Journal of Process Control 118 (2022) 170–184

Contents lists available at ScienceDirect

Journal of Process Control

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

A systematic elementary flux mode selection procedure for deriving macroscopic bioreaction models from metabolic networks

M. Maton [a], Ph. Bogaerts [b], A. Vande Wouwer [a,*]

[a] Systems, Estimation, Control and Optimization (SECO), Université de Mons, 7000 Mons, Belgium
[b] 3-BIO-BioControl, Université Libre de Bruxelles, 1050 Brussels, Belgium

ARTICLE INFO

Article history:
Received 1 February 2022
Received in revised form 30 August 2022
Accepted 1 September 2022
Available online 19 September 2022

Keywords:
Model reduction
Dynamic models
Metabolic network
Elementary flux modes
Biotechnology
Reaction scheme

ABSTRACT

Elementary flux modes (EFMs) are an important concept in metabolic pathway analysis and for the derivation of macroscopic dynamic models. However, the computation of elementary flux vectors is facing combinatorial explosion with the size of the metabolic network, which hinders widespread application. This study proposes a systematic elementary flux mode reduction procedure to derive reduced-order dynamic models starting from an initial set of EFMs either generated by complete enumeration or subset selection. The procedure proceeds in several steps, including geometric and optimization-based criteria. The methodology ends up with a macroscopic bioreaction scheme with a reaction number smaller than that of the measured species, and shows very satisfactory prediction results, as illustrated with data of batch cultures of CHO-320 cells.

1. Introduction

Therapeutic proteins, such as monoclonal antibodies, play a key role for diagnosis purposes and in the treatment of genetic diseases [1–3]. This explains the prominent place of mammalian cell cultures in the healthcare sector. However, large-scale cell cultures face some engineering challenges including high cell densities and process reproducibility, among others [4]. In this connection, the development of dynamic models of the bioprocesses, which could serve as a basis for the construction of digital twins, and the design of control strategies is of paramount importance.

Mathematical models can be classified into structured models and unstructured models. While the first ones contain an elaborate description of the cell and a precise depiction of the interactions between intracellular compounds, the second neglect the intracellular activity and describe only the evolution of extracellular variables. Recent studies [5–8] emphasize the connection between both approaches and several methods make use of the intracellular information contained in a metabolic network to develop small size macroscopic models. Some of these techniques rely on the concept of elementary flux modes [9] or extreme pathways [10] defining a set of vectors derived from the stoichiometric matrix of the network by means of a convex analysis. This approach exploits the biological knowledge about the organism [11] and enables obtaining a set of macroscopic bioreactions linking extracellular substrates/nutrients to products. However, although the value of elementary modes for the development of dynamical models is no longer questioned, the combinatorial explosion in the number of modes with the size of the metabolic network remains an issue and still a major research focus.

For metabolic networks of modest sizes, the elementary flux modes can usually be enumerated (they can be computed using software tools such as MetaTool [12] or EFMtool [13]). The underlying metabolic networks are obtained through reduction of detailed networks achieved through metabolic flux analysis [11] or by removing all insignificant fluxes [14] yielding small number of EFMs. Conversely, when the EFMs enumeration becomes computationally prohibitive [15], pathway analysis is still achievable by computing only a subset of EFMs. To this end, several techniques have emerged in the last decades, including the computation of the shortest elementary flux modes [16], subsystem analysis [17], decomposition of the flux distribution into a minimal number of elementary flux modes [18], random sampling where random subsets of new combinations of modes are selected on the basis of a given probability function [19,20], or MFA-based optimization [21]. In the same context, for purposes of reducing the dimensionality of the solution space and building dynamic models, [22,23] add constraints related to cell-specific uptake- or secretion rates. [24,25] develop an optimization criterion compromising error, efficiency of the modes and model size and the selection of the elementary flux vectors relies on ranking or controlled random search. Subsequently, [26] makes use of the cosine-similarity algorithm to reduce roughly the number of pathways and a relevant set of EFMs is selected thanks to

* Corresponding author.
E-mail address: Alain.VANDEWOUWER@umons.ac.be (A. Vande Wouwer).

https://doi.org/10.1016/j.jprocont.2022.09.002
0959-1524/

the definition of a multi-objective genetic algorithm. Methods using dynamic metabolic flux analysis, flux balance analysis and dynamical flux balance analysis are also proposed in [8]. Recently, [21,27] have applied column generation techniques to determine subsets of elementary flux modes in metabolic networks and [28] introduces the poly-pathway model approach to account for the metabolic behavior of multiple experimental conditions. Recently, our research group has developed two procedures, e.g., [29] picks the best EFMs candidates based on a linear optimization problem, and [30] uses a two-step reduction based on cosine-linearity and optimization, to identify and retain the most informative EFMs to develop macroscopic models.

This study extends this latter work by proposing several significant improvements to the original algorithm, which now proceeds in four steps, in order to tackle problems such as the enumeration of the initial set of elementary flux modes, the differentiate consideration of positivity constraints on the fluxes, and the prediction error of reduced macroscopic reaction sets (reduced below the number of measured components). The algorithm is also tested with data of batch cultures of CHO-320 cells, on the basis of a larger metabolic network than the one considered in [30]. Finally, a discussion about the construction of dynamic macroscopic models including the prediction of the biomass is included.

In contrast with recent algorithms such as [21,27,31], it is difficult to claim that our algorithm would be presumably more efficient, and we certainly do not want to make this claim. As a matter of fact, benchmark comparisons are delicate to achieve in view of the limited access to software packages implementing the various approaches. Two comments can however be made:

- the approaches are different. For instance, the column generation algorithm aims at generating a set of basis vectors, whereas our approach aims at reducing a large initial set of EFMs in order to extract the most relevant vectors;
- our approach is modular, fast and simple as it combines several pre-existing tools, making the methodology and the code easy to understand and apply for non-expert users. The Matlab code of our algorithm is available at www.umons.ac.be/seco (at the time of publication of this article).

The paper is organized as follows. The next section introduces the main concepts relative to metabolic network analysis. In Section 3, the improved version of the reduction algorithm is presented. Information about experimental data and the considered metabolic network are described in Section 4 as well as different case studies according to the number of measured extracellular species. Section 5 proposes a simple simulator to provide a correct prediction of the data, following the method used in [22]. Finally, conclusions are drawn in Section 6.

2. Metabolic network analysis

In metabolic engineering, a network is a matrix representation of the intracellular reactions. In other words, this is a $m \times n$ stoichiometric matrix, denoted N, where m represents the number of internal metabolites and n stands for the number of reactions taking place inside the cell. Accounting for the quasi steady-state paradigm of metabolic flux analysis, the following homogeneous system of linear equations can be stated:

$$N\underline{v} = 0 \tag{1}$$

where $\underline{v} \in \mathbb{R}^n$ represents the vector of metabolic fluxes. The latter are often subject to positivity constraints expressing that the reactions have a net direction (resulting from the forward and reverse reactions):

$$\underline{v} \geq 0 \tag{2}$$

Henceforth, the set of possible flux distributions is the set of vectors $\underline{v}$ satisfying the linear equations in Eq. (1) under the positivity constraint in Eq. (2). The solution space is a pointed polyhedral cone, denoted $\mathscr{S}$, in the positive orthant. The edges of $\mathscr{S}$ are called elementary flux vectors and represent, from a biochemical point of view, the simplest metabolic pathways connecting extracellular substrates to final products.

To further constrain the solution space, an additional set of linear constraints might be added by considering experimental measurements of excretion and uptake rates $\underline{v}_m$:

$$\begin{pmatrix} N \\ N_e \end{pmatrix} \underline{v} = \begin{pmatrix} 0 \\ \underline{v}_m \end{pmatrix} \tag{3}$$

under the same positivity constraint in Eq. (2). In the previous relation, N_e is the stoichiometric matrix of measured extracellular species, a $m_e \times n$ matrix where m_e corresponds to the number of extracellular measurements. In this case, the set of solutions, denoted $\mathscr{F}$, is smaller than $\mathscr{S}$ ($\mathscr{F} \subset \mathscr{S}$) and only certain combinations of modes provide a solution.

3. EFM reduction algorithm

This section focuses on the reduction algorithm which starts from a usually large number of EFMs to end up into a small set of macroreactions which can be the basis for a macroscopic bioreaction model. The target is therefore a bioreaction scheme with less reactions than components, i.e., measured species. Biomass is included in the measured species so as to ensure that the reduced macroreaction model can be used for process simulation.

3.1. Generation of EFMs

The first step consists in the generation of the elementary flux modes on the basis of a metabolic network. As mentioned in Section 1, depending on the size of the metabolic network, two cases can be distinguished: on the one hand, when the enumeration of all possible pathways is achievable and on the other hand, the situation where only subsets of modes can be identified.

3.1.1. Enumeration

When the enumeration of all the modes is possible, a matrix of elementary flux modes, denoted E, can be obtained using software tools such as EFMtool. The matrix of EFMs is a $n \times n_{EFM}$ matrix where n_{EFM} represents the number of modes. From a geometrical point of view, elementary flux modes represent the edges of a polyhedral cone $\mathscr{S}$ constructed by the intersection of the hyperplanes defined by the rows of the stoichiometric matrix N. For the enumeration of all extreme rays of the polyhedral cone, softwares such as EFMtool make use of the double description method [13] and requires only the definition of the matrix N as input.

3.1.2. Generation of subsets

When the enumeration of EFMs becomes impossible or their number difficult to handle, the use of alternative methods for computing reduced sets, such as the ones reviewed in the introduction section, is required. In our algorithm we have selected the method proposed in [18], where a fast algorithm for computing randomly a decomposition of admissible flux vectors in a minimal number of elementary flux modes is proposed. Indeed, it can be shown that the decomposition of any vector v, inside the polytope $\mathscr{F}$, in the convex basis formed by elementary flux vectors is not unique. The objective of the fast algorithm is to determine one of the minimal decompositions, and to this end, it requires the knowledge of the stoichiometric matrix N, the matrix

of extracellular measurements N_e and the vector of excretion and uptake rates $\underline{\nu}_m$. If the algorithm is used only one time, a minimal set of EFMs (with a number of EFMs equal to the number of measurements) is directly obtained, which can be further reduced in the following steps of our algorithm to get a number of macroreactions lower than the number of measured species. However, it might be interesting to generate a larger pool of candidate EFMs before proceeding to the reduction, and in this case, the generation algorithm is executed a number of times until the desired pool is built.

3.2. Biological interpretation of the EFMs

Although the reduction algorithm is applied to the matrix of elementary flux modes E obtained in the previous step, the stoichiometric matrix representing the macroreactions is given by:

$$K = N_e E \tag{4}$$

which is a $m_e \times n_{EFM}$ matrix, whose columns provide the stoichiometry of each macroreaction. A biological interpretation of this matrix, e.g., detecting macroreactions with no reactants or no products, might be useful to reduce the matrix to a matrix K^*, corresponding to a reduced matrix of EFMs, denoted E^*.

3.3. Main reduction of EFMs

The next step consists in drastically reducing the number of modes, with a target number Ω, which is generally set to a number close (maybe slightly greater than) the number of extracellular measurements. This reduction is an iterative procedure based on the following optimization problem:

$$\Xi = \sum_{k=1}^{M} \left(K_e \, \underline{\phi}_k - \underline{\nu}_{m,k} \right) W^{-1} \left(K_e \, \underline{\phi}_k - \underline{\nu}_{m,k} \right)^T$$

$$\min_{\underline{\phi}_k} \Xi \qquad s.t. \qquad \underline{\phi}_k \geq 0 \tag{5}$$

where $\underline{\nu}_{m,k}$ represents the vector of uptake and excretion rates for every step-time k, $\underline{\phi}(t)$ corresponds to a time-varying decomposition of the flux $\underline{\nu}_m(t)$ into a reduced set of vectors stored in K_e, and W is a weighting diagonal matrix whose diagonal elements are $(\underline{\nu}_{m,k}^{max})^2$. Note that $K_e = N_e E_e$ where E_e denotes a reduced matrix of EFMs. The index Ξ should be equal to zero when the number of modes is greater or equal to the number of extracellular measurements ($n_{EFM} \geq m_e$), K_e is full row rank and there are no constraints on $\underline{\phi}(t)$. However, the positivity constraints can be active on some time periods depending on the compatibility of the measurements (which are always affected by experimental errors) with the network. The index Ξ therefore indicates how well the positivity constraints are satisfied, and the minimization of the index selects the flux values which are achieving the best constraint satisfaction. This is illustrated in the top part of Fig. 2 where the time evolution of one component of the index is plotted, for the measurement of isoleucine (one of the essential amino acids) in the experimental case study presented in Section 4. This graph depicts the situation where the number of elementary flux modes n_{EFM} is equal to or greater than the number of extracellular measurements m_e, and it is apparent that each time the index is non-zero, a positivity constraint is active.

The reduction of EFMs can be split into two steps : (i) the first is a pre-filtering of the set of modes relying on the concept of cosine-similarity introduced in [26] and (ii) the second step is based on the optimization problem in Eq. (5) and includes several checks where a user-defined tolerance value guides the selection.

3.3.1. Cosine-similarity algorithm

This algorithm consists in computing the collinearity between the elementary flux vectors and removing the collinear modes. In other words, if the cosine of the angle between two extreme rays is greater than a defined threshold $\chi \in [0,1]$, then the two elementary flux vectors point in the same direction and one of them can be removed. The reduction is a compromise between the value of the indicator Ξ and the number of EFMs. Finally, the remaining modes are stored in a reduced matrix E_e and the optimization problem in Eq. (5) can be formulated.

3.3.2. Optimization-based reduction

Although the previous step allows a significant cut in the number of modes, especially when starting from the complete set of elementary flux vectors, the cosine criterion alone does not enable a reduction to the target number Ω. For this purpose, an optimization-based reduction is proposed, where a randomly selected vector could be removed if the corresponding reduced elementary flux vectors collection in E_e satisfies the following condition:

$$| \, \Xi^* - \Xi \, | < tol \tag{6}$$

Otherwise, the selected EFM is kept. In Eq. (6), Ξ^* is the performance index for the candidate elimination whereas Ξ is the prior value of the performance index (prior to the random reduction), and *tol* is a tolerance defined by the user.

Since the reduction is based on a random selection of the modes, the reduced set with Ω EFMs is not unique and may not be the best over all combinations but achieves nevertheless an acceptable satisfaction of the positivity constraints.

3.4. Generation of the final macroreaction set

The final step of the algorithm consists in reducing the number of EFMs below the number of extracellular measurements from the previous set containing Ω modes. To achieve this objective, the choice of a target number Λ is necessary ($\Lambda < m_e$) and the reduction is still based on the value of the indicator Ξ (i.e. based on the optimization problem in Eq. (5)), which however has now another interpretation and can be seen as a sum of squared residuals measuring the distance to the experimental data. This is illustrated in the lower graph of Fig. 2. When the number of EFMs is smaller than m_e, the non-zero value of the indicator corresponds to the combined impact of the positivity constraints and of the fitting of the experimental measurements with a model which has now a reduced number of degrees of freedom. Indeed, with a set of EFMs smaller than the number of measured species, it is no longer possible to explain all the variance in the data, but a satisfactory model can be obtained in a classical least-squares sense. As mentioned in the previous section, the set of retained modes is not unique, leading to equivalent models. Indeed, as explained in [32], even when the number of EFMs is limited (either initially or after the reduction), the resulting set of bioreactions is usually redundant for the design of dynamic models that fully explain the available experimental data. There exists a set of equivalent bioreaction models because they all provide exactly the same values of observed uptake and production rates. Our reduction procedure aims at finding a minimal set of bioreactions, but it is important to keep in mind that this set is not unique, both regarding the number of bioreactions (a number of EFMs equal to the number of measured species allows an exact representation of the fluxes, and a smaller number allows an approximate representation) and regarding the equivalent selections of EFMs

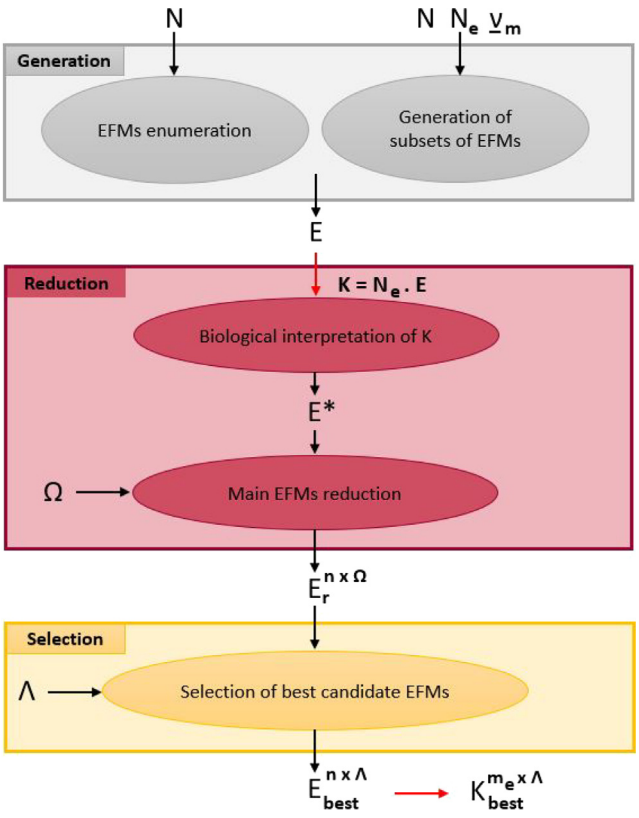

Fig. 1. The complete algorithm.

for a given number of them. In terms of input–output behavior, these equivalent models will therefore lead to equivalent control options. Some of the models might however be preferable in terms of biological interpretability and exploitation, as discussed in Section 3.2.

For the sake of clarity, it is worth noting that the optimization-based reduction in Section 3.3.2 and the selection of the final macroreaction set in Section 3.4 are different. While the first is a reduction based on several random successive eliminations of one mode so as to satisfy Eq. (6), the other selects the best combination of Λ EFMs among the reduced set of Ω modes according to the smallest value of the performance index. This last step does no longer rely on random eliminations of modes but on the selection of the best subset of EFMs. As a matter of fact, the final set obtained in this way could differ from directly choosing Ω equal to Λ in the previous step.

Furthermore, both steps are necessary as the enumeration of all EFMs combinations may become computationally prohibitive, their number being given by:

$$n_{comb} = \frac{\Omega!}{\Lambda!\,(\Omega - \Lambda)!} \tag{7}$$

Hence the importance of reducing first the number of modes to a small value Ω before going through the final step.

4. Application to Chinese Hamster Ovary CHO-320 cells in batch cultures

The reduction algorithm is now applied to experimental data from batch cultures of a Chinese Hamster Ovary cell line CHO-320. This section gives some information on the cell line, the life phases of these cells, introduces the metabolic network under consideration and addresses different case studies.

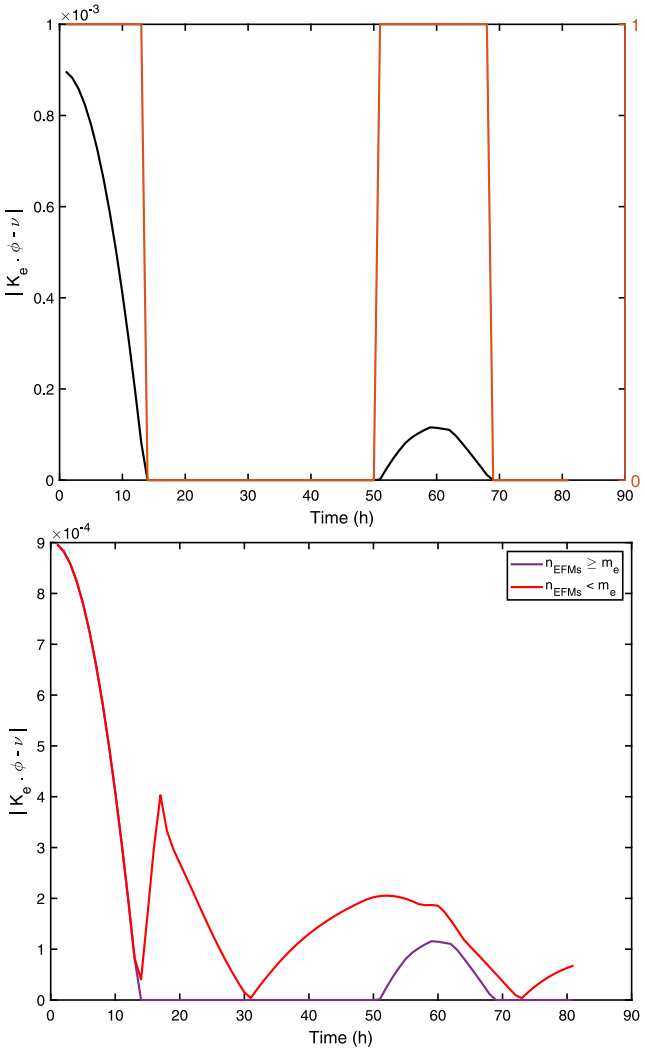

Fig. 2. Time evolution of the index $|K_e\phi - \nu|$ $(mM.h^{-1})$ for the measurement of isoleucine in the experimental case study—upper graph: positivity constraint violation (orange curve) when $n_{EFM} \geq m_e$ - lower graph: model discrepancy when $n_{EFM} < m_e$.

4.1. Cell line and metabolic network

The CHO-320 cell line is one of the first mammalian cell lines used in the production of therapeutically valuable proteins. The experiments exploited in this paper have been performed by a research team of UCLouvain, Belgium (courtesy of Prof. Yves-Jacques Schneider) and information relative to the cell line, the media, the bioreactor operation mode and the analysis methods can be found in [22].

As depicted in Fig. 1, the reduction algorithm starts from the definition of a metabolic network, which will depend on the phase of the cell during its life cycle. In Fig. 3, the growth phase, which corresponds to 80 h of culture, is used as data set. The metabolic network corresponding to the growth phase of the cells has been studied and developed in [23] and includes the pathways of glycolysis, the tricarboxylic acid cycle, the pentose phosphate pathways, the nucleotides synthesis pathways and several amino acids synthesis routes. The network contains 100 reactions and 72 internal metabolites. It is necessary to include a reaction of synthesis of the biomass in the metabolic network so as to consider its measurement in the EFM reduction procedure. According to [23], the cellular composition of the studied CHO

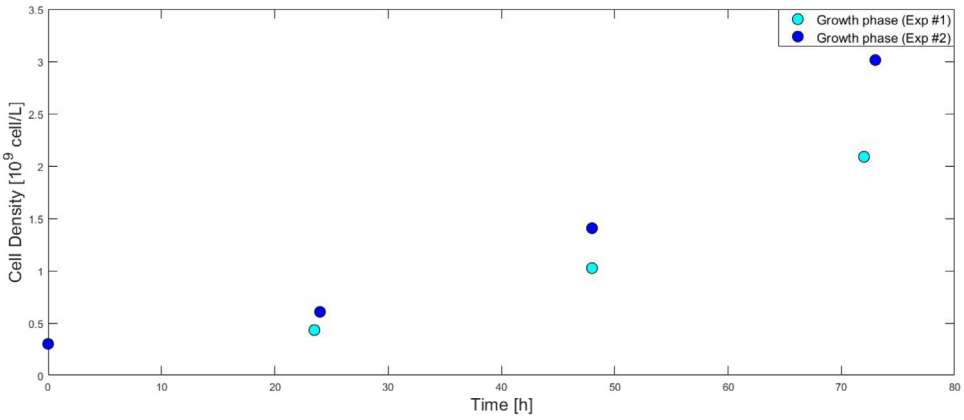

Fig. 3. Growth phase of CHO cells.

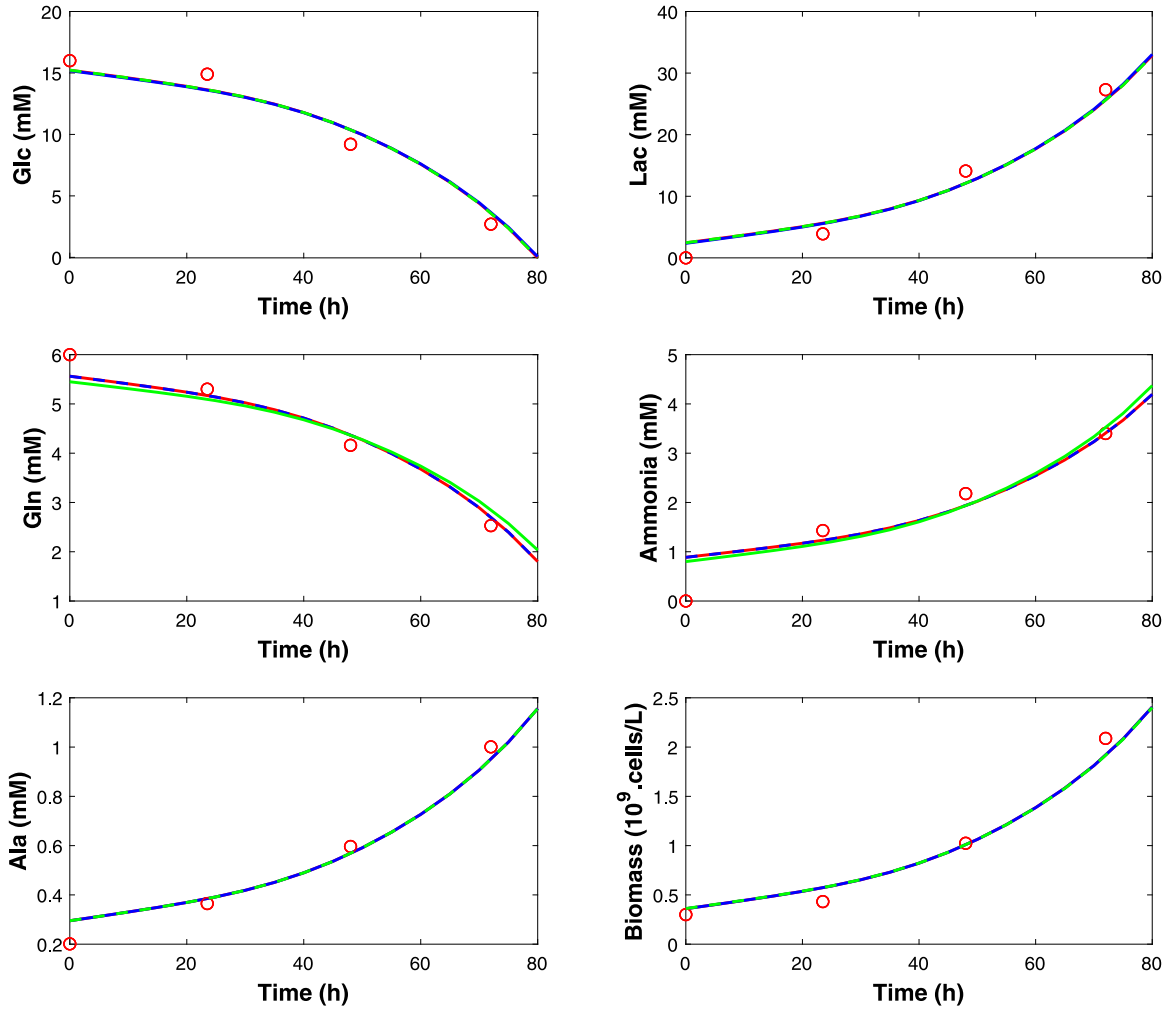

Fig. 4. Time evolution of the measured concentrations in dataset # 1 - results for $n_{EFM} = m_e$ (red curves), for $\Lambda_1 = 5$ (blue curves) and for $\Lambda_2 = 4$ (green curves).

cells is the following : 92.26% of proteins, 1.3% of RNA, 0.52% of DNA and 2.97% of lipids.

4.2. Case studies: Different measurement scenarios

Two measurement scenarios will be considered in the EFM reduction procedure, i.e., either 6 or 20 extracellular measurements. Before going more into technical details related to the two situations, a first insight consists in computing the number of elementary vectors on the basis of the metabolic network definition. This can be done independently of the measurement scenario using EFMtool, which provides 993203 EFMs. This very large number of EFMs is quite difficult to manipulate as such, and a first approach would be to partition this initial set into several smaller subsets and to apply the reduction procedure on the subsets. This approach has been tested and validated and gives good results in terms of reduction, but at the expense of a high computation time.

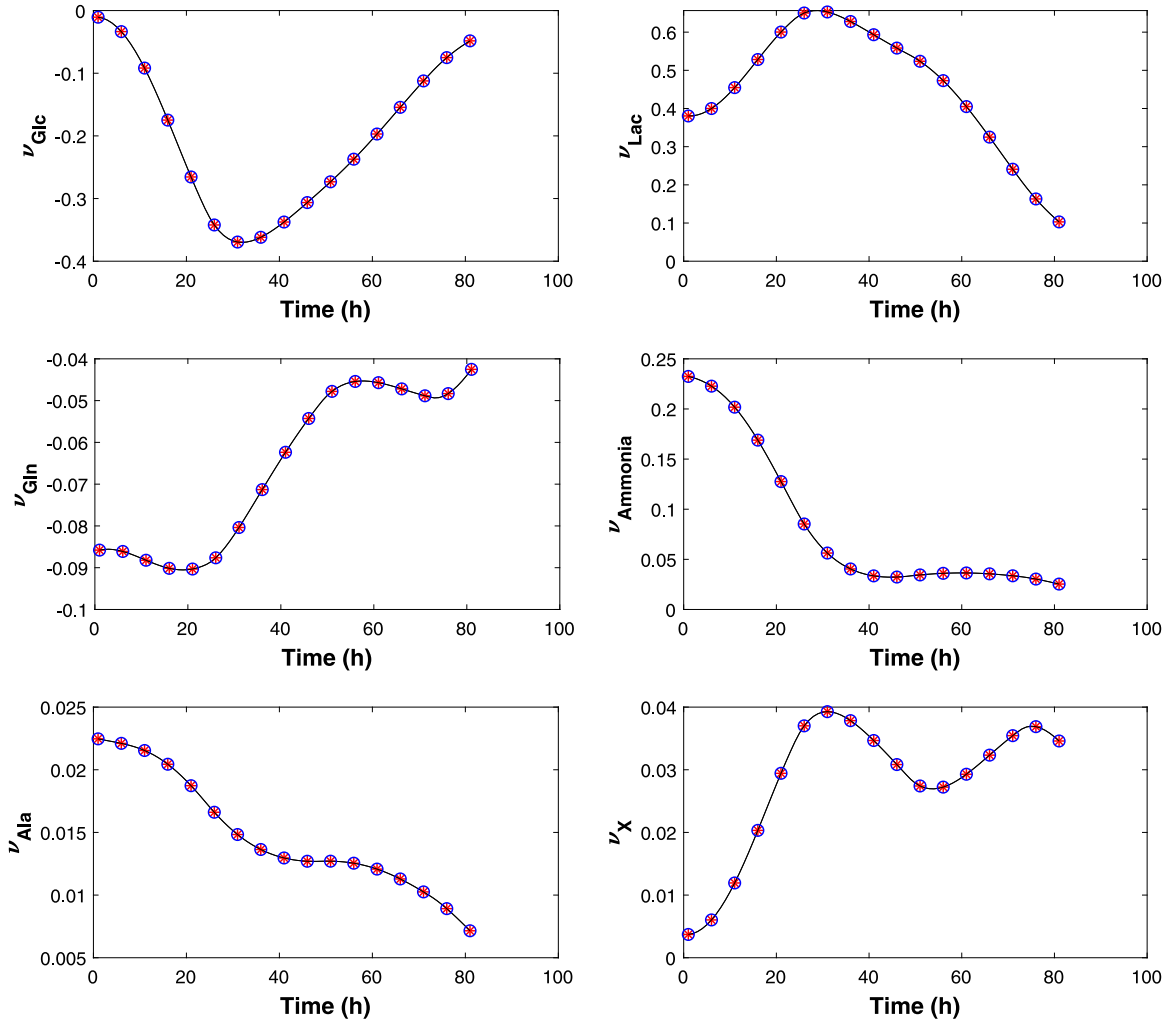

Fig. 5. Time evolution of the reaction rates (in $mM.h^{-1}$) in dataset # 1 - numerical results (black curves), $K_e\Phi$ for $n_{EFM} = 173$ (red crosses) and $K_e\Phi$ for $\Omega = 10$ (blue circles).

An alternative approach is to generate a smaller initial set of EFMs using a subset generation algorithm, and in our procedure we have selected the fast algorithm introduced in [18]. This algorithm can be executed an arbitrary number of times so as to generate the desired initial number of EFMs.

4.3. 6-Measurement case

In this case, only six extracellular measurements are considered, i.e., the concentration profiles of glucose, lactate, glutamine, ammonia, alanine and biomass.

The first step of the algorithm is the generation of an initial set of elementary flux vectors using the above-mentioned fast algorithm, which requires as inputs the stoichiometric matrix N, the matrix of extracellular measurements N_e, and the vector of specific rates $\underline{\nu_m}$. In first approximation, the specific uptake and excretion rates can be assumed constant during the exponential growth phase and are simply evaluated by linear interpolation (using the linear least-square method). On this basis, the fast generation algorithm allows obtaining $m_e = 6$ EFMs at once. Starting with this small initial set, the intermediate steps of the reduction algorithm can be bypassed, and the final step immediately applied to reduce the number of EFMs below m_e. Two target numbers are considered in what follows: (i) $\Lambda_1 = 5$ and (ii) $\Lambda_2 = 4$.

Table 1
Estimated value of the reaction rates (in $mM.h^{-1}$) for dataset #1.

ν_m	Linear regression	$\Lambda_1 = 5$	$\Lambda_2 = 4$
ν_{Glc}	−0.2072	-0.2054	-0.2075
ν_{Lac}	0.4132	0.4168	0.4127
ν_{Gln}	−0.0511	-0.0510	-0.0464
$\nu_{Ammonia}$	0.0449	0.0449	0.0485
ν_{Ala}	0.0117	0.0117	0.0117
ν_X	0.0278	0.0278	0.0277

A direct validation of the reduction can be achieved by comparing the constant reaction rates $\underline{\nu_m}$ obtained by regression with the product $K_e\Phi$ deduced from the optimization problem, K_e corresponding to the reduced matrix of modes. The rates being constant, this validation can be summarized in Table 1. This validation can be pursued by a comparison of the evolution of the concentrations by integration of the computed fluxes and identification of the most likely initial conditions of the measured species. This is illustrated in Fig. 4. The results are quite satisfactory, despite the simplifying assumption of constant rates, and even when the number of modes is reduced up to Λ = 4.

In a second evaluation, we now consider that the specific extracellular rates can vary along time, as it is usually the case during a batch or fed-batch culture. These rates can be evaluated by

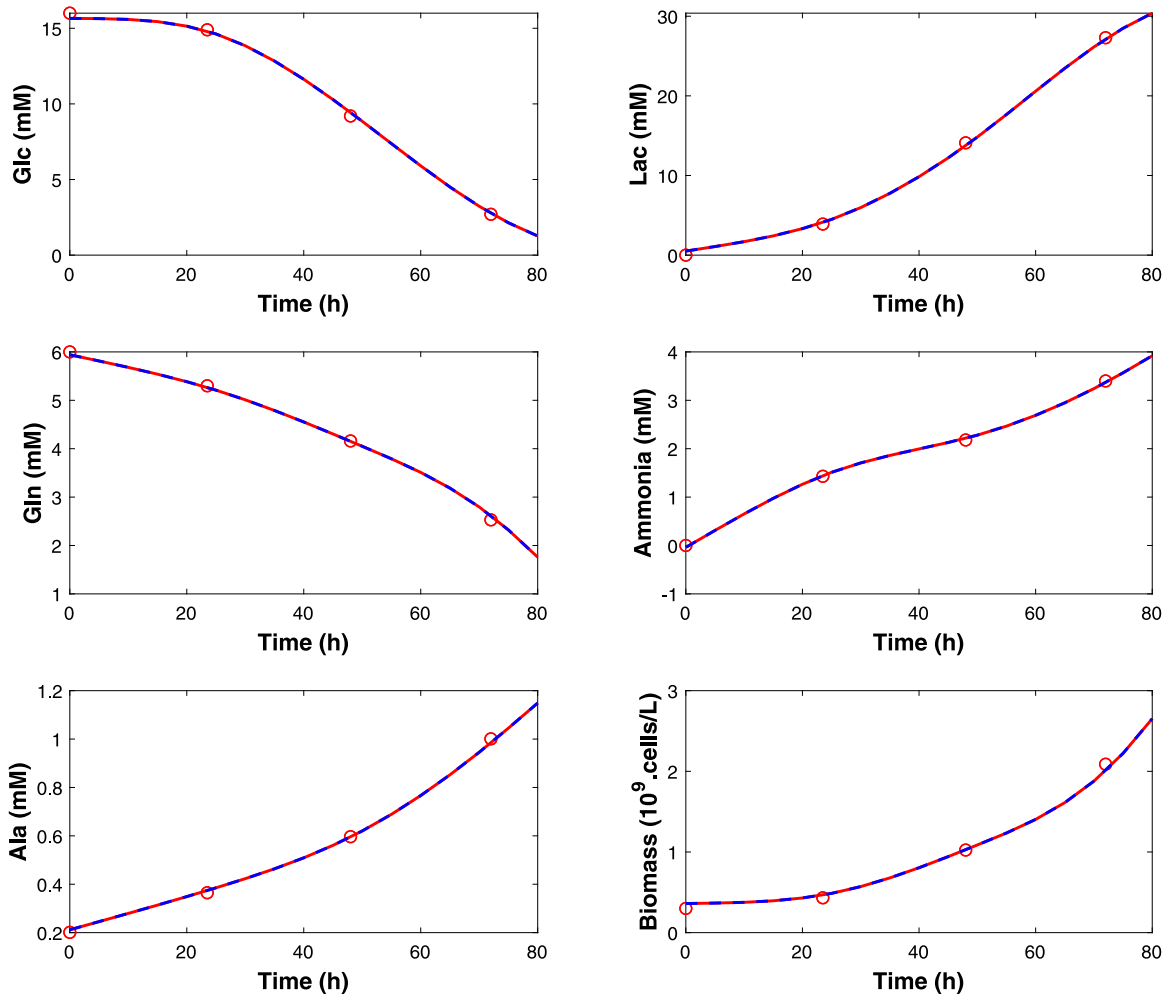

Fig. 6. Time evolution of the measured concentrations in dataset # 1 - results for $n_{EFM} = 173$ (red curves) and for $\Omega = 10$ (blue curves).

applying smoothing splines to the measured concentration profiles and numerical differentiation. Based on these time-varying rates, the execution of the fast generation algorithm produces an initial set of 206 elementary flux vectors. The complete reduction procedure can therefore now be applied. First, the test relative to the interpretability of the stoichiometric matrix K yields a reduced set of 173 modes. Second, the collinearity test could be used to reduce further this set, but as the number of EFMs is quite modest, it is preferable to use the optimization-based reduction. Usually the collinearity test is reserved to situations where the initial set is much larger and a preliminary, relatively drastic (but unsupervised) reduction is required, as discussed in [30]. The reduction is therefore continued using random elimination based on an optimization criterion ensuring the positivity constraints, with a target value $\Omega = 10$ (this latter number can of course be chosen differently, but we recommend a target slightly above m_e). Finally, the last step consists in finding the best combination of $\Lambda_1 = 5$ or $\Lambda_2 = 4$ modes among these 10 EFMs, minimizing a least-squares deviation from the measured data. Figs. 5 and 6 show the time evolution of the reaction rates and the concentrations, respectively, as predicted on the basis of 173 elementary flux modes or $\Omega = 10$ EFMs. Figs. 7 and 8 extend the illustration to the final sets of $\Lambda_1 = 5$ or $\Lambda_2 = 4$ modes. For $\Lambda_1 = 5$, the results are very satisfactory, whereas, for $\Lambda_2 = 4$, significant deviations appear in the reaction rates, which however have limited impact on the prediction of the time evolution of the concentrations. Mostly the prediction of ammonia is adversely affected, but the results remain quite acceptable in view of establishing simple dynamic models for bioprocess control purposes.

On the other hand, the reconstruction of the time evolution of the concentrations in Fig. 8 is better than in Fig. 4, which highlights the benefit of computing the reaction rates from smoothing splines and differentiation instead of the assumption of constant rates.

4.4. 20-Measurement case

It is now considered that most of the amino acids can be measured, leading to 20 extracellular measured species, including the biomass measurement. As outlined previously, the initial set of elementary flux modes is computed using the fast generation algorithm of [18], which yields 664 elementary flux modes (considering time-varying specific uptake and excretion rates obtained through smoothing splines and differentiation). The reduction procedure first analyses the biological interpretability of the macro-reactions described by K. In the present application, the modes leading to a macro-reaction where a product is generated without the consumption of any substrate are discarded. This step allows a slight reduction to 620 EFMs. Then, a collinearity test allowing a cut to 210 vectors is performed and an optimization based reduction can be achieved with a target value $\Omega = m_e = 20$. This value is fixed equal to the number of extracellular measured species to avoid computational issues in the calculation of all possible combinations during the last step of the procedure.

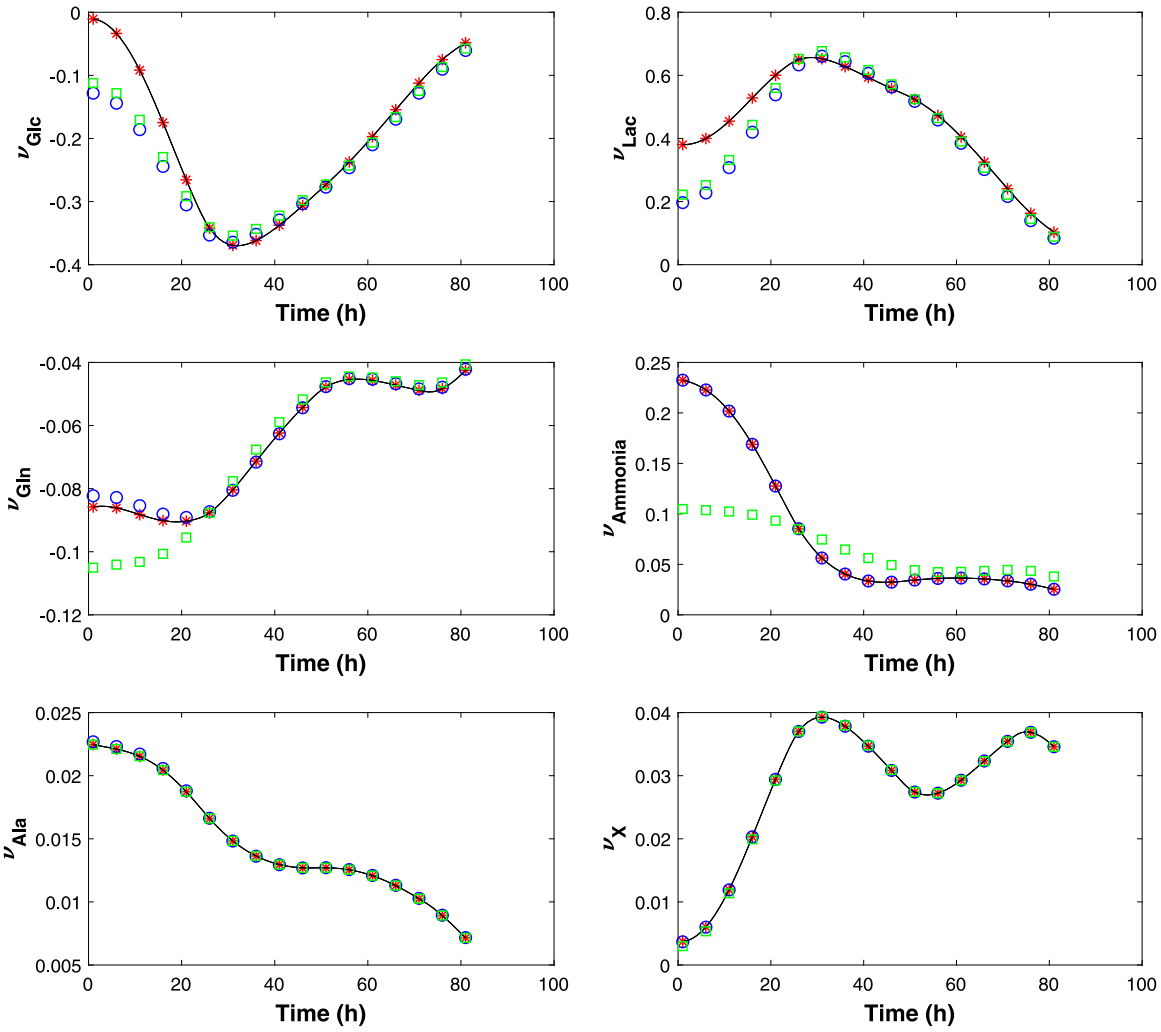

Fig. 7. Time evolution of the reaction rates (in $mM.h^{-1}$) in dataset # 1 - numerical results (black curves), $K_e\Phi$ for $\Omega = 10$ (red crosses), $K_e\Phi$ for $\Lambda_1 = 5$ (blue circles) and $K_e\Phi$ for $\Lambda_2 = 4$ (green squares).

Figs. 9 and 10 and Figs. 11 and 12 show the reconstruction of the time-evolution of the reaction rates and the concentrations, respectively.

Finally, the number of EFMs is reduced below the number of extracellular measurements m_e, e.g., (i) $\Lambda_1 = 19$ and (ii) $\Lambda_2 = 16$. The outcomes are given in Figs. 13 to 16. For most of the measured species, the reproduction of the experimental data remains quite acceptable. Small deviations can nonetheless be highlighted for ammonia, and larger deviations for the biomass, which are however acceptable.

5. Dynamical simulator

In order to assess the merits of the reduction, in addition to the previous validations, it might be interesting to examine the set of macro-reactions and to consider the following mass-balance dynamic model:

$$\frac{d\underline{\xi}}{dt} = K\Phi + D(\underline{\xi_{in}} - \underline{\xi}) \tag{8}$$

In this equation, $\underline{\xi}$ is the vector of concentrations of the extracellular species in the culture medium, K represents the stoichiometry of the reaction network built on the basis of the reduced set of EFMs, Φ is the vector of reaction rates computed in the optimization problem solved in the reduction procedure, D is the dilution rate and $\underline{\xi_{in}}$ denotes the inflow concentrations. In our case studies, the cultures operate in batch and the second term on the RHS vanishes ($D = 0$).

As an illustration, the case of 6 extracellular measured species, i.e., glucose, lactate, glutamine, ammonia, alanine and biomass, is considered, with the resulting stoichiometric matrix K:

$$K = \begin{bmatrix} -\alpha_1 & 0 & -\alpha_3 & 0 \\ 0 & 0 & \beta_3 & 0 \\ -\gamma_1 & -\gamma_2 & 0 & -\gamma_4 \\ 0 & \delta_2 & 0 & \delta_4 \\ 0 & 0 & 0 & \epsilon_4 \\ \sigma_1 & 0 & 0 & 0 \end{bmatrix} \tag{9}$$

where α_i, β_i, γ_i, δ_i, ϵ_i and σ_i are the stoichiometric coefficients. K has dimension $m_e \times \Lambda$ with $m_e = 6$ and $\Lambda = 4$. Equivalently, a macro-reaction scheme can be drawn:

$$\alpha_1\ Glc + \gamma_1\ Gln \rightarrow \sigma_1\ X \tag{10}$$

$$\gamma_2\ Gln \rightarrow \delta_2\ N \tag{11}$$

$$\alpha_3\ Glc \rightarrow \beta_3\ Lac \tag{12}$$

$$\gamma_4\ Gln \rightarrow \delta_4\ N + \epsilon_4\ Ala \tag{13}$$

This reaction scheme makes sense since glucose and glutamine are consumed in the growth phase, while lactate, ammonia, alanine and biomass are produced, as confirmed by the

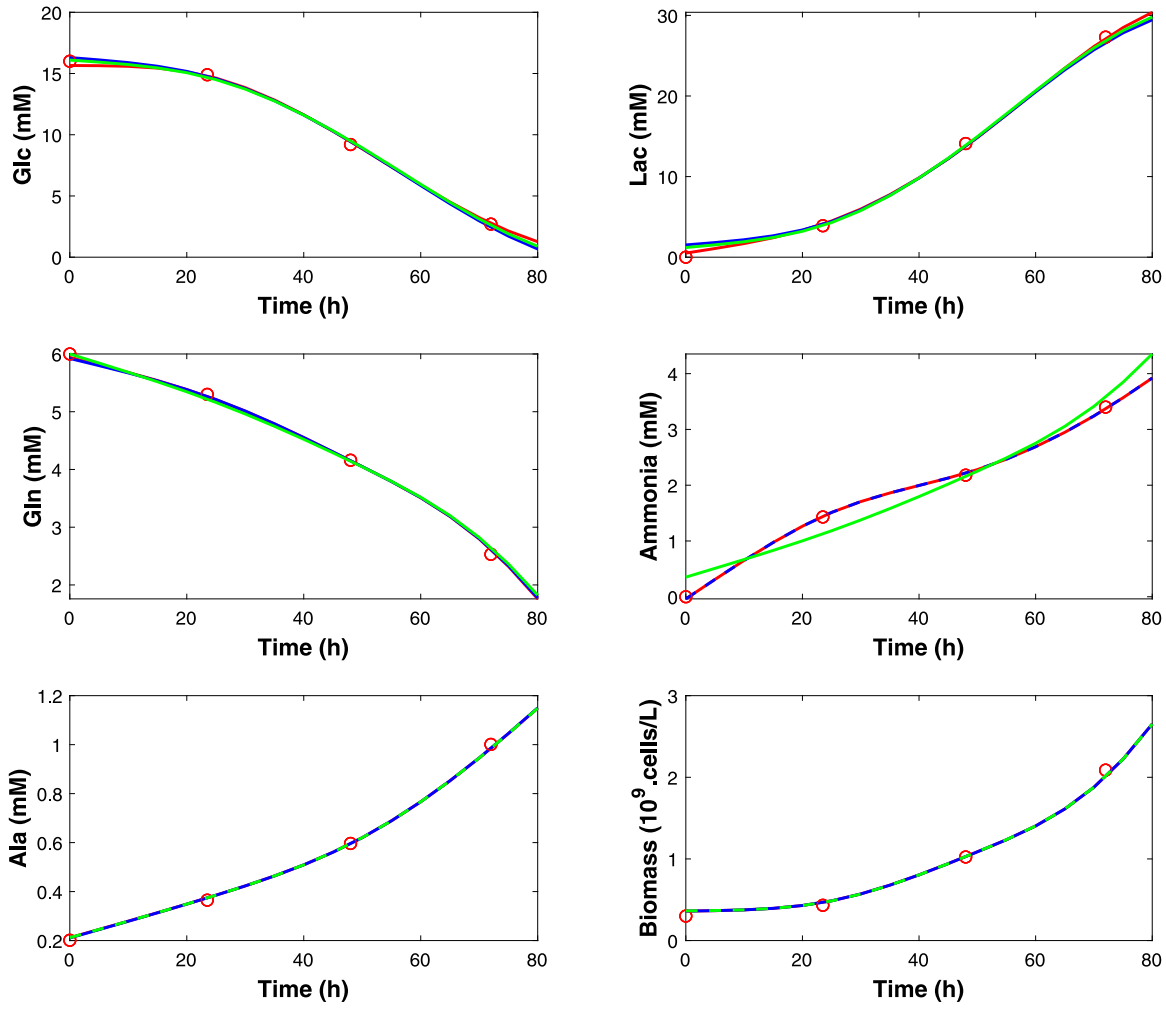

Fig. 8. Time evolution of the measured concentrations in dataset # 1 - results for $\Omega = 10$ (red curves), for $\Lambda_1 = 5$ (blue curves) and for $\Lambda_2 = 4$ (green curves).

time evolution of these species. The first reaction translates the growth of biomass by consumption of glucose and glutamine, the second reaction indicates that the consumption of glutamine leads to the production of ammonia, the third reaction shows that lactate is produced from glucose and the last reaction involves the consumption of glutamine to produce ammonia and alanine.

Without diving into the problem of the formulation of kinetic laws and the identification of their parameters, a basic dynamic simulator is proposed based on the assumption that the macro-reactions proceed at their maximal rate during the exponential growth phase. The reaction rates are modeled using Michaelis–Menten kinetics :

$$\phi_1 = \mu_{max_1} \frac{Glc}{Glc + \kappa} \frac{Gln}{Gln + \kappa} X \tag{14}$$

$$\phi_2 = \mu_{max_2} \frac{Gln}{Gln + \kappa} X \tag{15}$$

$$\phi_3 = \mu_{max_3} \frac{Glc}{Glc + \kappa} X \tag{16}$$

$$\phi_4 = \mu_{max_4} \frac{Gln}{Gln + \kappa} X \tag{17}$$

The parameters μ_{max_i} can be obtained by selecting the maximum entries in the vector $\Phi(t)$. On the other hand, the half-saturation constants κ (which would not be identifiable based on the limited data at disposal) are taken as small numbers to ensure the positivity of the model, but not too small so as to avoid excessive stiffness of the system of ordinary differential equations. The value of the parameters are given in Table 2.

Table 2
Parameters value.

	Units	Value
μ_{max_1}	h^{-1}	0.06
μ_{max_2}	h^{-1}	0.1
μ_{max_3}	h^{-1}	0.71
μ_{max_4}	h^{-1}	0.05
κ	mM	0.01

Simulation results are shown in Fig. 17, which are very satisfactory. Thus, we have demonstrated that the EFM reduction procedure opens the door to the fast development of dynamic macroscopic models. The estimation of kinetic laws requires larger quantities of informative data and could follow the procedure proposed in [33].

6. Conclusion

Starting from the preliminary work in [30], this study proposes several significant improvements to a procedure aimed at systematically reducing the number of elementary flux modes up to a number chosen below the number of measured extracellular species. Particularly, the situation of a detailed metabolic network possibly leading to an intractable initial set of EFMs, is tackled

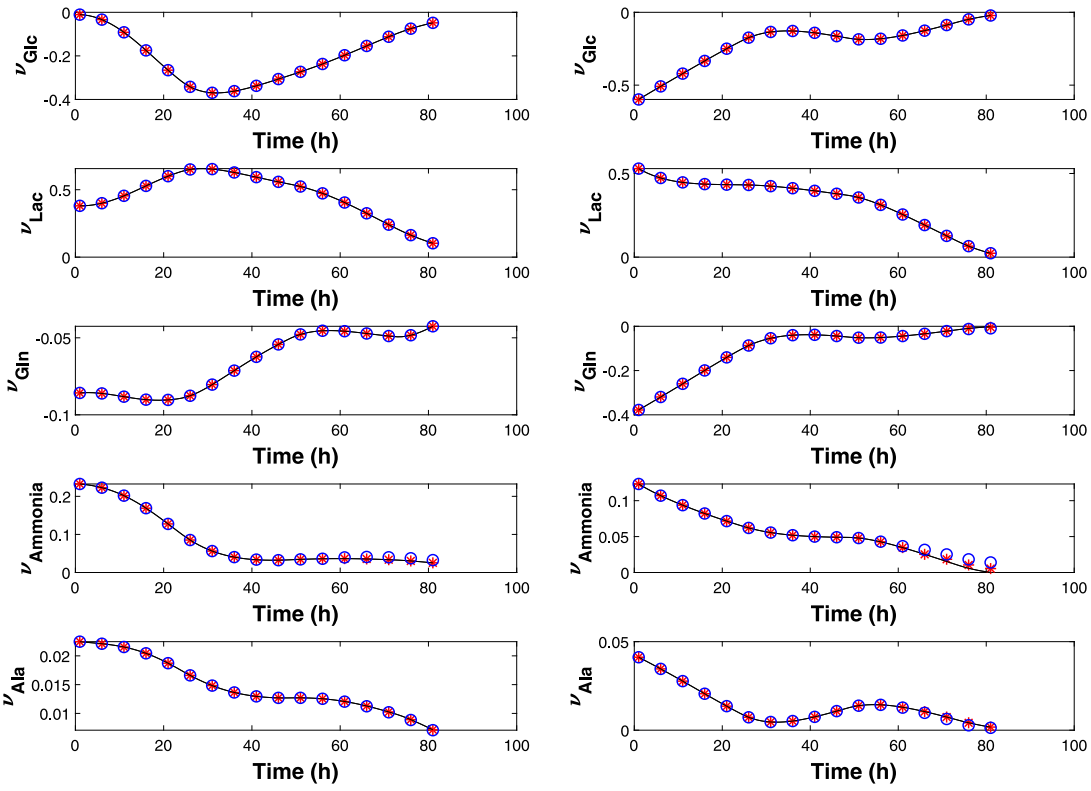

Fig. 9. Time evolution of the reaction rates (in $mM.h^{-1}$) in dataset # 1 (left side) and dataset # 2 (right side) - numerical results (black curves), $K_e\Phi$ for $n_{EFM} = 620$ (red crosses) and $K_e\Phi$ for $\Omega = m_e = 20$ (blue circles).

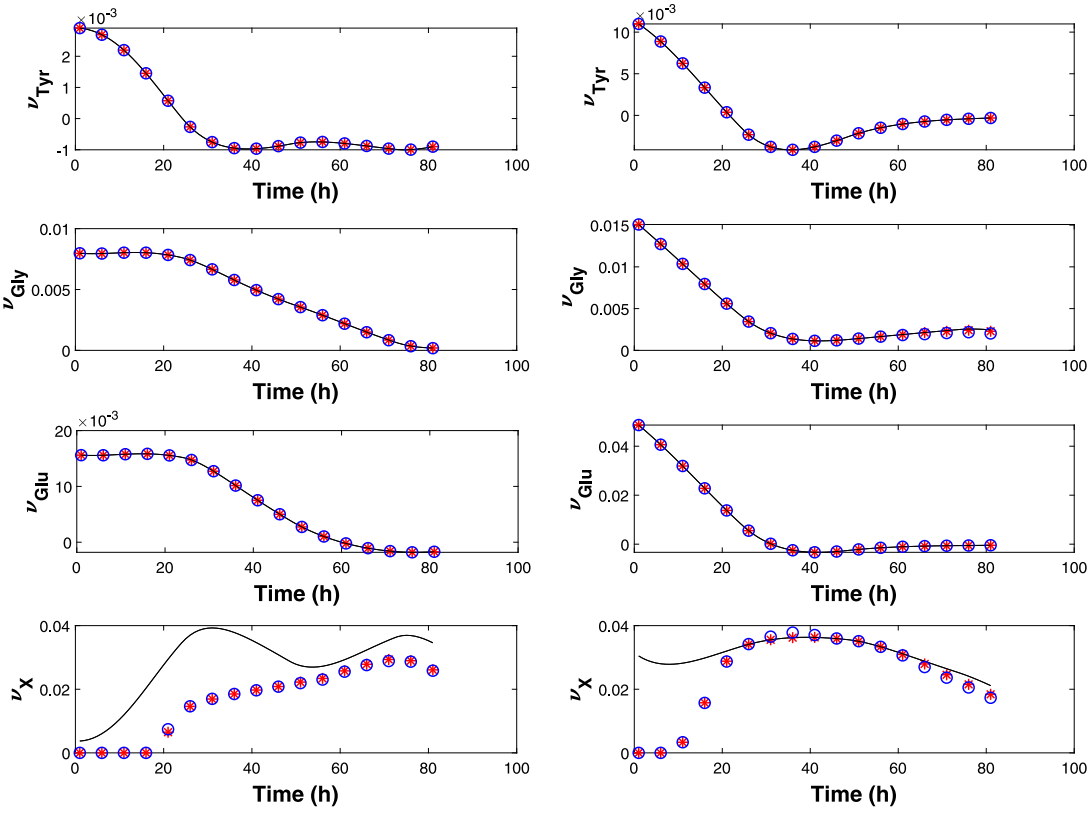

Fig. 10. Time evolution of the reaction rates (in $mM.h^{-1}$) in dataset # 1 (left side) and dataset # 2 (right side) - numerical results (black curves), $K_e\Phi$ for $n_{EFM} = 620$ (red crosses) and $K_e\Phi$ for $\Omega = m_e = 20$ (blue circles).

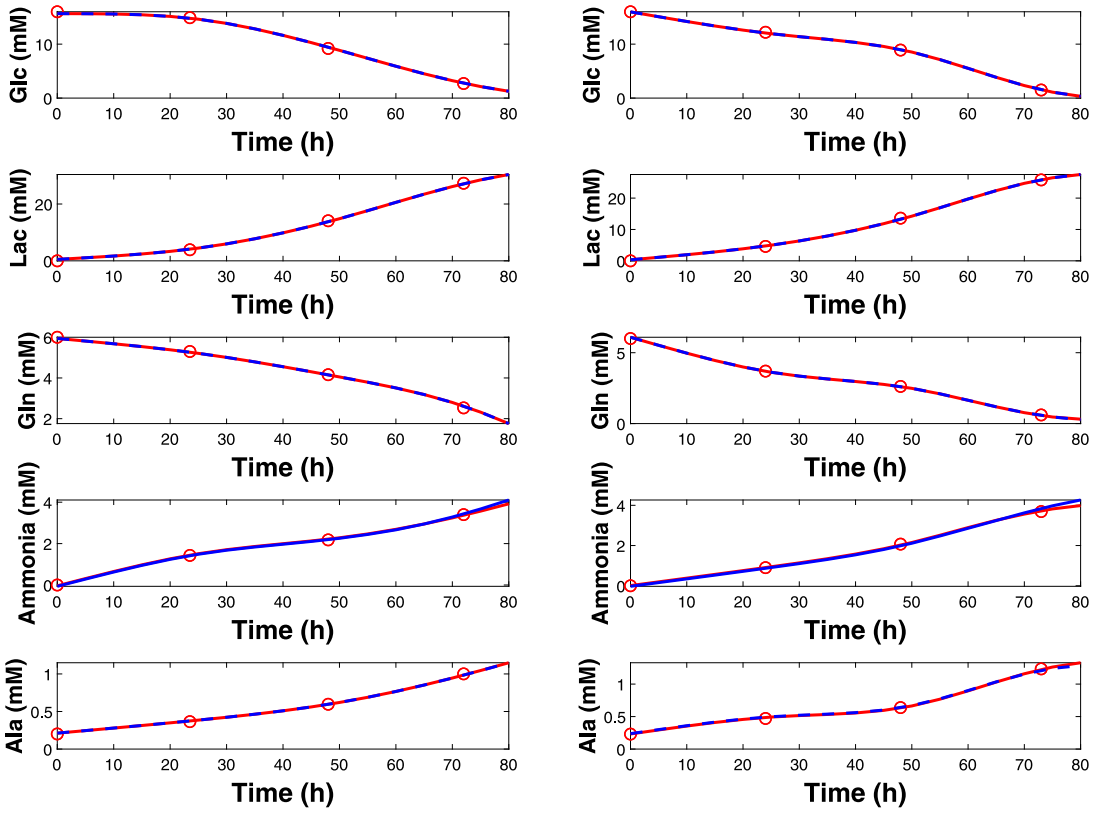

Fig. 11. Time evolution of the measured concentrations in dataset # 1 (left side) and dataset # 2 (right side) - results for $n_{EFM} = 620$ (red curves) and for $\Omega = m_e = 20$ (blue curves).

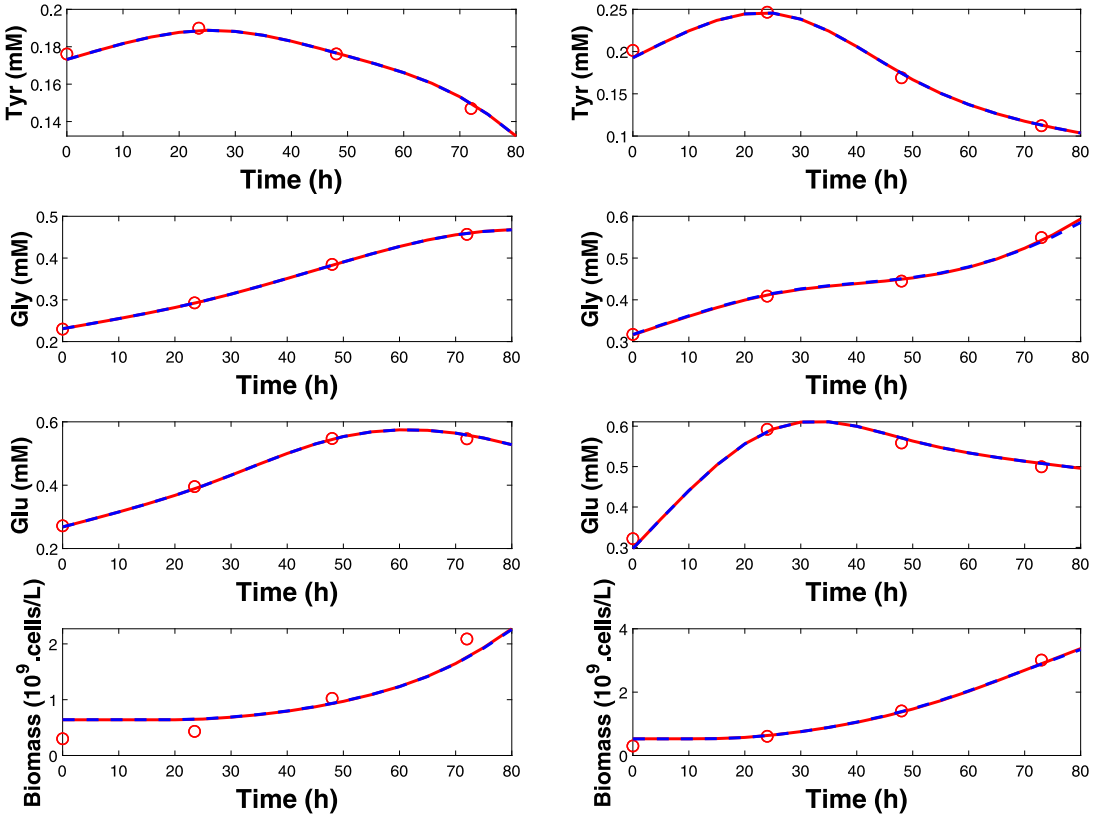

Fig. 12. Time evolution of the measured concentrations in dataset # 1 (left side) and dataset # 2 (right side) - results for $n_{EFM} = 620$ (red curves) and for $\Omega = m_e = 20$ (blue curves).

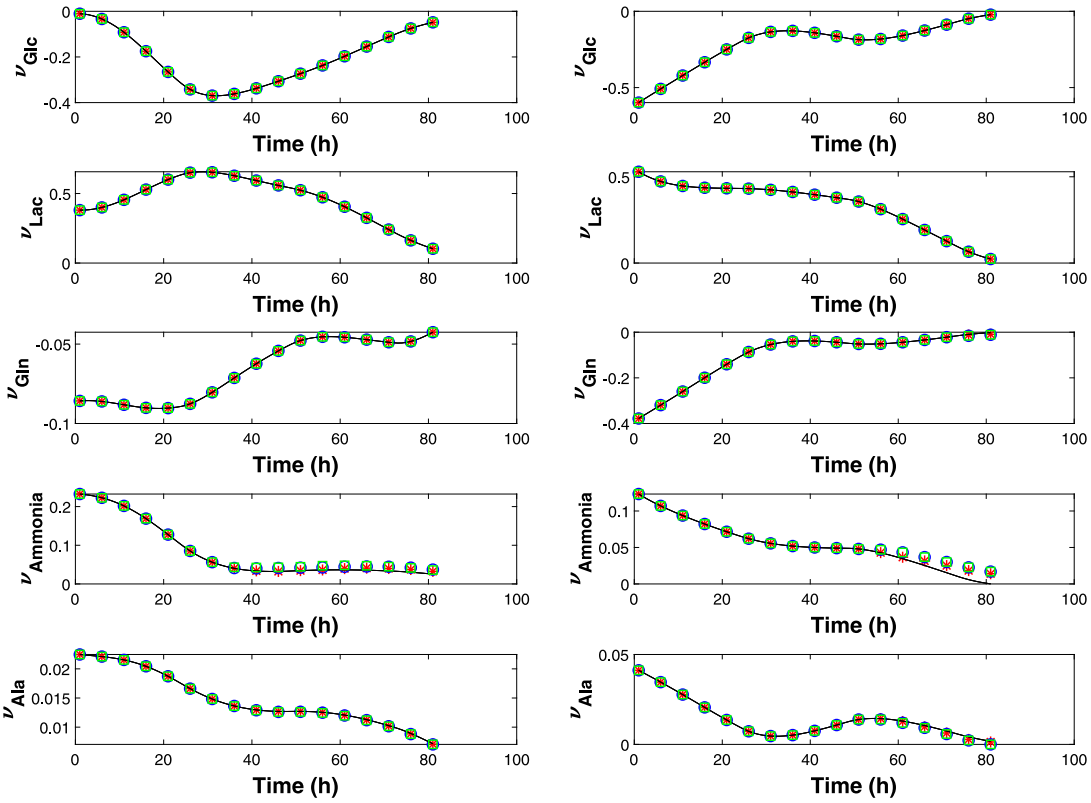

Fig. 13. Time evolution of the reaction rates (in $mM.h^{-1}$) in dataset # 1 (left side) and dataset # 2 (right side) - numerical results (black curves), $K_e\Phi$ for $\Omega = m_e = 20$ (red crosses), $K_e\Phi$ for $\Lambda_1 = 19$ (blue circles) and $K_e\Phi$ for $\Lambda_2 = 16$ (green squares).

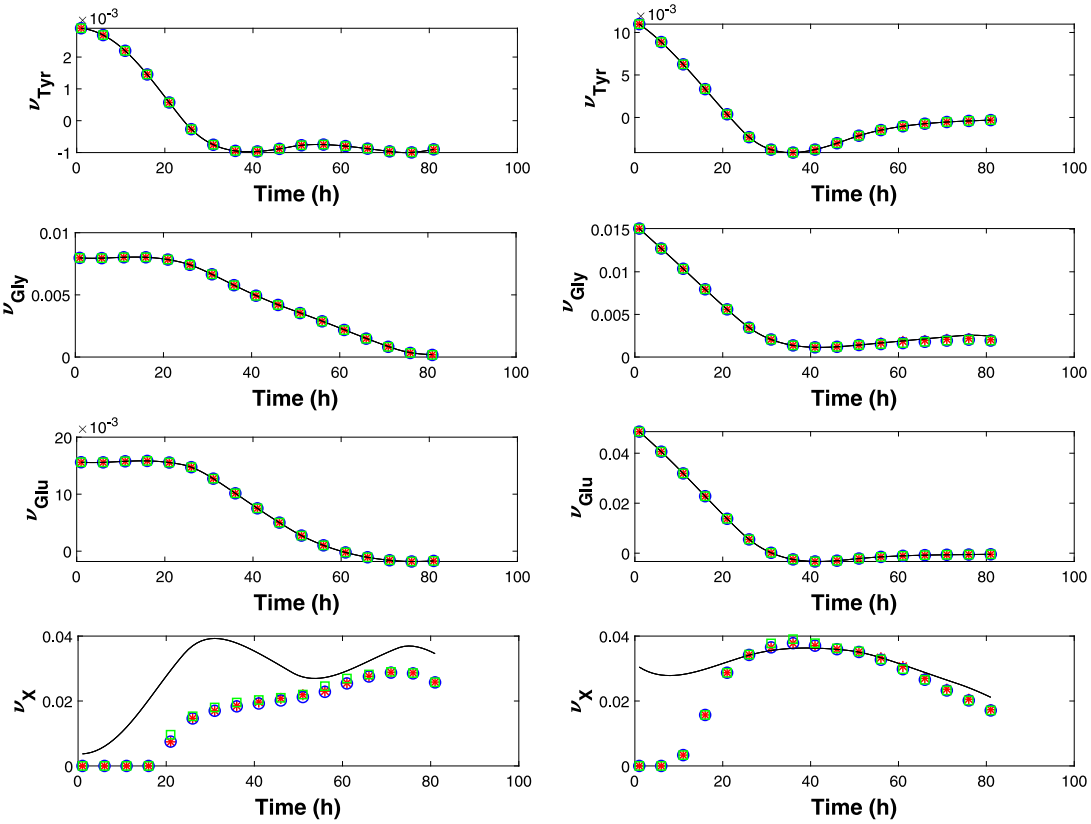

Fig. 14. Time evolution of the reaction rates (in $mM.h^{-1}$) in dataset # 1 (left side) and dataset # 2 (right side) - numerical results (black curves), $K_e\Phi$ for $\Omega = m_e = 20$ (red crosses), $K_e\Phi$ for $\Lambda_1 = 19$ (blue circles) and $K_e\Phi$ for $\Lambda_2 = 16$ (green squares).

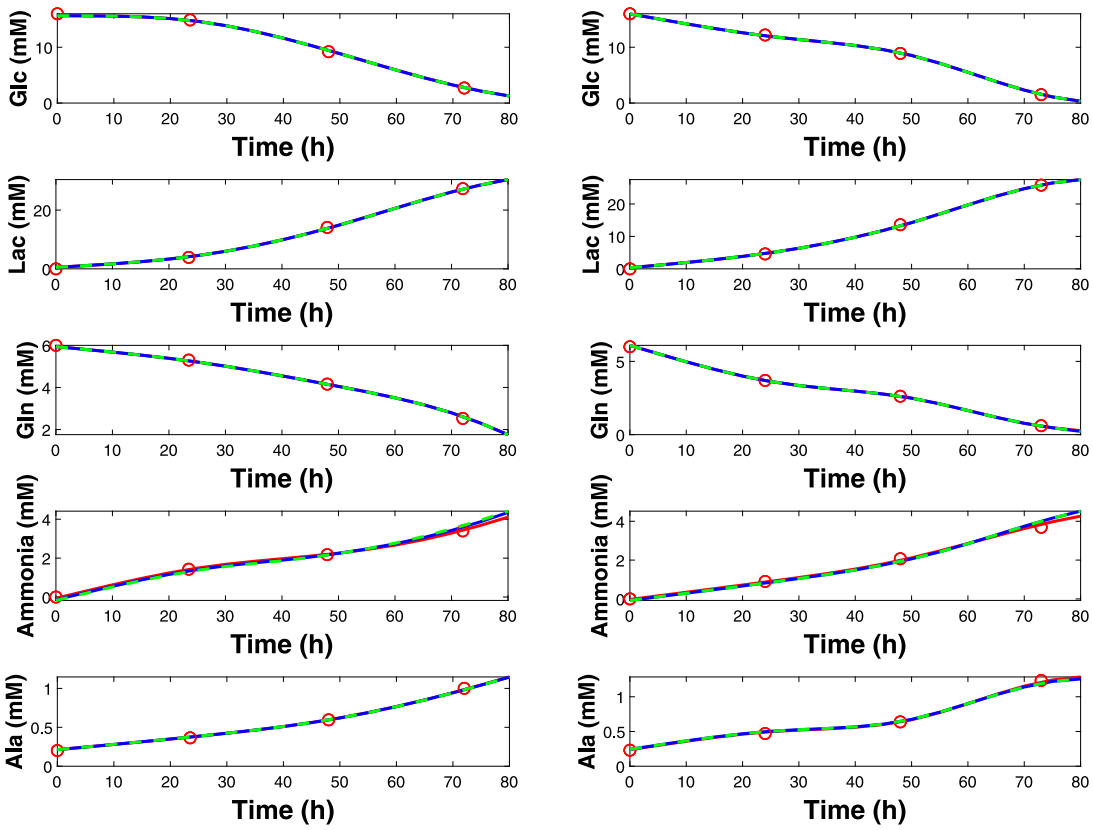

Fig. 15. Time evolution of the measured concentrations in dataset # 1 (left side) and dataset # 2 (right side) - results for $\Omega = m_e = 20$ (red curves), for $\Lambda_1 = 19$ (blue curves) and for $\Lambda_2 = 16$ (green curves).

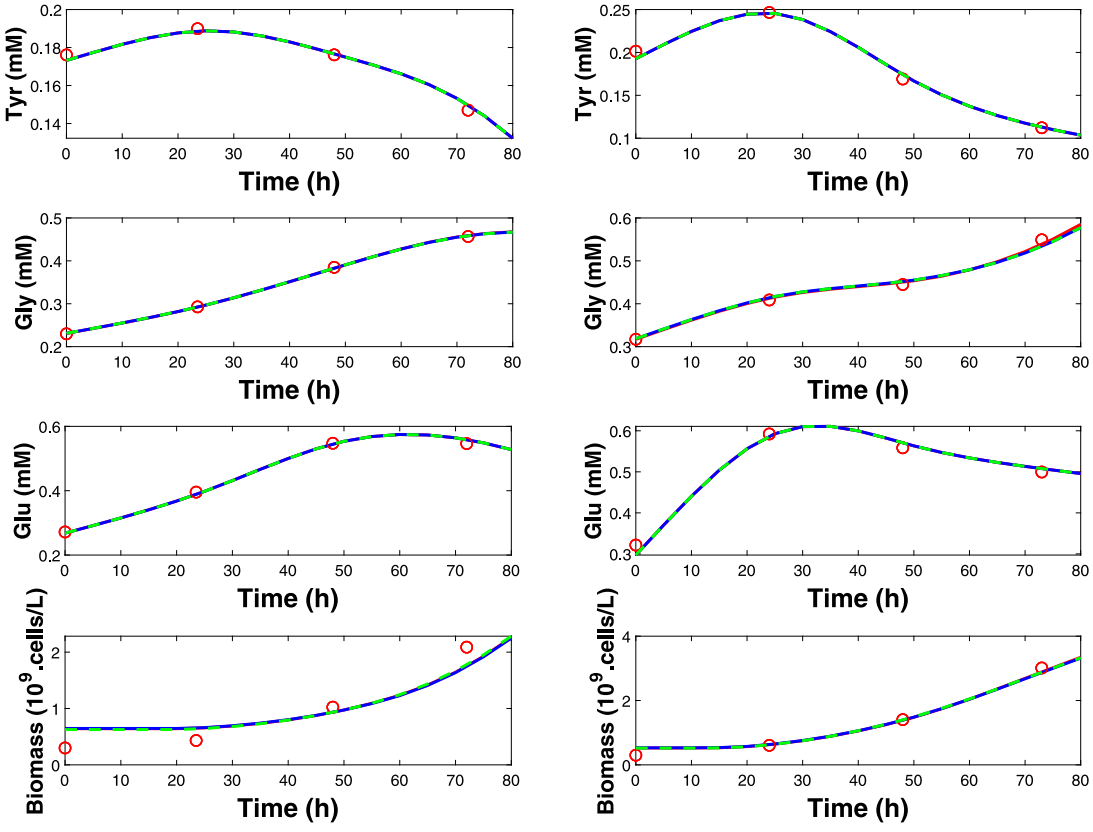

Fig. 16. Time evolution of the measured concentrations in dataset # 1 (left side) and dataset # 2 (right side) - results for $\Omega = m_e = 20$ (red curves), for $\Lambda_1 = 19$ (blue curves) and for $\Lambda_2 = 16$ (green curves).

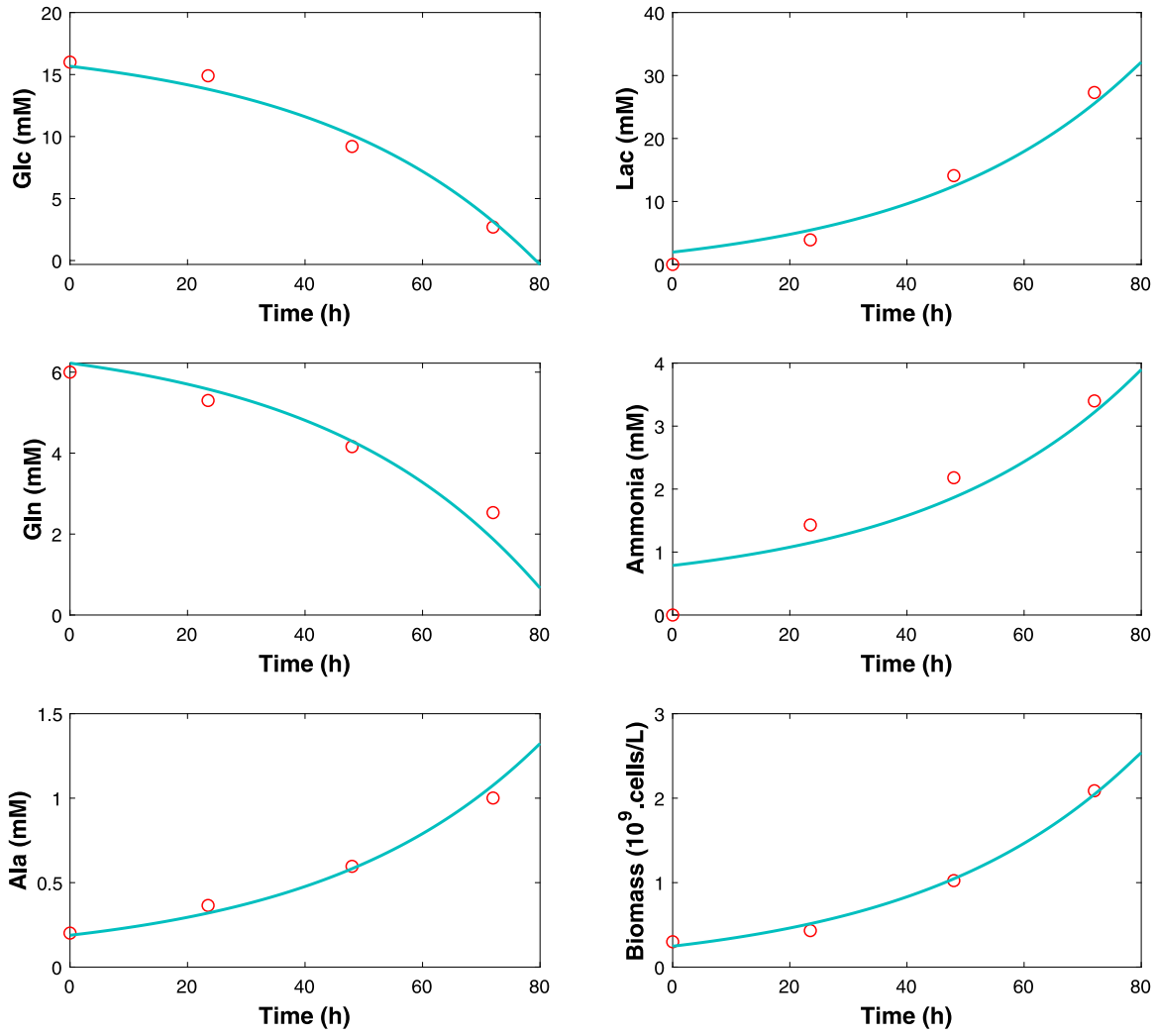

Fig. 17. Time evolution of the measured concentrations in dataset # 1 - results from the dynamical simulator (cyan curves).

by the use of an algorithm generating a minimal decomposition of a feasible solution. This algorithm can be applied once, for a particular value of the extracellular fluxes, or repeatedly following the time evolution of the extracellular fluxes, yielding an initial set of EFMs of modest size but with sufficient variety to represent the metabolic activity. Subsequently, the reduction procedure is made of several steps, which can all be activated or on the contrary bypassed: (a) the elimination of the modes leading to a stoichiometric matrix with macro-reactions lacking a biological interpretation, (b) the elimination of modes which are close to collinearity based on a cosine-criterion and (c) an optimization-based reduction, first targeting the positivity constraints and then the satisfaction of a least-squares deviation from the experimental data. The procedure is further tested by considering data collected in batch cultures of CHO cells, and two different measurement configurations, showing very satisfactory results. The generation of macroscopic bioreaction schemes is of great importance for the development of simple dynamic models of bioprocesses. Further research entails the development of systematic procedure for the identification of kinetic laws based on the fluxes estimated in our reduction procedure and further validation tests with different metabolic networks and data sets.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] F. Sidoli, A. Mantalaris, S. Asprey, Modelling of mammalian cells and cell culture processes, Cytotechnology 44 (1-2) (2004) 27–46.
[2] D. Vester, E. Rapp, S. Kluge, Y. Genzel, U. Reichl, Virus-host cell interactions in vaccine production cell lines infected with different human influenza a virus variants : A proteomic approach, J. Proteomics 73 (9) (2010) 1656–1669.
[3] J. Niklas, E. Heinzle, Metabolic flux analysis in systems biology of mammalian cells, Genom. Syst. Biol. Mammalian Cell Culture, Springer (2012) 109–132.
[4] H. Niu, Z. Amribt, P. Fickers, W. Tan, P. Bogaerts, Metabolic pathway analysis and reduction for mammalian cell cultures - towards macroscopic modeling, Chem. Eng. Sci. 102 (2013) 461–473.
[5] B. Hodgson, C. Taylor, J. Ushio, J. Leigh, Intelligent modelling of bioprocesses : A comparison of structured and unstructured approaches, Bioprocess Biosyst. Eng. 26 (6) (2005) 353–359.
[6] J. Haag, A. Vande Wouwer, P. Bogaerts, Systematic procedure for the reduction of complex biological reaction pathways and the generation of macroscopic equivalents, Chem. Eng. Sci. 60 (2) (2005) 459–465.

[7] J. Haag, A. Vande Wouwer, P. Bogaerts, Dynamic modeling of complex biological systems : a link between metabolic and macroscopic description, Math. Biosci. 193 (1) (2005) 25–49.
[8] C. Baroukh, O. Bernard, Metabolic modeling of c. sorokiniana diauxic heterotrophic growth, IFAC-PapersOnLine 49 (26) (2016) 330–335.
[9] S. Schuster, C. Hilgetag, On elementary flux modes in biochemical reaction systems at steady state, J. Biol. Systems 2 (2) (1994) 165–182.
[10] J. Stelling, S. Klamt, K. Bettenbrock, S. Schuster, E.D. Gilles, Metabolic network structure determines key aspects of functionality and regulation, Nature 420 (6912) (2002) 190–193.
[11] J. Gao, V. Gorenflo, J. Scharer, H. Budman, Dynamic metabolic modeling for a mab bioprocess, Biotechnol. Prog. 23 (1) (2007) 168–181.
[12] A. von Kamp, S. Schuster, Metatool 5.0 : fast and flexible elementary modes analysis, Bioinformatics 22 (2006) 1930–1931.
[13] M. Terzer, J. Stelling, Large-scale computation of elementary flux modes with bit pattern trees, Bioinformatics 24 (19) (2008) 2229–2235.
[14] S. Naderi, M. Meshram, C. Wei, B. McConkey, B. Ingalls, H. Budman, J. Scharer, Metabolic flux and nutrient uptake modeling of normal and apoptotic CHO cells, IFAC Proc. Vol. 43 (6) (2010) 395–400.
[15] S. Klamt, J. Stelling, Combinatorial complexity of pathway analysis in metabolic networks, Mol. Biol. Rep. 29 (2002) 233–236.
[16] L. Figueiredo, A. Podhorski, A. Rubio, C. Kaleta, J. Beasley, S. Schuster, F. Planes, Computing the shortest elementary flux modes in genome-scale metabolic networks, Bioinformatics 25 (2009) 3158–3165.
[17] C. Kaleta, L. de Figueiredo, S. Schuster, Can the whole be less than the sum of its parts ? Pathway analysis in genome-scale metabolic networks using elementary flux patterns, Genome Res. (2009) 1872–1883.
[18] R. Jungers, F. Zamorano, V. Blondel, A. Vande Wouwer, G. Bastin, Fast computation of minimal elementary decompositions of metabolic vectors, Automatica 47 (6) (2011) 1255–1259.
[19] D. Machado, Z. Soons, K.R. Patil, E.C. Ferreira, I. Rocha, Random sampling of elementary flux modes in large-scale metabolic networks, Bioinformatics 28 (18) (2012) i515–i521.
[20] S. Tabe-Bordbar, S. Marashi, Finding elementary flux modes in metabolic networks based on flux balance analysis and flux coupling analysis : application to the analysis of escherichia coli metabolism, Biotechnology Letters 35 (2013) 2039–2044.
[21] H. Oddsdottir, E. Hagrot, V. Chotteau, A. Forsgren, On dynamically generating relevant elementary flux modes in a metabolic network using optimization, J. Math. Biol. 71 (4) (2015) 903–920.
[22] A. Provost, G. Bastin, Dynamic metabolic modelling under the balanced growth condition, J. Process Control 14 (7) (2004) 717–728.
[23] F. Zamorano, A.V. Vande Wouwer, R.M. Jungers, G. Bastin, Dynamic metabolic models of CHO cell cultures through minimal sets of elementary flux modes, J. Biotechnol. 164 (3) (2013) 409–422.
[24] Z. Soons, E. Rocha, I. Ferreira, Selection of elementary modes for bioprocess control, Comput. Appl. Biotechnol. 11 (2010) 156–161.
[25] Z. Soons, I. Ferreira, E. Rocha, Identification of minimal metabolic pathway models consistent with phenotypic data, J. Process Control 21 (2011) 1483–1492.
[26] L. Hebing, T. Neymann, T. Thüte, A. Jockwer, S. Engell, Efficient generation of models of fed-batch fermentations for process design and control, IFAC-PapersOnLine 49 (7) (2016) 621–626.
[27] H. Oddsdottir, E. Hagrot, V. Chotteau, A. Forsgren, Robustness analysis of elementary flux modes generated by column generation, Math. Biosci. 273 (2016) 45–56.
[28] E. Hagrot, H. Oddsdottir, J. Hosta, E. Jacobsen, V. Chotteau, Poly-pathway model, a novel approach to simulate multiple metabolic states by reaction network-based model - application to amino acid depletion in CHO cell culture, J. Biotechnol. 259 (2017) 235–247.
[29] T. Abbate, S. Fernandes de Sousa, L. Dewasme, G. Bastin, A. Vande Wouwer, Inference of dynamical macroscopic models of cell metabolism based on elementary flux modes analysis, Biochem. Eng. J. 151 (2019) 1–11.
[30] M. Maton, A. Vande Wouwer, P. Bogaerts, Selection of a minimal suboptimal set of EFMs for dynamic metabolic modelling, IFAC-PapersOnLine 54–3 (2021) 667–672.
[31] E. Hagrot, H. Oddsdottir, M. Makinen, A. Forsgren, V. Chotteau, Novel column generation-based optimization approach for poly-pathway kinetic model applied to CHO cell culture, Metab. Eng. Commun. 8 (2019) e00083.
[32] G. Bastin, Metabolic flux analysis and metabolic design of bioreaction systems, Tutorial (2019).
[33] A. Richelle, P. Bogaerts, Systematic methodology for bioprocess model identification based on generalized kinetic functions, Biochem. Eng. J. 100 (2015) 41–49.