Computers and Chemical Engineering 55 (2013) 181–197

Contents lists available at SciVerse ScienceDirect

Computers and Chemical Engineering

journal homepage: www.elsevier.com/locate/compchemeng

Accelerating the parameters identifiability procedure: Set by set selection

Kese P.F. Alberton[a], André Luís Alberton[b], Jimena Andréa Di Maggio[c], María Soledad Díaz[c], Argimiro R. Secchi[a,*]

[a] Programa de Engenharia Química-COPPE, Universidade Federal do Rio de Janeiro, P.O. Box 68502, 21941-972 Rio de Janeiro, RJ, Brazil
[b] Instituto de Química, Universidade Estadual do Rio de Janeiro-UERJ, São Francisco Xavier 524, 20550-900 Rio de Janeiro, RJ, Brazil
[c] Planta Piloto de Ingeniería Química-CONICET, Universidad Nacional del Sur, Camino La Carrindanga Km 7, Bahía Blanca, Argentina

ARTICLE INFO

Article history:
Received 17 January 2013
Received in revised form 23 March 2013
Accepted 1 April 2013
Available online 1 May 2013

Keywords:
Parameters identifiability
Binary search
Parameters estimation
Scarce experimental data

ABSTRACT

In this paper, a numerical procedure based on the binary search is proposed for accelerating the parameters identifiability procedure. Basically, the parameters are selected set by set using a given criterion for ranking the parameters. Since parameters identifiability procedures are strongly dependent on the initial estimates of parameters values, simultaneous parameters re-estimation step has been proposed in this paper. Two examples were used to evaluate the performance of the proposed criterion. In both cases, a significant reduction of the computational time was observed, and the results regard to the model fit are similar to those criteria based on the selection of parameters one by one, as usually presented in the literature.

1. Introduction

Several areas of human knowledge make use of complex mathematical models containing a large number of parameters. However, statistically consistent estimation of all these parameters is very often not possible, since it would demand a prohibitive number of experimental data or would require inexistent/unavailable measurement devices. In fact, experiments are usually time consuming, difficult and expensive. A common approach to overcome this scenario of ill-posed estimation is selecting a subset of parameters for estimation, while other parameters are set at their initial estimates; these procedures are commonly known as parameters identifiability.

Briefly, parameters identifiability procedures seek to determine a set of parameters of the model that can be consistently estimated using the available experimental data. Theoretically, these procedures should take into account the influence of the selected parameters on the model predictions and/or the correlation with other parameters. Such analyses can be performed based on the structure of the model, the experimental measurements and uncertainties, and the uncertainties of the initial estimates (Thompson, McAuley, & McLellan, 2009), which are not easy tasks (Yue et al., 2006).

Identifiability procedures presents two important features (Kravaris, Chu, & Hahn, 2012): (1) interpretability – indicates which parameters are important for a model dynamic behavior and they also provide insight about the correlation of the effect that different parameters have on the outputs and (2) simplification – reduces the numbers of decision variables in parameter estimation problem since non-selected parameters are all fixed at in initial estimates. Thus, several parameters identifiability approaches have been proposed in the literature. A common sense is the use of sensitivity analysis, as a natural choice, being a valuable tool for verifying the parameters importance. In the last decade, several methodologies were developed based on local sensitivity analysis (Brun, Reichert, & Kunsch, 2001; Chu & Hahn, 2007; Li, Henson, & Kurtz, 2004; Lund & Foss, 2008; Sandink, McAuley, & McLellan, 2001; Secchi, Cardozo, Almeida Neto, & Finkler, 2006; Sun & Hahn, 2006; Yao, Shaw, Kou, McAuley, & Bacon, 2003) and less usual global sensitivity methods (Chu, Huang, & Hahn, 2011). In most of these methodologies, parameters are selected from the most estimable to the least estimable. Nonetheless, this parameters selection is established by evaluating the parameters one by one, which can be a time consuming procedure for a complex mathematical model containing a large number of parameters. In such scenario, an important question to be answered is: what is the smallest number of necessary evaluations to be performed to determine a set of identifiable parameters?

In this context, this work presents a numerical procedure for parameters selection based on the binary search that could be

* Corresponding author. Tel.: +55 21 25628307; fax: +55 21 25628300.
E-mail address: arge@peq.coppe.ufrj.br (A.R. Secchi).

0098-1354/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compchemeng.2013.04.014

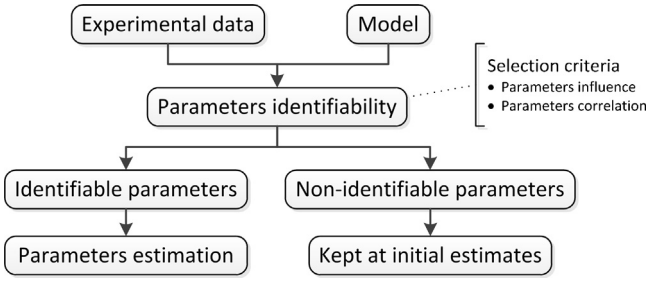

Fig. 1. Classical scheme of parameters identifiability procedures.

applied using any criteria proposed in the literature for ranking the parameters according to their identifiability (Brun et al., 2001; Chu & Hahn, 2007; Li et al., 2004; Lund & Foss, 2008; Sandink et al., 2001; Secchi et al., 2006; Sun & Hahn, 2006; Yao et al., 2003). Basically, the proposed procedure makes the parameters selection evaluating subsets of parameters divided in half according to the selected criterion to order the parameters. This proposal leads to a significant reduction in the computational cost compared to the usual procedures. Two examples illustrate the performance of the proposed procedure: (1) a hypothetical case study – a simply problem of parameters identifiability with known solution: a linear model contained large number of parameters and orthogonal experimental data, and (2) a real case study – a complex problem of parameters identifiability: a dynamic model of the microorganism *Escherichia coli* K-12 W3110 metabolic networks (Di Maggio, Ricci, & Díaz, 2009) contained 131 parameters and scarce experimental data.

2. Parameters identifiability

Based on the structure of the mathematical model and on the available experimental data, a parameter identifiability procedure splits the parameters set in two subsets: the parameters that can be estimated, called identifiable parameters, and the parameters that cannot be estimated and are kept in their initial estimates, called non-identifiable parameters. According to Kravaris et al. (2012), selection procedures can be divided into the following categories: heuristic methods and optimization-based methods. Optimization methods consist on a combinatorial problem requiring high computational efforts, for such reason, are less used than heuristic approaches.

Based on sensitivity analysis, heuristic methods for parameter selection are derived based on the effects that the variations in the parameters have on the model outputs and so take into account two criteria for parameters evaluation (Kravaris et al., 2012): (1) magnitude of the effect, and (2) uncorrelated effect. Non-significant and/or correlated effects make the parameters estimation procedure ill-posed, even when experimental data would be sufficient and consistent for identifiability of all parameters of the model. Although usually these criteria are used together, the performance of the resulting estimation problem can often not be guaranteed.

Usually, parameters estimation is performed after the set of estimable parameters be obtained in the identifiability procedure. Thus, the quality of the initial estimates values admitted to the parameters is fundamental for the adequate selection. Unfortunately, good parameters initial estimates are seldom known; some case they can be obtained from the literature or more commonly assumed arbitrarily (Kou, McAuley, Hsu, Bacon, & Yao, 2005). Classical scheme employed to parameters identifiability is presented in Fig. 1.

In the classical scheme, the parameters are included one by one in the set of selected parameters, verifying if this set is estimable; otherwise, the last included parameter is removed from the set and the procedure stops. The first parameter is included in the set of selected parameters, $\Theta^{(S)}$, based on its effect on model prediction. For the next parameters, a measure of correlation with the parameters already included in $\Theta^{(S)}$ is also taken into account for re-ranking the remaining parameters. Thus, at each step of the selection procedure, one should evaluate the correlation between the parameters, which requires costly and complex operations with large matrices (Kravaris et al., 2012) such as: (1) orthogonalization (Li et al., 2004; Lund & Foss, 2008; Yao et al., 2003), (2) collinearity index (Brun et al., 2001), (3) relative gain array – RGA (Sandink et al., 2001), (4) principal component analysis – PCA (Li et al., 2004), (5) optimum experimental design criteria – D-optimal, A-optimal, E-optimal (Chu & Hahn, 2007; Machado, Tapia, Gabriel, Lafuente, & Baeza, 2009; Weijers & Vanrolleghem, 1997), and (6) Hankel singular value (Sun & Hahn, 2006).

Being the most popular, the parameters identifiability procedure based on orthogonalization methods is simple and the results are easier to interpret, having been widely used for the analysis of complex models (Kravaris et al., 2012). Among them, identifiability procedure proposed by Yao et al. (2003) seems to be most used. Yao et al. (2003) proposed the parameters identifiability based on well-known Gram–Schmidt orthogonalization method, using a cut-off value previously specified as stop criterion of selection, being this criterion replaced by Thompson et al. (2009) for the point in what the parameters estimation problem becomes ill-posed. Chu and Hahn (2007) also use Gram–Schmidt orthogonalization method proposing a forward selection which maximizes the D-criterion. Also this orthogonalization method is employed by Lund and Foss (2008) that suggest a methodology that produces very similar results to those presented by Yao et al. (2003), but presents great computational complexity. Other approach of parameters identifiability problem employing orthogonalization methods uses Householder transforms as in Hiskens (2001) and Velez-Reyes and Verghese (1995). Both Gram–Schmidt and Householder transforms orthogonalization methods produce identical results, but Householder transforms is more stable numerically (Golub & van Loan, 1996) and the Gram–Schmidt is simpler to interpret (Kravaris et al., 2012).

Most of the parameters identifiability procedures based on orthogonalization methods make use of the Fisher Information Matrix (FIM), defined as:

$$\text{FIM} = (V_\Theta)^{-1} = B^T (V_Y)^{-1} B \tag{1}$$

in which $B = \partial Y/\partial \Theta$, more commonly known as local sensitivity matrix, V_Θ and V_Y represent, respectively, the parameters and experimental covariance uncertainty matrices.

Most commonly, B matrix is normalized (B_N), as presented in Eq. (2), to avoid distortions due to the different magnitude orders of output variables and parameters values. The literature reports several forms for normalization, being one of the most commonly presented as follows.

$$B_N = diag(\bar{Y}^E)^{-1} \left(\frac{\partial Y}{\partial \Theta} \right) diag(\Theta) \tag{2}$$

in which $\bar{Y}^E$ is the vector of mean experimental output variables, Θ is the vector of parameters values, and $diag(x)$ denotes the diagonal matrix of the vector x. Fig. 2 illustrates such normalization. In the local sensitivity matrix, B or B_N, each column is related to one parameter and each row is related to one output variable. Thus, each column i can be considered as a sensitivity vector corresponding to the parameter i, b_{θ_i}. The norm of these sensitivity vectors, $||b_{\theta_i}||$, can be used to evaluate the parameters effects on the prediction.

Surely, procedures for selecting parameters evaluating one by one are computationally costly and time consuming. A first

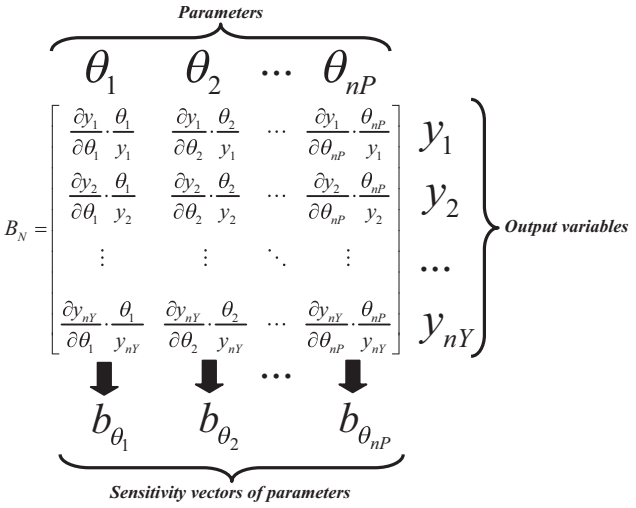

Fig. 2. Normalized local sensitivity matrix B_N and sensitivity vectors of parameters.

proposition to overcome this drawback was done recently by Chu, Huang, Hahn, and Hahn (2012), based on heuristics for assumption of a number of parameters that can be selected at each iteration. According to the authors, at each iteration, new parameters are included in the set of selected parameters, determined by an optimization procedure, making use of a metric based on the orthogonalization of the sensitivity matrix. Although appealing, parameters identifiability using optimization-based methods results is a combinatorial problem, where $nP!/[nSet!\,(nP-nSet)!]$ represents the total number of combinations for selecting $nSet$ parameters from a set of nP parameters (Kravaris et al., 2012). According to Kravaris et al. (2012), these techniques include the sequential selection procedures, stochastic search techniques, and heuristic reduction approaches.

3. Binary search

Binary search is an optimized way to determine an element (e.g., character or number) in a sorted sequence with no additional information. Given a set of sorted numbers, $\boldsymbol{\lambda} = \{1, 2, 3, 4, 5, 6, 7, 8\}$, where one number is randomly selected. How many questions are necessary to obtain the selected number making only questions that can be answered as "yes" or "no"? The most guaranteed fast way to get this answer with no any additional information is ask if the number is in the first half of the set. Thus, given the answer, one can eliminate half of the numbers where the selected number is not present. The remaining set is divided by half again, asking if the number is in the first half of the remaining set. This procedure characterizes a tree search, in which the branches with no selected number are discharged from the search, reducing rapidly the size of the set of numbers to be investigated. In Fig. 3, the query tree of this illustrative example of binary search is presented. Note that whatever is the selected number in the set $\boldsymbol{\lambda}$, the number of evaluations required is equal to 3.

In binary search the number of necessary evaluations n_{Eval} is given by:

$$n_{Eval} = \log_2 n \tag{3}$$

in which n is the number of possible values in the original set.

4. Parameters identifiability based on the binary search

An important aspect in search procedures regarding to the computational cost is the numbers of required evaluations to find

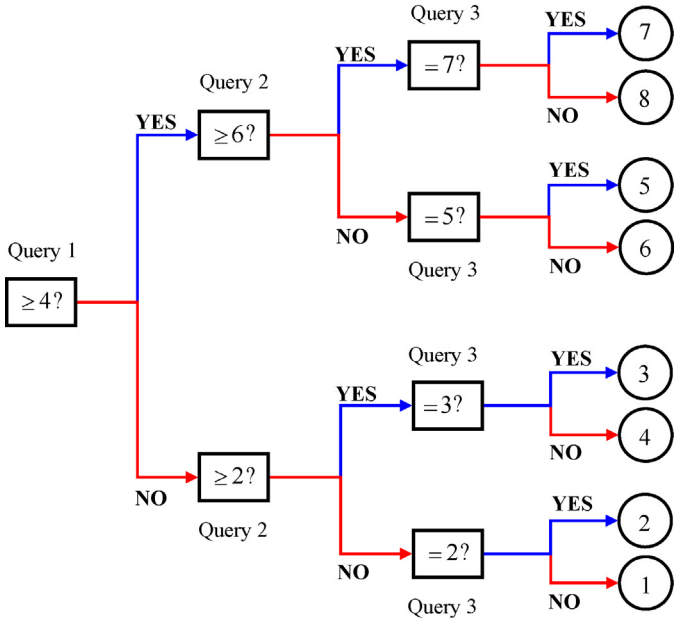

Fig. 3. Query tree of the example of set of numbers $\boldsymbol{\lambda}$ using the binary search.

a solution. In classical parameters identifiability procedures the number of evaluations may become very high, because the one by one selection of parameters criteria are chosen.

The numerical procedure proposed in this paper is presented in Fig. 4. In the algorithm, nP_{SS} represents the number of successfully selected parameters as identifiable, which is updated at each well-succeed selection. The algorithm tries to include more parameters in this set, being nP_S the number of parameters that are temporarily included in the set of identifiable parameters. However, if the set containing nP_S is not identifiable, the set will be reduced and nP_S will be updated accordingly. In the case of successive fails until nP_S reduces to nP_{SS}, no additional parameters could be selected and the procedure stops. The procedure can also stop when all the parameters are identifiable ($\Theta^{(NS)}$ is empty), or no parameters are identifiable. This last case only happens if some problem occurs for identifying even one parameter, indicating either a bad range established for the most important parameter or a modeling problem.

In order to explain the proposed procedure, in Fig. 5 it is applied to a model with four parameters, showing all possible combinations for this case. The parameters are sorted according to their importance for the model prediction (as in: Brun et al., 2001; Chu & Hahn, 2007; Li et al., 2004; Lund & Foss, 2008; Sandink et al., 2001; Secchi et al., 2006; Sun & Hahn, 2006; Weijers & Vanrolleghem, 1997; Yao et al., 2003, and others). According to the proposed procedure, the first half of the most relevant parameters of the rank are included in the set of selected parameters, represented by filled squares in Fig. 5. When adding parameters to the set of selected parameters, the conditioning of the estimation procedure should be evaluated in order to verify if such parameters can be estimated simultaneously, the invertibility of the matrix FIM can be used for such purpose. In this numerical procedure, the singularity of the FIM leads to the reduction of the set of selected parameters.

The main difference on the number of evaluations of the proposed procedure regard to the classical binary search is that, when a new parameter is included in the set of selected parameters and the FIM matrix is invertible, then the sort of the non-selected parameters may change (in the classical binary search, the sorted sequence do not change). Specially in the case where re-estimation of parameters is performed, even parameters that have already tried to

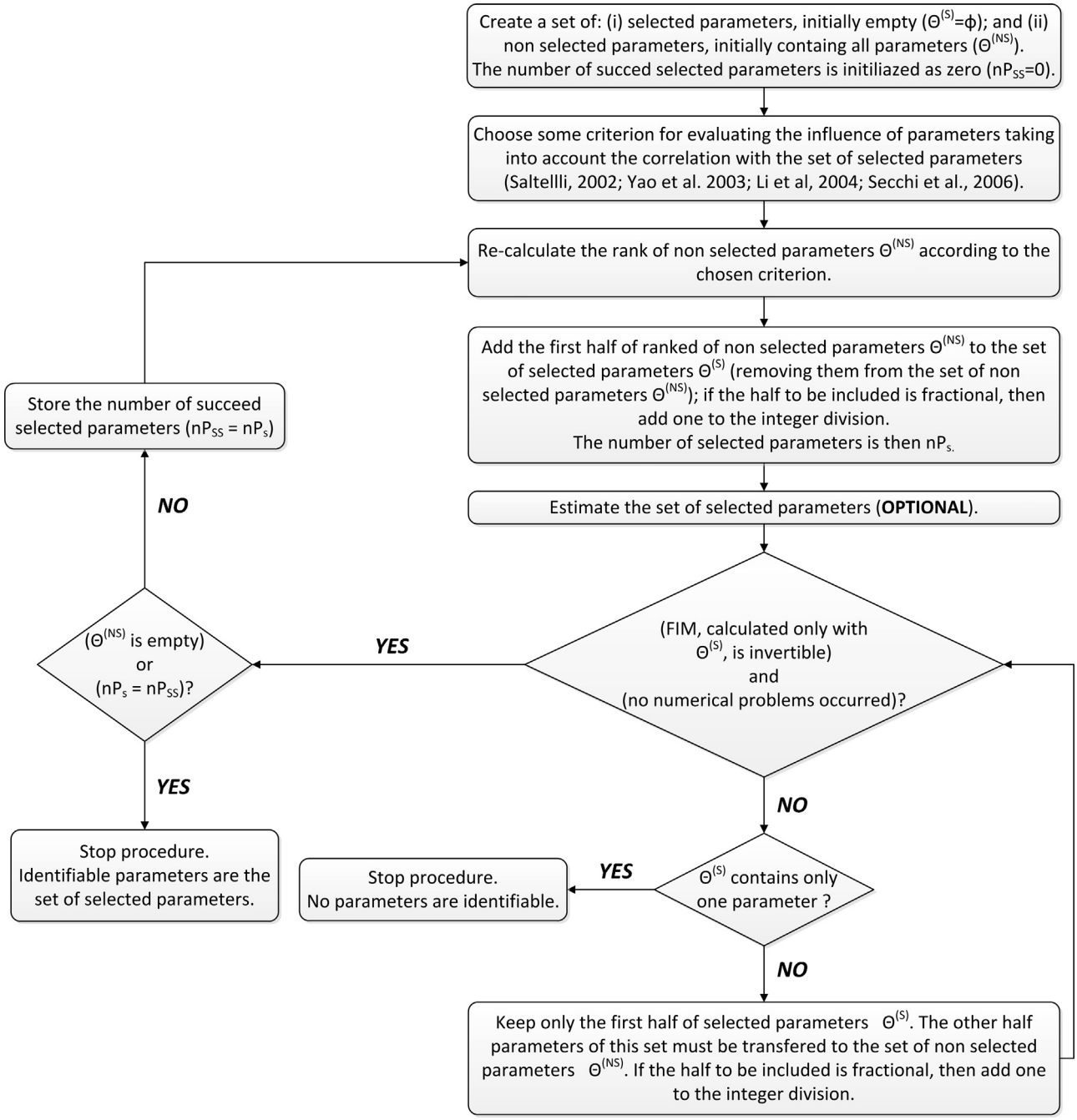

Fig. 4. Numerical procedure proposed for parameter identifiability: set-by-set selection.

been included without success can be now included, which may happen since the values of parameters change, and so the matrix FIM may change sufficiently to become invertible. Thus, the set of non-selected parameters should be re-evaluated according to the chosen criterion when a new set of parameters is successfully selected. It must be emphasized that, in the proposed procedure, once a parameter is successfully selected it will never be removed from the set of selected parameters.

In most procedures presented in the literature, the selected parameters re-estimation is only performed after the identifiability procedure. Without performing simultaneously with the identifiability procedure the estimation of the selected parameters, all the analyses are done based on the initial estimates of the parameters values, carrying great uncertainties out to the results. The simultaneous parameters re-estimation reduces the dependence of the identifiability procedure regarding to the initial estimates of the parameters values (Secchi et al., 2006). Moreover, since one of the main purposes of the parameters identifiability is to select the set of parameters that can be estimable by optimization algorithms that usually require the inversion of the Hessian matrix, with the simultaneous parameters re-estimation, it is possible to improve model fit to experimental data, resulting in a better prediction. Nevertheless, the computational time increases significantly when parameters re-estimation is performed simultaneously with identifiability procedure even so, the importance of this step should not be despised, becoming relevant the development

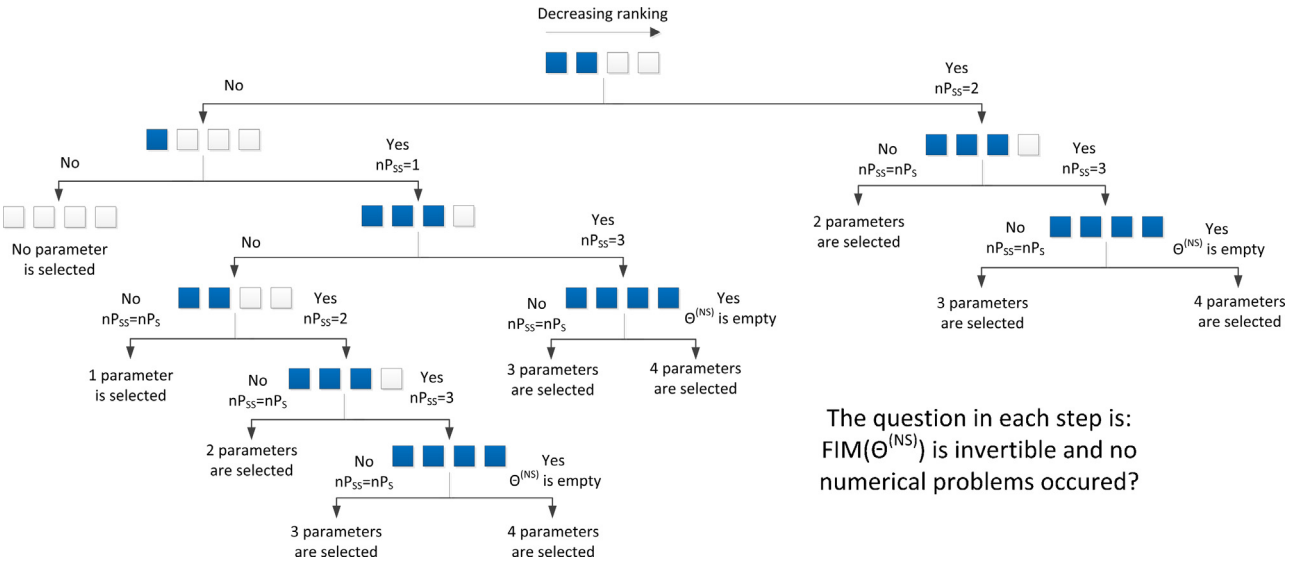

Fig. 5. Numerical procedure for 4 parameters to be investigated; nP_{SS}, nP_S represent the number of succeed selected parameters and the number of parameters in the selected set $\Theta^{(S)}$, $\Theta^{(NS)}$ is the set of non-selected parameters.

of identifiability procedures with less parameters estimation stages.

However, if one still decides to perform the parameters estimation only after the identifiability procedure, numerical problems may arise due to modeling errors and/or uncertainties in the initial estimates of the parameters. As previously mentioned, the quality of the initial estimates of the parameters is one of the major problems to the identifiability procedures. There is no guarantee that the resulting estimation problems will become well-posed, especially during the estimation step, where the parameters values may exceed pre-established parameters ranges, and also during the model evaluation step, where some combination of model parameters values may lead to infeasible solution. Although these problems are very common in parameters identifiability, they have been almost completely neglected in the literature.

When performing these steps simultaneously, as described in Fig. 4, it is possible to overcome problems that may occur during model solving or parameters re-estimation reducing the set of selected parameters. If the step of parameters re-estimation is performed in each selection, it is possible that the parameters discarded at this step may be selected in a posterior step, when an evaluation of the correlation between the set of selected and non-selected parameters is performed using more suitable values of the selected parameters.

Global optimization procedures that do not make use of the Hessian matrix could be used to obtain good initial estimates for the parameters values. However, in many cases of interest, one deal with complex dynamical models, which present significant numerical restrictions regarding to the parameters values without falling into numerical problems. Therefore, even derivative-free optimization methods, such as particle swarm optimization (Kennedy & Eberhart, 1995) or genetic algorithm (Goldberg, 1989), for obtaining good initial estimates, may fail in obtaining such values (Alberton et al., 2013).

An important question about the identifiability procedure is: what is the number of necessary evaluations? For the binary search (bs) (Eq. (3)) since there is the possibility of re-sorting the non-selected parameters when new parameters are selected, in the worst case, only one parameter would be included each time, and the re-estimation could make possible the inclusion of parameters that have also tried to be included without success. Therefore, the maximum number of re-estimation procedure calls or FIM matrix evaluation is given by:

$$N_{Eval}^{bs} = \underbrace{I_{\sup}(\log_2(nP))}_{\text{Evaluations to include only the first parameter}} + \underbrace{I_{\sup}(\log_2(nP-1))}_{\text{Evaluations to include only the second parameter}} + \cdots + \underbrace{I_{\sup}(\log_2(2))}_{\text{Evaluations to include only the penult parameter}} + \underbrace{1}_{\text{Evaluation to include or remove the last parameter}} \quad (4)$$

$$N_{Eval}^{bs} = 1 + \sum_{k=0}^{nP-2} I_{\sup}(\log_2(nP-k))$$

For the conventional procedures (cp) of parameters identifiability, where only one parameter is selected each time, in the worst case, where all the remaining parameters are evaluated and only at the last evaluation one parameter is included, allowing the re-evaluation of all remaining parameters again, the number of re-estimation procedure calls or FIM matrix evaluation is given by:

$$N_{Eval}^{cp} = \underbrace{nP}_{\text{Evaluations to include only the first parameter}} + \underbrace{(nP-1)}_{\text{Evaluations to include only the second parameter}} + \underbrace{(nP-2)}_{\text{Evaluations to include only the penult parameter}} + \cdots + \underbrace{1}_{\text{Evaluation to include or remove the last parameter}} \quad (5)$$

Fig. 6 presents the number of necessary evaluations for the worst case in procedures that try to select parameters one by one and for the proposed procedure. Although one may expect that the worst case will seldom happen, the proposed procedure really leads to a significant lower number of required evaluations. It is especially

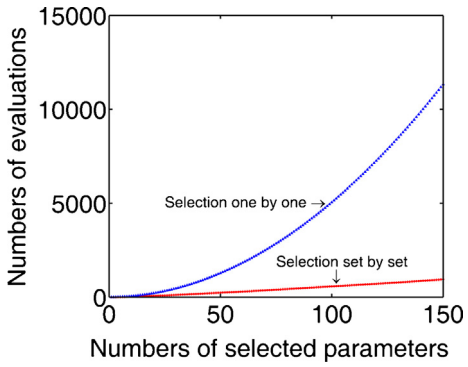

Fig. 6. Comparing between numbers of necessary evaluations for the worst case in procedures that try to select parameters one by one (▲) and the proposed numerical procedure (●).

important when simultaneous parameters re-estimation is performed together with identifiability procedure, reducing drastically the computational cost.

Most of the available criteria to sort the first set of parameters in the proposed procedure are based on the importance of the parameters to the model prediction. Usually, the parametric correlation is evaluated between set of selected parameters $\Theta^{(S)}$ and set of non-selected parameters $\Theta^{(NS)}$, avoiding the selection of parameters correlated with the parameters previously selected. Methods such as relative gain array – RGA (Bristol, 1966; Cao & Rossiter, 1997) and non-square relative gain array – NSRGA (Cao & Rossiter, 1997) can be used as criteria for parameters correlation evaluation previously to the first selection (Botelho, 2012; Sandink et al., 2001). In parameters identifiability problems, the use of NSRGA (Eq. (6)) is more indicated because the number of parameters is usually different from the number of measured variables of the model.

$$\text{NSRGA} = B \otimes (B^{+})^{T} \tag{6}$$

in which B^{+} is the pseudo inverse of the sensitivity matrix and $\otimes$ denotes the Hadamard (or element-wise) product.

In the proposed numerical procedure, the parameters correlation is evaluated at each step of selection, where parameters are definitively considered as identifiable parameters if the FIM is invertible. Therefore, if there are correlated parameters in the first set, then these parameters will be removed from this set during the FIM analysis. Moreover, the sum of each column of the NSRGA matrix is equivalent to the projection of the sensitivity vector of the corresponding parameter on the parametric space of the selected parameters, which is the criterion of linear independence used by Li et al. (2004) and Secchi et al. (2006).

Two examples to illustrate the application of the proposed procedure are presented in the next section. The results demonstrate that computational time and cost are lower than addressing parameters identifiability problem where the selection is made by evaluating the parameters one by one.

5. Illustrative examples

In the first example, a hypothetical problem of parameters identifiability with known solution is used for verifying if the proposed numerical procedure is able to make an adequate selection. In order to evaluate the robustness of the proposed numerical procedure when facing real problems, the second example represents a complex dynamic model containing 131 parameters that should be selected for estimation with scarce experimental data; in this problem the solution is unknown. On both cases, parameters re-estimation is performed.

The proposed numerical procedure was applied using the parameters identifiability criterion of Yao et al. (2003) for evaluation the importance and correlation of the parameters. Such criterion was selected due to its easy implementation and low computational cost. Nonetheless, any parameters identifiability criterion can be employed in the procedure.

The algorithm presented in Fig. 4 was implemented in FORTRAN 95, and all case studies were run in an Intel Core 2 Duo with 3 GB RAM. The parameter estimation was performed using the computational code Estima&Planeja (Schwaab et al., 2011), based on the Gauss–Newton method.

Parameters estimation with Gauss–Newton method is an optimization procedure that includes: (1) model evaluation, (2) linesearch for stepsize calculation with checking of parameters limits, (3) evaluation of FIM matrix and its inverse. As each of these steps may fail interrupting the estimation, some modifications were introduced in this code to improve the convergence for complex parameters identifiability problems that have scarce experimental data.

For both examples, the results are presented with the values of the re-estimated parameters together with the relative deviation, calculated as the ratio between the standard deviation of the parameter and its estimated value (σ_θ/θ).

5.1. Example 1 – a simple model

Let us consider the mathematical model presented in Eq. (7).

$$y = \sum_{k=1}^{nX} x_k \cdot \theta_k + \sum_{i=1}^{nX-1} \sum_{j=1}^{nX} x_i \cdot x_j \cdot \theta_{i,j} \tag{7}$$

in which θ denotes the parameter, x the independent variable vector of dimension nX, and y the dependent variable.

The original mathematical model is described only by main effects, where the number of parameters, nP, is equal to the number of independent variables, $nX = 31$. The crossed effects were added to the original model to create a super-parameterized model and increase the complexity of the identifiability problem. It is expected that the $31 \times 30 = 930$ parameters relating to the second-order terms become not identifiable because they are correlated with each other and were not employed to generate the simulated experimental data (y). In this case, the crossed effects are confounded with the main effects.

The values of the independent variables x for generating the simulated experimental data were obtained using the software Statistica 6.0 by Taguchi's experimental design. The dependent variable y was generated using the output of the original model added by a random noise with normal distribution, zero mean and variance $\sigma_Y^2 = 0.25$. The true parameters values of the original model form an arithmetic progression with the first parameter $\theta_1 = 5$, the last parameter $\theta_{31} = 155$, and the common difference of successive members equal to 5. In this scenario the simulated experimental data are orthogonal and the mathematical model structure is linear.

Two cases were investigated: without and with re-estimation. Moreover, since the parameters identifiability procedures are strongly dependent on the initial estimates of parameters values (Chu & Hahn, 2007; Omlin & Reichert, 1999), different values were given for the parameters as initial estimates. For linear models on the parameters this dependence is generated by using the normalized sensitivity matrix and also by the non-selected parameters when a re-estimation step is applied. In both cases, the results are presented in Table 1.

Table 1
Results for parameters identifiability procedure based on binary search for cases (a) and (b) of the first example. The superscripts represent the positions of the parameters.

Symbol	Exact value	Initial estimates (a.1 and b.1)	Initial estimates (a.2 and b.2)	Estimated values (b.1)	Estimated values (b.2)
θ_1	5	1.2	31	$159.265^{(31)}$	$14.1113^{(1)}$
θ_2	10	2	30	$152.752^{(30)}$	$13.8686^{(2)}$
θ_3	15	3	29	$144.396^{(29)}$	$17.6081^{(3)}$
θ_4	20	4	28	$145.181^{(28)}$	$23.0792^{(4)}$
θ_5	25	5	27	$138.824^{(27)}$	$26.5707^{(5)}$
θ_6	30	6	26	$134.280^{(26)}$	$31.0258^{(6)}$
θ_7	35	7	25	$129.384^{(25)}$	$39.2003^{(7)}$
θ_8	40	8	24	$120.687^{(24)}$	$45.6158^{(8)}$
θ_9	45	9	23	$123.449^{(23)}$	$46.3685^{(9)}$
θ_{10}	50	10	22	$115.398^{(22)}$	$57.3666^{(10)}$
θ_{11}	55	11	21	$108.222^{(21)}$	$63.0269^{(11)}$
θ_{12}	60	12	20	$104.095^{(20)}$	$64.8253^{(12)}$
θ_{13}	65	13	19	$103.921^{(19)}$	$72.3390^{(13)}$
θ_{14}	70	14	18	$88.3366^{(18)}$	$68.5984^{(14)}$
θ_{15}	75	15	17	$85.0125^{(17)}$	$76.4169^{(15)}$
θ_{16}	80	16	16	$86.7994^{(16)}$	$86.7994^{(16)}$
θ_{17}	85	17	15	$76.4169^{(15)}$	$85.0125^{(17)}$
θ_{18}	90	18	14	$68.5984^{(14)}$	$88.3366^{(18)}$
θ_{19}	95	19	13	$72.3390^{(13)}$	$103.921^{(19)}$
θ_{20}	100	20	12	$64.8253^{(12)}$	$104.095^{(20)}$
θ_{21}	105	21	11	$63.0269^{(11)}$	$108.222^{(21)}$
θ_{22}	110	22	10	$57.3666^{(10)}$	$115.398^{(22)}$
θ_{23}	115	23	9	$46.3685^{(9)}$	$123.449^{(23)}$
θ_{24}	120	24	8	$45.6158^{(8)}$	$120.687^{(24)}$
θ_{25}	125	25	7	$39.2003^{(7)}$	$129.384^{(25)}$
θ_{26}	130	26	6	$31.0258^{(6)}$	$134.280^{(26)}$
θ_{27}	135	27	5	$26.5707^{(5)}$	$138.824^{(27)}$
θ_{28}	140	28	4	$23.0792^{(4)}$	$145.181^{(28)}$
θ_{29}	145	29	3	$17.6081^{(3)}$	$144.396^{(29)}$
θ_{30}	150	30	2	$13.8686^{(2)}$	$152.752^{(30)}$
θ_{31}	155	31	1.2	$14.1113^{(1)}$	$159.265^{(31)}$
$\theta_{1,1}$ to $\theta_{30,31}$	0	1	1	[a]	[a]

[a] Non-selected parameters.

5.1.1. Case a – without the parameters re-estimation step

The main objective of this case is to demonstrate the lower time and computational cost of the proposed numerical procedure compared to usual procedures described in the literature. In fact, only 4 evaluations were required for selection of 31 parameters, while with conventional parameters identifiability procedures it would be required at least 31 evaluations. Note that the worst case of the binary search would require 8588 evaluations (Eq. (4)), and the worst case of the conventional procedures would require 461,281 evaluations (Eq. (5)).

In order to demonstrate the dependence on the initial estimates of the parameters values, two initial estimates were investigated:

- Case a.1 – values of the parameters 1–31 in crescent order of magnitude, with parameters related to the crossed effects kept equal to one.
- Case a.2 – values of the parameters 1–31 in decreasing order of magnitude, with parameters related to the crossed effects kept equal to one.

As expected for the case a.1, the parameters were selected in decreasing order according to their values, being obtained the set of the selected parameters $\Theta^{[S]} = \{\theta_{31}, \theta_{30}, \ldots, \theta_1\}$. For the case a.2, the parameters were selected according to their values, that is, $\Theta^{[S]} = \{\theta_1, \theta_2, \ldots, \theta_{31}\}$. It is important to empathize that for more complex problems (e.g., scarce experimental data, and non-linear models) the results of parameters selection are highly dependent on the employed parameters identifiability criteria. Despite the seriousness of the example, this problem of the dependence on the initial estimates also occurs in the conventional parameters identifiability procedures.

The CPU time spent for the selections is the same in both cases.

5.1.2. Case b – with the parameters re-estimation step

Similar to the case (a), investigations were performed with different initial estimates as follows:

- Case b.1 – values of the parameters 1–31 in crescent order of magnitude, with parameters related to the crossed effects kept equal to one.
- Case b.2 – values of the parameters 1–31 in decreasing order of magnitude, with parameters related to the crossed effects kept equal to one.

Introducing the step of parameters re-estimation, 31 parameters were selected, where the parameters selection order was the same as obtained without re-estimation. The estimated parameters values are presented in Table 1; for both cases, the relative deviations were about 1×10^{-2}, indicating a good accuracy of the estimated parameters. The CPU time spent in case b.1 was 1.30 min and in case b.2 was 1.84 min. In both cases, the results of estimation were the same, as expected, and the re-estimations were able to get good estimates for the true parameters values. The difference between re-estimated parameters and true parameters values is mainly associated with the crossed effects, that are zero for the exact model and equal to one for the initial estimates, and such values were kept unchanged during the estimation since the crossed effects were not included in the set of estimable parameters.

In this example, the NSRGA was also applied to select the first set of parameters, in order to avoid correlated parameters, leading to the same results.

5.2. Example 2 – a complex dynamic model

Representing a complex real case of parameters identifiability problems (complex dynamic model, large number of parameters,

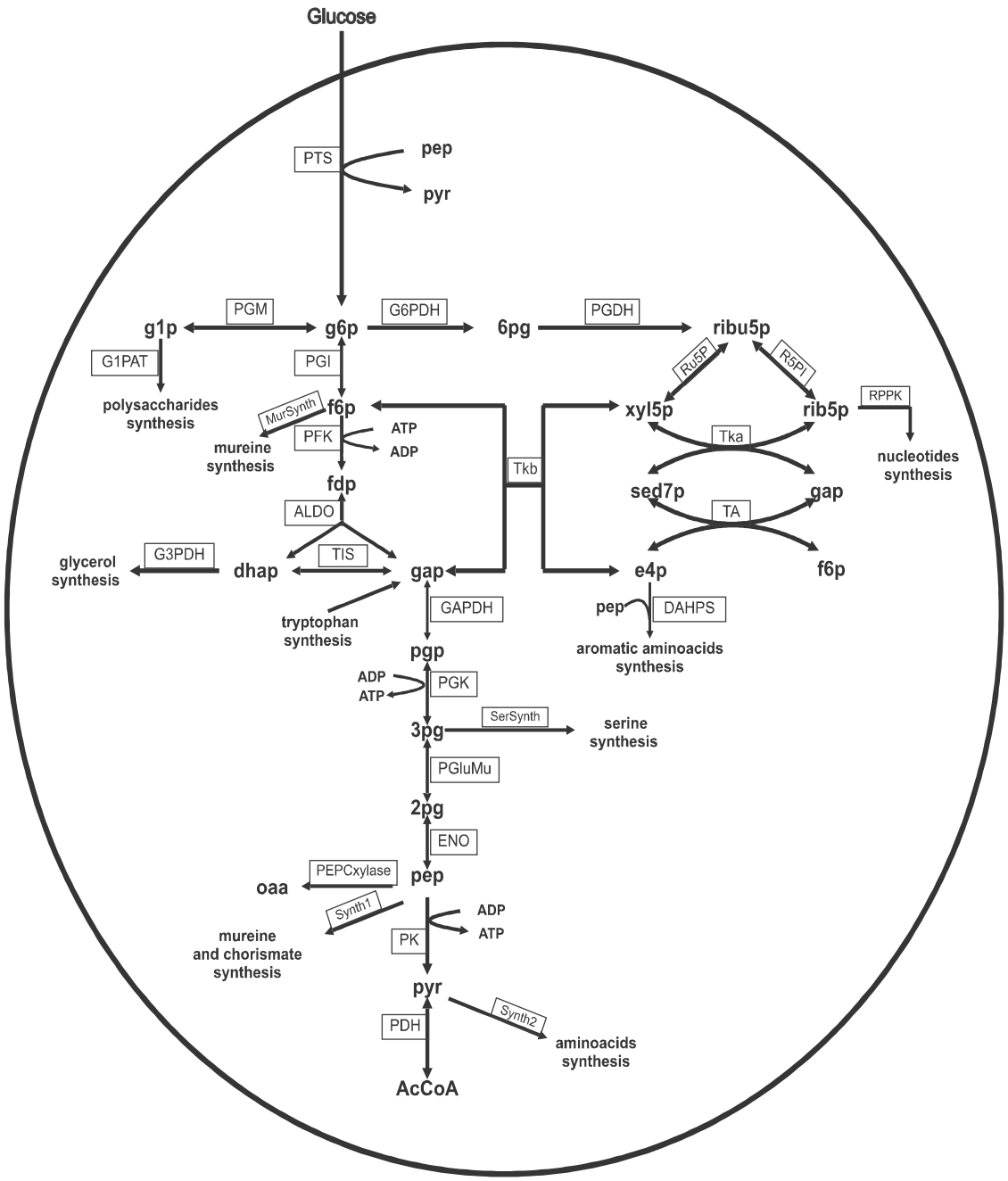

Fig. 7. Escherichia coli central carbon metabolism.

parameters with low influence for model prediction and/or correlated, and scarce experimental data), the *E. coli* K-12 W3110 metabolic network model (Chassagnole, Noisommit-Rizzi, Schimd, Mauch, & Reuss, 2002; Di Maggio et al., 2009) is used in this example. Additionally, conducting experiments for such system are costly, time consuming and difficult. Consequently, some important variables are not measured and their initial conditions are unknown.

Large scale metabolic network of the glycolysis, the pentose-phosphate-pathway and the phosphotransferase system of *E. coli* K-12 W3110 is presented in Fig. 7.

The mathematical model of *E. coli* K-12 W3110 metabolic networks is described by a system of 18 ordinary differential equations that represent mass balances for extracellular glycolysis and intracellular metabolites (Eqs. (8)–(25)), 7 forcing functions that represent the co-metabolites (Eqs. (A1)–(A7)) and 30 kinetic rates that represent the enzymes (Eqs. (A8)–(A37)) (Chassagnole et al., 2002). This dynamic model has 131 candidate parameters to be selected.

Modifications were introduced by Di Maggio, Diaz Ricci, and Diaz (2010) for some enzyme kinetics. In this sense, kinetic expression for the activity of phosphofructokinase was taken from Diaz Ricci (1996) (Eq. (A5)) and other modifications are given by Ratushny et al. (2006) for modeling the activity of glucose-6-phosphate dehydrogenase (Eq. (A6)).

$$\frac{dC_{\mathrm{glc}}^{\mathrm{extracellular}}}{dt} = D(C_{\mathrm{glc}}^{\mathrm{feed}} - C_{\mathrm{glc}}^{\mathrm{extracellular}}) + f_{\mathrm{pulse}} - \frac{C_x r_{\mathrm{PTS}}}{\rho_x} \tag{8}$$

$$\frac{dC_{\mathrm{g6p}}}{dt} = r_{\mathrm{PTS}} - r_{\mathrm{PGI}} - r_{\mathrm{G6PDH}} - r_{\mathrm{PGM}} - \mu C_{\mathrm{g6p}} \tag{9}$$

$$\frac{dC_{f6p}}{dt} = r_{PGI} - r_{PFK} + r_{TKb} + r_{TA} - 2r_{MurSynth} - \mu C_{f6p} \quad (10)$$

$$\frac{dC_{fdp}}{dt} = r_{PKF} - r_{ALDO} - \mu C_{fdp} \quad (11)$$

$$\frac{dC_{gap}}{dt} = r_{ALDO} + r_{TIS} - r_{GAPDH} + r_{TKa} + r_{TKb} - r_{TA} + r_{TrpSynth} - \mu C_{gap} \quad (12)$$

$$\frac{dC_{dhap}}{dt} = r_{ALDO} - r_{TIS} - r_{G3PDH} - \mu C_{dhap} \quad (13)$$

$$\frac{dC_{pgp}}{dt} = r_{GAPDH} - r_{PGK} - \mu C_{pgp} \quad (14)$$

$$\frac{dC_{3pg}}{dt} = r_{PGK} - r_{PGGluMu} - r_{SerSynth} - \mu C_{3pg} \quad (15)$$

$$\frac{dC_{2pg}}{dt} = r_{PGluMu} - r_{ENO} - \mu C_{2pg} \quad (16)$$

$$\frac{dC_{pep}}{dt} = r_{ENO} - r_{PK} - r_{PTS} - r_{PEPCxylase} - r_{DAHPS} - r_{Synth1} - \mu C_{pep} \quad (17)$$

$$\frac{dC_{pyr}}{dt} = r_{PK} + r_{PTS} - r_{PDH} - r_{Synth2} + r_{MetSynth} + r_{TrpSynth} - \mu C_{pyr} \quad (18)$$

$$\frac{dC_{6pg}}{dt} = r_{G6PDH} - r_{PGDH} - \mu C_{6pg} \quad (19)$$

$$\frac{dC_{ribu5p}}{dt} = r_{PGDH} - r_{Ru5P} - r_{R5PI} - \mu C_{ribu5p} \quad (20)$$

$$\frac{dC_{xyl5p}}{dt} = r_{Ru5P} - r_{TKa} - r_{TKb} - \mu C_{xyl5p} \quad (21)$$

$$\frac{dC_{sed7p}}{dt} = r_{TKa} - r_{TA} - \mu C_{sed7p} \quad (22)$$

$$\frac{dC_{rib5p}}{dt} = r_{R5PI} - r_{TKa} - r_{RPPK} - \mu C_{rib5p} \quad (23)$$

$$\frac{dC_{e4p}}{dt} = r_{TA} - r_{TKb} - r_{DAHPS} - \mu C_{e4p} \quad (24)$$

$$\frac{dC_{glp}}{dt} = r_{PGM} - r_{GlPAT} - \mu C_{glp} \quad (25)$$

The computational code DASSL (Petzold, 1989) was employed for the numerical integration of the dynamic model.

The experimental data were obtained in Hoque, Ushiyama, Tomita, and Shimizu (2005) for a time horizont of 300 s and for the following measured variables: glucose (glc), dihydroxyacetonephosphate (dhap), erythrose-4-phosphate (e4p), phosphoenolpyruvate (pep), pyruvate (pyr), fructose-1,6-diphosphate (fdp), glyceraldehyde-3-phosphate (gap), ribose-5-phosphate (rib5p), ribulose-5-phosphate (ribu5p), 2-phosphoglicerate (2pg), glucose-6-phosphate (g6p), fructose-6-phosphate (f6p), and 6-phosphogluconate (6pg). The other intracellular metabolites were not measured and thus, their initial conditions are unknown. Since these initial conditions are important information for the mathematical model, their values should be estimated together with the parameters. The results will be presented for two cases: without and with the parameters re-estimation step. As usual, for both cases, the experimental uncertainty is admitted proportional to the measured values of the independent variable (y_i) in the form:

$$\sigma_{y_i} = a \cdot y_i + b \quad (26)$$

where a and b can be arbitrarily chosen or obtained experimentally, or even estimated.

The knowledge of experimental uncertainty is a significant problem for the majority of investigated systems in chemical engineering. The appropriate knowledge would require replications for all experiments. It is important information, but often it is neglected and arbitrarily chosen due to the hard work that would be to characterize the experimental errors. In our case, we admitted $a = 10^{-2}$ and $b = 10^{-6}$ (particularly the value of b was chosen based on the magnitude of experimental data values).

Table 2
Results for parameters identifiability based on the information theory of *E. coli* metabolic networks.

Symbol	Nominal value	Symbol	Nominal value
Initial conditions			
C_{xyl5p}	5.5406×10^{-1}	C_{3gp}	1.2851
C_{g1p}	6.7016×10^{-2}	C_{sed7p}	4.0874×10^{-2}
C_{pgp}	3.7000×10^{-5}		
Parameters			
$K_{PGI,g6p}$	4.0000×10^{-1}	$K_{GAPDH,pgp}$	5.0000×10^{-5}
r^{max}_{PGI}	1.0519×10^{3}	$K_{G6PDH,nad,ihn}$	9.0000×10^{-3}
r^{max}_{G6PDH}	1.2873×101	$r^{max}_{Synthesis2}$	5.6900×10^{-1}
$K_{PGI,g6p2}$	2.0000×10^{-1}	r^{max}_{PK}	5.6700×10^{-1}
$K_{TIS,eq}$	1.3900×10^{-1}	r^{max}_{PTS}	8.21073×10^{4}
$K_{PEPCxylase,fdp}$	7.0000×10^{-1}	K_{PTS1}	3.0823×10^{3}
$n_{PEPCxylase,fdp}$	4.0000	$K_{Synthesis2,pyr}$	1.0000
$K_{RU5P,eq}$	1.4000	r^{max}_{PGDH}	5.22076
r^{max}_{GAPDH}	50.0000	$K_{PGDH,nadp}$	5.0600×10^{-2}
$K_{GAPDH,gap}$	1.2000	$K_{PGDH,nadph}$	1.3769×10^{-2}

5.2.1. Case a – without the parameters re-estimation step

Considering 5 unknown initial conditions and 131 model parameters, 136 parameters were evaluated by the proposed numerical procedure using the available experimental data for intracellular metabolites. The results are summarized in Table 2, where 20 parameters and 5 initial conditions were selected as identifiable, being required only 13 evaluations. The high number of non-selected parameters may indicate that a large number of parameters present low influence for model prediction and/or are correlated with the selected parameters, and that the non-measured intracellular metabolites store relevant information for describing *E. coli* metabolic behavior. Other important factor is related to the initial estimates of parameters values that can be far from the true parameters values, and are used in the evaluation of parameters identifiability criterion.

Table 3
Results of the estimation of initial conditions for non-measured intracellular metabolites using the proposed numerical procedure with parameters re-estimation.

Symbol	Nominal value	Estimated value	Relative deviation
C_{xyl5p}	5.5406×10^{-1}	5.2742×10^{-1}	8.9433×10^{3}
C_{g1p}	6.7016×10^{-2}	6.4919×10^{-2}	9.8674×10^{1}
C_{pgp}	3.7000×10^{-5}	3.0857×10^{-3}	3.3957
C_{3gp}	1.2851	1.7901	3.5516
C_{sed7p}	4.0874×10^{-2}	4.0850×10^{-2}	1.4311×10^{4}

Table 4
Results of parameters selection using the proposed numerical procedure with parameters re-estimation.

Parameter	Nominal value	Estimated value	Relative deviation	Parameter	Nominal value	Estimated value	Relative deviation
$K_{PGI,g6p}$	4.0000×10^{-1}	1.0769×10^{2}	3.0814	$K_{PGK,adp}$	1.8500×10^{-1}	1.8500×10^{-1}	1.9104×10^{2}
r^{max}_{PGI}	1.0519×10^{3}	5.9234×10^{2}	1.2569	$K_{GAPDH,nad}$	2.5200×10^{-1}	2.5213×10^{-1}	6.7565×10^{1}
r^{max}_{G6PDH}	1.2873×10^{1}	7.2217	2.0269×10^{1}	$K_{G6PDH,nad,inh}$	9.0000×10^{-3}	8.6536×10^{-3}	6.1194×10^{1}
$K_{PGI,f6p2}$	2.0000×10^{-1}	1.1109×10^{-5}	7.6247×10^{5}	r^{max}_{RU5P}	5.1816	2.2516	2.5393×10^{1}
$K_{PGK,3pg}$	5.1700×10^{-1}	4.5490×10^{-1}	4.4129	r^{max}_{R5PI}	3.7333	1.9924	4.6917
$K_{PEPCxylase,fdp}$	7.0000×10^{-1}	4.4571×10^{-1}	3.7857	r^{max}_{ALDO}	3.0423×10^{1}	4.0000	1.7730×10^{3}
$n_{PEPCxylase,fdp}$	4.0000	9.2161	2.3282	$K_{ENO,2pg}$	1.0000×10^{-1}	1.0000×10^{-1}	1.6760×10^{1}
r^{max}_{PGDH}	5.2208	9.4885×10^{-1}	1.2338×10^{3}	$K_{TIS,dhap}$	2.8000	2.5225	2.2587×10^{3}
r^{max}_{PFK}	1.9670	1.4669×10^{-1}	2.5160	$K_{R5PI,eq}$	4.0000	1.4200	3.5450×10^{1}
r^{max}_{GAPDH}	5.0000×10^{2}	9.4187	9.5864	r^{max}_{PDH}	4.0104	6.2336×10^{-1}	7.4603×10^{-1}
$K_{RU5P,eq}$	1.4000	1.4855	8.9487×10^{3}	$K_{PGLUMU,3pg}$	2.0000×10^{-1}	2.0000×10^{-1}	6.9664×10^{2}
$K_{GAPDH,gap}$	1.2000	2.2312×10^{-2}	1.3122×10^{3}	$H_{G6PDH,nadh}$	1.2000×10^{-1}	5.1769×10^{-1}	5.1116×10^{3}
$K_{GAPDH,pgp}$	5.0000×10^{-5}	8.3295×10^{-5}	1.3434×10^{3}	$K_{TKa,eq}$	1.3500×10^{-1}	1.1685×10^{-1}	2.3249×10^{4}
$K_{PGDH,nadp}$	5.0600×10^{-2}	1.1772	1.3215×10^{3}	$K_{ENO,pep}$	5.2056×10^{1}	1.3500×10^{-1}	3.9731×10^{1}
$K_{PGDH,nadph}$	1.3769×10^{-2}	7.3728×10^{-2}	4.0426×10^{2}	r^{max}_{TIS}	2.2000	8.7199×10^{-2}	1.9057×10^{3}
$K_{PGDH,atp}$	2.0800×10^{2}	5.4786×10^{1}	4.2688×10^{3}	$n_{DAHPS,pep}$	3.6900×10^{-1}	2.4990	2.6254×10^{1}
$K_{PGM,eq}$	1.5800×10^{-1}	1.3716×10^{-1}	8.7306×10^{1}	$K_{PGLUMU,2pg}$	3.0000×10^{-2}	3.6900×10^{-1}	6.3087×10^{2}
$K_{ALDO,eq}$	1.4438×10^{-2}	1.4935×10^{-2}	2.6863	$K_{G6PDH,nadph}$	1.4000	2.8818×10^{-2}	7.5260×10^{3}
r^{max}_{Synth2}	5.6900×10^{-1}	4.2606×10^{-2}	2.6446×10^{1}	$H_{G6PDH,nadph}$	8.8572×10^{-3}	1.4032	8.6816×10^{3}
r^{max}_{PK}	5.6700×10^{-1}	3.5304×10^{-2}	6.0122×10^{1}	$K_{R,f6p}$	2.0000×10^{-2}	8.8801×10^{-3}	9.6690×10^{3}
r^{max}_{PTS}	1.0000	1.0745	3.5897×10^{3}	$K_{MG6PDH,nadp}$	1.0500	1.9406×10^{-2}	3.7786×10^{3}
K_{PTS1}	2.6000×10^{-1}	2.3386×10^{3}	3.6104×10^{3}	$K_{TA,eq}$	1.3900×10^{-1}	1.0507	1.4314×10^{4}
$K_{Synthesis2,pyr}$	4.6800×10^{-2}	1.8140×10^{-1}	1.5514×10^{2}	$K_{TIS,eq}$	9.5000×10^{-2}	1.3896×10^{-1}	2.9680×10^{2}
$K_{PK,adp}$	2.6000	2.5944×10^{-1}	2.9091×10^{2}	$r^{max}_{PEPCxylase}$	8.8000×10^{-2}	1.0256×10^{-2}	6.2233×10^{1}
$K_{PGK,pgp}$	1.8700×10^{-1}	5.4125×10^{-2}	3.5774	$K_{ALDO,dhap}$	9.9000×10^{-1}	8.8000×10^{-2}	1.4290×10^{2}
$n_{DAHP,se4p}$	3.1000×10^{-2}	3.4499	3.5791	e	5.6180×10^{-1}	4.0301×10^{-1}	9.6704×10^{3}
$K_{PGLUMU,eq}$	1.0000	9.4505×10^{-1}	7.0688	r^{max}_{PGM}	5.4491	9.0650×10^{-2}	5.7009×10^{1}
$K_{PK,pep}$	4.3000×10^{-2}	3.0968×10^{-2}	1.4155×10^{2}	$K_{PGDH,6pg}$	3.0000×10^{-1}	3.6871	1.1154×10^{2}
n_{PDH}	4.0700	9.2200	1.5091	$K_{G6PDH,nadh}$	3.0000×10^{-1}	3.0070×10^{-1}	1.0491×10^{4}
$K_{PGI,eq}$	6.7000×10^{-1}	4.2994×10^{-2}	7.1124×10^{-1}	$K_{G6PDH,dtn}$	7.0000×10^{-3}	3.0613×10^{-1}	4.4899×10^{4}
$K_{PEPCxylase,pep}$	8.4613	8.0500	6.6521×10^{1}	$K_{MG6PDH,g6p}$	1.8000×10^{3}	6.9652×10^{-3}	2.2991×10^{2}
$K_{ENO,eq}$	2.0000×10^{-1}	7.1033×10^{-1}	2.0301	$K_{PGK,eq}$	1.0000	8.9599×10^{-3}	2.2371×10^{3}
r^{max}_{TA}	4.2260×10^{3}	3.2541×10^{2}	1.4314×10^{4}	$K_{TIS,gap}$	1.5000×10^{-3}	3.4455×10^{-1}	3.0763×10^{2}
$K_{PGI,g6p2}$	2.5274×10^{2}	5.9448×10^{-1}	4.9639	r^{max}_{G3PDH}	5.3000×10^{-3}	1.0700×10^{-2}	5.4517×10^{4}
r^{max}_{PGK}	7.3718	2.7500×10^{1}	2.3841×10^{3}	$K_{G3PDH,dhap}$	1.8500×10^{-1}	6.5999	6.1520×10^{4}
r^{max}_{ENO}	9.6972×10^{1}	2.5000	1.0603×10^{1}	$r^{max}_{MurSynth}$	2.5200×10^{-1}	1.4961×10^{-3}	1.6909×10^{2}
r^{max}_{TKa}	4.0000×10^{-1}	5.7700	8.9491×10^{3}	$K_{DAHPS,pep}$	9.0000×10^{-3}	5.3000×10^{-3}	1.5787×10^{2}
r^{max}_{PGLUMU}	1.0519×10^{3}	7.0600×10^{-1}	6.8782×10^{1}				

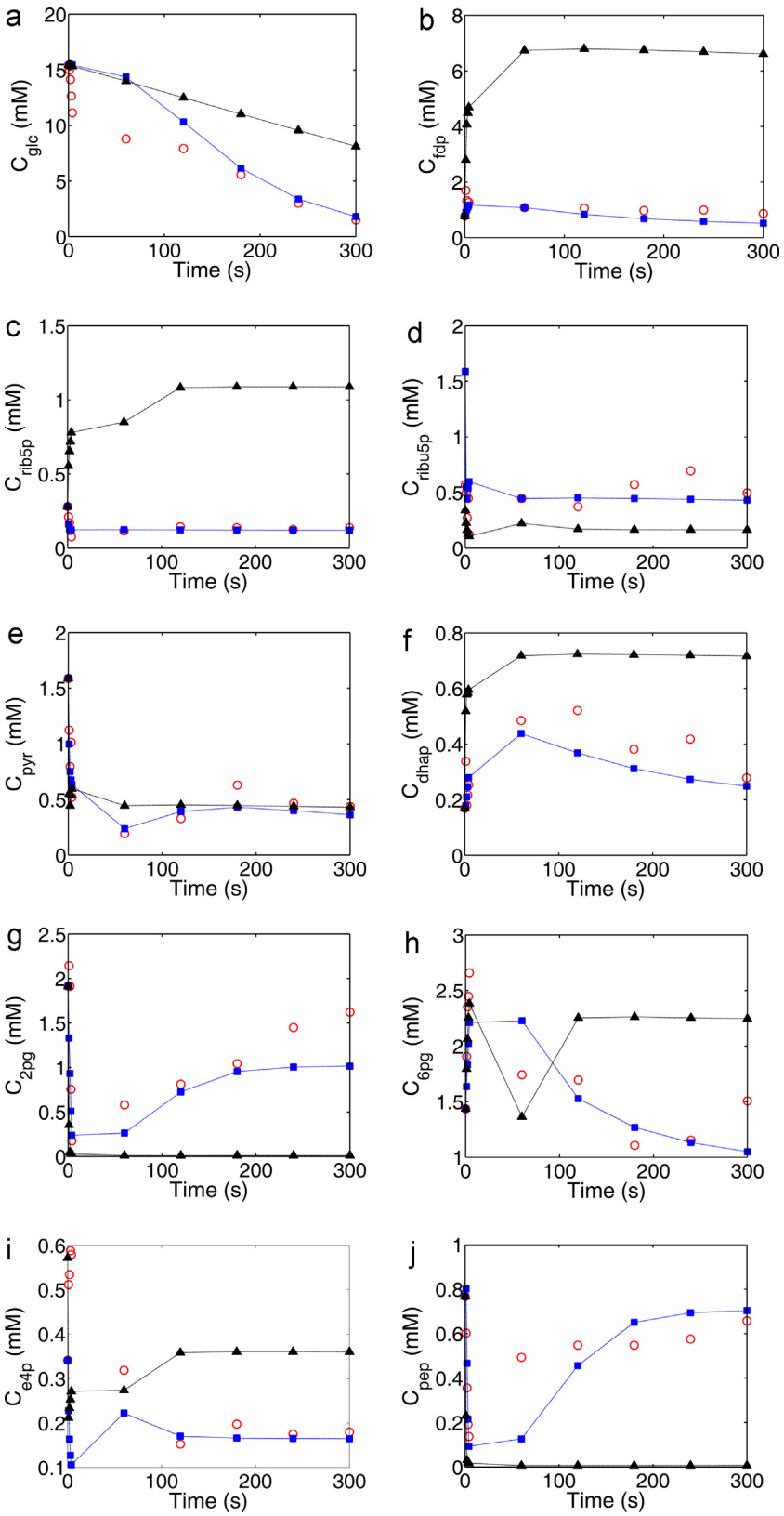

Fig. 8. Experimental and predicted metabolites concentrations as function of the time: (O) experimental value, (-▲-) predicted value using initial estimates and (-■-) predicted value after parameter identifiability.

5.2.2. Case b – with the parameters re-estimation step

Including the step of parameters re-estimation, the proposed numerical procedure was able to select 75 parameters and 5 unknown initial conditions, demanding only 19 iterations. Therefore, 55 additional parameters were selected by introducing the parameters re-estimation step with only 6 additional evaluations.

It must be emphasized that the selection of the parameters using the one by one procedure would require at least 80 iterations. Actually, the save of computational time is much higher, since the usual procedure demands much higher iterations, forasmuch the inclusion of one parameter at each iteration often is not succeed, and several parameters must be tested before one additional parameter is definitively included. The CPU time spent with the proposed procedure was only 3.24 min. For comparison, we have performed the identification procedure adding one parameter each time, and obtained the results in 2.28 h, with similar results. The results are summarized in Tables 3 and 4, respectively, for both parameters and initial conditions.

The initial estimates of the initial conditions, presented in Table 3, were obtained in two steps: (1) integrate the model with a small timestep and null initial concentrations of the non-measured intracellular metabolites, and (2) take the calculated values for the non-measured intracellular metabolites concentrations at the end of the first timestep as initial estimates for the initial conditions. With the parameters re-estimation step, these values do not present a significant influence.

Most of the estimated parameters values presented in Table 4 present difference of at least one order of magnitude with respect to the initial estimates. This difference demonstrates that initial estimates of the parameters could be significantly improved, as observed in Fig. 8, comparing experimental data with model prediction using initial estimates and estimated values after the parameters identifiability. It is important to emphasize that the estimated values of the selected parameters can be significantly influenced by the initial estimates of the non-selected parameters. Special attention should be given at Fig. 8(b), (c) and (g)–(j), for which the mathematical model behavior with re-estimated parameters values was able to lead to a better fit to the experimental data.

Applying the NSRGA criteria to select the first set of parameters in order to avoid correlated parameters did not change the results. Therefore, the additional overhead to select the first set of parameters seems to be unnecessary.

6. Conclusion

In this work, a numerical procedure for parameters selection based on binary search was proposed in order to solve the problem of parameters identifiability with minimum effort. Two examples were presented to illustrate the performance of the procedure. In the first example a simple case was presented showing that the numerical procedure is able to select the parameters according to their potential of estimation. The second example is a scarce experimental data scenario, using a complex mathematical model of *E. coli* metabolic network, containing 131 parameters and 5 unknown initial conditions of metabolites concentration. In both cases, the proposed procedure leads to significant reduction of the computational effort and similar results when compared to the traditional procedures of one by one parameter selection.

The use of NSRGA for previous evaluation of parameters correlation did not improve the results for the investigated examples. In fact, the results obtained with the NSRGA are overlapped by evaluating the correlation parameter performed in each step of the selection.

A discussion about the simultaneous parameters re-estimation together with parameters identifiability procedures was presented, and it was shown that such simultaneous step is very important for reducing the influence of uncertainties of initial parameters estimates.

Acknowledgments

The authors thank CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), National Research Council (CONICET), Universidad Nacional del Sur and ANPCYT, for providing scholarships and for supporting this work.

Appendix A.

See Tables A1–A4.

Table A1
Nomenclature.

Metabolites	
glc	Glycose
g6p	Glycose-6-phosphate
f6p	Glycose-6-phosphate
fdp	Fructose-1,6-biphosphate
gap	Glyceraldehyde-3-phosphate
dhap	Dihydroxyacetonephosphate
pgp	1,3-Diphosphoglycerate
3pg	3-Phosphoglicerate
2pg	2-Phosphoglicerate
pep	Phosphoenolpyruvate
pyr	Pyruvate
6pg	6-Phosphogluconate
ribu5p	Ribulose-5-phosphate
xyl5p	Xylulose-5-phosphate
sed7p	Sedoheptulose-7-phosphate
xyl5p	Xylulose-5-phosphate
sed7p	Sedoheptulose-7-phosphate
rib5p	Ribose-5-phosphate
e4p	Erythrose-4-phosphate
glp	Glycose-1-phosphate
Co-metabolites	
atp	Adenosintriphosphate
adp	Adenosindiphosphate
nad	Diphosphopyridindinucleotide, oxized
nadh	Diphosphopyridindinucleotide, reduced
nadp	Diphosphopyridindinucleotide-phosphate, oxized
nadph	Diphosphopyridindinucleotide-phosphate, reduced
amp	Adenosinmonophosphate
Enzymes	
ALDO	Aldolase
DAHPS	DAHPS synthase
ENO	Enolase
G1PAT	Glycose-1-phosphate adenyltransferase
G3PDH	Glycerol-3-phosphate dehydrogenase
G6PDH	Glycose-6-phosphate dehydrogenase
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
MetSynth	Methionine synthesis
MurSynth	Mureine synthesis
PDH	Pyruvate dehydrogenase
PEPCxylase	PEP carboxylase
PFK	Phosphofructokinase
PGDH	6-Phosphogluconate dehydrogenase
PGI	Phosphoglucoisomerase
PGK	Phosphoglycerate kinase
PGM	Phosphoglucomutase
PGluMu	Phosphoglycerate mutase
PK	Pyruvate kinase
PTS	Phosphotransferase system
R5PI	Ribose phosphate isomerase
RPPK	Ribose phosphate pyrophosphokinase
RU5P	Ribulose phosphate epimerase
SerSynth	Serine synthesis
Synth1	Synthesis 1
Synth2	Synthesis 2
TA	Transaldolase
TIS	Triosephosphate isomerase
TKa	Transketolase a
TKb	Transketolase b
TrpSynth	Tryptophan synthesis

Table A2
Forcing functions for co-metabolites concentrations (mM) (Chassagnole et al., 2002).

$$C_{atp} = 4.27 - 4.163\frac{t}{0.657 + 1.43t + 0.364t^2} \quad (A1)$$

$$C_{adp} = 0.582 + 1.73(2.731^{-0.15t})(0.12t + 0.000214t^3) \quad (A2)$$

$$C_{amp} = 0.123 + 7.25\frac{t}{7.25 + 1.47t + 0.17t^2} + 1.073\frac{t}{1.29 + 8.05t} \quad (A3)$$

$$C_{nadph} = 0.062 + 0.332(2.718^{-0.46t})(0.0166t^{1.58} + 0.000166t^{4.73} + 1.16\cdot 10^{-10}t^{7.89} + 1.36\cdot 10^{-13}t^{11} + 1.23\cdot 10^{-16}t^{14.2}) \quad (A4)$$

$$C_{nadp} = 0.159 + 0.00554\frac{t}{2.8 + 0.271t + 0.01t^2} + 0.182\frac{t}{4.81 + 0.526t} \quad (A5)$$

$$C_{nadh} = 0.0934 + 0.0011(2.371^{-0.123t})(0.844t + 0.104t^3) \quad (A6)$$

$$C_{nad} = 1.314 + 1.314(2.73^{(-0.0435t-0.342)}) - \frac{(t + 7.871)(2.73^{(-0.0218t-0.171)})}{8.481 + t} \quad (A7)$$

Table A3
Kinetics rates for enzymes (Chassagnole et al., 2002).

$$r_{PTS} = \frac{r_{PTS}^{max} C_{glc}^{extracelular}(C_{pep}/C_{pyr})}{(K_{PTS1} + K_{PTS2}(C_{pep}/C_{pyr}) + K_{PTS,3}C_{glc}^{extracelular} + C_{glc}^{extracelular}(C_{pep}/C_{pyr}))(1 + (C_{g6p}^{n_{PTS,g6p}}/K_{PTS,g6p}))} \quad (A8)$$

$$r_{PGI} = \frac{r_{PGI}^{max}(C_{g6p} - (C_{f6p}/K_{PGI,eq}))}{K_{PGI,g6p}(1 + (C_{f6p}/(K_{PGI,f6p}(1 + (C_{6pg}/K_{PGI,f6p,6pginh}))))) + (C_{6pg}/K_{PGI,g6p,6pginh})) + C_{g6p}} \quad (A9)$$

$$r_{PFK} = \frac{r_{PFK}^{max}\left[(eC_{f6p}/K_{Rf6p})(1 + (eC_{f6p}/K_{Rf6p}))^{n_{PFK}-1}(1 + (C_{adp}/K_{Radp}))^{n_{PFK}} + L_0[((1 + (C_{pep}/K_{Tpep}))(1 + (C_{adp}/K_{Tadp}))/((1 + (C_{pep}/K_{Rpep}))(1 + (C_{adp}/K_{Radp}))))]^{n_{PFK}}\theta ce'(C_{f6p}/K_{Rf6p})(1 + ce'(C_{f6p}/K_{Rf6p}))^{n_{PFK}-1}\right]}{(1 + (eC_{f6p}/K_{Rf6p}))^{n_{PFK}}(1 + (C_{adp}/K_{Radp}))^{n_{PFK}} + L_0[((1 + (C_{pep}/K_{Tpep}))(1 + (C_{adp}/K_{Tadp})))/((1 + (C_{pep}/K_{Rpep}))(1 + (C_{adp}/K_{Radp})))]^{n_{PFK}}(1 + ce'(C_{f6p}/K_{Rf6p}))^{C_{pep}/K_{Tpep}}} \quad (A10)$$

$$r_{ALDO} = \frac{r_{ALDO}^{max}(C_{fdp} - (C_{gap}C_{dhap}/K_{ALDO,eq}))}{K_{ALDO,fdp} + C_{fdp} + (K_{ALDO,gap}C_{dhap}/K_{ALDO,eq}V_{ALDO,blf}) + (K_{ALDO,dhap}C_{gap}/K_{ALDO,eq}V_{ALDO,blf}) + (C_{fdp}C_{gap}/K_{ALDO,gap,inh}) + (C_{dhap}C_{gap}/K_{ALDO,eq}V_{ALDO,blf})} \quad (A11)$$

$$r_{TIS} = \frac{r_{TIS}^{max}(C_{dhap} - (C_{gap}/K_{TIS,eq}))}{K_{TIS,dhap}(1 + (C_{gap}/K_{TIS,gap})) + C_{dhap}} \quad (A12)$$

$$r_{GAPDH} = \frac{r_{GAPDH}^{max}(C_{gap}C_{nad} - (C_{pgp}C_{nadh}/K_{GAPDH,eq}))}{(K_{GAPDH,gap}(1 + (C_{pgp}/K_{GAPDH,pgp})) + C_{gap})(K_{GAPDH,nad}(1 + (C_{nadh}/K_{GAPDH,nadh})) + C_{nad})} \quad (A13)$$

$$r_{PGK} = \frac{r_{PGK}^{max}(C_{adp}C_{pgp} - (C_{atp}C_{3pg}/K_{PGK,eq}))}{(K_{PGK,adp}(1 + (C_{atp}/K_{PGK,atp})) + C_{adp})(K_{PGK,pgp}(1 + (C_{3pg}/K_{PGK,3pg})) + C_{pgp})} \quad (A14)$$

$$r_{PGluMu} = \frac{r_{PGluMu}^{max}(C_{3pg} - (C_{2pg}/K_{PGluMu,eq}))}{K_{PGluMu,3pg}(1 + (C_{2pg}/K_{PGluMu,2pg}) + C_{3pg}} \quad (A15)$$

$$r_{ENO} = \frac{r_{ENO}^{max}(C_{2pg} - (C_{pep}/K_{ENO,eq}))}{K_{ENO,2pg}(1 + (C_{pep}/K_{ENO,pep})) + C_{2pg}} \quad (A16)$$

$$r_{PK} = \frac{r_{PK}^{max}C_{pep}C_{adp}((C_{pep}/K_{PK,pep}) + 1)^{(n_{PK}-1)}}{K_{PK,pep}(L_{PK}(((1 + (C_{atp}/K_{PK,atp}))/((C_{fdp}/K_{PK,fdp}) + (C_{amp}/K_{PK,amp}) + 1)))^{n_{PK}} + ((C_{pep}/K_{PK,pep}) + 1)^{n_{PK}})(C_{adp} + K_{PK,adp})} \quad (A17)$$

Table A3 (*Continued*)

$$r_{PDH} = \frac{r^{max}_{PDH} C^{n_{PDH}}_{pyr}}{K_{PDH,pyr} + C^{n_{PDH}}_{pyr}} \tag{A18}$$

$$r_{G6PDH} = \frac{r^{max}_{GPDH} C_{g6p}}{(K_{G6PDH,g6p}(1 + k_{tgn}(C^{h_{tg}}_{nadp}/(k^{h_{tg}}_{tg} + C^{h_{tg}}_{nadp}))) + C_{g6p})(1 + (C_{nadph}/K_{G6PDH,nadphinh}) + (C_{nadh}/K_{G6PDH,nadhinh}))((C_{nadp}/K_{G6PDH,nadp})^{1+k_{dtn}(C^{h_{dt}}_{nadh}/(k^{h_{dt}}_{dt} + C^{h_{dt}}_{nadh}))})/ ((1 + (C_{nadp}/K_{G6PDH,nadp})^{1+k_{dtn}(C^{h_{dt}}_{nadh}/(k^{h_{dt}}_{dt} + C^{h_{dt}}_{nadh}))} + (C_{nadph}/K_{G6PDH,nadphinh2})^{h_{nadph}})(1 + (C_{nadh}/K_{G6PDH,nadh})^{h_{nadh}}))} \tag{A19}$$

$$r_{PGDH} = \frac{r^{max}_{PGDH} C_{6pg} C_{nadp}}{(C_{6pg} + K_{PGDH,6pg})(C_{nadp} + K_{PGDH,nadp}(1 + (C_{nadph}/K_{PGDH,nadph,inh}))(1 + (C_{atp}/K_{PGDH,atp,inh})))} \tag{A20}$$

$$r_{R5PI} = r^{max}_{R5PI}\left(C_{ribu5p} - \frac{C_{rib5p}}{K_{R5PI,eq}}\right) \tag{A21}$$

$$r_{RU5P} = r^{max}_{RU5P}\left(C_{ribu5p} - \frac{C_{xyl5p}}{K_{RU5P,eq}}\right) \tag{A22}$$

$$r_{TKa} = r^{max}_{TKa}\left(C_{rib5p}C_{xyl5p} - \frac{C_{sed7p}C_{gap}}{K_{TKa,eq}}\right) \tag{A23}$$

$$r_{TKb} = r^{max}_{TKb}\left(C_{xyl5p}C_{e4p} - \frac{C_{f6p}C_{gap}}{K_{TKb,eq}}\right) \tag{A24}$$

$$r_{TA} = r^{max}_{TA}\left(C_{gap}C_{sed7p} - \frac{C_{e4p}C_{f6p}}{K_{TA,eq}}\right) \tag{A25}$$

$$r_{PEPCxylase} = \frac{r^{max}_{PEPCxylase} C_{pep}(1 + (C_{fdp}/K_{PEPCxylase,fdp})^{n_{PEPCxylase,fdp}})}{K_{PEPCxylase,pep} + C_{pep}} \tag{A26}$$

$$r_{Synth1} = \frac{r^{max}_{Synth1} C_{pep}}{K_{Synth1,pep} + C_{pep}} \tag{A27}$$

$$r_{Synth2} = \frac{r^{max}_{Synth2} C_{pyr}}{K_{Synth2,pyr} + C_{pyr}} \tag{A28}$$

$$r_{SerSynth} = \frac{r^{max}_{SerSynth} C_{3pg}}{K_{SerSynth,3pg} + C_{3pg}} \tag{A29}$$

$$r_{RPPK} = \frac{r^{max}_{RPPK} C_{rib5p}}{K_{RPPK,rib5p} + C_{rib5p}} \tag{A30}$$

$$r_{G3PDH} = \frac{r^{max}_{G3PDH} C_{dhap}}{K_{G3PDH,dhap} + C_{dhap}} \tag{A31}$$

$$r_{DAHPS} = \frac{r^{max}_{DAHPS} C^{n_{DAHPS,e4p}}_{e4p} C^{n_{DAHPS,pep}}_{pep}}{(K_{DAHPS,e4p} + C^{n_{DAHPS,e4p}}_{e4p})(K_{DAHPS,pep} + C^{n_{DAHPS,pep}}_{pep})} \tag{A32}$$

$$r_{PGM} = \frac{r^{max}_{PGM}(C_{g6p} - (C_{g1p}/K_{PGM,eq}))}{K_{PGM,g6p}(1 + (C_{g1p}/K_{PGM,g1p})) + C_{g6p}} \tag{A33}$$

$$r_{G1PAT} = \frac{r^{max}_{G1PAT} C_{g1p} C_{atp}(1 + (C_{fdp}/K_{G1PAT,fdp})^{n_{G1PAT,fdp}})}{(K_{G1PAT,g1p} + C_{g1p})(K_{G1PAT,atp} + C_{atp})} \tag{A34}$$

$$r_{MurSynth} = r^{max}_{MurSynth} \tag{A35}$$

$$r_{TrpSynth} = r^{max}_{TrpSynth} \tag{A36}$$

$$r_{MetSynth} = r^{max}_{MetSynth} \tag{A37}$$

Table A4
Kinetics parameters of the model.

Enzyme	Parameters	Description
Phosphotransferase system (PTS)	K_{PTS1}	M–M half-saturation constant (mM)
	K_{PTS2}	Constant (mM)
	K_{PTS3}	Constant
	$K_{PTS,g6p}$	Inhibition constant (mM)
	$n_{PTS,g6p}$	Constant
	r_{PTS}^{max}	Maximum reaction rate (mM s^{-1})
Phosphoglucoisomerase (PGI)	$K_{PGI,g6p}$	M–M half-saturation constant (mM)
	$K_{PGI,f6p}$	Inhibition constant (mM)
	$K_{PGI,eq}$	Equilibrium Constant
	$K_{PGI,g6p,6pg,inh}$	Inhibition constant (mM)
	$K_{PGI,f6p,6pg,inh}$	Inhibition constant (mM)
	r_{PGI}^{max}	Maximum reaction rate (mM s^{-1})
Phosphofructokinase (PFK)	$K_{PFK,f6p,s}$	M–M half-saturation constant (mM)
	$K_{PFK,atp,s}$	M–M half-saturation constant (mM)
	$K_{PFK,adp,a}$	Activation constant (mM)
	$K_{PFK,adp,b}$	Activation constant (mM)
	$K_{PFK,adp,c}$	Activation constant (mM)
	$K_{PFK,amp,a}$	Activation constant (mM)
	$K_{PFK,amp,b}$	Activation constant (mM)
	$K_{PFK,pep}$	Inhibition constant (mM)
	L_{PFK}	Allosteric constant
	n_{PFK}	Number of binding sites
	r_{PFK}^{max}	Maximum reaction rate (mM s^{-1})
Aldolase (ALDO)	$K_{ALDO,fdp}$	M–M half-saturation constant (mM)
	$K_{ALDO,dhap}$	M–M half-saturation constant (mM)
	$K_{ALDO,gap}$	M–M half-saturation constant (mM)
	$K_{ALDO,gap,inh}$	Inhibition constant (mM)
	$V_{ALDO,blf}$	Back-forward reaction rate relation
	$K_{ALDO,eq}$	Equilibrium constant (mM)
	r_{ALDO}^{max}	Maximum reaction rate (mM s^{-1})
Triosephosphate isomerase (TIS)	$K_{TIS,dhap}$	M–M half-saturation constant (mM)
	$K_{TIS,gap}$	M–M half-saturation constant (mM)
	$K_{TIS,eq}$	Equilibrium constant
	r_{TIS}^{max}	Maximum reaction rate (mM s^{-1})
Glyceraldehyde-3-phosphate dehydrogenase (GAPDH)	$K_{GAPDH,gap}$	M–M half-saturation constant (mM)
	$K_{GAPDH,pgp}$	Inhibition constant (mM)
	$K_{GAPDH,nad}$	M–M half-saturation constant (mM)
	$K_{GAPDH,nadh}$	Inhibition constant (mM)
	$K_{GAPDH,eq}$	Equilibrium constant
	r_{GAPDH}^{max}	Maximum reaction rate (mM s^{-1})
Phosphoglycerate kinase (PGK)	$K_{PGK,pgp}$	M–M half-saturation constant (mM)
	$K_{PGK,3pg}$	Inhibition constant (mM)
	$K_{PGK,adp}$	M–M half-saturation constant (mM)
	$K_{PGK,atp}$	Inhibition constant (mM)
	$K_{PGK,eq}$	Equilibrium constant
	r_{PGK}^{max}	Maximum reaction rate (mM s^{-1})
Phosphoglycerate mutase (PGluMu)	$K_{PGluMu,3pg}$	M–M half-saturation constant (mM)
	$K_{PGluMu,2pg}$	Inhibition constant (mM)
	$K_{PGluMu,eq}$	Equilibrium constant
	r_{PGluMu}^{max}	Maximum reaction rate (mM s^{-1})
Enolase (ENO)	$K_{ENO,2pg}$	M–M half-saturation constant (mM)
	$K_{ENO,pep}$	Inhibition constant (mM)
	$K_{ENO,eq}$	Equilibrium constant
	r_{ENO}^{max}	Maximum reaction rate (mM s^{-1})
Pyruvate kinase (PK)	$K_{PK,pep}$	M–M half-saturation constant (mM)
	$K_{PK,adp}$	M–M half-saturation constant (mM)
	$K_{PK,atp}$	Inhibition constant (mM)
	$K_{PK,fdp}$	Activation constant (mM)
	$K_{PK,amp}$	Activation constant (mM)
	L_{PK}	Allosteric constant
	n_{PK}	Number of binding sites
	r_{PK}^{max}	Maximum reaction rate (mM s^{-1})
Pyruvate dehydrogenase (PDH)	$K_{PDH,pyr}$	M–M half-saturation constant (mM)
	n_{PDH}	Number of binding sites
	r_{PDH}^{max}	Maximum reaction rate (mM s^{-1})
Phoenolpyruvate carboxylase (PEPCxylase)	$K_{PEPCxylase,pep}$	M–M half-saturation constant (Mm)
	$K_{PEPCxylase,fdp}$	Activation constant (mM)
	$n_{PEPCxylase,fdp}$	Number of binding sites
	$r_{PEPCxylase}^{max}$	Maximum reaction rate (mM s^{-1})

Table A4 (*Continued*)

Enzyme	Parameters	Description
Phosphoglucomutase (PGM)	$K_{PGM,g6p}$	M–M half-saturation constant (mM)
	$K_{PGM,g1p}$	Inhibition constant (mM)
	$K_{PGM,eq}$	Equilibrium constant
	r^{max}_{PGM}	Maximum reaction rate (mM s^{-1})
Glucose-1-phosphate adenyltransferase (G1PAT)	$K_{G1PAT,g1p}$	M–M half-saturation constant (mM)
	$K_{G1PAT,atp}$	M–M half-saturation constant (mM)
	$K_{G1PAT,fdp}$	Activation constant (mM)
	$n_{G1PAT,fdp}$	Number of binding sites
	r^{max}_{G1PAT}	Maximum reaction rate (mM s^{-1})
Ribose phosphate pyrophosphokinase (RPPK)	$K_{RPPK,rib5p}$	M–M half-saturation constant (mM)
	r^{max}_{RPPK}	Maximum reaction rate (mM s^{-1})
Glycerol-3-phosphate dehydrogenase (G3PDH)	$K_{G3PDH,dhap}$	M–M half-saturation constant (mM)
	r^{max}_{G3PDH}	Maximum reaction rate (mM s^{-1})
Serine synthesis (SerSynth)	$K_{SerSynth,3pg}$	M–M half-saturation constant (mM)
	$r^{max}_{SerSynth}$	Maximum reaction rate (mM s^{-1})
DAHP synthase (DAHPS)	$K_{DAHPS,e4p}$	M–M half-saturation constant (mM)
	$K_{DAHPS,pep}$	M–M half-saturation constant (mM)
	$n_{DAHPS,e4p}$	Number of binding sites
	$n_{DAHPS,pep}$	Number of binding sites
	r^{max}_{DAHPS}	Maximum reaction rate (mM s^{-1})
Glucose-6-phosphate dehydrogenase (G6PDH)	$K_{G6PDH,g6p}$	M–M half-saturation constant (mM)
	$K_{G6PDH,nadp}$	M–M half-saturation constant (mM)
	$K_{G6PDH,nadph,nadph,inh}$	Inhibition constant (mM)
	$K_{G6PDH,nadph,g6ph,inh}$	Inhibition constant (mM)
	r^{max}_{G6PDH}	Maximum reaction rate (mM s^{-1})
6-Phosphogluconate dehydrogenase (PGDH)	$K_{PGDH,6pg}$	M–M half-saturation constant (mM)
	$K_{PGDH,nadp}$	M–M half-saturation constant (mM)
	$K_{PGDH,nadph,inh}$	Inhibition constant (mM)
	$K_{PGDH,atp,inh}$	Inhibition constant (mM)
	r^{max}_{PGDH}	Maximum reaction rate (mM s^{-1})
Ribulose phosphate epimerase (RU5P)	$K_{RU5P,EQ}$	Equilibrium constant (mM)
	r^{max}_{RU5PI}	Maximum reaction rate (mM s^{-1})
Ribose phosphate isomerase (R5PI)	$K_{R5PI,eq}$	Equilibrium constant (mM)
	r^{max}_{R5PI}	Maximum reaction rate (mM s^{-1})
Transketolase a (TKa)	$K_{TKa,eq}$	Equilibrium constant (mM)
	r^{max}_{TKa}	Maximum reaction rate (mM s^{-1})
Transketolase b (TKb)	$K_{TKb,eq}$	Equilibrium constant (mM)
	r^{max}_{TKb}	Maximum reaction rate (mM s^{-1})
Transaldolase (TA)	$K_{TA,eq}$	Equilibrium constant (mM)
	r^{max}_{TA}	Maximum reaction rate (mM s^{-1})
Synthesis 1	$K_{Synth1,pep}$	M–M half-saturation constant (mM)
	r^{max}_{Synth1}	Maximum reaction rate (mM s^{-1})
Synthesis 2	$K_{Synth2,pyr}$	M–M half-saturation constant (mM)
	r^{max}_{Synth2}	Maximum reaction rate (mM s^{-1})
Mureine synthesis (MurSynth)	$r^{max}_{MurSynth}$	Maximum reaction rate (mM s^{-1})
Tryptophan synthesis (TrpSynth)	$r^{max}_{TrpSynth}$	Maximum reaction rate (mM s^{-1})
Methionine synthesis (MetSynth)	$r^{max}_{MetSynth}$	Maximum reaction rate (mM s^{-1})

References

Botelho, V. (2012). *Nova metodologia para análise de identificabilidade e estimação de parâmetros de modelos fenomenológicos*. Brazil: Universidade Federal do Rio Grande do Sul (Masters thesis, in Portuguese).

Bristol, E. (1966). On a new measure of interaction for multivariable process control. *IEEE Transactions on Automatic Control*, *11*, 133–134.

Brun, R., Reichert, P., & Kunsch, H. R. (2001). Practical identifiability analysis of large environmental simulation models. *Water Resources Research*, *37*, 1015–1030.

Cao, Y., & Rossiter, D. (1997). An input pre-screening technique for control structure selection. *Computers Chemical Engineering*, *21*, 563–569.

Chassagnole, C., Noisommit-Rizzi, N., Schimd, J. W., Mauch, K., & Reuss, M. (2002). Dynamic modeling of the central carbon metabolism of *Escherichia coli*. *Biotechnology and Bioengineering*, *79*, 53–72.

Chu, Y., & Hahn, J. (2007). Parameter set selection for estimation of nonlinear dynamic systems. *AIChE Journal*, *53*, 2858–2870.

Chu, Y., Huang, Z., & Hahn, J. (2011). Global sensitivity analysis procedure accounting for effect of available experimental data. *Industrial & Engineering Chemistry Research*, *50*, 1294–1304.

Chu, Y., Huang, Z., Hahn, J., Chu, Y., & Hahn, J. (2012). Generalization of a parameter set selection procedure based upon orthogonal projections and the D-optimality criterion. *AIChE Journal*, *58*, 2085–2096.

Diaz Ricci, J. C. (1996). Influence of phosphoenolpyruvate on the dynamic behavior of phosphofructokinase of *Escherichia coli*. *Journal of Theoretical Biology*, *178*, 145–150.

Di Maggio, J. A., Ricci, J. D., & Díaz, M. S. (2009). Global sensitivity analysis in dynamic metabolic networks. *Computers Chemical Engineering*, *34*, 770–781.

Di Maggio, J., Diaz Ricci, J. C., & Diaz, M. S. (2010). Parameter estimation in kinetic models for large scale metabolic networks with advanced mathematical programming techniques. *Computer Aided Chemical Engineering*, *28*, 355–360 (ISSN: 1570-7946).

Goldberg, D. E. (1989). *Genetic algorithms in search, optimization & machine learning*. New York: Addison-Wesley.

Golub, G. H., & van Loan, C. F. (1996). *Matrix computations* (3rd ed.). London: Johns Hopkins University Press.

Hiskens, I. A. (2001). Nonlinear dynamic model evaluation from disturbance measurements. *IEEE Transactions on Power Systems*, *16*, 702–710.

Hoque, M. A., Ushiyama, H., Tomita, M., & Shimizu, K. (2005). Dynamic responses of the intracellular metabolite concentrations of the wild type and pykA mutant *Escherichia coli* against pulse addition of glucose or NH_3 under those limiting continuous cultures. *Biochemical Engineering Journal*, *26*, 38–49.

Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. In *Proceedings of the IEEE international conference on neural networks* Perth, Australia.

Kou, B., McAuley, K. B., Hsu, C. C., Bacon, D. W., & Yao, K. Z. (2005). Mathematical model and parameter estimation for gas-phase ethylene/hexene copolymerization with supported metallocene catalyst. *Macromolecular Materials and Engineering*, *290*, 537–557.

Kravaris, C., Chu, Y., & Hahn, J. (2012). Advances and selected recent developments in state and parameter estimation. *Computers and Chemical Engineering*, *51*, 111–123.

Li, R. J., Henson, M. A., & Kurtz, M. J. (2004). Selection of model parameters for off-line parameter estimation. *IEEE Transactions on Control Systems Technology*, *12*, 402–412.

Lund, B. F., & Foss, B. A. (2008). Parameter ranking by orthogonalization—Applied to nonlinear mechanistic models. *Automatica*, *44*, 278–281.

Machado, V. C., Tapia, G., Gabriel, D., Lafuente, J., & Baeza, J. A. (2009). Systematic identifiability study based on the Fisher information matrix for reducing the number of parameters calibration of an activated sludge model. *Environmental Modelling & Software*, *24*, 1274–1284.

Omlin, M., & Reichert, P. (1999). A comparison of techniques for the estimation of model prediction uncertainty. *Ecological Modelling*, *115*(1), 45–59.

Petzold, L. R. (1989). *DASSL: A differential/algebraic system solver* (1st ed., pp.). Computer and Mathematical Research Division, Lawrence Livermore National Laboratory.

Ratushny, A., Smirnova, O., Usuda, Y., & Matsui, K. (2006). Regulation of the pentose phosphate pathway in *Escherichia coli*: Gene network reconstruction and mathematical modeling of metabolic reactions. In *Proceedings of the fifth international conference on bioinformatics of genome regulation and structure (BGRS)* Novosibirsk, Russia.

Sandink, C. A., McAuley, K. B., & McLellan, P. J. (2001). Selection of parameters for updating in on-line models. *Industrial & Engineering Chemistry Research*, *40*, 3936–3950.

Secchi, A. R., Cardozo, N. S. M., Almeida Neto, E., & Finkler, T. F. (2006). An algorithm for automatic selection and estimation of model parameters. In *Proceedings of the international symposium on advanced control of chemical processes (ADCHEM2006)* Gramado, Brazil.

Sun, C. L., & Hahn, J. (2006). Parameter reduction for stable dynamical systems based on Hankel singular values and sensitivity analysis. *Chemical Engineering Science*, *61*, 5393–5403.

Schwaab, M., Alberton, A. L., & Pinto, J. C. (2011). ESTIMA&PLANEJA: Pacote computacional para estimação de parâmetros de planejamento de experimentos. In *Internal Technical Report*. Brazil: Universidade Federal do Rio de Janeiro.

Thompson, D. E., McAuley, K. B., & McLellan, P. J. (2009). Parameter estimation in a simplified MWD model for HDPE produced by a Ziegler–Natta catalyst. *Macromolecular Reaction Engineering*, *3*, 160–177.

Velez-Reyes, M., & Verghese, G. C. (1995). Subset selection in identification, and application to speed and parameter estimation for induction machines. In *Proceedings of the 4th IEEE Conference on control applications* New York, EUA.

Weijers, S. R., & Vanrolleghem, P. A. (1997). A procedure for selecting best identifiable parameters in calibrating activated sludge model no 1 to full-scale plant data. *Water Science and Technology*, *36*, 69–79.

Yao, K. Z., Shaw, B. M., Kou, B., McAuley, K. B., & Bacon, D. W. (2003). Modeling ethylene/butene copolymerization with multi-site catalysts: Parameter estimability and experimental design. *Polymer Reaction Engineering*, *11*, 563–588.

Yue, H., Brown, M., Knowles, J., Wang, H., Broomhead, D. S., & Kellab, D. B. (2006). Insights into the behaviour of systems biology models from dynamic sensitivity and identifiability analysis: A case study of an NF-kB signaling pathway. *Molecular BioSystem*, *2*, 640–649.