Proceedings of the 11th International Symposium on
Computer Applications in Biotechnology (CAB 2010),
Leuven, Belgium, July 7-9, 2010
Julio R. Banga, Philippe Bogaerts, Jan Van Impe,
Denis Dochain, Ilse Smets (Eds.)

Possibilistic estimation of metabolic fluxes during a batch process accounting for extracellular dynamics

F. Llaneras*, A. Sala*
J Picó*

**AI2, Universidad Politécnica de Valencia, Camino de Vera s/n, 46022 Valencia, Spain (frallaes@upvnet.upv.es; jpico@ai2.upv.es)*

Abstract: Constraint-based models use the available knowledge about the operating constraints (e.g., mass balances and thermodynamic laws) to define a space of feasible states for cell cultures. Predictions can then be obtained incorporating experimental measurements of metabolite concentrations to perform a metabolic flux analysis. Although these predictions are typically static, aimed to study cells at given state, several works accounting for extracellular dynamics can be found in literature. In this work we formulate these predictions of time-varying fluxes and metabolites as possibilistic constraint satisfaction problems. The benefit of the described approach is that richer estimates are obtained —not only point-wise ones—, while considering uncertainty and even in scenarios of data scarcity. The method could also be the basis for on-line fault detection in industrial processes.

Keywords: constraints, possibility theory, uncertainty, metabolic flux analysis.

1. INTRODUCTION

To study cell systems with a system-level approach, the elementary biochemical reactions taking place within living cells are being collected in networks. Then, these networks can be used to assemble constraint-based models (Llaneras, 2008; Palsson, 2006).

Given a metabolic network with m metabolites and n reactions, the mass balances around the metabolites can be represented by the equation:

$$\frac{d\mathbf{c}}{dt} = \mathbf{S_c} \cdot \mathbf{v} \qquad (1)$$

where $\mathbf{c}$ is the vector of metabolite concentrations, $\mathbf{v}$ is the n-dimensional vector of reaction fluxes and $\mathbf{S_c}$ is the $m \times n$ stoichiometric matrix.

Nevertheless, as reaction kinetics are still rarely known, internal metabolites are often assumed to be at steady-state. Under this assumption, cells during a batch process can be represented as follows:

$$\mathbf{0} = \mathbf{S} \cdot \mathbf{v} \qquad (2a)$$

$$\frac{d\mathbf{e}}{dt} = \mathbf{S_e} \cdot \mathbf{v} \qquad (2b)$$

where $\mathbf{e}$ is a vector of external metabolite concentrations, substrates and products, $\mathbf{S}$ is the stoichiometric matrix correspondent to the intracellular metabolites, and $\mathbf{S_e}$ is a selection matrix linking each external metabolite with its flux (without loss of generality, each extracellular metabolite can be represented with two nodes, one internal and one external, so that there is only one reaction in $\mathbf{v}$ accounting for its total uptake/consumption. An example of these matrices is given below). Biomass, if considered, can be treated as an external metabolite and its synthesis represented with a flux in $\mathbf{v}$.

Hence, equation (2a) defines a space of stoichiometrically feasible intracellular flux states, and (2b) links this internal state with the cell environment through the uptake of substrates and the outflow of products.

Along with mass balances, other inequality constraints can be imposed, such as the irreversibility of certain reactions:

$$\mathbf{D} \cdot \mathbf{v} \geq \mathbf{0} \qquad (3)$$

where $\mathbf{D}$ is a diagonal matrix with $\mathbf{D_{ii}} = 1$ if the flux i is irreversible (otherwise 0).

The resultant constraint-based models (2-3) are typically used under a static point of view to analyze the cells at a given set of circumstances; extracellular dynamics are thus not considered and derivatives are replaced by constant uptake or production rates in (2b). Then, to get predictions from these models, experimental measurements of the extracellular fluxes can be incorporated as constraints to perform a metabolic flux analysis (MFA) (Heijden, 1994).

This contribution explores the use of constraint-based models and MFA when extracellular dynamics are taken into account, as it has been done before in several works (Herwig, 2002; Takiguchi, 1997; Henry, 2007; Mahadevan, 2002; Hjersted, 2009). In particular, we discuss the benefits that a possibilistic approach could bring in this case.

2. PRELIMINARIES: STATIC PMFA

Possibilistic metabolic flux analysis (PMFA) was recently introduced (Llaneras, 2009) as a variant of traditional MFA

providing richer estimates, well suited for scenarios of data scarcity, and computationally efficient. The approach uses possibility theory, instead of probability, as framework for handling uncertainty in constraint satisfaction problems (Dubois, 1996). The relationship between probability and possibility was briefly discussed in (Llaneras, 2009).

Consider a constraint-based model representing the feasible (steady) states of cells at a given moment:

$$\mathcal{MOC} = \begin{cases} \mathbf{S}\cdot\mathbf{v} = \mathbf{0} \\ \mathbf{D}\cdot\mathbf{v} \geq \mathbf{0} \end{cases} \quad (4)$$

A set of experimentally measured fluxes are incorporated, accounting for imprecision by means of two vectors of suitable artificial slack variables ε_1 and μ_1:

$$\mathcal{MEC} = \begin{cases} \hat{\mathbf{v}}_\mathbf{m} = \mathbf{v}_\mathbf{m} + \varepsilon_1 - \mu_1 \\ \varepsilon_1, \mu_1 \geq 0 \end{cases} \quad (5)$$

These slack variables are penalized in a cost index J to generate a possibility distribution that relax the basic assertion $\mathbf{v'_m} = \mathbf{v_m}$. Under a non-interactivity assumption, an overall cost index J reflecting the log-possibility of a particular flux vector $\mathbf{v}$ is defined as follows:

$$\mathrm{J} = \alpha\cdot\varepsilon_\mathbf{1} + \beta\cdot\mu_\mathbf{1} \quad (6)$$

where α and ß are row vectors of user-defined, sensor accuracy, or reliability, coefficients.

The possibility π of each solution δ of (4-5) is given by:

$$\pi(\delta) = \mathrm{e}^{-\mathrm{J}(\delta)} \quad \delta \in \mathcal{MEC} \cap \mathcal{MOC} \quad (7)$$

In this way, equations (4) and (5) define a constraint, satisfaction problem, where the "degree of possibility" of each candidate solution —or flux vector $\mathbf{v}$— is given by the cost index (6) as in (7). This basic setting allows performing flux estimations, which can be conveniently cast as linear programming optimization problems (LP). See (Llaneras, 2009) for details.

3. DYNAMIC PMFA: PROBLEM SETTING

In this section we show how a constraint, satisfaction problem can be formulated to incorporate extracellular dynamics in PMFA.

Consider a time period [0, T] divided in t intervals by the sampling rate of the measurements. First, we define the constraints conforming the model at successive time instants k, hereinafter referred as $\mathcal{MOC}(k)$:

$$\mathbf{0} = \mathbf{S}\cdot\mathbf{v}(k) \quad (8a)$$

$$\frac{\mathbf{e}(k) - \mathbf{e}(k-1)}{\Delta\mathrm{T}} = \mathbf{S_e}\cdot\mathbf{v}(k) \quad (8b)$$

$$\mathbf{D}\cdot\mathbf{v}(k) \geq \mathbf{0} \quad (8c)$$

$$\mathbf{e}(k) \geq \mathbf{0} \quad (8d)$$

Initial conditions should be given, at least, for each non-measured metabolite. Notice that a backward approximation of derivatives is used, but other alternative might be chosen.

The measured concentrations of extracellular metabolites are then incorporated as constraints, $\mathcal{MEC}(k)$:

$$\begin{aligned} &\mathbf{e_m}(k) = \mathbf{e'_m}(k) + \varepsilon_1(k) - \mu_1(k) + \varepsilon_2(k) - \mu_2(k) \\ &\varepsilon_1(k), \mu_1(k) \geq 0 \\ &0 \leq \varepsilon_2(k) \leq \varepsilon_2^{\max}(k) \\ &0 \leq \mu_2(k) \leq \mu_2^{\max}(k) \end{aligned} \quad (9)$$

where $\mathbf{e_m}(k)$ represent the actual concentrations of each metabolite and $\mathbf{e'_m}(k)$ the measured values, the slack variables ε and μ are introduced to consider its uncertainty.

The overall cost index J(k) at each time instant k can be defined as follows:

$$\mathrm{J}(k) = \alpha(k)\cdot\varepsilon_\mathbf{1}(k) + \beta(k)\cdot\mu_\mathbf{1}(k) \quad (10)$$

The index J(k) reflects the log-possibility of the values for each measured metabolite $\mathbf{e_m}(k)$, as given by (7). The interpretation of (9-10) may be: "$\mathbf{e'_m}(k)$=$\mathbf{e_m}(k)$ is fully possible; the more $\mathbf{e'_m}(k)$ differs from $\mathbf{e_m}(k)$, the less possible such situation is". In this way, each measured metabolite is represented with a possibility distribution defined by the user (see examples below).

The user has to define the bounds ε_2^{max} and μ_2^{max} and the weights α(k) and β(k) to describe each measured metabolite in possibilistic terms. The bounds ε_2^{max} and μ_2^{max} define an interval of "fully possible" values (possibility π=1). For instance, the user can choose a band of 10% around the measured value to capture random errors. The values α(k) and β(k) define the decreasing possibility that the user assigns to values out of this interval. For instance, α(k) and β(k) can be choose so that an error of 30% has π=0.5. Herein two pairs of slack variables have been defined to represent each measurement, but slack variables can be added to achieve more complex representations.

Notice that uncertainty in the model constraints (4) (e.g., imprecision of the values of $\mathbf{S}$ and $\mathbf{S_e}$) can be sometimes considered in an analogous way.

4. DYNAMIC PMFA: ESTIMATIONS

Once the PMFA problem has been formulated, all the variables of the system —fluxes and metabolites— can be estimated solving linear programming (LP) problems. Two basic estimates can be obtained: the maximum possibility solution, which contains the most possible values for each variable and gives an indication of consistency, and the

possibilistic intervals, which provide a richer and more reliable prediction.

4.1 Maximum possibility solution

The maximum possibility (minimum-cost) solution of the constraint satisfaction problem (8-9) can be obtained solving a LP problem:

$$\begin{aligned} \min \quad & \mathrm{J_T} = \sum_{\mathrm{k=1}}^{t} \mathrm{J}(k) \\ \text{s.t.} \quad & \begin{cases} \mathcal{MOC}(k) & \forall k \\ \mathcal{MEC}(k) & \forall k \end{cases} \end{aligned} \tag{11}$$

with a maximum possibility $\pi^{\mathrm{mp}} = \exp(\mathrm{J}_{\mathrm{T}}^{\min})$.

The maximum possibility π^{mp} is an indicator of the consistency between model and measurements. Possibility equal to one must be interpreted as complete agreement, whereas lower values imply that there is some degree of error in the measurements (or in the model). Indeed, the slack variables ε_1 and μ_1 could be inspected to investigate which measurements may be causing the inconsistency.

The solution of (11) also provides the most possible estimate for each flux $\mathbf{v}(k)$ and metabolite $\mathbf{e}(k)$. However, even if uncertainty is indeed considered, these point-wise estimates can be unreliable, or insufficient, when multiple solutions are reasonably possible (a common situation due to a lack of measurements). As a better alternative, possibilistic intervals can be computed solving a set of LP problems.

4.2 Possibilistic estimates for fluxes and metabolites

The interval of values with a conditional possibility higher than γ for a given flux, denoted as $[\, v_{\mathrm{i},\gamma}^{\mathrm{m}}(k)\ v_{\mathrm{i},\gamma}^{\mathrm{M}}(k)\,]$, can be computed solving two LP problems:

$$\begin{aligned} & v_{i,\gamma}^{m}(k) = \min v_{\mathrm{i}}(k) \\ \text{s.t.} \quad & \begin{cases} \mathcal{MOC}(k) & \forall k \\ \mathcal{MEC}(k) & \forall k \\ \sum \mathrm{J}(k) - \log \pi(v_{mp}) < -\log \gamma \end{cases} \end{aligned} \tag{12}$$

The upper bound would be obtained by replacing minimum by maximum. Notice that possibilistic intervals can be computed for metabolite concentrations in a analogous way, simply replacing v(k) for e(k).

The possibilistic intervals have a similar interpretation to "confidence intervals" ("credible intervals") in Bayesian statistics. They enclose all the values for a given variable with the desired degree of possibility at k, and thus they capture the uncertainty of each estimate. This uncertainty will be caused by the imprecision of the measurements — which is translated to the estimates in non-trivial ways due to the model structure and the constraints—, but also by a lack of knowledge —measurements will be usually insufficient to offset the underdeterminacy of the model.

In summary, equation (12) allows to estimate all the variables on the system, metabolites and fluxes, along the duration of the batch process, based on the knowledge embedded in the constraint-based model and the set of available measurements. Furthermore, the estimation will be reliable, because uncertainty is considered, but also conservative: it is not point-wise estimate, which could be deviated from the actual value when several candidates are reasonably possible, but an interval, which in this situation will be wider, but not biased.

5. EXAMPLE I

This section is devoted to illustrate the use of dynamic PMFA with a simple example. Consider the metabolic network depicted in fig. 1. The network has three internal metabolites (M1, M2 and M3), three external ones (E1 E2 and E3), and six fluxes (v_1-v_6) connecting them.

Hence, the vector of extracellular metabolite concentrations and the vector of metabolic fluxes at k are the following:

$$\mathbf{e}(k) = \begin{pmatrix} \mathrm{e}_1(k) \\ \mathrm{e}_2(k) \\ \mathrm{e}_3(k) \end{pmatrix} \qquad \mathbf{v}(k) = \begin{pmatrix} \mathrm{v}_1(k) \\ \mathrm{v}_2(k) \\ \vdots \\ \mathrm{v}_6(k) \end{pmatrix}$$

5.1 Model-based constraints

The information embedded in the network defines the model-based constraints (4). The stoichiometric matrix **S** is the following:

$$\mathbf{S} = \begin{pmatrix} -2 & 1 & 0 & -1 & 0 & 0 \\ 1 & 0 & 1 & 0 & -1 & 0 \\ 0 & -1 & -1 & 0 & 0 & 1 \end{pmatrix}$$

The selection matrix $\mathbf{S_e}$, which links the extracellular metabolites with the intracellular fluxes, is defined so that substrates uptakes and production rates are positive:

$$\mathbf{S_e} = \begin{pmatrix} 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

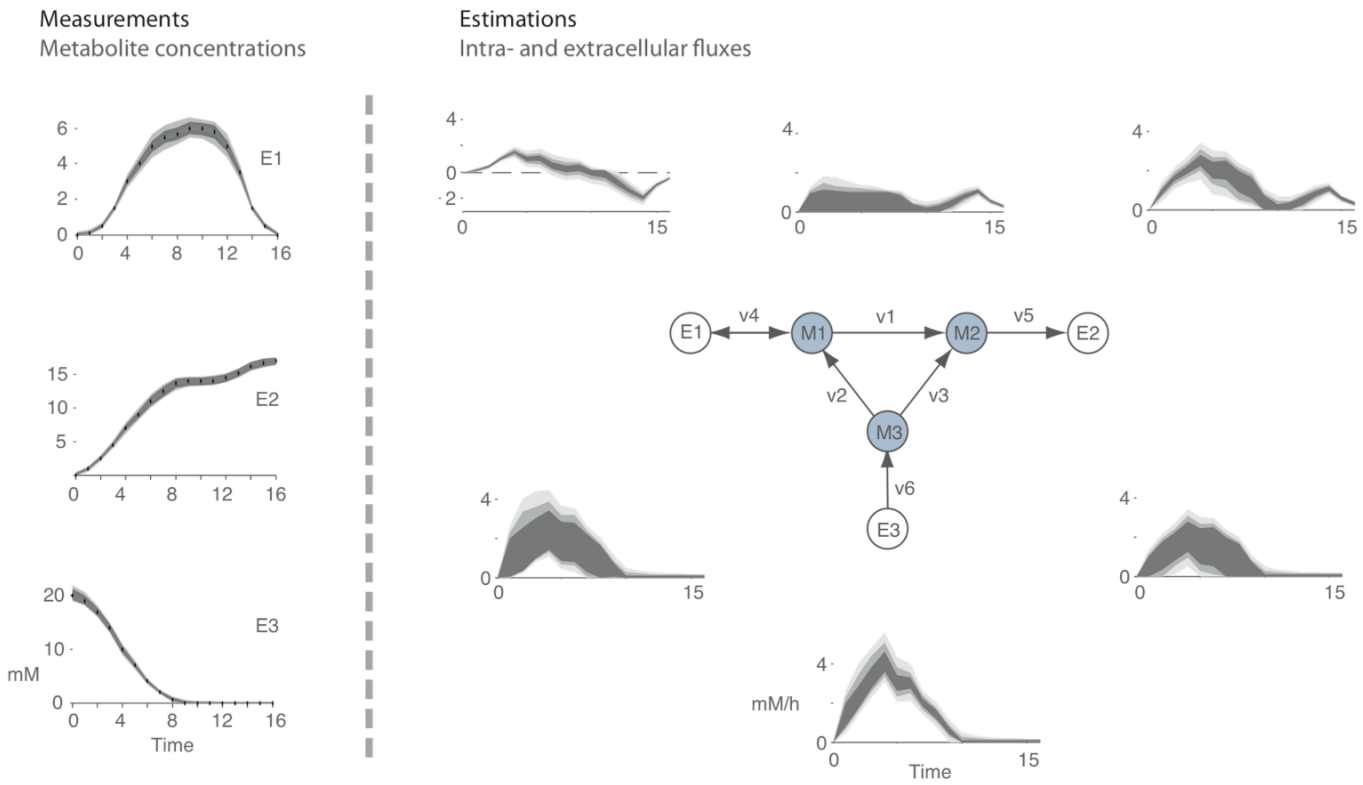

Figure 1. Example I. The measured concentrations for three metabolites (M1, M2 and M3) are depicted on the left (black dots). Concentrations estimated with PMFA for three chosen degrees of possibility (π=1, π=0.5 and π=0.15) are also depicted there (grey areas). The six fluxes of the network (v1-v6) estimated with PMFA for the three degrees of possibility (π=1, π=0.5 and π=0.15) are depicted on the right (grey areas).

And the following diagonal matrix defines the constraints due to reaction reversibilities:

$$\mathbf{D} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

5.2 Measurement-based constraints

We assume that the concentration of the three external metabolites is measured along time (0h-16h). Then, to account for their intrinsic imprecision, measurements are represented in possibilistic terms, as follows:

- Values near the measured ones, within ±2% deviation, are considered fully possible (to account for *systemic errors*).

- A decreasing possibility is assigned to larger deviations; values with a deviation of ±5% have a possibility of 0.5 and those with a deviation of ±10% a possibility of 0.15.

This is achieved choosing the necessary weights (α and ß) and bounds (ε_2^{max}, μ_2^{max}) at each time instant k. Notice also that possibility has been defined by conjunction; thus, if, for instance, two measurements are deviated with possibilities 0.8 and 0.5 respectively, their joint possibility will be 0.4. The measurements are represented in fig. 1 (left).

5.3 Estimation of fluxes and metabolites along time

We have used (12) to estimate the fluxes $\mathbf{v}(k)$ and the metabolites concentrations $\mathbf{e}(k)$ along the duration of the batch process (0-16h). The possibilistic intervals for three degrees of possibility were computed at each time instant, so 2·3·17 LP problems have been solved for each flux/metabolite. Results are depicted in fig. 1 (right).

For this example we get estimates for three degrees of possibility: π=1, to capture all the estimates that are equally and fully possible, π=0.5, to enclose estimates considered highly possible (an error of ±5% in only one measurement or smaller errors in several measurements) and π=0.15, to capture values that are still reasonably possible (an error of ±5% in only one measurement or smaller errors in several measurements). Notice, however, that is the user who has to choose these intervals and interpret them taking into account how measurements uncertainty was represented in 5.2.

6. EXAMPLE II: CHO CELLS

We have also considered a real-scale example of Chinese Hamster Ovary (CHO) cells cultivated in batch mode in stirred flasks. The metabolic network was taken from (Bastin,

2008) and describes the metabolism concerned with the two main energetic nutrients, glucose and glutamine.

Measurements of concentration for glucose (G), alanine (A), lactate (L), glutamine (Q) and ammonia (NH4) and the growth rate (μ) were taken from (Provost, 2006). Those data were collected with a sample rate of 24h. Measurements are represented in possibilistic terms as in the first example, values near the measured ones (±2% deviation) are considered fully possible, while a decreasing possibility is assigned to larger deviations as in example 1.

Now we solved the LP problems (12) to estimate the fluxes and the metabolite concentrations along the batch process.

6.1 Estimation of metabolite concentrations along time

The estimated evolution of the metabolite concentrations is depicted in fig. 2. It can be observed how the method is able to estimate a metabolite that was not measured (CO2). Moreover, the uncertainty of the measured metabolites might have been reduced using the other measurements.

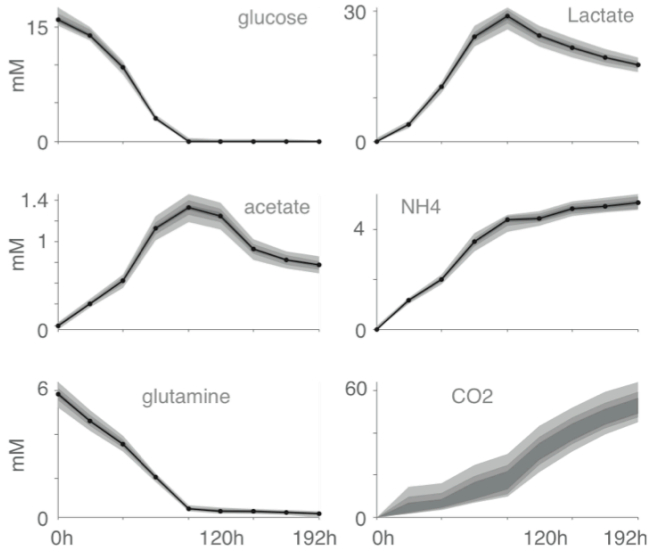

Figure 2. Measured and estimated metabolite concentrations during a cultivation of CHO cells. Measurements are denoted with black dots. The estimations for the metabolites are given by three possibilistic intervals (π=1, π=0.5 and π=0.15) represented which grey areas.

6.2 Estimation of fluxes along time

The estimated intracellular fluxes are presented in fig. 3 (for brevity, only 6 out of 31 are shown). It can be observed that some of them are estimated with precision (v_5 or v_7), whereas other estimates are wider (v_8 or v_{12}). However, even the wider ones can be valuable: for instance, v_{12} indicates that reaction 12 is always active during exponential growth (0h-120h). Uptake or production rates for the extracellular metabolites can also be estimated (v_{25} or v_{26}).

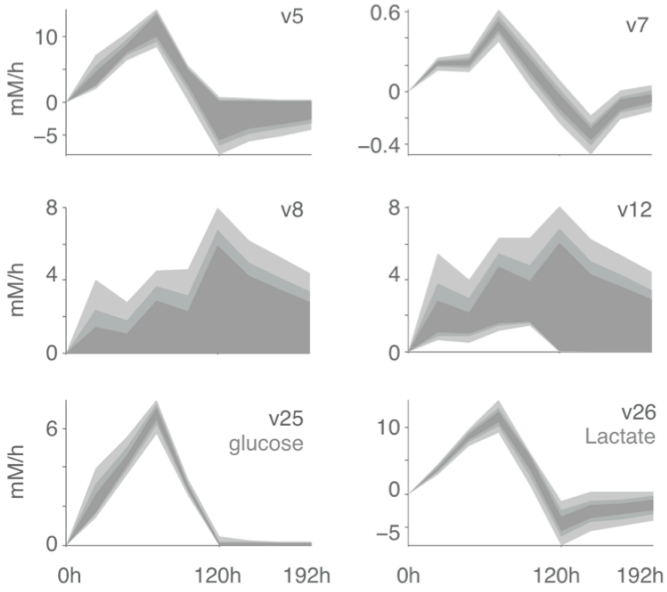

Figure 3. Estimated fluxes during a cultivation of CHO cells. The estimates for 6 out of 31 fluxes are given for three possibilistic intervals (π=1, π=0.5 and π=0.15).

7. CONCLUSIONS

We have explored the benefits of a possibilistic approach to MFA when extracellular dynamics are accounted for. We have shown that dynamic PMFA estimates the metabolic fluxes and the extracellular metabolite concentrations during a cultivation process. In comparison with other procedures based on MFA, our proposal considers measurements uncertainty and gives richer estimates —intervals instead of point-wise values. Notice also that this dynamic version inherits other benefits of static PMFA discussed in (Llaneras, 2009), for instance, it handles data scarcity and allows to represent model constraints in a flexible way to account for imprecision (for instance, one could consider uncertainty in matrices **S** and $\mathbf{S_e}$).

LP problems are very efficient and MFA is usually limited to relatively small networks (100-150 variables or less, and 5-15 measured metabolites), however, the approach we follow to perform the estimations will be computationally expensive, and even non solvable, if the sampling rate is too high (since constraints at every time instant *k* are simultaneously taken into account). Fortunately, this difficulty will be rare because extracellular dynamics are typically slow and measurements are scarce on time. Future work will be devoted to develop a more efficient approach, able to deal with larger sampling rates, and thus making it possible to apply the methodology to other problems.

Current work is being devoted to the use of dynamic PMFA to monitor measurements and model consistency during a running process. This information may be useful for on-line, or quasi on-line, fault detection in industrial processes.

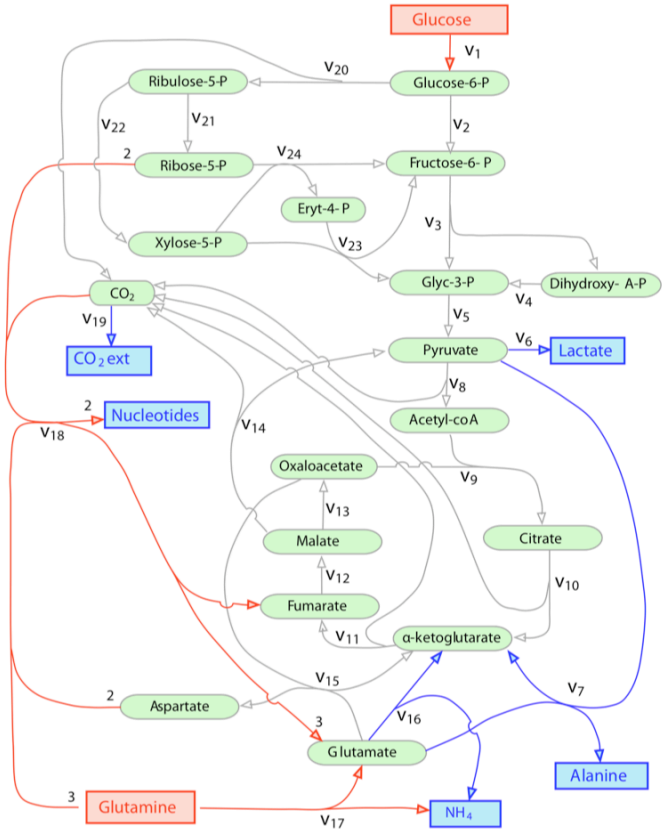

Figure 4. Metabolic network of CHO cells, adapted from (Bastin, 2008).

ACKNOWLEDGEMENTS

This research has been partially supported by the Spanish Government (1st and 3rd authors are grateful to grants DPI2008-06880-C03-01 and A/016560/08, 2nd author is grateful to DPI2008-06731-C02-01 and PROMETEO /2008/088). FLL is recipient of a fellowship from the Spanish Ministry of Science and Innovation (FPU AP2005-1442).

REFERENCES

Bastin G. (2008) 'Quantitative analysis of metabolic networks and design of minimal bioreaction models' International Conference in Honor of Claude Lobry, 9:41-45.

Dubois D., Fargier H., and Prade H. (1996) 'Possibility theory in constraint satisfaction problems: handling priority, preference and uncertainty' Applied Inteligence, 6(4):287–309.

Heijden R.T., Romein B., Heijnen J.J., Hellinga C., Luyben K.C. (1994) 'Linear Constraint Relations in Biochemical Reaction Systems: II. Diagnosis and Estimation of Gross' *Biotechnol Bioeng* 43 (1), 11-20.

Henry O., Kamen A., Perrier M. (2007) 'Monitoring the physiological state of mammalian cell perfusion processes by on-line estimation of intracellular fluxes' Journal of Process Control.

Herwig C., von Stockar U. (2002) 'A small metabolic flux model to identify transient metabolic regulations in Saccharomyces cerevisiae' Bioprocess and Biosystems 2002, 24:395-403.

Hjersted J.L. and Henson M.A. (2009) 'Steady-state and dynamic flux balance analysis of ethanol production by Saccharomyces cerevisiae' IET systems biology, vol. 3 (3) pp. 167-79.

Llaneras F.. and Picó J. (2008) 'Stoichiometric modelling of cell metabolism' *J Biosci Bioeng*, 105 (1), 1-11.

Llaneras F., Sala A. and Picó J. (2009). 'A possibilistic framework for constraint-based metabolic flux analysis.' *BMC Syst Biol,* 31 (3), 79.

Mahadevan R., Edwards J.S., Doyle F.J. (2002) 'Dynamic flux balance analysis of diauxic growth in Escherichia coli' Biophys J, vol. 83 (3) pp. 1331-40.

Palsson B. (2006) 'Systems Biology: Properties of Reconstructed Networks' Cambridge University Press, pp. 334.

Price N.D., Papin J.A., Schilling J.H., Bernhard B.O. (2003) 'Genome-scale microbial in silico models: the constraints-based approach' Trends in Biotechnology, vol. 21 (4) pp. 162-9.

Provost A., Bastin G., Agathos S.N., Schneider Y.J. (2006) 'Metabolic design of macroscopic bioreaction models: application to Chinese hamster ovary cells' Bioprocess and biosystems engineering, vol. 29 (5-6) pp. 349-66.

Takiguchi N., Shimizu H., Shioya S. (1997) 'An on-line physiological state recognition system for the lysine fermentation process based on a metabolic reaction model' Biotechnol Bioeng, 55:170-181.