Parameter Estimation in Kinetic Models for Large Scale Metabolic Networks with Advanced Mathematical Programming Techniques

Jimena Di Maggio\textsuperscript{a}, Juan C. Diaz Ricci\textsuperscript{a}, M. Soledad Diaz\textsuperscript{a}

\textsuperscript{a}Planta Piloto de Ingeniería Química-PLAPIQUI (UNS-CONICET), Camino la Carrindanga km 7, Bahía Blanca 8000, Argentina, sdiaz@plapiqui.edu.ar

\textsuperscript{b}Instituto Superior de Investigaciones Biológicas-INSIBIO (UNT-CONICET), Chacabuco 461, San Miguel de Tucumán 4000, Argentina, juan@fbqf.unt.edu.ar

Abstract

In this work, we formulate a parameter estimation problem for a large-scale dynamic metabolic network. The DAE system represents the dynamic model for the Embden-Meyerhof-Parnas pathway, the phosphotransferase system and the pentose-phosphate pathway of \textit{Escherichia coli} K-12 W3110 (Chassagnole et al., 2002), with modifications on several enzyme kinetics and the addition of fermentation reactions. Model parameters have been estimated based on recently published experimental data for this strain. Most sensitive parameters have been ranked by performing global sensitivity analysis on the dynamic metabolic network (Di Maggio et al., 2009a,b). Eleven kinetic parameters, including maximum reaction rates, inhibition and half-saturation constants, have been estimated with good agreement with available experimental data.

Keywords: dynamic metabolic network, dynamic optimization, control vector parametrization

1. Introduction

Intracellular and extracellular metabolite concentrations can now be measured, as well as protein levels and activities. The advances on experimental techniques and the consequent increase on the amount of accessible data on the dynamics of functioning cells pave the way to building dynamic models for metabolic networks, which in turn can predict microbial behavior and constitute important tools in metabolic engineering. Dynamic models provide time profiles for the concentration of metabolites involved in the metabolic network under study. They comprise a nonlinear differential algebraic system of equations, which arise from mass balances for metabolites and have a large number of kinetic parameters that require tuning for a specific growth condition.

Chassagnole et al. (2002) applied simulated annealing to estimate all the parameters, around one hundred, in a model describing a large-scale metabolic network for \textit{Escherichia coli}. Ceric and Kurtanjek (2006) compared three different strategies, Nelder and Mead (1965) optimization, Simulated annealing and Differential evolution, to estimate kinetic parameters of a dynamic model of central carbon metabolism of \textit{Escherichia coli}.

In this work, we propose dynamic model extensions for the phosphotransferase system, the Embden-Meyerhof-Parnas pathway, pentose-phosphate pathway, as well as fermentation reactions of \textit{Escherichia coli} K-12 W3110 (Chassagnole et al., 2002) to formulate a parameter estimation problem subject to this differential algebraic system,
within a parameter optimization framework in gPROMS. Numerical results provide maximum reaction rates and half-saturation and inhibition constants for several enzymes from the main metabolic pathways.

2. Mathematical modeling

2.1. Dynamic model of metabolic network

In this work, we have introduced several modifications on the dynamic model representing the Embden-Meyerhof-Parnas pathway, the pentose-phosphate pathway and the phosphotransferase system of *Escherichia coli* K-12 W3110 by Chassagnole et al. (2002). Within the glycolysis pathway, the phosphofructokinase (PFK) kinetics was taken from Diaz Ricci (1996). We have added fermentation reactions, including acetate, formate, lactate, ethanol and succinate production. Kinetic expressions for fermentation pathways for lactate, ethanol and acetate were taken from Hoefnagel et al. (2002). Succinate pathway was modeled with Michaelis-Menten kinetics. In the formate synthesis pathway, the kinetics for pyruvate-formate lyase (PFL) was taken from Knappe et al. (1974). The resulting model comprises twenty five differential equations that represent dynamic mass balances of extracellular glucose, intracellular metabolites and fermentation products, thirty seven kinetic rate expressions and seven additional algebraic equations for co-metabolites. Equation 1 shows mass balance for extracellular glucose; Equations 2 and 3 show general expressions for mass balances on intracellular metabolites and fermentation products, respectively. As it can be seen from Equations 4 to 9 kinetic expressions have a large number of parameters that must be estimated from experimental data. However, not all parameters are feasible candidates for tuning: In this sense maximum reaction rates, which are related to enzyme concentration and inhibition and half-saturation constants could be estimated and, among them, the most influential parameters have been determined through global sensitivity analysis.

\[
\frac{dC_{glc_{ext}}}{dt} = D(C_{glc_{ext}} - C_{glc_{int}}) + f_{pulso_{ext}} - \frac{C_{glc_{int}}}{\rho_f} \tag{1}
\]

\[
\frac{dC_i}{dt} = \sum_{k=1}^{NC_i} v_{ik} r_k - \mu C_i \quad i = 1, \ldots, NC_i \tag{2}
\]

\[
\frac{dC_i}{dt} = r_i \left( \frac{C_i}{p_i} \right) - D C_i \quad i = 1, \ldots, NC_i \tag{3}
\]

where

- \( NC_{int} = 20 \), number of intracellular metabolites (g6p, f6p, fdp, dhap, gap, ppg, 3pg, 2pg, pep, pyr, g1p, 6pg, ribu5p, xyl5p, rib5p, sed7p, e4p, accoa, acetaldehyde, acetyl-P)
- \( NC_{f} = 5 \), number of fermentation products (acetate, formate, lactate, ethanol and succinate)

\[
\frac{r_{PFK}^{max}}{r_{PFK}} = \left[ \frac{e_{PFK}^{max}}{K_{PFK}^{max}} \left( 1 + \frac{e_{PFK}}{K_{PFK}} \right)^{\gamma_{PFK}} \left( 1 + \frac{C_{glc}}{K_{glc}} \right)^{\gamma_{PFK}} \right]^{\gamma_{PFK} - 1} + \frac{1 + \frac{C_{glc}}{K_{glc}}}{L \left( 1 + \frac{C_{glc}}{K_{glc}} \right)^{\gamma_{PFK}} + \frac{C_{PFK}}{K_{PFK}}^{\gamma_{PFK}}} \tag{4}
\]
Parameter Estimation in Kinetic Models for Large Scale Metabolic Networks with Advanced Mathematical Programming Techniques

\[
L = L_{ij} = \begin{bmatrix}
(1 + \frac{C_{\text{pyr}}}{K_{\text{pyr}}}) & (1 + \frac{C_{\text{pdh}}}{K_{\text{pdh}}}) \\
(1 + \frac{C_{\text{pdh}}}{K_{\text{pdh}}}) & (1 + \frac{C_{\text{pyr}}}{K_{\text{pyr}}})
\end{bmatrix}^{-1}
\]

\[
r_{\text{PDH}} = \frac{r_{\text{PDH, max}} C_{\text{form}} C_{\text{AcCoA}}}{K_{\text{PDH, max}} C_{\text{pyr}} + C_{\text{pyr}}^\ast}
\]

\[
r_{\text{PFK}} = \frac{r_{\text{PFK, max}} C_{\text{pyr}}}{K_{\text{PFK, max}} C_{\text{pyr}} + 1}
\]

\[
r_{\text{PGDH}} = \frac{r_{\text{PGDH, max}} C_{\text{pyr}}}{K_{\text{PGDH, max}} C_{\text{pyr}} + 1}
\]

2.2. Experimental data

Experimental data correspond to an *Escherichia coli* K-12 culture perturbed with a glucose pulse (Degnenring et al., 2004) throughout a time horizon of twenty seconds. Temporal profiles for glucose-6-phosphate (g6p), fructose-6-phosphate (f6p), fructose-1,6-diphosphate (fdp), pyruvate (pyr) and phosphoenolpyruvate (pep) were used for parameter estimation problem.

2.3. Parameter estimation problem

The parameter estimation problem has been formulated in g-PROMS (g-PROMS, 2007) as a Maximum Likelihood parameter estimation problem with constant variance. The problem has been formulated as follows:

\[
\frac{N}{2} \ln(2\pi) + \min \sum_{i=1}^{V} \sum_{j=1}^{V} \left[ \ln(\sigma_j) + \frac{(C_j - C_i)^2}{\sigma_j^2} \right]
\]

s.t.

\[
C(0) = C^0
p^\lower{0.5ex}l \leq p \leq p^U
\]

where the summation in the objective function is over \( NM \) measured state variables and \( NT \) data points for each measured variable; \( \sigma_j \) is the variance of the \( j \)th measurement of variable \( i \), which is determined by the measured variable's variance model whose elements correspond to variances of the measured variables. Vector \( p \) corresponds to estimated parameters.
3. Numerical results

Eleven model parameters have been identified as the most influential ones through a previous global sensitivity analysis applied to the DAE system representing the metabolic network for *E. coli* K-12 W3110 (Di Maggio et al., 2009a,b). These parameters stand for maximum reaction rates and half-saturation and inhibition constants for some enzymes that participate in the metabolic network.

Figure 1 shows dynamic sensitivity indices for fructose-6-phosphate (f6p) concentration; it can be noted that the most influential parameter within the first twenty seconds (time horizon for the parameter estimation), is $K_{PTS,f6p}$, which represents by-product inhibition constant for the phosphotransferase system. Regarding pyruvate (pyr) concentration (Fig. 2), it must be pointed out that the most influential parameter throughout the entire time horizon is $r_{max,PDH}$, the maximum reaction rate for pyruvate dehydrogenase (PDH), enzyme for which pyruvate is the substrate. Taking into account most important parameters for the entire set of intracellular metabolite concentrations, we have determined the set of eleven parameters for estimation that is shown in Table 1.

<table>
<thead>
<tr>
<th>Parameter Description</th>
<th>Nominal value</th>
<th>Calibrated value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{PTS,f6p}$</td>
<td>1.5</td>
<td>1.362</td>
</tr>
<tr>
<td>$K_{PTS}$</td>
<td>0.00156</td>
<td>0.00163</td>
</tr>
<tr>
<td>$r_{max,PDH}$</td>
<td>1.2</td>
<td>1.381</td>
</tr>
<tr>
<td>$K_{GAPDH,gap}$</td>
<td>0.569</td>
<td>0.652</td>
</tr>
<tr>
<td>$r_{max,GAPDH}$</td>
<td>0.000013</td>
<td>0.000012</td>
</tr>
<tr>
<td>$r_{max,PDH}$</td>
<td>1132.390</td>
<td>1079.976</td>
</tr>
<tr>
<td>$r_{max,PGDH}$</td>
<td>13.244</td>
<td>12.442</td>
</tr>
<tr>
<td>$K_{GAPDH,pep}$</td>
<td>12.5</td>
<td>11.898</td>
</tr>
<tr>
<td>$r_{max,GAPDH}$</td>
<td>31.25</td>
<td>25</td>
</tr>
<tr>
<td>$r_{max,PGDH}$</td>
<td>15.613</td>
<td>18.736</td>
</tr>
</tbody>
</table>
Parameter Estimation in Kinetic Models for Large Scale Metabolic Networks with Advanced Mathematical Programming Techniques

Table 1 shows nominal values taken from the literature and calibrated values for the eleven estimated parameters. Experimental data and simulated profiles for fructose-6-phosphate (f6p), fructose-1,6-diphosphate (fdp) and pyruvate (pyr) concentrations are shown in Figs. 3 to 5. A good fit between experimental data and predicted profiles for can be seen for both f6p and fdp concentrations, while in the case of pyruvate concentration, it can be noted that experimental data are rather disperse and the best fit we could determine is shown in Fig. 4.

4. Conclusions
We have formulated and solved a dynamic parameter estimation problem for a large-scale metabolic network for the phosphotransferase system, glycolysis, pentose-phosphate pathway and fermentation pathways. Eleven kinetic parameters have been estimated, which had been previously identified as the most influential ones through a global sensitivity analysis. To our knowledge, it is the first time the parameter estimation problem for a large-scale metabolic network has been addressed with advanced mathematical programming techniques (gPROMS, PSE Enterprise, 2007). A satisfactory fit was observed between the predicted profiles and the experimental data, not only the recently reported but also previously published ones (Chassagnole et al., 2002), not reported in this work. Current research includes the determination of additional experimental data for this system.
5. Acknowledgements
The authors gratefully acknowledge financial support from the National Research Council (CONICET), Universidad Nacional del Sur and ANPCYT, Argentina.

References