Bidirectional Reaction Steps in Metabolic Networks Part II: Flux Estimation and Statistical Analysis

Wolfgang Wiechert, Claudia Siefke, Albert A. de Graaf, and Achim Marx
Institut für Biotechnologie, Forschungszentrum Jülich,
52425 Jülich GmbH, Germany

Accepted for publication in “Biotechnology an Bioengineering”
November 15, 1996

Running title: Bidirectional Reaction Steps in Metabolic Networks (Part I)

Corresponding author: Wolfgang Wiechert
IMR, Abteilung Simulationstechnik
Universität-GH Siegen
57068 Siegen
Germany
Phone: -271 / 740-47 27, Fax: -23 65
E-mail: wiechert@simtec.imr.mb.uni-siegen.de

Abstract: Metabolic carbon labelling experiments enable a large amount of extracellular fluxes and intracellular carbon isotope enrichments to be measured. Since the relation between the measured quantities and the unknown intracellular metabolic fluxes is given by bilinear balance equations, flux determination from this data set requires the numerical solution of a nonlinear inverse problem. To this end a general algorithm for flux estimation from metabolic carbon labelling experiments based on the least squares approach is developed in this contribution and complemented by appropriate tools for statistical analysis. The linearization technique usually applied for the computation of nonlinear confidence regions is shown to be inappropriate in the case of large exchange fluxes. For this reason a sophisticated compactification transformation technique for nonlinear statistical analysis is developed. Statistical analysis is then performed by computing appropriate statistical quality measures like output sensitivities, parameter sensitivities and the parameter covariance matrix. This allows to determine the order of magnitude of exchange fluxes in most practical situations. An application study with a large data set from lysine producing *Corynebacterium glutamicum* demonstrates the power and limitations of the carbon labelling technique. It is shown that all intracellular fluxes in central metabolism can be quantitated without assumptions on intracellular energy yields. At the same time several exchange fluxes are determined which is an invaluable information for metabolic engineering.

Keywords: stationary flux estimation, sensitivity analysis, covariance analysis, non-linear statistics, *Corynebacterium glutamicum*

Introduction

In the preceding contribution [Wiechert (1996b)] (henceforth denoted as Part I) we developed a general modelling approach for stationary metabolic carbon isotope labelling experiments that extends the well established metabolite flux balancing technique [Vallino (1992), Varma (1994)]. Strong emphasis was laid on the description and analysis of bidirectional reaction steps and on the documentation and exploitation of biological assumptions made on intracellular fluxes. A simulation strategy for labelling experiments and the corresponding computational methods were introduced and some general properties of labelling systems were derived. In this contribution we now concentrate on flux estimation from given experimental data and the statistical analysis of the achieved results. The problems that have to be expected due to the bilinear structure of the carbon labelling balance equations with respect to fluxes and fractional labelling have already been illustrated in Part I.

Available measurement data

The experimental details of metabolic carbon labelling experiments are described e.g. in [Anderson (1983), Wiechert (1995c), Wiechert (1996a)]. The set of measured data obtained with such experiments is always subdivided into two parts:

1. *Extracellular metabolite fluxes* between the cell interior and the surrounding medium or the biomass like substrate uptake, product formation, incorporation of precursor metabolites into biomass or gas efflux are measured with standard bioreactor instrumentation using the ideas from [Holms (1986), Neidhardt (1990), Vallino (1991)]. The metabolite flux balancing technique is solely based on this data [Vallino (1992), Goel (1993), Varma (1994), Jorgensen (1995)].

2. *Fractional enrichments* of ^{13}C label within certain carbon atom positions of intracellular metabolite pools are measured by NMR. In particular the usage of hydrolysed intracellular polymers recently led to a dramatically increased amount of available labelling data [Marx (1996), Wiechert (1996a), Wiechert (1996b)].

For the experiment described in [Marx (1996)] a total of 14 extracellular fluxes and 26 fractional carbon isotope enrichments were determined (including the fractional enrichments of the ribose-5-phosphate pool that has now also become available). This large amount of measured data requires the development of sophisticated methods for flux estimation and statistical analysis.

Flux estimation and statistical analysis

The ultimate goal of our carbon labelling experiments is the estimation of the intracellular metabolite fluxes that were present during the experiment from the available measurement data. While this can be achieved by direct matrix computations in the case of metabolite flux balancing [Lawson (1974), Vallino (1991), van Heijden (1994a)] the bilinearity of carbon labelling systems prevents this simple approach. The same holds for the statistical analysis of the estimated fluxes, i.e. classical results of linear statistics [Chatterjee (1988)] can be readily applied to metabolite flux balancing [Wang (1983), Vallino (1991), van Heijden (1994b)] but must be appropriately extended to the nonlinear situation.

Currently, only rudimentary results (mostly for special metabolic systems) on flux estimation and statistical analysis for carbon labelling experiments can be found in the literature. Three basic approaches can be distinguished:

1. Most authors [Walsh (1984), Jans (1989), Sharfstein (1994), Chauvin (1994), Rollin (1995)] derived explicit formulas for flux determination based on only a single or a few metabolites as has been done in Part I for a simple example. This approach is well suited for solving the general identifiability problem [Wiechert (1995a), Wiechert (1995b)], i.e. for deciding whether the available data contains sufficient information for the determination of all unknown intracellular fluxes. Its main disadvantage is that redundant measurement information cannot be used for improving the statistical quality of the estimated fluxes. Moreover, it turns out that for complex networks the algebraic complexity of the involved bilinear balance equations is much too high for the application of general algebraic solution algorithms [Wiechert (1995b)]. In particular when bidirectional steps are incorporated into the model an explicit solution that may have been worked out for the unidirectional case will be much more complicated if not even impossible in the bidirectional case.

2. The graphical technique of contour plot superposition [Zupke (1994)] already used in Part I for representation purposes enables the uniqueness and well-determinedness of the flux estimate to be quickly decided and the statistical quality of the estimates can be judged in a simple way by graphing sensitivities and confidence regions (see below). However, this method is restricted to low-dimensional parameter spaces and is again not suited for the exploitation of redundant information contained in additional measurements.

3. A generally applicable numerical approach that can utilize redundant measurements and enables statistical quality measures to be derived as well is the familiar parameter fitting approach [Crawford (1983), Rabkin (1985), Chatham (1995), Marx (1996)]. However, it is well known that this approach may suffer from the existence of multiple solutions or ill determined estimates due to non-identifiable parameters. In Part I an example from the cyclic pentose phosphate pathway was given, for which two alternative flux solutions can be found (cf. [Zupke (1994)] for another example).

From these approaches only the last one can be universally applied to any carbon labelling experiment because it does not require any model-specific work. Moreover (as is explained below) this method can always accompanied by linearized statistical analysis in a canonical way. Finally, the existence of possible multiple solutions can in practice be detected by a multiple offset of the parameter fitting algorithm (see below). For these reasons we have pursued a parameter fitting approach. The reader interested in general algebraic approaches for explicit flux determination and identifiability analysis is referred to [Wiechert (1995a), Wiechert (1995b)].

Aims of this contribution

Our aim was to extend the well established computational and statistical theory for linear metabolite flux balancing [Vallino (1991), van Heijden (1994a), van Heijden (1994b)] to the more general non-linear case of carbon labelling systems based on the general model described in Part I. The result is a universal method for metabolic flux estimation and statistical analysis by stationary carbon isotope labelling experiments. After a comprehensive summary of the general model equations from Part I this contribution concentrates on the following aspects:

i) The development of a powerful generally applicable algorithm for flux estimation that can be applied to any metabolic network, respects all imposed flux constraints and is numerically stable when large exchange fluxes occur (cf. Part I).

ii) The exemplification and in depth discussion of the nonlinear statistical problems caused by measurement errors leading to a powerful generally applicable nonlinear transformation method for estimating confidence regions.

iii) The demonstration of the power and limitations of the statistical methods using a complex application example concerned with lysine producing *Corynebacterium glutamicum*. As in Part I the main emphasis will be on the question to what extent exchange fluxes can be estimated from the given data set.

iv) The implementation of all described algorithms within a flexible software framework as documented in Appendix B.

This distinguishes the current approach from formerly applied flux estimation methods including the preliminary parameter fitting approach applied in [Marx (1996)]. Another important aspect of the availability of redundant measurements is the possibility to improve the statistical quality of the computed flux estimate, to test the models' ability to describe the measured data, to detect gross measurement errors and to perform validation studies by comparing different models [Wang (1983), Vallino (1991), van Heijden (1994a), van Heijden (1994b)]. These techniques will be shortly addressed although we cannot elaborate on them in this context.

Summary of model equations

The introduced symbols and general model equations derived in Part I are summarized below to supply all structures required for the solution of the flux determination problem.

Transformations, balances and constraints

1. The state variables are given by the following vectors (where both flux vectors $\mathbf{v}^{\rightarrow}, \mathbf{v}^{\leftarrow}$ have the same dimension):

fractional labelling state of all enumerated carbon pools	$\mathbf{x}$	$[\%]$
labelling state of the enumerated input carbon atoms	$\mathbf{x}^{\text{inp}}$	$[\%]$
enumerated forward metabolite fluxes	$\mathbf{v}^{\rightarrow}$	$[\text{mol}/(\text{h} \cdot \text{g DW})]$
enumerated backward metabolite fluxes	$\mathbf{v}^{\leftarrow}$	$[\text{mol}/(\text{h} \cdot \text{g DW})]$

2. Fluxes are alternatively represented within three different coordinate systems by

natural flux coordinates	as forward and backward fluxes	$\mathbf{v}^{\rightarrow}, \mathbf{v}^{\leftarrow}$
application flux coordinates	as net and exchange fluxes	$\mathbf{v}^{\text{net}}$,$\mathbf{v}^{\text{xch}}$
numerical flux coordinates	as net and $[0,1]$-rescaled exchange fluxes	$\mathbf{v}^{\text{net}}$,$\mathbf{v}^{\text{xch[0,1]}}$

where all flux coordinate vectors $\mathbf{v}^{\rightarrow}$, $\mathbf{v}^{\leftarrow}$, $\mathbf{v}^{\text{net}}$, $\mathbf{v}^{\text{xch}}$ have the same dimension and physical unit while $\mathbf{v}^{\text{xch[0,1]}}$ is dimensionless. The respective coordinate transformations defining these alternative coordinate systems are given by

$$\Phi : \begin{pmatrix} \mathbf{v}^{\text{net}} \\ \mathbf{v}^{\text{xch}} \end{pmatrix} \longrightarrow \begin{pmatrix} \mathbf{v}^{\rightarrow} \\ \mathbf{v}^{\leftarrow} \end{pmatrix} = \begin{pmatrix} \mathbf{v}^{\text{xch}} - \min\left(-\mathbf{v}^{\text{net}}, \mathbf{0}\right) \\ \mathbf{v}^{\text{xch}} - \min\left(\ \mathbf{v}^{\text{net}}, \mathbf{0}\right) \end{pmatrix} \tag{1}$$

where $\mathbf{0}$ is the zero vector and the minimum has to be taken component-wise and the compactification map

$$\Phi_\beta^{[0,1]} : \begin{pmatrix} \mathbf{v}^{\text{net}} \\ \mathbf{v}^{\text{xch}[0,1]} \end{pmatrix} \longrightarrow \begin{pmatrix} \mathbf{v}^{\text{net}} \\ \mathbf{v}^{\text{xch}} \end{pmatrix} = \begin{pmatrix} \mathbf{v}^{\text{net}} \\ \beta \cdot \mathbf{v}^{\text{xch}[0,1]} / (\mathbf{1} - \mathbf{v}^{\text{xch}[0,1]}) \end{pmatrix} \quad . \tag{2}$$

Herein all entries of the vector $\mathbf{1}$ are 1, β is a fixed constant (a good choice being the rate of substrate uptake into the system) and the vector division has to be taken component-wise.

3. The carbon label flux balance equations exhibit the general structure

$$\left(\sum_i \mathbf{v}_i^{\rightarrow} \cdot \mathbf{P}_i^{\rightarrow} + \sum_i \mathbf{v}_i^{\leftarrow} \cdot \mathbf{P}_i^{\leftarrow} \right) \cdot \mathbf{x} + \left(\sum_i \mathbf{v}_i^{\rightarrow} \cdot \mathbf{P}_i^{\text{inp}} \right) \cdot \mathbf{x}^{\text{inp}} = \mathbf{0} \tag{3}$$

with the square $\dim \mathbf{x} \times \dim \mathbf{x}$ atom transition coefficient matrices $\mathbf{P}_i^{\rightarrow}, \mathbf{P}_i^{\leftarrow}, i = 1, ..., \dim \mathbf{v}^{\rightarrow}$, and the $\dim \mathbf{x} \times \dim \mathbf{x}^{\text{inp}}$ input atom transition matrices $\mathbf{P}_i^{\text{inp}}, i = 1, ..., \dim \mathbf{v}^{\rightarrow}$.

4. Various assumptions about metabolite fluxes like metabolic stationarity, unidirectionality of reaction steps or rapid equilibria caused by large exchange fluxes are expressed by two linear flux constraint equations

$$\begin{array}{lllll} & \mathbf{N}^{\text{net}} & \cdot \mathbf{v}^{\text{net}} & = & \mathbf{n}^{\text{net}} \\ \text{and} & \mathbf{N}^{\text{xch}[0,1]} & \cdot \mathbf{v}^{\text{xch}[0,1]} & = & \mathbf{n}^{\text{xch}[0,1]} \end{array} \quad , \tag{4}$$

where the $\dim \mathbf{n}^{\text{net}} \times \dim \mathbf{v}^{\text{net}}$ net flux constraint matrix $\mathbf{N}^{\text{net}}$, the $\dim \mathbf{n}^{\text{xch}[0,1]} \times \dim \mathbf{v}^{\text{xch}[0,1]}$ exchange flux constraint matrix $\mathbf{N}^{\text{xch}[0,1]}$ and the constraint value vectors $\mathbf{n}^{\text{net}}, \mathbf{n}^{\text{xch}[0,1]}$ are given and fixed.

5. When redundancies of equations are excluded (which should always be the case for a correct model formulation) the Equations (4) leave $\dim \mathbf{v}^{\text{net}} + \dim \mathbf{v}^{\text{xch}[0,1]} - \dim \mathbf{n}^{\text{net}} - \dim \mathbf{n}^{\text{xch}[0,1]}$ degrees of freedom for the determination of all flux coordinates. The free flux coordinates representing these degrees of freedom are defined by an arbitrarily given third constraint equation

$$\mathbf{N}^{\text{free}} \cdot \begin{pmatrix} \mathbf{v}^{\text{net}} \\ \mathbf{v}^{\text{xch}[0,1]} \end{pmatrix} = \mathbf{n}^{\text{free}} \quad . \tag{5}$$

The combined constraint matrix and the combined constraint value vector are then defined by

$$\mathbf{N} = \begin{pmatrix} \mathbf{N}^{\text{net}} & \mathbf{0} \\ \mathbf{0} & \mathbf{N}^{\text{xch}[0,1]} \\ & \mathbf{N}^{\text{free}} \end{pmatrix} \quad \text{and} \quad \mathbf{n} = \begin{pmatrix} \mathbf{n}^{\text{net}} \\ \mathbf{n}^{\text{xch}[0,1]} \\ \mathbf{n}^{\text{free}} \end{pmatrix} \quad .$$

The choice of $\mathbf{N}^{\text{free}}$ is only restricted by the requirement that $\mathbf{N}$ is square and invertible. By combining Equations (4) and (5) we then get the overall linear constraint equation than can be formulated within different coordinate systems by

$$\mathbf{N} \cdot \Phi^{-1} \circ (\Phi_\beta^{[0,1]})^{-1} \begin{pmatrix} \mathbf{v}^{\rightarrow} \\ \mathbf{v}^{\leftarrow} \end{pmatrix} = \mathbf{N} \cdot \begin{pmatrix} \mathbf{v}^{\text{net}} \\ \mathbf{v}^{\text{xch}[0,1]} \end{pmatrix} = \mathbf{n} \quad . \tag{6}$$

Notice that this equation is linear in the $(\mathbf{v}^{\text{net}}, \mathbf{v}^{\text{xch}[0,1]})$ coordinate system but non-linear with respect to $(\mathbf{v}^{\rightarrow}, \mathbf{v}^{\leftarrow})$.

6. Similarly a combined inequality constraint equation for specifying range restrictions or directionality assumptions is given by

$$\mathbf{U} \cdot \Phi^{-1} \circ (\Phi_\beta^{[0,1]})^{-1} \begin{pmatrix} \mathbf{v}^{\rightarrow} \\ \mathbf{v}^{\leftarrow} \end{pmatrix} = \mathbf{U} \cdot \begin{pmatrix} \mathbf{v}^{\text{net}} \\ \mathbf{v}^{\text{xch}[0,1]} \end{pmatrix} \geq \mathbf{u} \tag{7}$$

with a $\dim \mathbf{u} \times (\dim \mathbf{v}^{\text{net}} + \dim \mathbf{v}^{\text{xch}[0,1]})$ matrix $\mathbf{U}$ and the inequality value vector $\mathbf{u}$. The fluxes $\mathbf{v}^{\text{net}}$, $\mathbf{v}^{\text{xch}[0,1]}$ (or $\mathbf{v}^{\rightarrow}$, $\mathbf{v}^{\leftarrow}$ respectively) satisfying these constraints are called feasible fluxes.

7. In this contribution the flux and label measurement equations given by Equations (14) and (15) will be added to the model equations.

Derived equations

Two important results that have been proven in Part I will be used in the following:

1. The labelling state $\mathbf{x}$ turns out to be a function Γ of the natural flux state obtained by solving Equation (3):

$$\mathbf{x} = \Gamma \begin{pmatrix} \mathbf{v}^{\rightarrow} \\ \mathbf{v}^{\leftarrow} \end{pmatrix} = -\left(\sum_i \mathbf{v}_i^{\rightarrow} \cdot \mathbf{P}_i^{\rightarrow} + \sum_i \mathbf{v}_i^{\leftarrow} \cdot \mathbf{P}_i^{\leftarrow} \right)^{-1} \cdot \left(\sum_i \mathbf{v}_i^{\rightarrow} \cdot \mathbf{P}_i^{\text{inp}} \right) \cdot \mathbf{x}^{\text{inp}} \tag{8}$$

2. There exists a matrix $\mathbf{K}^{\text{free}}$ and a vector $\mathbf{k}^{\text{free}}$ such that the solution of Equation (6) as a linear function Ψ of $\mathbf{n}^{\text{free}}$ is given by (see Equation (21) in Part I)

$$\begin{pmatrix} \mathbf{v}^{\rightarrow} \\ \mathbf{v}^{\leftarrow} \end{pmatrix} = \Psi\left(\mathbf{n}^{\text{free}}\right) = \mathbf{K}^{\text{free}} \cdot \mathbf{n}^{\text{free}} + \mathbf{k}^{\text{free}} \quad . \tag{9}$$

A Simple Example

The simple example network used in Part I to introduce the model equations will also serve as an example in this section to illustrate the statistical problems of flux determination and confidence region estimation from measurements. The main emphasis of this section is to demonstrate the strong non-linear statistical effects that are caused by large exchange fluxes and their mathematical treatment by non-linear rescaling.

Example network

Using the formal notation introduced in Part I the example network structure with the corresponding carbon atom transitions is given by

V_1:	A	$\rightarrow$	B	V_3:	B	$\rightarrow$	C
	#uv	$\rightarrow$	#uv ;		#wx	$\rightarrow$	#xw ;
V_2:	B	$\rightarrow$	C	V_4:	C	$\rightarrow$	D
	#wx	$\rightarrow$	#wx ;		#yz	$\rightarrow$	#yz ;

.

i.e. V_3 reverses the position of the two carbon atoms B_1 and B_2 involved while all other steps leave positions unchanged. A graphical network representation is shown in Part I. We then have $\mathbf{x} = (b_1, b_2, c_1, c_2)^T$, $\mathbf{v}^{\rightarrow} = (v_1^{\rightarrow}, v_2^{\rightarrow}, v_3^{\rightarrow}, v_4^{\rightarrow})$ and $\mathbf{v}^{\leftarrow} = (v_1^{\leftarrow}, v_2^{\leftarrow}, v_3^{\leftarrow}, v_4^{\leftarrow})$. The system matrix

$\sum_{i=1,..,4} \mathbf{v}_i^{\rightarrow} \cdot \mathbf{P}_i^{\rightarrow} + \sum_{i=1,..,4} \mathbf{v}_i^{\leftarrow} \cdot \mathbf{P}_i^{\leftarrow}$ and the label input matrix $\sum_{i=1,..,4} \mathbf{v}_i^{\rightarrow} \cdot \mathbf{P}_i^{\text{inp}}$ for carbon flux balancing is consequently given by

$$\begin{pmatrix} -v_2^{\rightarrow} - v_3^{\rightarrow} & . & v_2^{\leftarrow} & . \\ . & -v_2^{\rightarrow} - v_3^{\rightarrow} & . & v_2^{\leftarrow} \\ v_2^{\rightarrow} & v_3^{\rightarrow} & -v_2^{\leftarrow} - v_4^{\rightarrow} & . \\ v_3^{\rightarrow} & v_2^{\rightarrow} & . & -v_2^{\leftarrow} - v_4^{\rightarrow} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} v_1^{\rightarrow} & . \\ . & v_1^{\rightarrow} \\ . & . \\ . & . \end{pmatrix}$$

while the metabolite flux balances, unidirectionality of V_1, V_3, V_4 and the values for the chosen free flux coordinates $v_1^{\text{net}}, v_2^{\text{net}}, v_2^{\text{xch}[0,1]}$ are expressed by the combined constraint equation

$$\left(\begin{array}{cccc|cccc} 1 & -1 & -1 & . & . & . & . & . \\ . & 1 & 1 & -1 & . & . & . & . \\ \hline . & . & . & . & 1 & . & . & . \\ . & . & . & . & . & . & 1 & . \\ . & . & . & . & . & . & . & 1 \\ \hline 1 & . & . & . & . & . & . & . \\ . & 1 & . & . & . & . & . & . \\ . & . & . & . & . & 1 & . & . \end{array}\right) \cdot \left(\begin{array}{c} v_1^{\text{net}} \\ v_2^{\text{net}} \\ v_3^{\text{net}} \\ v_4^{\text{net}} \\ \hline v_1^{\text{xch}[0,1]} \\ v_2^{\text{xch}[0,1]} \\ v_3^{\text{xch}[0,1]} \\ v_4^{\text{xch}[0,1]} \end{array}\right) = \left(\begin{array}{c} 0 \\ 0 \\ \hline 0 \\ 0 \\ 0 \\ \hline c_1^{\text{net}} \\ c_2^{\text{net}} \\ c_2^{\text{xch}[0,1]} \end{array}\right)$$

This leaves three degrees of freedom for simulation, represented by the values of $c_1^{\text{net}}, c_2^{\text{net}}, c_2^{\text{xch}[0,1]}$.

Flux determination and exact confidence intervals

Suppose that the net flux v_1^{net} and the fractional labels b_1, c_1 have been measured. The aim of flux determination is the determination of values for the free fluxes $v_1^{\text{net}}, v_2^{\text{net}}, v_2^{\text{xch}[0,1]}$ that reproduce these measured data. Assuming for simplicity that the substrate uptake $v_1^{\rightarrow}$ has been directly measured as $v_1^{\rightarrow} = 1$ with good precision (cf. the analysis in Part I) we can concentrate on the relation between $(v_2^{\rightarrow}, v_2^{\leftarrow})$ and (b_1, c_1). From Part I the free fluxes $v_2^{\text{net}}, v_2^{\text{xch}}$ corresponding to given values of b_1, c_1 are computed as

$$\begin{aligned} (v_2^{\rightarrow}, v_2^{\leftarrow}) &= \gamma^{-1}(b_1, c_1) &&= \left(-\tfrac{(b_1+c_1-1)(1-c_1)}{(2b_1-1)(c_1-b_1)}, \quad -\tfrac{1-b_1}{c_1-b_1} \right) \\ (v_2^{\text{net}}, v_2^{\text{xch}}) &= \phi^{-1}(v_2^{\rightarrow}, v_2^{\leftarrow}) &&= \left(v_2^{\rightarrow} - v_2^{\leftarrow}, \ \min(v_2^{\rightarrow}, v_2^{\leftarrow}) \right) \end{aligned}$$

which gives rise to the one-to-one mappings γ and ϕ.

In practice only noisy measurements of b_1, c_1 are available. Denoting the measurement errors by $\varepsilon_b, \varepsilon_c$ respectively, the estimated fluxes based on noisy measurement data are given by

$$(\hat{v}_2^{\text{net}}, \hat{v}_2^{\text{xch}}) = \phi^{-1} \circ \gamma^{-1}(b_1 + \varepsilon_b, c_1 + \varepsilon_c) \quad . \tag{10}$$

Now the usual assumption of independent normally distributed errors with expectation 0 and variance σ^2 is made, i.e. $(\varepsilon_b, \varepsilon_c) \in N(\mathbf{0}, \sigma^2 \cdot \mathbf{1})$ (with $\mathbf{1}$ denoting the unit matrix). Given some specified confidence level α we consider the circle $C_{1-\alpha}(b_1, c_1)$ centred at (b_1, c_1) with radius $\sqrt{\chi_2^2(1-\alpha)}$ (where χ_n^2 denotes the chi-square distribution with n degrees of freedom):

$$C_{1-\alpha}(b_1, c_1) = \{(b_1 + \varepsilon_b, c_1 + \varepsilon_c) \,|\, \varepsilon_b^2 + \varepsilon_c^2 \leq \chi_2^2(1-\alpha)\}$$

By definition of χ_n^2 the probability of the event that $(b_1 + \varepsilon_b, c_1 + \varepsilon_c)$ lies in this circle is exactly [Arnold (1990)]

$$P\left[(b_1 + \varepsilon_b, c_1 + \varepsilon_c) \in C_{1-\alpha}(b_1, c_1)\right] = 1 - \alpha \quad .$$

Applying the mapping $\phi^{-1} \circ \gamma^{-1}$ on both sides of the $\in$-symbol and using Equation (10) we get

$$P\left[(\hat{v}_2^{\text{net}}, \hat{v}_2^{\text{xch}}) \in \phi^{-1} \circ \gamma^{-1}\left(C_{1-\alpha}\left(b_1, c_1\right)\right)\right] = 1 - \alpha \quad ,$$

i.e. the circles' pre-image under $\phi^{-1} \circ \gamma^{-1}$ in the $(v_2^{\text{net}}, v_2^{\text{xch}})$ space represents an exact $(1-\alpha)$-confidence region for the estimated parameters:

$$\text{Conf}_{1-\alpha}(v_2^{\text{net}}, v_2^{\text{xch}}) = \phi^{-1} \circ \gamma^{-1}\left(C_{1-\alpha}\left(b_1, c_1\right)\right)$$

Clearly, in practice the computation of confidence regions must be performed around the estimated fluxes $(\hat{v}_2^{\text{net}}, \hat{v}_2^{\text{xch}})$ because the true values are not available. However, all arguments remain valid when the circles are centred at the measured values $(b_1 + \varepsilon_b, c_1 + \varepsilon_c)$.

Properties of the exact confidence intervals

Figure 1a shows some representatively chosen circles $C_{1-\alpha}\left(b_1, c_1\right)$ for $1-\alpha$ =90% and $\sigma^2 = 0.008^2$ (i.e. a 0.8 % measurement error) in the (b_1, c_1) plane. Figure 1b then illustrates how these circles are mapped onto the $(v_2^{\text{net}}, v_2^{\text{xch}})$ plane by $\phi^{-1} \circ \gamma^{-1}$. A close inspection of these exact confidence regions $\text{Conf}_{1-\alpha}(v_2^{\text{net}}, v_2^{\text{xch}})$ leads to the following conclusions:

1. The exact confidence regions are irregularly shaped so that for higher-dimensional systems their geometric description will become increasingly difficult.

2. In most cases the net flux v_2^{net} is estimated within reasonable tolerance. In the worst case being a simultaneously large exchange flux v_2^{xch} it is still much better determined than the corresponding exchange flux.

3. Only small exchange fluxes v_2^{xch} can be determined with narrow confidence intervals from measurements while large exchanges are highly sensitive with respect to measurement errors.

4. For large exchange fluxes the exact confidence interval may include the possibility of an infinitely large flux which will lead to severe problems for a general numerical solution.

Summarizing, exchange fluxes can in most situations only be estimated within an order of magnitude while net fluxes remain always (comparatively) better determined. However, the tolerances may still be sufficient to distinguish large exchange fluxes from small ones. This is valuable information for metabolic modelling.

Linearized statistics

Due to the generally irregular shape of the exact non-linear confidence regions an approximation has to be found that comes close to the exact region and at the same time can be described with a small set of practically meaningful parameters. A similar situation is known from the characterization of probability distributions where expectations and covariances are often taken as characteristic parameters while higher order moments are neglected.

The usual approach for obtaining an approximation to $\text{Conf}_{1-\alpha}(v_2^{\text{net}}, v_2^{\text{xch}})$ is by computing a linearization of $\phi^{-1} \circ \gamma^{-1}$ around (b_1, c_1), i.e.

$$\text{Lin}_{(b_1,c_1)}^{\phi^{-1}\circ\gamma^{-1}}\left(b_1 + \varepsilon_b, c_1 + \varepsilon_c\right) = \phi^{-1} \circ \gamma^{-1}\left(b_1, c_1\right) + \frac{\partial\left(\phi^{-1} \circ \gamma^{-1}\right)}{\partial\left(b_1, c_1\right)}\left(b_1, c_1\right) \cdot \begin{pmatrix} \varepsilon_b \\ \varepsilon_c \end{pmatrix} \quad . \quad (11)$$

Replacing the original non-linear mapping $\phi^{-1} \circ \gamma^{-1}$ by this approximation the exact confidence region around $(v_2^{\text{net}}, v_2^{\text{xch}})$ can be approximated by the elliptical region:

$$\text{Conf}_{1-\alpha}(v_2^{\text{net}}, v_2^{\text{xch}}) \approx \text{Lin}_{(b_1,c_1)}^{\phi^{-1}\circ\gamma^{-1}}\left(C_{1-\alpha}\left(b_1, c_1\right)\right) \tag{12}$$

Since ellipsoids can be conveniently described by their centre and their principal axes [Press (1988)] this approximation can be practically interpreted even for higher-dimensional models.

Figure 1c shows the results for the circles of Figure 1a. It turns out that the approximation is practically useless in case of large exchange fluxes v_2^{xch}. Compared with the exact confidence regions the approximating ellipses are then displaced towards low exchange fluxes. For practical application this result is disappointing because the linearization leads to a significantly reduced ability to distinguish large exchange fluxes from small ones.

To explain the mathematical reason for this poor result we reduce the dimension by fixing the rather well determined flux v_2^{net} to its estimated value and focus upon the variation of v_2^{xch}. Figure 2a then illustrates the dependency of c_1 on v_2^{xch} for fixed v_2^{net}. As has already been pointed out in Part I, the values of b_1, c_1 tend to a limit value for large exchange fluxes. Consequently, the labelling state becomes more and more insensitive with respect to v_2^{xch} when $v_2^{\text{xch}} \to \infty$. The tangent line to the curves (i.e. the linearization) in Figure 2a then tends to a horizontal line. This line is an extremely poor global approximation to the original curve over the whole value range $[0, \infty]$.

Thus the reason for the observed poor approximation results is that the linear approximation of the function $\phi^{-1} \circ \gamma^{-1}$ given by Equation (11) is not able to follow the curvature of the original function well enough. In general, from the theory of non-linear statistics it is known [Pázman (1993)] that the approximation quality of a linearization depends on the curvature of the linearized function. Functions with low curvature (so called "flat models" [Pázman (1993)]) tend to exhibit good linearization results.

Non-linear rescaling

In order to obtain a better approximation for the exact confidence region $\text{Conf}_{1-\alpha}(v_2^{\text{net}}, v_2^{\text{xch}})$ a non-linear transformation of the mapping $\phi^{-1} \circ \gamma^{-1}$ must be found that reduces the curvature and thus is more suitable for linearization. Our solution is given by the compactification mapping between the application flux coordinate $v_2^{\text{xch}} \in [0, \infty]$ and the numerical flux coordinate $v_2^{\text{xch[0,1]}} \in [0, 1]$ that has already been introduced in Part I and Equation (2). Using $\beta = v_1^{\text{net}} = 1$ (as in Part I) a flux estimate in the new coordinate system is obtained by

$$\begin{aligned}(\hat{v}_2^{\text{net}}, \hat{v}_2^{\text{xch[0,1]}}) =& (\phi_\beta^{[0,1]})^{-1} \circ \phi^{-1} \circ \gamma^{-1}\left(b_1 + \varepsilon_b, c_1 + \varepsilon_c\right) \\ \text{with} \qquad & (\phi_\beta^{[0,1]})^{-1}\left(v_2^{\text{net}}, v_2^{\text{xch}}\right) = \left(v_2^{\text{net}}, v_2^{\text{xch}}/(1 + v_2^{\text{xch}})\right)\end{aligned} \quad .$$

The dependency of c_1 on $v_2^{\text{xch[0,1]}} \in [0, 1]$ for fixed v_2^{net} is shown in Figure 2b. Obviously this mapping has an extremely low curvature and thus is much better linearizable than that shown in Figure 2a. Consequently, the tangent line to this curve will produce a good approximation over the complete value range $[0, 1]$ of v_2^{xch}.

Of course, the approximated elliptical confidence region based on the linearization $\text{Lin}_{(b_1,c_1)}^{(\phi_\beta^{[0,1]})^{-1}\circ\phi^{-1}\circ\gamma^{-1}}$ (computed analogously to Equations (11) and (12)) is situated in the $(v_2^{\text{net}}, v_2^{\text{xch[0,1]}})$ space. In order to compare the quality of this approximation with that of Equation (12) this elliptical confidence region must be transformed back to the $(v_2^{\text{net}}, v_2^{\text{xch}})$ space via $\phi_\beta^{[0,1]}$. The result shown in Figure 1d turns out to correspond extremely well with the exact confidence region $\text{Conf}_{1-\alpha}(v_2^{\text{net}}, v_2^{\text{xch}})$

for all circles in Figure 1a! Therefore the non-linear coordinate change to the numerical flux coordinates enables an almost exact region to be computed that is defined by the characteristic geometric parameters of an ellipse together with the known transformation $\phi_\beta^{[0,1]}$.

In many application cases it is sufficient to compute approximative confidence intervals for single parameters only. Assuming that a confidence interval $[\hat{v}_2^{\text{xch}[0,1]} - \Delta, \hat{v}_2^{\text{xch}[0,1]} + \Delta]$ has been computed in the $(v_2^{\text{net}}, v_2^{\text{xch}[0,1]})$ space using the methods presented below it is easily transformed back via $\phi_\beta^{[0,1]}$ to the unsymmetrical interval

$$\left[\beta \cdot \frac{\hat{v}_2^{\text{xch}[0,1]} - \Delta}{1 - (\hat{v}_2^{\text{xch}[0,1]} - \Delta)}, \beta \cdot \frac{\hat{v}_2^{\text{xch}[0,1]} + \Delta}{1 - (\hat{v}_2^{\text{xch}[0,1]} + \Delta)}\right] \quad . \tag{13}$$

This should at least enable the order of magnitude of v_2^{xch} to be determined. The general message of the studied example is that flux estimation as well as the computation of statistical quality measures should always be performed in a suitable coordinate system like $(\mathbf{v}^{\text{net}}, \mathbf{v}^{\text{xch}[0,1]})$. An interpretation in the $(\mathbf{v}^{\text{net}}, \mathbf{v}^{\text{xch}})$ system (which is more convenient for practical purposes) can then be obtained by back-transformation using $\phi_\beta^{[0,1]}$.

Flux Estimation

The main difference between the general situation and the example is that redundant measurement information may now be available. The determination of the intracellular fluxes then poses an inverse problem associated with the general model Equations (3), (6), (7) that is usually solved by a least squares approach [Bates (1988), Seber (1989)].

The general theory of flux estimation and the computation of statistical quality measures is developed in this and the following section. For that purpose all concepts presented in the preceding section have to be generalized appropriately. The mathematical foundation for this generalization is given by the fact that the compactification method developed for the simple example can be readily taken over to the general case. The mathematically rather involved proof can be taken from [Wiechert (1995c)].

Measurement equations

The relation between the fluxes and labelling fractions predicted by the model and their noisy measurements is given by the measurement equations that we develop now. To start with, two measurement matrices $\mathbf{M_w}$ (for net fluxes) and $\mathbf{M_y}$ (for labels) are introduced that indicate which coordinate entries of $\mathbf{v}^{\text{net}}$ and $\mathbf{x}$ are actually measured.

In the example the extracellular net flux v_1^{net} and the fractional labels b_1, c_1 (enumerated as $\mathbf{x}_1$ and $\mathbf{x}_3$) are assumed to be measured, which is expressed by the matrices

$$\mathbf{M_w} = \begin{pmatrix} 1 & . & . & . \end{pmatrix} \quad \text{and} \quad \mathbf{M_y} = \begin{pmatrix} 1 & . & . & . \\ . & . & 1 & . \end{pmatrix} \quad .$$

If no measurement noise is present the resulting vectors of measured net fluxes and measured label fractions are given by $\mathbf{w} = \mathbf{M_w} \cdot \mathbf{v}^{\text{net}}$ and $\mathbf{y} = \mathbf{M_y} \cdot \mathbf{x}$. Denoting the corresponding measurement noise vectors by $\varepsilon_{\mathbf{w}}$ and $\varepsilon_{\mathbf{y}}$ and keeping in mind that $\mathbf{v}^{\text{net}} = \mathbf{v}^{\rightarrow} - \mathbf{v}^{\leftarrow}$ we end up

with two measurement equations:

$$\begin{array}{lrcccccl} \text{the flux measurement equation} & \mathbf{w} & = & \mathbf{M_w} & \cdot & (\mathbf{v}^{\rightarrow} - \mathbf{v}^{\leftarrow}) & + & \varepsilon_{\mathbf{w}} \\ & & = & (\mathbf{M_w}, -\mathbf{M_w}) & \cdot & \begin{pmatrix} \mathbf{v}^{\rightarrow} \\ \mathbf{v}^{\leftarrow} \end{pmatrix} & + & \varepsilon_{\mathbf{w}} \\ \text{the label measurement equation} & \mathbf{y} & = & \mathbf{M_y} & \cdot & \mathbf{x} & + & \varepsilon_{\mathbf{y}} \end{array} \tag{14}$$

Additionally, the statistical properties of the measurements have to be expressed. The usual assumption is that the noise terms $\varepsilon_{\mathbf{w}}, \varepsilon_{\mathbf{y}}$ are normally distributed with expectation vector $\mathbf{0}$ and covariance matrices $\mathbf{\Sigma_w}, \mathbf{\Sigma_y}$:

$$\begin{array}{rcl} \varepsilon_{\mathbf{w}} & = & N(\mathbf{0}, \mathbf{\Sigma_w}) \\ \varepsilon_{\mathbf{y}} & = & N(\mathbf{0}, \mathbf{\Sigma_y}) \end{array} \tag{15}$$

In the example the measurements were assumed to be independently distributed with the same variance σ^2 for the label measurements and some other variance τ^2 for the flux measurement. The associated covariance matrices are:

$$\mathbf{\Sigma_w} = (\tau^2) \quad \text{and} \quad \mathbf{\Sigma_y} = \begin{pmatrix} \sigma^2 & . \\ . & \sigma^2 \end{pmatrix}$$

It should be noticed that the diagonal shape need not always be the case because the measurements can exhibit correlations which may not be negligible. An in depth discussion of this problem and the choice of appropriate covariance matrices can be found in [Wiechert (1995c)].

Sum of squares function

The discrepancy between the system state predicted by the model and the measured values is quantitated by the familiar sum of squares function. In order to obtain a proper weighting of measurement errors the covariance matrices must be incorporated within this function by using a squared weighted norm that is defined now.

Since a covariance matrix $\mathbf{\Sigma}$ is always square, symmetric, and positively definite, its square root $\sqrt{\mathbf{\Sigma}}$ can be computed (e.g. by using a Cholesky factorization [Horn (1985)]) satisfying the condition $\sqrt{\mathbf{\Sigma}}^T \cdot \sqrt{\mathbf{\Sigma}} = \mathbf{\Sigma}$. The squared weighted norm associated with a covariance matrix $\mathbf{\Sigma}$ enables a measurement error vector ε to be appropriately weighted as:

$$||\varepsilon||^2_{\mathbf{\Sigma}} = \varepsilon^T \cdot \mathbf{\Sigma}^{-1} \cdot \varepsilon = (\sqrt{\mathbf{\Sigma}}^{-1} \varepsilon)^T \cdot (\sqrt{\mathbf{\Sigma}}^{-1} \varepsilon)$$

Using this notation the sum of squares function is given in terms of the natural flux variables $\mathbf{v}^{\rightarrow}, \mathbf{v}^{\leftarrow}$ and the labelling state $\mathbf{x}$ by:

$$\kappa(\mathbf{v}^{\rightarrow}, \mathbf{v}^{\leftarrow}, \mathbf{x}) = ||\mathbf{w} - \mathbf{M_w} \cdot (\mathbf{v}^{\rightarrow} - \mathbf{v}^{\leftarrow})||^2_{\mathbf{\Sigma_w}} + ||\mathbf{y} - \mathbf{M_y} \cdot \mathbf{x}||^2_{\mathbf{\Sigma_y}} \tag{16}$$

In the example we have the weighted sum of squares

$$\kappa(\mathbf{v}^{\rightarrow}, \mathbf{v}^{\leftarrow}, \mathbf{x}) = \left(\frac{\varepsilon_b}{\sigma}\right)^2 + \left(\frac{\varepsilon_c}{\sigma}\right)^2 + \left(\frac{\varepsilon_v}{\tau}\right)^2$$

where $\varepsilon_b, \varepsilon_c, \varepsilon_v$ denote the respective measurement errors for $b_1, c_1, v_1^{\text{net}}$.

Least squares estimation

The general least squares estimate (which equals the maximum likelihood estimate for normally distributed errors [Seber (1989)]) is now obtained by constrained minimization of the sum of squares function from Equation (16):

$$\begin{aligned} &\text{minimize } \kappa\left(\mathbf{v}^{\rightarrow}, \mathbf{v}^{\leftarrow}, \mathbf{x}\right) \\ &\text{subject to the constraints given by Equations (3), (6) and (7)} \end{aligned} \tag{17}$$

This poses a quadratic minimization problem with non-linear equality and inequality constraints.

Aiming at a simplified representation of this problem the mappings Ψ from Equation (9) and Γ from Equation (8) can be used. All equality constraints are thereby reduced by parametrizing their solution space with $\mathbf{n}^{\text{free}}$. Replacing the corresponding terms in Equation (14) the minimization problem from Equation (17) transforms to:

$$\begin{aligned} \text{minimize} \quad \kappa\left(\mathbf{n}^{\text{free}}\right) &= \left|\left|\mathbf{w} - \left(\mathbf{M}_{\mathbf{w}}, -\mathbf{M}_{\mathbf{w}}\right) \cdot \quad \Phi \circ \Phi_{\beta}^{[0,1]} \circ \Psi\left(\mathbf{n}^{\text{free}}\right)\right|\right|^2_{\boldsymbol{\Sigma}_{\mathbf{w}}} \\ &+ \left|\left|\mathbf{y} - \quad \mathbf{M}_{\mathbf{y}} \cdot \Gamma \circ \Phi \circ \Phi_{\beta}^{[0,1]} \circ \Psi\left(\mathbf{n}^{\text{free}}\right)\right|\right|^2_{\boldsymbol{\Sigma}_{\mathbf{y}}} \\ \text{subject to the inequality constraints} \quad & \mathbf{U} \cdot \Psi\left(\mathbf{n}^{\text{free}}\right) \geq \mathbf{u} \end{aligned} \tag{18}$$

Some details on the numerical solution of this linearly constrained non-linear minimization problem are given in Appendix B. Having computed an estimate $\hat{\mathbf{n}}^{\text{free}}$, the corresponding estimates for fluxes in the application coordinate system and the labelling state can be computed from

$$\begin{pmatrix} \hat{\mathbf{v}}^{\text{net}} \\ \hat{\mathbf{v}}^{\text{xch}} \end{pmatrix} = \Phi_{\beta}^{[0,1]} \circ \Psi\left(\mathbf{n}^{\text{free}}\right) \qquad \text{and} \qquad \hat{\mathbf{x}} = \Gamma \circ \Phi \begin{pmatrix} \hat{\mathbf{v}}^{\text{net}} \\ \hat{\mathbf{v}}^{\text{xch}} \end{pmatrix} \tag{19}$$

Statistical Analysis

Having computed $\hat{\mathbf{n}}^{\text{free}}$ its statistical quality must be judged. To this end statistical quality measures like the output and parameter sensitivities, the covariance matrix and parameter confidence regions must be derived. Since $\mathbf{n}^{\text{free}}$ is actually composed from coordinate entries of the numerical flux vectors the model linearization in the numerical flux coordinate space as described for the simple example will lead to good approximation results. Thus standard procedures based on linear regression theory [Pázman (1993)] are applicable.

Statistical notation

To obtain a very compact notation of the constrained least squares problem (18) we introduce the following new symbols:

1. The parameter vector $\mathbf{n}^{\text{free}}$ is replaced by Θ which in regression theory is the familiar symbol for the estimated parameter vector.

2. The combined measured output vector η and its covariance matrix $\boldsymbol{\Sigma}$ are represented by

$$\eta = \begin{pmatrix} \mathbf{w} \\ \mathbf{y} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{\mathbf{w}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{\mathbf{y}} \end{pmatrix} \quad .$$

3. The overall input-output function is then given by (cf. Equation (18))

$$F(\Theta) = \begin{pmatrix} (\mathbf{M_w}, -\mathbf{M_w}) \cdot & \Phi \circ \Phi_\beta^{[0,1]} \circ \Psi(\Theta) \\ \mathbf{M_y} \cdot \Gamma \circ & \Phi \circ \Phi_\beta^{[0,1]} \circ \Psi(\Theta) \end{pmatrix} \tag{20}$$

4. Finally with $\mathbf{A} = \mathbf{U} \cdot \mathbf{K}^{\text{free}}$ and $\mathbf{a} = \mathbf{u} - \mathbf{U} \cdot \mathbf{k}^{\text{free}}$ (cf. Equation 9) the inequality constraints are written as

$$\mathbf{A} \cdot \Theta \geq \mathbf{a} \quad . \tag{21}$$

With these abbreviations we end up with the general non-linear constrained regression model:

$$\eta = F(\Theta) + \varepsilon \quad \text{subject to} \quad \mathbf{A} \cdot \Theta \geq \mathbf{a} \quad \text{and} \quad \varepsilon \in N(\mathbf{0}, \mathbf{\Sigma}) \quad . \tag{22}$$

The corresponding least squares estimator is denoted as usual by $\hat{\Theta}$.

Statistical quality measures

The key to the computation of statistical quality measures for the estimated parameter $\hat{\Theta}$ is the evaluation of the model Jacobian matrix $\partial F / \partial \Theta(\hat{\Theta})$, i.e. the linear approximation of the non-linear model Equation (22) around $\hat{\Theta}$. This model linearization is well known as the (absolute) output sensitivity matrix of the system [Chatterjee (1988)]. The output sensitivity reveals how the measured state variables will be influenced by a differential change in Θ. More details concerning its computation can be taken from Appendix B. Having once linearized the model, all the results from linear statistical theory can be (approximately) applied [Pázman (1993)]. In particular we compute the following quantities:

1. For a proper linear approximation of the estimator's statistical properties the output sensitivities are weighted by the measurement covariance matrices to obtain the weighted output sensitivity matrix:

$$\text{Sens}_\Theta^{\mathbf{w},\mathbf{y}}(\hat{\Theta}) = \sqrt{\mathbf{\Sigma}}^{-1} \cdot \frac{\partial \mathbf{F}}{\partial \Theta}(\hat{\Theta})$$

2. From the weighted output sensitivity at $\hat{\Theta}$ the estimator's covariance matrix is approximated:

$$\text{Cov}(\hat{\Theta}) \approx \left[\text{Sens}_\Theta^{\mathbf{w},\mathbf{y}}(\hat{\Theta})^T \cdot \text{Sens}_\Theta^{\mathbf{w},\mathbf{y}}(\hat{\Theta}) \right]^{-1} \quad .$$

   In particular the parameter variance estimates are given by the diagonal vector, i.e. $\text{Var}(\hat{\Theta}) = \text{diag}\,\text{Cov}(\hat{\Theta})$.

3. Elliptical parameter confidence regions in the Θ-space for a given confidence level α can now be computed from $\text{Cov}(\hat{\Theta})$ by:

$$\text{Conf}_{1-\alpha}(\hat{\Theta}) \approx \{ \Theta \,|\, (\Theta - \hat{\Theta})^T \cdot \text{Cov}(\hat{\Theta})^{-1} \cdot (\Theta - \hat{\Theta}) \leq \chi^2_{\dim \Theta}(1-\alpha) \} \tag{23}$$

   A confidence interval for a single parameter $\hat{\Theta}_i$ is given by

$$[\hat{\Theta}_i - \Delta, \hat{\Theta}_i + \Delta] \quad \text{with} \quad \Delta = \chi^2_1(1-\alpha) \cdot \text{Var}(\hat{\Theta}_i) \quad . \tag{24}$$

4. Finally, the weighted parameter sensitivity matrix tells us how the parameter estimates are influenced by a change in some measured value. It is obtained by neglecting second order terms as

$$\text{Sens}_{\mathbf{w},\mathbf{y}}^{\hat{\Theta}} \approx \text{Cov}(\hat{\Theta}) \cdot \text{Sens}_{\mathbf{v}}^{\mathbf{w},\mathbf{y}}(\hat{\Theta})^T \, .$$

All formulas have been given for the free parameter estimator $\hat{\Theta}$ in the numerical flux coordinate system. Clearly, these results can be immediately transferred to results on $\hat{\mathbf{v}}^{\text{net}}, \hat{\mathbf{v}}^{\text{xch}}$ using Equation (19). This produces confidence intervals for all interesting parameters in the system. The numerically stable computation of these quantities by a singular value decomposition is described in [Press (1988)].

Analysis of the Central Metabolism of *Corynebacterium glutamicum*

The complex data set discussed below is taken from [Marx (1996)]. It consists of the directly measured extracellular fluxes and fractional labelling values from a continuous culture of *Corynebacterium glutamicum* under lysine-producing conditions. The underlying metabolic network with the flux names is presented in Figure 3a and the corresponding carbon atom transitions can be taken from Appendix A. An important addition compared to [Marx (1996)] is that the ^{13}C enrichments of the ribose-5-phosphate pool have meanwhile become available from the ribonucleotides isolated from RNA. The details of the preparation will be presented elsewhere.

A very important fact with respect to the investigation of bidirectional reaction steps is that [1-^{13}C] glucose was taken as a substrate. The labelled carbon atom is immediately split off as CO_2 in the oxidative pentose phosphate pathway. Thus when all reaction steps are assumed to be unidirectional no ^{13}C enrichment is expected in the intermediates of the non-oxidative pentose phosphate pathway. However high fractional labelling values were measured in erythrose-4-phosphate and the pentose-phosphate pools (see Table II). This is already evidence for the existence of significant exchange fluxes in the pentose phosphate pathway.

Whereas in the original publication [Marx (1996)] no statistical data were given, the newly developed software tools allowed the statistical quality of the flux estimates to be investigated for the first time. Moreover, the additionally available measurements were used for further improvement of the estimate's statistical quality. Since the focus of this contribution is on modelling and data analysis the biological implications of this experiment are not discussed here (see [Marx (1996)]).

Metabolic network and assumptions

The metabolic network model from Figure 3a developed by us has the following most important features:

- Some reaction steps are assumed to be unidirectional. Those steps which are assumed to be bidirectional are labelled with two values in Figure 3b.

- Since at this stage we cannot discriminate between fluxes involving the malate and the oxaloacetate pool, respectively, all anaplerotic carboxylation reactions are represented by one single bidirectional reaction step from PEP/Pyr to Mal/OAA. It must be pointed out that this local simplification does not influence the other flux estimates in the system as has been shown in [Wiechert (1995a)] by algebraic identifiability analysis.

- In contrast to [Marx (1996)] the pentose-phosphate pools (ribulose-5-phosphate, xylulose-5-phosphate, ribose-5-phosphate) are now merged together in one pool by rapid equilibrium assumptions (cf. [Wiechert (1996b)]). This is probably justified because, with the new ribose-5-phosphate measurements available, flux estimations showing high exchanges

between the three pentose-5-phosphate pools consistently led to the lowest sum-of-squares deviation.

- It should be noticed that lysine is produced via two parallel pathways in *Corynebacterium glutamicum* [Sonntag (1993)]. These are the diaminopimelate dehydrogenase pathway (ddh) and the succinyl-diaminopimelate dehydrogenase pathway (succ-dap) which are distinguished by the different fates of the carbon atoms.

Measured data and flux estimation

The substrate uptake, the biomass effluxes and the product formation of CO_2 and lysine were directly measured as presented in Table I. All measured fractional carbon enrichments can be taken from Table II.

The metabolite flux balances now leave three degrees of freedom for net fluxes. As the free net fluxes we have chosen:

ppp_1^{net}	in the oxidative pentose phosphate pathway
gc_1^{net}	in the glyoxylate cycle
lp_2^{net}	in the ddh lysine production pathway

The set of free fluxes is completed by all exchange fluxes of reaction steps that are assumed to be bidirectional, i.e.

$ppp_2^{\text{xch}}, ppp_3^{\text{xch}}, ppp_4^{\text{xch}}$	in the pentose phosphate pathway
$gly_1^{\text{xch}}, gly_3^{\text{xch}}$	in glycolysis
cac_4^{xch}	in the citric acid cycle
ac^{xch}	in the anaplerotic section

Based on this data extensive simulation studies and a repeated offset of the parameter-fitting algorithm led to a uniquely determined minimum of the least squares problem (18). The local optimality has been verified graphically and by by inspecting the computed gradient of $\kappa(\theta)$. Moreover, no constraints became active for the computed flux estimates as presented in Figure 3b, where all fluxes are normalized to a substrate uptake rate of 100 %.

In Table III the resulting estimates for the free fluxes together with their unsymmetrical 90% single parameter confidence intervals computed from Equations (24) and (13) are given. The parameter $\beta = 1.0$ (corresponding to 100 % substrate uptake) was used for the flux transformation $\Phi_\beta^{[0,1]}$. The variance estimates obtained for substrate uptake, product formation and biomass effluxes are omitted in Table III for the sake of brevity. As can be seen in Table I their estimated values are almost identical to the measured values considering the given measurement standard deviation. Moreover, each net flux estimate turned out to be almost uncorrelated to all other net flux estimates. For that reason the variances of all non-free net fluxes can be easily computed from Equation (19) because variances have simply to be wheighted with their squared linear coefficients and then summed up. For this reason all linearly dependent fluxes have confidence regions in the same order of magnitude as the involved free net fluxes. Because this paper concentrates on the statistically critical aspects we now focus on the analysis of the intracellular free fluxes.

Analysis of the covariance matrix

As becomes clear from Table III all free net fluxes are well determined from the measured data with the largest confidence interval in ppp_1^{net}. Of course the *absolute* size of the confidence rea-

gions has to be compared here because small fluxes (like gc_1^{net}) can in general not be estimated with smaller confidence regions than large fluxes. Contrastingly, the exchange fluxes can only be given within their order of magnitude. Compared with the corresponding net fluxes the exchange fluxes in Gly_1 and PPP_2 can be considered close to equilibrium while those in PPP_3, PPP_4 and AC are medium sized and Gly_3, CAC_4 is rather low.

The next question is whether there are correlations between the flux estimates, which of course has to be decided in the numerical flux coordinate system where the linearization of Equation (22) takes place. A detailed insight into these correlations can be achieved by an inspection of the parameter confidence ellipsoid as given by Equation (23). Its longest principal axes in the Θ-space are reproduced in Table IV. The axis length varies within three orders of magnitude with two exceptionally long axes. The longest axis is almost completely determined by the entries for $gly_1^{\text{xch}[0,1]}$ and ppp_1^{net}, i.e. these estimates are highly correlated. However, the variation in the $gly_1^{\text{xch}[0,1]}$ direction is much larger, which leads to the large confidence interval for $gly_1^{\text{xch}[0,1]}$ while that for ppp_1^{net} is substantially smaller. The second long axis reveals a similar dependency between $gly_3^{\text{xch}[0,1]}$ and $ac^{\text{xch}[0,1]}$ where now $gly_3^{\text{xch}[0,1]}$ is not well determined. All other axes are at least one order of magnitude shorter, i.e. not critical. A further group of correlated fluxes can be identified with $ppp_2^{\text{xch}[0,1]}, ppp_3^{\text{xch}[0,1]}, ppp_4^{\text{xch}[0,1]}$ (i.e. the pentose phosphate exchange fluxes).

The principal component analysis proves that the covariance matrix is not singular, i.e. the required information for a complete flux determination is in principle contained in the data set. However, from a statistical viewpoint this information is not sufficient to estimate $gly_1^{\text{xch}[0,1]}$ and $gly_3^{\text{xch}[0,1]}$ with good statistical quality.

Finally, it is of some interest to find out which measured values influence the estimates of those fluxes with the largest confidence intervals. A closer look at the parameter sensitivity matrix computed by Equation (23) (not shown here) reveals that $gly_1^{\text{xch}[0,1]}$ and ppp_1^{net} are both strongly influenced by $p5p_2, gap_1, gap_2$ while gly_3^{xch} and $ac^{\text{xch}[0,1]}$ are both influenced by gap_2, gap_3, lys_6. Moreover, the respective sensitivities are almost linearly dependent which explains the high correlations. Finally, the occurring parameter sensitivities with respect to the pentose-phosphate label stress the importance of measuring these pools.

Error analysis

An in-depth analysis of the discrepancy between the flux and labelling state predicted by the model and the measured values (i.e. the residuals) would exceed the scope of this contribution. However, all methods developed for linear models [van Heijden (1994a), van Heijden (1994b)] are applicable. Only a short discussion of the most important facts is given here. Other statistical methods (e.g. for testing model validity or finding gross measurement errors) would exceed the scope of this contribution (see [Wang (1983), Vallino (1991), van Heijden (1994a), van Heijden (1994b)]).

As a fact, the χ^2 test for the goodness of fit [Pázman (1993)] fails (as is nearly always the case for biological experiments). The computed sum of squares is 137 while 25 is tolerable for 90% confidence (with 40-23=17 degrees of freedom). The reason might be a wrong model structure, a badly quantitated measurement error or gross measurement errors [van Heijden (1994a), Wiechert (1995c)]. Tables I and II show that by far the major contribution to the sum of squares is given by the deviations of lys_3 (46.0), akg_2 (25.1), $co2_1$ (17.3), akg_4 (8.7), akg_3 (7.7) and lys_5 (7.2). Omitting these values already would reduce the sum of squares to 25.0. Thus gross measurement errors might be the case. On the other hand, the measurement standard deviations may be given too optimistically. Doubling them would reduce the sum of squares by a factor of four at the cost of doubled parameter confidence intervals.

In general the model fits much better in the glycolysis and pentose phosphate pathway than in

the other sections of the metabolic network. Systematic measurement errors in the citric acid cycle pools (in particular in the lysine and α-ketoglutarate labels) must therefore be sought. Indeed these pools are strongly influenced by the glutamate pool which has an extremely high intracellular concentration of approximately 160-200 mM [Schrumpf (1995)]. Thus the duration of the experiment has to be sufficiently long to compensate the labelling capacity effects of this pool (see [Wiechert (1995c), Wiechert (1996a)] for an in-depth discussion of this problem). An immediate consequence of this observation is that in the future carbon labelling experiments with *Corynebacterium glutamicum* will be extended to 5 cell residence times in the bioreactor. Indeed the fit for a recently performed sequel experiment that will be published elsewhere passed the χ^2 test.

Discussion and Conclusions

The importance of bidirectional fluxes

The key concept of exchange fluxes has already been stressed in Part I. From the achieved experimental results another example of the influence of bidirectional reaction steps can now be given. In most previous investigations bidirectional reaction steps could not be considered to a large extent because little labelling data was available. In particular the glycolysis and pentose phosphate pathway reactions have been frequently assumed to be unidirectional [Portais (1993), Sharfstein (1994), Rollin (1995)]. In this situation glycolysis and the pentose phosphate pathway can be essentially reduced to two parallel reaction steps

$$\begin{array}{lrcl} \text{Gly:} & \text{3 Glu} & \rightarrow & \text{6 Pyr} \\ \text{PPP:} & \text{3 Glu} & \rightarrow & \text{5 GAP + 3 CO2} \end{array} .$$

with net fluxes $gly^{\text{net}} = gly_1^{\text{net}}/3$ and $ppp^{\text{net}} = ppp_1^{\text{net}}/3$. As can be easily shown the ratio of these fluxes can then be simply calculated from gap_3 by using the formula:

$$\frac{ppp_1^{\text{net}}}{gly_1^{\text{net}}} = \frac{3 - 6\,gap_3}{5\,gap_3}$$

Including the biomass effluxes and the natural ^{13}C enrichment of 1.13% results in a more complicated formula (not given here) but does not lead to significantly different results. From this extended formula and the measured data (Table II) we calculate $ppp_1^{\text{net}}/gly_1^{\text{net}} = 1.0$ with a 90 % confidence interval of $\pm 2\%$. This is a significant difference of 200 % to the flux ratio of 2.0 computed from Figure 3b!

Thus, if no further measurement data would be available the simplified view would produce wrong but statistically well determined results! Of course if the whole data set is used to fit the simplified model it turns out that the unidirectional model cannot explain the measured pentose phosphate pathway labelling. This illustrates how the availability of redundant measurement data enables the validity of a given model to be tested. The example stresses once more that bidirectional reaction steps are a key concept for the evaluation of carbon labelling experiments. Significantly biased results will be produced if they are not considered.

Flux estimation from labelling data

This contribution proves that the carbon labelling technique enables all stationary net fluxes in a complex reaction network to be quantified and at the same time also the order of magnitude of

most exchange fluxes. To this end, the availability of a large amount of high-precision fractional labelling measurements is of great importance for estimating all parameters with reasonably small confidence intervals. No assumptions about the stoichiometry of the energy metabolism were required for flux determination. On the contrary, such parameters can now be derived from the estimated fluxes.

On the other hand, due to the inherent non-linearities of the carbon label flux balance equations the determination and statistical analysis of flux estimates requires highly sophisticated mathematical tools that exploit the special structure of such systems. Only the parameter-fitting approach is currently suited for evaluating such complex data sets while explicit calculations and graphical displays are only applicable to certain subproblems.

The required numerical methods for treating the non-linear problems have been developed in this contribution and successfully applied to a complex experimental data set. The numerical and statistical problems arising have been exposed and appropriately solved. All algorithms have been implemented within a flexible software framework for flux analysis. Because all those tools are available now, flux estimation from carbon labelling experiments has reached the same state of maturity as metabolite flux balancing from a computational and statistical viewpoint.

Further work

This contribution focused on the estimation of confidence intervals and the analysis of the covariance matrix for the estimated parameters while other tools of regression analysis were only briefly discussed. Clearly, all established linear methods for identifying gross measurement or modelling errors [Wang (1983), Chatterjee (1988), van Heijden (1994a), van Heijden (1994b)] can be readily extended to carbon labelling systems because our statistical analysis is essentially based on a special linearizing transformation.

Another question of great practical interest is that of choosing an input substrate (or even a mixture of differently labelled substrates) that maximizes the obtained information about the intracellular fluxes. This typical problem of optimal experimental design is usually solved on the basis of the covariance matrix Cov $\hat{\Theta}$ [Pázman (1986)]. Further work will concentrate on an improved estimation of the pentose phosphate pathway influx ppp_1^{net} and the glycolysis exchange fluxes $gly_1^{\text{xch}[0,1]}, gly_3^{\text{xch}[0,1]}$.

A problem that has not yet been extensively treated is the fine-tuning of the transformation parameter β in Equation (2) that influences the goodness of the linearization approximation for the exact confidence intervals. Although this approximation turned out to be rather insensitive with respect to β we are developing a numerical method for an optimal a posteriori choice of this constant.

Finally, a source of information that has not been used in this framework is isotopomer measurements [Wiechert (1995c), Wiechert (1996a)] that can be easily obtained with the experimental technique presented in [Marx (1996)]. For a given network this leads to a much larger amount of measurement data while leaving the number of parameters in the system unchanged. It was shown in [Wiechert (1996a)] that isotopomer measurements have the potential for identifying even more fluxes in metabolic networks. Fortunately, the framework presented above can be readily extended to isotopomer networks using the concepts and algorithms presented in [Wiechert (1995c)]. To this end only one higher order term has to be added to the carbon labelling Equations (3).

A further question that is of great interest for the mechanistic modelling of metabolic networks using concepts from enzyme kinetics is the significance of exchange fluxes for the regulation of metabolism. In particular the question has to be answered whether a large exchange flux permits one to an enzymatic step as non-rate-controlling.

Acknowledgements

We thank the Stiftung Volkswagenwerk for partly funding this project.

Appendix A: Network Structure of *Corynebacterium glutamicum*

The carbon atom transitions of the metabolic network shown in Figure 3 are subsequently given. The trivial transitions of biomass effluxes are not given here (cf. Table I). Using the formal notation introduced in the appendix of Part I the remaining reaction steps are:

Upt :	Glu	+	GAP	→	G6P	+	PYR
	#ABCDEF	+	#abc	→	#ABCDEF	+	#abc
Gly1 :	F6P			→	G6P		
	#ABCDEF			→	#ABCDEF		
Gly2 :	G6P			→	GAP	+	GAP
	#ABCDEF			→	#CBA	+	#DEF
Gly3 :	GAP			→	Pyr		
	#ABC			→	#ABC		
PPP1 :	G6P			→	CO2	+	P5P
	#ABCDEF			→	#A	+	#BCDEF
PPP2 :	P5P	+	P5P	→	S7P	+	GAP
	#ABCDE	+	#abcde	→	#ABabcde	+	#CDE
PPP3 :	GAP	+	S7P	→	E4P	+	F6P
	#ABC	+	#abcdefg	→	#defg	+	#abcABC
PPP4 :	P5P	+	E4P	→	GAP	+	F6P
	#ABCDE	+	#abcd	→	#CDE	+	#ABabcd
CAC1:	Pyr	+	OAA	→	ICit	+	CO2
	#abc	+	#ABCD	→	#DCBAcb	+	#a
CAC2:	ICit			→	AKG	+	CO2
	#ABCDEF			→	#ABCEF	+	#D
CAC3:	AKG			→	Fum	+	CO2
	#ABCDE			→	#BCDE	+	#A
CAC4A:	Fum			→	OAA		
	#ABCD			→	#ABCD		
CAC4B:	Fum			→	OAA		
	#ABCD			→	#DCBA		
GC1:	ICit			→	GlyOx	+	Fum
	#ABCDEF			→	#AB	+	#CDEF
GC2:	GlyOx	+	AcCoA	→	OAA		
	#AB	+	#ab	→	#ABba		
AC:	PYR	+	CO2	→	OAA		
	#ABC	+	#a	→	#ABCa		
LP1A:	OAA	+	Pyr	→	Lys	+	CO2
	#ABCD	+	#abc	→	#ABCDcb	+	#a
LP1B:	OAA	+	Pyr	→	Lys	+	CO2
	#ABCD	+	#abc	→	#abcDCB	+	#A
LP2:	OAA	+	Pyr	→	Lys	+	CO2
	#ABCD	+	#abc	→	#ABCDcb	+	#a

Those steps that are assumed to be unidirectional can be taken from Figure 3. Complete scrambling was assumed for the fumarase reactions CAC4A and CAC4B and lycine production via the succ-dap pathway (LP1A and LP1B).

Appendix B: Computational Details

All computational methods in this contribution have been implemented within a universal software framework for flux analysis written in C++ [Wiechert (1994)]. Its first component is the textual input compiler described in the appendix of Part I. For incorporating measured data, this textual input format has been appropriately augmented. In this appendix some technical details on the implemented algorithms for simulation and computation of the flux estimates are given. More information on the programs can be obtained from the authors.

Solving the model equations

The linear constraint Equations (6) are treated by using a numerically stable singular value decomposition. Clearly, this has to be done only once for a given model structure. When inconsistencies or indeterminacies occur in the constraint equations this can be immediately detected from the singular values [Press (1988)]. It has not been attempted to automate the finding of free fluxes because according to experience the program user usually has has a suitable intuitive suggestion (see the choice of free fluxes in the examples).

Solving the carbon isotope Equations (3) is substantially more complicated because this has to be done in each iteration step of the parameter-fitting algorithm. Unfortunately, it can be proven that the system matrix $(\sum_i \mathbf{v}_i^{\rightarrow} \cdot \mathbf{P}_i^{\rightarrow} + \sum_i \mathbf{v}_i^{\leftarrow} \cdot \mathbf{P}_i^{\leftarrow})$ becomes more and more ill conditioned when exchange fluxes tend to infinity (i.e. $\mathbf{v}_i^{\text{xch}[0,1]} \rightarrow 1$). It was explained in Part I that this situation may well be expected and data analysis for other experiments showed that this really happens. Due to the poor condition neither QR decomposition [Horn (1985)] nor iterative methods (like the Gauss Seidel iteration [Hackbusch (1993)] suggested in [Zupke (1994)]) are then appropriate.

We solved the problem by an appropriate preconditioning of the system matrix using an idea introduced in [Schuster (1992)]. Moreover, this technique is also used for solving the equations "at infinity". This mathematically rather involved technique would exceed the scope of this contribution. More details can be taken from [Wiechert (1995c), Siefke (1996)]. After preconditioning both QR decomposition and iterative methods are applicable. While QR decomposition turned out to be much faster for the initial computation step of a parameter-fitting run, an iterative refinement requires only a few steps when flux parameters are only slightly changed in the course of the optimization run. Exploiting the sparsity of the occurring matrices led to a further speed up of the iterative solution. For these reasons the solution algorithm is automatically switched within our implementation based on a cost estimation.

Differentiating the model equations

The key to fast parameter-fitting algorithms as well as linearized statistical analysis is the efficient computation of the model's Jacobian matrix, i.e. $\partial F / \partial \Theta \, (\hat{\Theta})$. According to Equation (20) this means the derivation of $\mathbf{M}_{\mathbf{w}} \cdot \Phi \circ \Phi_{\beta}^{[0,1]} \circ \Psi \, (\Theta)$ and $\mathbf{M}_{\mathbf{y}} \cdot \Gamma \circ \Phi \circ \Phi_{\beta}^{[0,1]} \circ \Psi \, (\Theta)$ with respect to Θ. Clearly this can be achieved by applying the chain rule. While most incorporated derivatives are computed straightforwardly, the derivation of Γ is obtained by implicit differentiation from Equation (8) as:

$$\frac{\partial \mathbf{x}}{\partial \mathbf{v}_j^{\rightarrow}} \;=\; -\left(\sum_i \mathbf{v}_i^{\rightarrow} \cdot \mathbf{P}_i^{\rightarrow} + \sum_i \mathbf{v}_i^{\leftarrow} \cdot \mathbf{P}_i^{\leftarrow} \right)^{-1} \cdot (\mathbf{P}_j^{\rightarrow} \cdot \mathbf{x} + \mathbf{P}_j^{\text{inp}} \cdot \mathbf{x}^{\text{inp}})$$

$$\frac{\partial \mathbf{x}}{\partial \mathbf{v}_j^{\leftarrow}} = -\left(\sum_i \mathbf{v}_i^{\rightarrow} \cdot \mathbf{P}_i^{\rightarrow} + \sum_i \mathbf{v}_i^{\leftarrow} \cdot \mathbf{P}_i^{\leftarrow}\right)^{-1} \cdot \mathbf{P}_j^{\leftarrow} \cdot \mathbf{x}$$

This shows that Γ as well as its Jacobian can be computed using the same matrix factorization (cf. Equation 8). Thus the computing time for the derivatives is almost negligible.

As is the case for the solution of the model equations, the sketched procedure for computing $\partial F / \partial \Theta \, (\hat{\Theta})$ is not applicable when large or even infinite exchange fluxes occur because the system matrix will then become ill conditioned. Again a specially implemented preconditioner solves the problem but cannot be explained here in detail (see [Wiechert (1995c), Siefke (1996)]).

Handling non-differentiable terms

The reader may have recognized that the transformation Φ in Equation (1) is not differentiable everywhere because terms of type $v^{\text{xch}} = \min(v^{\rightarrow}, v^{\leftarrow})$ are involved which are not differentiable for $v^{\rightarrow} = v^{\leftarrow}$. This merely technical problem can be resolved by a differentiable approximation of the minimum function. The key idea is that

$$\min(v^{\rightarrow}, v^{\leftarrow}) = \frac{1}{2} \cdot (v^{\rightarrow} + v^{\leftarrow} - |v^{\rightarrow} - v^{\leftarrow}|) \quad .$$

Now the absolute value $|v^{\rightarrow} - v^{\leftarrow}|$ can be differentiably approximated by

$$|v^{\rightarrow} - v^{\leftarrow}| = \sqrt{(v^{\rightarrow} - v^{\leftarrow})^2} \approx \sqrt{(v^{\rightarrow} - v^{\leftarrow})^2 + \delta^2}$$

with a smoothing parameter δ. For $\delta \to 0$ the approximation error becomes arbitrarily small.

For parameter estimation the minimum function is replaced by its approximation with a reasonably small δ. The resulting flux estimate will be biased as an effect of the approximation. However, this first guess will be sufficient for determining at least the sign of the involved net fluxes. Knowing these signs the flux directions can be fixed by adding the constraints $\mathbf{v}_i^{\text{net}} \geq 0$ or $\mathbf{v}_i^{\text{net}} \leq 0$ to the inequalities (7). Afterwards δ can be set to 0 so that the non-differentiable points are removed from the set of feasible fluxes. The optimization algorithm described below will then experience no difficulties.

Non-linear minimization algorithm

The non-linear constrained least squares problem (18) is numerically solved by using a hybridization of the Levenberg-Marquardt approach for unconstrained parameter-fitting and the Sequential Quadratic Programming (SQP) approach for general constrained optimization [Fletcher (1987), Boggs (1995)]. The key idea of the SQP algorithm is to replace the optimization problem locally by its second order Taylor approximation. With the Gauss-Newton approximation of the Hessian as used in the Levenberg Marquardt algorithm [Fletcher (1987)] the approximative problem around an initial guess Θ for the true minimum is

$$\begin{array}{ll} \text{minimize} & \kappa(\Delta\Theta) = ||F(\Theta) + \frac{\partial F}{\partial \Theta}(\Theta) \cdot \Delta\Theta - \eta||^2_{\mathbf{\Sigma}} \\ \text{subject to the constraints} & \mathbf{A} \cdot (\Theta + \Delta\Theta) \geq \mathbf{a} \end{array} \quad .$$

This quadratic problem is solved using a numerically stable code [Goldfarb (1983)] to obtain the constrained minimum $\Delta\Theta$. Then $\Theta + \Delta\Theta$ is taken as the new guess which gives rise to an iterative algorithm.

Another key idea of the Levenberg-Marquardt algorithm is the maintenance of an elliptical trust region to prevent $\Delta\Theta$ from becoming too large, which would result in an inadmissible Taylor extrapolation [More (1983)]. This trust region approach is carried over to the hybrid code by replacing the elliptical region by a rectangular region which can be expressed by another set of linear inequality constraints. The details would exceed the scope of this contribution. The reader is referred to [Wiechert (1995c), Siefke (1996)].

As is the case with unconstrained optimization, the SQP algorithm converges to a local optimum which can be checked by verifying the so-called second order sufficient conditions [Boggs (1995)]. Restarting the algorithm with different starting values helps to ensure that the global optimum has been found.

Statistical treatment of active inequality constraints

As was demonstrated in Part I the inequality constraints in the nonlinear statistical model of Equation (22) must be strictly obeyed by the optimization procedure to obtain meaningful results. The reason for this is that due to measurement errors the measured labels $\mathbf{y}$ may be outside the accessible labelling states of the system (see Figure 1a). In this situation the computed estimator runs to the boundary of the feasible region $\mathbf{A} \cdot \Theta \geq \mathbf{a}$ from Equation (21).

In the situation, where $\hat{\Theta}$ lies on the boundary of the feasible region $\mathbf{A} \cdot \Theta \geq \mathbf{a}$ the computation of statistical quality measures for $\hat{\Theta}$ is a delicate problem that is currently not adequately covered by statistical theory. A first practical approach for treating this problem is now outlined. Assume for simplicity that only one constraint is given, i.e. $\mathbf{A}$ is a transposed vector. In most cases this equation will be of type $\mathbf{v}_j^{\text{xch[0,1]}} = 0$ or $\mathbf{v}_j^{\text{xch[0,1]}} = 1$. Now consider the situation where the parameter-fitting algorithm has run to the corresponding boundary plane, i.e. $\mathbf{A} \cdot \Theta = \mathbf{a}$.

In this situation it is advisable to take this equation for granted, i.e. the inequality constraint is a posteriori replaced by the equality $\mathbf{A} \cdot \Theta = \mathbf{a}$. This equation can then also be included within the equality constraints section of the model (i.e. Equation (6)) from the very start. Clearly this reduces the degree of freedom by one but the problem reduced in this way has exactly the same flux solutions as the original problem.

This a posteriori procedure makes it possible to remove those inequality constraints that have become active. Since all the others pose no problem the inequality constraints $\mathbf{A} \cdot \Theta \geq \mathbf{a}$ can be completely removed a posteriori from the problem formulation. Of course, this ad hoc procedure deserves further statistical research which is currently going on. In particular the significance of a constraint to be active should be statistically tested. This would exceed the scope of this contribution (see [Wiechert (1995c)] for more details). Fortunately, active inequality constraints did not occur in the presented application example (although we frequently encountered them with other experimental data sets).

References

[Anderson (1983)] Anderson, D.H. 1983. Compartmental Modelling and Tracer Kinetics, Vol. 50 of Lecture Notes in Biomathematics. Springer

[Arnold (1990)] Arnold, S.F. 1990. Mathematical Statistics. Ellis Horwood.

[Bates (1988)] Bates, D.M., Watts, D.G. 1988. Nonlinear Regression Analysis and its Applications. Wiley.

[Boggs (1995)] Boggs, P.T., Tolle, J.W. 1995. Sequential quadratic programming. Acta Numerica, 1–51.

[Chatham (1995)] Chatham, J.C., Forder, J.R., Glickson, J.D., Chance, E.M. 1995. Calculation of absolute metabolic fluxes and the elucidation of the pathways of glutamate labeling in perfused rat heart by ^{13}C NMR spectroscopy and nonlinear least squares analysis. The Journal of Biological Chemistry **270**:7999–8008.

[Chatterjee (1988)] Chatterjee, S., Hadi, A.S. 1988. Sensitivity Analysis in Linear Regression. Wiley.

[Chauvin (1994)] Chauvin, M.-F., Megnin-Chanet, F., Martin, G., Lhoste, J.-M., and Baverel, G. 1994. The rabbit kidney tubule utilizes glucose for glutamine synthesis. J. Biol. Chem. **269**:26025–26033, 1994.

[Crawford (1983)] Crawford, J.M., Blum, J.J. 1983. Quantitative analysis of flux along the gluconeogenic, glycolytic and pentose phosphate pathways under reducing conditions in hepatocytes isolated from fed rats. Biochem. J. **212**:595–598.

[Fletcher (1987)] Fletcher, R. 1987. Practical Methods of Optimization. 2nd edition, Wiley.

[Goel (1993)] Goel, A., Ferrance, J., Jeong, J., Ataai, M.M. 1993. Analysis of metabolic fluxes in batch and continuous cultures of *Bacillus subtilis*. Biotechnology and Bioengineering **42**:686–696.

[Goldfarb (1983)] Goldfarb, D., Idnani, A. 1998. A numerically stable dual method for solving strictly convex quadratic programs. Math.Progr. **27**:1–33.

[Hackbusch (1993)] Hackbusch, W. 1993. Iterative Solution of Large Sparse Systems of Equations, Vol. 95 of Applied Mathematics. Springer.

[Holms (1986)] Holms, W.H. 1986. The central metabolic pathways of *Escherichia coli*: relationship between flux and control at a branch point, efficiency of conversion to biomass, and excretion of acetate. Curr.Topics Cell.Regul. **28**:69–104.

[van Heijden (1994a)] van Heijden, R.T.J.M., Heijnen, J.J., Hellinga, C., Romein, B., and Luyben, K.C.A.M. 1994. Linear constraint relations in biochemical reaction systems: I. Classification of the calculability and the balanceability of conversion rates. Biotechnol.Bioeng. **43**:3–10.

[van Heijden (1994b)] van Heijden, R.T.J.M., Romein, B., Heijnen, J.J., Hellinga, C., and Luyben, K.C.A.M. 1994. Linear constraint relations in biochemical reaction systems: II. Diagnosis and estimation of gross errors. Biotechnol.Bioeng. **43**:11–20.

[Horn (1985)] Horn, R.A. 1985. Matrix Analysis. Cambridge University Press.

[Jans (1989)] Jans, A.W.H., Winkel, C., Buitenhuis, L., Lugtenburg, J. 1989. ^{13}C-n.m.r. study of citrate metabolism in rabbit renal proximal-tubule cells. Biochem. J. **257**:425–429.

[Jorgensen (1995)] Jorgensen, H. 1995. Metabolic flux distributions in *Penicillium chrysogenum* during fed-batch cultivations. Biotechnol.Bioeng. **46**:117–131.

[Lawson (1974)] Lawson, C.L., Hanson, R.J. 1974. Solving Least Squares Problems. Prentice Hall.

[Marx (1996)] Marx, A., de Graaf, A.A., Wiechert, W., Eggeling, L., Sahm, H. 1996. Determination of the fluxes in central metabolism of *Corynebacterium glutamicum* by NMR spectroscopy combined with metabolite balancing. Biotechnol.Bioeng. **49**:111–129.

[More (1983)] More, J.J., Sorensen, D.C. 1983. Computing a trust region step. SIAM J.Sci.-Statist.Comput. **4**:553–572.

[Neidhardt (1990)] Neidhardt, F.C., Ingraham, J.L., Schaechter, M. 1990. Physiology of the Bacterial Cell — A Molecular Approach. Sinauer Associates.

[Pázman (1986)] Pázman, A. 1986. Foundations of Optimum Experimental Design. Kluwer Academic Publishing.

[Pázman (1993)] Pázman, A. 1993. Nonlinear Statistical Models. Kluwer Academic Publishing.

[Portais (1993)] Portais, J.-C., Schuster, R., Merle, M., Canioni, P. 1993. Metabolic flux determination in C6 glioma cells using carbon-13 distribution upon [1-^{13}C] glucose incubation. Eur.J.Biochem. **217**:457–468.

[Press (1988)] Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T. 1988. Numerical Recipes in C - The Art of Scientific Computing. Cambridge University Press.

[Rabkin (1985)] Rabkin, M., Blum, J.J. 1985. Quantitative analysis of intermediary metabolism in hepatocytes incubated in the presence and absence of glucagon with a substrate mixture containing glucose, ribose, fructose, alanine and acetate. Biochem. J. **225**:761–786.

[Rollin (1995)] Rollin, C., Morgant, V., Guyonvarch, A., Guerquin-Kern, J.-L. 1995. ^{13}C-NMR studies of *Corynebacterium melassecola* metabolic pathways. Eur. J. Biochem. **227**:488–493.

[Schrumpf (1995)] Schrumpf, B. 1995. Lysinbildung in *Corynebacterium glutamicum*: Analyse des Metabolitflusses durch Bestimmung zellinterner Aminosäurekonzentrationen und enzymatische Untersuchungen. Ph.D. thesis, Universität Düsseldorf.

[Schuster (1992)] Schuster, R., Schuster, S., Holzhütter, H.-G. 1992. Simplification of complex kinetic models used for the quantitative analysis of nuclear magnetic resonance or radioactive tracer studies. J.Chem.Soc.Faraday Trans. **88**:2837–2844.

[Seber (1989)] Seber, G.A.F., Wild, C.J. 1989. Nonlinear Regression. Wiley.

[Sharfstein (1994)] Sharfstein, S.T., Tucker, S.N., Mancuso, A., Blanch, H.W., Clark, D.S. 1994. Quantitative in vivo nuclear magnetic resonance studies of hybridoma metabolism. Biotechnol.Bioeng. **43**:1059–1074.

[Siefke (1996)] Siefke, C. 1996. Ein numerisches Verfahren zur Flußschätzung bei metabolischen ^{13}C-Markierungsexperimenten. Diploma Thesis, University of Bonn.

[Sonntag (1993)] Sonntag, K., Eggeling, L., de Graaf, A.A., Sahm, H. 1993. Flux partitioning in the split pathway of lysine synthesis in *Corynebacterium glutamicum* — quantification by ^{13}C- and ^{1}H-NMR spectroscopy. Eur.J.Biochem. **213**:1325–1331.

[Vallino (1991)] Vallino, J.J. 1991. Identification of Branch-Point Restrictions in Microbial Metabolism through Metabolic Flux Analysis and local Network Perturbations. Ph.D. thesis, Massachusetts Institute of Technology.

[Vallino (1992)] Vallino, J.J., Stephanopoulos, G. 1993. Metabolic flux distribution in *Corynebacterium glutamicum* during growth and lysine overproduction. Biotechnol.Bioeng. **41**:633–646.

[Varma (1994)] Varma, A., Palsson, B.O. 1994. Metabolic flux balancing: basic concepts, scientific and practical use. Bio/Technol. **12**:994–998.

[Walsh (1984)] Walsh, K., Koshland, D.E. 1984. Determination of flux through the branch point of two metabolic cycles — the tricarboxylic acid cycle and the glyoxylate shunt. J.Biol.Chem. **259**:9646–9654.

[Wang (1983)] Wang, N.S., Stephanopoulos, G. 1983. Application of macroscopic balances to the identification of gross measurement errors. Biotechnol.Bioeng. **25**:2177–2208.

[Wiechert (1994)] Wiechert, W. 1994. Design of a software framework for flux determination by ^{13}C NMR isotope labelling experiments, pp. 305–310. In: Gnaiger E., Gellerich F.N., Wyss M. (eds.), What is Controlling Life?, Vol. 3 of Modern Trends in BioThermoKinetics, Innsbruck University Press.

[Wiechert (1995a)] Wiechert, W. 1995. Algebraic methods for the analysis of redundancy and identifiability in metabolic ^{13}C labelling systems, pp. 169–184. In: Schomburg D., Lessel U. (eds.), Bioinformatics; From Nucleic Acids and Proteins to Cell Metabolism. Verlag Chemie.

[Wiechert (1995b)] Wiechert, W. 1996. Metabolic flux determination by stationary ^{13}C tracer experiments: Analysis of sensitivity, identifiability and redundancy, pp. 128–135. In: J. Dolezal, J. Fidler (eds.), System Modelling and Optimization. Chapman and Hall.

[Wiechert (1995c)] Wiechert, W. 1996. Metabolische Kohlenstoff-Markierungssysteme – Modellierung, Simulation, Analyse, Datenauswertung. Habilitationsschrift, University of Bonn.

[Wiechert (1996a)] Wiechert, W., de Graaf, A.A. 1996. In vivo stationary flux analysis by ^{13}C labelling experiments. Adv.Biochem.Eng.Biotechnol. **54**:109–154.

[Wiechert (1996b)] Wiechert, W., de Graaf, A.A. 1996. Bidirectional reaction steps in metabolic networks – Part I: Modelling and simulation of carbon labelling experiments. Biotechnol.Bioeng., 1996. This volume.

[Zupke (1994)] Zupke, C., Stephanopoulos, G. 1994. Modeling of isotope distributions and intracellular fluxes in metabolic networks using atom mapping matrices. Biotechnol.Prog. **10**:489–498.

Figures and Tables

A high quality camera ready copy of all Figures and Tables is attached seperately. All Figures are magnified by 200 %. The 100 % sized figures will fit exactly into one column of “Biotechnology & Bioengineering”.

Table I: Measured and estimated extracellular fluxes in *C. glutamicum* MH20-22B under lysine producing conditions. All values are scaled to a substrate uptake rate of 100 % corresponding to 1.49 mmol/(gTM $\cdot$ h). The biomass effluxes $pyreffl_2$, *e4peffl* and $lyseffl_1$ are coupled to CO_2 formation while $p5peffl_1$ and $p5peffl_2$ refixate CO_2. The flux symbols can be taken from Figure 3a. All measuring errors were taken to be 2% except for CO_2 with 5%. The right column shows the corresponding weighted deviation.

Table II: Measured and estimated fractional labels with assumed measurement errors depending on the quality of the corresponding NMR spectrum. The metabolite symbols can be taken from Figure 3a. Labelled CO_2 was measured by mass spectrometry. The right column shows the corresponding weighted deviation.

Table III: Estimated values and 90% unsymmetric confidence intervals for the estimated free fluxes.

Table IV: The longest principal axes of the 1σ-confidence ellipsoid. Only the largest entries are reproduced.

Flux	Measured value [%]	Estimated value [%]	Assumed standard error	Weighted deviation
Substrate uptake:				
upt	100.0	98.7	2.000	-0.63
Biomass effluxes:				
g6peffl	1.3	1.3	0.026	0.04
f6peffl	0.5	0.5	0.010	0.01
gapeffl	0.9	0.9	0.018	0.02
$pyreffl_1$	18.0	18.0	0.360	0.37
$pyreffl_2$	23.0	23.1	0.460	0.15
e4peffl	1.8	1.8	0.036	0.06
$p5peffl_1$	1.0	1.0	0.020	0.05
$p5peffl_2$	4.9	4.9	0.098	0.21
$akgeffl_1$	7.0	7.0	0.140	0.07
$akgeffl_2$	1.2	1.2	0.024	0.01
oaaeffl	11.6	11.6	0.232	0.17
Product formation:				
lyseffl	18.3	18.5	0.370	0.40
co2effl	275.1	239.0	13.800	-2.62

Table

Carbon atom	Measured value [%]	Estimated Value [%]	Assumed standard error	Weighted deviation
$p5p_1$	12.6	12.5	0.2	-0.14
$p5p_2$	2.9	2.8	0.5	-0.28
$p5p_3$	2.1	2.3	2.0	0.11
$p5p_4$	1.5	2.1	0.5	1.28
$p5p_5$	19.3	18.4	1.3	-0.67
$e4p_1$	2.0	3.0	1.0	1.01
$e4p_2$	3.6	2.1	1.0	-1.49
$e4p_3$	2.0	1.9	1.0	-0.08
$e4p_4$	16.7	15.5	2.0	-0.59
gap_1	2.9	2.9	0.2	-0.11
gap_2	2.6	2.5	0.1	-0.60
gap_3	26.7	26.4	0.2	-1.65
pyr_2	3.0	2.7	1.0	-0.33
pyr_3	26.4	26.3	0.5	-0.20
akg_2	24.1	22.5	0.3	-5.10
akg_3	11.1	9.7	0.5	-2.78
akg_4	28.1	26.3	0.6	-2.95
oaa_2	7.6	9.7	2.0	1.05
oaa_3	20.9	22.5	2.0	0.81
oaa_4	16.8	17.3	2.7	0.17
lys_2	6.8	7.1	0.2	1.36
lys_3	21.9	23.9	0.3	6.78
lys_4	18.9	17.3	1.0	-1.63
lys_5	22.2	24.9	1.0	2.69
lys_6	5.6	5.3	0.3	-1.00
$co2_1$	23.0	21.6	0.4	-4.16

Table

Table

Flux	Estimated value [%]	Estimated 90% confidence interval
ppp_1^{net}	65.3	[53, 78]
gc_1^{net}	1.2	[0, 5]
lp_2^{net}	4.7	[3, 6]
gly_1^{xch}	313.2	[59, ∞]
gly_3^{xch}	14.5	[0, 77]
ppp_2^{xch}	84.2	[51, 137]
ppp_3^{xch}	5.7	[0, 18]
ppp_4^{xch}	11.0	[4, 19]
ac^{xch}	30.4	[14, 53]
cac_4^{xch}	3.2	[0, 15]

Table

Flux	Axis Length							
	0.256	0.206	0.085	0.058	0.032	0.019	0.016	0.001
upt^{net}	.	.	.	.	.	−0.32	−0.75	.
ppp_1^{net}	0.30	.	.	.	−0.27	.	−0.44	.
gc_1^{net}	.	.	.	.	−0.24	.	.	−0.31
lp_2^{net}	.	.	.	.	.	.	.	0.92
$gly_1^{xch[0,1]}$	0.94	.	.	.	.	.	.	.
$gly_3^{xch[0,1]}$	.	−0.93	.	.	.	−0.30	.	.
$ppp_2^{xch[0,1]}$	.	.	0.79	−0.61	.	.	.	.
$ppp_3^{xch[0,1]}$	.	.	−0.53	−0.66	.	.	.	.
$ppp_4^{xch[0,1]}$	.	.	0.28	0.40	.	.	.	.
$ac^{xch[0,1]}$	.	0.34	.	.	.	−0.87	0.27	.
$cac_4^{xch[0,1]}$	.	.	.	.	0.91	.	−0.23	.

Figure 1: Comparison of different approximations for the exact 90 % confidence region of $(v_2^{\text{net}}, v_2^{\text{xch}})$ in the example network from Part I (compare to Figure 2b in Part I): a) six 90 % confidence regions in the (b_1, c_1) plane assuming a measurement error of 0.008 (cf. Figure 2a in Part I). b) the corresponding exact non-linear 90 % confidence regions obtained from the non-linear mapping of the original circles via $\phi^{-1} \circ \gamma^{-1}$. c) linearized elliptical 90 % confidence regions obtained from a linearization of $\phi^{-1} \circ \gamma^{-1}$ in the $(v_2^{\text{net}}, v_2^{\text{xch}})$ space compared to the exact regions. d) approximated 90 % confidence regions obtained from a linearization of $(\phi_\beta^{[0,1]})^{-1} \circ \phi^{-1} \circ \gamma^{-1}$ in the $(v_2^{\text{net}}, v_2^{\text{xch[0,1]}})$ space and subsequent back-transformation into the $(v_2^{\text{net}}, v_2^{\text{xch}})$ coordinate system via $\phi_\beta^{[0,1]}$.

Figure 2: Behaviour of the example system's labelling state c_1 for fixed net flux v_2^{net} and a) increasing v_2^{xch}, b) increasing $v_2^{\text{xch[0,1]}}$. For each case tangent lines and 90 % measurement confidence intervals corresponding to a measurement standard deviation of 0.008 are drawn for small, intermediate and large exchange values. The much lower curvature is apparent from b).

Figure 3: Network model for the central metabolism for *Corynebacterium glutamicum*: a) chosen flux names and free net fluxes (shaded), b) determined net fluxes (boxes with corners) and exchange fluxes (boxes with rounded corners) based on the data of Tables and . Biomass and CO_2 effluxes (except for the ppp_1 step) are not shown for simplicity. Confidence intervals for the flux estimates as given in Table show that the glycolysis exchange fluxes are rather undetermined. i.e. these values should be taken with care.

Figure 1a

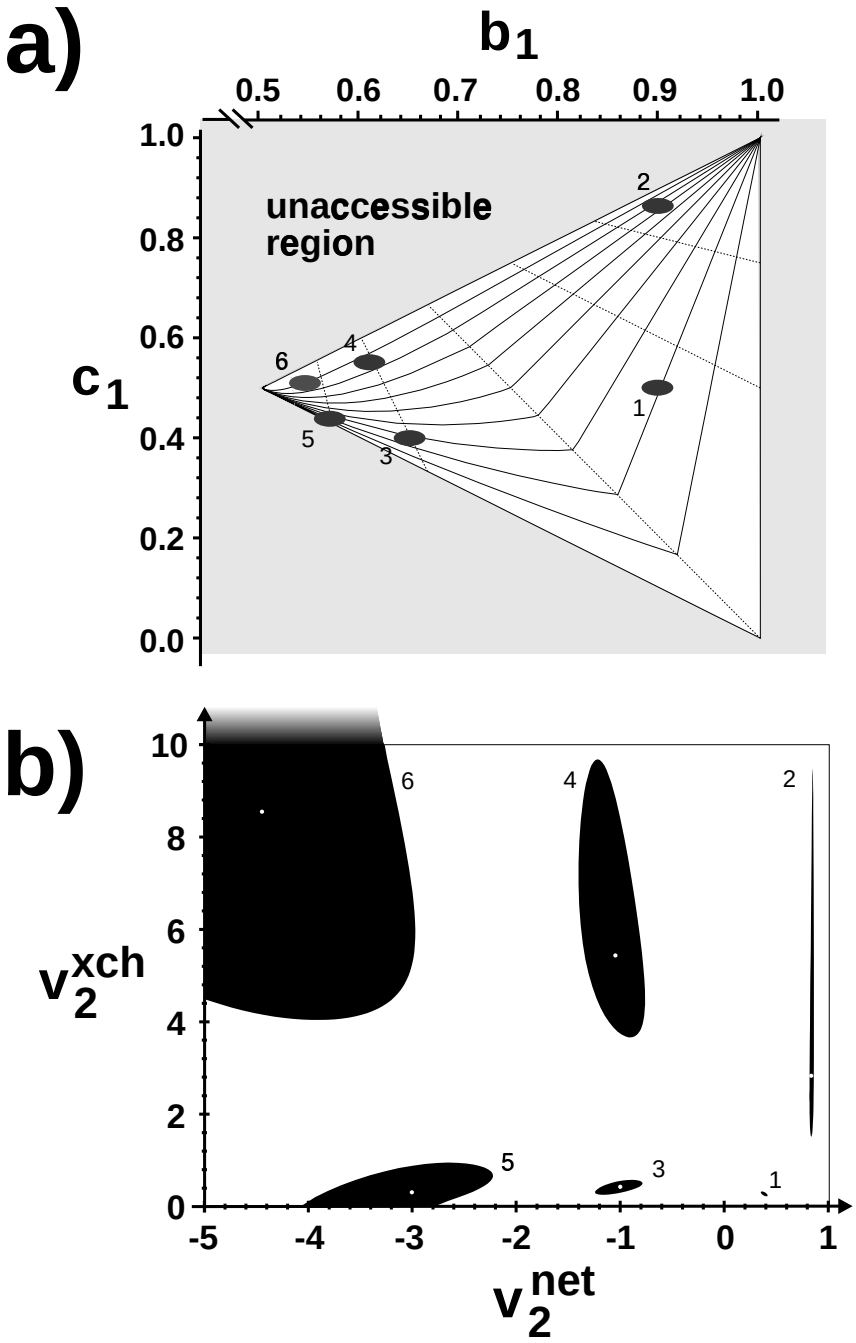

a)
b1
0.5
0.6
0.7
0.8
0.9
1.0
1.0
0.8
0.6
0.4
0.2
0.0
c1
unaccessible
region
1
2
3
4
5
6
b)
10
8
6
4
2
0
-5
-4
-3
-2
-1
0
1
v2xch
v2net
1
2
3
4
5
6

Figure 1b

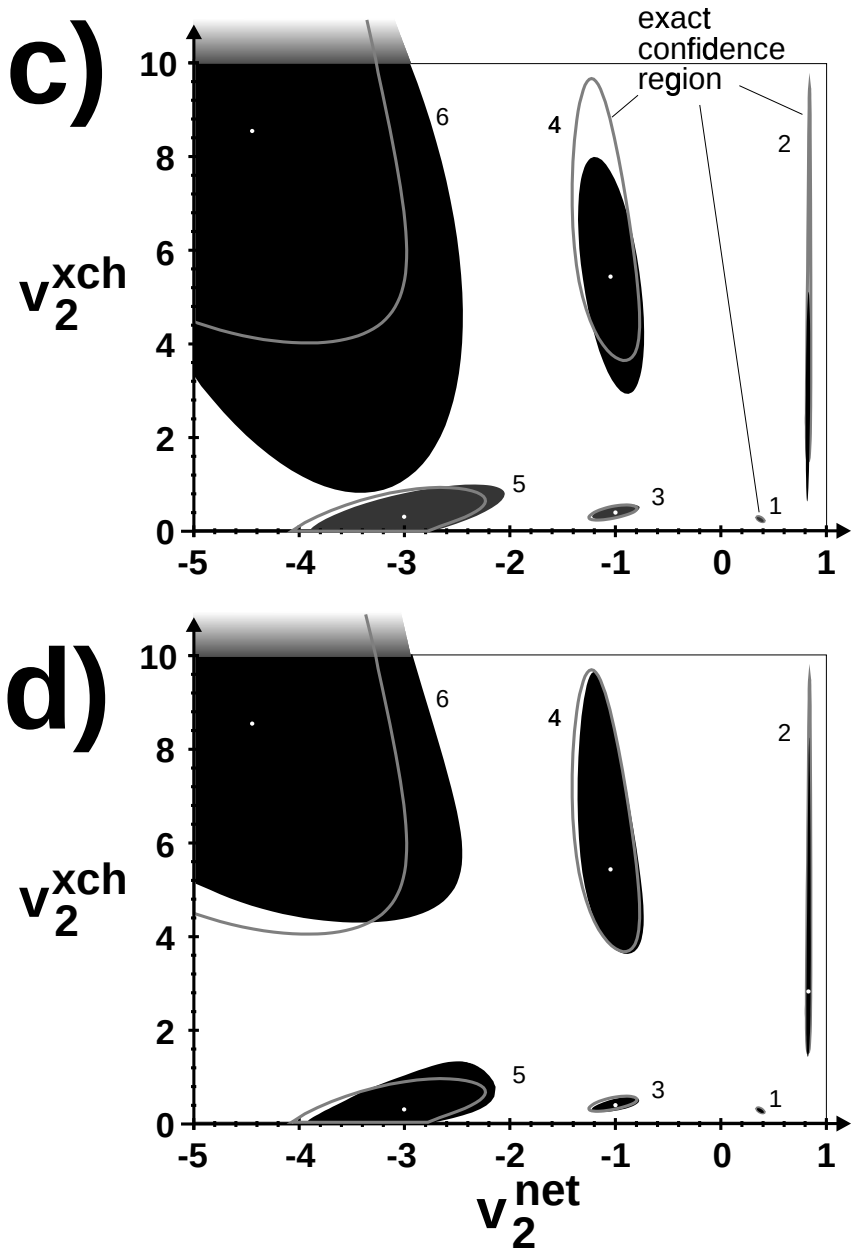

c)
exact
confidence
region
6
4
2
5
3
1
10
8
6
4
2
0
-5
-4
-3
-2
-1
0
1
v_2^{xch}
d)
6
4
2
5
3
1
10
8
6
4
2
0
-5
-4
-3
-2
-1
0
1
v_2^{xch}
v_2^{net}

Figure 2

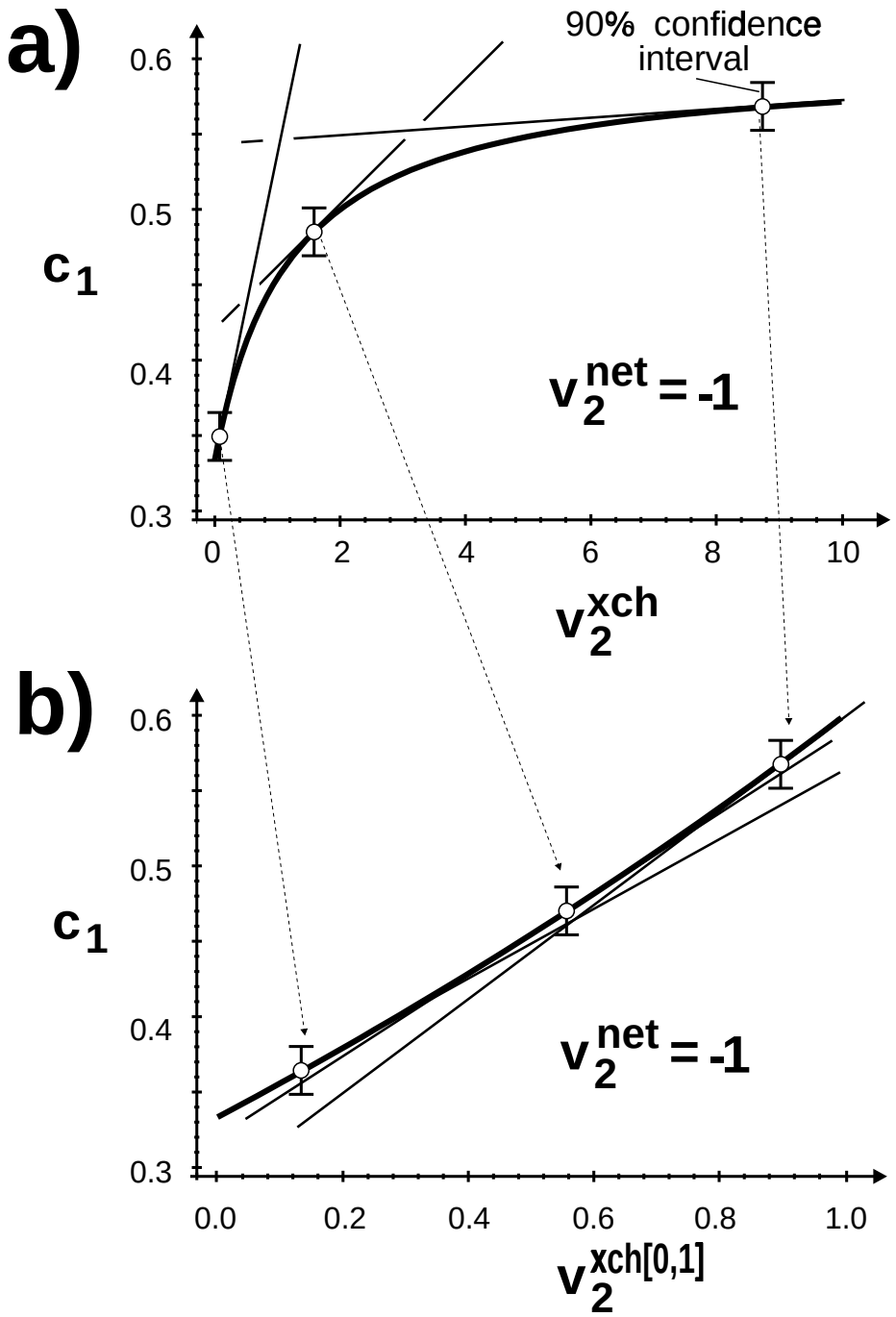

a)
90% confidence interval
0.6
0.5
0.4
0.3
c1
v2net = -1
0
2
4
6
8
10
v2xch
b)
0.6
0.5
0.4
0.3
c1
v2net = -1
0.0
0.2
0.4
0.6
0.8
1.0
v2xch[0,1]

Figure 3a

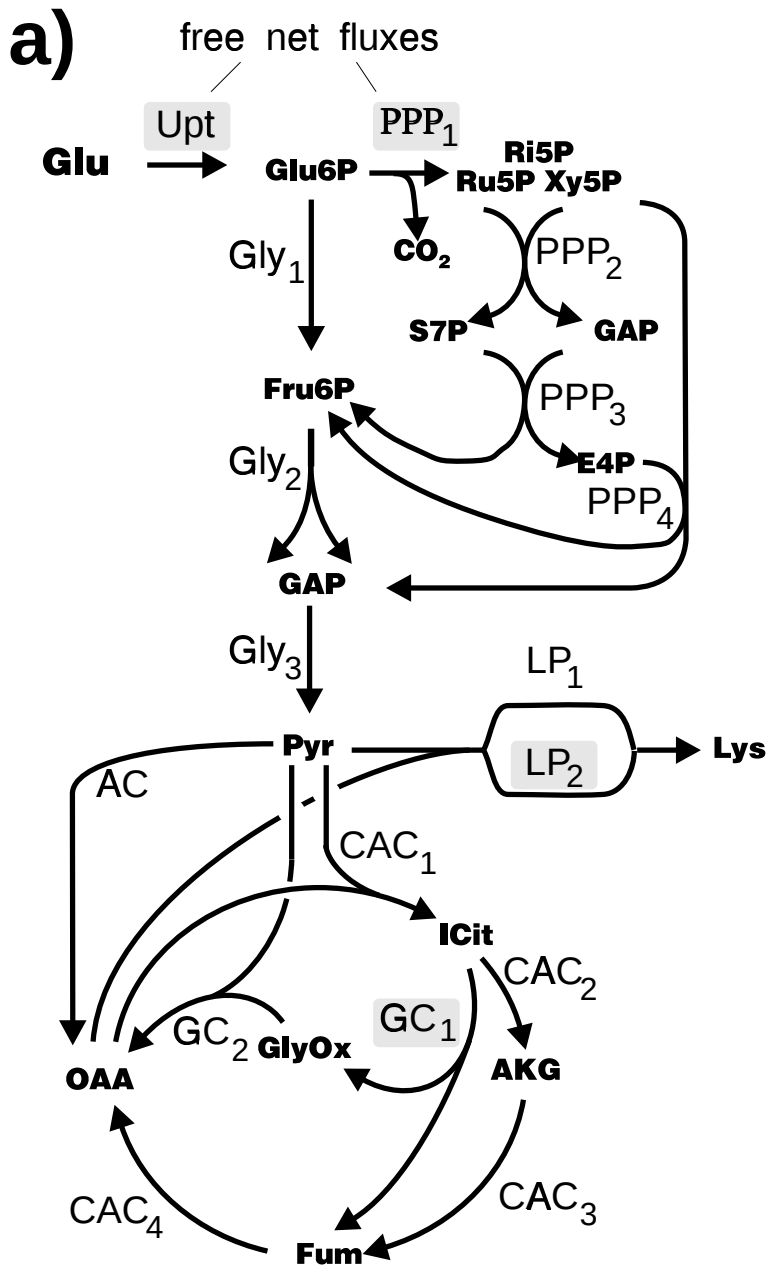
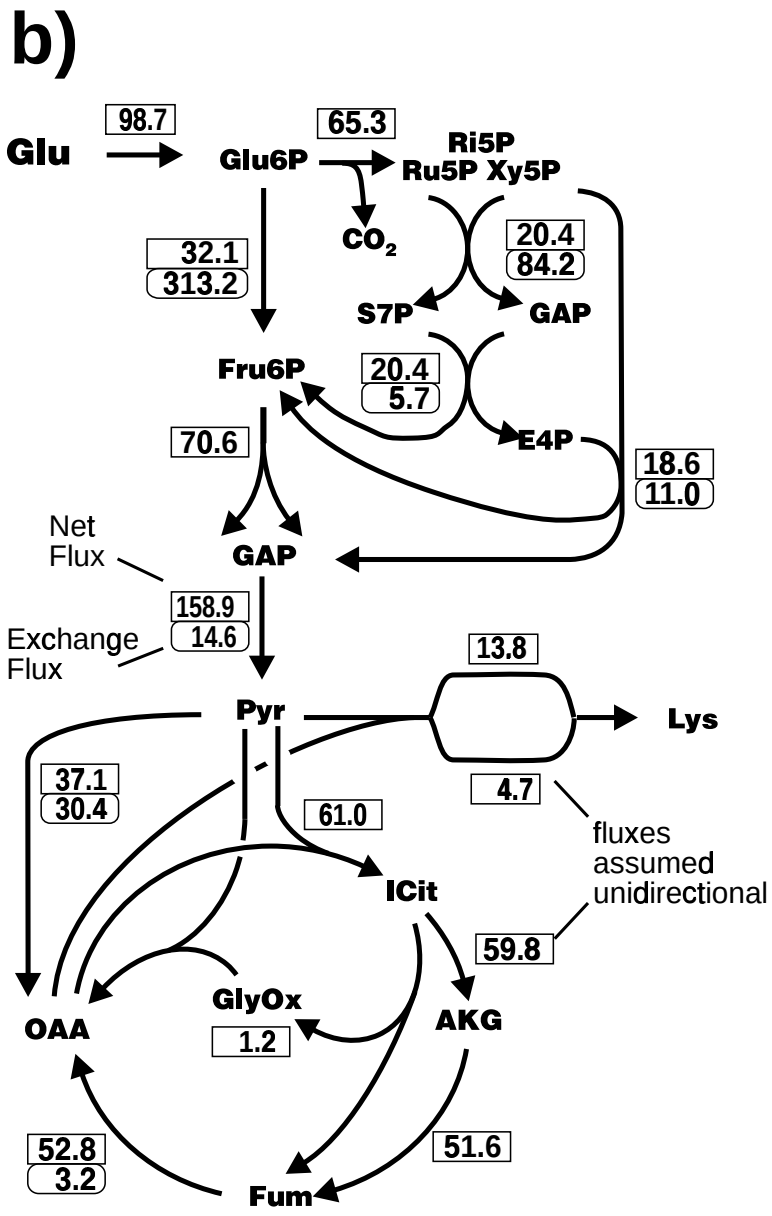

a)
free net fluxes
Upt
PPP1
Glu
Glu6P
Ri5P
Ru5P Xy5P
Gly1
CO2
PPP2
S7P
GAP
Fru6P
PPP3
Gly2
E4P
PPP4
GAP
Gly3
LP1
Pyr
LP2
Lys
AC
CAC1
ICit
CAC2
GC2
GlyOx
GC1
OAA
AKG
CAC4
CAC3
Fum

Figure 3b


b)
Glu
98.7
Glu6P
65.3
Ri5P
Ru5P Xy5P
CO2
20.4
84.2
32.1
313.2
S7P
GAP
Fru6P
20.4
5.7
E4P
70.6
18.6
11.0
Net
Flux
GAP
158.9
14.6
Exchange
Flux
13.8
Pyr
Lys
37.1
30.4
61.0
4.7
fluxes
assumed
unidirectional
ICit
59.8
GlyOx
1.2
OAA
AKG
52.8
3.2
51.6
Fum