

Determining a unique solution to underdetermined metabolic networks via a systematic path through the Most Accurate Fluxes

Khadija Mhallem Gziri*, Philippe Bogaerts*

*3BIO-BioControl, Université Libre de Bruxelles
Av. F.-D. Roosevelt 50 C.P. 165/61, B-1050 Brussels, Belgium
(e-mail: kmhallem@ulb.ac.be; philippe.bogaerts@ulb.ac.be)

Abstract: Metabolic Flux Analysis (MFA) and Flux Balance Analysis (FBA) often lead to underdetermined problems in the sense that there are more unknown fluxes in the metabolic network than the number of available equations which represent balanced metabolites and measured fluxes. Even the additional inequality constraints, e.g. flux positivity, and/or the use of an objective function in FBA do not allow obtaining a unique solution in many cases. This contribution aims at determining a simple, systematic and computationally efficient algorithm for obtaining a unique solution based on the iterative determination of the Most Accurate Fluxes (MAF). A measure of accuracy is introduced and the systematic algorithm is proposed and illustrated on a case study aiming at the determination of unique dynamic flux values within hybridoma cells. It is also shown that the MAF distribution is similar to the mean values obtained from a uniform sampling of admissible solutions.

© 2019, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved.

Keywords: Metabolic network, Metabolic Flux Analysis, Flux Balance Analysis, Flux Variability Analysis, underdetermined systems, constrained-based modeling, mammalian cells, Most Accurate Fluxes

1. INTRODUCTION

There is an ever-increasing use of metabolic networks for determining the values of intracellular reaction rates (or fluxes) under specific culture conditions. These latter are often corresponding to measured output and/or input fluxes. Metabolic Flux Analysis (MFA) (Stephanopoulos et al., 1998) consists in determining intracellular flux values based on the assumption that intracellular metabolites do not accumulate and are therefore at quasi-steady state. The mass balances of these internal metabolites reduce then to a simple system of algebraic homogenous equations. Additional equality constraints are given by the available measurements, at a given time instant, of the output and/or input fluxes and inequality constraints are given by minimum/maximum values for some fluxes and/or flux positivity if one assumes direct reactions being predominant in comparison with the reverse ones. This method corresponds to constrained-based modeling. Dynamic Metabolic Flux Analysis (DMFA) consists in solving the former problem at each time instant for which external flux measurements are available (Antoniewicz, 2013). One of the key issues is that this system of equations to be solved is most of the time underdetermined, meaning that the number of equations (mass balances of internal metabolites + measurements of extracellular fluxes) is less than the number of unknowns (the intracellular fluxes). Admissible flux intervals can be obtained by solving two linear programs for each intracellular flux in order to determine its lower and upper bounds. This method is called Flux Variability Analysis (FVA) (Mahadevan and Schilling, 2003) or Flux Spectrum Approach (FSA) (Llaneras and Picó,

2007). Other methods are based on convex analysis and provide the admissible solutions as combinations of Elementary Flux Modes or Extreme Pathways (Klamt and Stelling, 2003), even in a dynamical framework (Fernandes de Sousa et al., 2016).

One of the most famous methods to try to overcome the underdeterminacy is Flux Balance Analysis (FBA) (Orth et al., 2010) which consists in determining the flux distribution which minimizes an objective cost function under the above-mentioned equality and inequality constraints. This cost function, made of a linear combination of some metabolic fluxes, is often chosen for representing cell growth rate maximization. Despite the use of a cost function to be minimized, the problem often remains underdetermined with several admissible flux distributions leading to the same minimum cost. Several methods have been proposed to find and analyze multiple optimal solutions (Motamedian and Naeimpoor, 2018).

In case of underdetermined systems, it can be useful not only to determine the admissible flux intervals but to compute a specific solution with the following features: (i) unique, (ii) obtained with a simple, systematic and computationally efficient algorithm, and (iii) derived from the relative accuracy which results from the admissible flux intervals. Such a systematic unique solution could be used for instance in MFA- or FBA-based simulators (Richelle et al., 2016; Bogaerts et al., 2017) which require computing unique values of all the internal fluxes at each simulation step time. We propose an iterative algorithm for obtaining this unique

solution via a systematic path through Most Accurate Fluxes (MAF). These latter are defined by using a measure of accuracy which is derived from the admissible flux intervals.

This paper is organized as follows. Section 2 summarizes the basics of MFA and FVA, as well as the way to reduce the problem to inequality constraints by elimination of equality constraints. Section 3 describes the MAF algorithm and defines a flux accuracy measure. Section 4 illustrates the methodology on a case study consisting of hybridoma cell metabolic flux determination. The MAF solution is also compared with mean values obtained from a uniform sampling of admissible solutions. Conclusions and perspectives are given in Section 5.

2. MFA, FVA AND ELIMINATION OF EQUALITY CONSTRAINTS

2.1 Metabolic Flux Analysis (MFA)

Consider a metabolic network described with m internal metabolites involved in n metabolic reactions. $c \in \mathfrak{R}^m$ contains the intracellular metabolite concentrations (in mol.cell⁻¹) and $v \in \mathfrak{R}^n$ the metabolic fluxes (in mol.cell⁻¹.h⁻¹). The mass balances of the internal metabolites are described by the following ODEs

$$\dot{c} = Nv - \mu c \tag{1}$$

where $N \in \mathfrak{R}^{m \times n}$ is the stoichiometric matrix and μ is the specific cell growth rate (in h⁻¹). With the usual assumptions that (i) the internal metabolites do not accumulate in cells (quasi-steady state $\dot{c} = 0$) and (ii) the dilution term - μc may be neglected in comparison with reaction term Nv , (1) reduces to the system of m algebraic homogenous equations with n unknowns

$$Nv = 0 \tag{2}$$

Usually, direct reactions are considered predominant in comparison with their reverse counterparts, hence leading to the positive sign constraints

$$v \geq 0 \tag{3}$$

Some upper bounds may also be considered

$$v \leq v_{\max} \tag{4}$$

All kinds of inequality constraints, e.g. (3) and/or (4), will be grouped in the set of n_i linear inequalities

$$Av \leq b \tag{5}$$

with $A \in \mathfrak{R}^{n_i \times n}$ and $b \in \mathfrak{R}^{n_i}$. Equality constraints may correspond to external flux measurements $v_{\text{ext}} \in \mathfrak{R}^{n_{\text{ext}}}$

$$N_{\text{ext}}v = v_{\text{ext}} \tag{6}$$

with $N_{\text{ext}} \in \mathfrak{R}^{n_{\text{ext}} \times n}$. All kinds of equality constraints, e.g. (2) and (6), will be grouped in the set of n_e linear equalities

$$A_e v = b_e \tag{7}$$

with $A_e \in \mathfrak{R}^{n_e \times n}$ and $b_e \in \mathfrak{R}^{n_e}$. We will assume that these equalities are linearly independent and, consequently, that A_e is a full row rank matrix. Note that measurements of external fluxes v_{ext} can be described with (6) and hence be included in equality constraints (7) but, to cope with measurement noise which could lead to unfeasible solutions when solving system $\{(2),(6)\}$, it is sometimes necessary to describe measurements with inequality constraints

$$\begin{aligned} N_{\text{ext}}v &\geq (1 - e_v)v_{\text{ext}} \\ N_{\text{ext}}v &\leq (1 + e_v)v_{\text{ext}} \end{aligned} \tag{8}$$

with e_v being a variation coefficient linked to the measurement uncertainties. In this case, measurements (8) would be included in the inequality constraints (5). Finally, MFA consists in solving system $\{(5),(7)\}$, i.e. determining the n unknown metabolic fluxes in v based on n_e linear equations under the n_i linear inequality constraints. In many cases, system $\{(5),(7)\}$ is underdetermined in the sense that the number of equations (n_e) is less than the number of unknown fluxes (n).

2.2 Flux Variability Analysis (FVA)

In case system $\{(5),(7)\}$ is underdetermined, lower and upper bounds can be computed for each flux $v(i)$, $i \in [1, n]$, by solving $2n$ linear programs (LPs)

$$v_{\text{MIN}, \text{MAX}}(i) = \underset{v}{\text{Min, Max}} v(i) \quad \forall i \in [1, n] \tag{9}$$

subject to $\{(5),(7)\}$. The lower and upper bounds $v_{\text{MIN}}(i)$ and $v_{\text{MAX}}(i)$ define the interval of admissible flux values for $v(i)$.

2.3 Elimination of the Equality Constraints

Instead of considering the n -dimensional space of fluxes v subject to the underdetermined system $\{(5),(7)\}$ of n_e equalities and n_i inequalities, a reduced n' -dimensional space of fluxes v' (with $n' = n - n_e$), subject to n_i inequalities, can be defined. Indeed, let $A_0 \in \mathfrak{R}^{n \times (n - n_e)}$ be the matrix whose columns define the orthonormal basis for the null space of A_e , i.e. the set of all $v \in \mathfrak{R}^n$ such that $A_e v = 0$. Given that $A_e A_0 = 0$, any flux distribution v satisfying the set of equalities (7) can be redefined as

$$v = v_0 + A_0 v' \tag{10}$$

where v_0 is a particular solution of (7), such that $A_e v_0 = b_e$, and $v' \in \mathfrak{R}^{n'}$. In this new space of reduced fluxes v' , the inequalities (5) are now given by

$$A'v' \leq b' \tag{11}$$

with

$$A' = AA_0 \quad (12)$$

$$b' = b - Av_0 \quad (13)$$

An example of particular solution v_0 is the parsimonious solution which can be obtained by solving the quadratic program (QP)

$$v_0 = \underset{v}{\text{Min}} v^T v \quad (14)$$

subject to $\{(5),(7)\}$. Note that this particular solution is not at all representative of the most probable metabolism in the cell (on the contrary to the unique solution we are seeking) but it has just to satisfy (7).

While MFA in the v space consisted in solving system $\{(5),(7)\}$, in the v' space it consists now in solving system (11), i.e. a set of n_i inequalities. The application of FVA, as summarized in Section 2.2, in this reduced v' space is straightforward:

$$v'_{\text{MIN},\text{MAX}}(i) = \underset{v'}{\text{Min},\text{Max}} v'(i) \quad \forall i \in [1, n'] \quad (15)$$

subject to (11).

3. MOST ACCURATE FLUXES (MAF) ALGORITHM

3.1 Basic Principles

In order to minimize the number of iterations in the algorithm, we first eliminate the equality constraints as explained in Section 2.3. The following two-step procedure is repeated in a loop until the number of identified most accurate fluxes (MAF) is equal to n' , the dimension of v' :

- STEP 1: the lower and upper bounds of the flux distribution v' (v'_{MIN} and v'_{MAX}) are computed with the LPs (15) subject to (11) and to the additional constraints $v'(i) = v'_{\text{MAF}}(i)$ corresponding to the so far identified MAF (no such equality constraints at the first iteration); the mean values v'_{MEAN} of the admissible intervals defined by v'_{MIN} and v'_{MAX} are computed too, as well as their corresponding accuracy measurement v'_{ACC} as defined in the next Section 3.2;

- STEP 2: among the fluxes $v'(i)$ which have not yet been identified as MAF, the next MAF is determined with the lowest value of the accuracy measure $v'_{\text{ACC}}(i)$ and its value is set to its corresponding mean $v'_{\text{MAF}}(i) = v'_{\text{MEAN}}(i)$;

When exiting that sequence, all the most accurate fluxes v'_{MAF} have been identified iteratively and the corresponding v_{MAF} in the original space can be obtained with (10). The way to define and compute the flux accuracy measure v'_{ACC} and the detailed MAF algorithm are proposed, respectively, in Sections 3.2 and 3.3.

3.2 A Flux Accuracy Measure

Remind that, in the reduced space v' , fluxes can be positive or negative. Let $v'_{\text{ACC}}(i)$ be a flux accuracy measure for $v'(i)$ such that the higher the accuracy the lower the value of $v'_{\text{ACC}}(i)$. Let $v'_{\text{MIN}}(i)$ and $v'_{\text{MAX}}(i)$ be the minimum and maximum values defining the admissible interval for $v'(i)$, as computed with (15) subject to (11). The proposed accuracy measure is

$$v'_{\text{ACC}}(i) = \frac{v'_{\text{MAX}}(i) - v'_{\text{MIN}}(i)}{1 + \left| \frac{v'_{\text{MAX}}(i) + v'_{\text{MIN}}(i)}{2} \right|} \quad (16)$$

The basic idea of this relative measure is that the accuracy increases if

- the admissible interval length ($v'_{\text{MAX}}(i) - v'_{\text{MIN}}(i)$) decreases;
- for a given interval length, the absolute value of the interval center increases.

3.3 MAF Algorithm

Based on the two-step procedure proposed in Section 3.1 and the flux accuracy measurement (16) given in Section 3.2, the MAF algorithm is depicted in Fig. 1. Based on the output v'_{MAF} of this algorithm, the final v_{MAF} can be computed with (10).

4. CASE STUDY: METABOLIC FLUXES WITHIN HYBRIDOMA CELLS

4.1 Materials and Experimental Methods

Two cultures (one batch and one fed-batch) of hybridoma cells (HB-58 cell line, American Culture Collection – ATCC) are used in this case study. They were conducted at the State Key Laboratory of Bioreactor Engineering, East China University of Science and Technology (ECUST), Shanghai (Niu et al., 2013). A 2L bubble free bioreactor was used for the batch experiment Exp. 1 in controlled environment (37 °C, 40% DO, pH 7.1, stirring rate 120 rpm) and a 1L bioreactor for the fed-batch experiment Exp. 2 (37 °C, 50% DO, pH 7.0, stirring rate 120 rpm). After a 35h batch phase, fed-batch Exp. 2 was fed with a constant flow rate of 0.1 L/day containing 9.3 mM glutamine and 15 mM glucose. Measurement samples of 10 mL were taken about every 12 h for viable cell density and glucose, glutamine, lactate, ammonium and alanine concentrations. More details on these experiments, e.g. initial concentrations and initial volumes, are given in Amribt et al. (2013).

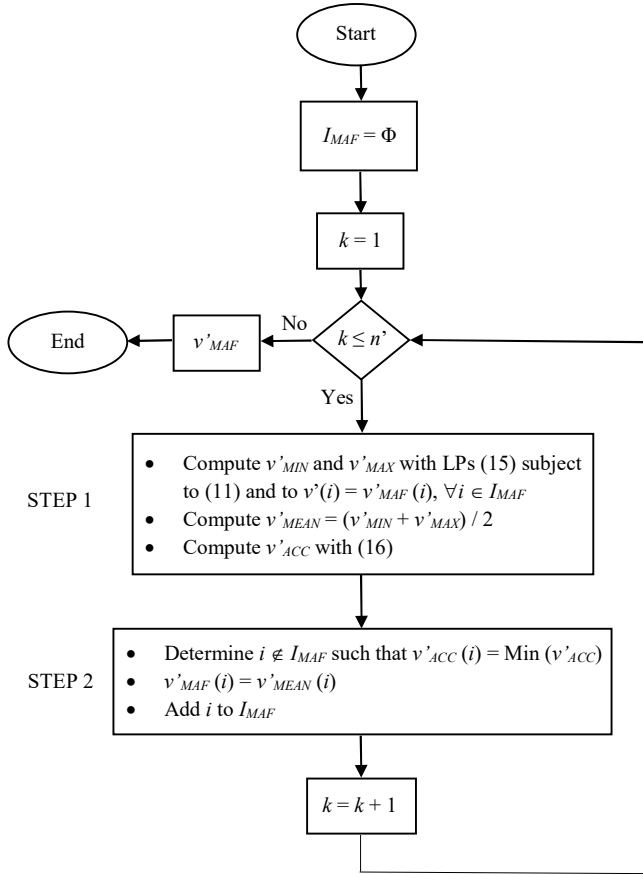


Fig. 1. MAF algorithm. Notations are given in Section 2.3, 3.1 and 3.2. At each iteration k , I_{MAF} is the set of the indexes i corresponding to the most accurate fluxes $v'_{MAF}(i)$ identified up to that iteration.

4.2 FVA and MAF results

The metabolic network used in this case study has been proposed by Provost et al. (2006) and describes the basic mammalian cell intracellular fluxes: glycolysis, citric acid cycle, pentose phosphate pathways, nucleotide synthesis. It contains 24 internal fluxes involving 17 intracellular metabolites which are assumed balanced. Positive sign constraints (3) are considered for all fluxes. To the 17 equality constraints representing the quasi-steady state, one additional equality accounts for the assumption that the purine and pyrimidine nucleotide synthesis fluxes are equal. The network stoichiometry matrix in equ. (2) is therefore $N \in \mathfrak{R}^{18 \times 24}$. Four measured external fluxes are available as a function of time and correspond to glucose and glutamine specific uptake rates and to lactate and ammonium specific production rates. Details about their estimation from extracellular concentration measurements are given in Bogaerts et al. (2017). We will consider either the use of the four measurements ($N_{ext} \in \mathfrak{R}^{4 \times 24}$) or the use of only the two uptake rate measurements ($N_{ext} \in \mathfrak{R}^{2 \times 24}$). These measurements are taken into account through inequality constraints (8) where the minimum variation coefficient linked to measurement uncertainties has been determined as $e_v = 13\%$ and 5% for,

respectively, Exp. 1 and Exp. 2. Putting the inequality constraints (3) and (8) in the general equation (5) leads to a matrix $A \in \mathfrak{R}^{28 \times 24}$ when only the 2 measured uptake fluxes are used and $A \in \mathfrak{R}^{32 \times 24}$ when the 4 measured fluxes are used. Putting the equality constraints (2) in the general equation (7) leads to a matrix $A_e \in \mathfrak{R}^{18 \times 24}$. After eliminating these equality constraints as explained in section 2.3, the reduced flux vector v' defined in (10) belongs to \mathfrak{R}^6 . The particular solution v_0 used in (10) is the parsimonious one defined in (14).

Upper and lower bounds of the internal metabolic fluxes time profiles are computed from FVA based on LPs $\{(15), (11)\}$ (and the final computation of v based on (10)). The MAF solution is computed with the algorithm presented in Fig. 1 (still using (10) for computing v from v'). Results are given (red curves) in Fig. 2 and Fig. 3 for batch Exp. 1, respectively with 2 and 4 measured external fluxes and in Fig. 4 and Fig. 5 for fed-batch Exp. 2, respectively with 2 and 4 measured external fluxes. MAF solutions may differ significantly from the mean of the corresponding upper and lower bounds, e.g., fluxes linked to the citric acid cycle (v_8 to v_{12}) which are closer to the lower bounds due to overflow metabolism.

4.3 Comparison with Mean of Uniform Samples

Besides the already mentioned advantages of the unique MAF solution, it is interesting to compare it to a representative solution derived from the admissible flux distribution. This can be achieved by computing the mean value of a large collection of admissible solutions uniformly sampled in the solution space. One of the most recent algorithms for the uniform sampling of admissible metabolic flux solutions is the Coordinate Hit-and-Run with Rounding (CHRR) algorithm proposed by Haraldsdóttir et al. (2017) and implemented in the COBRA toolbox (Schellenberger et al., 2011). The CHRR solution corresponding to the mean of 10^4 samples is compared in Fig. 2 to Fig. 5 (green curves) with the MAF solutions (red curves). The Normalized Root-Mean-Square Error (NRMSE) has been computed, for each of the 24 fluxes in each of the four cases depicted in Fig. 2 to Fig. 5 (i.e., Exp. 1 or Exp. 2 with 2 or 4 external measured fluxes):

$$NRMSE(v(i)) = \frac{\sqrt{\frac{1}{T} \sum_{k=1}^T (v_{MAF}(i, t_k) - v_{CHRR}(i, t_k))^2}}{\frac{1}{T} \sum_{k=1}^T v_{CHRR}(i, t_k)} \quad (20)$$

where $v_{MAF}(i, t_k)$ and $v_{CHRR}(i, t_k)$ are respectively the MAF and CHRR solutions for the i^{th} flux ($i \in [1, 24]$) at time t_k , and T is the total number of time instants (21 in this case). The mean values (on the 24 fluxes) of the NRMSE are equal to 17% (Exp. 1 with 2 measured output fluxes), 10% (Exp. 1 with 4 measured output fluxes), 28% (Exp. 2 with 2 measured output fluxes) and 9% (Exp. 2 with 4 measured output fluxes). This confirms the similarity between MAF and CHRR solutions which can be observed in Fig. 2 to Fig. 5. Note that the computational time is of course significantly lower for

computing the MAF solution than for obtaining the 10^4 samples of the CHRR solution (e.g., about 30 times less in the case of Exp. 2. with 2 measured output fluxes).

5. CONCLUSIONS AND PERSPECTIVES

This contribution proposes a new efficient algorithm computing a unique flux distribution for underdetermined metabolic networks. A common point with FBA is that it consists of an optimal solution which can be easily obtained by solving a series of LPs. A first key advantage is that the uniqueness of the solution is fully guaranteed on the contrary to FBA which often leads to multiple admissible solutions corresponding to the same cost function optimum. A second key advantage is that the optimality criterion is only based on the accuracy of the iteratively estimated fluxes (based on the proposed concept of MAF) without requiring any biological assumption regarding the definition of an objective cost function. Future work will consist in testing the MAF algorithm on other and larger metabolic networks.

REFERENCES

- Amribt, Z., Niu, H., and Bogaerts, Ph. (2013). Macroscopic modelling of overflow metabolism and model based optimization of hybridoma cell fed-batch cultures. *Biochem. Eng. J.*, 70, 196-209.
- Antoniewicz, M. (2013). Dynamic metabolic flux analysis – tools for probing transient states of metabolic networks. *Curr. Opin. Biotechnol.*, 24, 973-978.
- Bogaerts, Ph., Mhallem Gziri, K., and Richelle, A. (2017). From MFA to FBA: Defining linear constraints accounting for overflow metabolism in a macroscopic FBA-based dynamical model of cell cultures in bioreactor. *J. Process Control*, 60, 34-47.
- Fernandes de Sousa, S., Bastin, G., Jolicoeur, M., and Vande Wouwer, A. (2016). Dynamic metabolic flux analysis using a convex analysis approach: Application to hybridoma cell cultures in perfusion. *Biotechnol. Bioeng.*, 113, 1102-1112.
- Haraldsdóttir, H., Cousins, B., Thiele, I., Fleming, R., Vempala, S. (2017). CHRR: coordinate hit-and-run with rounding for uniform sampling of constraint-based models. *Bioinformatics*, 33 (11), 1741-1743.
- Klamt, S., and Stelling, J. (2003). Two approaches for metabolic pathway analysis? *Trends in Biotech.*, 21 (2), 64-69
- Llaneras, F., and Picó, J. (2007). An interval approach for dealing with flux distributions and elementary modes activity patterns. *J. of Theor. Biol.*, 246, 290-308.
- Mahadevan, R., and Schilling, C. (2003). The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab. Eng.*, 5, 264-276.
- Motamedian, E., and Naeimpoor, F. (2018). LAMOS: A linear algorithm to identify the origin of multiple optimal flux distributions in metabolic networks. *Comput. Chem. Eng.*, 117, 372-377.
- Niu, H., Amribt, Z., Fickers, P., Tan, W., and Bogaerts, Ph. (2013). Metabolic pathway analysis and reduction for mammalian cell cultures: Towards macroscopic modeling. *Chem. Eng. Sci.*, 102, 461-473.
- Orth, J., Thiele, I., and Palsson, B. (2010). What is flux balance analysis? *Nature Biotechnol.*, 28 (3), 245-248.
- Provost, A., Bastin, G., Agathos, S., and Schneider, Y.-J. (2006). Metabolic design of macroscopic bioreaction models: application to Chinese hamster ovary cells. *Bioproc. Biosyst. Eng.*, 29, 349-366.
- Richelle, A., Mhallem Gziri, K., and Bogaerts, Ph. (2016). A methodology for building a macroscopic FBA-based dynamical simulator of cell cultures through flux variability analysis. *Biochem. Eng. J.*, 114, 50-61.
- Schellenberger, J., Que, R., Fleming, R., Thiele, I., Orth, J, Feist, A., Zielinski, D., Bordbar, A., Lewis, N., Rahmanian, S., Kang, J., Hyduke, D., Palsson, B. (2011). Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat. Protoc.*, 6 (9), 1290-1307.
- Stephanopoulos, G., Aristidou, A., and Nielsen, J. (1998). *Metabolic engineering: Principles and methodologies*, chapter 8. Academic Press, San Diego.

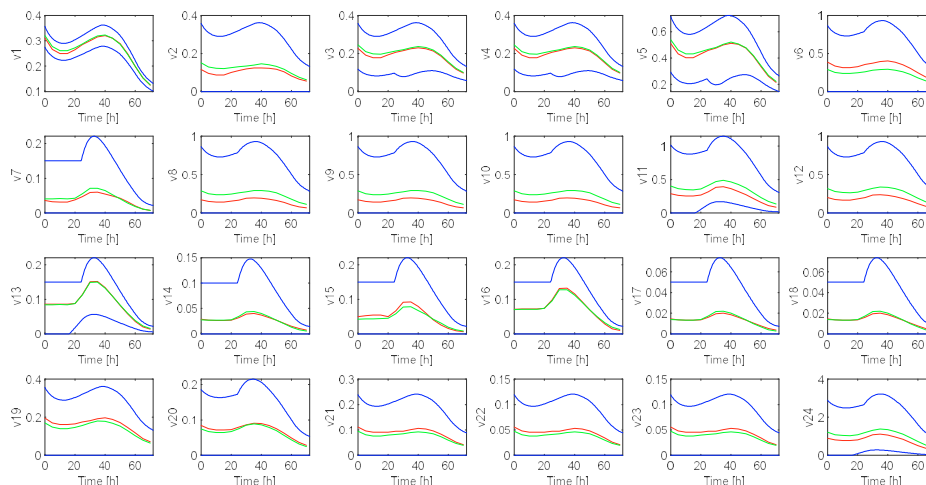


Fig. 2. Time profiles of the internal metabolic fluxes for batch Exp. 1 with 2 external measured fluxes (glucose and glutamine specific uptake rates) : upper and lower bounds from FVA (blue), MAF solution (red) and mean of uniform sampling (10^4 samples) with CHRR (green).

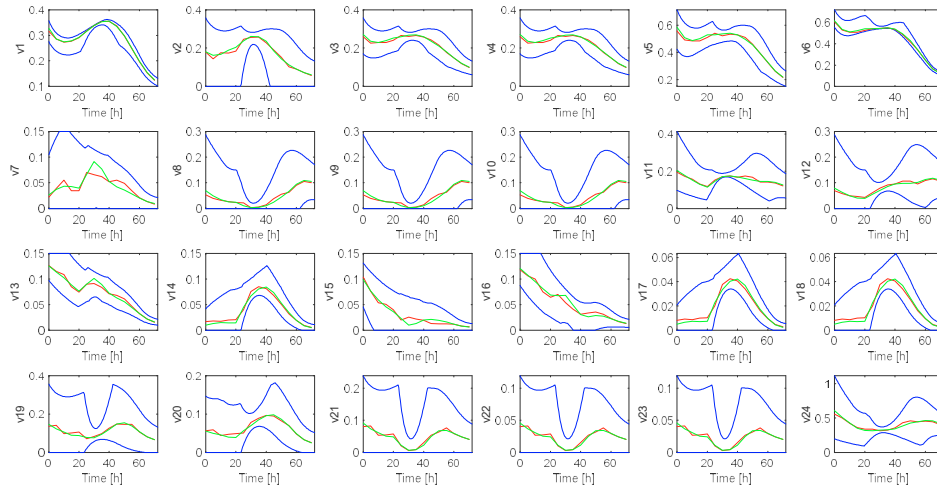


Fig. 3. Time profiles of the internal metabolic fluxes for batch Exp. 1 with 4 external measured fluxes (glucose and glutamine specific uptake rates and lactate and ammonium specific production rates) : upper and lower bounds from FVA (blue), MAF solution (red) and mean of uniform sampling (10^4 samples) with CHRR (green).

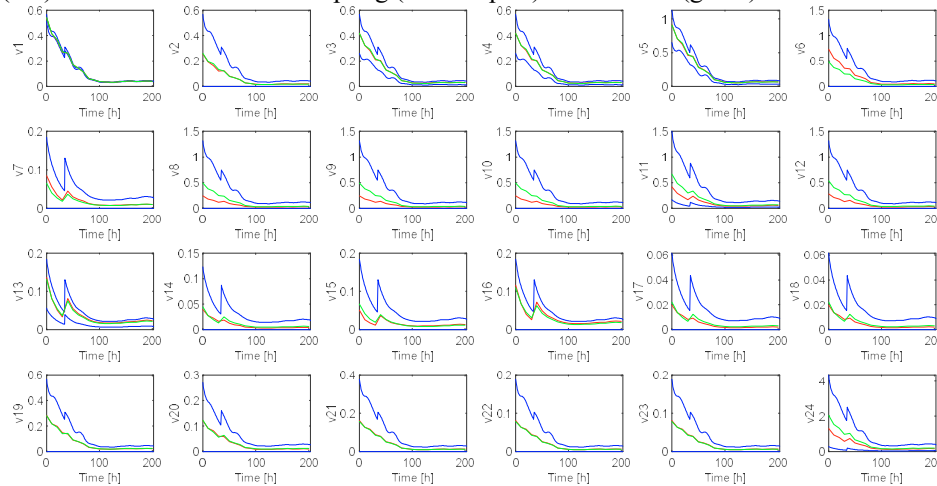


Fig. 4. Time profiles of the internal metabolic fluxes for fed-batch Exp. 2 with 2 external measured fluxes (glucose and glutamine specific uptake rates) : upper and lower bounds from FVA (blue), MAF solution (red) and mean of uniform sampling (10^4 samples) with CHRR (green).

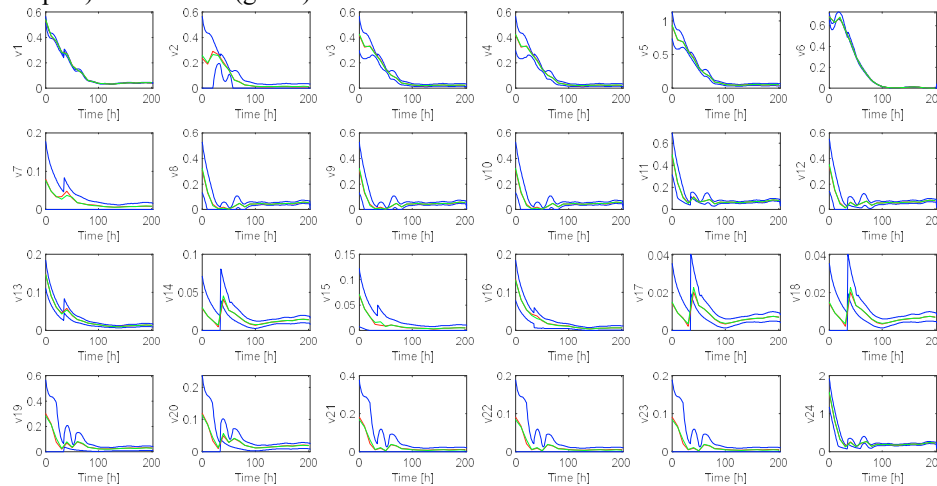


Fig. 5. Time profiles of the internal metabolic fluxes for fed-batch Exp. 2 with 4 external measured fluxes (glucose and glutamine specific uptake rates and lactate and ammonium specific production rates) : upper and lower bounds from FVA (blue), MAF solution (red) and mean of uniform (10^4 samples) sampling with CHRR (green).