Consistent Parameter Estimation in Metabolic Networks: A case study

Alexandre Domingues,[1] Susana Vinga,[1,2] Paula Gaspar,[3] Ana Rute Neves,[3] Helena Santos,[3] and João Miranda Lemos[1,4,5]

[1]*INESC-ID - R Alves Redol 9, 1000-029 Lisboa, Portugal*
[2]*FCM-UNL - C. Mártires Pátria 130, 1169-056 Lisboa, Portugal*
[3]*ITQB-UNL, 2780-156 Oeiras, Portugal*
[4]*IST-UTL - Avenida Rovisco Pais 1, 1049-001 Lisboa, Portugal*
[5]*Email: jlml@inesc-id.pt*

Valid mathematical models for metabolic networks can avoid time consuming and expensive experiments when testing and acquiring data from these networks. The process of formulating these models and estimating its parameters is complex and there is no defined and straightforward framework to obtain valid solutions. This paper presents a procedure to estimate parameters using data sets from different experiments, ensuring the consistency of the estimates. The procedure is illustrated by a case study on the effect of Nisin on Mannitol production in *Lactococcus lactis*.

INTRODUCTION

Genetic engineering of certain organisms has proven useful to maximize product yield [3] or even to produce something that would not be normally present in wild-type strains. Such an example is provided in [2], where a genetically modified strain of *L. lactis* was able to produce Mannitol. Fine tunning of the experiment parameters to maximize the product yield is complex, since it, not only involves simple parameters, like the pH or temperature, but also complex variables such as the use of inductors where the time of induction and quantity drastically change the final yield. In [6] Nisin was used as an inductor in different time points that resulted in different Mannitol yields.
This paper presents two simple models for the production of Mannitol in two genetically manipulated strains. The *L. lactis* FI10089 [2] and the strain *L. lactis* FI10089Pase+ [6] with a Nisin inducible gene. It also describes a consistent parameter identification process using two different sets of data simultaneously. Having a valid model for the Mannitol production in the presence of Nisin, the maximization of the Mannitol yield becomes a control problem, where the control variable can be one of the possible manipulated parameters, such as time of induction, quantity of Nisin or pH.

State of the art

The use of Nisin as an inductor to control a certain product yield has been tested several times. In [4] an optimization strategy, relying on practical experiments, is formulated to maximize yield controlling variables such as the pH, type of neutralizing agent, fermentation temperature, point of induction, among others.

Paper contributions and organization

The contribution of this paper consists on a procedure to estimate parameters using data sets from different experiments, while ensuring the consistency of the estimates. This is illustrated by a case study on parameter estimation in metabolic networks using data taken in different conditions. An initial model for Mannitol production is suggested and a sub-model is later added to account for the Nisin induction. Although the initial model does not predict Nisin induction, the data taken using induction is also used to identify the model's parameters. The resulting model can later be used to optimize the product yield without the need of complex and expensive experiences.

The paper is organized as follows: after the Introduction, Section I, in which the problem is introduced and motivated and the state of the art revised, the problem is formulated in Section II, where the two models are presented and the parameter estimation problem is described. In Section III the estimation methods are described in detail and in Section IV the results are presented. Finally, conclusions are drawn in Section V.

PROBLEM FORMULATION

Mannitol model

The complete metabolic pathway of *Lactococcus lactis* that leads to the production of Mannitol is yet to be fully understood. Recent studies [2] suggest that the Mannitol production is highly dependent on the available substrate and total amount of biomass. A simple model with these three variables, Mannitol, biomass and substrate, was formulated ans is going to be referred as model A through the rest o this paper. The model uses the S-System formalism [1, 9, 10] and is described by the following set of ordinary differential equations:

$$\begin{aligned} \frac{dx_1}{dt} &= -\beta_1 x_1^{h_{11}} x_2^{h_{12}} x_3^{h_{13}} \\ \frac{dx_2}{dt} &= \alpha_2 x_1^{g_{21}} x_3^{g_{23}} - \beta_2 x_2^{h_{22}} \\ \frac{dx_3}{dt} &= \alpha_3 x_1^{g_{31}} x_3^{g_{33}} - \beta_3 x_3^{h_{33}} \end{aligned} \tag{1}$$

Here x_1 represents the amount o available Glucose, x_2 the amount of Mannitol and x_3 the biomass, measured in terms of its dry weight. The formation of biomass depends on the amount of available Glucose and on the amount of the biomass itself. The production of Mannitol depends on the amount of biomass and on the available Glucose. The set of parameters that are part of the S-System is going to be referred as δ on the rest of this paper.

$$\delta = \{\alpha_2, \alpha_3, g_{21}, g_{23}, g_{31}, g_{33}, \beta_1, \beta_2, \beta_3, h_{11}, h_{12}, h_{13}, h_{22}, h_{33}\}$$

Mannitol model with Nisin induction

In order to incorporate the induction using Nisin, model A was modified. The time profile of the Nisin concentration is unknown, but it is assumed that the maximum concentration in the solution is reached shortly after the addition. It is also assumed that this concentration remains constant throughout the experience. Given these assumptions, a Hill-type Function was used to approximate the Nisin concentration. A Hill function has the form (2), where n controls the steepness of the curve and θ is the point where $f(t) = \frac{f_{max} - f_{min}}{2}$.

$$f(t) = \frac{t^n}{\theta^n + t^n} \tag{2}$$

Model A, described in (1), was adapted to the induction using Nisin by multiplying each metabolite with a scaled Hill Function. Thus, model B becomes:

$$\begin{aligned} \frac{dx_1}{dt} &= -\beta_1 x_1^{h_{11}} x_2^{h_{12}} x_3^{h_{13}} \alpha_{1n} \left(1 + (\frac{t^n}{\theta^n + t^n})^{h_{1n}}\right) \\ \frac{dx_2}{dt} &= \alpha_2 x_1^{g_{21}} x_3^{g_{23}} \alpha_{2n} \left(1 + (\frac{t^n}{\theta^n + t^n})^{h_{2n}}\right) - \beta_2 x_2^{h_{22}} \\ \frac{dx_3}{dt} &= \alpha_3 x_1^{g_{31}} x_3^{g_{33}} \alpha_{3n} \left(1 + (\frac{t^n}{\theta^n + t^n})^{h_{3n}}\right) - \beta_3 x_3^{h_{33}} \end{aligned} \tag{3}$$

and is schematically represented in Fig. 1. The set of parameters of the Hill-type Function and scaling are going to be referred as σ through the rest of the paper.

$$\sigma = \{\alpha_{1n}, \alpha_{2n}, \alpha_{3n}, h_{1n}, h_{2n}, h_{3n}, \theta, n\}$$

It is important to point that (3) does not obey to the S-System formalism. Fig. 1 schematically represents model B which includes Nisin induction. The initial model A, without Nisin induction, is similar but without the positive and negative feedback from Nisin.

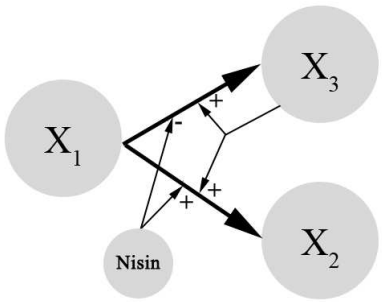

FIG. 1: Mannitol Model - Model B adds Nisin induction to model A.

Data sets

Three data sets were used on the parameter estimations. The first data set, from [2], describes the Glucose and Mannitol concentrations and Biomass dry weight, in a 26 hours period, in a non-Nisin inducible strain of *L. lactis*. The remaining two data sets [6], describe the Glucose and Mannitol concentrations and Biomass in a Nisin inducible strain with Nisin added in two distinct time points.

Parameters estimation

The suggested Mannitol model A, has 14 parameters (δ) to be estimated, the second model, B, including the Nisin induction, adds 8 more parameters (σ).

The estimation problem consists, on a first stage, estimating the common parameters of model A and model B, (δ), using only the data acquired without the addition of Nisin.

On a second stage, the parameters (δ) are estimated using both the data acquired with and without Nisin.

Finally, the parameters of the Nisin part of mode B, (σ), are estimated and fine tunned using only the data acquired with Nisin added.

PARAMETER ESTIMATION METHODS

The estimations of the parameters of both models were made using Matlab. On a first approach the estimations were made using the freely available toolbox SBTOOLBOX and SBPD [5], on a second stage using scripts written for the effect. The parameters estimation is resumed as follows:

- Initial parameters and constraints are defined. The initial parameters were set to 1 most of the times, the constraints of the S-System parameters were set to -4 on the lower bound and 4 on the higher bound.
- An optimizing function is called. Three functions were tested, *FMINUNC*, *FMINCON* and *simannealingSB*. The first two function belong to Matlab's optimization toolbox, and do unconstrained and constrained optimizations respectively. The last function belongs to the SBTOOLBOX and performs minimization by simulated annealing.
    1. Inside the optimization function the set of differential equations of the model is integrated, using the initial set of parameters.
    2. After the integration a cost function calculates the cost, normally using the minimum sum of squares.
    3. If the obtained cost obeys the stop condition of the optimization function, the optimization is stopped. Otherwise, a new set of parameters is tested.

The estimations were made on a laptop with 4GB of RAM and a dual processor. Estimation times varied between less than a minute and several minutes.

Estimation using one data set

The first estimation uses the data set obtained without the addition of Nisin and model A. A first and rough estimation used SBTOOLBOX, with the initial parameters set to 1. After this estimation a script written for the effect was used. The cost function was defined as:

$$J = \sum_{i=1}^{3} \left[\sum_{j=1}^{\text{timepoints}} \left[\frac{y_{ij}(\delta) - \widehat{y}_{ij}}{\omega_i} \right]^2 \right]$$

Where $i = 1, 2, 3$ refers to the three metabolites, Glucose, Mannitol and Biomass, $y_{ij}(\delta)$ refers to sample j of metabolite i of the model data, integrated with the set of parameters δ and $\widehat{y}_{ij}$ refers to experimental data, sample j of metabolite i. The weighting factor ω_i allows to give more or less weight to each metabolite during the estimation process. This estimation identifies the 14 parameters, δ, belonging to the S-System.

Estimation using multiple data sets

On the second estimation both models A and B are used, as well as the three data sets, the ones with and without the addition of Nisin. It is important to note that the data sets obtained with Nisin will not fit model A (1) with the parameters δ obtained in the previous section, but since model B (3) is an extension of model A (1) the common parameters δ have to be equal. The δ parameter set is forced to be common by doing its estimation using both models simultaneously. The cost function becomes:

$$J = J_1 + J_2 \tag{4}$$

where

$$J_1 = \sum_{i=1}^{3} \left[\sum_{j=1}^{\text{timepoints}} \left[\frac{y_{ij}(\delta) - \widehat{y}_{ij}}{\omega_i} \right]^2 \right]$$

$$J_2 = \sum_{i=1}^{3} \left[\sum_{j=1}^{\text{timepoints}} \left[\frac{y_{ij}(\delta, \sigma) - \widehat{y}_{ij}}{\omega_i} \right]^2 \right]$$

and σ is the set of parameters of the Hill Function and scaling. Thus, the parameters of the S-System, δ, are estimated and the ones from the Hill-type function and scaling, σ, are left free. Since two Nisin data sets are used, two distinct σ are estimated. Thus, the estimation algorithm identifies $\delta + 2\sigma = 30$ parameters.

Estimation using Nisin data sets

Having estimated the set of parameters of the S-System, δ, the Hill Function and the respective scaling parameters, σ can be fine tunned. This estimation process only uses model B.

Even though the two set of parameters σ obtained on the previous section are able to fit the experimental data, from a control point of view it is useful to reduce the control variables. Thus, σ is divided, and the set $\{\alpha_{1n}, \alpha_{2n}, \alpha_{3n}, h_{1n}, h_{2n}, h_{3n}\}$ is forced to be common to the two Nisin data sets. The only variables/parameters left free are $\{\theta, n\}$.

This decision is based in the fact that $\{\theta, n\}$ directly manipulate the shape and position of the Hill-type function, as seen in (2), more specifically, varying θ changes the position of the function in the time axis, creating a time control variable. This time control variable subsequently model the time of addition of the Nisin.

The estimation algorithm estimates $\{\alpha_{1n}, \alpha_{2n}, \alpha_{3n}, h_{1n}, h_{2n}, h_{3n}\} + \{\theta, n\} * 2 = 10$ parameters.

RESULTS

On the first estimation, parameters δ were estimated using the no Nisin data set. Model A was then integrated using the obtained δ set and both modeled and experimental data were plotted, Fig. 2. The estimated parameters allow the model to fit the experimental data. Without Nisin added the Mannitol production is residual and almost all energy goes to the formation of biomass.

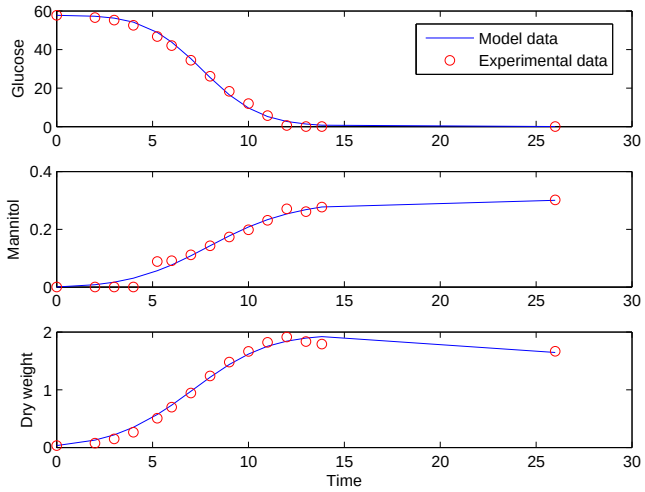

FIG. 2: Mannitol Model A - The integration results were similar using the two different δ sets

On the second estimation, using the three data sets and both models A and B, the initial δ parameters were set to the previously obtained values and $\sigma_{1..8} = 1$. Model A (1) was then integrated with the new δ and the obtained results were similar to Fig. 2.
While the two δ sets (first and second estimation) seem to fit real data, the second set (second estimation) is more consistent and more likely to be valid on future data sets. Not only because it was obtained using more data sets but also because the used data was relative to different experimental conditions.

On the third estimation only model B and the Nisin data sets were used. The δ set obtained on the previous estimation was used and the set σ was fine tuned by forcing the parameters $\{\alpha_{1n}, \alpha_{2n}, \alpha_{3n}, h_{1n}, h_{2n}, h_{3n}\}$ to be common to both data sets and estimating the pair $\{\theta, n\}$ for each data set. Fig. 3 shows the obtained results for Nisin induced at $\mathrm{OD}_{600} = 0.1$ and Nisin induced at $\mathrm{OD}_{600} = 0.3$. The modeled data fits the real data in both cases.

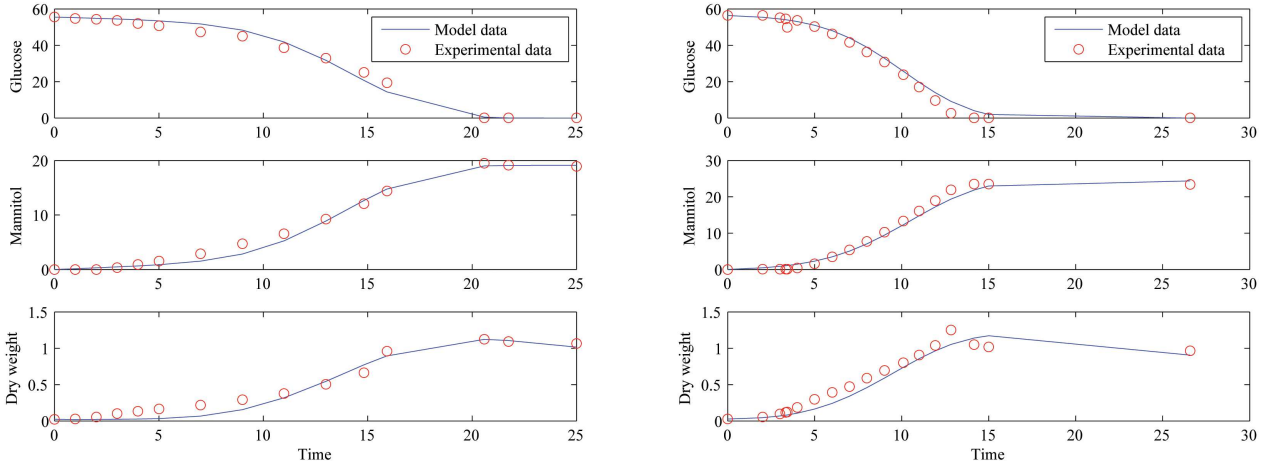

FIG. 3: Nisin added at $OD_{600} = 0.1$ (left) and $OD_{600} = 0.3$ (right)

As in the first estimation, by forcing a sub-set of σ to be common to both data sets we provide consistency to the estimation process.

The described estimation sequence succeeded in finding a set δ common to the Nisin induced and non-Nisin induced data sets. It was also successful in expressing Nisin induction time as a function of a fixed sub-set of σ and a varying $\{\theta, n\}$ pair. This fact suggests that a control strategy could be proposed to maximize the production of Mannitol.

Further work proved that the relation between $\{\theta, n\}$ and the time of induction was not as straightforward as expected and the plot of 2 with the obtained $\{\theta, n\}$ was approximately a straight line, suggesting that a simple scalar could be used to model the addition of Nisin. Future work, using more data sets, with different induction times, will be done in order to clarify this issue.

CONCLUSION

The work presented in this paper shows that parameter estimation, using data sets from different experiments, can provide consistency to the identification process. The ability to find parameter sets, common to different experiments and even different models is seen as a step toward the formulation of predictive models instead of descriptive models.

The use of custom cost functions provides an extra degree of freedom on the estimation process. Different weights can be easily given to different parameters/data sets/time points and strategies as the one described, where data sets from different models are used, are easily implemented.

The described work is still a work in progress, with several items needing further work. The suggested models, while able to fit the experimental data, should be simplified in order to reduce the number of parameters. The used data sets suffer of lack of time resolution, specially in the last hours of experiment. The reduced number of experiments is also a bottleneck on the definition of a control strategy. Finally, the data obtained without the addition of Nisin was obtained using a slightly different strain of *L. lactis* than the rest of the data sets.

[1] A. Sorribas, B. Hernández-Bermejo, E. Vilaprinyo, R. Alves. (2006). "Cooperativity and saturation in biochemical networks: A saturable formalism using Taylor series approximations." *Biotechnology and Bioengineering*, vol. 97, no. 5, pp. 1259-1277.

[2] Gaspar, Paula, Neves, Ana Rute, Ramos, Ana, Gasson, Michael J., Shearman, Claire A., Santos, Helena "Engineering Lactococcus lactis for Production of Mannitol: High Yields from Food-Grade Strains Deficient in Lactate Dehydrogenase and the Mannitol Transport System" Appl. Environ. Microbiol. 2004 70: 1466-1474

[3] Kapil G. Gadkar, F. J. Doyle III, J. S. Edwards, R. Mahadevan. (2005). "Estimating optimal profiles of genetic alterations using constraint-based models." *Biotechnology and Bioengineering* 89(2): 243-251.

[4] Igor Mierau, Kees Olieman, James Mond, and Eddy J Smid. "'Optimization of the Lactococcus lactis nisin-controlled gene expression system NICE for industrial applications." Microb Cell Fact. 2005; 4: 16. Published online 2005 May 30. doi:10.1186/1475-2859-4-16.

[5] Systems Biology Toolbox for MATLAB: A computational platform for research in Systems Biology, Bioinformatics, 22(4), 514-515, 2006. (www.sbtoolbox2.org).

[6] Paula Gaspar. "Metabolic Engineering of Lactococcus lactis for Mannitol Production - The Role of NADH-dependent Dehydrogenases in Directing Carbon Fluxes"

[7] Stphane Hazebrouck, Laetitia Pothelune, Vasco Azevedo, Grard Corthier, Jean-Michel Wal, and Philippe Langella. "Efficient production and secretion of bovine ?-lactoglobulin by Lactobacillus casei." Microb Cell Fact. 2007; 6: 12. Published online 2007 April 6. doi: 10.1186/1475-2859-6-12.

[8] Maksat Ashyraliyev, Yves Fomekong-Nanfack, Jaap A. Kaandorp, Joke G. Blom. "Systems biology: parameter estimation for biochemical models." FEBS Journal, Vol. 276, No. 4., pp. 886-902.

[9] Savageau, M. A. (1969). "Biochemical systems analysis. II. The steady-state solutions for an n-pool system using a power-law approximation." Journal Of Theoretical Biology 25(3): 370-379.

[10] Savageau, M. A. (1970). "Biochemical systems analysis. III. Dynamic solutions using a power-law approximation." Journal Of Theoretical Biology 26(2): 215-226.