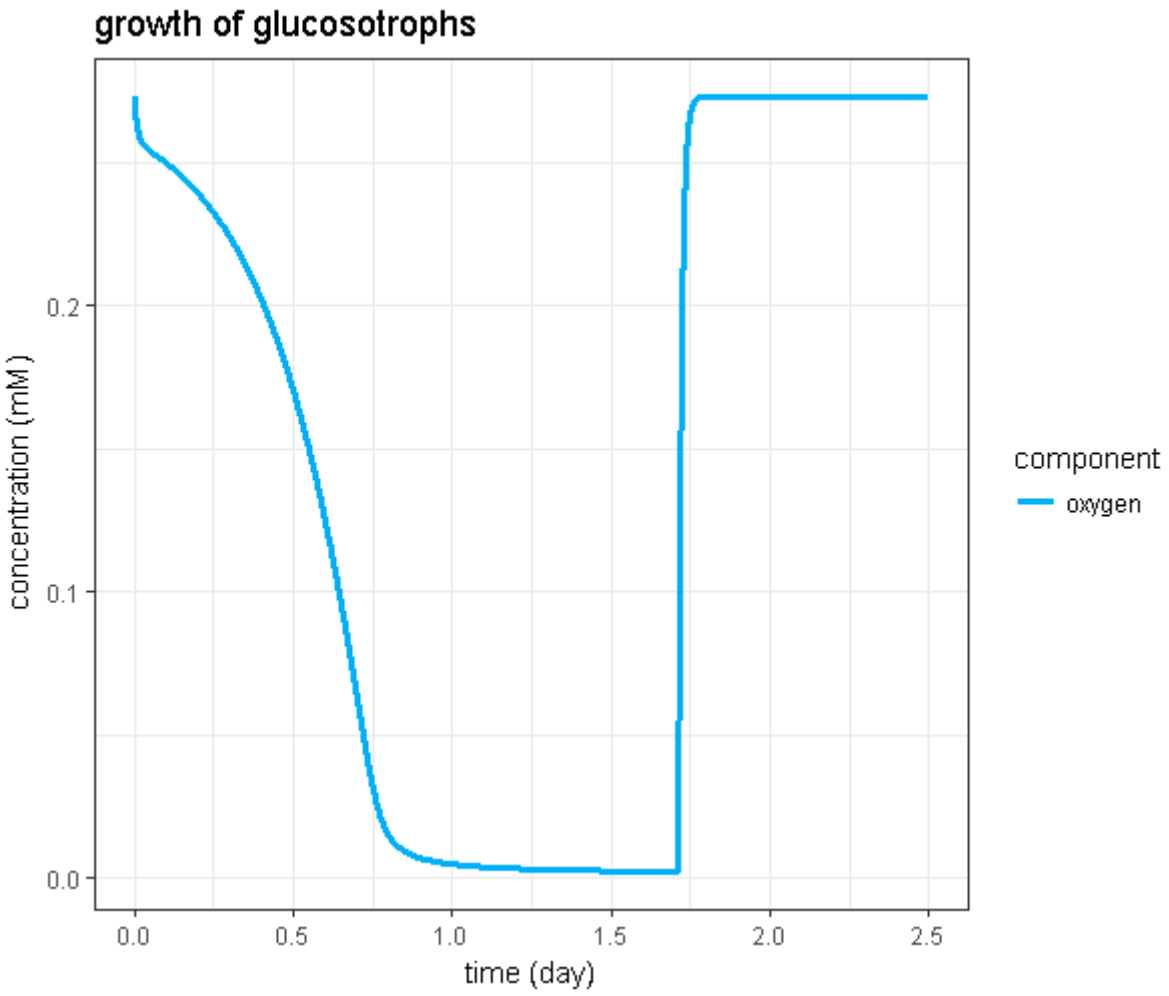


Figure 3 : Concentration d'oxygène au cours de la simulation de la croissance d'une population glucosotrophe par le modèle MTS. La concentration est exprimée en mM et le temps en jour.

Les limitations associées à chaque ressource consommée par la population au cours du temps sont montrées dans la figure 4.

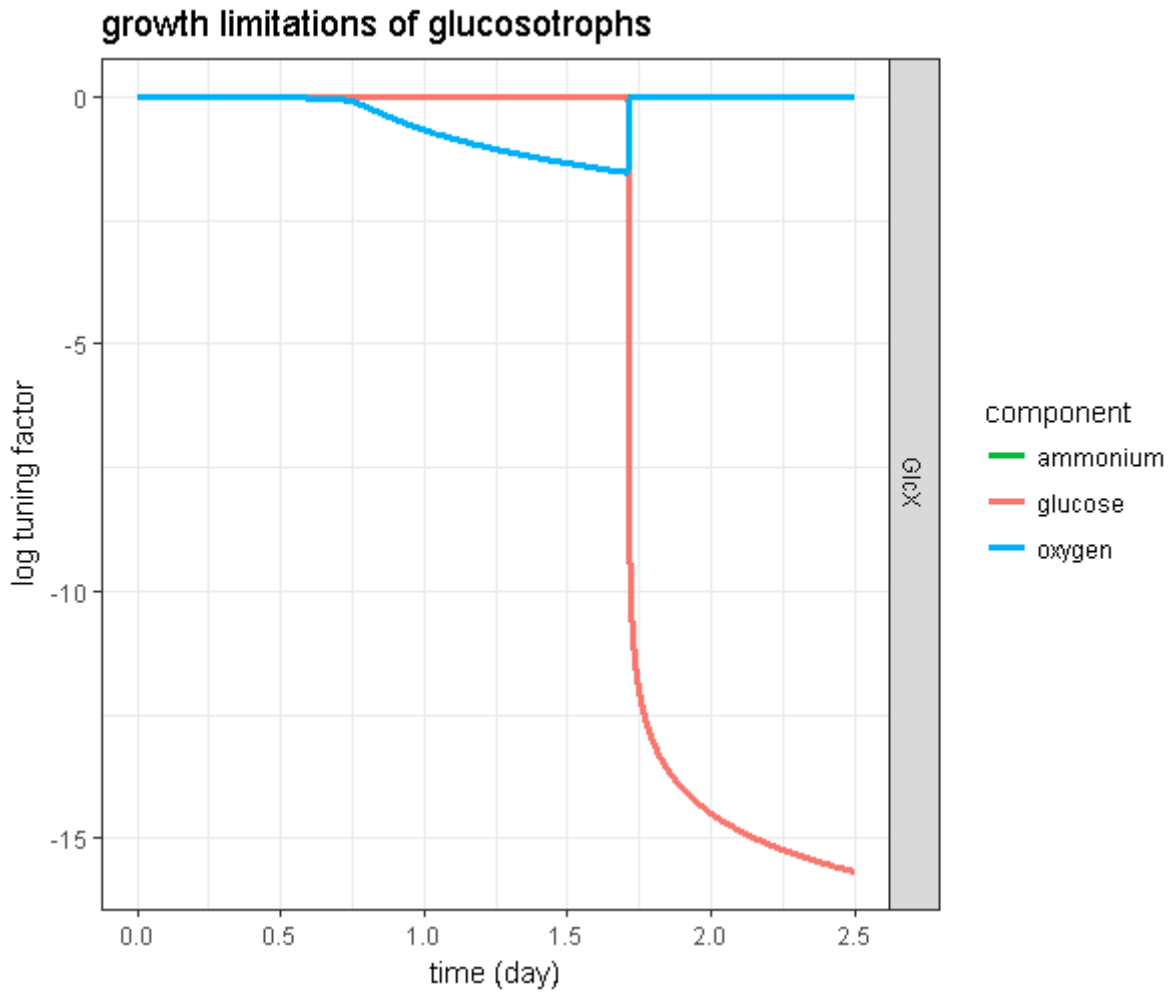


Figure 4 : valeur des exponentielles négatives associées à chaque molécule requise par la croissance de la population glucosotrophe telle que simulée avec le modèle MTS. La courbe verte représente l'ammonium, la courbe rose représente le glucose et la courbe bleue représente l'oxygène. L'axe des Y est en log. Le temps est exprimé en jours.

On peut voir dans ce résultat que la croissance se produit en deux phases distinctes, une première limitée par l'oxygène et une seconde limitée par le glucose.

Le fait que la fonction de croissance du modèle MTS donne lieu à des motifs de croissance évoquant la loi de Liebig suggère que les principes premiers du modèle MTS sont une possible explication à l'observation de la loi de Liebig dans les motifs de croissance des populations microbiennes. En même temps, le modèle MTS n'est pas strictement tenu à la reproduction de la loi de Liebig et est également capable de produire des résultats plus nuancés montrant des colimitations de la croissance par plusieurs ressources.

Reproduction d'une redox tower

Un des résultats les plus importants obtenus avec le modèle MTS durant cette thèse est la reproduction d'une "redox tower".

Une redox tower est le nom consacré d'une succession écologique déterminée par le potentiel redox des substrats dans le milieu de culture. La redox tower se produit lorsque toutes les guildes sont en compétition pour une ressource, telle qu'un donneur d'électron. La population qui catalyse la

réaction ayant le potentiel redox le plus élevé en valeur absolu exclue alors compétitivement les autres populations, ce qui donne lieu à la succession écologique.

Un résultat exposé dans le chapitre 4 de cette thèse reproduit une redox tower. Dans ce résultat, une série de simulations de culture en chemostat est effectuée. Un seul donneur d'électron (l'acétate) est présent et quatre populations microbiennes sont simulées, catalysant le transfert des électrons de l'acétate vers l'oxygène, le nitrate, le fer et le sulfate, respectivement.

Les équations métaboliques de chaque population sont les suivantes ;

- Population aérobie
  - Catabolism: $C_2H_3O_2^- + 2\ O_2 \rightarrow + 2\ HCO_3^- + 1\ H^+$ ($\Delta G^{0'} = -844.4\ kJ.mol_D^{-1}$)
  - Anabolisme: $0.503\ C_2H_3O_2^- + 0.158\ NH_4^+ + 0.338\ H^+ \rightarrow 0.4305\ H_2O + 0.0063\ HCO_3^- + C_1H_{1.613}O_{0.557}N_{0.158}$ ($\Delta G^{0'} = 23.9\ kJ.C\text{-}mol\text{-}X^{-1}$)
  - Energie dissipée: $432.12\ kJ.C\text{-}mol\text{-}X^{-1}$
- Population dénitrifiante
  - Catabolisme: $C_2H_3O_2^- + 1.6\ NO_3^- + 0.6\ H^+ \rightarrow 0.8\ N_2 + 2\ HCO_3^- + 0.8\ H_2O$ ($\Delta G^{0'} = -792.1\ kJ.mol_D^{-1}$)
  - Anabolisme: $0.503\ C_2H_3O_2^- + 0.158\ NH_4^+ + 0.338\ H^+ \rightarrow 0.4305\ H_2O + 0.0063\ HCO_3^- + C_1H_{1.613}O_{0.557}N_{0.158}$ ($\Delta G^{0'} = 23.9\ kJ.C\text{-}mol\text{-}X^{-1}$)
  - Energie dissipée: $432.12\ kJ.C\text{-}mol\text{-}X^{-1}$
- Population ferroreductrice
  - Catabolisme: $C_2H_3O_2^- + 4\ H_2O + 8\ Fe^{+3} \rightarrow 9\ H^+ + 2\ HCO_3^- + 8\ Fe^{+2}$ ($\Delta G^{0'} = -809.6\ kJ.mol_D^{-1}$)
  - Anabolisme: $0.503\ C_2H_3O_2^- + 0.158\ NH_4^+ + 0.338\ H^+ \rightarrow 0.4305\ H_2O + 0.0063\ HCO_3^- + C_1H_{1.613}O_{0.557}N_{0.158}$ ($\Delta G^{0'} = 23.9\ kJ.C\text{-}mol\text{-}X^{-1}$)
  - Energie dissipée: $432.12\ kJ.C\text{-}mol\text{-}X^{-1}$
- Population sulfatoreductrice
  - Catabolisme: $1\ C_2H_3O_2^- + 1\ SO_4^{-2} \rightarrow 2\ HCO_3^- + 1\ HS^-$ ($\Delta G^{0'} = -47.7\ kJ.mol_D^{-1}$)
  - Anabolisme: $0.503\ C_2H_3O_2^- + 0.158\ NH_4^+ + 0.338\ H^+ \rightarrow 0.4305\ H_2O + 0.0063\ HCO_3^- + C_1H_{1.613}O_{0.557}N_{0.158}$ ($\Delta G^{0'} = 23.9\ kJ.C\text{-}mol\text{-}X^{-1}$)
  - Energie dissipée: $432.12\ kJ.C\text{-}mol\text{-}X^{-1}$

Dans chaque simulation, la concentration des accepteurs d'électron est toujours la même, seul la concentration du donneur d'électron (acétate) change. La proportion de chaque population (en terme de concentration de biomasse) présente lors de l'état stable du système est alors mesurée, et mise en relation avec la concentration de donneur d'électron en entrée du chemostat (voir figure 5).

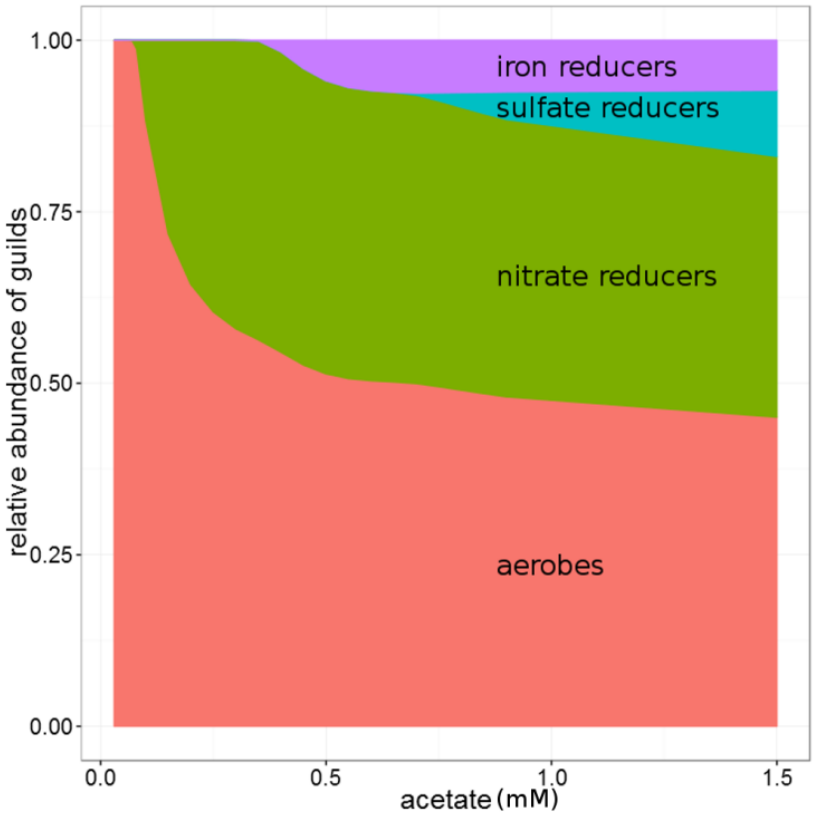


Figure 5 : concentration relative de biomasse de chaque population à l'état stable du chémostat, tel que prédite par le modèle MTS, en fonction de la concentration d'acétate en entrée (en mM). L'aire rose représente la population aérobie, l'aire verte représente la population dénitrifiante, l'aire violette représente la population ferroreductrice, et l'aire bleue représente la population sulfatoreductrice.

Il apparait alors que l'ordre dans lequel les populations microbiennes arrivent à se stabiliser dans le chémostat correspond à l'ordre qui pouvait déjà être prédit d'après le différentiel d'énergie de Gibbs de chaque catabolisme. Aux plus faibles concentrations de donneur d'électron, la population aerobie, qui catalyse le différentiel d'énergie de Gibbs le plus important en valeur absolue, exclue compétitivement toutes les autres populations. Lorsqu'il commence à y avoir trop de donneur d'électron en entrée pour que tous les molécules soient oxydées par les aerobies, la population ayant le deuxième métabolisme le plus avantageux thermodynamiquement ; les dénitrifiantes ; arrive à se stabiliser en consommant les molécules de donneur d'électron restantes. Ainsi de suite, les différentes populations parviennent à se stabiliser dans le chémostat. Ce résultat montre donc que l'issue des competitions entre les populations, telle que prédite par le modèle MTS, reflete une succession écologique connue et prédictible par la thermodynamique.

Ce redox tower phenomenon est une phénomène connu et observé depuis une cinquantaine d'années. D'autres modèles ont déjà montré qu'il était possible de le reproduire à l'aide d'expressions phénoménologiques du taux de croissance microbien couplé à des expressions du rendement de croissance calculé à partir du modèle thermodynamique de Heijnen et collaborateurs

(Gonzalez-Cabaleiro et al., 2015). L'intérêt du résultat de reproduction d'une redox tower par le modèle MTS, tel que présenté dans ce mémoire, n'est donc pas de montrer que de telles successions écologiques s'expliquent à partir de grandeurs thermodynamiques. En effet, cela a déjà été montré par des approches précédentes. L'intérêt du résultat présenté dans ce mémoire est de montrer que l'équation de croissance propre au modèle MTS est à même de produire de telles successions écologiques. Cette équation de croissance se démarque en effet des équations phénoménologiques dans la mesure où elle se déduit d'hypothèses fondamentales sur la division cellulaire, à l'échelle microscopique. Un tel résultat donne donc du support aux principes premiers sur lesquels est fondée l'approche MTS.

Première estimation de la valeur des paramètres du modèle MTS

Durant la première partie de la thèse, le modèle MTS a été simulé en considérant des valeurs de paramètres cinétiques μ_{max} et V_h par défaut, issues de considérations théoriques. Ces valeurs sont les suivantes ;

$$\begin{cases} \mu_{max} = \dfrac{k_B T}{h} = 5.35e17\ day^{-1} \\ V_h = 1\ m^3.mol^{-1} \end{cases}$$

Avec $\frac{k_B T}{h}$ la constant de Boltzman multipliée par la temperature divisée par la constant de Planck, par analogie avec la fréquence maximum de réaction du modèle des états de transition définition par Eyring (Eyring, 1935), sur lequel est basé le modèle MTS.

De telles valeurs ont été choisies car il n'existait pas, à ce moment, de référence quant à la valeur que doivent prendre ces paramètres. De plus, il a dabord été considéré que toutes les populations avaient les mêmes valeurs de μ_{max} et V_h pour des raisons de parcimonie. Afin d'obtenir une première estimation de l'ordre de grandeur de ces paramètres, les paramètres du modèle MTS ont été calibrés sur des données expérimentales.

Les données expérimentales sur lesquelles les paramètres du modèle MTS a été calibré sont des données de respirométrie de boue activée. Une expérience de respirométrie consiste à introduire un donneur d'électron dans une culture microbienne, et à mesurer la consommation d'oxygène au cours du temps, occasionnée par la consommation du substrat par les populations microbiennes. Des populations microbiennes de boue activée ont donc été formalisées de façon à ce que leur croissance puisse être modélisée par le modèle MTS ;

- Ordinary Heterotroph Bacteria (OHO)
    - Catabolism: $C_2H_3O_2^-$ + 2 O_2 → + 2 HCO_3^- + 1 H^+ ($\Delta G^{0'}$ = -844.4 kJ.mol_D^{-1})
    - Anabolisme: 0.503 $C_2H_3O_2^-$ + 0.158 NH_4^+ + 0.338 H^+ → 0.4305 H_2O + 0.0063 HCO_3^- + $C_1H_{1.613}O_{0.557}N_{0.158}$ ($\Delta G^{0'}$ = 23.9 kJ.C-mol-X^{-1})
    - Energie dissipée: 432.12 kJ.C-mol-X^{-1}
- Ammonium Oxidizing Bacteria (AOB):
    - Catabolisme: NH_4^+ + 1.5 O_2 → 1 NO_2^- + 1 H_2O + 2 H^+ ($\Delta G^{0'}$ = -269.9 kJ.mol_D^{-1})

- Anabolisme: HCO_3^- + 0.828 NH_4^+ → 1.101 H_2O + 0.670 NO_2^- + 0.499 H^+ + $C_1H_{1.613}O_{0.557}N_{0.158}$ ($\Delta G^{0'}$ = 267.7 kJ.C-mol-X^{-1})
- Energie dissipée: 3500 kJ.C-mol-X^{-1}

- Nitrite Oxidizing Bacteria (NOB);
  - Catabolisme: $NO2^-$ + 0.5 O_2 → 1 NO_3^- ($\Delta G^{0'}$ = -79.1 kJ.mol_D^{-1})
  - Anabolisme: HCO_3^- + 2.64 NO_2^- + 1.16 H^+ → 0.27 H_2O + 2.49 NO_3^- + $C_1H_{1.613}O_{0.557}N_{0.158}$ ($\Delta G^{0'}$ = 241.3 kJ.C-mol-X^{-1})
  - Energie dissipée: 3500 kJ.C-mol-X^{-1}

Deux expériences de respirométrie ont été effectuées. Dans une première expérience, de l'acétate est ajouté à l'échantillon de boue activée. De l'AllylThioUrée est préalablement ajouté au milieu de culture, de façon à empêcher la croissance des populations autotrophes (AOB et NOB), de façon à ce que la consommation d'oxygène soit le seul fait de la population hétérotrophe (OHO). Dans la deuxième expérience, de l'ammonium est ajouté à l'échantillon de boue activée ; la consommation d'oxygène est alors le fait des populations autotrophes (AOB et NOB). Toutes les expériences ont été réalisées par E. Paul et ses étudiants, de LISBP Toulouse.
Le calage des paramètres cinétiques du modèle MTS sur ces données expérimentales permet alors de définir une valeur de μ_{max} et de V_h d'une part pour les OHO et d'autre part pour les AOB et les NOB. Le calage consiste à déterminer le couple de valeurs de μ_{max} et V_h qui minimise la somme du carré des différences entre la consommation d'oxygène au cours du temps prédite par MTS et celle expérimentalement mesurée. Il est à noter qu'il s'agit d'un calage partiel dans la mesure où les populations AOB et NOB se voient attribuer la même valeur pour chaque paramètre cinétique, et une seule valeur de V_h est considérée pour tous les substrats d'une population. La calibration du plus petit nombre possible de paramètres a été choisie pour conserver la parcimonie du modèle.

Les valeurs suivantes de paramètre ont alors été déterminées par calage pour la population OHO ;

$$\begin{cases} \mu_{max} = 3.75\ day^{-1} \\ V_h = 147\ m^3.mol^{-1} \end{cases}$$

Et les valeurs suivantes ont été déterminées pour les populations autotrophes ;

$$\begin{cases} \mu_{max} = 2.08\ day^{-1} \\ V_h = 296\ m^3.mol^{-1} \end{cases}$$

Afin d'évaluer la sensibilité des prédictions du modèle MTS a la valeur de ses paramètres, la somme des carrés des écarts entre la prédiction MTS et les données respirométriques a été calculé en fonction des valeurs de μ_{max} et V_h (figure 6).

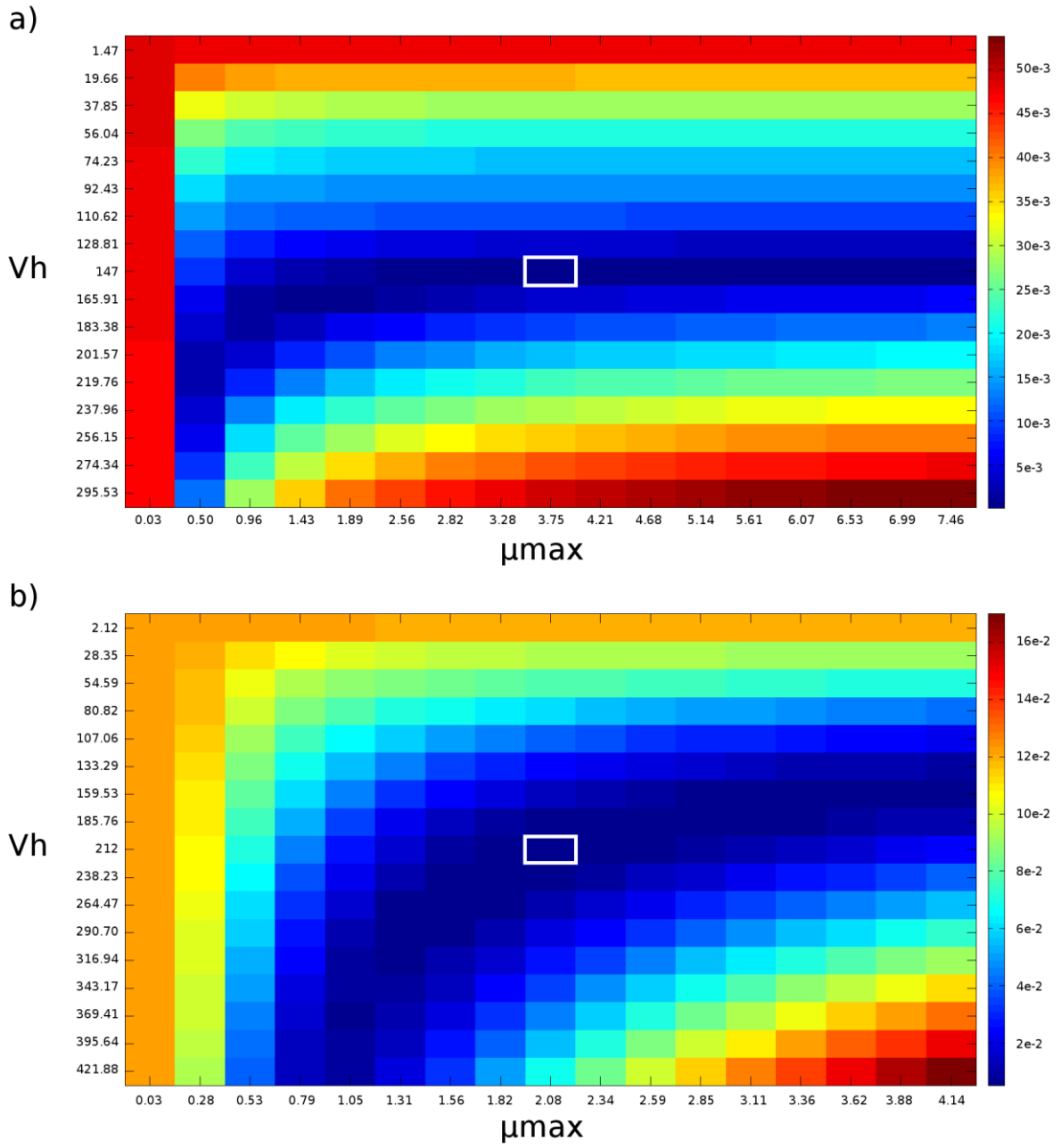


Figure 6 : Somme des carrés des écarts entre les données respirométriques et les prédictions du modèle MTS en fonction de la valeur de µmax et Vh considérée pour le modèle MTS. L'axe des abscisses correspond au paramètre µmax et l'axe des ordonnées correspond au paramètre Vh. Le centre de chaque graph correspond au couple de paramètres calibrés. a) Données respirométriques après ajout d'acétate. b) Données respirométriques après ajout d'ammonium.

Pour chaque cas de figure, il est visible sur la figure 6 que le couple (μ_{max} ; V_h) minimisant les carrés des écarts entre la prédiction faite par le modèle MTS et les données respirométriques se trouve dans une « vallée » de très faible déclivité. Pour chaque population ou groupe de populations testé, un couple (μ_{max} ; V_h) optimal existe bel et bien, cependant il existe une incertitude sur la valeur exacte de ces paramètres.

Cette incertitude qui existe sur la valeur des paramètres du modèle MTS indique que le modèle est peu sensible à la valeur de ses paramètres cinétiques. Dans le modèle MTS, le rendement de croissance des populations est déterminé par la thermodynamique (via la méthode de Heijnen et collaborateurs), et la relation liant le taux de croissance aux gradients d'énergie présents dans le

milieu de culture dépend des principes premiers sur lesquels est définit le modèle MTS. Ces deux éléments représentent l'influence des gradients d'énergie sur les dynamiques de croissance ; ils ne peuvent pas faire l'objet d'une calibration. Bien que les paramètres cinétiques μ_{max} et V_h sont supposés avoir un sens physique, celui-ci n'est pas encore suffisamment bien définit pour que la valeur de ces paramètres puisse être déterminé a priori, en amont de toute expérience. Par conséquent, ils représentent la part « calibrable » du modèle MTS, et donc la part des dynamiques qui n'est pas directement imputable à la thermodynamique. Le constat de la faible sensibilité des prédictions du modèle MTS à la valeur des paramètres μ_{max} et V_h indique alors que la « résiduelle » qu'ils représentent (par rapport à l'explication thermodynamique apportée par la partie non-calibrable du modèle) a peu d'influence sur les dynamiques du système simulé.

Afin d'estimer l'impact de la valeur des paramètres cinétiques sur les prédictions du modèle MTS pour un cas concret ; une simulation de croissance des populations d'une boue activée en batch aéré a été effectuée avec le modèle MTS en considérant le jeu de paramètres par défaut puis le jeu de paramètres partiellement calibrés sur les données respirométriques. Ces deux simulations ont été comparées avec le même système simulé à partir des équations de croissance du modèle ASMN (Hiatt and Grady, 2008).

Le modèle ASMN est une version modifiée du modèle ASM pour la simulation des dynamiques de population d'une boue activée en incluant les deux populations AOB et NOB distinctes (elles ne sont pas distinctes dans le modèle ASM classique). Il est à noter que le modèle ASMN (de même que les autres itérations du modèle ASM) comportent normalement des dynamiques additionnelles décrivant des phénomènes d'hydrolyse et de mortalité cellulaires, qui sont exclues de la simulation effectuée ici. Il s'agit donc de comparer les simulations du modèle MTS avec une version dégradée du modèle ASMN, ne comprenant que ses formules de croissance, afin de constituer un modèle de référence. Les résultats sont visibles sur la figure 7.

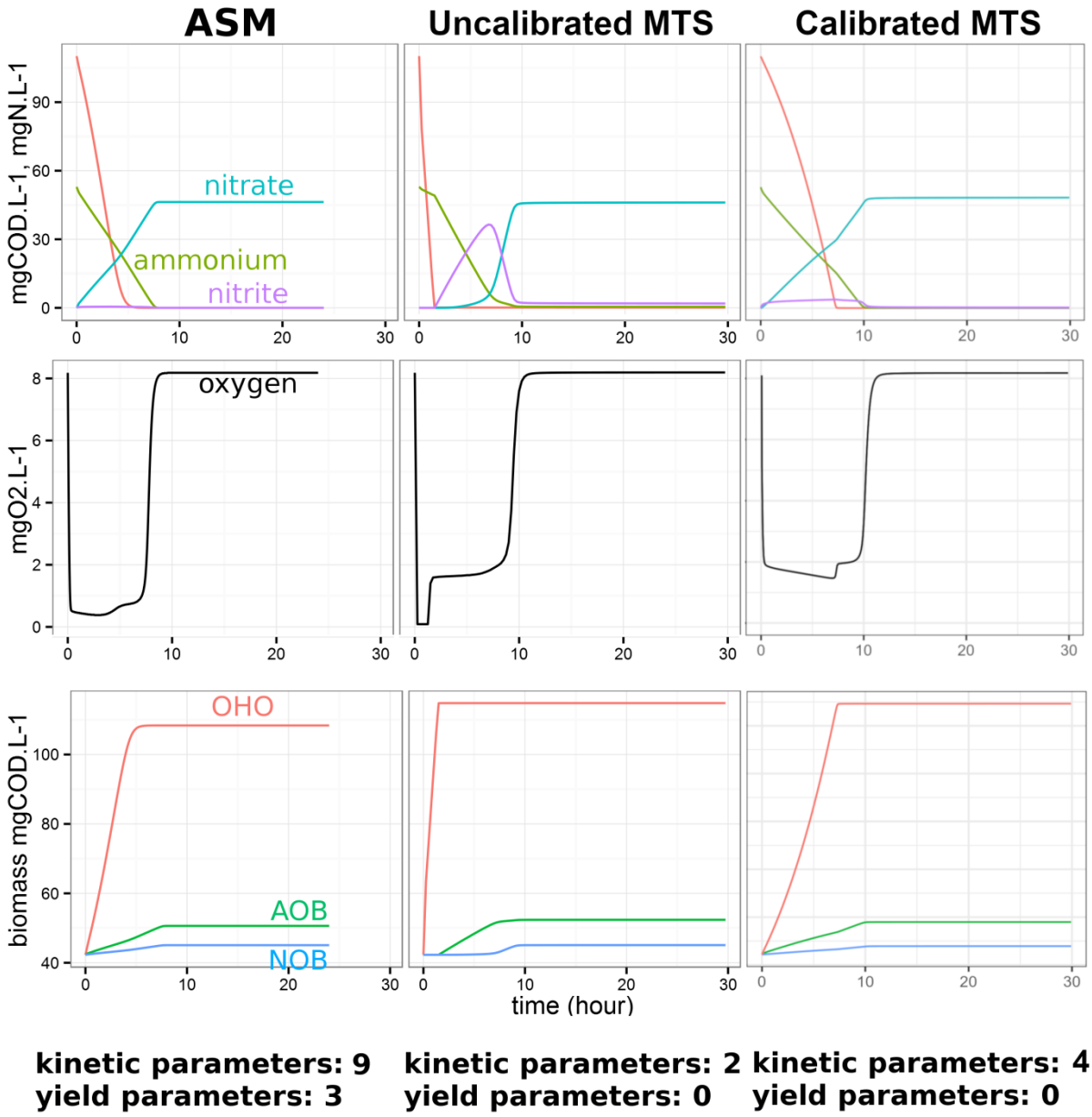


Figure 7 : comparaison entre la concentration de plusieurs espèces chimiques au cours du temps, telle que simulée par différents modèles. « uncalibrated MTS » est le modèle MTS avec µmax = 5.35e17 jour-1 et Vh = 10 m3.molX-1 pour chaque populations, tandis que "calibrated MTS" représente le modèle MTS avec µmax = 3.75 jour-1 pour la population OHO et 2.08 jour-1 pour les populations AOB et NOB, et Vh = 147 m3.molX-1 pour la population OHO et Vh = 296 m3.molX-1 pour les populations AOB and NOB.

Il est visible sur la figure 7 que des différences qualitatives existent entre les prédictions du modèle MTS et ASMN, et ce même après calibration partielle du modèle MTS sur des données respirométriques. Il est tout de même à noter que les tendances générales prédites par le modèle ASMN sont reproduites par le modèle MTS. Tout l'intérêt scientifique de cette simulation comparée est de montrer que le modèle MTS est capable de capturer une partie des dynamiques de population ayant lieu dans le contexte des boues activées. En effet, il est remarquable que le modèle MTS soit capable de reproduire les grandes tendances des dynamiques de population avec seulement 2 à 4 paramètres calibrés, alors que le modèle ASMN les génère à grâce à 12 paramètres calibrés.

Avec ses 12 paramètres calibrés, le modèle ASMN dégradé tel qu'utilisé ici prend en compte la totalité des phénomènes affectant les dynamiques de croissance des populations OHO, AOB et NOB. En proposant une description théorique de l'influence de la stœchiométrie et des gradients d'énergie sur les dynamiques de croissance, le modèle MTS permet visiblement de capturer une part importante de ces dynamiques. Ce modèle n'a pas été conçu pour - et ne prétend pas – prendre en compte la totalité des phénomènes affectant les dynamiques de croissance microbienne, mais il apparait qu'une bonne part de celles-ci est déterminée par la fermeture des bilans de matière et d'énergie dans le système, couplée aux principes premiers introduits par le modèle MTS. La théorisation de la relation entre les gradients d'énergie du milieu de culture et ses dynamiques de population, telle que proposée par le modèle MTS, semble alors pouvoir permettre une réduction du nombre de paramètres calibrés nécessaires dans l'ingénierie des bioprocédés.

Nouvelle estimation de la relation entre l'énergie dissipée et les caractéristiques des métabolismes microbiens

Durant cette thèse, un travail a été effectué sur l'énergie dissipée par un métabolisme par unité de biomasse produite. Contrairement aux autres travaux effectués durant cette thèse, il ne porte pas directement sur l'implémentation et la simulation du modèle MTS. Pour comprendre son lien avec le modèle MTS, il est nécessaire de bien comprendre la façon dont le taux de croissance du modèle MTS est exprimé.

Le modèle MTS exprime le lien entre les gradients d'énergie présents dans le milieu de culture et le taux de croissance d'une population microbienne. Il présuppose que le bilan d'énergie de la croissance de la population modélisée est fermé. La méthode employée pour fermer le bilan d'énergie de la croissance consiste à calculer le facteur λ. Dans l'implémentation du modèle MTS effectuée durant cette thèse, le calcul du facteur λ se fait grâce à l'estimation de l'énergie dissipée par unité de biomasse produite par la croissance (cf equation 1). Cette estimation est faite par le biais d'une formule empirique proposée par Heijnen et collaborateurs (Heijnen *et al.*, 1991). Cette formule a été calibrée sur un certain nombre d'observations expérimentales de rendement de croissance, de manière à produire des prédictions de rendement satisfaisantes. Il n'existe alors pas, à l'heure actuelle, de théorie permettant d'expliquer pourquoi un métabolisme donné dissipe une certaine quantité d'énergie par unité de biomasse produite plutôt qu'une autre. Des lacunes existent donc dans la compréhension de l'énergie dissipée par un métabolisme.

Afin d'améliorer cette connaissance, un travail effectué durant cette thèse a consisté à calibrer une nouvelle formule de prédiction de l'énergie dissipée d'un métabolisme à partir de ses propriétés (degré de réduction du donneur d'électron, énergie catabolique…). Alors que les formules déjà proposées dans la littérature donnent des prédictions plutôt satisfaisantes du point de vue de la prédiction de rendement (Kleerebezem et Van Loosdrecht, 2010), l'objectif de la démarche entreprise ici est de déterminer si l'énergie dissipée de tous types de métabolismes semble obéir à une unique relation, ou si l'énergie dissipée semble suivre des lois distinctes pour plusieurs sous-groupes de métabolismes. Sans proposer de théorisation de l'énergie dissipée (ce qui est totalement hors de portée de cette thèse), l'approche pragmatique consistant à calibrer une relation statistique liant des rendements métaboliques observés à des propriétés des métabolismes tente donc de répondre à une question scientifique fondamentale.

Afin de réaliser ce projet, les bases de données d'observations expérimentales de rendement métaboliques de plusieurs publications précédentes (Batstone *et al.*, 2001 ; Liu *et al.*, 2007 ; Roden *et al.*, 2011) ainsi que d'autres observations issues de la littérature ont été combinées. Il est à noter que l'important travail d'homogénisation ayant mené à cette nouvelle base de données, la plus importante de ce type à ce jour, fait de cette base de données un résultat en lui-même.

A la suite de l'analyse de cette base de données, il apparait que les critères testés comme étant les plus pertinents pour prédire le rendement de croissance d'un métabolisme microbien sont l'énergie catabolique et le nombre de carbones du donneur d'électron et de la source de carbone du métabolisme. Les figures 8, 9 et 10 montrent respectivement les performances de prédiction de rendement du prédicteur proposé par Heijnen et collaborateurs (Heijnen *et al.*, 1991), celles d'un prédicteur générique (même formule pour tous les métabolismes) calibré sur la base de données, et un prédicteur partitionné (mêmes variables explicatives, mais des coefficients différents en fonction de si le métabolisme est une respiration ou une fermentation).

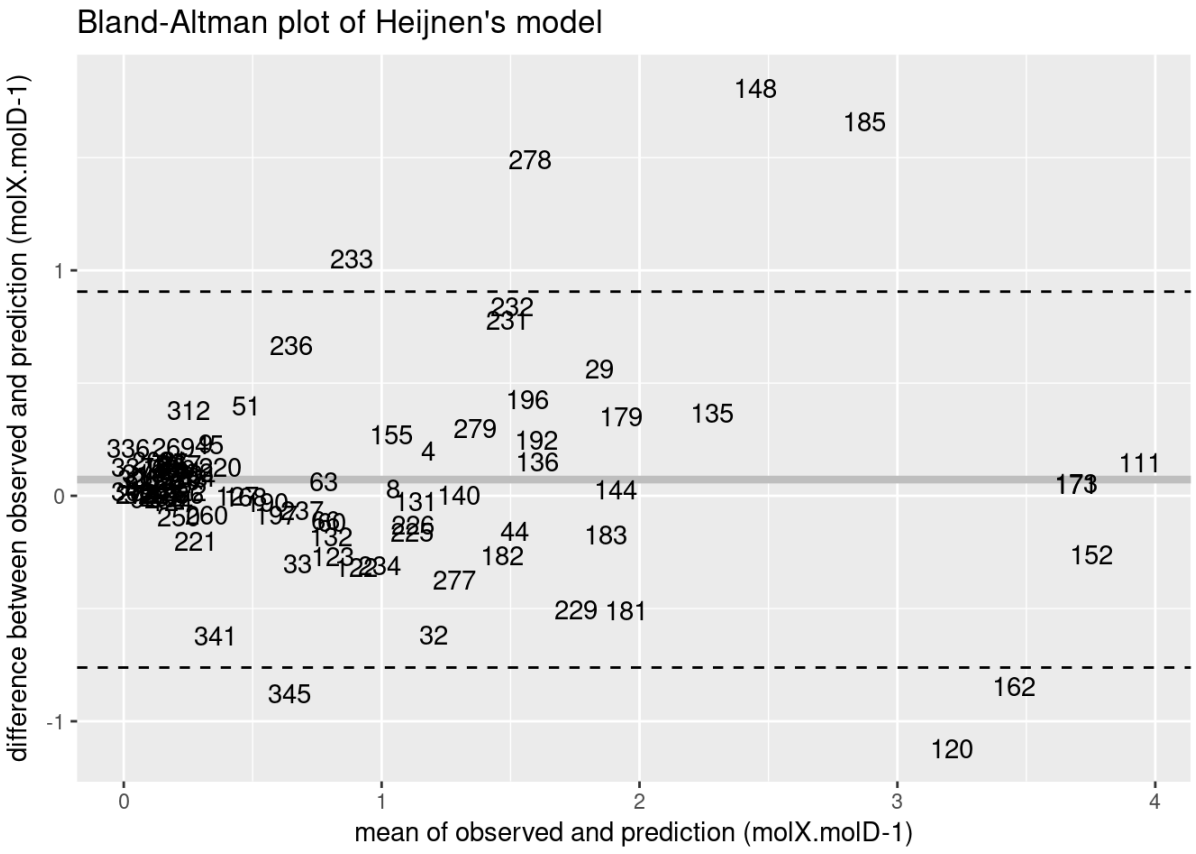


Figure 8 : figure de Bland-Altman montrant les prédictions de rendement du modèle de Heijnen sur le jeu de données collecté pour ce projet. L'axe des abscisses représente la moyenne entre chaque valeur de rendement observée et sa prédiction, et l'axe des ordonnées montre la différence entre chaque valeur de rendement observé et sa prédiction. Chaque chiffre correspond à l'identifiant de l'observation considérée dans la base de données.

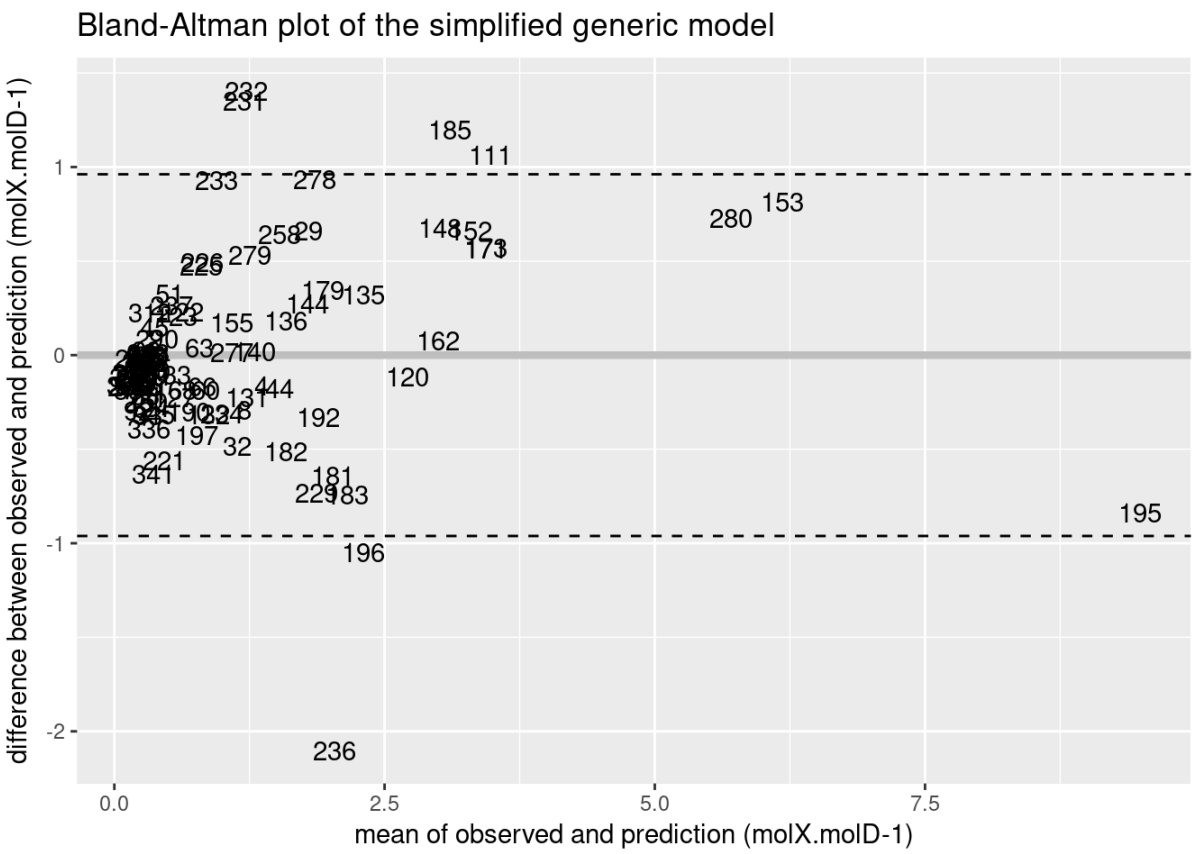


Figure 9 : figure de Bland-Altman montrant les prédictions de rendement du modèle générique déterminé sur le jeu de données collecté pour ce projet pour prédire le rendement de chaque métabolisme en fonction du nombre de carbone de sa source de carbone, du nombre de carbones de son donneur d'électron et de son énergie catabolique. L'axe des abscisses représente la moyenne entre chaque valeur de rendement observée et sa prédiction, et l'axe des ordonnées montre la différence entre chaque valeur de rendement observé et sa prédiction. Chaque chiffre correspond à l'identifiant de l'observation considérée dans la base de données.

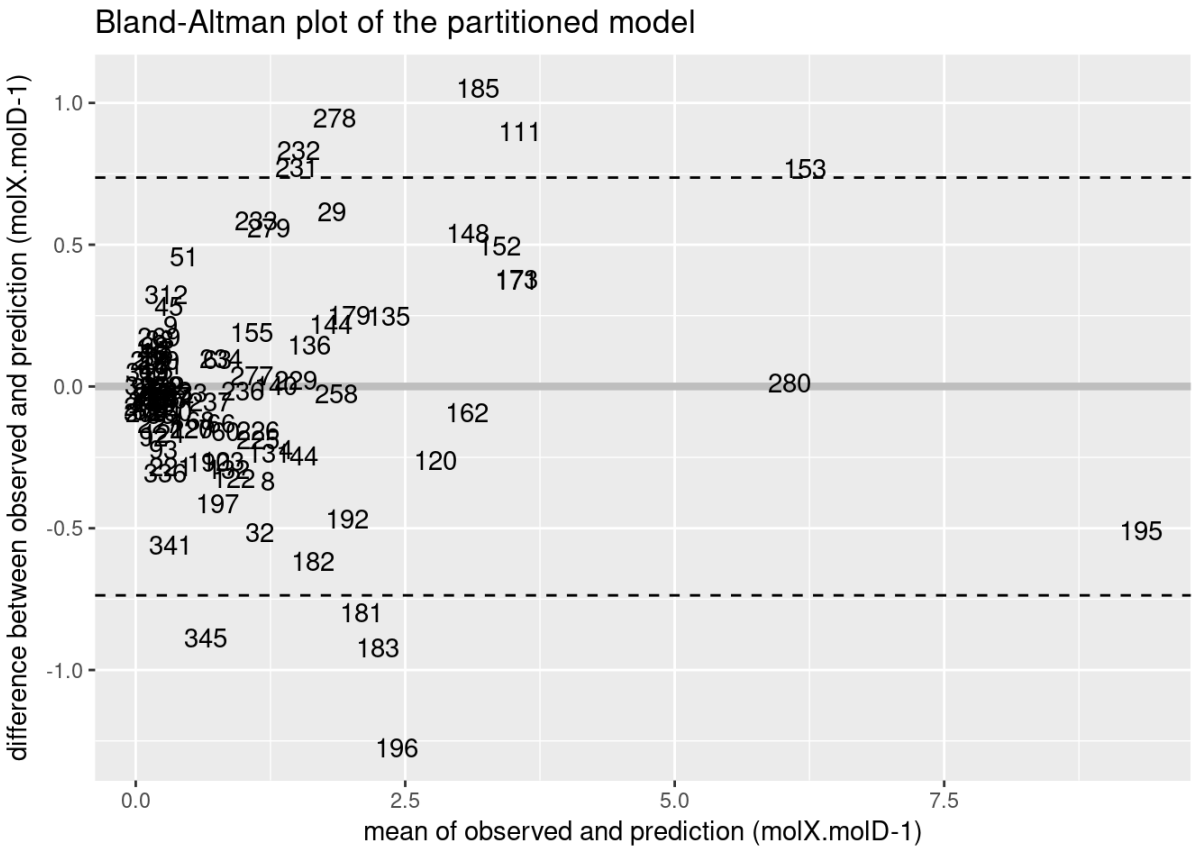


Figure 10 : figure de Bland-Altman montrant les prédictions de rendement du modèle partitionné déterminé sur le jeu de données collecté pour ce projet pour prédire le rendement de chaque métabolisme en fonction du nombre de carbone de sa source de carbone, du nombre de carbones de son donneur d'électron et de son énergie catabolique. Ce modèle utilise des coefficients différents en fonction de si le métabolisme est une respiration ou une fermentation. L'axe des abscisses représente la moyenne entre chaque valeur de rendement observée et sa prédiction, et l'axe des ordonnées montre la différence entre chaque valeur de rendement observé et sa prédiction. Chaque chiffre correspond à l'identifiant de l'observation considérée dans la base de données.

De plus, l'analyse statistique de cette base de données permet d'isoler un sous-groupe de métabolismes dont les rendements sont sensiblement moins bien prédits que les autres, et ce même avec une formule spécifique. Il s'agit des métabolismes fermentaires dont le degré de réduction du donneur d'électron est inférieur à 4.67 mol_e.C-mol_D^{-1}. La distinction ici faite entre respiration et fermentation est que les métabolismes respiratoires sont ceux qui reposent sur l'entretien d'un gradient membranaire, alors que les métabolismes fermentaires ne le font pas. Alors que le travail mené ici ne permet pas de donner une explication théorique à cette observation, elle ouvre néanmoins des pistes de réflexion intéressantes. Il est par exemple possible d'avancer deux explications à ce phénomène. D'une part les métabolismes fermentaires donnent lieu à une certaine diversité de sous-produit ; il est possible que les rendements de croissance dépendent de l'accumulation de ces sous-produits. D'autre part, les variables prédictives considérées lors de cette étude reposent sur une certaine façon de définir les métabolismes, avec un accepteur et un donneur d'électron clairement identifié. Cette façon de décrire les métabolismes convient particulièrement aux métabolismes respiratoires, mais est moins adaptée aux métabolismes fermentaires, où la distinction entre accepteur et donneur d'électron est plus complexe. Dans les deux cas, il est possible que le choix de nouvelles variables pour décrire les métabolismes, plus adaptées aux fermentations, permettent de donner lieu à des prédicteurs plus efficaces.

Dans tous les cas, l'étude statistique de la base de données a révélé des biais de représentation importants pour les métabolismes. Le nombre de mesures de rendement est inégale en fonction des métabolismes. En effet, certains métabolismes d'intérêt industriel ont fait l'objet d'un important effort de recherche et leur rendement est donc documenté par de nombreuses mesures (par exemple, la méthanogénèse hydrogénotrophe ; 31 mesures), alors que la plupart des autres sont très peu documentés (la médiane du nombre de mesures par métabolisme est de 1). Cette base de données étant la plus grande collection de données de ce type dans la littérature il est raisonnable de penser que le biais observé ici reflète un biais présent dans l'ensemble de la littérature scientifique. Ce biais affecte la portée des résultats statistiques qui peuvent être obtenus de ce type d'étude. Un des messages qui émerge est donc que les recherches portant sur l'énergie dissipée des métabolismes et sa prédiction verraient leur fiabilité et leur portée scientifique grandement améliorée si ils se basaient sur des expériences dédiées (couvrant l'étendu des métabolismes possibles dans les conditions les plus homogènes et contrôlées possibles) plutôt que sur des données agrégées à partir de la littérature.

Conclusions et perspectives

Le modèle MTS propose une formulation du lien entre le taux de croissance d'une population de microbes et les gradients d'énergie présents dans le milieu de culture à partir de principes premiers. Ces principes premiers consistent en une description probabiliste de la croissance microbienne à l'échelle de la cellule. L'aspect fondamental du raisonnement dont dérive l'expression du taux de croissance proposé par le modèle MTS constitue une nouveauté apportée par rapport aux autres modèles existants. Le modèle MTS tel qu'implémenté durant cette thèse ne prétend pas reproduire de façon quantitative les dynamiques de croissance. Il se présente davantage comme un outil permettant d'étudier le lien entre le taux de croissance et les gradients d'énergie en fournissant une équation de croissance incarnant une explication fondamentale à ce lien.

Dans un premier temps, le travail effectué durant cette thèse a permis de caractériser les propriétés de la version multi-substrat de la fonction de croissance du modèle MTS. En effet, lors de son introduction en 2014, peu de temps le début de la thèse, le modèle MTS avait été présenté dans une forme mono-substrat, et sans simulation dynamique. L'implémentation du modèle MTS multi-substrat durant cette thèse a permis de constater qu'il reproduit, sans que cela ait été initialement prévu, la loi de Liebig dans la plupart des conditions de culture. Ce résultat est scientifiquement intéressant dans la mesure où ce comportement est la conséquence directe des principes premiers sur lesquels est fondé le modèle, c'est-à-dire le raisonnement statistique décrivant la probabilité de capture de substrat par une cellule. Le modèle MTS fournit donc une explication parcimonieuse à la survenue de la loi de Liebig dans le contexte des dynamiques de croissance des populations microbiennes.

De plus, les simulations dynamiques du modèle MTS effectué durant cette thèse ont montré que la nouvelle fonction de croissance introduite par le modèle MTS était à même de reproduire les dynamiques de croissance à l'œuvre dans des successions écologiques que l'on sait déterminées par les gradients d'énergie (redox tower). Alors que de telles successions ont déjà été reproduites par des modèles basés sur la fonction de croissance de Monod (Gonzalez-Cabaleiro *et al.*, 2015), qui présuppose une relation entre la concentration de substrat et le taux de croissance dont la dérivée est monotone décroissante, il était nécessaire de montrer que la nouvelle fonction de croissance

introduite par le modèle MTS, de forme différente (sigmoïde) est également à même de décrire ces dynamiques.

En ce qui concerne la calibration des paramètres μ_{max} et V_h du modèle à partir de données expérimentales, il a déjà été remarqué plus haut qu'une incertitude subsiste quant à la valeur précise des paramètres cinétiques μ_{max} et V_h estimés à partir de données expérimentales. Cette incertitude existe à la fois du fait de la faible sensibilité des prédictions du modèle MTS à la valeur de ces paramètres (du moins, dans les conditions testées dans le cadre de cette étude), et du fait que la calibration des paramètres a été faite uniquement à partir de deux résultats expérimentaux. A ce stade, l'exploitation de ces résultats permet d'obtenir une première estimation de l'ordre de grandeur des paramètres μ_{max} et V_h, mais ils ne permettent pas de répondre à d'autres questions qui pourraient se poser. Par exemple, il serait scientifiquement intéressant d'estimer les paramètres cinétiques du modèle MTS avec d'avantage de précision de manière à pouvoir déterminer si la valeur de ces paramètres dépend de la population microbienne considérée, ou bien des conditions de culture. Puisque le modèle MTS est bâti sur une description microscopique de la croissance microbienne, il est possible d'attribuer à terme un sens physique à ces paramètres cinétiques. Il est par exemple attendu que le paramètre V_h agrège tous les phénomènes qui modifient la probabilité de rencontre d'une cellule avec une particule de substrat. De tels phénomènes regroupent par exemple la diffusivité, qui dépend du substrat, l'agitation, qui dépend des conditions de culture, et les stratégies spécifiques des microbes pour améliorer leur capture du substrat, telles que le chimiotactisme. De même, il est attendu que le taux de croissance maximum dépende de contraintes métaboliques, telles qu'un flux d'électron maximum, comme cela a déjà été avancé dans la littérature (Andersen and Von Meyenburg, 1980). Seule une calibration plus précise de ces paramètres cinétiques sur davantage de données expérimentales pourra permettre à l'avenir de donner un sens physique plus précis à ces paramètres.

Pour ce qui est de l'étude du lien entre l'énergie dissipée des métabolismes et leurs propriétés physicochimiques, l'analyse statistique effectuée durant cette thèse ne donne pas lieu à une formule de prédiction de rendement sensiblement meilleure que celles proposées dans la littérature, ce qui est rassurant quant à la fiabilité de celles-ci. L'objectif de ce travail était autre, dans la mesure où il a permis de souligner les limites de ce type d'approche et de mettre en évidence une observation intéressante d'un point de vue fondamentale ; il existe un groupe de métabolismes fermentaires dont les rendements ne sont pas prédits de façon satisfaisante à partir des variables explicatives considérées (énergie catabolique, degrés de réduction et nombre de carbone des substrats). Le choix de ces variables provient de la façon usuelle de décrire les métabolismes microbiens du point de vue de la thermodynamique microbienne « boite noire » (c'est-à-dire que les métabolismes sont décrits uniquement en fonction de ce qui entre et sort de la cellule, voir Kleerebezem et Van Loosdrecht, 2010). Il est alors possible qu'une meilleure prédiction des rendements de croissances de ces métabolismes passe par une description théorique différente de ce type de métabolismes.

En conclusion, cette thèse a permis de mettre en évidence que le modèle MTS est capable de reproduire des motifs de croissance à l'échelle des populations individuelles (loi de Liebig) et des communautés (redox tower, dynamiques des populations de boue activée). Son caractère innovant vient du fait qu'il produit ces résultats à partir d'une description de la croissance microbienne à l'échelle des cellules, en utilisant des principes premiers. Alors que les modèles de croissance microbienne basés sur la thermodynamique (sans même parler des modèles proposant une

théorisation) ne constituent pas l'approche la plus développée à présent pour représenter les dynamiques de communautés microbiennes, les capacités du modèle MTS, révélées au cours de cette thèse, constituent un signal positif dans cette direction. Les résultats de cette thèse encouragent le développement d'une nouvelle génération de modèles en écologie et en ingénierie des populations microbiennes, bâtis sur un socle théorique plus robuste. L'avantage de tels modèles est de pouvoir prédire le comportement de communautés microbiennes dans des situations qui n'ont pas encore été expérimentalement testées ; de tels modèles seront alors un atout pour résoudre les défis environnementaux et biotechnologiques qui vont se poser dans les décennies à venir.

agreenium
l'institut agronomique, vétérinaire & forestier de France

Ecole Doctorale ABIES

Title : How do energy gradients shape microbial communities? Study of the Microbial Transition State (MTS) approach for modelling microbial ecosystems dynamics and its application to environmental biotechnology processes

Keywords : microbial thermodynamics, modeling microbial dynamics, microbial kinetics, microbial growth, Gibbs free energy, metabolic model, microbial ecosystems, microbial community assembly patterns

Abstract :

Microbial communities play a key role in geochemical cycles and environmental bioprocesses. Despite their importance, the mechanisms involved in their structuration remain elusive and are poorly captured in current models. The modelling approach developed during this thesis stands as an alternative to the current empirical approaches. It relies on a novel theory of microbial growth (the MTS theory), which introduce a flux/force relationship between the microbial growth rate and the free energy gradients available in the biotope. The purpose of this thesis is to characterize the dynamic properties of the MTS model and to determine, through simulations, the part of the microbial communities' spatio-temporal structuration that is intrinsically captured by the MTS theory and which does not pertain to parameters adjustment.

Simulations firstly reveal that a characteristic of the MTS model is its ability to account for the simultaneous growth limitation by many resources of different kinds (electron acceptor/donor, but also nutrients), and to integrate them as stoichiometric limitations, giving rise to coherent populations dynamics.

In a second stage, the MTS model has been used to predict the dynamics of microbial communities. Those studies revealed that the thermodynamics constraints on which the MTS kinetic theory is built intrinsically give rise to consistent ecological successions without the need to adjust specifically the parameters of each population. In the case of a simplified activated sludge ecosystem, after calibration using respirometric data, the model was able to reproduce ecosystem dynamics quantitatively with a reduced number of parameters compared to current Activated Sludge Models (ASM).

In a third stage, a large database of experimental growth yield observations has been compiled from literature. The relationship between multiple physicochemical parameters characterizing the metabolisms (reduction degrees, catabolic energy...) and the growth yield has been investigated using statistical methods. This work confirms that microbial growth yields can be accurately predicted solely on thermodynamic properties of metabolic reactions. The growth yields predictor could be included in future developments of the MTS models.

More generally, the work undertaken during this thesis evidenced that the MTS model proposes a formalization of the coupling between thermodynamic and dynamic variables of a microbial ecosystem. The simulated microbial populations and ecosystems display coherent dynamic behaviors. The model is able to account, by construction, for well-known ecological successions, without specific parameter adjustment. This model is peculiarly adapted to the prediction of the functional structure of communities in ecosystems dominated by selection by competition, rather than on species dispersion, diversification or genetic drift.

Those results encourage the development of microbial ecosystems based on firmer theoretical grounds. Such models are necessary to the development of bioprocesses able to answer to the new technological and environmental challenges.

agreenium
l'institut agronomique, vétérinaire & forestier de France

Ecole Doctorale ABIES

Titre : Comment les gradients d'énergie façonnent-ils la structure des communautés microbiennes? Etude de la théorie des états de transition microbiens pour modéliser la dynamique des écosystèmes microbiens et application aux procédés de biotechnologie environnementale

Mots-clés : thermodynamique microbienne, modélisation de dynamiques microbiennes, cinétique microbienne, croissance microbienne, communautés microbiennes, écosystèmes microbiens, énergie libre de Gibbs, modèles métaboliques, motifs d'assemblage

Résumé :

Les communautés microbiennes jouent un rôle clef dans les cycles géochimiques et dans les bioprocédés environnementaux. Malgré leur importance, les mécanismes impliqués dans leur structuration restent méconnus et mal appréhendés par les modèles actuels. L'approche de modélisation développée au cours de cette thèse se présente comme une alternative aux approches empiriques actuelles. Elle repose sur une nouvelle théorie de la croissance microbienne (la théorie MTS), qui introduit une relation flux/force entre le taux de croissance microbien et les gradients d'énergie libre disponibles dans le biotope. L'objet de cette thèse est de déterminer par simulation les propriétés dynamiques des modèles MTS et dans quelle mesure la théorie est capable d'apporter une explication qualitative à la structuration spatio-temporelle des communautés microbiennes.

Les simulations ont premièrement révélé que les modèles MTS sont capables de tenir compte de la limitation sur la croissance exercée simultanément par plusieurs ressources de type différent (accepteur/donneur d'électron, nutriment etc.), et de les intégrer sous la forme de limitations stoichiométriques, donnant lieu à des dynamiques de population cohérentes.

Dans un deuxième temps, le modèle MTS a été utilisé pour prédire les dynamiques de communautés microbiennes simplifiées. Ce travail a révélé que les contraintes thermodynamiques sur lesquelles la théorie cinétique de MTS est construite donnent lieu à des successions écologiques cohérentes, sans qu'il n'y ai besoin d'ajuster spécifiquement les paramètres de chaque population. Dans le cas d'un écosystème de boues activées simplifié, après calibration sur des données respirométriques, le modèle MTS a été capable de reproduire quantitativement des dynamiques de l'écosystème, avec un nombre de paramètres moindre que celui de l'actuel Activated Sludge Model (ASM) faisant autorité dans le domaine.

Dans un troisième temps, une grande base de données d'observations expérimentales de taux de croissance a été compilée depuis la littérature. La relation entre plusieurs paramètres physicochimiques caractéristiques des métabolismes (degrés de réduction, énergie catabolique …) et le rendement de croissance microbien a été étudiée en utilisant des méthodes statistiques. Ce travail confirme que le rendement de croissance microbien peut être bien prédit à partir des seules propriétés thermodynamiques des réactions métaboliques.

Le travail entrepris durant cette thèse montre que le modèle MTS propose une formalisation du couplage entre des variables thermodynamiques et dynamiques d'un écosystème microbien. Les populations et écosystèmes microbiens simulés ont montré des dynamiques cohérentes. Le modèle est capable de rendre compte, par construction, de successions écologiques observées, sans nécessiter d'ajustement paramétrique spécifique. Ce modèle est particulièrement bien adapté pour prédire la structure fonctionnelle de communautés dans des écosystèmes dominés par la sélection sur la compétition, plutôt que sur la dispersion, la diversification ou la dérive génétique.

Ces résultats encouragent le développement de modèles d'écosystèmes microbiens construits sur des bases théoriques plus solides. De tels modèles sont nécessaires au développement de bioprocédés plus aptes à répondre aux nouveaux défis technologiques et environnementaux.