BioMed Central

Methodology article

A procedure for the estimation over time of metabolic fluxes in scenarios where measurements are uncertain and/or insufficient

Francisco Llaneras* and Jesús Picó

Address: Dept. of Systems Engineering and Control, Technical University of Valencia, Camino de Vera s/n, 46022 Valencia, Spain

Email: Francisco Llaneras* - frallaes@doctor.upv.es; Jesús Picó - jpico@ai2.upv.es

* Corresponding author

Abstract

Background: An indirect approach is usually used to estimate the metabolic fluxes of an organism: couple the available measurements with known biological constraints (e.g. stoichiometry). Typically this estimation is done under a static point of view. Therefore, the fluxes so obtained are only valid while the environmental conditions and the cell state remain stable. However, estimating the evolution over time of the metabolic fluxes is valuable to investigate the dynamic behaviour of an organism and also to monitor industrial processes. Although Metabolic Flux Analysis can be successfully applied with this aim, this approach has two drawbacks: i) sometimes it cannot be used because there is a lack of measurable fluxes, and ii) the uncertainty of experimental measurements cannot be considered. The Flux Balance Analysis could be used instead, but the assumption of optimal behaviour of the organism brings other difficulties.

Results: We propose a procedure to estimate the evolution of the metabolic fluxes that is structured as follows: 1) measure the concentrations of extracellular species and biomass, 2) convert this data to measured fluxes and 3) estimate the non-measured fluxes using the Flux Spectrum Approach, a variant of Metabolic Flux Analysis that overcomes the difficulties mentioned above without assuming optimal behaviour. We apply the procedure to a real problem taken from the literature: estimate the metabolic fluxes during a cultivation of CHO cells in batch mode. We show that it provides a reliable and rich estimation of the non-measured fluxes, thanks to considering measurements uncertainty and reversibility constraints. We also demonstrate that this procedure can estimate the non-measured fluxes even when there is a lack of measurable species. In addition, it offers a new method to deal with inconsistency.

Conclusion: This work introduces a procedure to estimate time-varying metabolic fluxes that copes with the insufficiency of measured species and with its intrinsic uncertainty. The procedure can be used as an off-line analysis of previously collected data, providing an insight into the dynamic behaviour of the organism. It can be also profitable to the on-line monitoring of a running process, mitigating the traditional lack of reliable on-line sensors in industrial environments.

Background

Fostered by the importance of studying the cell metabolism under a system-level approach [1,2], the set of metabolic pathways of organisms of interest are assembled in metabolic networks [3,4]. If it is assumed that the intracellular metabolites of a network are at pseudo steady-

state, mass balances around each metabolite can be described by means of a homogeneous system of linear equations [5]. These equations can be considered as stoichiometric constraints. Then, the constraints imposed by enzyme or transport capacities and thermodynamics (e.g. irreversibility of reactions) can be incorporated to the system [6]. Thereby the imposed constraints define a space where every feasible flux distribution lives [7]. Since the metabolic phenotype can be defined in terms of flux distributions through a metabolic network, this space represents (or at least contains) the set of feasible phenotypes [8]. The environmental conditions given at a certain time instant will determine which of these flux distributions corresponds to the actual one [9].

Coupling constraints with experimental measurements

Experimental measurements of fluxes can be incorporated as constraints, in order to determine the actual flux distribution or at least to reduce the space of possible flux distributions. However, it must be taken into account that measurements are not invariant constraints, but specific condition constraints [8]. There are several methodologies that use this approach with different purposes: estimate the non-measured fluxes, predict flux distributions, investigate the cell behaviour or monitor bioprocesses.

Metabolic Flux Analysis (MFA) provides a methodology to uniquely determine the actual flux distribution by using a metabolic network and a set of measured fluxes [5]. It has been intensively used in recent years with successful results [10-13]. As it can only consider stoichiometric constraints, a considerable number of fluxes need to be measured to determine the rest of the fluxes. Unfortunately, the available measurements are often insufficient [14].

The Flux Balance Analysis (FBA) can be used to predict metabolic flux distributions [15,16]. Firstly, a constraint-based model is defined as a set of invariant constraints: stoichiometrics, thermodynamics, etc. Then, only a few specific condition constraints (usually substrates uptakes) are imposed. Subject to these constraints, which define a region of possible flux distributions, an optimal flux distribution is calculated using linear programming. Yet, the optimal solution may not correspond to the actual flux distribution. It must be hypothesized that i) the cell has identified the optimal solution, ii) the objective sought by the cell is known, and iii) it can be mathematically expressed. However, FBA predictions based on different objective functions (e.g. maximize growth) are consistent with experimental data [17-19].

Estimating the evolution over time of flux distributions

Typically, calculation of a flux distribution (e.g. with MFA or FBA) is done under a static point of view: the measured

fluxes are assumed to be constant. That means that the obtained flux distribution will only be valid during a certain period of time, while the environmental conditions and the cell state remain steady (e.g. during the growth phase). However, if these conditions change along time, as it happens in an actual culture, the flux distribution will change. The estimation of the flux distribution over time can be useful to investigate the dynamic behaviour of the microorganism or to monitor the progress of industrial fermentations [20]. In [21], the classical FBA is extended to predict the dynamic evolution of flux distributions. In [22], an approach based on elementary modes and the assumption of optimal behaviour is used to estimate the flux distributions of *Corynebacterium glutamicum* at different temporal phases of fermentation. Elementary modes are also employed in [23], where the cell life is decomposed in a succession of phases, and then the time-varying intracellular fluxes are obtained by switching the flux distributions calculated at each phase. In [24], on-line MFA is successfully applied to quantify coupled intracellular fluxes. Takiguchi et al. [25] use a similar approach to recognize the physiological state of the cells culture. They also show how this information can be used to improve Lysine production yield. Very recently [26] has presented an on-line estimation of intracellular fluxes applying MFA to an over-determined metabolic network.

To calculate the succession of flux distributions, it is usually assumed that intracellular fluxes are in quasi-steady state within each measurement step. However, that does not mean that the intrinsic dynamic nature of the cultivation is being disregarded. Instead, the intracellular fluxes will follow the change of environmental conditions as mediated by the measured fluxes (e.g. substrate uptakes). Hence, steady states may undergo shifting from one state to another depending on the evolution of the measured fluxes [27]. Such assumption has been successfully applied in the works cited in above and in the development of several dynamic models [23,28-32]. This approach makes it possible to study the dynamic behaviour of the organism, without considering the still not well-known intracellular kinetics.

Using the flux spectrum approach to estimate the fluxes

MFA can be successively applied to estimate the evolution of a flux distribution over time. However, this approach has three main difficulties: *i) It cannot be used when measurements are scant* (i.e. when the system is underdetermined). This happens very often due to the lack of measurable fluxes. *ii) The uncertainty of the measured fluxes cannot be considered.* Not only gross errors may appear - which could be managed only in case there are redundant measured fluxes- but also most sources of measurements are intrinsically uncertain and the propagation of this uncertainty to the estimated fluxes is not controlled, and

iii) only equalities can be used as constraints. For instance, reversibility constraints or maximum flux values cannot be taken into account. FBA solves the first difficulty and provides a framework to deal with the other ones. But the use of FBA in this context could be problematic due to the appearance of a time-variant metabolic objective [22]. For these reasons, the procedure introduced in this work uses the Flux Spectrum Approach (FSA) [33]. It is a variant of MFA that includes some characteristics of FBA (e.g. it is not restricted to stoichiometric constraints) and provides some additional benefits (e.g. it allows to consider measurements uncertainty). The use of FSA will make it possible to face the difficulties described above without assuming an optimal behaviour of the organism.

Although FSA is capable of considering a wide range of constraints, in this work we will only use stoichiometric relationships and simple thermodynamic constraints (reactions directions), and we assume them to be known a priori. However, it must be noticed that the incorporation of thermodynamic constraints -based on measurements or estimations of the standard Gibbs free energy change of reactions- is capturing attention in recent times. A genome-scale thermodynamic analysis of *Escherichia coli* has been recently carried out [34]. Kümmel et al. have introduced an algorithm that -based on thermodynamics, network topology and heuristic rules- automatically assigns reaction directions in metabolic models such that the reaction network is thermodynamically feasible [35]. Interestingly, the reaction directions obtained can be incorporated as constraints before using FSA. Standard Gibbs free energy changes have been also used to incorporate thermodynamic realizability as constraint for FBA [36] -or in an analogous manner to FSA-, and to develop

a new form of MFA with the capability of generating thermodynamically feasible fluxes [37].

The objectives of this article are twofold: first, introduce a procedure for the estimation of the metabolic fluxes over time by using a metabolic network as a constraint-based model and a reduced set of measurable species. This procedure is capable of coping with lack of measured species and with its intrinsic uncertainty, thanks to the use of the Flux Spectrum Approach (FSA). Second, illustrate this procedure with a real example: the estimation of non-measured fluxes during a cultivation of CHO cells in batch mode in stirred flasks.

Results and discussion
Procedure overview
In most cases, only a few extracellular species are measurable during fermentation processes. This is the reason for use an indirect approach to estimate the fluxes that cannot be measured: couple the available measurements with known biological constraints. Under this philosophy, the proposed procedure is structured as follows (Figure 1): 1) obtain experimental measurements of the concentration of some extracellular species and biomass, 2) convert these concentrations to *measured fluxes* and 3) estimate the non-measured fluxes using the Flux Spectrum Approach (FSA).

It is sometimes overlooked that extracellular fluxes are not directly measured. Instead, the concentrations of a set of species are measured (step 1), and those data are converted to flux units or measured fluxes (step 2). The importance of a good conversion should not be disregarded: error in the measurements of concentrations may
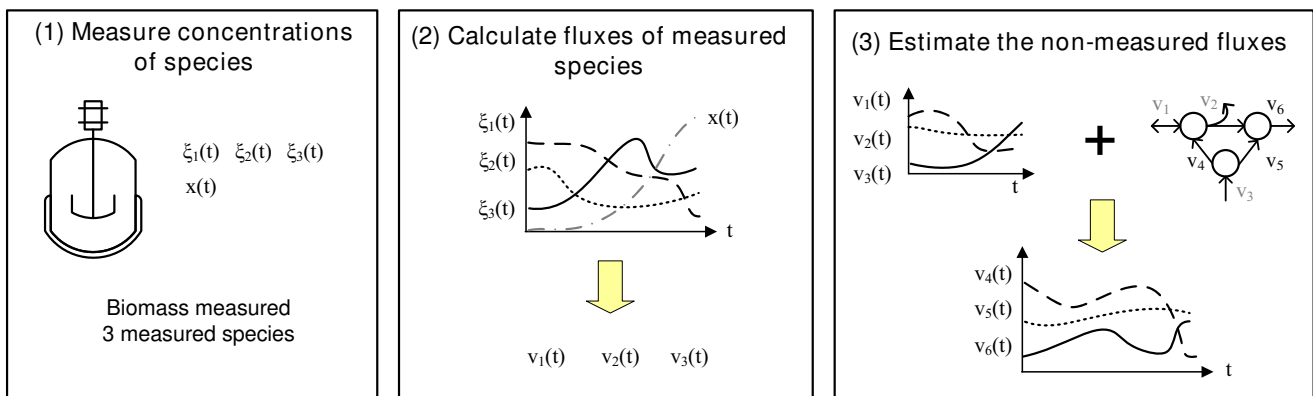

Figure 1
Procedure overview. Step 1: get experimental measurements of concentration of some extracellular species and biomass. Step 2: convert this concentrations to measured fluxes. Step 3: estimate the non-measured fluxes by using the Flux Spectrum Approach (FSA). $\xi(t)$ is the concentration of an extracellular specie and $v(t)$ its flux. $x(t)$ is the biomass concentration. Subindexes 1, 2 and 3 denote the measured fluxes and 4, 5 and 6 the non-measured ones.

be amplified through the conversion, incorporated into the measured fluxes, and then propagated to the estimation of the non-measured fluxes. To minimize this hitch, the conversion should be done carefully. Afterwards, the non-measured fluxes can be estimated by coupling the metabolic network and the measured fluxes (step 3). This has been done before by means of the MFA methodology [24-26]. Yet, this approach has certain limitations. We will overcome some of them using FSA.

It must be remarked that the procedure can be used in two main scenarios: as an off-line analysis of previously collected data or as an on-line monitoring of an industrial process. The structure of the procedure and its fundamental step (step 3) are exactly the same in both cases. Nevertheless, there are several differences concerning step 2. These differences will be briefly described along the article and illustrated in an additional file [additional file 3].

Preliminaries: choice and analysis of the metabolic network

A metabolic network can be represented with a stoichiometric matrix S, where rows correspond to the m metabolites and columns to the n fluxes. Assuming that the intracellular metabolites are at pseudo-steady state, material balances around them can be formulated as follows [38,39]:

$$S \cdot \mathbf{v} = 0 \qquad (1)$$

where $\mathbf{v}$ is a flux distribution. Assuming that S has full row rank, the number of independent equations is m. As typically n becomes larger than m, the system (1) is underdetermined (n-m degrees of freedom). That means that there is not a unique flux distribution fulfilling (1), but an infinite number of feasible flux distributions. In order to determine which of these feasible flux distributions is the current one, the constraints imposed by the measured fluxes will be incorporated -latter on it will be shown that other constraints, for example the reversibility constraints, can be added.

Thereby, when choosing the metabolic network to be used through the procedure, it must be taken into account that its degree of detail needs to be compatible with the number of available measurements -i.e. the available measurements must be sufficient to offset the underdeterminacy of the network. In order to study this, we can analyze the system formed by the stoichiometric constraints given by (1) and the constraints imposed by the measured fluxes. This system -which constitutes the fundamental equation of MFA- can be obtained making a partition between measured (subindex m) and non-measured or unknown fluxes (subindex u):

$$S_u \cdot \mathbf{v_u} = -S_m \cdot \boldsymbol{v}_{\mathbf{m}} \qquad (2)$$

System Determinacy and Calculability of Fluxes

System (2) is determined when there are enough linearly independent constraints for uniquely calculate all non-measured fluxes $\mathbf{v_u}$; i.e., when $rank(S_u) = u$ (u is the number of non-measured fluxes). On the contrary, when $rank(S_u)>u$, the system is classified as underdetermined because at least one non-measured flux, and probably most of them, is non calculable [14]. If the system is underdetermined, the traditional MFA methodology cannot be used to calculate the non-measured fluxes. Fortunately, the use of FSA may provide an estimation of the non-measured fluxes even in this situation. However, it must be taken into account that the likelihood of obtaining a precise estimation increases as the underdeterminancy reduces, as the set of flux distributions compatible with the measured values will be smaller.

System Redundancy and Consistency of Measurements

System (2) is redundant when some rows in S_u can be expressed as linear combinations of other rows; i.e., when $rank(S_u)<m$. This can lead to an inconsistent system if the vector $\mathbf{v_m}$ contains such values that no $\mathbf{v_u}$ exists that exactly solves (2). Therefore, when the system is redundant, the inconsistency of the measurements can be checked and its importance can be estimated (see methods). Unfortunately some measured fluxes have no impact on the consistency of the system, so they cannot be considered in the analysis of consistency. These fluxes are called non-balanceable. The balanceable fluxes can be detected as explained in [14], and they should be adjusted (or balanced) in case the system is inconsistent (see methods). All these methods are commonly applied when MFA is used [12,24,26]. They can also be used within our procedure, but in addition the use of FSA provides new methods to deal with inconsistency as it will be shown in a subsequent section.

Step 1: Getting experimental measurements of species

There are several alternatives to measure the concentration of species -e.g. on-line sensors, isotopic tracer experiments or laboratory procedures- but providing a detailed description of each one is out of the scope of this work. In any case, it must be remembered that the more measurements are available, the more non-measured fluxes may be accurately estimated. However, it is necessary to be prepared to overcome a lack of measurements, especially when the procedure is done on-line (due to the lack of reliable on-line sensors).

Step 2: conversion of measured concentrations in measured fluxes

A mass balance around each extracellular species whose concentration is measurable can be stated as:

$$\frac{d\xi}{dt} = v_\xi \cdot X - D \cdot \xi + F_\xi \qquad (3)$$

where ξ is the specie concentration, v_ξ its flux (substrate uptake or product formation), X the biomass concentration, D the dilution term and F_ξ the net exchange of the specie with the outside. Notice that this equation is only valid for extracellular species; however, the biomass growth and the mass balance around an internal metabolite not assumed to be at pseudo-steady state can be represented in a similar way [40,41].

It is possible to calculate v_ξ as a function of ξ, X, D, F_ξ and $d\xi/dt$. But this presents two main difficulties: i) approximate a derivative (directly or indirectly) and ii) deal with the presence of errors and noise in the measurements of the concentration ξ [42]. The underlying problem is how precision can be combined with robustness with respect to measurement errors. The most straightforward approach is to approximate the derivative with a simple method (e.g. Euler or Runge-Kutta methods) and then solve (3) [42]. Very often this straight approximation needs to be combined with the use of filters to eliminate - or at least to reduce- the presence of noise. This approach provides very good results when centred methods can be used to approximate the derivative and to filter the resultant signal, i.e. when the whole procedure is done off-line, or when it is done on-line but certain delay in the calculation of the fluxes is allowable (i.e. when past, *k-i*, and future information, *k+i*, is available for the calculation of $v_\xi(k)$). Furthermore, there are methods especially aimed to the on-line approximation of the derivative. If the noise signal is well characterized (e.g. the frequency band or a stochastic feature is known) a linear differentiator [43] or even a Luenberger observer may be used [44]. If nothing is known on the structure of the signal, then sliding mode techniques are profitable. For example, the method intro-

duced in [45] combines exact differentiation for a large class of input signals with robustness with respect to any small noises. Finally, there are other approaches to calculate the extracellular fluxes that avoid the approximation of the derivative, as for example the use of extended Kalman filters [26,46] or the observers based on concepts from nonlinear systems theory, such as the high gain estimators described in [40,47]. These methods do not use future information because they are aimed to the on-line operation mode.

The importance of the use of filters should be remarked: not only the signal of measured concentrations should be filtered to reduce its noise, but also the calculated extracellular fluxes may be filtered to get a smooth signal. Filters based on the moving average will be used in this work since they are simple and versatile. Basically, the filtered value at time *k* is calculated by averaging the values of the original signal within a time window. There are several versions that differ in the time window used (backward or centred) and in the distribution of weight over the averaged values (uniform or exponential). Interestingly, this kind of filters has already been successfully applied to the calculation of metabolic fluxes [42].

To provide a complete description of our procedure, two conversion approaches are described in the methods section: the combination of an Euler method with a moving average filter, and the use of a nonlinear observer (see Figure 2). The first one is especially suitable when the procedure is done off-line, while the second one is aimed to work on-line. Nevertheless, it must be taken into account that there is not a universal solution for the conversion problem. In real applications, the particularities of the concentrations measurements (accuracy, sample rate, importance and characteristics of the noise, etc.) and the operation mode (off-line, on-line with an acceptable delay or purely on-line) will determine which method is

Figure 2
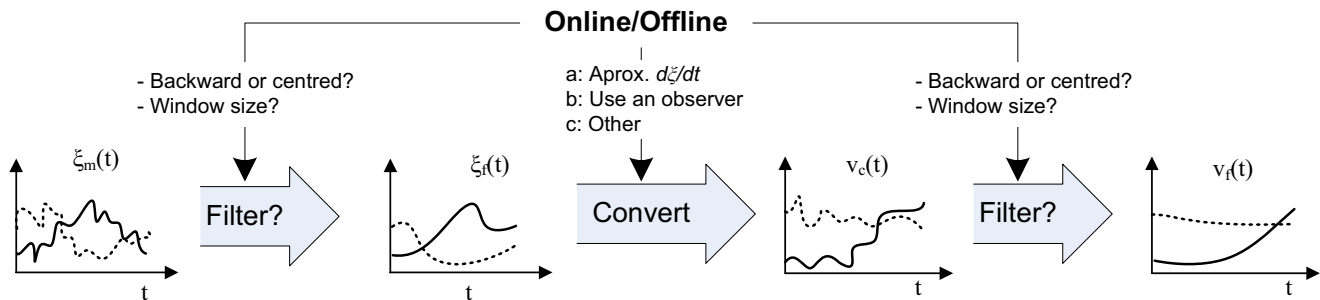**Conversion of measured concentrations to measured fluxes**. First, the measured concentrations should be filtered. Then, fluxes are calculated from the concentration data (e.g. approximating the derivative or using a dynamic observer). Finally, the calculated fluxes may be filtered to get a smooth signal. Each step is conditioned by the operation mode (on-line or off-line).

the most suitable one. A real off-line conversion is described below, but the most illustrative example of the step 2 is given in the Additional File 3, which addresses the on-line and the off-line operation modes and the use of filters. A practical guide about step 2 is also given in the mentioned file.

Step 3: estimation of the non-measured fluxes using FSA

Finally, the measured fluxes obtained in step 2 are coupled with known biological constraints in order to estimate the non-measured fluxes (Figure 1). Basically, this implies that a solution for system (2) has to be found at each time instant k. Traditionally MFA was successively applied with this purpose. Unfortunately, as mentioned in the background section, this has some limitations - which become especially critical if the procedure is done on-line, due to the traditional lack of reliable on-line sensors. To overcome them, the Flux Spectrum Approach (FSA) will be used instead in the third step of our procedure.

Using FSA, the estimation of the non-measured fluxes at each time instant k is obtained as follows [33]: 3.1) impose the set of constraints given by (2) and the reversibility constraints. They define a region where the actual fluxes may live. 3.2) calculate the interval of possible values for each non-measured flux by solving two linear programming problems, one to compute its maximum value within the region and the other one to compute its minimum (details are given in the methods section). Thus, at each time k, and for each non-measured flux, an interval bracketing its possible values will be obtained: $v_{uj}(k) = [v_{uj, min}, v_{uj, max}]$. The size of the intervals (i.e. the imprecision of the estimation) depends on the number of non-measured fluxes, the irreversible reactions, the available measurements and the degree of uncertainty considered. Of course, the more constraints are available, the tighter intervals are obtained. If uncertainty is not considered,

reversibility constraints are not used, and the system (2) is determined, FSA gives the same unique solution as MFA [33]. But in addition, the use of FSA provides several advantages to the estimation procedure (see Table 1):

• It makes it possible to consider the uncertainty of experimental measurements and even qualitative knowledge (e.g. maximum values of certain fluxes). Hence, if measurements uncertainty is indeed present and it is well characterized, the estimation of non-measured fluxes will be more reliable (Figure 3E). FSA provides not only a prediction of the fluxes, but also an indication of the reliability of this prediction.

• It considers the reversibility constraints of certain fluxes. This provides an estimation of the non-measured fluxes even when measurements are insufficient, i.e. when (2) is underdetermined (Figure 3A). This estimation will be precise if the degree of underdeterminancy is limited and there are irreversible fluxes. On the contrary, the estimation could be poor and some intervals may be unbounded. The reversibility constraints will also restrict the intervals of the estimated fluxes when uncertainty is considered (Figure 3C). Finally, the reversibility constraints can also provide a means to detect inconsistencies even when the system is not redundant (Figure 3F).

• It provides a straight method for coping with inconsistency: a band of uncertainty is used instead of adjusting the inconsistent measurements. As any inconsistent set of measurements is necessarily uncertain, it seems reasonably to define a band of uncertainty around the measured values trying to enclose nearby consistent sets of measurements. Thus, every consistent set of measurements enclosed by the band will be taken into account in the estimation of the non-measured fluxes (Figure 3B). Furthermore, the band size needed to find the nearest consist-

Table 1: Comparison between MFA and FSA

|	Traditional Metabolic Flux Analysis (MFA)		Flux Spectrum Approach (FSA)	
	Flux estimation	*Consistency*	*Flux estimation*	*Consistency*
Determined *Redundant*	Yes.	Consistency check χ^2. Flux adjustment.	Yes. Considers uncertainty. Detects sensitivity problems.	Consistency check χ^2. Flux adjustment or use of a band of uncertainty.
Determined *Not Redundant*	Yes.	No. No.	Yes. Considers uncertainty. Detects sensitivity problems.	Detect some inconsistencies.
Underdetermined *Redundant*	No.	Consistency check χ^2. Flux adjustment.	Yes (not guaranteed). Considers uncertainty. Detects sensitivity problems.	Consistency check χ^2. Flux adjustment or use of a band of uncertainty.
Underdetermined *Not Redundant*	No.	No. No.	Yes (not guaranteed). Considers uncertainty. Detects sensitivity problems.	Detect some inconsistencies.
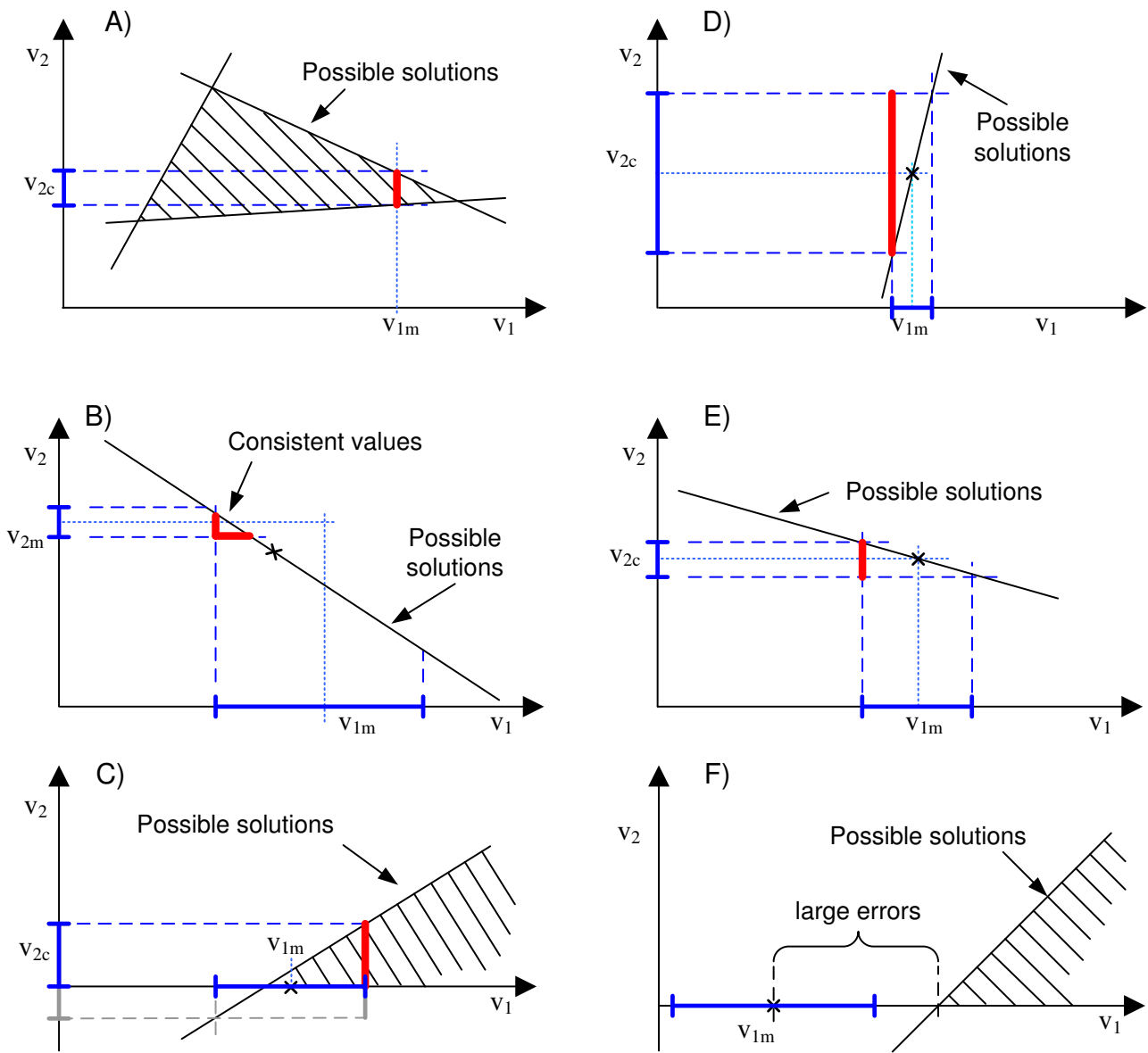
Figure 3
Flux spectrum approach in use. Each figure shows a schematic projection of a high-dimensional flux space into two dimensions. The space of possible solutions (before taken into account v_1 and v_2) is represented by a polygon or a solid black line and tagged with a label. Subindex m denotes a measured flux, and c a calculated one. The band of uncertainty around measured fluxes is represented with a blue, solid interval in the axis. The estimations provided by FSA are represented with red, thick lines. Dotted lines are just auxiliary projections. (A) Underdetermined case. The interval of possible values for v_2 is computed even when the system is underdetermined. (B) Determined and redundant case. Both fluxes are measured, but its values are inconsistent. With the band of uncertainty used, all the values of v_1 and v_2 within the red line are considered valid. Due to the shape of the band, the values given by a least squares adjustment (denoted with an x) are not considered as a valid solution. (C) Reversibility constraints. A reversibility constraint reduces the interval of possible values for v_2. (D) Detection of sensitivity problems. The strangely large interval of possible values for v_2 may indicate a problem of sensitivity. (E) Representation of uncertainty. A non-measured flux is estimated from an uncertain measurement. (F) Detection of large errors. A large error in the measured flux v_{1m} is detected with the reversibility constraint.

ent flux distribution gives an indication of the degree of inconsistency.

Additional advantages arise when FSA is used in a successive way to estimate the temporal evolution of the metabolic fluxes:

• It may detect sensitivity problems. Assume that a band of uncertainty is being used and that the measured fluxes change smoothly over time. If the interval of values for an estimated flux is strangely large at a certain instant k, it indicates that a slight change in the measured fluxes has a big effect over the estimated flux, i.e. that a sensitivity problem exists (Figure 3D). Thereby, an analysis of sensitivity is incorporated in the estimation procedure.

• The peak values at certain time instants k -which may appear when MFA is used- are avoided with FSA. These peaks are consequence of slight errors in the measurements (which are common due to the lack of reliable sources of measurements and due to the uncertainty of the conversion of concentration data into measured fluxes). Since FSA considers a band of uncertainty around the measured values, it avoids, or at least reduces, this phenomenon.

• The estimation given by FSA for a certain flux at time k (an interval of possible values), combined with the inspection of past and future estimations and with our qualitative knowledge about cell behaviour, may be used to hypothesize which of the possible temporal evolutions corresponds to the actual one. That is to say, the richness of the estimation given by FSA makes it possible to exploit our qualitative knowledge to support certain hypothesis without being confused by measurements uncertainty.

Application: estimation of the fluxes during a cultivation of CHO cells

The three-step procedure described in the previous section is now applied to a real problem taken from the literature: the estimation of the intracellular fluxes of CHO cells cultivated in batch mode in stirred flasks. The available experimental data are the typical data measured off-line (accurate measurements of the concentration of a few species but with a low sample rate), and therefore this example will be approached assuming that the procedure is done off-line. This assumption is important during the second step of the procedure, and for this reason an example has been included in the Additional File 3 that illustrates the differences between the on-line and the off-line operation modes. However, hereinafter we will pay special attention to the third step of the procedure because it is the most important one. In particular, the benefits provided by the use of FSA will be compared with those obtained with the well-established MFA methodology,

which is the basis of most of the similar procedures [24-26]. This comparison illustrates the advantages of the new estimation procedure in three different scenarios:

S1. When measurements are *almost* sufficient. The number of measured fluxes is almost sufficient when there are enough to determine all the non-measured fluxes but there are not redundant measurements (i.e. when the system (2) is determined and not redundant).

S2. When measurements are sufficient, i.e. when the measured fluxes are enough to determine the non-measured fluxes and there are also redundant measurements (the system (2) is determined and redundant).

S3. When measurements are insufficient. The number of measured fluxes is insufficient when there are not enough to determine all the non-measured fluxes (i.e. when the system (2) is underdetermined and not redundant).

For completeness, the most uncommon case (when the system is underdetermined but redundant) is illustrated with a toy example in an appendix [Additional File 2]. In the three scenarios, the intrinsic uncertainty of the measured fluxes is taken into account.

Metabolic network of CHO Cells

The metabolic network (Figure 4) has been taken from [48]. The network describes only the metabolism concerned with the two main energetic nutrients, glucose and glutamine. Thus, the metabolism of the amino-acids provided by the culture medium is not included. Four pathways are considered: the glycolysis, the glutaminolysis, the TCA cycle and the nucleotides synthesis. All reactions are assumed to carry flux only in only one direction, except reactions 2, 4, 5, 6 and 7 that are reversible (e.g. when glucose is exhausted lactate and alanine are consumed instead of produced). The complete lists of species and reactions are given in the Additional File 1.

The mass balance around intracellular metabolites at pseudo-steady state is given by eq. 1 (the stoichiometric matrix S is given in the Additional File 1). There are 12 metabolites (m) and 18 intracellular fluxes (n). Therefore, the system is underdetermined and has 6 degrees of freedom. The extracellular fluxes v_G, v_L and v_A coincide with the fluxes $-v_1$, v_6 and v_7. Three equations that link v_{NH4}, v_Q and v_{CO2} with the intracellular fluxes can be obtained by inspection of the metabolic network. Moreover, it is a natural assumption to consider that the formation of purine and pyrimidine nucleotides is the same. As a result, four equations are incorporated by the authors [48]:
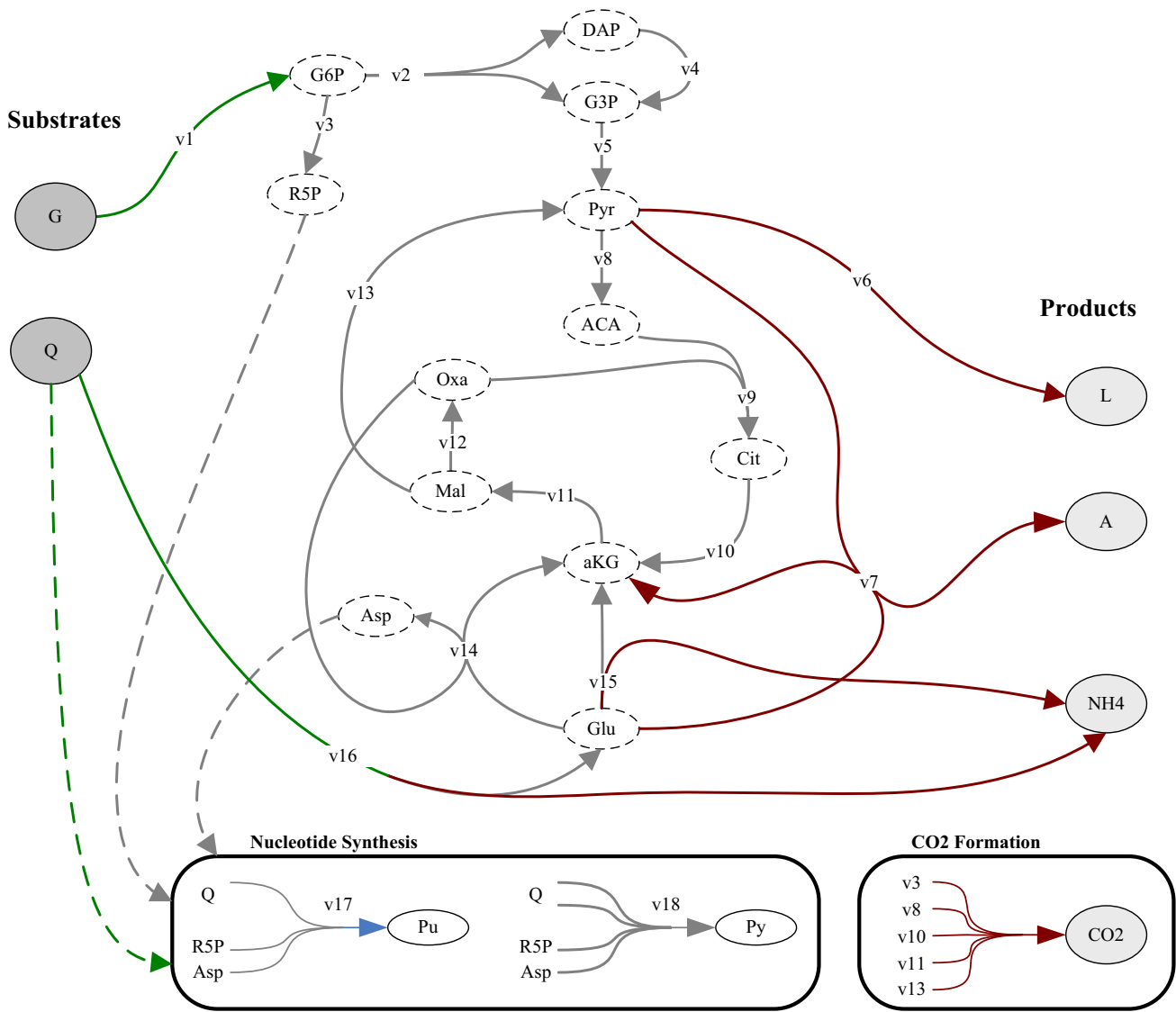
Figure 4
Metabolic Network of CHO cells. Extracted from [41]. Initial substrates (dark grey ovals), extracellular products (light grey ovals), terminal internal metabolites (white ovals) and internal metabolites (white ovals with dashed line). The CO_2 formation and the nucleotide synthesis are described separately. The nomenclature is given in the additional file 1.

$$v_{NH4} : v_{19} = v_{15} + v_{16}$$
$$-v_Q : v_{20} = v_{16} + v_{17} + 2 \cdot v_{18}$$
$$v_{CO2} : v_{21} = v_3 + v_8 + v_{10} + v_{11} + v_{13} \quad (4)$$
$$v_{Extra} : v_{22} = 0 = v_{17} - v_{18}$$

These constraints can be represented with a *4×18* matrix S_ξ fulfilling (11). Then, (11) and (1) can be joined to define an extended homogeneous system of linear equations (see methods). The extended system has 16 metabolites (*mx*) and 22 reactions (*nx*).

The mathematical model, formed by the stoichiometric matrixes S and S_ξ, is given in a Matlab script and a standard SBML file [see Additional File 4].

Step 1: getting experimental measurements of species
The experimental data taken from [28] is given in Figure 5. The cell density (X) and the concentration of 5 extracellular species are measured; two substrates, glucose (G) and glutamine (Q), and three excreted products, lactate (L), alanine (A) and ammonia (NH4). This data was collected with a sample rate of 24 h. These measurements cannot be filtered because -due to the low sample rate- it
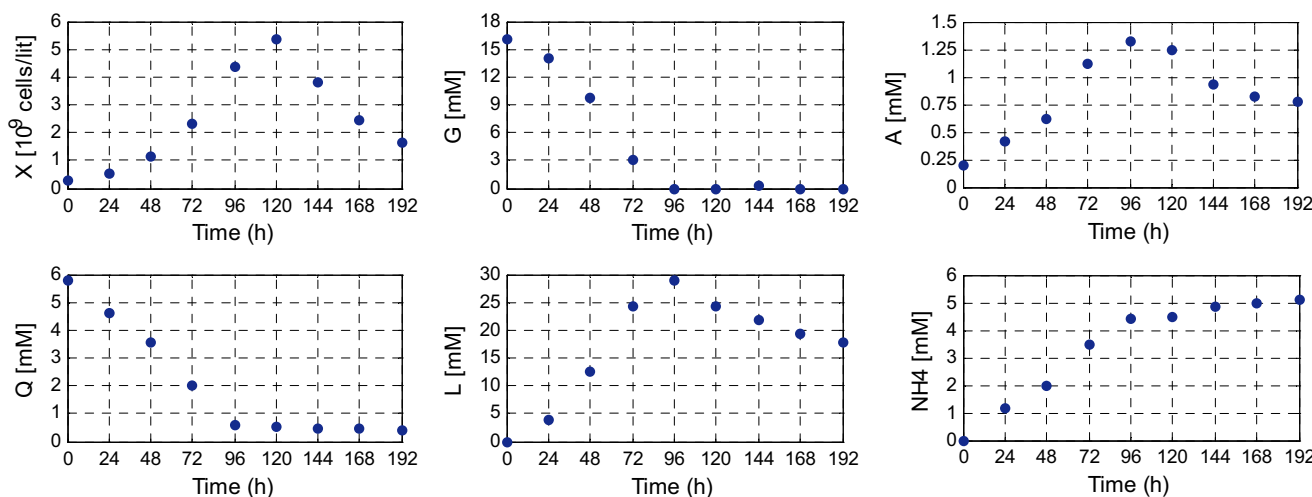
Figure 5
Concentration of measured extracellular species and biomass during a cultivation of CHO cells. The measurements correspond to cell density (X), glucose (G), glutamine (Q), lactate (L), alanine (A) and ammonia (NH4).

is impossible to distinguish between noise and true changes of the signal.

Step 2: conversion of measured concentrations in measured fluxes
The second step of the procedure is the conversion of the measured concentrations in measured fluxes. The measured fluxes (and the biomass growth) calculated with

three different approximations of the derivative are depicted in Figure 6 (see methods). Since the procedure is being done off-line, a centred approximation is the most advisable choice. Therefore, the fluxes calculated with the middle point Euler approximation will be used hereinafter. We obtained similar results (not shown) when the complete example was done using a backward Euler
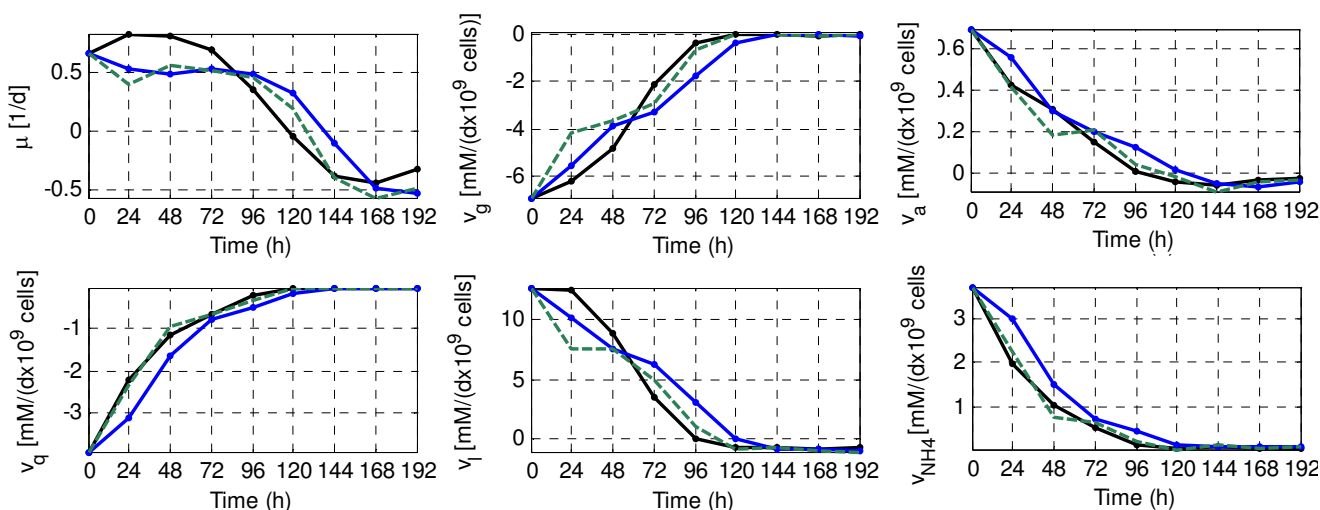
Figure 6
Extracellular fluxes and growth rate calculated from the measured concentrations. μ is the biomass growth rate, v_g, the flux of glucose, v_Q the flux of glutamine, v_L the flux of lactate, v_A the flux of alanine and v_{NH4} the flux of ammonia. Fluxes are calculated with the middle point Euler approximation (black solid line) and the backward Euler approximation (green dotted line). In addition, fluxes calculated with the backward Euler approximation and filtered with a standard moving average of order 2 are also depicted (blue solid line).

approximation (which would be more suitable in case the procedure were done on-line). It is also remarkable that Figure 6 already gives the idea of uncertainty -differences between the conversions obtained with different methods are significant. In fact, the different conversions, along with the precision of the sensors and the protocols used to measure the concentration of species, could be used to characterize the uncertainty in the measured fluxes.

Step 3 (S1): estimation of fluxes if measurements are almost sufficient and uncertain
If the five measured fluxes are used (v_1 (G), v_6 (L), v_7 (A), v_{19} (NH_4) and v_{20} (Q)) and it is assumed that the formation of purine and pyrimidine nucleotides is the same (v_{22} = 0), the rank of S_u (16) is equal to the number of unknown fluxes (22-5-1). Thereby the system (2) is determined but not redundant. In this case we could use MFA to determine the non-measured fluxes. More precisely, at each time instant k, the unique flux distribution fulfilling (2) can be obtained by using the inverse matrix of S_m (see methods). However, as it can be observed in Figure 7 (green solid line) the results obtained are not very satisfactory:

• The estimated values at time 24 h an 168 h for fluxes v_8, v_9, v_{10}, v_{11}, v_{12} and v_{21} seem unreasonable: the measured fluxes evolve in a smooth way, but these fluxes show peak values.

• The estimated fluxes v_8, v_9 and v_{10} do not fulfil the reversibility constraints (they are not considered by MFA).

• MFA assumes that there is not any kind of error in the measurements, which is unlikely, and therefore the estimated fluxes are unreliable.

A new estimation has been done at time 24 h, where the measured values for fluxes v_1 and v_6 are slightly modified (+2% and -5% respectively). In a similar way, a new estimation at time 168 h assumes a slight variation of the measured values for v_1 and v_6 (-0.05 and +0.05 mM/(d · 10^9 · cells), respectively). As it can be observed in Figure 7 (red crosses), the peak values in fluxes v_8, v_9, v_{10}, v_{11}, v_{12} and v_{21} are eliminated or reduced, while the values of the rest of non-measured fluxes remain almost unchanged. This demonstrates that the peak values at times 24 h and 168 h could be caused by slight errors in the measured fluxes. The same issue is illustrated with figure A1 (Additional File 7). Hence, the main weakness of MFA in the determined case is pointed out: the effect of slight errors in the measured fluxes is not under control. These slight errors will exist in virtually all the measured fluxes (none sensor has a precision of 100%). Moreover, even the conversion of the measured concentrations into measured fluxes may introduce slight errors. For this rea-

son, the fluxes estimated with MFA are unreliable in this scenario.

The same scenario is now approached following the procedure introduced in this paper, i.e. using FSA instead of MFA in the third step. If uncertainty is not considered and all reactions are assumed to be reversible, FSA provides the same solution that MFA (results not shown). But it is possible to include the reversibility constraints for those reactions classified as irreversible. By using these constraints, FSA has detected a high inconsistency at 24 h and a lower one at 144 h (i.e. the region defined by the imposed constraints does not contain any solution at these time instants). It must be highlighted that the system is not redundant, so methods to check consistency based on redundancy cannot be used; however, FSA is detecting inconsistencies thanks to the reversibility constraints. Afterwards, it is also interesting to consider the intrinsic uncertainty of the measurements. We will define a band of uncertainty around the measured values, and then we will use FSA to estimate the non-measured fluxes. The most common ways to define a band of uncertainty are the use of a relative error around the measured values (e.g. of the 5%) and the use of an absolute one (e.g. 0.05 mM/(d•10^9•cells)). Herein, we use a mixed approach. For each measured flux v_m, at each time instant k, the band is defined as:

$$
\begin{aligned}
\text{If } relErr \cdot v_m > absErr &\quad \rightarrow \quad band = v_m \pm relErr \cdot v_m \\
\text{Else} &\quad \rightarrow \quad band = v_m \pm absErr
\end{aligned}
$$

$$(5)$$

With this expression the relative error (*relErr*) will be considered when the measured value is high, and the absolute one (*absErr*) when it is near to zero (see figure A2 in the Additional File 7). If more information about the measurements sources were available, the range of uncertainty of each measured flux could be defined accordingly. For example, if a commercial sensor is employed, its technical specifications can be used to define the band.

The non-measured fluxes estimated with FSA -when the band of uncertainty is considered and the reversibility constraints are incorporated- are shown in Figure 7 (black intervals). If they are compared with those obtained when MFA was used, several conclusions can be pointed out:

• The peaks at time 24 h an 168 h for fluxes v_8, v_9, v_{10}, v_{11}, v_{12} and v_{21} -which appeared when MFA was used- are avoided with FSA. As it was shown, when the measurements were slightly modified, these peak-values were replaced by more sensible predictions. Since these modified measurements are included in the band of uncertainty, the obtained intervals for v_8, v_9, v_{10}, v_{11}, v_{12} and v_{21}

Figure 7
FSA and MFA in the determined and not redundant case (S1). Known fluxes are: $v_1(G)$, $v_6(L)$, $v_7(A)$, v_{19} (NH_4), $v_{20}(Q)$ and v_{22}. The measured fluxes have a grey background and its band of uncertainty is represented with a black interval. The non-measured fluxes estimated with FSA are represented with a black interval, and the non-measured fluxes estimated with MFA with a green line. Two additional estimations with MFA are given at times 24 h and 168 h, where fluxes have been estimated from slight variations of the measured values for v_1 and v_6 (red x).

contain the sensible predictions. In principle, the peak-values would be within the intervals. However, a peak value could not satisfy the reversibility constraints and therefore it will not be considered a valid solution by FSA -this is the case at k = 24 h.

• The uncertainty of experimental measurements is non-trivially propagated to the non-measured fluxes. Hence, the use of FSA provides not only a prediction of the non-measured fluxes, but also an indication of the reliability of this prediction. For example, the predicted v_8, v_9 and v_{10} are highly influenced by measurements uncertainty, while v_2, v_4, and v_5 are quite insensitive. Although all fluxes can be determined, FSA highlights that the estimated values for v_8, v_9 and v_{10} are less reliable (or less precise) than those assigned to v_2, v_4 and v_5. This issue is more deeply analyzed in a subsequent section.

• Reversibility constraints provide a method to detect inconsistencies. For example, it can be easily checked that the solution provided by MFA do not satisfy the reversibility constraints at 24 h (a negative value is given to the irreversible fluxes v_8, v_9 and v_{10}). This inconsistency is detected and avoided with FSA.

• The underdeterminancy introduced as uncertainty in the measurements can be partially neutralized with the reversibility constraints. Hence, the estimated fluxes are more reliable but not necessarily highly imprecise.

This example shows that the procedure provides a reliable and rich estimation of the evolution along time of the non-measured fluxes when the measurements are only *almost* sufficient, i.e. when the system is determined but not redundant. In particular, the use of FSA in the third step of the procedure -instead of the well-established MFA- provides several benefits, thanks to taking into account the uncertainty of measurements and considering the reversibility constraints.

Step 3 (S2): estimation of fluxes if measurements are sufficient and uncertain

When the system (2) is determined and redundant, an estimation based on MFA will work as follows (approach 1): firstly, the importance of the inconsistency is checked and the measured flux values are adjusted; then, the pseudo-inverse matrix is used to estimate the non-measured fluxes. These two properties -checkable consistency and adjustable measurements- are responsible of the success of MFA in this scenario. However, FSA provides a new approach (approach 2) which holds the property of checkable consistency, but replaces the adjustment of the measurements by the definition of a band of uncertainty. We will apply both alternatives to our example.

The system (2) was determined and not redundant when six fluxes were known. If another independent flux is measured, the system will be redundant because the rank of S_u (15) will be less than m (16). Since no more fluxes were measured in [28], we will assume that the evolution of $v_{21}(CO2)$ has been measured -we chose it because it is a well-known extracellular flux. We assume that v_{21} evolves smoothly and that its values are within the intervals estimated with FSA in the previous section. Hence, at each time instant k, except 24 h and 168 h, the values given by MFA in the previous section are used as measured values (they lay within the intervals). The values at 24 h and 168 h are calculated by the approximation of a spline curve (see Figure A3 in the Additional File 7).

First of all, we apply the χ-square method to estimate the importance of the inconsistency at each time instant k (see methods). The data fails the consistency check at time 168 h, what indicates that the set of measurements contains gross errors at this point (see table A1 in the Additional File 7). Afterwards, we estimate the non-measured fluxes at each time instant k with the two approaches described above. In the first one, the measured values are adjusted to be consistent (as explained in methods). Then, the non-measured fluxes are estimated with MFA. In the second one, a band of uncertainty around the measured values is defined trying to enclose some nearby consistent sets of measured fluxes (the band is the same that in the previous section). Then, the non-measured fluxes are estimated with FSA. The results (shown in Figure 8) illustrate the benefits of using FSA in this scenario:

• All the consistent sets of measured values enclosed by the band of uncertainty are considered by FSA. That guarantees that the intervals obtained enclose the actual values of the fluxes if the band was correctly chosen. Contrarily, when MFA is used (approach 1), the actual values of the measured fluxes need to be *exactly* found to ensure that the estimations fit in with the actual fluxes. To illustrate this idea a consistent flux distribution within the band of uncertainty has been highlighted in Figure 8 (dotted line). This flux distribution corresponds to a set of measured values very near to the original ones; nevertheless the evolution of v_8, v_9, v_{10} and v_{12} is quite different to the estimation given by MFA. That proves that the values estimated with MFA may be deviated from the actual ones, even if there are only slight errors in the measured fluxes. Conversely, FSA shows that two qualitatively different interpretations of fluxes v_8, v_9 and v_{10} are possible: they can be stable around 0.6 or evolve from 0.2 to 0.7 mM/(d * 10^9 cells). If there were other evidences supporting one alternative over the other one, we could hypothesize which of these two scenarios corresponds to the actual one. Hence, FSA not only reduces the number of wrong predictions,
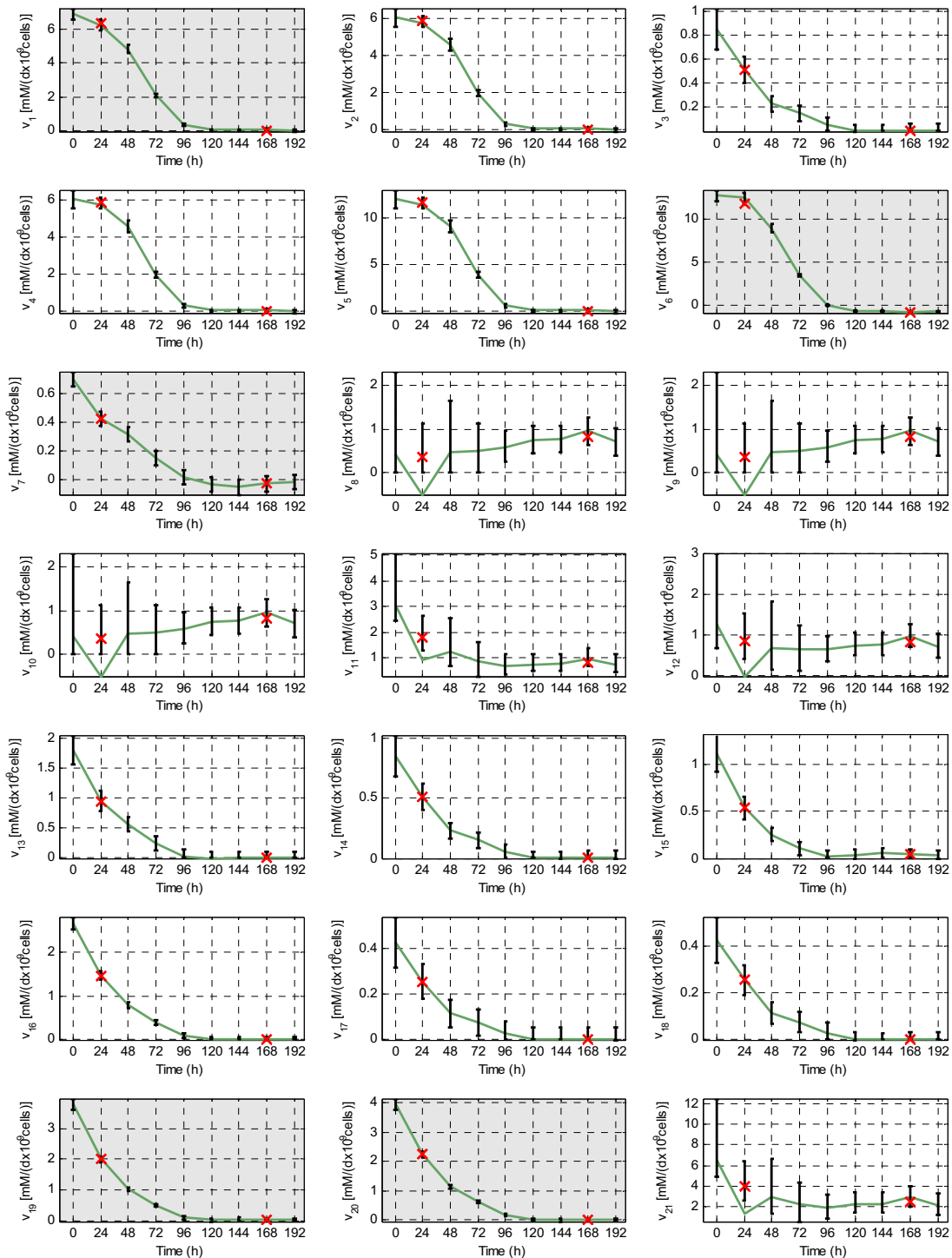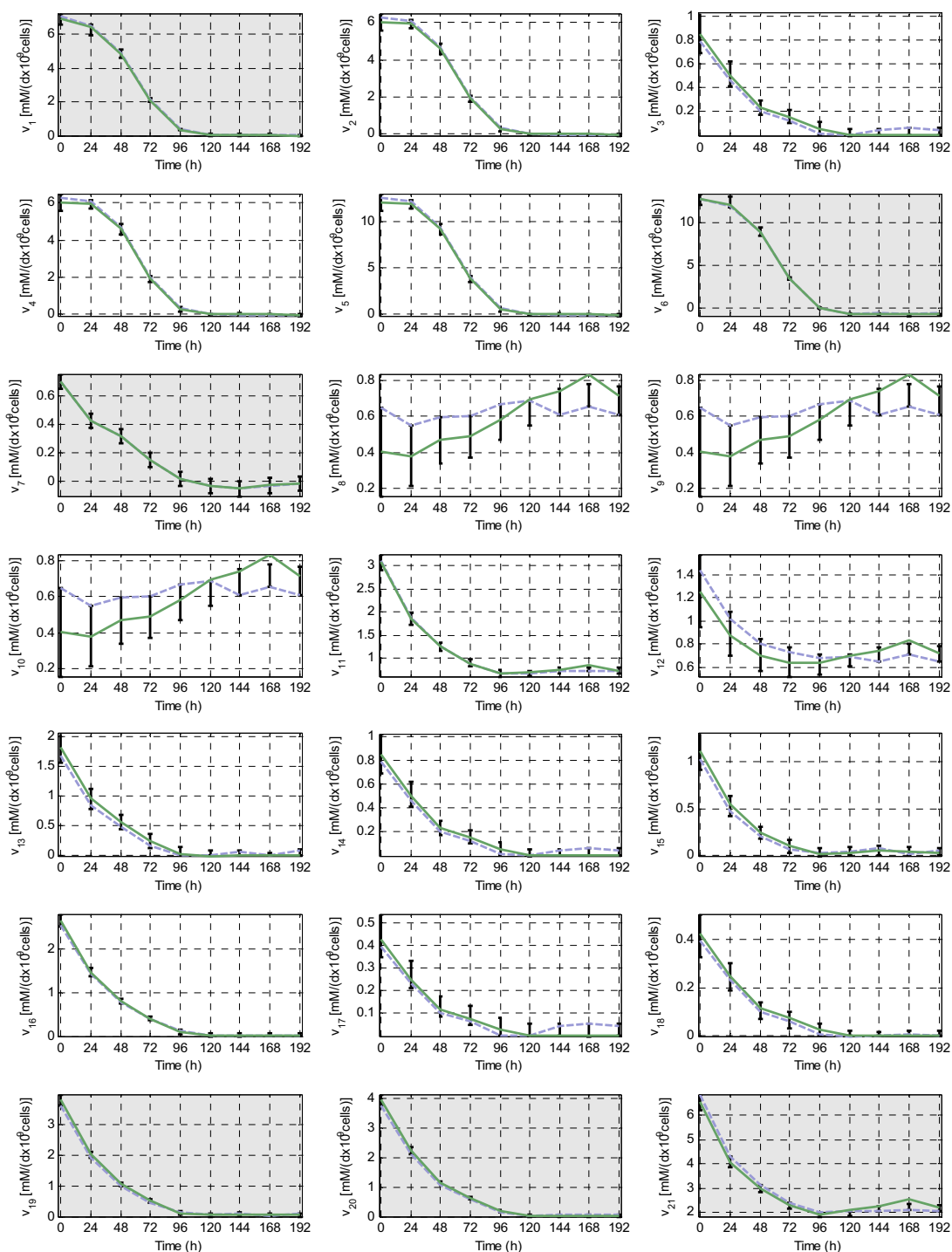
Figure 8
FSA and MFA in the determined and redundant case (S2). The known fluxes are: $v_1(G)$, $v_6(L)$, $v_7(A)$, $v_{19}(NH_4)$, $v_{20}(Q)$, $v_{21}(CO2)$ and v_{22}. The measured fluxes have a grey background and its band of uncertainty is represented with a black interval. The non-measured fluxes estimated by FSA are denoted with a black interval, and the non-measured fluxes estimated with MFA with a green line. One consistent flux distribution within the intervals given by FSA has been highlighted (blue dotted line) to show its discrepancy with the one calculated with MFA. Notice that this flux distribution corresponds to a set of measurements very close to the original ones (± 5% or ± 0.05 mM/(d•10^9•cells)).

but may also provide a quantitative support for our qualitative knowledge.

• Although there is a gross error in the measurements at 168 h, FSA finds at least one consistent set of measured values within the band of uncertainty (providing an error bound that complements the χ-square method). The estimations provided by FSA at 168 h seem sensible: the measurements are only slightly adjusted (the adjustment is limited by the band size) and the peak values are avoided. On the contrary, the fluxes estimated with MFA are very sensitive to the gross error. The value of v_{21} is significantly changed by the adjustment method resulting in a peak. Moreover, this insensible peak also appears in the estimated values of v_8, v_9, v_{10}, v_{11} and v_{12}. In fact, the fluxes calculated with MFA are generally discarded when the measurements fail the χ-square method.

• When FSA is used, the uncertainty of experimental measurements is non-trivially translated to the non-measured fluxes. Again, FSA provides not only a prediction of the non-measured fluxes, but also an indication of the reliability of this prediction.

This example has shown that the procedure can be useful to estimate the evolution of the fluxes even when measurements are sufficient but uncertain, i.e. when the system is determined and redundant. Although this is the scenarios were the procedures based on the use of MFA are most successful, the use of FSA provides a more reliable estimation of the non-measured fluxes and offers an interesting approach to cope with inconsistency.

Step 3 (S3): estimation of fluxes if measurements are insufficient and uncertain
Finally, it will be shown that our procedure can be used even when the available measurements are insufficient (i.e. when system (2) is underdetermined). In this situation procedures based on MFA cannot be applied, but the use of FSA allows our procedure to estimate the interval of possible values for each non-measured flux. In particular, the non-measured fluxes will be estimated by using different sets of 5 and 4 measured fluxes -remember that 6 were necessary to get a determined system. In all cases, uncertainty has been considered using the band described above. All results are given in Table 2 and two illustrative cases are depicted in Figure 9.

With four sets of 5 measurements (G, F, E and C) the evolution over time of all the non-measured fluxes can be estimated. Case G, where v_{22} is not known, provides the best results. There is a mean interval increment of 39% with respect to the determined case and the increment is minor than 25% for 12 fluxes (out of 17). This case is depicted in Figure 9 (in green). The intervals are practi-

cally the same than in the determined case for most fluxes (v_2, v_4, v_5, v_8, v_9, v_{10}, v_{11}, v_{12}, v_{13}, v_{15} and v_{21}). Intervals for v_3 and v_{14} are larger, but still accurate, and only the estimations of v_{16}, v_{17} and v_{18} are highly imprecise. Moreover, the temporal evolution -that can be characterized by using the middle point of the interval- is almost the same than of the determined case even for these fluxes (see figure A4 in the Additional File 7). Case C, where v_7 is not measured, provides very good results. All fluxes are predicted (with a mean interval increment of 122%), the interval increment is minor than 25% for 5 fluxes and minor than 100% for 9 fluxes. Case F, where v_{20} is not measured, provides good results too. There is a mean interval increment of 155% and the interval increment is minor than 100% for 11 fluxes (out 17). Case E, where v_{19} is not measured, provides slightly worse results than F. With the other sets of 5 measurements (B and A), some non-measured fluxes cannot be estimated. Nevertheless the intervals of the fluxes that can be estimated (10 and 7 fluxes respectively) are exactly the same that in the determined case.

Two sets of 4 measurements have been studied (I and H). Case I, where v_{20} and v_{22} are not measured, provides remarkable results. There is a mean interval increment of 180% with respect to the determined case and the increment is minor than 100% for 11 fluxes (out 18). This case is depicted in Figure 9 (in blue). For most fluxes the intervals are similar to the determined case (v_2, v_4, v_5, v_8, v_9, v_{10}, v_{11}, v_{12}, v_{13}, v_{15} and v_{21}). Intervals for v_{16} and v_{20} are larger but still useful, and only v_3, v_{14}, v_{17} and v_{18} are highly imprecise. Again, the temporal evolution of the estimated fluxes is similar to the determined case (see figure A4 in the Additional File 7).

This scenario has illustrated an important feature of the introduced procedure: it can estimate the evolution of the non-measured fluxes even when there is a lack of measurable species (i.e. the system is underdetermined) and the available measurements are uncertain.

Unbalanced propagation of measurements uncertainty
As it has been shown in previous sections, the uncertainty of the experimentally measured fluxes is not equally propagated to the estimated fluxes (i.e. not all the estimated fluxes are equally affected by measurements uncertainty). On the contrary, the structure of the metabolic network (the stoichiometric relations and the reversibility constraints) will determine how the uncertainty is propagated from the measured fluxes to the estimated ones.

A convenient way of measuring this effect is to calculate the interval size for each estimated flux at each time instant -in absolute and relative terms. The complete dataset has been included in the Additional File 6, but, as a summary, the average interval size (AIS) for each esti-

Table 2: Comparison of different estimations of the non-measured fluxes

Reactions [d]	Ref (v_1, v_6, v_7, v_{19}, v_{20}, v_{22}) MIa [b]	G (no v_{22}) MI [a]	[%c]	F (no v_{20}) MI [b]	[%]	E (no v_{19}) MI [b]	[%]	B (no v_6) MI [b]	[%]	A (no v_1) MI [b]	[%]	C (no v_7) MI [b]	[%]	I (no v_{20} v_{22}) MI [b]	[%]	H (no v_{19} v_{22}) MI [b]	[%]
1: G→G6P	0.267e	0.267	-	0.267	-	0.267	-	0.267	-	x	-	0.267	-	0.267	-	0.267	-
2: G6P→G3P+DAP	0.367	0.387	5%	0.628	71%	0.541	47%	0.398	8%	x	-	0.627	71%	0.628	71%	0.572	56%
3: G6P→R5P+CO2	0.131	0.199	53%	0.526	303%	0.340	160%	0.131	0%	0.131	0%	0.401	207%	0.526	303%	0.383	193%
4: DAP→G3P	0.367	0.387	5%	0.628	71%	0.541	47%	0.398	8%	x	-	0.627	71%	0.628	71%	0.572	56%
5: G3P→Pyr	0.735	0.774	5%	1.256	71%	1.082	47%	0.795	8%	x	-	1.253	71%	1.256	71%	1.144	56%
6: Pyr→L	0.475	0.475	-	0.475	-	0.475	-	x	-	0.475	-	0.475	-	0.475	-	0.475	-
7: Pyr+Glu→A+aKG	0.100	0.100	-	0.100	-	0.100	-	0.100	-	0.100	-	1.488	inf	0.100	-	0.100	-
8: Pyr→ACA+CO2	1.031	1.031	0%	1.562	51%	1.901	84%	x	-	x	-	0.957	-7%	1.562	51%	1.906	85%
9: Oxa+ACA→Cit	1.031	1.031	0%	1.562	51%	1.901	84%	x	-	x	-	0.957	-7%	1.562	51%	1.906	85%
10: Cit→aKG+CO2	1.031	1.031	0%	1.562	51%	1.901	84%	x	-	x	-	0.957	-7%	1.562	51%	1.906	85%
11: aKG→Mal+CO2	1.156	1.156	0%	1.604	39%	2.532	119%	x	-	x	-	1.443	25%	1.604	39%	2.530	119%
12: Mal→Oxa	0.994	0.994	0%	1.398	41%	1.769	78%	x	-	x	-	1.093	10%	1.398	41%	1.769	78%
13: Mal→Pyr+CO2	0.209	0.240	15%	0.352	68%	0.920	341%	0.209	0%	0.209	0%	0.903	332%	0.352	68%	0.918	340%
14: Oxa+Glu→Asp+aKG	0.131	0.199	53%	0.526	303%	0.340	160%	0.131	0%	0.131	0%	0.401	207%	0.526	303%	0.383	193%
15: Glu→aKG+NH4	0.150	0.182	21%	0.298	98%	0.870	479%	0.150	0%	0.150	0%	0.586	289%	0.298	98%	0.870	479%
16: Q→Glu+NH4	0.117	0.145	23%	0.325	177%	0.553	372%	0.117	0%	0.117	0%	0.569	386%	0.325	177%	0.548	367%
17: R5P+Asp+Q→Pu	0.104	0.277	165%	0.293	181%	0.200	91%	0.104	0%	0.104	0%	0.225	116%	0.526	404%	0.383	267%
18: R5P+Asp+2Q→Py	0.078	0.132	69%	0.283	262%	0.177	126%	0.078	0%	0.078	0%	0.209	168%	0.263	237%	0.163	108%
19: →NH4	0.141	0.141	-	0.141	-	1.419	904%	0.141	-	0.141	-	0.141	-	0.141	-	1.412	899%
20: →Q	0.132	0.132	-	1.127	752%	0.132	-	0.132	-	0.132	-	0.132	-	1.107	737%	0.132	-
21: →CO2	3.338	3.338	0%	4.770	43%	6.966	109%	x	-	x	-	3.843	15%	4.770	43%	6.966	109%
22: Pu-Py (constraint)	0.100	0.354	254%	0.100	-	0.100	-	0.100	-	0.100	-	0.100	-	0.526	426%	0.383	283%
Mean	0.554	0.587	39%	0.899	155%	1.138	196%	0.217	2%	0.156	0%	0.802	122%	0.927	180%	1.168	214%
Measured fluxes [number]	6	5		5		5		5		5		5		4		4	
Estimated fluxes [number]	16/16	17/17		17/17		17/17		7/17		10/17		17/17		18/18		18/18	
<25% (·Ref)		12		0		0		10		7		5		0		0	
25–100% (<2·Ref)		3		11		8		0		0		4		11		7	
100–300% (2–4·Ref)		2		3		5		0		0		5		2		7	
>300% (>4·Ref)		0		3		4		0		0		3		5		4	

Column Ref: FSA is applied by using the six available measurements (Determined case). Columns F, G, E B, A and C: FSA is applied by using a different set of 5 measurements in each case (underdetermined, 1 degree of freedom). Columns I and H: FSA is applied by using two different sets of 4 measurements (underdetermined, 2 degrees of freedom). In all cases the band of uncertainty described in the text has been used. [a] Mean interval size along time evolution; [b] in [mM/(d * 10^9 cells)]; [c] Intervals enlargement w.r.t. case Ref. (in percentage); [d] The nomenclature is given in the additional file 1; [e] Measured values are in bold.

mated flux is given in Table 3. It can be observed (determined case) that certain fluxes -such as v_{10}, v_{12} and v_{21}- are highly affected by the uncertainty of the measurements (they have an average interval size larger than 1 mM/$(d \cdot 10^9 \cdot cells)$), while other fluxes -such as v_{14} and v_{17}-are less sensitive (values around 0.1 mM/$(d \cdot 10^9 \cdot cells)$). Although it is obvious that in relative terms the smaller fluxes are usually more affected by the uncertainty, this phenomenon is not the only responsible for the unbalanced propagation of the uncertainty. For example, the calculated fluxes v_8 and v_{14} have a similar maximum value (around 1 mM/$(d \cdot 10^9 \cdot cells)$), but the effect of the uncertainty over them is dramatically different: v_8 is the flux more influenced by the uncertainty (with an AIS of 90.3% in relative terms) whereas v_{14} is quite insensitive to it (an AIS of 15.12%). Another example is given by v_{21}: although being one of the fluxes with a bigger maximum value (8.6 mM/$(d \cdot 10^9 \cdot cells)$), it is highly affected by the uncertainty (an AIS interval size of 3.4 mM/$(d \cdot 10^9 \cdot cells)$, which represents a 39.1%).

Furthermore, the data given in Table 3 provides a quantitative indication of the benefits of incorporating a redundant measurement. When seven fluxes are assumed to be measurable instead of six, the estimations of the non-measured fluxes are more precise (the interval sizes are reduced around 71% on average). This is particularly important for those fluxes that were poorly estimated in the determined case (reductions of 78% for v_8, v_9 and v_{10} and 76% for v_{12}).

Non-linear propagation of measurements uncertainty

In the previous section we analyzed the unbalanced propagation of the uncertainty from the measured fluxes to the estimated ones. Herein we investigate some characteristics of this propagation and, in particular, the interrelation between the uncertainty of the different measured fluxes and their combined effect over the estimated fluxes.

Again, the time series of the five measured species (G, L, A, NH_4 and Q) have been used, under the assumption that
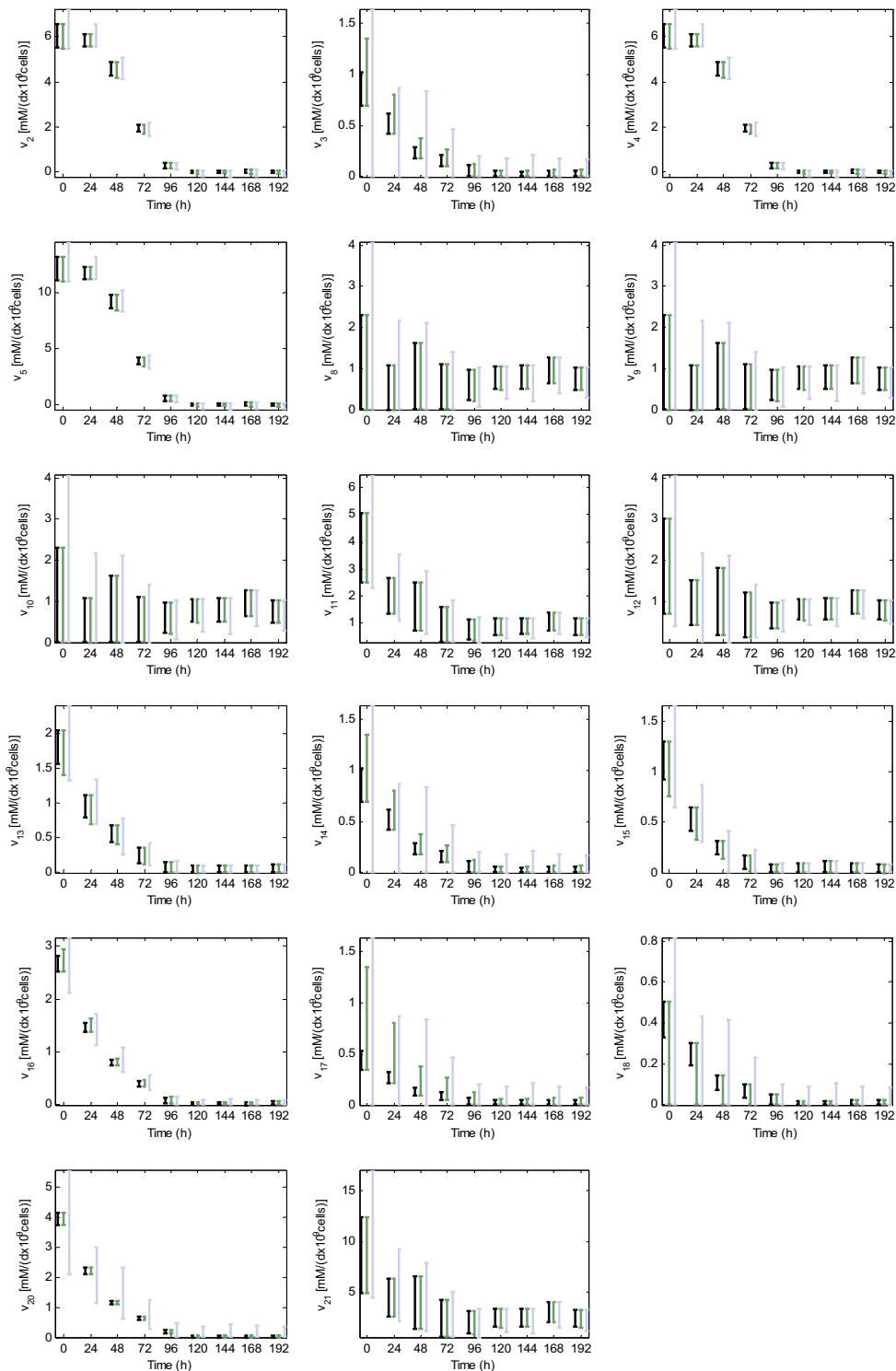
Figure 9
Non-measured fluxes estimated with FSA in two underdetermined cases (S3). The estimations when five fluxes are measured (v_1, v_6, v_7, v_{19} and v_{20}) are depicted in green (second interval). The estimations when four fluxes are measured (v_1, v_6, v_7 and v_{19}) are depicted in blue (third interval). The estimations obtained in the determined case (when six fluxes were measured) are included for the shake of comparison (in black).

Table 3: Imprecision of the estimated fluxes caused by measurements uncertainty

	Determined case			Determined/Redundant case			Comparative	
	Max [^{a}]	AIS [^{a}]	AIS [%^{b}]	Max [^{a}]	AIS [^{a}]	AIS [%^{b}]	Diff. [^{a}]	Diff. [%]
v_2	6,041	0,377	6,25%	6,032	0,321	5,32%	0,057	14,97%
v_3	0,853	0,129	15,12%	0,859	0,123	14,35%	0,006	4,41%
v_4	6,041	0,377	6,25%	6,032	0,321	5,32%	0,057	14,97%
v_5	12,081	0,755	6,25%	12,065	0,642	5,32%	0,113	14,98%
v_8	1,166	1,053	90,37%	0,715	0,231	32,32%	0,822	78,07%
v_9	1,166	1,053	90,37%	0,715	0,231	32,32%	0,822	78,07%
v_{10}	1,166	1,053	90,37%	0,715	0,231	32,32%	0,822	78,07%
v_{11}	3,769	1,180	31,30%	3,073	0,165	5,37%	1,015	86,02%
v_{12}	1,854	1,017	54,89%	1,263	0,241	19,05%	0,777	76,34%
v_{13}	1,813	0,209	11,52%	1,809	0,195	10,78%	0,014	6,58%
v_{14}	0,853	0,129	15,12%	0,859	0,123	14,35%	0,006	4,41%
v_{15}	1,113	0,150	13,52%	1,109	0,147	13,27%	0,003	2,11%
v_{16}	2,665	0,117	4,39%	2,668	0,114	4,26%	0,003	2,91%
v_{17}	0,426	0,101	23,64%	0,442	0,087	19,60%	0,014	14,10%
v_{18}	0,426	0,079	18,42%	0,417	0,063	15,17%	0,015	19,48%
v_{21}	8,698	3,407	39,17%	-	-	-	-	-
Mean		0,699	32,31%		0,202	14,32%	0,497	71,09%

Max: Maximum value of the estimated flux along time; AIS: Averaged interval size for each estimated fluxes (average of its interval sizes along time); Diff: Difference between determined and overdetermined cases; ^{a} in [mM/(d × 109 × cells)]; ^{b} the interval size for each estimated flux is expressed w.r.t. its maximum value. The complete dataset is given in the additional file 6.

the formation of purine and pyrimidine are equal (v_{22} = 0). Then, $15 \cdot 15$ executions of the estimation procedure have been carried out with different degrees of uncertainty for the measured fluxes v_1 and v_6 (between ± 2% and ± 30%). Afterwards, the averaged interval size for each estimated flux was calculated. This makes it possible to analyze how the different combinations of uncertainty in v_1 and v_6 affect to the estimated fluxes.

Figure 10 shows the averaged interval size (AIS) of one of the estimated fluxes (v_2) for each execution (similar figures are given in the Additional File 7). As it was predictable, the interval size tends to increase as the uncertainty of the measurements is increased. Therefore, the less precise estimation (i.e. the biggest AIS) corresponds to the execution with maximum uncertainty for v_1 and v_6. It is also seen that, as it was expected, the uncertainty of all the measured fluxes has not the same effect over the estimated ones. For instance, the uncertainty of v_6 has a bigger effect over v_2 than the uncertainty of v_1. More even, the figures illustrate two important properties of the propagation of the measurements uncertainty to the estimated fluxes.

On the one hand, *the propagation of the uncertainty does not satisfy the principle of superposition.* Let $f(u_i)$ be the interval size of a calculated flux when the degree of measurements uncertainty is u_i, then: $f(u_1) + f(u_2) \neq f(u_1 + u_2)$. To remark this, the result of summing up the independent effect of the uncertainty of v_6 and v_1 has been depicted in Figure 10

(black dots). Interestingly, if the uncertainty of one of the two measured fluxes is kept low then $f(u_1) + f(u_2) > f(u_1 + u_2)$; but if the uncertainty of both fluxes is increased, this is inverted: $f(u_1) + f(u_2) < f(u_1 + u_2)$. This implies that the net result of considering the uncertainty of two (or more) measured fluxes can not be predicted just by summing up the results of considering only the uncertainty of one measured flux at a time. On the contrary, the net result may be given by a complex non-linear function, as it happens in the example of Figure 10, where:

• The uncertainty in v_6 is always translated to the estimated v_2 (see right bottom figure). When v_6 uncertainty increases, the AIS of v_2 increases -even if there is not v_1 uncertainty. However, the bigger v_1 uncertainty gets, the bigger the effect of adding v_6 uncertainty is.

• When v_6 uncertainty is low, the addition of v_1 uncertainty has a low effect over the estimated v_2 (see right bottom figure) More precisely, the first small addition of v_1 uncertainty has a slight effect, but the subsequent additions do not (there is a saturation). However, the saturation limit increases with v_6 uncertainty, and, therefore, the more uncertainty there is in v_6, the more important effect has addition of uncertainty in v_1. In summary, v_1 uncertainty has not an important effect itself, but its combination with v_6 uncertainty boosts it.
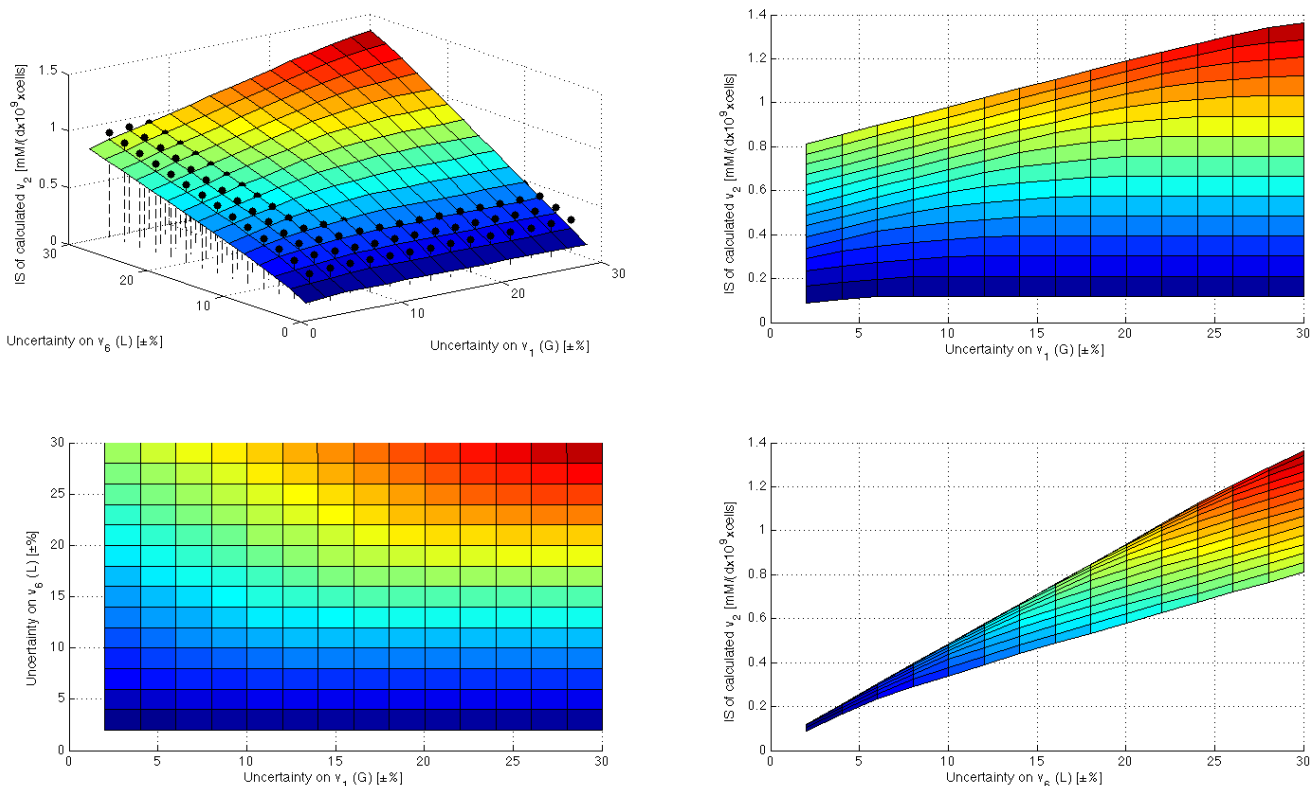
Figure 10
Effect over the estimated flux v_2 of the uncertainty on the measured values of v_1 and v_6. The surface (and its projections) represents the averaged interval size (IS) of the estimated v_2 when different degrees of uncertainty are considered for the measured fluxes v_1 and v_6. In the top left figure the results of summing up the independents effects of v_6 uncertainty and v_1 uncertainty are depicted with black dots.

On the other hand, as the superposition principle is not fulfilled *the effect of the uncertainty of one measured flux over one estimated flux is not linear*, i.e. $f(k \cdot u_1) \neq k \cdot f(u_1)$ For example, assume that the uncertainty of v_6 is fixed in 10% (fourth row in the right top figure). It can be observed that the effect of adding a first 4% of uncertainty to v_1 is higher than the effect of adding a second one. In fact, when the uncertainty of v_1 is bigger than 16%, the addition of more uncertainty has practically zero effect (there is a saturation phenomenon).

In the last two sections it has been shown that the relationship between the uncertainty of the measurements and the precision of the estimation is a complex one. On the one hand, the propagation of measurements uncertainty to each estimated flux will be different. On the other hand, the net effect of considering the uncertainty of two (or more) measured fluxes simultaneously does not correspond to the sum of the effects of considering the uncertainty of each measured flux one at a time. Finally,

the effect of the uncertainty of one measured flux over one estimated flux is not linear.

Therefore, when the procedure introduced in this paper considers the propagation of the uncertainty from the measurements to the estimated fluxes, it provides non-trivial information.

Analysis of the effect of the uncertainty of each measured flux

In this section we analyze the effect of the uncertainty of each measured flux over the imprecision of the estimated fluxes. Basically, we can apply the estimation procedure over previously logged data -considering the uncertainty of each measured flux one at a time- in order to determine which measured fluxes have the more critical uncertainty. There are two similar approaches to carry out the analysis:

a) Direct approach. Calculate the *increase* of the imprecision of the estimated fluxes when the uncertainty of one

measured flux is *increased*. This calculation is repeated for each measured flux.

b) Indirect approach. Calculate the *reduction* of the imprecision of the estimated fluxes when the uncertainty of one measured flux is *decreased*. This is repeated for each measured flux. Notice that the effect of decreasing the uncertainty is not the inverse of increasing it, i.e. $f(u + x) + f(u + x) \neq f(0)$.

The direct approach (a) informs about the effect of considering the uncertainty of each measured flux over the estimated ones (it is similar to a classical analysis of sensitivity). This information may be useful during the setting-up of a process plant in order to choose the sources of measurements (the equipment and the protocols). Nevertheless, the indirect approach (b) is probably more promising. It calculates how much the imprecision of the estimated fluxes will be reduced, if we reduce the uncertainty of one of the measured fluxes. Given the characteristics of our current equipment and our measuring protocols (e.g. our sensors provide measurements with a ± 5% of uncertainty), we can calculate which of the measured fluxes should be more accurately measured in order to improve the precision of the estimations (e.g. using a

more accurate sensor or taking redundant measurements).

The indirect analysis has been applied to the cultivation of CHO cells (using the set of 5 measurements described above and the assumption of equal formation of purine and pyrimidine). Figure 11 shows the reduction of the imprecision of each estimated flux when the uncertainty of a measured flux is decreased a 3%. This is repeated for each measured flux (v_1, v_6, v_7, v_{19} and v_{20}). Those data provide valuable information. For example, it is shown that the maximum reduction of the imprecision occurs when the uncertainty of v_{20} is reduced at time 144 h: the imprecision of v_3, v_{14}, v_{16} and v_{18} is reduced more than 85%. It can be also observed that during the first 96 h removing the uncertainty of v_{20} slightly reduces the imprecision of the estimated v_{16}, but this reduction is very important between 120 h and 192 h. Those data are summarized in Figure 12. The left figure shows the averaged reduction at each time instant, and the one on the right the averaged reduction for each estimated flux. For example, it can be observed that removing the uncertainty of v_1 or v_6 has no effect over the estimations of v_3, v_{14}, v_{15}, v_{16} v_{17} and v_{18}. This information can be used to improve our estimations in a rational manner. For example, one could be interested in increasing the precision of the estimation of v_3. In
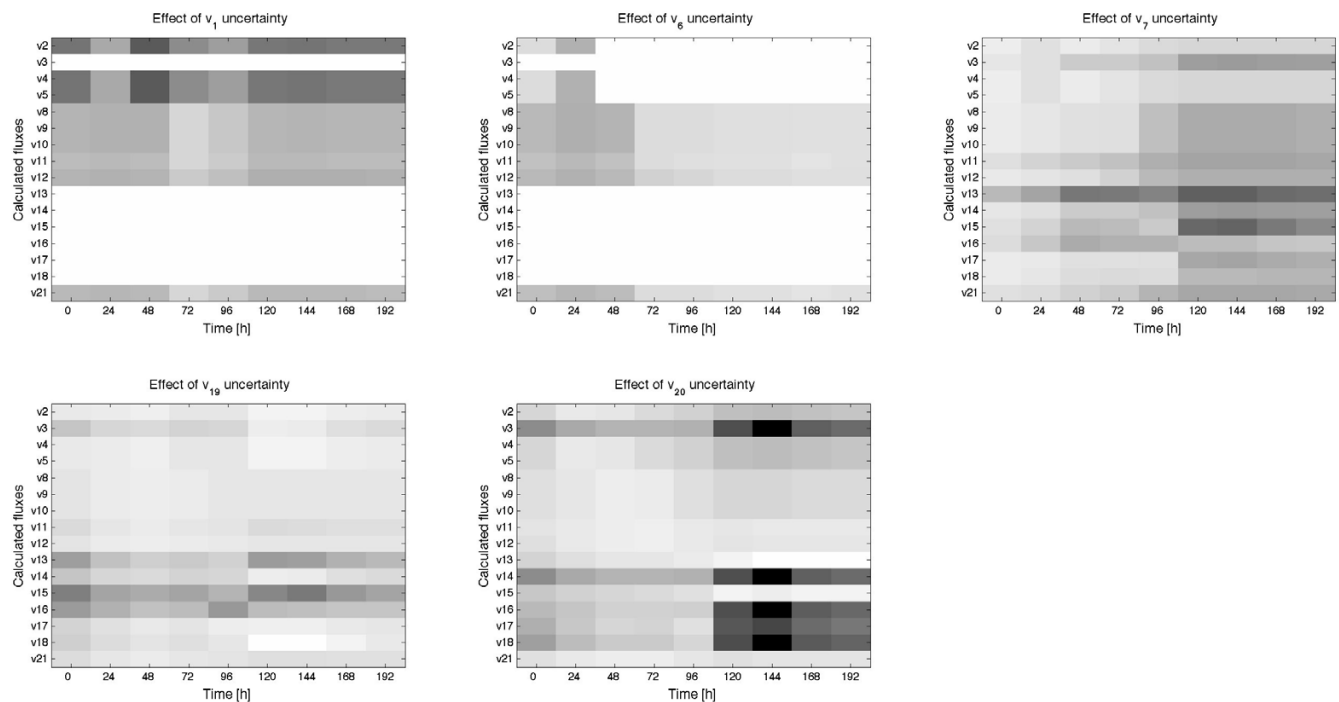

Figure 11
Effect of the uncertainty of each measured flux over the imprecision of the estimated fluxes along time. The figures show the reduction of the imprecision of the estimated (or calculated) fluxes when the uncertainty of the measured flux is decreased a 3%. The reductions are quantified between 0% (white colour) and 100% (black).
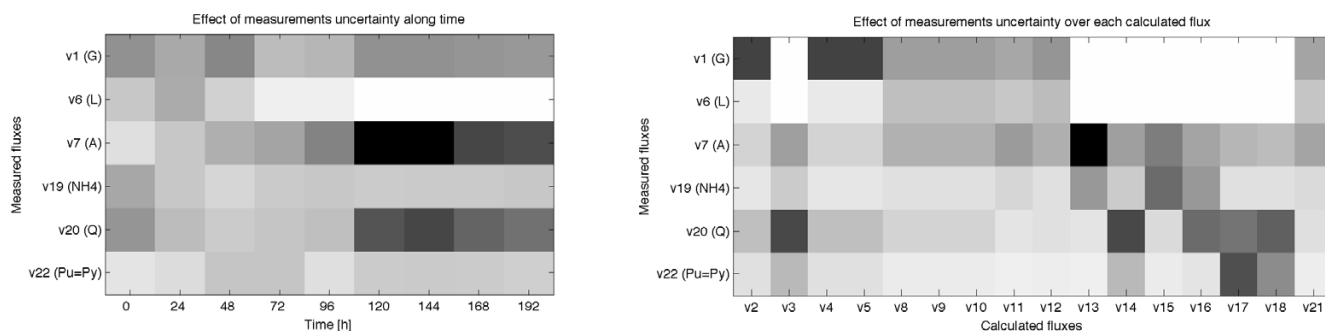
Figure 12
Summary of the effect of the uncertainty of each measured flux over the imprecision of the estimated fluxes.
(Left) Averaged reduction at each time instant of the imprecision of the estimated (or calculated) fluxes when the uncertainty
of the measured flux is decreased a 3% (Right) Averaged reduction of the imprecision of each estimated (or calculated) flux
when the uncertainty of the measured flux is decreased a 3%. The reductions are quantified between 0% (white colour) and
100% (black).

this case the indirect analysis indicates that the best option is to reduce the uncertainty of the measured v_{20}. However, if we want to improve the estimations during the transition phase (between 72 h and 120 h) we should reduce the uncertainty of v_7. Finally, if we prefer to improve the overall precision of the estimations we should reduce the uncertainty of v_7, although reducing the uncertainty of v_1 or v_{20} brings similar benefits.

More details about this analysis -including the complete dataset- are given in the Additional File 8. In addition, the results obtained with the direct analysis are also included there.

Conclusion
In this contribution we have presented a new procedure to estimate the temporal evolution of the metabolic fluxes. It copes with the intrinsic uncertainty of experimental measurements and with the lack of measurable species by means of the use of the Flux Spectrum Approach (FSA). The potential of the procedure has been demonstrated using a real problem: the estimation of the intracellular fluxes of CHO cells cultivated in batch mode in stirred flasks. Using this example, the benefits that the use of FSA brings to the whole procedure have been illustrated through a comparison with the use of Metabolic Flux Analysis (MFA), a well-established methodology that is the basis of related procedures [24-26]. When the available measurements are only *almost* sufficient (i.e. the system is determined but not redundant), the procedure provides a more reliable and richer estimation of the evolution of the fluxes, because it takes into account measurements uncertainty and it considers the reversibility constraints. It has been also shown that even when measurements are insufficient (i.e. the system is underdetermined), the procedure is capable of estimating the

evolution of the non-measured fluxes. Finally, when measurements are sufficient (i.e. the system is determined and redundant), the procedure provides a reliable estimation of the non-measured fluxes -because it considers measurements uncertainty- and offers an interesting approach to cope with inconsistency.

The procedure to estimate the metabolic fluxes can be applied off-line (with previously collected data), providing an insight into the time-varying behaviour of the organism. This can help in the understanding of its dynamic metabolic regulation and its adaptation to the environmental conditions. It can also be useful for physiological studies, strain characterization tasks, and to guide research to improve strain and processes. On the other hand, the procedure is a promising tool for on-line monitoring processes in industrial environments, where there is still a lack of reliable on-line sensors. The features of the procedure are especially suitable for this application.

In summary, it has been shown that the temporal evolution of non-measured fluxes can be estimated by using a set of measurable species and a set of known biological constraints. Moreover, the procedure proposed considers the intrinsic uncertainty of the experimental measurements and could be applied even if there is a lack of measurable species.

Methods
Calculation of the measured fluxes by approximating the derivative
If the concentration of an extracellular specie is measured, the value of its corresponding flux can be worked out from eq. 3. But that means the derivative $d\xi/dt$ has to be approximated. The Euler methods provide the most straightforward approximations:

$$\text{Backward}: \quad \frac{df(k)}{dt} \approx \frac{f(k)-f(k-1)}{t(k)-t(k-1)}$$
$$\text{Middlepoint}: \quad \frac{df(k)}{dt} \approx \frac{f(k+1)-f(k-1)}{t(k+1)-t(k-1)} \tag{6}$$

The backward version does not introduce an intrinsic delay (the derivative of *f(.)* at *k* is calculated with the values of *f(.)* at *k* and *k-1*), but the middle point provides a less noisy approximation. In any case, usually this approach needs to be combined with the use of filters.

Moving average filters

These filters calculate each value of a new signal by averaging the values of the original signal within a time window. Thus, the new signal becomes smoother. The centred moving average (CMA) provides the best results because it uses past and future information. The filtered value for instant *k* (CMA_k) is calculated by averaging the values of the original signal (*S*) between *k-n* and *k+n*:

$$CMA_k = \frac{\sum\limits_{1}^{n} S_{k-i} + S_k + \sum\limits_{1}^{n} S_{k+i}}{2 \cdot n + 1} \tag{7}$$

If only past values of the original signal are available, the standard moving average (SMA) can be used instead:

$$SMA_k = \frac{\sum\limits_{0}^{n} S_{k-i}}{n+1} \tag{8}$$

The key parameter of moving average filters is the size of the window (i.e., the number of averaged values of the original signal). The optimal size would be one observation in order to be as close as possible to the original signal. However, as noise rejection is desired, the window size needs to be increased. Hence, there is a trade-off between sensitivity to noise and delay with respect to the original signal. A typical variant of these filters includes multiplying factors to give a different weight to each value within the time window (e.g. the exponential moving average).

Calculation of the measured fluxes with a nonlinear observer

A high-gain nonlinear observer of the extracellular fluxes can be directly synthesized from (3) by using the method proposed in [47]:

$$\frac{d\xi_e}{dt} = v_e \cdot X - D \cdot \xi_e - 2 \cdot \theta \cdot (\xi_e - \xi)$$
$$\frac{dv_e}{dt} = -\frac{\theta^2 \cdot (\xi_e - \xi)}{X} \tag{9}$$

where ξ_e denotes the observed concentration of the extracellular specie and v_e the observed flux. The unique adjustable parameter is θ. Not only these observers are proved to be stable, but also its asymptotic error can be made arbitrarily small by choosing sufficiently large values of θ. However, very large values need to be avoided in practice since the observer may become noise sensitive. Thereby the choice of θ represents a trade-off between fast convergence (minor delay) and sensitivity to noise.

Flux spectrum approach

The non-measured fluxes at a certain time instant *k* can be estimated using the Flux Spectrum Approach (FSA). The method works as follows [33]:

1. Impose the set of constraints given by (2) and the reversibility constraints for the irreversible reactions, $v_i \geq 0$. Then, in order to consider the uncertainty of the measured fluxes, the unique value of each measured flux v_m can be replaced by an interval $[v_{m,\ min},\ v_{m,\ max}]$. Accordingly, each equation of (2) is substituted by two inequalities. The resultant constraints define a region where the actual flux distribution may live.

2. Calculate the minimum and maximum values within the region for each non-measured flux v_{uj}, by solving a set of min/max linear programming problems (n_u minimizations and n_u maximizations):

$$\forall v_{uj}, j = 1, \ldots, nu$$
$$Min\{v_{uj}\} \quad subject:$$
$$S_u \cdot \mathbf{v_u} \geq \{-S_m \cdot [\mathbf{v_m}]\}^{\min} \quad S_u \cdot \mathbf{v_u} \leq \{-S_m \cdot [\mathbf{v_m}]\}^{\max} \quad v_i \geq 0$$
$$Max\{v_{uj}\} \quad subject:$$
$$S_u \cdot \mathbf{v_u} \geq \{-S_m \cdot [\mathbf{v_m}]\}^{\min} \quad S_u \cdot \mathbf{v_u} \leq \{-S_m \cdot [\mathbf{v_m}]\}^{\max} \quad v_i \geq 0 \tag{10}$$

where n_u is the number of unknown fluxes, $\mathbf{v_u}$ is the vector of non-measured fluxes, v_i represents each irreversible flux and $[\mathbf{v_m}]$ represents the two vectors $[\mathbf{v_{m,\ min}},\ \mathbf{v_{m,\ max}}]$ formed with the uncertain representation of the measured fluxes. $\{-S_m \cdot [\mathbf{v_m}]\}^{\min}$ and $\{-S_m \cdot [\mathbf{v_m}]\}^{\max}$ are vectors formed with the maximum and minimum row values of the product $-S_m \cdot [\mathbf{v_m}]$. To calculate each row, maximum and minimum values of $\mathbf{v_m}$ need to be combined taking into account the signs of the elements of the corresponding row of S_m (you can find a detailed description of the algorithm in the Additional File 5).

The obtained $v_{uj,\,min}$ and $v_{uj,\,max}$ for each non-measured flux define an interval bracketing its possible values: $v_{uj}(k) = [v_{uj,\,min},\,v_{uj,\,max}]$.

Auxiliary: link between extracellular and intracellular fluxes

A set of *ne* extracellular fluxes can be linked with the intracellular fluxes by using a matrix S_ξ fulfilling:

$$\mathbf{v}_\xi = S_\xi \cdot \mathbf{v} \qquad (11)$$

where $\mathbf{v}_\xi$ is the vector of extracellular fluxes and $\mathbf{v}$ the vector of intracellular fluxes. Eq. 11 can be easily joined with eq. 1 as follows:

$$\begin{bmatrix} S_{mxn} & 0_{mxne} \\ S\xi_{nexn} & I_{nexne} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{v} \\ \mathbf{v}_\xi \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad (12)$$

Hence, the extended system holds the structure of a homogeneous system of linear equations.

Auxiliary: inconsistency of the measured fluxes

A redundant system will be consistent if it fulfils the consistency condition:

$$R \cdot \mathbf{v_m} = 0 \quad R = S_m - S_u \cdot S_u^{\#} \cdot S_m \qquad (13)$$

where R is the redundancy matrix and $S_u^{\#}$ the Penrose pseudo-inverse of S_u. In case inconsistency is detected, the method described in [38,39] can be used to estimate its importance. It is based upon statistical hypothesis testing to determine if redundancies are satisfied to within expected experimental error. The test is performed by calculating a consistency index *h* as follows:

$$\begin{aligned} h &= \varepsilon^T \cdot P^{-1} \cdot \varepsilon \\ \varepsilon &= -R_r \cdot \mathbf{v_m} \\ P &= R_r \cdot F \cdot R_r^T \end{aligned} \qquad (14)$$

where R_r is the reduced redundancy matrix and *F* the variances-covariances matrix of the measurements in $\mathbf{v_m}$. If a given set of measured fluxes $\mathbf{v_m}$ fails the consistency check ($h > \chi^2$), then there is a (confidence level)% chance that either $\mathbf{v_m}$ contains gross errors or the assumed stoichiometric matrix is incorrect. The χ^2 values for two confidence levels are given in Table 4. It must be noticed that some measured fluxes have no impact on the consistency of the system, so they are not considered in the analysis of consistency. These fluxes are called non-balanceable. On the contrary, a measured flux is called balanceable if the consistency of the system depends on its value. They can be detected as explained in [14]. The balanceable fluxes can be adjusted (or balanced) if they are inconsistent. Fol-

Table 4: Chi-square values for two confident levels

Degrees of freedom	90%	95%
1	2.71	3.84
2	4.61	5.99
3	6.25	7.81
4	7.78	9.49

lowing the method described in [38,39], the adjusted fluxes can be calculated as:

$$\hat{\mathbf{v}}_\mathbf{m} = \left(I - F \cdot R_r^T \cdot P^{-1} \cdot R_r^{\#} \right) \cdot \mathbf{v_m} \qquad (15)$$

where $R_r^{\#}$ denoted the Penrose pseudo-inverse of the matrix R_r. This equation provides the adjusted values for the balanceable fluxes and the original values for the non-balanceable ones.

Auxiliary: Metabolic flux analysis

When the system (2) is determined but not redundant the unique solution can be calculated by using the inverse matrix of S_u:

$$\mathbf{v_u} = -S_u^{-1} \cdot S_m \cdot \mathbf{v_m} \qquad (16)$$

When the system is determined but redundant, matrix S_u is not invertible so the Penrose pseudo-inverse is used instead (providing a least squares solution):

$$\mathbf{v_u} = -S_u^{\#} \cdot S_m \cdot \mathbf{v_m} \qquad (17)$$

Finally, if system (2) is underdetermined, Metabolic Flux Analysis cannot be used. Only some fluxes may be uniquely calculable by using the method explained in [14].

List of abbreviations

MFA: Metabolic flux analysis

FSA: Flux spectrum approach

FBA: Flux balance analysis

CHO: Chinese hamster ovary (cells)

AIS: Average along time of the interval size of an estimated flux.

Authors' contributions

FL and JP designed the research, analyzed the results and conceptualized the manuscript. FL performed the estimation of non-measured fluxes and drafted the manuscript.

JP supervised and coordinated the project. All the authors read and approved the final manuscript.

Additional material

Additional file 1
Metabolic network description. *List of metabolites, reactions and stoichiometric matrixes.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-8-421-S1.doc]

Additional file 2
Flux spectrum approach in the underdetermined and redundant case. *Example of the estimation of fluxes with the Flux Spectrum Approach in an Underdetermined and redundant case.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-8-421-S2.doc]

Additional file 3
Conversion of measured concentrations to measured fluxes. *Example of the calculation of the measured fluxes when the measurements of concentration are obtained at a high sample rate.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-8-421-S3.doc]

Additional file 4
Mathematical model (SBML and matlab). *The zip file contains the mathematical model of the metabolism of CHO cells in three different formats: a matlab script, an SBML model and a metatool file.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-8-421-S4.zip]

Additional file 5
Implementation of the Flux Spectrum Approach. *The zip file contains a matlab script which implements the method to estimate the non-measured fluxes with the Flux Spectrum Approach and a simple example that illustrates how to use it.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-8-421-S5.zip]

Additional file 6
Analysis of the unbalanced propagation of the uncertainty. *Dataset with the interval size of the estimations of each non-measured flux at each time instant.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-8-421-S6.doc]

Additional file 7
Additional figures and tables. *Additional figures cited in the manuscript.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-8-421-S7.doc]

Additional file 8
Analysis of the effect of the uncertainty of each measured flux. *Details and datasets of the indirect and direct analysis of the effect of the uncertainty of each measured flux.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2105-8-421-S8.doc]

References
1. Palsson B: **The challenges of in silico biology.** *Nat Biotechnol* 2000, **18:**1147-1150.
2. Kitano H: **Computational systems biology.** *Nature* 2002, **420:**206-210.
3. Papin JA, Price ND, Wiback SJ, Fell DA, Palsson BO: **Metabolic pathways in the post-genome era.** *Trends Biochem Sci* 2003, **28:**250-258.
4. Cornish-Bowden A, Cardenas ML: **From genome to cellular phenotype – a role for metabolic flux analysis?** *Nat Biotechnol* 2000, **18:**267-268.
5. Stephanopoulos GN, Aristidou AA, Nielsen J: **Metabolic Engineering: Principles and Methodologies.** San Diego: Academic Press; 1998.
6. Bonarius HPJ, Schmid G, Tramper J: **Flux Analysis of underdetermined metabolic networks: the quest for the missing constraints.** *Tibtech* 1997, **15(308):**309-314.
7. Wiback SJ, Famili I, Greenberg HJ, Palsson BO: **Monte Carlo sampling can be used to determine the size and shape of the steady-state flux space.** *J Theor Biol* 2004, **228:**437-447.
8. Edwards JS, Covert M, Palsson B: **Metabolic modelling of microbes: the flux-balance approach.** *Environ Microbiol* 2002, **4:**133-140.
9. Duarte NC, Palsson BØ, Fu P: **Integrated analysis of metabolic phenotypes in Saccharomyces cerevisiae.** *BMC Genomics* 2004, **5:**63.
10. Bonarius HPJ, Hatzimanikatis V, Meesters KPH, de Gooijer CD, Schmid G, Tramper J: **Metabolic flux analysis of hybridoma cells in different culture media using mass balances.** *Biotechnol Bioeng* 1996, **50:**299-318.
11. Follstad BD, Balcarcel RR, Stephanopoulos G, Wang DI: **Metabolic flux analysis of hybridoma continuous culture steady state multiplicity.** *Biotechnol Bioeng* 1999, **63:**675-683.
12. Nyberg GB, Balcarcel RR, Follstad BD, Stephanopoulos G, Wang DI: **Metabolism of peptide amino acids by Chinese hamster ovary cells grown in a complex medium.** *Biotechnol Bioeng* 1999, **62:**324-335.
13. Fischer E, Zamboni N, Sauer U: **High-throughput metabolic flux analysis based on gas chromatography-mass spectrometry derived 13C constraints.** *Anal Biochem* 2004, **325:**308-316.
14. Klamt S, Schuster S, Gilles ED: **Calculability analysis in underdetermined metabolic networks illustrated by a model of the central metabolism in purple nonsulfur bacteria.** *Biotechnol Bioeng* 2002, **77:**734-751.
15. Kauffman KJ, Prakash P, Edwards JS: **Advances in flux balance analysis.** *Curr Opin Biotechnol* 2003, **14:**491-496.
16. Price ND, Papin JA, Schilling CH, Palsson BO: **Genome-scale microbial in silico models: the constraints-based approach.** *Trends Biotechnol* 2003, **21:**162-169.
17. Edwards JS, Ibarra RU, Palsson BO: **In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data.** *Nat Biotechnol* 2001, **19:**125-130.
18. Price ND, Papin JA, Palsson BO: **Determination of redundancy and systems properties of the metabolic network of Helico-**

bacter pylori using genome-scale extreme pathway analysis. *Genome Res* 2002, **12**:760-769.
19. Schilling CH, Covert MW, Famili I, Church GM, Edwards JS, Palsson BO: **Genome-Scale Metabolic Model of Helicobacter pylori 26695.** *J Bacteriol* 2002, **184**:4582-4593.
20. Mahadevan R, Burgard A, Famili I, Van Dien S, Schilling C: **Applications of Metabolic Modeling to Drive Bioprocess Development for the Production of Value-added Chemicals.** *BIOTECHNOLOGY AND BIOPROCESS ENGINEERING* 2005, **10**:408.
21. Mahadevan R, Edwards JS, Doyle FJ: **Dynamic Flux Balance Analysis of Diauxic Growth in Escherichia coli.** *Biophys J* 2002, **83**:1331-1340.
22. Gayen K, Venkatesh K: **Analysis of optimal phenotypic space using elementary modes as applied to Corynebacterium glutamicum.** *BMC Bioinformatics* 2006, **7**:445.
23. Provost A: **Metabolic design of dynamic bioreaction models.** Faculté des Sciences Appliquées, Université catholique de Louvain, Louvain-la-Neuve, Louvain-la-Neuve; 2006.
24. Herwig C, von Stockar U: **A small metabolic flux model to identify transient metabolic regulations in *Saccharomyces cerevisiae.*** *Bioprocess and Biosystems Engineering* 2002, **24**:395-403.
25. Takiguchi N, Shimizu H, Shioya S: **An on-line physiological state recognition system for the lysine fermentation process based on a metabolic reaction model.** *Biotechnol Bioeng* 1997, **55**:170-181.
26. Henry O, Kamen A, Perrier M: **Monitoring the physiological state of mammalian cell perfusion processes by on-line estimation of intracellular fluxes.** *Journal of Process Control* 2007, **17**:241-251.
27. Ren HT, Yuan JQ, Bellgardt KH: **Macrokinetic model for methylotrophic Pichia pastoris based on stoichiometric balance.** *J Biotechnol* 2003, **106**:53-68.
28. Provost A, Bastin G, Agathos SN, Schneider YJ: **Metabolic Design of Macroscopic Models: Application to CHO cells.** *Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC'05.44th IEEE Conference on 2005* :2982-2989.
29. d'Anjou MC, Daugulis AJ: **A rational approach to improving productivity in recombinant Pichia pastoris fermentation.** *Biotechnol Bioeng* 2001, **72**:1-11.
30. Lei F, Rotbøll M, Jørgensen SB: **A biochemically structured model for *Saccharomyces cerevisiae.*** *J Biotechnol* 2001, **88**:205-221.
31. Teixeira AP, Alves C, Alves PM, Carrondo MJ, Oliveira R: **Hybrid elementary flux analysis/nonparametric modeling: application for bioprocess control.** *BMC Bioinformatics* 2007, **8**:30.
32. Sainz J, Pizarro F, Perez-Correa JR, Agosin E: **Modeling of yeast metabolism and process dynamics in batch fermentation.** *Biotechnol Bioeng* 2003, **81**:818-828.
33. Llaneras F, Pico J: **An interval approach for dealing with flux distributions and elementary modes activity patterns.** *J Theor Biol* 2007, **246**:290-308.
34. Henry CS, Jankowski MD, Broadbelt LJ, Hatzimanikatis V: **Genome-Scale Thermodynamic Analysis of *Escherichia coli* Metabolism.** *Biophys J* 2006, **90**:1453-1461.
35. Kummel A, Panke S, Heinemann M: **Systematic assignment of thermodynamic constraints in metabolic network models.** *BMC Bioinformatics* 2006, **7**:512.
36. Hoppe A, Hoffmann S, Holzhutter HG: **Including metabolite concentrations into flux balance analysis: thermodynamic realizability as a constraint on flux distributions in metabolic networks.** *BMC Syst Biol* 2007, **1**:23.
37. Henry CS, Broadbelt LJ, Hatzimanikatis V: **Thermodynamics-based metabolic flux analysis.** *Biophys J* 2007, **92**:1792-1805.
38. Van der Heijden RTJM, Romein B, Heijnen JJ, Hellinga C, Luyben K, Ch AM: **Linear Constraint Relations in Biochemical Reaction Systems: I. Classification of the Calculability and the Balanceability of Conversion Rates.** *Biotechnol Bioeng* 1994, **43**:3-10.
39. Van der Heijden RTJM, Romein B, Heijnen JJ, Hellinga C, Luyben K, Ch AM: **Linear Constraint Relations in Biochemical Reaction Systems: II. Diagnosis and Estimation of Gross Errors.** *Biotechnol Bioeng* 1994, **43**:11-20.
40. Bastin G, Dochain D: **On-line Estimation and Adaptative Control of Bioreactors.** Amsterdam: Elsevier; 1990.
41. Schügerl K, Bellgardt KH: **Bioreaction Engineering: Modeling and Control.** Springer; 2000.
42. Herwig C, Marison I, von Stockar U: **On-line stoichiometry and identification of metabolic state under dynamic process conditions.** *Biotechnol Bioeng* 2001, **75**:345-354.
43. Pei SC, Shyu JJ: **Eigenfilter design of higher-order digital differentiators.** *Acoustics, Speech, and Signal Processing [see also IEEE Transactions on Signal Processing], IEEE Transactions on 1989* **37**:505-511.
44. Luenberger D: **An introduction to observers.** *Automatic Control, IEEE Transactions on 1971* **16**:596-602.
45. Levant A: **Robust exact differentiation via sliding mode technique.** *Automatica* 1998, **34**:379-384.
46. Dochain D, Pauss A: **On-line Estimation of Microbial Specific Growth-Rates: An Illustrative Case Study.** *The Can J Chem Eng* 1988, **66**:626.
47. Farza M, Busawon K, Hammouri H: **Simple nonlinear observers for on-line estimation of kinetic rates in bioreactors.** *Automatica* 1998, **34**:301-318.
48. Provost A, Bastin G: **Dynamic metabolic modelling under the balanced growth condition.** *Journal of Process Control* 2004, **14**:717-728.