Ecological Modelling 335 (2016) 74–86

Contents lists available at ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel

Reaction-centric modeling of microbial ecosystems

Stilianos Louca [a,*], Michael Doebeli [b,c]

[a] Institute of Applied Mathematics, University of British Columbia, 311-6356 Agricultural Road, Vancouver V6T1Z2, Canada
[b] Department of Zoology, University of British Columbia, Canada
[c] Department of Mathematics, University of British Columbia, Canada

ARTICLE INFO

Article history:
Received 27 January 2016
Received in revised form 18 May 2016
Accepted 20 May 2016

Keywords:
Biochemical reaction network
Bioprocess monitoring
Bioreactor
Pathway-centric
Microbial ecology

ABSTRACT

The growth of microbial populations catalyzing biochemical reactions leads to positive feedback loops and self-amplifying process dynamics at ecosystem scales. Hence, the state of a biocatalyzed process is not completely determined by its physicochemical state, but also depends on current cell or enzyme concentrations that are often unknown. Here we propose a generic approach to modeling reaction networks of natural and engineered microbial ecosystems, that is able to capture the self-amplifying nature of biochemical reactions without explicit reference to the underlying microbial populations. This is achieved by keeping track of a system's "capacity" to perform particular reactions, rather than the cell populations actually catalyzing them. Our reaction-centric approach minimizes the need for cell-physiological parameters such as yield factors and provides a suitable framework for describing a system's dynamics purely in terms of chemical concentrations and fluxes. We demonstrate our approach using data from an incubation experiment involving urea hydrolysis and nitrification, as well as time series from a long-term nitrifying bioreactor experiment. We show that reaction-centric models can capture the dynamical character of microbially catalyzed reaction kinetics and enable the reconstruction of bioprocess states using solely chemical data, hence reducing the need for laborious biotic measurements in environmental and industrial process monitoring.

1. Introduction

Microbial metabolism powers biochemical fluxes in natural and engineered ecosystems (Falkowski et al., 2008; McDuffie, 1991). Reciprocally, biochemical fluxes sustain biosynthesis and thus drive microbial population dynamics (Jin and Bethke, 2007). Changes in the microbial populations, in turn, influence the reaction kinetics at ecosystem scales because system-wide reaction rates depend not only on substrate concentrations but also on the density of catalyzing cells or of extracellular enzymes (Simkins and Alexander, 1984). Thus, the dynamics of microbial communities emerge from the continuous interplay between metabolic activity, changes in the extracellular metabolite pool and microbial population growth (Song et al., 2014). In particular, and in contrast to purely abiotic chemical processes (Marjanovic et al., 2006), the state and future trajectory of a biocatalyzed process cannot be determined solely based on the system's chemical state (Simkins and Alexander, 1984; Jin and Bethke, 2007). For example, empirical mineralization curves that describe the degradation rate of organic matter as a function of substrate density can vary strongly in shape, and this variation historically resulted partly from the interaction of substrate concentrations and cell population densities in experiments (Simkins and Alexander, 1984).

In deterministic or stochastic differential equation models (Resat et al., 2009; Khatri et al., 2012; Song et al., 2014), the dynamical character of microbially catalyzed reaction kinetics is typically incorporated by including additional variables representing cell densities, whose growth is proportional to the rates of the processes that they catalyze and determined by cell-per-substrate (or sometimes biomass-per-substrate) yield factors (Jin and Bethke, 2007). In turn, system-wide reaction kinetics are modulated by current cell densities and extracellular metabolite concentrations. Such cell-centric models are widely used and can capture the typical self-amplifying character of biocatalyzed processes (Cheyns et al., 2010). Likewise, deterministic as well as stochastic individual-based models, which keep track of multiple individual organisms and their metabolic activity, can also capture the feedback loops within microbial metabolic networks because the metabolic or trophic activity of each organism eventually

Abbreviations: AOB, ammonium oxidizing bacteria; NOB, nitrite oxidizing bacteria; *ure*, urea hydrolysis (gene or pathway); *amo*, aerobic ammonium oxidation (gene or pathway); *nxr*, aerobic nitrite oxidation (gene or pathway).

* Corresponding author.
E-mail address: louca@math.ubc.ca (S. Louca).

http://dx.doi.org/10.1016/j.ecolmodel.2016.05.011
0304-3800/

leads to the production of new copies of that organism (Ferrer et al., 2008; Larsen et al., 2012). All of these cell-centric models, however, depend on physiological parameters such as yield factors, cell masses or maximum cell-specific reaction rates, and require knowledge of cell or enzyme concentrations (in addition to physicochemical variables) for describing a system's current state. As we explain below, some of these parameters also introduce redundancies from a reaction kinetic point of view that can lead to strong uncertainties in parameter estimation (Simkins and Alexander, 1984; Knightes and Peters, 2000).

Flux-balance models, a popular alternative to dynamical models (Orth et al., 2010), reduce the number of required parameters by ignoring cell population dynamics and by assuming that metabolite concentrations are constant through time (i.e. metabolite fluxes are "balanced"). In these models, reaction rates (and sometimes metabolite turnover rates; Chung and Lee, 2009) are the only independent variables, and their values are calculated by optimizing some objective function (e.g. ATP production) in the presence of constraints (e.g. on maximum reaction rates). Flux balance models have been very successful in elucidating metabolic network properties such as the feasibility of certain reactions or the prediction of metabolic interactions between species (Stolyar et al., 2007; Zomorrodi and Maranas, 2012; Klitgord and Segrè, 2010) but – being steady-state models – they fail to capture the dynamical nature of microbial communities. Hence, current model frameworks either ignore the temporal and self-amplifying character of biocatalyzed processes or require an extensive set of – often poorly estimated – physiological parameters.

To address the above limitations, here we develop a new framework for dynamical bioprocess modeling with a focus on system-wide reaction kinetics. Our objective was to reduce the reliance on physiological parameters and to reduce the need for biotic measurements for state reconstruction and model calibration, while still accounting for the self-amplifying character of metabolic reactions at the ecosystem level. Such a "reaction-centric" model would ideally make predictions purely in terms of metabolite concentrations and reaction rates at the ecosystem level, without the need to consider the underlying cell populations. As we show below, this can be achieved by keeping track of a system's "capacity" to perform particular reactions (or pathways), rather than the cell populations actually catalyzing them. Microbial ecosystem metabolism can then be described similarly to abiotic reaction networks, with the addition of so-called self- and cross-amplification factors between reactions. These amplification factors are specific to a particular microbial community and translate the system's metabolic fluxes into changes of the system's reaction capacities. Hence, a system's state and dynamics can be inferred using solely physicochemical measurements, bypassing laborious biotic measurements for example in environmental and industrial process monitoring. Furthermore, reaction-centric models minimize the reliance on cell-physiological parameters, allowing for model calibration even when biotic data are scarce. Reaction-centric models thus provide an elegant alternative to many conventional cell-centric models, particularly when the ultimate focus is on a system's reaction kinetics.

We begin with a derivation of the reaction-centric framework and show how it relates to conventional, cell-centric models. We focus on differential equation models, however we note that our reasoning can also be applied to other cell-centric frameworks. We demonstrate the potential of reaction-centric models using data from a previous short-term incubation experiment with a ureolytic and nitrifying microbial community (de Boer and Laanbroek, 1989), as well as long-term time series from a flow-through nitrifying bioreactor (Dumont et al., 2009). Bioreactors provide ideal model ecosystems for testing new theories for microbial ecology, due to their higher controllability and measurability when compared to natural ecosystems. Ureolysis and nitrification were chosen as examples because of their conceptual simplicity as well as their great relevance to ecosystem productivity, industry and agriculture (Wiesmann, 1994; Prosser, 2005). Our entire analysis was performed with a recently published computational tool for modeling microbial ecosystems (Louca and Doebeli, 2015a), which we extended to accommodate reaction-centric models.

2. Methods

2.1. Derivation of reaction-centric models: one reaction per cell

Conventional cell-centric microbial ecosystem models consider the extracellular concentrations of metabolites as well as the cell densities of microbial populations catalyzing various reactions. In the simplest and most common case each reaction is catalyzed by a distinct microbial population, the growth of which is proportional to the rate of the reaction (Simkins and Alexander, 1984; Larsen et al., 2012; Jin and Bethke, 2007). More precisely, the population density of cells catalyzing reaction r (N_r, cells per volume) and the concentration (C_m) of each metabolite m are described by differential equations similar to the following:

$$\frac{dN_r}{dt} = N_r Y_r V_r h_r(\mathbf{C}) - \lambda_r N_r, \tag{1}$$

$$\frac{dC_m}{dt} = F_m(t, \mathbf{C}) + \sum_r S_{mr} N_r V_r h_r(\mathbf{C}). \tag{2}$$

In Eq. (1), Y_r is a cell yield factor (cells produced per substrate used), V_r is the maximum cell-specific reaction rate (flux per cell per time) and $\mathbf{C}$ is the vector representing all metabolite concentrations (overview of symbols in Table 1). We note that in models where N_r is alternatively measured in biomass (rather than cells) per volume, Y_r is typically a biomass yield factor and V_r is a maximum biomass-specific reaction rate. The dependence of cell-specific reaction kinetics on $\mathbf{C}$ is encoded by the unitless function $h_r(\mathbf{C})$, which is normalized to unity at those $\mathbf{C}$ that maximize the cell-specific reaction rate. The last term in Eq. (1) corresponds to the decay of biomass at an exponential rate λ_r (with units time^{-1}), for example due to cell death. Alternatively, λ_r can account for reduced biosynthesis due to maintenance energy requirements, in which case it is sometimes called the "specific maintenance rate" (Jin and Bethke, 2007). In Eq. (2), F_m accounts for abiotic metabolite fluxes, such as substrate supply in a bioreactor, and S_{mr} is the stoichiometric coefficient of metabolite m in reaction r. The sum in Eq. (2) iterates through all reactions and accounts for microbial metabolic fluxes.

In the above cell-centric model the system's state depends on the current metabolite concentrations (C_m) as well as the current cell densities (N_r), the prediction of which, in turn, requires knowledge of physiological parameters such as Y_r and V_r. As we show below, this focus on cell populations can be avoided if one is solely interested in the system's reaction kinetics. Observe that the product $M_r = N_r V_r$, henceforth referred to as the system's current "reaction capacity", is the maximum system-wide rate of reaction r (flux per volume per time) that could possibly be attained at favorable metabolite concentrations (i.e. when $h_r(\mathbf{C}) = 1$). On the other hand, the product $H_r = N_r V_r h_r = M_r h_r$ is the actual system-wide rate of reaction r. Note that H_r depends both on the reaction capacity M_r as well as the normalized kinetics $h_r(\mathbf{C})$, which encodes the dependence of the reaction rate on the system's chemical state. Rewriting Eqs. (1) and (2) in terms of the reaction capacities M_r yields the reaction-centric model

$$\frac{dM_r}{dt} = A_r M_r h_r(\mathbf{C}) - M_r \lambda_r, \tag{3}$$

Table 1
Overview of symbols and units used in this study. The indices r and q enumerate reactions or cell species, while m enumerates metabolites. Parameters or variables specific to cell-centric models are indicated by "†", those specific to reaction-centric models are indicated by "⋆". Parameter values used in the examples are given in Tables A.1 and B.1.

Symbol and description			Units	Used as
t	Time		days (d)	–
N_r	Cell density	†	cells/L	Independent variable
$\mathbf{N}$	All cell densities (vector)	†	cells/L	Independent variable
C_m	Metabolite concentration		mol/L	Independent variable
$\mathbf{C}$	All metabolite concentrations (vector)		mol/L	Independent variable
C_m^o	Initial metabolite concentration		mol/L	Parameter
Y_r	Cell yield factor	†	cells/mol	Parameter
V_r	Maximum cell-specific reaction rate	†	mol/(cell d)	Parameter
h_r	Normalized cell-specific reaction rate		–	Function of $\mathbf{C}$
λ_r	Exponential biomass decay rate		1/d	Parameter
K_r	Substrate half-saturation constant		mol/L	Parameter
F_m	Abiotic net metabolite influx		mol/(L d)	Function of t and $\mathbf{C}$
S_{mr}	Stoichiometric coefficient		–	Parameter
H_r	Reaction rate		mol/(L d)	Dependent variable
A_r	Self-amplification factor	⋆	1/d	Parameter
A_{rq}	Cross-amplification factor	⋆	1/d	Parameter
M_r	Reaction capacity	⋆	mol/(L d)	Independent variable
$\mathbf{M}$	All reaction capacities (vector)	⋆	mol/(L d)	Independent variable
M_r^o	Initial reaction capacity	⋆	mol/(L d)	Parameter
ρ_{ure}	Ammonia recycling fraction		–	Parameter
T_r	Substrate assimilation factor		–	Parameter
Symbols specific to example 2 (flow-through bioreactor)				
C_m^{in}	Metabolite concentration in inflow		mol/L	Function of t
$\hat{M}_r$	Reconstructed reaction capacity	⋆	mol/(L d)	Estimated variable
$\hat{C}_m$	Reconstructed metabolite concentration		mol/(L d)	Estimated variable
μ	Hydraulic dilution rate		1/d	Function of t

$$\frac{dC_m}{dt} = F_m(t, \mathbf{C}) + \sum_r S_{mr} H_r(\mathbf{C}), \tag{4}$$

$$H_r = M_r h_r, \tag{5}$$

where we introduced the so called self-amplification factor $A_r = V_r Y_r$ in Eq. (3). This model describes biochemical reactions at the scale of the ecosystem, without explicit reference to biotic quantities such as cell densities or physiological parameters such as yield factors.

The structure of Eq. (3) emphasizes the self-amplifying nature of biochemical reactions at the ecosystem level, with the self-amplification factors A_r mediating the conversion of reaction rates to a growth of reaction capacities. In the context of cell-centric models, A_r is the maximum specific growth rate of cells performing reaction r (in units time^{-1}). In the reaction-centric model, A_r becomes the maximum exponential growth rate of the reaction capacity M_r. Note that A_r only depends on the product $V_r Y_r$, but not on the individual V_r or Y_r. Hence, the system's biochemical dynamics can be modeled without knowledge of the V_r and Y_r because the system's trajectory is completely determined by the self-amplification factors and the reaction capacities at some point in time. This collapse of unknown parameters into fewer ones, without loosing any predictive power, means that fewer parameters are needed for practical purposes than often assumed. In fact, the redundancy inherent to the simultaneous inclusion of V_r and Y_r in conventional models was previously pointed out by Simkins and Alexander (1984). This redundancy can lead to strong negative correlations between estimated Y_r and V_r, particularly when parameter estimation is based solely on non-biotic chemical time series, because such time series cannot differentiate between alternative combinations of V_r and Y_r yielding the same product $V_r Y_r$ (Knightes and Peters, 2000).

2.2. Derivation of reaction-centric models: multiple reactions per cell

So far we assumed that each cell performs exactly one reaction, which means that each modeled reaction only induces the growth of its own capacity. While this assumption is widespread in ecosystem modeling (Larsen et al., 2012; Reed et al., 2014), in reality several alternative pathways may be performed by the same cells. For example, members of the ammonium oxidizing *Nitrosospira* genus are also able to hydrolyze urea (Marsh et al., 2005), and urea hydrolysis in incubation experiments with *Nitrosospira* was shown to promote ammonium oxidation by the same population (de Boer and Laanbroek, 1989). In the simplest case, the combined effects of several metabolic pathways on cell population growth can be assumed to be additive, so that each reaction r has a contribution $Y_r H_r$ to the total growth of the cell population:

$$\frac{dN_r}{dt} = Y_r H_r + \sum_q Y_q H_q - N_r \lambda_r. \tag{6}$$

Here, the sum in Eq. (6) iterates over all additional reactions attributable to cells performing reaction r. If two reactions r and q are performed by the same population then $N_r = N_q$ and, reciprocally, if two populations share a common reaction, that reaction will need to be represented twice using two separate indices r. The assumption of additive effects on growth is common in conventional microbial population models. For example, Courtin and Spoelstra (1990) model a population of acetic acid bacteria utilizing multiple organic substrates by assuming that each pathway has an additive effect on the total population growth. More sophisticated models of microbial metabolism based on flux balance analysis and optimization of a linear utility function also assume additive effects of various metabolic fluxes, although the functions $h_r(\mathbf{C})$ may not be explicit in $\mathbf{C}$, but instead specified in terms of an optimization algorithm (Orth et al., 2010).

The cell-centric model in Eq. (6) corresponds to a reaction-centric model in which multiple reactions amplify each other's capacities whenever they are performed by the same cells:

$$\frac{dM_r}{dt} = A_r H_r + \sum_q A_{rq} H_q - \lambda_r M_r. \tag{7}$$

Here, the so-called "cross-amplification" factors $A_{rq} = V_r Y_q$ correspond to the positive effects of the flux through some reaction q

on the capacity of some other reaction r and hence, the sum in Eq. (7) iterates through all additional reactions q performed by the same cell population as reaction r. The amplification matrix, whose diagonal entries are the self-amplification factors A_r and whose off-diagonal entries are the cross-amplification factors A_{rq}, defines a linear transformation of the vector containing all reaction rates to a vector containing changes in reaction capacities. Note that regardless of any amplifications of the reaction capacities, actual rates may still be limited by low substrate concentrations or the presence of inhibitors, as determined by the normalized kinetics $h_r(\mathbf{C})$. Also note that since $A_{rq} = V_r Y_q$ for any two reactions q and r performed by the same cells, the following consistency conditions apply:

$$A_{qr} = \frac{A_r A_q}{A_{rq}}, \quad M_r = M_q \frac{A_{rq}}{A_q}. \tag{8}$$

Regardless of any cell-centric interpretation, the system's reaction dynamics only depend on the amplification factors A_{rq}, but not on any Y_q or V_r.

The above discussion illustrates how conventional cell-centric models can be used to derive reaction-centric models and foster confidence in their realism. For example, amplification factors can be seen as a combination of – and a replacement for – cell-centric parameters. However, as we demonstrate below, in practice a reaction-centric model can be taken as an alternative self-contained description of a system's reaction kinetics. Under such a paradigm, the amplification matrix becomes a set of standalone system-specific parameters and the reaction capacities become independent state variables whose dynamics are shaped by the amplification matrix. We note that while here we focus on linear growth dynamics, non-linear generalizations are also possible with additional amplification coefficients mediating the higher order effects of biochemical fluxes on reaction capacities.

Apart from the elegance of a reaction-centric description, an added benefit is that all parameters and state variables can be inferred from purely physicochemical time series. For example, at high substrate concentrations and in the absence of inhibitors, reaction capacities (M_r) are approximately equal to actual reaction rates (H_r) and can thus be estimated directly from the derivative (slope) of chemical concentration time series. Similarly, if the normalized reaction kinetics h_r (or equivalently, the half-saturation constants in case of Monod kinetics) are known, then reaction rates estimated from concentration time series can be divided by h_r to yield the reaction capacities. In general, however, reaction capacities may constitute unknown system variables which must be estimated indirectly, for example by repeated observation of metabolite concentrations (as demonstrated below).

2.3. Validation of the reaction-centric framework

To exemplify our approach, we constructed reaction-centric models for two separate engineered microbial ecosystems used in previously published experiments. Specifically, in the first example we consider urea hydrolysis and nitrification in a batch-fed incubation experiment previously described by de Boer and Laanbroek (1989). The structure of our model, described in detail below, was chosen to closely resemble the physicochemical conditions in the experiment as well as the metabolic network involved in dissimilatory nitrogen transformations – as inferred from the experiment. We test the adequacy of our reaction-centric model by assessing its "goodness of fit" after calibrating unknown parameters to the experimental data. Further, we demonstrate the importance of cross-amplification factors for accounting for pathway co-occurrences in cells by comparing two variants of the model, namely one variant with and one variant without the required cross-amplification factors.

In the second example we consider a reaction-centric model for a flow-through ammonium-fed nitrifying bioreactor, operated under varying conditions over the course of several months (Dumont et al., 2009). Similarly to the first example, our model is constructed to closely resemble the physicochemical conditions of the bioreactor. In this example, we demonstrate how purely chemical time series can be used to calibrate a reaction-centric model and to infer the full biochemical state of the bioreactor (i.e. M_r and C_m) in "real-time". In addition, to further assess the fidelity of the model, we use independent biomass concentration measurements from the original experiment, which we compare to the hypothetical biomass concentrations that would correspond to the reaction capacities in the reaction-centric model.

2.4. Computation

All time series analysis, simulations and parameter calibrations in this study were performed using MCM (Microbial Community Modeler), a computational framework recently published by our lab (Louca and Doebeli, 2015a). The construction and analysis of the models in MCM is explained in Supplement S1.4. MCM, together with a thorough user manual and step-by-step examples, is available at: http://www.zoology.ubc.ca/MCM.

3. Results and discussion

3.1. Example 1: Urea hydrolysis and nitrification in a batch-fed incubator

3.1.1. Overview of experimental results

The microbial community in the incubator was dominated by *Nitrosospira* sp., which are ammonium oxidizing bacteria (AOB) also capable of hydrolyzing urea to ammonium, and *Nitrobacter* sp., which are nitrite oxidizing bacteria (NOB; Fig. 1a). The incubator was batch-fed with urea, the complete hydrolysis of which by the AOB led to a temporary accumulation of ammonium (NH_4^+) within roughly one week. Concurrently to its production, NH_4^+ was also oxidized by the AOB into nitrite (NO_2^-), which was in turn oxidized by the NOB into nitrate (NO_3^-). Nitrification continued after complete urea hydrolysis until NH_4^+ concentration dropped to about 0.5 mM. The high energy requirements for maintaining a more neutral internal pH than the external environment (pH 5) could presumably not be met at lower NH_4^+ concentrations, eventually leading to a halt of nitrification (de Boer and Laanbroek, 1989).

3.1.2. Inferred model structure

The model focuses on dissimilatory nitrogen fluxes encompassing urea hydrolysis (*ure*), ammonium oxidation (*amo*) and nitrite oxidation (*nxr*). All nitrogen metabolism is assumed to be entirely dissimilatory. Specifically, we assume that each mol urea is converted by *ure* to 2 mol NH_4^+, of which a small fraction ρ_{ure} is immediately oxidized ("recycled") to NO_2^- within the same cell, while the remaining NH_4^+ leaks to the extracellular medium:

$$ure : \text{urea} \rightarrow 2(1-\rho_{ure}) \times NH_4^+ + 2\rho_{ure} \times NO_2^-. \tag{9}$$

We assume that extracellular NH_4^+ taken up by AOB is completely oxidized to NO_2^-, and that all NO_2^- taken up by NOB is completely oxidized to NO_3^-:

$$amo : NH_4^+ \rightarrow NO_2^-, \quad nxr : NO_2^- \rightarrow NO_3^-. \tag{10}$$

The recycling term ρ_{ure} was included in order to explain the early appearance of NO_3^- in the incubator (Fig. 2c). Despite the increased model complexity (one additional free parameter), preliminary statistical model selection tests (based on AIC and BIC; Konishi and Kitagawa, 2008) showed a clear preference for the inclusion of ρ_{ure}

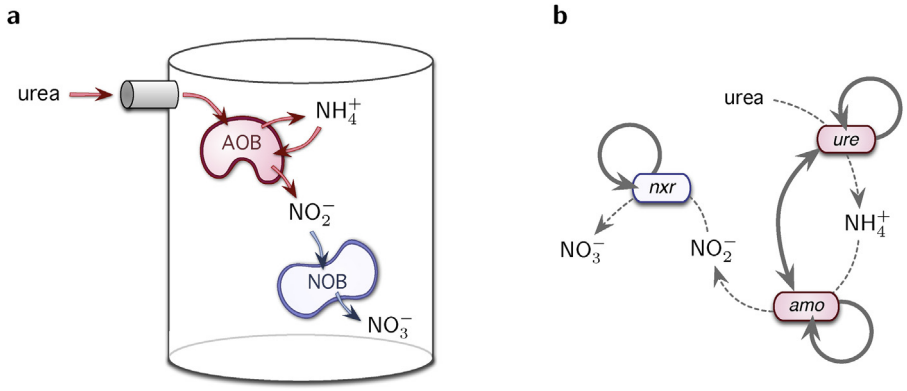

Fig. 1. Modeling urea hydrolysis and nitrification. (a) Microbial ecosystem model for urea hydrolysis and subsequent nitrification by ammonium (NH_4^+) oxidizing bacteria (AOB) and nitrite (NO_2^-) oxidizing bacteria (NOB), in a batch-fed incubator. (b) Corresponding reaction-centric model comprising urea hydrolysis (*ure*), ammonium oxidation (*amo*) and nitrite oxidation (*nxr*) with explicit self- and cross-amplifications (continuous arrows): A flux through each reaction (dashed arrows) increases the rate capacity of that reaction and of reactions catalyzed by the same cells.

(Supplementary Figure S1.5). *amo* rates were assumed to be limited by ammonia (NH_3), rather than NH_4^+, in accordance to findings by Suzuki et al. (1974). Due to a lack of further information, potential oxygen limitation in the incubator was ignored.

The co-occurrence of *ure* and *amo* genes in the same AOB cells leads to a direct coupling of the population dynamics of these genes and enzymes, and therefore the incubator's *amo* and *ure* reaction capacities (Fig. 1b). In the model, this coupling corresponds to positive cross-amplification factors that measure the mutual effects of *ure* flux on *amo* capacity and vice versa. Hence, based on the model structure introduced in Section 2.2, the differential equations for the reaction capacities M_{ure}, M_{amo} and M_{nxr} take the form

$$\frac{dM_{ure}}{dt} = A_{ure} \cdot H_{ure} + A_{ure,amo} \cdot H_{amo} - M_{ure} \cdot \lambda_{\mathrm{AOB}}, \tag{11}$$

$$\frac{dM_{amo}}{dt} = A_{amo} \cdot H_{amo} + A_{amo,ure} \cdot H_{ure} - M_{amo} \cdot \lambda_{\mathrm{AOB}}, \tag{12}$$

$$\frac{dM_{nxr}}{dt} = A_{nxr} \cdot H_{nxr} - M_{nxr} \cdot \lambda_{\mathrm{NOB}}. \tag{13}$$

Preliminary tests indicated that *nxr* decay could be omitted from the model because within the time span of the experiment NOB cell densities were mostly limited by NO_2^- supply, hence on grounds of parsimony we set $\lambda_{\mathrm{NOB}} = 0$. On the other hand, our tests indicated that the decay term λ_{AOB} was mostly attributable to AOB maintenance rates (Jin and Bethke, 2007), resulting in a reduced *ure* and *amo* growth and leading to a substrate threshold below which dissimilatory metabolism can no longer sustain growth. That threshold is reached when

$$A_{amo} h_{amo} + A_{ure} h_{ure} \leq \lambda_{\mathrm{AOB}}, \tag{14}$$

at which point we assumed a complete halt of *ure* and *amo* activity (Supplement S1.1). Note that care needs to be taken to ensure consistency between the cross-amplification terms $A_{amo,ure}$ and

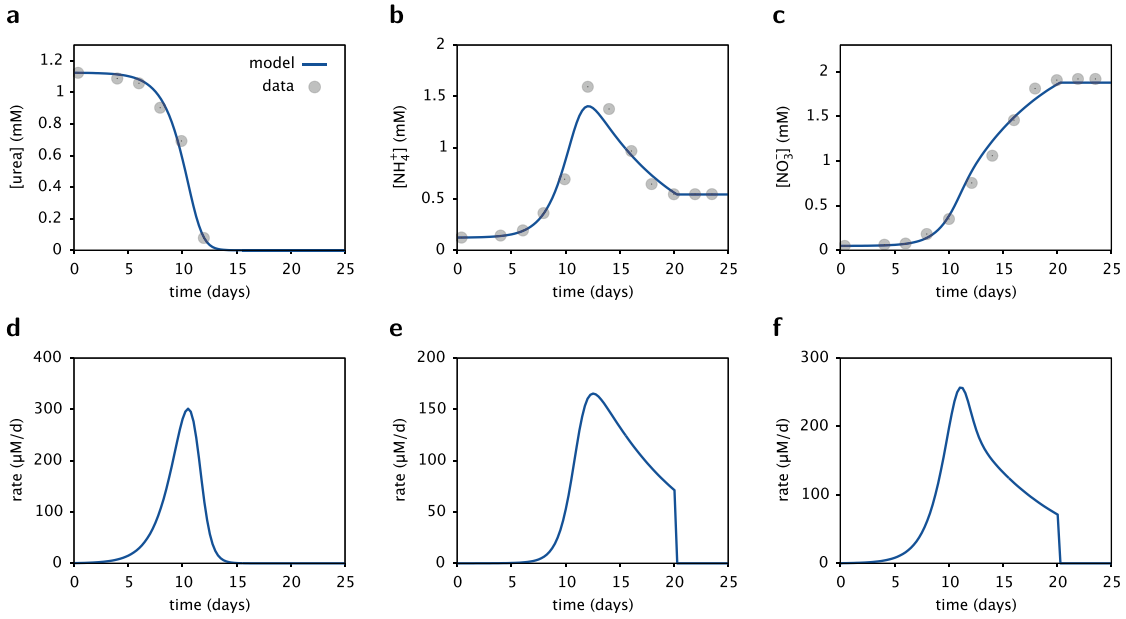

Fig. 2. Model predictions and data for Example 1. Model predictions and data from a batch-fed incubation experiment involving urea hydrolysis and nitrification: (a) Urea, (b) ammonium and (c) nitrate concentrations over time, following incubation of a mixed *Nitrosospira* AHB1 and *Nitrobacter* NHB1 community in a urea-enriched medium. Second row: (d) urea hydrolysis (*ure*), (e) ammonium oxidation (*amo*) and (f) nitrite oxidation (*nxr*) rates over time. The rapid halt of *amo* (and subsequently *nxr*) around day 20 occurs when ammonia concentration falls below the threshold imposed by the maintenance energy requirements of the cells (Eq. (14)). See Section 2 for details. Data from de Boer and Laanbroek (1989).

$A_{ure,amo}$, as well as between the initial reaction capacities M^o_{ure} and M^o_{amo}. As explained previously, we need to have

$$A_{ure,amo} = \frac{A_{amo}A_{ure}}{A_{amo,ure}}, \quad M^o_{amo} = M^o_{ure}\frac{A_{amo,ure}}{A_{ure}}. \tag{15}$$

The normalized reaction kinetics h_{ure}, h_{amo} and h_{nxr} are Monod-functions of substrate concentrations (Jin et al., 2013), i.e. linear at lower and saturating at higher concentrations:

$$h_{ure} = \frac{C_{\text{urea}}}{K_{ure} + C_{\text{urea}}}, \quad h_{amo} = \frac{C_{\text{NH}_3}}{K_{amo} + C_{\text{NH}_3}}, \quad h_{nxr} = \frac{C_{\text{NO}_2^-}}{K_{nxr} + C_{\text{NO}_2^-}}. \tag{16}$$

Here, K_{ure}, K_{amo} and K_{nxr} are half-saturation constants. Note that no cell yield factors or cell-specific rates appear in the model; instead, growth dynamics are completely captured by the cross-amplification factors A_{ure}, A_{amo}, A_{nxr} and $A_{amo,ure}$.

3.1.3. Model goodness of fit

We fixed 4 out of 11 model parameters to values from the literature. For example, the self-amplification factors A_{amo} and A_{nxr} were set to 1.2 d^{-1} and 1.03 d^{-1}, according to typical maximum growth rates of *Nitrosospira* (Belser and Schmidt, 1980) and *Nitrobacter* (Keen and Prosser, 1987), respectively. The initial *ure* rate, M^o_{ure}, was determined directly from the slope of the urea time series at time $t = 0$, eliminating the need for initial cell counts as in conventional cell-centric models. The remaining free parameters were calibrated to the experimental data using a maximum-likelihood approach (see Section 2 for details and Table A.1 for parameter values).

Upon calibration, the model largely explains the experimental data and is able to capture the self- and cross-amplifying character of the incubator's dynamics (Fig. 2). In particular, the *ure-amo* cross-amplification causes an increase of the system's *amo* capacity during urea hydrolysis, even when *amo* rates are still slow. This results in a fundamentally different behavior of the system than could have been explained by a cell- or reaction-centric model not accounting for the co-occurrence of *ure* and *amo* in the same cells. To verify this interpretation, we tested a variation of the model in which the cross-amplification factors $A_{ure,amo}$ and $A_{amo,ure}$ are set to zero. In this model variant, the initial capacities M^o_{ure} and M^o_{amo} became independent parameters. Similarly, λ_{AOB} was split into two independent maintenance rates, λ_{amo} and λ_{nxr}. The resulting larger set of free parameters was fitted to the same data as above. This model variant was unable to explain the NH_4^+ and NO_3^- time series, despite the higher number of calibrated parameters (Supplementary Figure S1.6). We concluded that the early increase of *amo* reaction capacity cannot be explained solely on grounds of *amo* self-amplification, but was indeed partly fueled by *ure* activity. This highlights the importance of taking into account pathway co-occurrences and interactions in cells, and suggests that cross-amplification factors in reaction-centric models may be an adequate means to that end.

3.2. Example 2: Nitrification in a flow-through bioreactor

3.2.1. The problem of state reconstruction

In principle, reaction-centric models predict future system trajectories ($\mathbf{M}(t)$, $\mathbf{C}(t)$) given initial conditions ($\mathbf{M}(0)$, $\mathbf{C}(0)$). In practice, uncertainty in initial conditions or model parameters, as well as neglected secondary processes, lead to uncertainties in the predicted system state that can increase with time. Selected measurements can provide crucial information to ensure model proximity to reality, however typically only a subset of state variables may be measurable. Inferring a system's full state from a smaller set of observations is a common problem, for example in oceanography or engineering, and generally multiple sequential measurements are used to gradually improve state reconstruction and model predictions (Bertino et al., 2003; Camacho and Bordons Alba, 2004). In this example we demonstrate how long-term, purely abiotic chemical time series can be combined with a reaction-centric model in order to infer the full state of a bioreactor in real time.

3.2.2. Model structure

The model describes a flow-through ammonium-fed nitrifying bioreactor, resembling the experimental setup by Dumont et al. (2009, Bioreactor B). In our model we assume that each mol NH_4^+ is oxidized to one mol NO_2^- by *amo* and subsequently to one mol NO_3^- by *nxr* (Wiesmann, 1994). The model thus keeps track of the bioreactor's *amo* and *nxr* reaction capacities as well as extracellular NH_4^+, NO_2^- and NO_3^- concentrations over time:

$$\frac{dM_{amo}}{dt} = A_{amo}H_{amo} - M_{amo}\cdot\mu, \tag{17}$$

$$\frac{dM_{nxr}}{dt} = A_{nxr}H_{nxr} - M_{nxr}\cdot\mu, \tag{18}$$

$$H_{amo} = M_{amo}h_{amo}, \quad H_{nxr} = M_{nxr}h_{nxr}, \tag{19}$$

$$\frac{dC_m}{dt} = S_{m,amo}H_{amo} + S_{m,nxr}H_{nxr} + (C^{\text{in}}_m - C_m)\cdot\mu. \tag{20}$$

Here, C_m is the concentration of the m-th metabolite (NH_4^+, NO_2^- or NO_3^-), $S_{m,amo}$ and $S_{m,nxr}$ are the stoichiometric coefficients for metabolite m in *amo* and *nxr*, respectively, μ is the bioreactor's dilution rate (causing the bulk of biomass decay in the bioreactor), and C^{in}_m is the metabolite concentration in the input medium (zero for all metabolites except NH_4^+). During the original experiment, the dilution rate as well as the input NH_4^+ concentration were varied on several occasions (Fig. 3c and d), resulting in non-equilibrium bioreactor dynamics. Hence, in our model both μ and $C^{\text{in}}_{\text{NH}_4^+}$ depend explicitly on time in the same way as in the original experiment (Fig. 3c and d). The normalized reaction kinetics, h_{amo} and h_{nxr}, are Monod-functions of NH_3 and NO_2^- concentrations, respectively, as in the previous example.

3.2.3. Model calibration and "real-time" state reconstruction

The concentrations of NH_4^+, NO_2^- and NO_3^- in the bioreactor were monitored throughout, providing a subset of the bioreactor's state variables. The remaining state variables (i.e. the reaction capacities) were inferred through gradual assimilation of these time series into the model, as follows. At each point in time the derivatives of the NH_4^+ and NO_3^- time series, which represent rates of change of NH_4^+ and NO_3^- concentrations, were used to infer the reaction rates (H_{amo} and H_{nxr}) after subtracting the part explained by the known dilution and substrate supply rates (Fig. 3c and d). Next, we inserted the inferred reaction rates into Eqs. (17) and (18) to predict the growth rates of *amo* and *nxr* capacities that would correspond to these reaction rates:

$$\frac{d\hat{M}_r}{dt} = A_rH_r - \hat{M}_r\cdot\mu. \tag{21}$$

Integrating Eq. (21) over time yields estimates, $\hat{M}_{amo}(t)$ and $\hat{M}_{nxr}(t)$, for the reaction capacities (Fig. 3h and i). Due to the decay rate μ, any initial discrepancies between the estimated and true capacities quickly decay exponentially regardless of initial conditions, provided that model parameters are correctly chosen (see below):

$$\frac{d}{dt}(\hat{M}_r - M_r) = -\mu\cdot(\hat{M}_r - M_r), \tag{22}$$

This method of gradual state reconstruction (Fig. 3b) is analogous to the use of so-called "observers" in control theory, which gradually approach the system's unknown state with time by combining sequential observations with concurrent model predictions

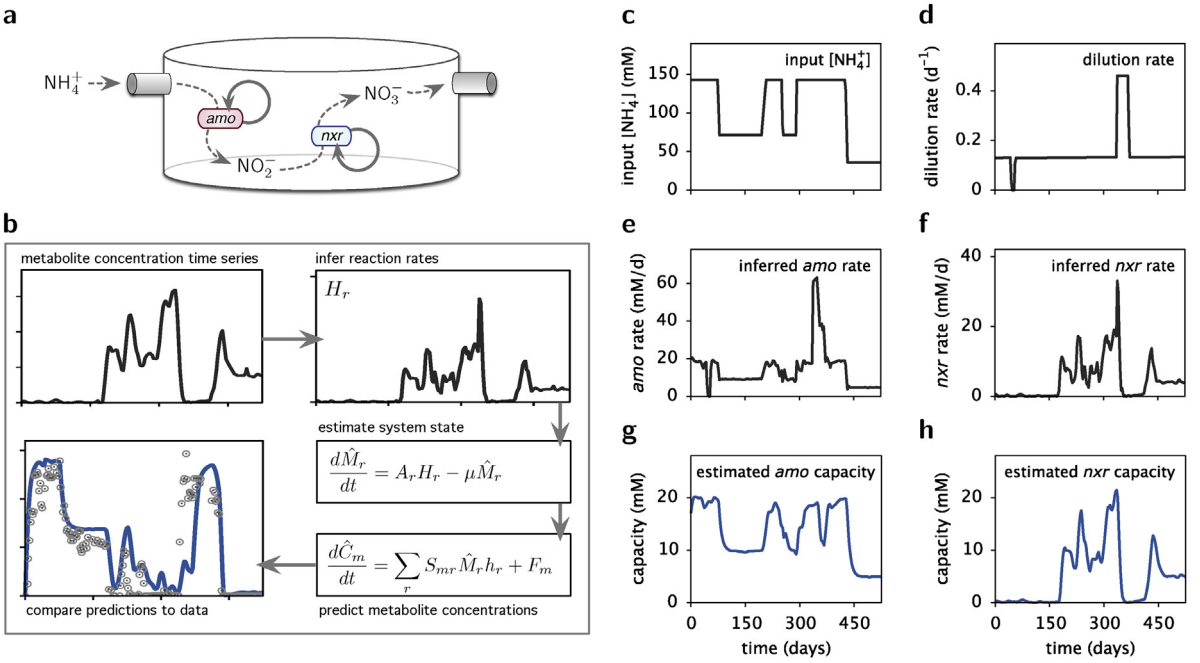

Fig. 3. Reconstructing a bioreactor's state using chemical time series. (a) Reaction-centric illustration of a flow-through nitrifying bioreactor, corresponding to experiments by Dumont et al. (2009). Continuous loop-arrows represent self-amplifications of ammonium oxidation (*amo*) and nitrite oxidation (*nxr*). (b) Methodological overview for model-based inference of the bioreactor's state using chemical time series, as performed in this paper. Reaction rates are inferred from the derivative of metabolite concentration time series. These reaction rates, in turn, are used in the reaction-centric model to predict the growth of the corresponding reaction capacities and hence the trajectory of the system's state. A comparison of predictions with the original data can be used to calibrate and validate the model. Right panel: (c) Input NH_4^+ concentration, (d) dilution rate, (e) inferred *amo* rate, (f) inferred *nxr* rate, (g) estimated *amo* capacity and (h) estimated *nxr* capacity over time.

(Sontag, 2013). In general, finding appropriate observes for the available data and ensuring their convergence can be challenging, and our example shows that the special structure of reaction-centric models mitigates this problem. Note that the temporal resolution of the chemical data, as opposed to single snapshots, is key to estimating the reaction rates needed for full-dimensional state reconstruction (Fig. 4). We note that our reaction-centric approach presents an alternative to the approach taken in the original experiment, where biomolecular time series data are assimilated into a cell-centric model (Dumont, 2008; Dumont et al., 2009).

To validate the estimated bioreactor state, we used Eqs. (19) and (20) to predict the time courses of the metabolite concentrations corresponding to the estimated $\hat{M}_r$, and these predictions were then compared to the measured NH_4^+, NO_2^- and NO_3^- concentrations. The *amo* and *nxr* half-saturation constants (K_{amo} and K_{nxr}), as well as the self-amplification factors (A_{amo} and A_{nxr}), were a priori unknown and were calibrated via least-squares fitting of the predicted metabolite concentrations to the data (see Appendix B for details and Table B.1 for fitted values). Hence, the chemical time series were used both for model calibration as well as state reconstruction. Only data from days 1–250 were used for the calibration; the remaining data (days 250–525) were used to assess the adequacy of the model for explaining the experimental observations.

Within the calibration period the model is able to reproduce most major patterns of NO_2^-, NO_3^- and, to a lesser extend, NH_4^+ concentrations (Fig. 5a–c). This indicates that the bioreactor's state is well estimated by the model during that time. The agreement

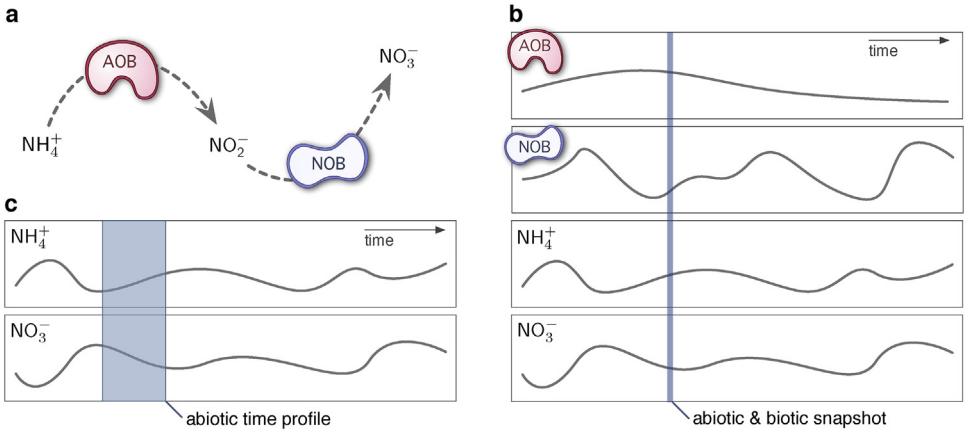

Fig. 4. Information needed to estimate the state of a reaction-centric model. (a) Illustration of a nitrifying microbial community: AOB oxidize ammonium (NH_4^+) to nitrite (NO_2^-), which is subsequently oxidized by NOB to nitrate (NO_3^-). (b) In a cell-centric framework, both abiotic (e.g. physicochemical) and biotic (e.g. cell density) measurements are required for a complete description of the system's state at any particular moment in time ("snapshot"). (c) In a reaction-centric framework, the system's state can be reproduced based on purely abiotic measurements, however measurements across multiple time points are needed ("time profile").

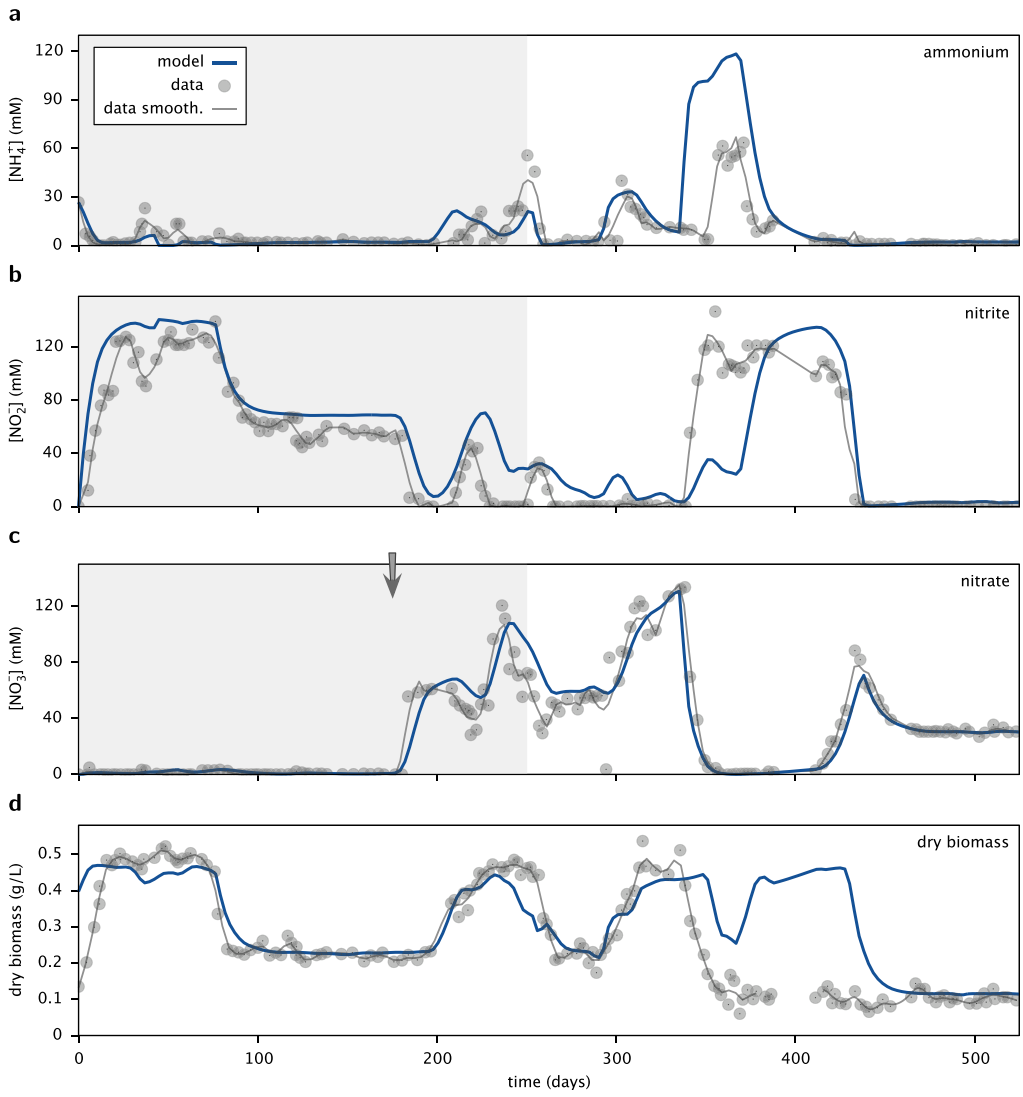

Fig. 5. Model predictions and data for Example 2. (a) Ammonium, (b) nitrite, (c) nitrate and (d) dry biomass concentration in the flow-through nitrifying bioreactor, as predicted by the data-driven model (thick blue curve) and compared to experimental data (dots). The thin grey curves show smoothened, i.e. noise-reduced, approximations of the data (see Section 2 for details). The shaded regions in (a–c) mark the data that were used for model calibration. Data in the white region were ignored during calibration and serve as an independent validation of the model. The arrow in (c) indicates the delayed onset of *nxr* after the temperature of the bioreactor was reduced from 30 °C to 25 °C on day 181. The unknown biomass yield factor, required for comparing the reaction-centric model to biomass measurements in (d), was calibrated using least-squares fitting (see the main text for details). Data by Dumont et al. (2009, Bioreactor B). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

between the model and the NH_4^+ and NO_2^- data decreases outside of the calibration period, although NO_3^- predictions remain accurate. In particular, the model overestimates the temporary accumulation of NH_4^+ on days 337–380, during which a higher dilution rate was applied to the bioreactor (Fig. 3d). An increase of residual substrate concentration at higher dilution rates, as predicted by our model, is consistent with standard bioreactor theory (McDuffie, 1991). An explanation for the absence of NH_4^+ accumulation in the data could be the potential appearance of an alternative opportunistic ammonium oxidizer that achieves faster growth rates at high substrate concentrations, thus maintaining the residual NH_4^+ below the model's predictions. Indeed, this scenario is supported by molecular analyses in the original experiment, which showed that a previously rare phylotype had emerged temporarily during that period (Dumont et al., 2009).

3.2.4. Comparison with biomass concentration profiles

The reaction-centric model in the above example does not, a priori, require or predict biomass concentrations or cell densities. After all, its purpose is to shift the focus towards system-wide reaction kinetics, and away from the microbial populations that catalyzed them. Nevertheless, biotic data (if available) can be used as an additional means to test the accuracy of a reaction-centric model. In the following we shall compare our model's predicted reaction capacities (which are proportional to biomass concentrations) to independent dry biomass concentrations measured during the original experiment (Dumont et al., 2009).

We assumed that the bulk of biomass can be attributed to ammonium oxidizers, an assumption typically met in practice (Wiesmann, 1994; Dumont et al., 2009). It then follows that $Y_{amo}M_{amo}/A_{amo}$ should be comparable to the biomass concentration, with Y_{amo} being an unknown biomass yield factor. Note that Y_{amo} simply rescales the predicted time profile of M_{amo}/A_{amo}. Hence, Y_{amo} can be estimated in retrospect by choosing Y_{amo} such that $Y_{amo}M_{amo}/A_{amo}$ best resembles the measured biomass profile. Ordinary linear least-squares fitting yields an estimate of $Y_{amo} \approx 3.2$ g dW/mol N (Fig. 5d). This estimate is greater than typical yield factors for AOB (e.g. 2.1 g dW/mol N for *Nitrosomonas*

europaea; Wiesmann, 1994), although higher yield factors have also been reported (Meurant, 1989). Other microbial groups such as NOB or non-nitrifiers likely also contribute to total biomass, resulting in an overestimate of Y_{amo}. For example, heterotrophic bacteria were detected in the original experiment using molecular methods (Dumont et al., 2009).

While the model is consistent with chemical measurements during most of the experiment as discussed previously (Fig. 5), it clearly overestimates biomass concentrations during days 380–420 (Fig. 5d). At that time, the input substrate concentration was high and the dilution rate was low (Fig. 3c and d), in principle allowing for high equilibrium cell densities. Previous models for this system based on molecular data show a similar discrepancy (Dumont, 2008). Both Dumont's and our model assume a constant yield factor, ignoring the fact that the microbial community is subject to continuous taxonomic turnover (Dumont et al., 2009). Previous bioreactor experiments have repeatedly revealed rapid taxonomic turnover and fluctuations in biomass densities, despite stable metabolic performance (Fernandez et al., 2000; Wittebolle et al., 2008). This discrepancy between reaction rates and community composition is often attributed to functional redundancy within microbial communities (Briones and Raskin, 2003; Louca and Doebeli, 2015b), and highlights an important limitation of reaction-centric models: Namely, reaction-centric models may explain ecosystem reaction rates, but they can fail to detect microbial community changes when functional performance remains stable. Multiple reaction capacities representing equivalent reactions may be included in a model to account for functional redundancy, however this will typically compromise parameter identifiability.

3.3. Estimating concentrations of other organic compounds

In the last example above we assessed the adequacy of our reaction-centric model using independent biomass concentration measurements by introducing an additional biomass yield factor, which related dissimilatory nitrogen fluxes to biosynthesis rates. Similarly, reaction-centric models may also predict the concentration of other organic compounds or elements, either for model validation using additional data or for addressing particular ecological questions. For example, organic nitrogen or carbon concentration profiles can yield insight into nitrogen fixation rates and productivity at ecosystem scales (Carlson et al., 1994; Sterner and Elser, 2002). The concentrations of various compounds in living cells (e.g., organic N) can be derived from the reaction capacities using so called assimilation factors, which represent the amount of compound assimilated or synthesized per reaction flux (Wiesmann, 1994). More precisely, the concentration of a particular organic compound is given by the matrix product

$$\mathbf{T}^{\mathrm{T}}\mathbb{A}^{-1}\mathbf{M}, \tag{23}$$

where $\mathbf{M}$ is the column vector containing all reaction capacities, $\mathbf{T}$ is the column-vector containing the assimilation factors for the compound for the various reactions, $\mathbf{T}^{\mathrm{T}}$ is the transpose of $\mathbf{T}$ and $\mathbb{A}^{-1}$ is the inverse of the amplification matrix (see Supplement S1.2 for a derivation). For example, the stoichiometries of N-metabolism and anabolism in the ammonium oxidizer *N. europaea* and nitrite oxidizer *Nitrobacter winogradskyi* are usually summarized by

$$55\mathrm{NH_4^+} + 76\mathrm{O_2} + 109\mathrm{HCO_3^-} \rightarrow \mathrm{C_5H_7NO_2} + 54\mathrm{NO_2^-} + 57\mathrm{H_2O} + 104\mathrm{H_2CO_3} \tag{24}$$

and

$$400\mathrm{NO_2^-} + \mathrm{NH_4^+} + 4\mathrm{H_2CO_3} + \mathrm{HCO_3^-} + 195\mathrm{O_2} \rightarrow \mathrm{C_5H_7NO_2} + 3\mathrm{H_2O} + 400\mathrm{NO_3^-}, \tag{25}$$

respectively (Wiesmann, 1994). Here, $\mathrm{C_5H_7NO_2}$ represents biomass. Hence, for organic N the assimilation factors are $T_{amo} = 1:55 \approx 0.018$ (1 mol N assimilated per $\mathrm{NH_4^+}$ consumed) for dissimilatory ammonium oxidation

$$amo: \mathrm{NH_4^+} + \frac{3}{2}\mathrm{O_2} \rightarrow \mathrm{NO_2^-} + \mathrm{H_2O} + 2\mathrm{H^+}, \tag{26}$$

and $T_{nxr} = 1:400 \approx 0.0025$ for dissimilatory nitrite oxidation

$$nxr: \mathrm{NO_2^-} + \frac{1}{2}\mathrm{O_2} \rightarrow \mathrm{NO_3^-}. \tag{27}$$

In other cases (e.g. when stoichiometries are unknown) assimilation factors may be estimated through linear least-squares fitting, as demonstrated above for total biomass.

3.4. Limitations and extensions of reaction-centric models

The reaction-centric models presented in this study were formulated in terms of ordinary differential equations that describe the temporal evolution of the chemical and reaction-kinetic state of a well-mixed (i.e. spatially homogenous) system. Spatial extensions, for example comprising multiple interacting compartments or formulated as partial (i.e. spatiotemporal) differential equations, are equally possible. Such extensions may be used to describe the biogeochemistry in the ocean water column (Reed et al., 2014) or in multi-stage industrial processes (Prokop et al., 1969).

For simplicity, we only considered Monod-type reaction kinetics, which capture the non-linear and saturating dependence of reaction rates on single substrate concentrations, but which ignore potential substrate inhibition effects or multi-substrate dependencies. For example, excess ammonia and nitrous acid concentrations in nitrifying bioreactors can cause inhibition of the very pathways that consume them (Anthonisen et al., 1976), and this substrate inhibition can result in reduced bioreactor performance (Sheintuch et al., 1995). Similarly, the accumulation of metabolic products can inhibit pathway activity, e.g. by rendering pathways energetically unfavorable (LaRowe et al., 2012), thereby slowing down reaction rates or even causing a decline of reaction capacities due to cell death (Kaspar and Wuhrmann, 1977; Conrad, 1999). In reaction-centric models, substrate or product inhibition as well as multi-substrate dependencies can be incorporated through appropriate normalized reaction kinetics, $h_r(\mathbf{C})$, for example in the form of multi-substrate Michaelis–Menten functions with inhibition terms (e.g. Thullner et al., 2007).

We note that reaction-centric models are not appropriate for capturing complex heterogeneities in the physiology or metabolic activity within populations that may be caused, for example, by stochastic regulatory switching (Ackermann, 2015). Simple heterogeneities, e.g. involving a small set of alternative metabolic phenotypes, may be accounted for by including multiple reactions whose capacities are coupled through cross-amplification factors. However, when variation between individual cells involves multiple traits or spans a continuum of values, individual-based models (Ferrer et al., 2008; Larsen et al., 2012) may be more appropriate for incorporating that variation. Moreover, reaction-centric descriptions eliminate cell-centric information (e.g. cell densities of particular species or strains) that is potentially needed to model additional community-level processes such as predation by protists (Güde, 1979) or bacteriophages (Shapiro and Kushmaro, 2011). For example, bacteriophages adapted to specific bacterial taxa can exert strong control on their host populations and can drive rapid turnover of competing bacterial taxa through "killing the winner" dynamics (Suttle, 2007; Shapiro and Kushmaro, 2011). Such taxonomic turnover within microbial "metabolic guilds" cannot be captured by reaction-centric models, although in several previous bioreactor experiments the overall biochemical throughput

remained constant despite rapid taxonomic turnover (Fernández et al., 1999; Hashsham et al., 2000; Wittebolle et al., 2008; Wang et al., 2011) and hence, reaction-centric models may be adequate for such systems. Other biotic interactions, such as chemical warfare (Riley and Wertz, 2002) or quorum sensing (Fuqua et al., 2001) may also necessitate the use of cell-centric (e.g. individual-based) models.

4. Conclusions

Marker gene profiling of taxonomic community composition has become a standard tool in microbial ecology and bioengineering (Wittebolle et al., 2008; Dumont et al., 2009). However, taxonomic profiles can lead to ambiguous conclusions about metabolic processes due to functional redundancy across microbial clades, fine-scale ecological differentiations and poor functional characterization of species (Kashtan et al., 2014; Dumont et al., 2009). In fact, microbial communities can have highly variable taxonomic composition while maintaining stable overall reaction rates, as has been repeatedly demonstrated in bioreactors (Fernandez et al., 2000; Wittebolle et al., 2008; Wang et al., 2011). Furthermore, the measurement of biotic variables such as enzyme concentrations and molecular profiles often presents practical challenges (Vojinović et al., 2006). These observations motivate the pursuit for reaction-centric descriptions of microbial ecosystems that can fully utilize abiotic physicochemical data and minimize the need for laborious biotic measurements. This is particularly important in bioprocess and environmental engineering, where the need for real-time and unambiguous state reconstruction imposes strong requirements on the data (Lazar et al., 2007).

Here we have shown how a reaction-centric model enables the inference of a bioreactor's state, from a reaction kinetic point of view, based solely on chemical data. Reaction-centric models can capture the self- and cross-amplifying nature of biocatalyzed processes that so strongly sets them apart from most non-living systems. This is achieved through an amplification matrix that translates system-wide reaction rates to changes in system-wide reaction capacities. Because the amplification matrix can contain off-diagonal entries it can account for pathway co-occurrences in cells, as we have demonstrated for the case of urea hydrolysis and ammonium oxidation. The elegance of reaction-centric models makes them a potentially powerful alternative to many conventional models describing microbial metabolic activity at ecosystem scales.

Conflict of interests

The authors declare that they have no conflicts of interest.

Acknowledgements

We thank Jérôme Harmand and Alain Rapaport (Institut National de la Recherche Agronomique, France) for additional data and details of their experiments (Dumont et al., 2009). S.L. acknowledges the financial support of the Department of Mathematics, UBC. S.L. and M.D. acknowledge the support of NSERC.

Appendix A. Details on example 1 (batch-fed incubator)

A.1. Parameterization

In the model, temperature was held constant at 20 °C and pH was held constant at 5, in accordance with the original incubation experiment (de Boer and Laanbroek, 1989). Ammonia and ammonium were assumed to be at dissociation equilibrium, determined by the pH and the standard ammonium dissociation constant 5.69×10^{-10} M (Clegg and Whitfield, 1995). The dissociation constant was corrected for the lower temperature in the experiment using the Van't Hoff equation (Atkins and de Paula, 2012).

The initial *ure* capacity M^o_{ure} was estimated from the derivative of the urea time series, assuming that the initial *ure* kinetics were saturated by high substrate concentration. Time series derivatives were estimated via 4th order Savitzky–Golay smoothening with a sliding window span of 10 days (Kantz and Schreiber, 2004), followed by centered finite differences. The initial urea, NH_4^+ and NO_3^- concentrations were set to 1.12 mM, 124 μM and 49.8 μM, respectively, according to the first sampling point in the measured time series. The initial NO_2^- concentration was assumed to be zero. The parameters K_{ure}, K_{nxr}, A_{amo} and A_{nxr} were taken from existing literature on *Nitrosospira* and *Nitrobacter* (Table A.1).

The remaining free parameters K_{amo}, M^o_{nxr}, A_{ure}, $A_{amo,ure}$, λ_{AOB} and ρ_{ure} were simultaneously calibrated to the urea, NH_4^+ and NO_3^- time series via maximum-likelihood estimation (Eliason, 1993). This approach estimates unknown parameters by maximizing the likelihood of observing the available data given a particular candidate choice of parameter values. Maximum likelihood estimation is widely used in statistical inference such as multilinear regression and physics (Lyons, 1986). In our case, the likelihood of the data was calculated on the basis of a mixed deterministic–stochastic

Table A.1
Fixed and fitted model parameters for the batch bioreactor incubated with *Nitrosospira* sp. and *Nitrobacter* sp. Parameters marked with an asterisk (⋆) were fitted to the time series, and are compared to literature values where available. SE refers to the standard error of the fitted value, in the same units. The initial metabolite concentrations C^o_{urea}, $C^o_{NH_4^+}$ and $C^o_{NO_3^-}$ were taken from the chemical time series on day 1. The initial reaction capacity M^o_{ure} was estimated from the slope of the time series at time zero. The remaining parameter values were taken from the indicated literature.

Param.		Value	SE	Comparison	Group	Literature
K_{ure}		670 μM urea	–	–	*Nitrosospira* L115	Jiang and Bakken (1999)
K_{amo}	⋆	4.59 μM NH_3	±0.27	6–11	*Nitrosospira* spp.	Jiang and Bakken (1999)
K_{nxr}		27.2 μM NO_2^-	–	–	*Nitrobacter* spp.	Blackburne et al. (2007)
A_{ure}	⋆	1.11 d^{-1}	±0.004	–	–	–
A_{amo}		1.2 d^{-1}	–	–	*Nitrosospira* AV2	Belser and Schmidt (1980)
A_{nxr}		1.03 d^{-1}	–	–	*Nitrobacter* sp.	Keen and Prosser (1987)
$A_{amo,ure}$	⋆	12.8 d^{-1}	±0.54	–	–	–
λ_{AOB}	⋆	0.0055 d^{-1}	±0.0005	0.027 d^{-1}	*N. europaea*	Tappe et al. (1999)
ρ_{ure}	⋆	0.26	±0.018	–	–	–
C^o_{urea}		1.12 mM	–	–	–	
$C^o_{NH_4^+}$		124 μM	–	–	–	
$C^o_{NO_2^-}$		0	–	–	–	
$C^o_{NO_3^-}$		49.8 μM	–	–	–	
M^o_{ure}		773 nM/d	–	–	–	–
M^o_{nxr}	⋆	35.8 mM/d	±1.4	–	–	–

Table B.1
Fixed and fitted model parameters for the flow-through bioreactor (Dumont et al., 2009). Parameters marked with an asterisk (⋆) were calibrated using data from days 1–250 and are compared to literature values. The initial metabolite concentrations $C^o_{NH_4^+}$, $C^o_{NO_2^-}$ and $C^o_{NO_3^-}$ were taken from the chemical time series on day 1. The initial reaction capacities M^o_{amo} and M^o_{nxr} were estimated from the slopes of the chemical time series on day 1. The parameters $C^{in}_{NH_4^+}$, μ, pH and temperature were controlled throughout the experiment.

Param.		Value	Comparison	Group	Literature
K_{amo}	⋆	3.21 μM NH_3	1.2–23	AOB	Suzuki et al. (1974), Ward (1987)
K_{nxr}	⋆	1.32 mM NO_2^-	0.01–1.68	NOB	Hunik et al. (1993), Maixner et al. (2006)
A_{amo}	⋆	0.145 d^{-1}	0.32–2.1	AOB	Jang et al. (2005), Prosser (2005)
A_{nxr}	⋆	0.176 d^{-1}	0.17–1.4	NOB	Belser (1979), Prosser (2005)
$C^o_{NH_4^+}$		26.7 mM	–	–	–
$C^o_{NO_2^-}$		0	–	–	–
$C^o_{NO_3^-}$		0	–	–	–
M^o_{amo}		17.1 mM/d	–	–	–
M^o_{nxr}		0	–	–	–
$C^{in}_{NH_4^+}$		35.7–143 mM	–	–	–
μ		0–0.46 d^{-1}	–	–	–
pH		7	–	–	–
Temperature		30–25 °C	–	–	–

structure, in which the deterministic part is given by the reaction-centric model and errors are assumed to be normally distributed on a logarithmic scale. The likelihood was maximized using the SBPLX optimization algorithm (Johnson, 2014), which uses repeated simulations and gradual exploration of parameter space. To reduce the possibility of only reaching a local maximum, fitting was repeated 100 times using random initial parameter values and the best fit among all 100 runs was used. Parameter confidence intervals were calculated using the inverse observed Fisher information, which is an estimator of the parameter covariance matrix (Davidson and MacKinnon, 2004). Fitted parameter values, their confidence intervals and a comparison to available literature are given in Table A.1.

A.2. Assessing the importance of ure-amo cross-amplification

To test the suitability of a model variant without *ure-amo* cross-amplification as outlined in the main text, we treated *ure* and *amo* as independent reactions performed by separate cell populations. Hence, we assumed $A_{amo,ure} = A_{ure,amo} = 0$ and $\rho_{ure} = 0$, and replaced the maintenance rate λ_{AOB} with two independent rates λ_{amo} and λ_{ure}. Furthermore, the initial capacities M^o_{ure} and M^o_{amo} were treated as independent parameters. The new set of free parameters thus comprised K_{amo}, M^o_{amo}, M^o_{nxr}, A_{ure}, A_{amo}, λ_{ure} and λ_{amo}, while the remaining parameters were fixed as described above. Fitting was performed as with the original model and yielded multiple local optima, none of which matched the data as well as the original model (Supplementary Figure S1.6).

Appendix B. Details on example 2 (flow-through bioreactor)

B.1. Assimilation of time series

Experimental time series of NH_4^+ and NO_3^- concentrations were noise-filtered using 4th order Savitzky-Golay smoothening with a sliding window time span of 30 days (Kantz and Schreiber, 2004). Derivatives of concentration profiles were estimated by applying a centered finite differences scheme to the noise-filtered profiles. *amo* and *nxr* rates were estimated from the derivatives of the NH_4^+ and NO_3^- concentration profiles, respectively, after accounting for substrate input and dilution. Estimated *amo* and *nxr* rates were then used in the growth model for the reaction capacities, Eqs. (17) and (18), as described in the main text.

B.2. Parameterization

In the experiment, pH was maintained around 7 by the automatic addition of an alkaline solution, and the bioreactor was maximally ventilated to ensure sufficient oxygenation (Dumont et al., 2009). In our model we thus assumed pH 7 and ignored oxygen limitation in the reaction kinetics. Temperature was assumed to be 30 °C until day 181 and 25 °C afterwards, in accordance with the original experiment. Bioreactor dilution rates and input substrate concentrations were obtained from the authors of the original experiment upon personal correspondence. NH_3 concentration was calculated from NH_4^+ concentration by assuming that the two are at acid-dissociation equilibrium, similarly to the first example.

The initial *amo* capacity, M^o_{amo}, was estimated from the NH_4^+ time series but had negligible effects on the simulations. The initial *nxr* capacity was set to zero based on the absence of NO_3^- accumulation. The *amo* and *nxr* half-saturation constants and the self-amplification factors A_{amo} and A_{nxr} were calibrated to the NH_4^+, NO_2^- and NO_3^- time series by maximizing the mean coefficient of determination (R^2) across all three data sets, which is analogous to weighted least-squares fitting in the univariate case. Only data from days 1–250 were used for the calibration. The mean R^2 was maximized using the SBPLX algorithm (Johnson, 2014). To reduce the possibility of only reaching a local maximum, fitting was repeated 100 times using random initial parameter values and the best fit among all 100 runs was used. Fitted parameter values and a comparison to available literature are given in Table B.1.

Appendix C. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ecolmodel.2016.05.011.

References

Ackermann, M., 2015. A functional perspective on phenotypic heterogeneity in microorganisms. Nat. Rev. Microbiol. 13 (8), 497–508.
Anthonisen, A.C., Loehr, R.C., Prakasam, T.B.S., Srinath, E.G., 1976. Inhibition of nitrification by ammonia and nitrous acid. J. Water Pollut. Control Federation 48 (5), 835–852.
Atkins, P., de Paula, J., 2012. Elements of Physical Chemistry. OUP Oxford.
Belser, L., Schmidt, E., 1980. Growth and oxidation kinetics of three genera of ammonia oxidizing nitrifiers. FEMS Microbiol. Lett. 7 (3), 213–216.
Belser, L.W., 1979. Population ecology of nitrifying bacteria. Annu. Rev. Microbiol. 33 (1), 309–333.

Bertino, L., Evensen, G., Wackernagel, H., 2003. Sequential data assimilation techniques in oceanography. Int. Stat. Rev. 71 (2), 223–241.
Blackburne, R., Vadivelu, V.M., Yuan, Z., Keller, J., 2007. Kinetic characterisation of an enriched *Nitrospira* culture with comparison to *Nitrobacter*. Water Res. 41 (14), 3033–3042.
Briones, A., Raskin, L., 2003. Diversity and dynamics of microbial communities in engineered environments and their implications for process stability. Curr. Opin. Biotechnol. 14 (3), 270–276.
Camacho, E.F., Bordons Alba, C., 2004. Model Predictive Control. Springer.
Carlson, C.A., Ducklow, H.W., Michaels, A.F., 1994. Annual flux of dissolved organic carbon from the euphotic zone in the northwestern Sargasso Sea. Nature 371 (6496), 405–408.
Cheyns, K., Mertens, J., Diels, J., Smolders, E., Springael, D., 2010. Monod kinetics rather than a first-order degradation model explains atrazine fate in soil mini-columns: implications for pesticide fate modelling. Environ. Pollut. 158 (5), 1405–1411.
Chung, B.K.S., Lee, D.-Y., 2009. Flux-sum analysis: a metabolite-centric approach for understanding the metabolic network. BMC Syst. Biol. 3 (1), 1–10.
Clegg, S.L., Whitfield, M., 1995. A chemical model of seawater including dissolved ammonia and the stoichiometric dissociation constant of ammonia in estuarine water and seawater from −2 to 40 °C. Geochim. Cosmochim. Acta 59 (12), 2403–2421.
Conrad, R., 1999. Contribution of hydrogen to methane production and control of hydrogen concentrations in methanogenic soils and sediments. FEMS Microbiol. Ecol. 28 (3), 193–202.
Courtin, M.G., Spoelstra, S.F., 1990. A simulation model of the microbiological and chemical changes accompanying the initial stage of aerobic deterioration of silage. Grass Forage Sci. 45 (2), 153–165.
Davidson, R., MacKinnon, J., 2004. Econometric Theory and Methods. Oxford University Press.
de Boer, W., Laanbroek, H., 1989. Ureolytic nitrification at low pH by *Nitrosospira sp.* Arch. Microbiol. 152 (2), 178–181.
Dumont, M., 2008. Apports de la modelisation des interactions pour une comprehension fonctionnelle d'un ecosysteme (Ph.D. thesis). Universite Montpellier II.
Dumont, M., Harmand, J., Rapaport, A., Godon, J.-J., 2009. Towards functional molecular fingerprints. Environ. Microbiol. 11 (7), 1717–1727.
Eliason, S.R., 1993. Maximum Likelihood Estimation: Logic and Practice. SAGE Publications, Newbury Park, CA.
Falkowski, P.G., Fenchel, T., Delong, E.F., 2008. The microbial engines that drive Earth's biogeochemical cycles. Science 320 (5879), 1034–1039.
Fernández, A., Huang, S., Seston, S., Xing, J., Hickey, R., Criddle, C., Tiedje, J., 1999. How stable is stable? Function versus community composition. Appl. Environ. Microbiol. 65 (8), 3697–3704.
Fernandez, A.S., Hashsham, S.A., Dollhopf, S.L., Raskin, L., Glagoleva, O., Dazzo, F.B., Hickey, R.F., Criddle, C.S., Tiedje, J.M., 2000. Flexible community structure correlates with stable community function in methanogenic bioreactor communities perturbed by glucose. Appl. Environ. Microbiol. 66 (9), 4058–4067.
Ferrer, J., Prats, C., López, D., 2008. Individual-based modelling: an essential tool for microbiology. J. Biol. Phys. 34 (1), 19–37.
Fuqua, C., Parsek, M.R., Greenberg, E.P., 2001. Regulation of gene expression by cell-to-cell communication: acyl-homoserine lactone quorum sensing. Annu. Rev. Genet. 35 (1), 439–468.
Güde, H., 1979. Grazing by protozoa as selection factor for activated sludge bacteria. Microb. Ecol. 5 (3), 225–237.
Hashsham, S.A., Fernandez, A.S., Dollhopf, S.L., Dazzo, F.B., Hickey, R.F., Tiedje, J.M., Criddle, C.S., 2000. Parallel processing of substrate correlates with greater functional stability in methanogenic bioreactor communities perturbed by glucose. Appl. Environ. Microbiol. 66 (9), 4050–4057.
Hunik, J., Meijer, H., Tramper, J., 1993. Kinetics of *Nitrobacter agilis* at extreme substrate, product and salt concentrations. Appl. Microbiol. Biotechnol. 40 (2–3), 442–448.
Jang, A., Okabe, S., Watanabe, Y., Kim, I.S., Bishop, P.L., 2005. Measurement of growth rate of ammonia oxidizing bacteria in partially submerged rotating biological contactor by fluorescent in situ hybridization (FISH). J. Environ. Eng. Sci. 4 (5), 413–420.
Jiang, Q.Q., Bakken, L.R., 1999. Comparison of *Nitrosospira* strains isolated from terrestrial environments. FEMS Microbiol. Ecol. 30 (2), 171–186.
Jin, Q., Bethke, C.M., 2007. The thermodynamics and kinetics of microbial metabolism. Am. J. Sci. 307 (4), 643–677.
Jin, Q., Roden, E.E., Giska, J.R., 2013. Geomicrobial kinetics: extrapolating laboratory studies to natural environments. Geomicrobiol. J. 30 (2), 173–185.
Johnson, S.G., 2014. The NLopt Nonlinear-Optimization Package. Software.
Kantz, H., Schreiber, T., 2004. Nonlinear Time Series Analysis, 2nd ed. Cambridge University Press.
Kashtan, N., Roggensack, S.E., Rodrigue, S., Thompson, J.W., Biller, S.J., Coe, A., Ding, H., Marttinen, P., Malmstrom, R.R., Stocker, R., et al., 2014. Single-cell genomics reveals hundreds of coexisting subpopulations in wild *Prochlorococcus*. Science 344 (6182), 416–420.
Kaspar, H., Wuhrmann, K., 1977. Product inhibition in sludge digestion. Microb. Ecol. 4 (3), 241–248.
Keen, G., Prosser, J., 1987. Steady state and transient growth of autotrophic nitrifying bacteria. Arch. Microbiol. 147 (1), 73–79.
Khatri, B.S., Free, A., Allen, R.J., 2012. Oscillating microbial dynamics driven by small populations, limited nutrient supply and high death rates. J. Theor. Biol. 314, 120–129.
Klitgord, N., Segrè, D., 2010. Environments that induce synthetic microbial ecosystems. PLOS Comput. Biol. 6 (11), e1001002.
Knightes, C.D., Peters, C.A., 2000. Statistical analysis of nonlinear parameter estimation for Monod biodegradation kinetics using bivariate data. Biotechnol. Bioeng. 69 (2), 160–170.
Konishi, S., Kitagawa, G., 2008. Information Criteria and Statistical Modeling. Springer Series in Statistics. Springer, New York.
LaRowe, D.E., Dale, A.W., Amend, J.P., Van Cappellen, P., 2012. Thermodynamic limitations on microbially catalyzed reaction rates. Geochim. Cosmochim. Acta 90, 96–109.
Larsen, P., Hamada, Y., Gilbert, J., 2012. Modeling microbial communities: current, developing, and future technologies for predicting microbial community interaction. J. Biotechnol. 160 (1–2), 17–24.
Lazar, I., Petrisor, I., Yen, T., 2007. Microbial enhanced oil recovery (MEOR). Petrol. Sci. Technol. 25, 1353–1366.
Louca, S., Doebeli, M., 2015a. Calibration and analysis of genome-based models for microbial ecology. eLife 4, e08208.
Louca, S., Doebeli, M., 2015b. Transient dynamics of competitive exclusion in microbial communities. Environ. Microbiol., http://dx.doi.org/10.1111/1462-2920.13058.
Lyons, L., 1986. Statistics for Nuclear and Particle Physicists. Cambridge University Press, Cambridge, UK.
Maixner, F., Noguera, D.R., Anneser, B., Stoecker, K., Wegl, G., Wagner, M., Daims, H., 2006. Nitrite concentration influences the population structure of *Nitrospira*-like bacteria. Environ. Microbiol. 8 (8), 1487–1495.
Marjanovic, O., Lennox, B., Sandoz, D., Smith, K., Crofts, M., 2006. Real-time monitoring of an industrial batch process. Comput. Chem. Eng. 30 (10–12), 1476–1481.
Marsh, K., Sims, G., Mulvaney, R., 2005. Availability of urea to autotrophic ammonia-oxidizing bacteria as related to the fate of ^{14}C-and ^{15}N-labeled urea added to soil. Biol. Fertil. Soils 42 (2), 137–145.
McDuffie, N.G., 1991. Bioreactor Design Fundamentals. Butterworth-Heinemann, USA.
Meurant, G., 1989. Advances in Microbial Physiology. Advances in Microbial Physiology, vol. 30. Academic Press.
Orth, J.D., Thiele, I., Palsson, B.O., 2010. What is flux balance analysis? Nat. Biotechnol. 28 (3), 245–248, 03.
Prokop, A., Erickson, L.E., Fernandez, J., Humphrey, A.E., 1969. Design and physical characteristics of a multistage, continuous tower fermentor. Biotechnol. Bioeng. 11 (5), 945–966.
Prosser, J., 2005. Nitrogen in soils: nitrification. In: Hillel, D. (Ed.), Encyclopedia of Soils in the Environment. Elsevier, Oxford, UK, pp. 31–39.
Reed, D.C., Algar, C.K., Huber, J.A., Dick, G.J., 2014. Gene-centric approach to integrating environmental genomics and biogeochemical models. Proc. Natl. Acad. Sci. 111 (5), 1879–1884.
Resat, H., Petzold, L., Pettigrew, M.F., 2009. Computational Systems Biology. Humana Press, Totowa, NJ, pp. 311–335, Chapter: Kinetic Modeling of Biological Systems.
Riley, M.A., Wertz, J.E., 2002. Bacteriocins: evolution, ecology, and application. Annu. Rev. Microbiol. 56 (1), 117–137.
Shapiro, O.H., Kushmaro, A., 2011. Bacteriophage ecology in environmental biotechnology processes. Curr. Opin. Biotechnol. 22 (3), 449–455.
Sheintuch, M., Tartakovsky, B., Narkis, N., Rebhun, M., 1995. Substrate inhibition and multiple states in a continuous nitrification process. Water Res. 29 (3), 953–963.
Simkins, S., Alexander, M., 1984. Models for mineralization kinetics with the variables of substrate concentration and population density. Appl. Environ. Microbiol. 47 (6), 1299–1306.
Song, H.-S., Cannon, W.R., Beliaev, A.S., Konopka, A., 2014. Mathematical modeling of microbial community dynamics: a methodological review. Processes 2 (4), 711–752.
Sontag, E., 2013. Mathematical Control Theory: Deterministic Finite Dimensional Systems. Texts in Applied Mathematics. Springer, New York.
Sterner, R., Elser, J., 2002. Ecological Stoichiometry: The Biology of Elements from Molecules to the Biosphere. Princeton University Press.
Stolyar, S., Van Dien, S., Hillesland, K.L., Pinel, N., Lie, T.J., Leigh, J.A., Stahl, D.A., 2007. Metabolic modeling of a mutualistic microbial community. Mol. Syst. Biol. 3 (1), 92.
Suttle, C.A., 2007. Marine viruses – major players in the global ecosystem. Nat. Rev. Microbiol. 5 (10), 801–812.
Suzuki, I., Dular, U., Kwok, S.C., 1974. Ammonia or ammonium ion as substrate for oxidation by *Nitrosomonas europaea* cells and extracts. J. Bacteriol. 120 (1), 556–558.
Tappe, W., Laverman, A., Bohland, M., Braster, M., Rittershaus, S., Groeneweg, J., van Verseveld, H.W., 1999. Maintenance energy demand and starvation recovery dynamics of *Nitrosomonas europaea* and *Nitrobacter winogradskyi* cultivated in a retentostat with complete biomass retention. Appl. Environ. Microbiol. 65 (6), 2471–2477.
Thullner, M., Regnier, P., Van Cappellen, P., 2007. Modeling microbially induced carbon degradation in redox-stratified subsurface environments: concepts and open questions. Geomicrobiol. J. 24 (3–4), 139–155.
Vojinović, V., Cabral, J., Fonseca, L., 2006. Real-time bioprocess monitoring: Part I: In situ sensors. Sens. Actuators B: Chem. 114 (2), 1083–1091.

Wang, X., Wen, X., Yan, H., Ding, K., Zhao, F., Hu, M., 2011. Bacterial community dynamics in a functionally stable pilot-scale wastewater treatment plant. Bioresour. Technol. 102 (3), 2352–2357.
Ward, B., 1987. Kinetic studies on ammonia and methane oxidation by *Nitrosococcus oceanus*. Arch. Microbiol. 147 (2), 126–133.
Wiesmann, U., 1994. Biological Nitrogen Removal from Wastewater. Advances in Biochemical Engineering and Biotechnology, vol. 51. Springer Berlin Heidelberg, pp. 113–154.
Wittebolle, L., Vervaeren, H., Verstraete, W., Boon, N., 2008. Quantifying community dynamics of nitrifiers in functionally stable reactors. Appl. Environ. Microbiol. 74 (1), 286–293.
Zomorrodi, A.R., Maranas, C.D., 2012. OptCom: a multi-level optimization framework for the metabolic modeling and analysis of microbial communities. PLoS Computat. Biol. 8 (2), e1002363.