BMC Bioinformatics

Methodology article

Novel metaheuristic for parameter estimation in nonlinear dynamic biological systems

Maria Rodriguez-Fernandez, Jose A Egea and Julio R Banga*

Address: Process Engineering Group, Instituto de Investigaciones Marinas (C.S.I.C.), Spanish Council for Scientific Research, C/Eduardo Cabello, 6. 36208 Vigo, Spain

Email: Maria Rodriguez-Fernandez - mrodriguez@iim.csic.es; Jose A Egea - jegea@iim.csic.es; Julio R Banga* - julio@iim.csic.es

* Corresponding author

Published: 02 November 2006

BMC Bioinformatics 2006, **7**:483 doi:10.1186/1471-2105-7-483

Received: 29 May 2006
Accepted: 02 November 2006

This article is available from: http://www.biomedcentral.com/1471-2105/7/483

Abstract

Background: We consider the problem of parameter estimation (model calibration) in nonlinear dynamic models of biological systems. Due to the frequent ill-conditioning and multi-modality of many of these problems, traditional local methods usually fail (unless initialized with very good guesses of the parameter vector). In order to surmount these difficulties, global optimization (GO) methods have been suggested as robust alternatives. Currently, deterministic GO methods can not solve problems of realistic size within this class in reasonable computation times. In contrast, certain types of stochastic GO methods have shown promising results, although the computational cost remains large. Rodriguez-Fernandez and coworkers have presented hybrid stochastic-deterministic GO methods which could reduce computation time by one order of magnitude while guaranteeing robustness. Our goal here was to further reduce the computational effort without loosing robustness.

Results: We have developed a new procedure based on the scatter search methodology for nonlinear optimization of dynamic models of arbitrary (or even unknown) structure (i.e. black-box models). In this contribution, we describe and apply this novel metaheuristic, inspired by recent developments in the field of operations research, to a set of complex identification problems and we make a critical comparison with respect to the previous (above mentioned) successful methods.

Conclusion: Robust and efficient methods for parameter estimation are of key importance in systems biology and related areas. The new metaheuristic presented in this paper aims to ensure the proper solution of these problems by adopting a global optimization approach, while keeping the computational effort under reasonable values. This new metaheuristic was applied to a set of three challenging parameter estimation problems of nonlinear dynamic biological systems, outperforming very significantly all the methods previously used for these benchmark problems.

Background

Modelling approaches are central in systems biology and provide new ways towards the analysis of omics data, ultimately leading to a greater understanding of the language of cells and organisms [1-3]. Further, these approaches provide systematic strategies for key issues in medicine [4] and the pharmaceutical and biotechnological industries. For example, model-based approaches can provide a rational framework to guide drug development, taking into account the effects of possible new drugs on bio-

chemical pathways and physiology [5]. A common approach to model inter- and intra-cellular dynamic processes is by means of dynamic models, usually consisting of sets of differential equations [6].

The general area of system identification deals with the development of mathematical models of dynamic systems from input/output data [7,8]. When building mathematical models, one starts from the definition of the purpose of the model and uses the a priori available knowledge (i.e. physical, chemical or biological laws, initial hypothesis and/or preliminary data) to choose a model framework and to propose a model structure. This model contain parameters and we need to know whether is it possible to uniquely determine their values (identifiability analysis) and if so, to estimate them with maximum precision and accuracy. This leads to a first working model that must be validated with new experiments, revealing in most cases a number of deficiencies. In this case, a new model structure and/or a new experimental design must be planned, and the process is repeated iteratively until the validation step is considered satisfactory. This iterative process (i.e. the model building cycle) should also contain other elements like optimal experimental design and model discrimination steps [9-13].

This work is focused on the key step of parameter estimation, assuming the structure of the nonlinear dynamic model as given. Parameter estimation (also known as model calibration) aims to find the parameters of the model which give the best fit to a set of experimental data. Examples of recent efforts in the particular case of biochemical pathways are the works of Sugimoto and coworkers [14], Voit and Almeida [15], Rodriguez-Fernandez and coworkers [13] and Polisetty and coworkers [16]. The key issues considered here in this work were to ensure reliable and accurate parameter estimation, paying especial attention to the computational cost, and also to perform the identifiability analysis of the models.

Parameter estimation in nonlinear dynamic biological models

Estimating the parameters of a nonlinear dynamic model is more difficult than for the linear case, as no general analytic result exists. Biological models are often dynamic and highly nonlinear, thus, in order to find the estimates, we must resort to nonlinear optimization techniques where a measure of the distance between model predictions and experimental data ($Z = \tilde{Y} - Y$) is used as the optimality criterion to be minimized. The criterion selection will depend on the assumptions about the data disturbance and on the amount of information provided by the user. As explained in detail in the *Methods* section, the maximum likelihood estimator maximizes the probability of the occurrence of the observed measurements. If we make the assumption that the residuals are normally distributed and independent with the same variance σ^2, then the maximum likelihood criterion is equivalent to the least squares and we aim to find $\hat{p}$ which minimizes the sum of squared residuals of all the responses. That is, the estimation criterion would be to minimize the *trace* of $Z^T Z$ [17]. This is subject to the dynamics of the system, plus possibly other algebraic constraints, and model parameters are also subject to upper and lower bounds. This formulation is that of a non-linear programming problem (NLP) with differential-algebraic (DAEs) constraints. In this work, we have followed the so-called single shooting approach [18], where an initial value problem (IVP, i.e., the systems dynamics) is solved as an inner problem of the master NLP problem. When estimating parameters of dynamical systems a number of difficulties may arise, like e.g. convergence to local solutions if standard local methods are used, very flat objective function in the neighborhood of the solution, over-determined models, badly scaled model functions or non-differentiable terms in the systems dynamics [18].

Due to the nonlinear and constrained nature of the systems dynamics, these problems are very often multimodal (non-convex). Thus, traditional gradient based methods, like Levenberg-Marquardt or Gauss-Newton, may fail to identify the global solution and may converge to a local minimum when a better solution exists just a small distance away. Moreover, in the presence of a bad fit, there is no way of knowing if it is due to a wrong model formulation, or if it is simply a consequence of local convergence. Thus, there is a distinct need for using global optimization methods which provide more guarantees of converging to the globally optimal solution, as shown in [19]. The importance of using global optimization methods for parameter estimation in systems biology has been increasingly recognized in recent years [16,20-23]. Global optimization methods can be roughly classified as deterministic, stochastic and hybrid strategies. Deterministic methods can guarantee, under some conditions and for certain problems, the location of the global optimum solution. Nevertheless, no deterministic algorithm can solve global optimization problems of the class considered here with certainty in finite time. Actually, computational effort increases very rapidly (often exponentially) with the problem size. Although very significant advances have been recently made [24-26], these methods have a number of requirements about the dynamics of the system, and currently they do not seem to be applicable to problems with a relatively large number of parameters. Stochastic methods are based on probabilistic algorithms,

and they rely on statistical arguments to prove their convergence in a weak way. However, many stochastic methods can locate the vicinity of global solutions in modest computational times [27]. Additionally, stochastic methods do not require transformation of the original problem, which can be treated as a black-box.

In our group, and during the last decade, we have compared a number of different stochastic and deterministic global optimization methods. The overall conclusions from these studies indicate that modern evolution strategies have several key advantages over genetic algorithms and simulated annealing, namely better efficiency/robustness ratio, good scaling properties, an inherent parallel nature and an almost self-tuning mechanism for the search parameters of the method. Our tests and comparisons indicate that DE [28] and SRES [29] are one of the most competitive algorithms, with the additional advantage of being able to handle arbitrary constraints if needed. The main problem presented by these methods is that they require too many evaluations of the objective function [19]. In order to surmount this difficulty, we have recently proposed a hybrid method [13] that speeds up these methodologies while retaining their robustness.

The key concept behind hybrid methods is the well known idea of synergy, that is, a mutually advantageous conjunction of distinct elements. There are several non-trivial questions associated with the actual development of such method, namely choosing which methods to combine, and how to structure such combination. Our work is then focused on selecting more efficient stochastic GO methods and designing better hybrid methods in order to improve the ratio efficiency/robustness. Rodriguez-Fernandez and coworkers [13] combined a global and a local optimization method in a sequential, two-phase hybrid method in order to speedup the stochastic global optimization methods while retaining their robustness. However, computational times were still rather significant, especially if one considers its possible application to larger scale problems.

In order to further increase computational efficiency, in the present work we present a novel metaheuristic approach based on extensions of scatter search combined with various local methods. As it will be shown below, this metaheuristic shows speeds up of between one and two orders of magnitude with respect to previous results obtained with the above mentioned methods. Moreover, this method eliminates the delicate task of deciding where to set the switching point from the global to the local method.

Global optimization: novel metaheuristic

Scatter search (SS) is a population-based method based on formulations originally proposed in the 1960s for combining decision rules and problem constraints, such as the surrogate constraint method. It was first introduced by Glover [30] as a heuristic for integer programming, although it has been extended for other problem classes more recently [31,32]. Scatter search orients its explorations systematically relative to a set of reference points that typically consist of good solutions obtained by prior problem solving efforts.

The justification for choosing this algorithm as the starting framework for our metaheuristic was based on a recent review comparing a number of global optimization solvers over a set of over 1000 constrained GO problems [33], in which a solver based on scatter search proved to be the best among all the stochastic solvers, and the best among all methods for black-box problems. Furthermore, for problems with a large number of decision variables, this solver also proved to be the most reliable.

Scatter search, when the local search is activated, can be defined as a hybrid method since it combines a global search with an intensification phase (i.e. local search). The algorithm uses different heuristics to efficiently choose suitable initial points for the local search, based on merit and distance filters as well as a memory term. This feature helps overcome the problem of switching from global to local search as explained in [13]. The user does not have to worry about stopping the global search and starting the local solver since the algorithm performs this work automatically.

A scatter search framework in a five-step template is given by Laguna and Martí [31] to describe the basic steps of the algorithm (see Figure 1):

• A *diversification generation method* to generate a collection of diverse trial solutions.

• An *improvement method* to transform a trial solution into one or more enhanced trial solutions.

• A *reference set update method* to build and maintain a reference set consisting of the *b* "best" solutions found (where the value of *b* is typically small, e.g., no more than 20), organized to provide efficient accessing by other parts of the method. Solutions gain membership to the reference set according to their quality or their diversity.

• A *subset generation method* to operate on the reference set, to produce several subsets of its solutions as a basis for creating combined solutions.

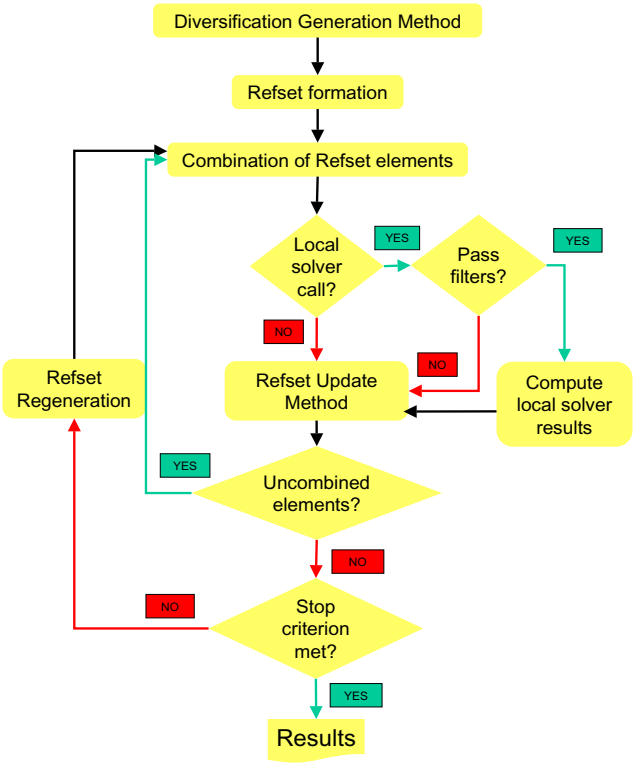

Figure 1
Scatter search pseudo-code diagram.

• A *solution combination method* to transform a given subset of solutions produced by the *subset generation method* into one or more combined solution vectors.

Of the five methods in the SS methodology, only four are strictly required. The *improvement method* is usually needed if high quality outcomes are desired, but a scatter search procedure can be implemented without it. Differences among scatter search implementations are based in the level of sophistication in which these steps are implemented, not in the presence or absence of other steps. In the algorithm presented here, we have added some advanced features in order to improve its performance when solving parameter estimation problems:

• A logarithmic distribution for generating initial trial solutions can be chosen by the user to favor their presence close to the bounds in terms of Euclidean distance, since the location of the global optimum near or even touching the bounds (i.e. being active at any of the bounds) is quite usual in parameter estimation problems.

• Mechanisms to avoid flat zones (also frequent in parameter estimation problems), as well as others to avoid getting stuck in local solutions, have been added. In every iteration the algorithm analyzes if the elite solutions have very similar objective function values regardless their Euclidean distances. If the variance of these values is too low, our procedure considers that the search is located in a flat zone and resets the elite solutions to explore different (and scattered) areas. This mechanism also prevents the algorithm getting stuck in a local solution when the elite solutions have converged to that minimum.

• A new solution combination method allows to explore deeper the search space. Apart from the traditional method of linear combination between solutions, another method based on building hypercubes around the solutions to generate new solutions inside them has been implemented. Now new points around elite solutions (and not only on the segments joining these elite solutions) can be generated, which favors the intensification and accelerates the convergence.

• When all the combinations among elite solutions have been done, the algorithm can stop or continue by partially rebuilding the set of the elite solutions. A new strategy for rebuilding this set, based in orthogonal search directions has been implemented. Instead of simply maximizing the Euclidean distances between the new elite solutions to be generated and the remaining ones, the algorithm takes into account the directions generated by every pair of elite solutions and force the generator to build new solutions that create new search directions.

• The user can choose a number of different local solvers such us SQP methods like fmincon (The MathWorks Inc.), SOLNP [34], SNOPT [35], direct methods like Nomad [36] for the cases of very noisy data, and others specifically designed for parameter estimation problems such as N2FB/DN2FB [37].

It is interesting to observe similarities and contrasts between scatter search and the original genetic algorithm proposals. Both are instances of what are sometimes called "population based" or "evolutionary" approaches. Both incorporate the idea that a key aspect of producing new elements is to generate some form of combination of existing elements. However, genetic algorithm approaches are predicated on the idea of choosing parents randomly to produce offspring, and further on introducing randomization to determine which components of the parents should be combined. By contrast, the scatter search approach does not emphasize randomization, particularly in the sense of being indifferent to choices among alternatives. Instead, the approach is designed to incorporate strategic responses, both deterministic and probabil-

istic, that take account of evaluations and history. Scatter search focuses on generating relevant outcomes without losing the ability to produce diverse solutions, due to the way the generation process is implemented.

A detailed description of the method is given in the *Methodology* section.

Confidence intervals

Determining the parameter values with the maximum likelihood of being correct is only part of the parameter estimation problem. Moreover, it is equally important to find a realistic measure of the precision of those parameters [38,39].

It should be noted that, unlike for the linear case for which a neat theory already exists, there is no exact theory for the evaluation of confidence intervals for systems which are nonlinear in the parameters. An approximate method based on a local linearization of the output function is often used and was also adopted in this study, thus the confidence region is evaluated as a function of the parameter covariance matrix *C*, based on the Fisher information matrix (see details in the *Methods* section).

However, the confidence intervals obtained with the Fisher method are statistically optimistic due to the use of a linear approximation of the non-linear model in the neighborhood of the best parameter estimates [40].

Alternatively, more robust techniques such as the *jackknife* and *bootstrap* methods produce parameter variances that are more realistic. As a drawback, one should mention that these methods are very computing intensive. Another way to obtain the true confidence region of the parameters in non-linear models is by a systematic exploration of the objective functional for an extensive number of parameter combinations. This is a computing intensive task as well, because the number of evaluations increases as a power function of the number of parameters. Therefore, in this study we will make use of the method based on the FIM.

Precision of parameter estimates

Many difficulties found during parameter estimation are due to a poor identifiability of the model parameters. Parameter identifiability tests should be performed prior to the estimation process to ensure that the parameter estimation problem is well-posed [11]. The identifiability analysis investigates if the unknown parameters of the postulated model can be estimated in a unique way.

Regarding this problem, we can distinguish between structural and practical (or a posteriori) identifiability [41]. Structural identifiability is a theoretical property of the model structure depending only on the observation function and the input function. The parameters of a model are structurally globally identifiable if, under ideal conditions of noise-free observations and error-free model structure, and independently of the particular values of the parameters, they can be uniquely estimated from the designed experiment [8].

The requirements for global structural identifiability are rather strict, since we can find realistic situations where the parameters are not identifiable according to this definition, but nevertheless they would be identifiable for a reasonably restricted set of all possible parameters. This leads to the definition of local structural identifiability, where the requirement for the parameters is to be identifiable in a ε neighborhood of a parameter set. Although necessary, structural identifiability is obviously not sufficient to guarantee successful parameter estimation from real data, and this is when the concept of practical identifiability comes into play. In contrast to the theoretical properties of structural identifiability, the practical identifiability is limited by the finite amount of data and the observational noise. Hence, in the presence of large observation errors and/or few data, no reliable estimate is possible and these parameters are called practical non-identifiable.

Assessing a priori global identifiability is very difficult for nonlinear dynamic models, although techniques based on differential algebra have shown very promising results [42]. However, it has been argued that these techniques have somewhat limited applicability [43,44]. These limitations, taken in conjunction with the need for practical methods, provides a key argument for emphasizing the use of practical identifiability despite its limitations derived from its local nature. The question addressed in the a posteriori or practical identifiability analysis is the following: with the available experimental data, can the parameters be uniquely estimated? Or, in other words, if a small deviation of the parameter set occurs, does this have a great impact on the quality of the fit?

The output sensitivity functions (partial derivatives of the measured states with respect to the parameters), are central to the evaluation of practical identifiability. If the sensitivity functions are linearly dependent the model is not identifiable, and sensitivity functions that are nearly linearly dependent, result in parameter estimates that are highly correlated. An easy way to study the practical identifiability of a simple model is to plot the sensitivity functions calculated for a given parameter set. However, this straightforward procedure becomes intractable when the number of measured states and parameters is of realistic size. In the *Methods* section, a numerical procedure to test practical identifiability based on the Fisher information

matrix (FIM), as well as an approximate computation of the correlation matrix, are described.

The correlation matrix measures the interrelationship between the parameters and gives an idea of the compensation effects of changes in the parameter values on the model output. If two parameters are highly correlated, a change in the model output caused by a change in a model parameter can be (nearly) compensated by an appropriate change in the other parameter value. This prevents the parameters from being uniquely identifiable even if the model output is very sensitive to changes in the individual parameters.

In order to perform the practical identifiability analysis, prior knowledge of the model parameters is required. In an experimental situation, the parameters values will not be known a priori, and the identifiability analysis will be an important step in an iterative process involving experimental design and parameter estimation [45].

In this work, the new global optimization metaheuristic described above has been coupled with a computational procedure to check identifiability and related measures. This has resulted in an integrated environment to perform robust parameter estimation and identifiability analysis.

Results and discussion

In order to evaluate the performance and reliability of the novel metaheuristic presented here, which we will denote SSm (scatter search method), we have considered three challenging benchmark problems of increasing order of complexity. All the computations were carried out using a PC/Pentium 4 (1.80 GHz).

Isomerization of α-pinene

In this case study, we want to estimate 5 rate constants (p_1,...,p_5) of a complex biochemical reaction originally studied by Box and coworkers [46], which is also part of COPS (Collection of large-scale Constrained Optimization ProblemS) maintained by Dolan and coworkers [47]. Figure 2 contains the proposed reaction scheme for this homogeneous chemical reaction describing the thermal isomerization of α-pinene (y_1) to dipentene (y_2) and allo-ocimen (y_3) which in turn yields α- and β-pyronene (y_4) and a dimer (y_5). This process was studied by Fuguitt and Hawkins [48], who reported the concentrations of the reactant and the four products at eight time intervals (z). If the chemical reaction orders are known, then mathematical models can be derived giving the concentration of the various species as a function of time. Hunter and MacGregor [49] assumed first-order kinetics and derived the following linear equations for the five responses:

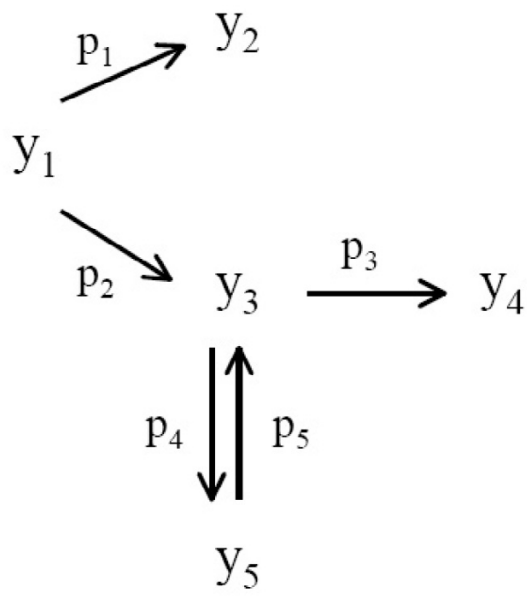

Figure 2
Mechanism for thermal isomerization of α-pinene. Reaction scheme for the thermal isomerization of a-pinene (y_1) to dipentene (y_2) and allo-ocimen (y_3) which in turn yields α- and β-pyronene (y_4) and a dimer (y_5).

$$\frac{dy_1}{dt} = -(p_1 + p_2)y_1 \qquad (1)$$

$$\frac{dy_2}{dt} = p_1 y_1 \qquad (2)$$

$$\frac{dy_3}{dt} = p_2 y_1 - (p_3 + p_4)y_3 + p_5 y_5 \qquad (3)$$

$$\frac{dy_4}{dt} = p_3 y_3 \qquad (4)$$

$$\frac{dy_5}{dt} = -p_4 y_3 + p_5 y_5 \qquad (5)$$

Assuming this model to be appropriate, the initial conditions for the five species are known, and we can estimate the unknown coefficients p_1,...,p_5 by minimization of a cost function which is usually a weighted distance measure between the experimental values corresponding to the measured variables and the predicted values for those variables. For this problem the cost function can be formulated as:

$$J(p) = \sum_{j=1}^{5}\sum_{i=1}^{8}(y_j(p,t_i) - \tilde{y}_{ji})^2 \qquad (7)$$

Box and coworkers [46] tried, in a first instance, to solve this problem without analyzing the multiresponse data, finding parameter values which provided an unsatisfactory data fit. Since ignoring possible dependencies among the responses can lead to difficulties when estimating the parameters (e.g. multiple local minima, very flat objective function, etc.), Box and coworkers described a method for detecting and handling these linear relationships. They showed that there are dependencies in the data and thus only three independent linear combinations of the five responses are used in the identification improving significantly the fit of the data. This analysis of multiresponse data, although efficient, requires a considerable effort specially to uncover the dependencies causes once they have been found, and a deep understanding of the model (that can no longer be considered as a black-box) is essential. Moreover, it becomes unaffordable when the model complexity increases.

Tjoa and Biegler [50] also considered this problem and used a robust local estimation approach to estimate the unknown parameters. They considered the entire data set in order to asses the performance of this method with dependencies in the data, finally reaching the same optimal parameters reported by Box et al. However, the initial value considered for the parameters was very close to the truly optimal solution, which explains why this local method reached the global optimum without getting trapped in a local solution. As pointed out by Averick and coworkers [51], the solution of this problem is not difficult to obtain from initial values of p which are close to the global solution, but becomes increasingly difficult to solve from more remote starting points.

In order to avoid the convergence to local solutions without a good initialization value for the parameters and/or further analysis of the multiresponse data, the use of a global optimization approach is proposed here. The lower bounds considered for the five parameters arise from physical considerations, $p_i \geq 0$, and we took the upper bounds to be $p_i \leq 1$, very far from the best known solution (p_1 = 5.93e - 5, p_2 = 2.96e - 5, p_3 = 2.05e - 5, p_4 = 27.5e - 5, p_5 = 4.00e - 5). As initial point, we chose p_i = 0.5. It should be noted that all the local solvers that we tried with this initial point failed to converge, or converged to bad local solutions.

Figure 3 (value of cost function versus computation time, the latter in log scale) clearly shows that SSm always converged to the global solution after a short computational time, while two other highly reputed global optimization methods (SRES and DE) failed, or converged in a much larger computational time. In order to help the visualization, the convergence curve corresponding to SSm is represented in a different subplot (with log-log scales), since SRES and DE got trapped in local solutions close to the initial point while SSm converged to the global optimum without difficulties.

Figure 4 shows a comparison between the model predicted values and the experimental data reported by Fuguitt and Hawkins [48] corresponding to the concentration of the reactant and the four products. The estimated parameters allow to reproduce almost exactly the experimental data. Furthermore, the homoscedasticity assumption is confirmed by the lack of correlation between the residuals and time (see Figure 5).

The confidence intervals obtained for the optimal parameters, presented in Table 1, are small, indicating a precise estimation. Moreover, the color plot of the correlation matrix in Figure 6 shows a good identifiability at the optimal value with a maximum correlation coefficient of 0.82 between parameters p_4 and p_5. This fact leads us to think that the existence of multiple local minima is the cause of the identification problems experienced in most of the previous studies. These difficulties can be surmounted by proper global optimization methods, as shown here.

Inhibition of HIV proteinase

This problem consists of the estimation of a number of rate constants of a model for the reaction mechanism of irreversible inhibition of HIV proteinase as originally

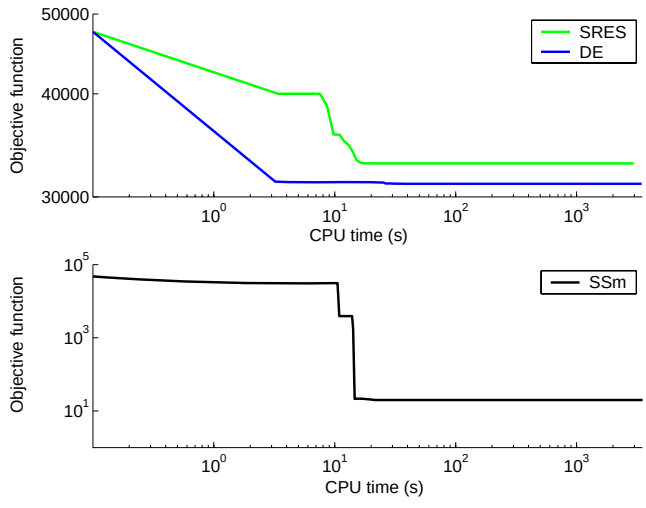

Figure 3
Convergence curves for the alpha-pinene case study. Value of cost function versus computation time for SSm, SRES and DE.

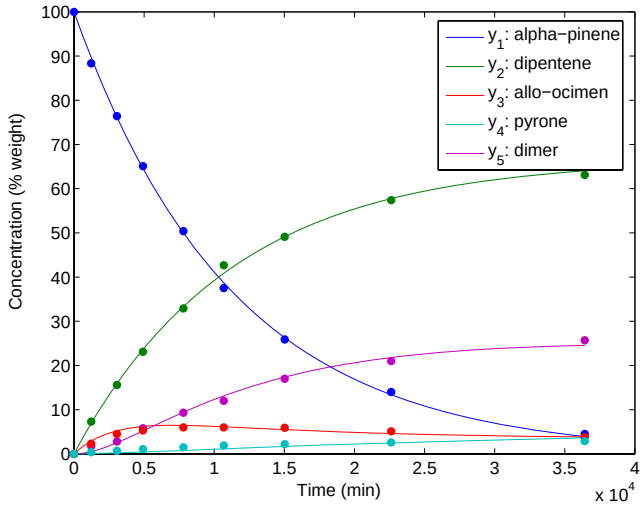

Figure 4
Experimental data versus model prediction for the alpha-pinene case study.

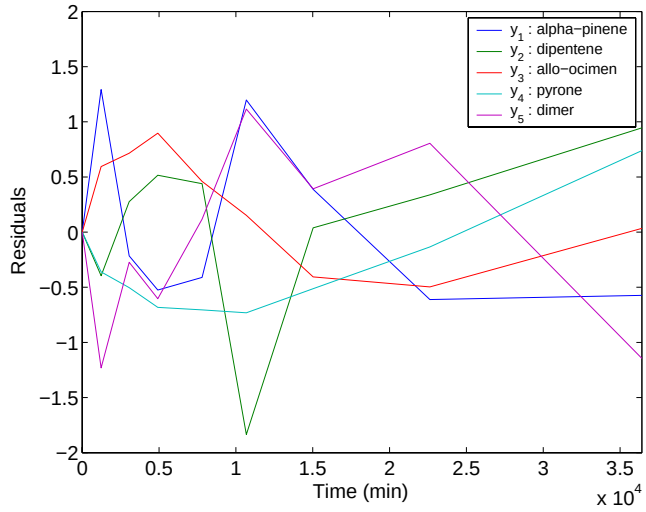

Figure 5
Residuals for the alpha-pinene case study.

studied by Kuzmic [52] (Figure 7). The problem considers an experiment where HIV proteinase (assay concentration 0.004 μM) was added to a solution of an irreversible inhibitor and a fluorogenic substrate (25 μM). The fluorescence changes were monitored for 1 h in each of the five experiments conducted at four different concentrations of the inhibitor (0, 0.0015, 0.003, and 0.004 μM in replicate).

We considered the same problem solved by Kuzmic [52] and Mendes and Kell [53] fitting five of the rate constants to the experimental data. In this fit, a certain degree of uncertainty (± 50 %) in the value of the initial concentrations of substrate and enzyme (titration errors) was also assumed. In addition, the offset (baseline) of the fluorimeter was also considered as a degree of freedom. Given that there are five time course curves, there are a total of 20 adjustable parameters: the five rate constants, five initial concentrations of enzyme, five initial concentrations of substrate and five offset values.

Table 1: Optimal parameters for the alpha pinene isomerization problem

Optimal parameters (J = 19.87)	
Parameter	Optimal value
P_1	5.9259e-5 ± 1.4391e-6
P_2	2.9634e-5 ± 1.3039e-6
P_3	2.0473e-5 ± 6.6657e-6
P_4	2.7449e-4 ± 5.5314e-5
P_5	3.9980e-5 ± 1.9514e-5

By minimization of the sum-of-squares function of the residuals between the measured and the simulated data, the best known solution was obtained by Mendes and Kell using simulated annealing, with a computational cost of 3 million simulations. The next best solution was obtained using a Levenberg-Marquardt method in a considerable shorter computational time (4000 simulations) although this method is only guaranteed to converge to the global minimum if started in its vicinity.

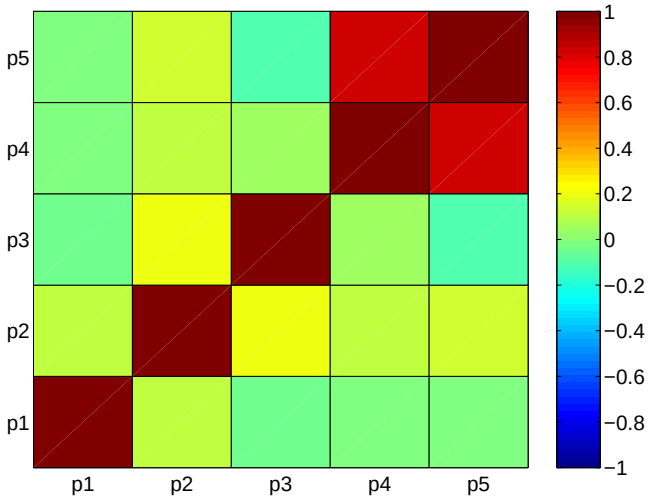

Figure 6
Correlation matrix for the alpha-pinene case study.

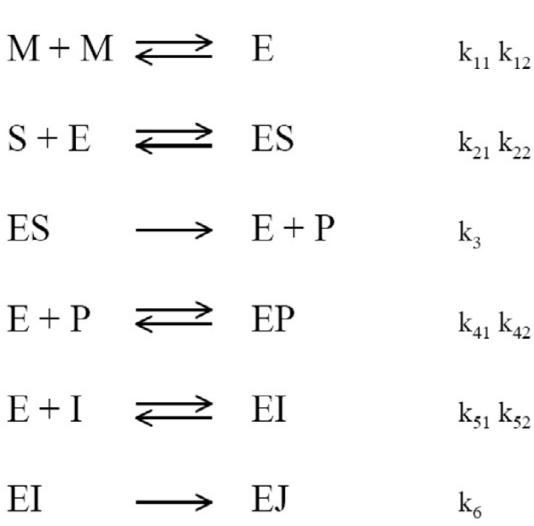

Figure 7
Mechanism of irreversible inhibition of HIV proteinase. The HIV proteinase (E) was added to a solution of an irreversible inhibitor (I) and a fluorogenic substrate (S). The enzyme is only active in a dimer form, the product is a competitive inhibitor for the substrate and the inhibitor is irreversible.

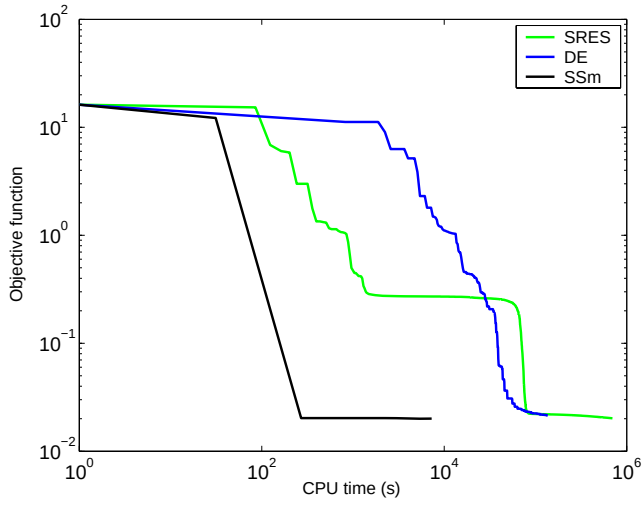

Figure 8
Convergence curves for the inhibition of the HIV proteinase. Value of cost function versus computation time (in log scale) for SSm, SRES and DE.

In our study, SSm converged to a better solution in less that 1500 simulations, which confirms the good performance of this method even with challenging parameter estimation problems. Moreover, when compared with other performant stochastic methods such as SRES or DE, SSm reached better solutions with speed-ups of almost 3 orders of magnitude (see Figure 8).

Despite SSm converged in every run to solutions with a very good values of the cost function (always lower than the best value previously published), the values of the parameters were not always the same (see examples in Table 2) indicating a very flat objective function in the region of parameter space near the optimum. The correlation matrix (see Figure 9) helps to explain this fact since there are correlation values of 0.9999 between some pairs of parameters, (like k_{42} and k_{22}) indicating the lack of identifiability for this problem. This characteristic is first reported here and explains the difficulties (i.e. multiple solutions almost equivalent) experienced by previous researches, confirming the importance of coupling identifiability tests with parameter estimation procedures.

However, it is worth noting the very good correlation between the experimental and predicted data for the best decision vector and the lack of correlation between the residuals and time (see Figure 10 and Figure 11).

Three-step biochemical pathway

This case study, considered as a challenging becnhmark problem by Moles and coworkers [19], involves a biochemical pathway with three enzymatic steps, including the enzymes and mRNAs explicitly. Figure 12 contains a diagram illustrating the network of reactions and kinetics effects (feedback loops).

The identification problem consists of the estimation of 36 kinetic parameters of the nonlinear biochemical dynamic model (8 nonlinear ODEs) which describes the variation of the metabolite concentration with time. Moles and coworkers tried to solve this problem using several deterministic and stochastic global optimization algorithms. They found that only a certain type of stochastic algorithms, evolution strategies (implemented as the SRES code), was able to successfully solve it, although at a rather large computational cost. In Figure 13 we can see how the two-phase hybrid method recently presented by Rodriguez-Fernandez and coworkers [13] converged to better solutions, with speeds up of more than one order of magnitude with respect to the previous results.

The novel metaheuristic, SSm, presented in this work was able to improve this result in an additional order of magnitude regarding the computational time. Moreover, SSm had the additional advantage of not requiring preliminary runs, or any user inputs, for tuning the method, making it

Table 2: Optimal parameters for the HIV proteinase inhibition problem

Results SSm		
Parameter	Parameter value (J = 0.0199)	Parameter value (J = 0.0203)
k_3	6.235 ± 3.2546	5.656 ± 1.953
k_{42}	8772 ± 46120	688.4 ± 3436
k_{22}	473.0 ± 624.6	120.6 ± 508.1
k_{52}	0.09726 ± 0.1288	4.615 ± 583.4
k_6	0.01417 ± 0.01032	3.531 ± 455.4
S_0 (exp 1)	24.63 ± 0.07817	24.69 ± 0.08049
S_0 (exp 2)	23.32 ± 1.349	23.43 ± 0.1541
S_0 (exp 3)	26.93 ± 1.222	27.11 ± 0.1672
S_0 (exp 4)	13.34 ± 1.822	17.07 ± 1.986
S_0 (exp 5)	12.50 ± 1.812	14.49 ± 1.757
E_0 (exp 1)	0.005516 ± 0.001968	0.005397 ± 0.0009091
E_0 (exp 2)	0.005321 ± 0.001309	0.005199 ± 0.0005520
E_0 (exp 3)	0.006000 ± 0.001111	0.006000 ± 0.0005489
E_0 (exp 4)	0.004391 ± 0.00008686	0.004264 ± 0.00005821
E_0 (exp 5)	0.003981 ± 0.00008844	0.003973 ± 0.00005648
offset (exp 1)	-0.004339 ± 0.001788	-0.005611 ± 0.001836
offset (exp 2)	-0.001577 ± 0.002966	-0.004247 ± 0.003432
offset (exp 3)	-0.01117 ± 0.002734	-0.01522 ± 0.003865
offset (exp 4)	-0.001661 ± 0.001881	-0.009649 ± 0.003277
offset (exp 5)	0.007133 ± 0.001764	0.001329 ± 0.003178

a very easy to use strategy. In short, using SSm we have reduced the computation time from two days [19] to a couple of minutes, while ensuring robustness.

Figure 14 shows a comparison (between the predicted and experimental data) for one of the experiments evidencing the accuracy of the fit. Figure 15 confirms that the residuals are white. The representation of the dynamic behavior for the other experiments is quite similar and is not included here for the sake of brevity.

It is sometimes argued that a multistart local method can solve almost all global optimization problems. This can be false for even small problems [54]. The histogram in Figure 16 represents the frequency of the solutions for a multistart of 100 runs using N2FB. The global optimum is in this region close to zero but we can see that it was never reached while a very large number of solutions are far from the global optimum. Despite the identifiability difficulties of this problem, which make most of the solvers fail when trying to solve it, the confidence intervals of the global solution are small indicating a precise parameter estimation. This fact is discussed in more detail in [13].

Conclusion

Parameter estimation from experimental data remains a bottleneck for a major breakthrough in systems biology. Traditional global optimization methods can ensure proper solutions, but suffer from the large computational burden required for large-scale model identification. In this contribution, we have presented a novel global optimization metaheuristic, SSm, which increases very significantly the efficiency of the estimation while keeping robustness. Its capabilities were tested considering three challenging benchmark problems. This new method was able to successfully find the best known solutions for these problems while reducing the computation time by several orders of magnitude with respect to previous approaches.

Methods

Problem statement

In this work, we consider deterministic, nonlinear dynamic models of biochemical systems, i.e. those described by deterministic ordinary differential equations (ODEs), or differential-algebraic equations (DAEs). In the case of ODEs, a popular statement is the so called state-space formulation:

$$\dot{x}(p,t) = f[x(p,t), u(t),p], x(0) = x_0, \quad (8)$$

$$y(p,t) = g[x(p,t), u(p,t), p] \quad (9)$$

where x is the vector of N_x state variables and p the vector of N_p model parameters. Note that f specify the model, u specifies the vector of inputs (i.e. for a particular experiment) and y the vector of N_y measured states. An experiment is specified by the initial conditions $x(0)$, the inputs u chosen from among some set of possible inputs U and

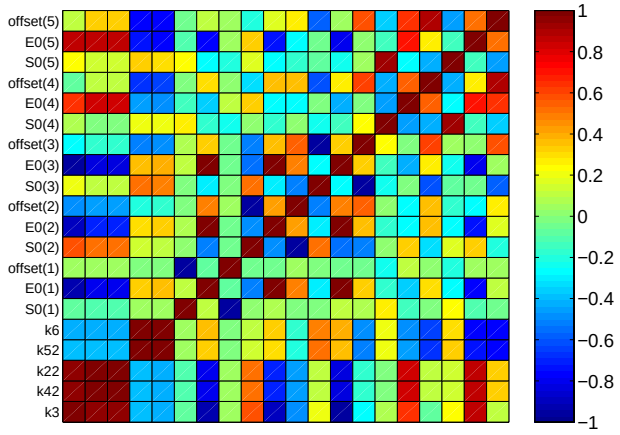

Figure 9
Correlation matrix for the inhibition of the HIV proteinase.

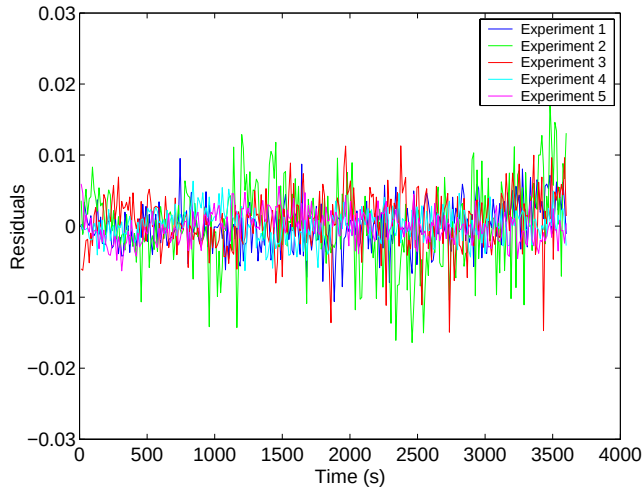

Figure 11
Residuals for the HIV proteinase case study.

the observations y. Note that the inputs can be time dependent.

Given a model structure and a set of experimental data, the goal of the parameter estimation problem is to calibrate the model so as to reproduce the experimental results in the best possible way. This is performed by minimizing a cost function that measures the goodness of the fit. Several estimator functions have been suggested as metrics, standing out the maximum likelihood estimator introduced by Fisher (1912), for being the one that maximizes the probability of the observed event.

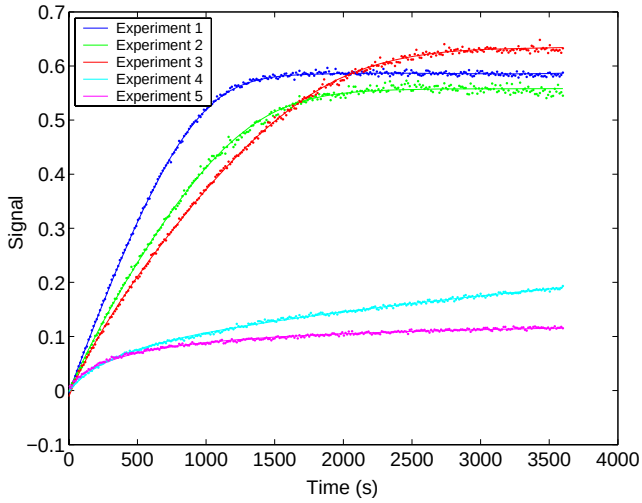

Figure 10
Experimental data versus model prediction for the HIV proteinase case study.

Maximum likelihood estimation consists of maximizing the so-called likelihood function, J_{ml}, which is the probability density of a model for the occurrence of the measurements for given parameters. The likelihood function depends on the probability of the measurements. Assuming these to be uncorrelated normal distributions, the log-likelihood function (which yields the same estimate that

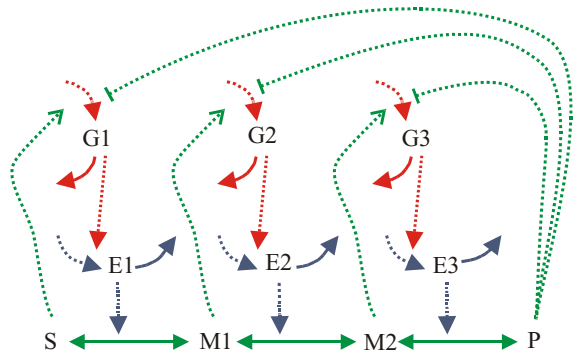

Figure 12
Three-step biochemical pathway scheme. The pathway substrate (S) and the product (P) are held at constant concentrations; Ml and M2 are intermediate metabolites of the pathway; El, E2, and E3 are the enzymes and G1, G2, and G3 are the mRNA species for the enzymes. Solid arrows indicate mass transfer reactions and point to the positive direction of flux but are chemical reversible. Dashed arrows indicate activation and dashed curves with blunt ends indicate inhibition.

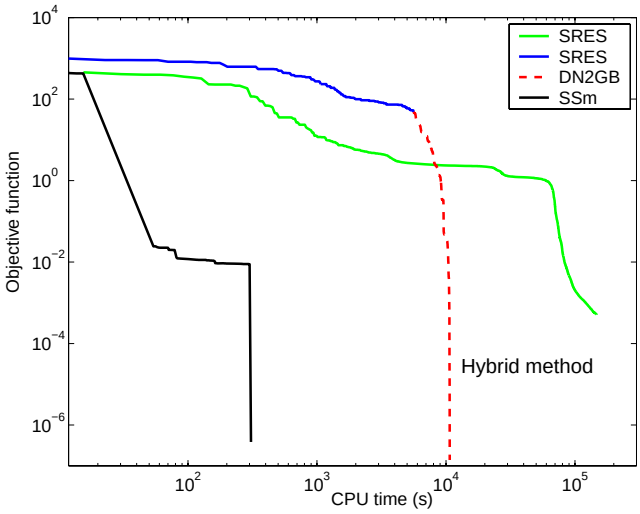

Figure 13
Convergence curves for the three-step biochemical pathway case study. Value of cost function versus computation time (in log scale) for SSm, SRES and two-phase hybrid method formed by SRES+DN2FB.

the likelihood function but is easier to handle in practice) is given as:

$$J_{ml}(p) = \frac{N}{2} ln(2\pi) + \frac{1}{2} \sum_{i=1}^{N} \left[ln(\sigma_i^2) + \frac{(\tilde{y}_i - y_i(p))^2}{\sigma_i^2} \right] \quad (10)$$

For given measurements $\tilde{y}_i$, the maximum likelihood estimates of the parameters are those values of p for which

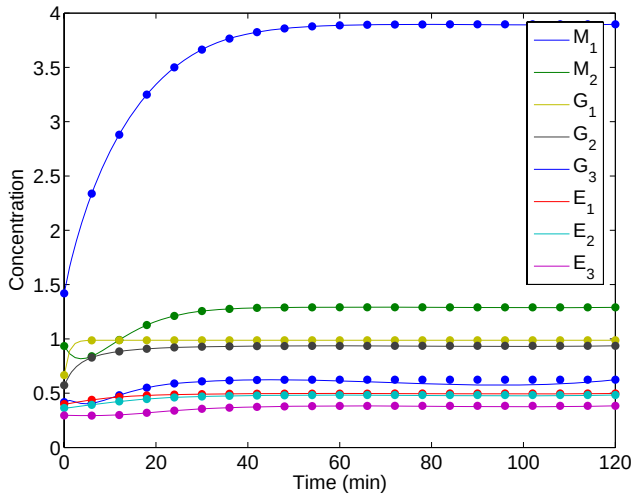

Figure 14
Experimental data versus model prediction for the three-step biochemical pathway case study.

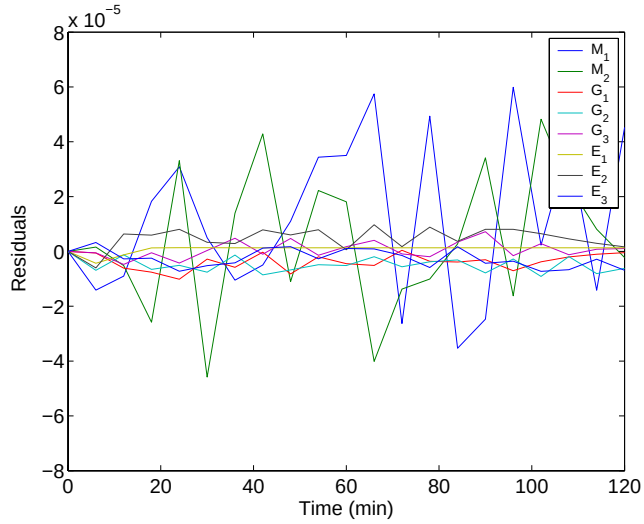

Figure 15
Residuals for the three-step biochemical pathway case study.

the likelihood function has its minimum. Moreover, if we assume the noise to be Gaussian with known of constant (homoscedastic) variance, minimizing J_{ml} (Equation 10) is equivalent to minimizing the function:

$$J_{ls}(p) = w_i[\tilde{y}_i - y_i(p)]^2 \quad (11)$$

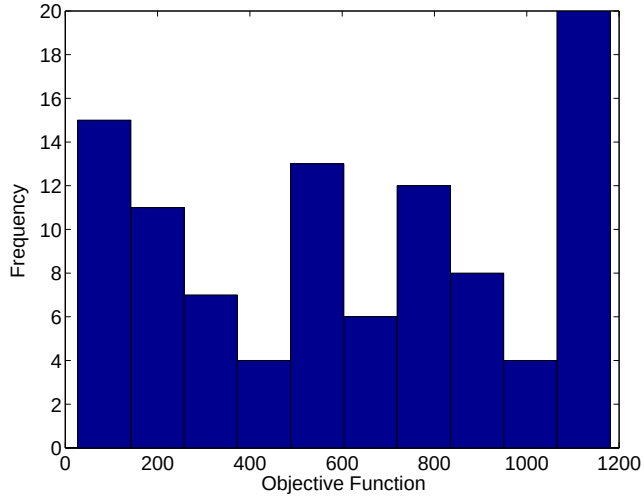

Figure 16
Multistart for the three-step biochemical pathway case study. Frequency of the solutions for a multistart using N2FB.

with the weights $W_i = \frac{1}{\sigma^2}$. One thus obtains a weighted least-squares estimator. If all σ_i's are equal, unweighed least-squares should be used ($w_i = 1$) and the noise variance do not need to be known a priori and can be estimated a posteriori from the residuals [7,8].

Confidence intervals

In general, confidence regions can be expressed as:

$$\{p : (p-p)^T C^{-1}(p-p)^T \leq n_p F^{1-\alpha}_{n_p, N-n_p}\} \quad (12)$$

The covariance matrix obtained for a linear case can be extended for nonlinear models leading an approximate covariance matrix as:

$$C_J(p) = \frac{J(p)}{N-n_p}\left[\sum_{i=1}^{N}\left(\frac{\partial \gamma_i}{\partial p}(p)\right)^T V^{-1}\left(\frac{\partial \gamma_i}{\partial p}(p)\right)\right]^{-1} \quad (13)$$

where the term $\frac{J(p)}{N-n_p}$ is an unbiased approximation of the residual variance σ^2 and $\frac{\partial \gamma}{\partial p}(p)$ the sensitivity functions with respect to the parameters evaluated at $\hat{p}$.

Under the assumption of uncorrelated measurement noise with Gaussian distribution with a mean of zero, the approximation of the covariance matrix C_J given by the Equation 13 is just the inverse of the Fisher information matrix of the estimation problem defined as:

$$FIM(p) = \sum_{i=1}^{N}\left(\frac{\partial \gamma_i}{\partial p}(p)\right)^T V^{-2}\left(\frac{\partial \gamma_i}{\partial p}(p)\right) \quad (14)$$

According to the Cramèr-Rao theorem, $C_J(\hat{p}) = FIM^{-1}$ represents the error covariance matrix of the minimum variance unbiased estimator, thus substituting C_J from Equation 13 into Equation 12, yields the approximate confidence ellipsoids.

Therefore, a lower bound for the individual parameter confidence interval σ_i ($i = 1,...,n_p$) can be obtained from the diagonal of the covariance matrix as:

$$\delta_i = \pm t^{1-(\alpha/2)}_{N-n_p}\sqrt{C_{ii}} \quad (15)$$

where $t^{1-(\alpha/2)}_{N-n_p}$ is the two-tails Student's t distribution for the given confidence level α and N - n_p degrees of freedom which converges to a linear distribution when the number of measurements N is high. Assuming that the measurement noise is white and uncorrelated we consider the error correlation matrix as diagonal, neglecting the off-diagonal elements of C, that is, the covariances among the parameters. When parameters are simultaneously determined they usually have a significant covariance thus the confidence intervals might be underestimated. That is why these confidence intervals obtained from the FIM can only be taken as lower bounds and never as an exact confidence region.

A posteriori local identifiability analysis

Under the assumption of uncorrelated measurement noise with Gaussian distribution with a mean of zero, the covariance matrix can be approximated by the inverse of the Fisher information matrix involving the output sensitivity functions. If the sensitivity equations shows linear dependence at the experimental data points, the FIM becomes singular and the model is not identifiable.

Useful information about the correlation between estimated parameters can be also obtained from the covariance matrix. The correlation matrix, which elements are the approximate correlation coefficients between the i-th and the j-th parameter, is defined by:

$$R_{ij} = \frac{C_{ij}}{\sqrt{C_{ii}C_{jj}}}, i \neq j, \quad (16)$$

$$R_{ij} = 1,\ i = j, \quad (17)$$

A singular FIM indicates the presence of unidentifiable parameters, and correlations between parameters that are greater than 0.99 may lead to singular FIM.

SSm algorithm

SSm is an advanced design of the scatter search algorithm for real variables. The method uses a relatively short number, b, of *elite* decision vectors in a so-called *reference set* (*refset*). These *elite* vectors are combined in pairs to generate new ones that may enter the *refset* replacing existing vectors in it (i.e. the *refset* always maintains a fixed number of vectors). This evolutionary approach is combined with local searches from selected vectors.

In a n-dimensional problem, vectors of decision variables are represented by, $x \in \mathbb{R}^n$ so that a particular decision variable in the population of size NP can be symbolized as

x_i^r, where $i = 1, 2,...,n$ and $r = 1, 2,...,NP$. The *refset* will have $NP = b$, whose default value is 10.

The main steps of the algorithm are shown below, with a diagram presented in Figure 1.

Generation of diverse vectors within the search space

The first step consists of generating a set S of m (default $m = 10 \cdot b$) diverse vectors in the search space. Unlike other diversification strategies, SSm does not only generate vectors with their components uniformly distributed within the search space, but also drives the generation of values for each decision variable onto parts of the space where they have not appeared very often during the search process. For that, the method makes use of memory taking into account the number of times that every decision variable appears in different parts of the search space.

Initially, the range of every decision variable, defined by its lower and upper bounds, xl_i and xu_i respectively, is divided in p (default $p = 4$) subranges of equal size, $(xu_i - xl_i)/p$. Therefore, the limits that define each subrange $j \in [1, 2,...,p]$ for the variable i are given by

• Lower bound:

$$lb_{ij} = xl_i + \frac{xu_i - xl_i}{p}(j-1) \qquad (18)$$

• Upper bound:

$$ub_{ij} = xl_i + \frac{xu_i - xl_i}{p} j \qquad (19)$$

Frequencies f_{ij} are defined as the number of times that the variable i is in the sub range j along all the generated vectors, with $i \in [1,2,...,n]$ and $j \in [1,2,...,p]$.

To initialize all the frequencies to a value of 1, p vectors are first generated, each of them having all their variables randomly generated in the same sub range using a uniform distribution (e.g. vector 1, x^1, has all its variables in sub range 1, and every decision variable i is randomly generated using a uniform distribution within the bounds xl_i and $xl_i + \frac{(xu_i - xl_i)}{p}$). This first set of vectors forms the initial matrix of diverse vectors $S^{p \times n}$ that will be extended up to a size of $S^{m \times n}$ by adding new diverse vectors.

$$S = \begin{bmatrix} x^1 \\ x^2 \\ \vdots \\ x^p \end{bmatrix} \; with \; f = \begin{pmatrix} f_{11} & f_{12} & \cdots & f_{1p} \\ f_{21} & f_{22} & \cdots & f_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ f_{n1} & f_{n2} & \cdots & f_{np} \end{pmatrix} = ones(n,p) \qquad (20)$$

New vectors will be generated using the following procedure:

For each new vector x^{p+t} to be generated, the probability of having its decision variable i in the sub range j is calculated as

$$prob_{i,j}^{p+t} = \frac{\frac{1}{f_{ij}}}{\sum_{k=1}^{p} \frac{1}{f_{ik}}} \qquad (21)$$

with $t \in [1,2,...,m - p]$, $i \in [1,2,...,n]$ and $j \in [1,2,...,p]$.

Then, a uniformly distributed random number, *rnd*, in the interval [0 1] is generated. The next generated vector x^{p+t} will have its i-th component in the subrange $j = a$ for the first value of a that accomplishes

$$rnd \le \sum_{j=1}^{a} prob_{i,j}^{p+t} \quad a = 1,2,...,p \qquad (22)$$

Each component, x_i^{p+t}, will take a value randomly selected using an uniform distribution in the range $[lb_{ij}, ub_{ij}]$.

Thus, for a new vector to be generated, the probability of having the variable i in the subrange j is inversely proportional to the frequency of appearance of the variables i in this subrange along the already created vectors. Therefore, the method has to "remember" and update these frequencies to enhance diversity. As new vectors x^{p+t} are generated, they are added to the matrix S in rows until it becomes m-by-n dimensional.

Refset formation

When the diverse vectors have been generated a selected number of them will create the first *refset*, R. There are two strategies to do it.

The first one consists of evaluating the *fitness* $f(x)$ (i.e. the cost function) of all diverse vectors and select the $b/2$ best ones in term of *fitness*. For example in a minimization problem, provided the diverse vectors are sorted according to their *fitness* (the best one first), the initial selection is $R^{b/2 \times n} = [x^1, x^2,...,x^{b/2}]$ such that

$$f(x^i) \leq f(x^j)\ \forall\ j > i,\ i \in [1,2,...,b/2],\ j \in [2,3,...,m] \quad (23)$$

The current number of vectors present in the *refset* is computed as h. Therefore, in this stage $h = b/2$ and the maximum value of h is b. We complete the *refset* with the remaining diverse vectors not yet included by maximizing the minimum Euclidean distance to the included vectors in the *refset*.

For every diverse vector not yet included in the *refset*, x^d with $d \in [h + 1, h + 2,...,m]$, Euclidean distances to all current *refset* vectors are computed. The minimum of these distances, d_{min}, is stored for each vector **x**:

$$d_{min}(x^d) = min\{d(x^d, R\} \quad (24)$$

where $d(x, R)$, represents a vector whose components are the Euclidean distances between vector **x** and all the vectors in the matrix R.

Then, the vector **x** having the highest minimum distance will join the *refset*, $R = R \cup x$ such that

$$d_{min}(x) = max(d_{min}(x^d))\ \forall\ d = h + 1,\ h + 2,...,m \quad (25)$$

and the value of h is increased one unit since a new vector has been added to the *refset*. This is repeated until the *refset* is filled with b vectors (i.e. $h = b$) so that $R \in \mathbb{R}^{b \times n}$.

The second strategy does not take into account the *fitness* of the diverse vectors. The initial *refset* is formed by 3 vectors: one having all the variables in their lower bounds, one having all the variables in their upper bounds and the middle point between these two vectors. This initial *refset* $R \in \mathbb{R}^{3 \times n}$ is completed using the same distance criterion described in the first strategy until it is composed of b decision vectors. Please note that the first strategy involves a higher computational cost since the *fitness* of all the diverse vectors has to be evaluated. However, this strategy ensures a better quality of the initial *refset* which can help to converge faster to the global solution.

Combination

Unlike genetic algorithms or other evolutionary strategies, scatter search does not use mutation or crossover operators among its members, but combinations among them. SSm combines all the vectors in the *refset* in pairs, making use of memory to avoid the combination of two vectors that have already been combined. The number of vectors created from each pair of *elite* vectors depends on the quality of the latter. These combinations are of the following three types, assuming x' and x'' being the *elite* vectors to be combined and being x' superior in quality to x'':

• Type 1: $c_1 = x' - d$

• Type 2: $c_2 = x' + d$

• Type 3: $c_3 = x'' + d$

where $d = r. \cdot (x'' - x')/2$

And r is a vector of dimension n with all its components being uniform random numbers in the interval [0 1].

Please note that the notation $.\cdot$ above indicates that the vectors are multiplied component by component, thus that is not an scalar product.

The vector has the form

$$d = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix} = \begin{bmatrix} \frac{r_1(x_1'' - x_1')}{2} \\ \frac{r_2(x_2'' - x_2')}{2} \\ \vdots \\ \frac{r_n(x_n'' - x_n')}{2} \end{bmatrix} \quad (26)$$

If both x' and x'' belong to the best $b/2$ elements of the *refset* in terms of *fitness*, then 4 vectors are generated: one of type 1, one of type 3 and two of type 2.

If only x' belongs to the best $b/2$ elements of the *refset* in terms of *fitness*, then 3 vectors are generated: one of each type.

If neither x'' nor x' belong to the best $b/2$ elements of the *refset* in terms of *fitness*, then 2 vectors are generated: one of type 2 and another one of type 1 or 3 (randomly chosen).

This type of combination allows more diversity in the generated vectors than the linear combination used in classical implementations of scatter search. These vectors generated by combination of *refset* members will be named x^c with $c \in [1, 2,..., nc]$, and form a matrix $C \in \mathbb{R}^{nc \times n}$ where nc is the total number of vectors generated by combination, which is not a fixed number. It may change every iteration depending on the number of combinations made among *refset* members (remember that the method avoids doing combinations with pairs of vectors already combined).

Refset update

Once the combinations have been done, the new vectors generated may replace the *elite* vectors if the *refset* can increase its quality. Each new vector created by combination which is better than the worst vector in *refset* is compared with all the *elite* vectors. If new vectors outperform

elite vectors in terms of *fitness* they replace them as long as they comply with a minimum diversity (i.e. the method avoids vector duplication in the *refset* by computing Euclidean distances among all vectors).

The best generated vector by combination is compared with the worst vector in *refset*. If the former outperforms the latter and is not included in the *refset*, the replacement is carried out. Otherwise, the algorithm tries to find another vector in the *refset* to accomplish both conditions and do the replacement.

The first candidate vector to join the *refset* among the *nc* generated vectors by combination is *z* such that

$$f(z) \leq f(x^i) \ \forall \ i = 1,2,...,nc \quad (27)$$

The possible vector to be replaced in the *refset* is the worst in the *refset*, x^w such that

$$f(x^w) \geq f(x^j) \ \forall \ j = 1,2,...,b \quad (28)$$

The replacement will be carried out if

$$f(z) \leq f(x^w) \ and \ z \notin R \quad (29)$$

Regardless the replacement is done or not, *z* is deleted from the matrix *C*, therefore *nc* is decreased in one unit. This is repeated with every generated vector by combination until no new vectors are better in quality than the current worst vector in *refset*.

There is an exception to these rules: if one vector has the best *fitness* in terms of quality found so far, it will join the *refset* replacing the worst vector in it or, in case that the diversity condition can not be achieved, the closest *elite* vector to it.

A mechanism to avoid flat zones is added to the *refset* update. In flat areas, many vectors with very similar (and sometimes the same) *fitness* can appear. To avoid including vectors from the same flat area, new vectors can only join the *refset* if the candidate vector has a different *fitness* value apart from being diverse enough. This prevents vectors in the same flat area from joining the *refset* at the same time.

Provided the diversity criterion is accomplished, the candidate vector *z* will join the *refset* only if

$$f(z) < f(x^i)(1 - \varepsilon) \ \forall \ x^i \in R \quad (30)$$

where ε is a small value defined by the user.

Refset regeneration

When all possible new combinations have been done and none of the generated vectors can replace any of the *elite* vectors, the algorithm can either stop or continue by regenerating the *refset*. The latter strategy is used in our algorithm. The worst *g elite* vectors (in terms of *fitness*) are deleted. New diverse vectors are generated (see above) and the *refset* is refilled according to a diversity criterion as the one described in the *refset* formation.

Normally $g = b/2$ but in aggressive implementations it can be set to $b - 1$ (i.e. all the vectors in the *refset* except the best one are deleted).

A new strategy for regenerating the *refset* has been implemented in SSm. Because the classical diversity criterion based on Euclidean distances described above does not ensure that the search will be performed along the different dimensions of the space. In our new strategy the vectors refilling the *refset* are chosen to maximize the number of relative directions defined by them and the existing vectors in the *refset*.

After deleting the *g* worst solutions the *refset* is $(b - g) \times n$ dimensional. Again, we compute *h* as the number of existing vectors in the current *refset* thus when the regeneration starts $h = b - g$.

A new matrix *M* containing the vectors that define the segments formed by the best vector in *refset* and the rest of vector in it is defined as

$$M^{h-1 \times n} = x^1 - x^k \ \forall \ k = 2,3,...,h \quad (31)$$

with x^1 being the best element not deleted in *refset* in terms of *fitness* and x^k the rest of the elements in it (note that the *refset* is ordered according to *fitness*).

For every diverse vector x^v with $v \in [1, 2,...,m]$ to join the *refset* in the regeneration phase a vector *P* of scalar products is also defined:

$$P^v = (x^1 - x^v) \cdot M^T \quad (32)$$

where x^1 is again the best not deleted element in *refset* and M^T is the transpose matrix of *M*. For every x^v the maximum value of its vector P^v is computed as $msp(x^v)$.

The solution $v \in x^v$ will join the *refset* in the regeneration phase if

$$msp(v) = min\{msp(x^v)\} \quad (33)$$

with $v \in x^v$. In this stage, the value of *h* is increased one unit and the process continues until $h = b$. The application

of this strategy results in a maximum diversity in search directions on the regenerated *refset*.

Local search – filters

Local searches are carried out from different vectors as initial points to accelerate the convergence to the minima as shown in Figure 1. The user can use a different set of local solvers (see list above) to solve their problems. When a local (maybe global) solution provided by a local search outperforms the vector used as initial point to start the local search in terms of *fitness*, the former replaces the latter and becomes a member to join the *refset*. Otherwise, the solution obtained in the local search is discarded.

To avoid doing too many local searches or start from different initial points that might provide the same local solutions, two filters are implemented in the routine. The first one is a *merit* filter that takes into account the *fitness* of the vectors so that a local search is not started from bad vectors in terms of *fitness*. The other filter takes into account distances from initial points to the local solutions they provide, thus it avoids starting local searches from the area of influence of already found minima.

In principle, both filters must be passed to start a local search, but depending on the characteristics of the problem, any of them (or both) can be deactivated. Furthermore, they can be relaxed if no vectors passing them are found after a number of consecutive iterations.

Stopping criterion

The stopping criterion is taken as a combination of three conditions:

- maximum number of evaluations exceeded
- maximum computational time exceeded
- value to reach of the cost function satisfied

By default, the algorithm will stop when any of these conditions is satisfied.

Authors' contributions

MRF performed the parameter estimation and the identifiability analysis and drafted the manuscript. JAE developed the novel metaheuristic and assisted in the preparation of the manuscript. JRB conceived of the study and participated in its design and coordination. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by the European Community as part of the FP6 COSBICS Project (STREP FP6-512060) and by Xunta de Galicia (PGIDIT05PXIC40201PM). Author JAE gratefully acknowledges financial support (FPU fellowship) from the Spanish Ministry of Education and Science.

References

1. Kell DB: **Metabolomics and systems biology: making sense of the soup.** *Current Opinion in Microbiology* 2004, **7(3):**296-307.
2. Mendes P, Camacho D, de la Fuente A: **Modelling and simulation for metabolomics data analysis.** *Biochemical Society Transactions* 2005, **33:**1427-1429.
3. Kell DB: **Metabolomics, modelling and machine learning in systems biology – towards an understanding of the languages of cells.** *FEBS Journal* 2006, **273(5):**873-894.
4. Carson E, Cobelli C: *Modelling methodology for physiology and medicine* Academic Press; 2001.
5. Aksenov S, Church Dhiman AB, Georgieva A, Sarangapani R, Helmlinger G, Khalil I: **An integrated approach for inference and mechanistic modeling for advancing drug development.** *FEBS Letters* 2005, **579(8):**1878-1883.
6. Wolkenhauer O, Ullah M, Wellstead P, Cho K: **The dynamic systems approach to control and regulation of intracellular networks.** *FEBS Letters* 2005, **579(8):**1846-1853.
7. Ljung L: *System Identification: Theory for the User* Prentice Hall; 1999.
8. Walter E, Pronzato L: *Identification of Parametric Models from Experimental Data* Springer; 1997.
9. Cho K, Shin S, Kolch W, Wolkenhauer O: **Experimental design in systems biology, based on parameter sensitivity analysis using a Monte Carlo method: A case study for the TNF a-mediated NF-kB signal transduction pathway.** *Simulation – Transactions of the Society for Modeling and Simulation International* 2003, **79:**726-739.
10. Gadkar K, Gunawan R, Doyle F III: **Iterative approach to model identification of biological networks.** *BMC Bioinformatics* 2005, **6:**155.
11. Gadkar K, Varner J, Doyle F III: **Model identification of signal transduction networks from data using a state regulator problem.** *IEE Systems Biology* 2005, **2(1):**17-30.
12. Kremling A, Fischer S, Gadkar K, Doyle FJ, Sauter T, Bullinger E, Allgöwer F, Gilles ED: **A benchmark for Methods in Reverse Engineering and Model Discrimination: Problem Formulation and Solutions.** *Genome Res* 2004, **14:**1773-1785.
13. Rodriguez-Fernandez M, Mendes P, Banga J: **A hybrid approach for efficient and robust parameter estimation in biochemical pathways.** *BioSystems* 2006, **83:**248-265.
14. Sugimoto M, Kikuchi S, Tomita M: **Reverse engineering of biochemical equations from time-course data by means of genetic programming.** *BioSystems* 2005, **80:**155-164.
15. Voit EO, Almeida J: **Decoupling dynamical systems for pathway identification from metabolic profiles.** *Bioinformatics* 2004, **20:**1670-1681.
16. Polisetty PK, Voit EO, Gatzke EP: **Identification of metabolic system parameters using global optimization methods.** *Theoretical Biology and Medical Modelling* 2006, **3:**4.
17. Bates DM, Watts DG: *Nonlinear Regression Analysis and its Applications* Wiley; 1988.
18. Schittkowski K: *Numerical data fitting in dynamical systems – A practical introduction with applications and software* Kluwer Academic Publishers; 2002.
19. Moles C, Mendes P, Banga J: **Parameter estimation in biochemical pathways: a comparison of global optimization methods.** *Genome Research* 2003, **13:**2467-2474.
20. Zwolak J, Tyson J, Watson L: **Globally optimised parameters for a model of mitotic control in frog egg extracts.** *IEE Proceedings Systems Biology* 2005, **152(2):**81-92.
21. Tsai K, Wang F: **Evolutionary optimization with data collocation for reverse engineering of biological networks.** *Bioinformatics* 2005, **21(7):**1180-1188.
22. Dhar P, Meng T, Somani S, Ye L, Sakharkar K, Krishnan A, Ridwan A, Wah S, Chitre M, Hao Z: **Grid Cellware: the first grid-enabled tool for modelling and simulating cellular processes.** *Bioinformatics* 2005, **21(7):**1284-1287.
23. Ji X, Xu Y: **libSRES: a C library for stochastic ranking evolution strategy for parameter estimation.** *Bioinformatics* 2006, **22(1):**124-126.

24. Esposito WR, Floudas CA: **Global Optimization for the Parameter Estimation of Differential-Algebraic Systems.** *Ind Eng Chem Res* 2000, **39(5):**1291-1310.
25. Singer AB, Bok JK, Barton PI: **Convex Underestimators for Variational and Optimal Control Problems.** *Comp Aided Chem Eng* 2001, **9:**767-772.
26. Papamichail I, Adjiman CS: **A Rigorous Global Optimization Algorithm for Problems with Ordinary Differential Equations.** *J Global Optim* 2002, **24:**1-33.
27. Banga J, Moles C, Alonso A: **Global optimization of bioprocesses using stochastic and hybrid methods.** In *Nonconvex Optimization and Its Applications. Frontiers In Global Optimization Volume 74*. Edited by: Floudas C, PM Pardalos E. Kluwer Academic Publishers; 2003:45-70.
28. Storn R, Price K: **Differential Evolution – a simple and efficient heuristic for global optimization over continuous spaces.** *J Global Optim* 1997, **11:**341-359.
29. Runarsson T, Yao X: **Stochastic Ranking for Constrained Evolutionary Optimization.** *IEEE Trans Evol Comp* 2000, **4:**284-294.
30. Glover F: **Heuristics for integer programming using surrogate constraints.** *Decision Sciences* 1977, **8(1):**156-166.
31. Laguna M, Marti R: *Scatter Search: Methodology and Implementations in C* The Netherlands: Kluwer Academic Publishers; 2003.
32. Laguna M, Marti R: **Experimental testing of advanced scatter search designs for global optimization of multimodal functions.** *J Global Optim* 2005, **33(2):**235-255.
33. Neumaier A, Shcherbina O, Huyer W, Vinko T: **A comparison of complete global optimization solvers.** *Math Program* 2005, **103(2):**335-356.
34. Ye Y: **Interior algorithms for linear, quadratic and linearly constrained non-linear programming.** In *PhD thesis* Department of ESS, Stanford University; 1987.
35. Gill PE, Murray W, Saunders MA: **SNOPT: An SQP algorithm for large-scale constrained optimization.** *SIAM J Optim* 2002, **12(4):**979-1006.
36. Abramson MA: **Pattern Search Algorithms for Mixed Variable General Constrained Optimization Problems.** In *PhD thesis* Houston, Texas, Rice University; 2002.
37. Dennis J, Gay D, Welsch R: **Algorithm 573, NLZSOL – An adaptive nonlinear least-squares algorithm.** *ACM Trans Math Software* 1993, **7:**369-383.
38. Jonhson ML: **Why, When, and How Biochemist Should Use Least Squares.** *Analytical Biochemistry* 1992, **206:**215-225.
39. Marsili-Libelli S, Guerrizio S, Checchi N: **Confidence regions of estimated parameters for ecological systems.** *Ecological Modelling* 2003, **165:**127-146.
40. Vanrolleghem P, Dochain D: **Bioprocess Model Identification.** In *Advanced Instrumentation, data interpretation, and control of biotechnological process* Edited by: Van Impe JFF, Vanrolleghem PE, Iserentant DM. Kluwer Academic Publishers; 1998:251-318.
41. Faller D, Klingmüller U, Timmer J: **Simulation Methods for Optimal Experimental Design in Systems Biology.** *Simulation* 2003, **79:**717-725.
42. Audoly S, Bellu G, D'Angio L, Saccomani M, Cobelli C: **Global identifiability of nonlinear models of biological systems.** *IEEE Trans Biomedical Engineering* 2001, **48(I):**55-65.
43. Dokos S, Lovell NH: **Parameter estimation in cardiac ionic models.** *Progress in Biophysics and Molecular Biology* 2004, **85:**407-431.
44. Baker CTH, Bocharov GA, Paul CAH, Rihan FA: **Computational modelling with functional differential equations: Identification, selection, and sensitivity.** *Applied Numerical Mathematics* 2005, **53:**107-129.
45. Zak DE, Gonye GE, Schwaber JS, Doyle FJ III: **Importance of Input Perturbations and Stochastic Gene Expression in the Reverse Engineering of Genetic Regulatory Networks: Insights From an Identifiability Analysis of an In Silico Network.** *Genome Res* 2003, **13:**2396-2405.
46. Box GEP, Hunter WG, MacGregor JF, Erjavec J: **Some problems associated with the analysis of multiresponse data.** *Technometrics* 1973, **15:**33-51.
47. Dolan ED, Moré JJ, Munson TS: **Benchmarking optimization problems with COPS 3.0.** *Technical Report ANL/MCS-TM-273. Argonne National Laboratory* 2004.
48. Fuguitt R, Hawkins JE: **Rate of Thermal Isomerization of α-Pinene in the Liquid Phase.** *JACS* 1947, **69:**461.
49. Hunter WG, McGregor JF: **The Estimation of Common Parameters from Several Responses: Some Actual Examples.** In *Unpublished Report* The Department of Statistics. University of Wisconsin; 1967.
50. Tjoa IB, Biegler LT: **Simultaneous solution and optimization strategies for parameter estimation of differential-algebraic equation systems.** *Ind Eng Chem Res* 1991, **30(2):**376-385.
51. Averick BM, Carter RG, Moré JJ: **The MINPACK-2 test problem collection.** *Technical Report ANL/MCS-TM-273, Argonne National Laboratory* 1991.
52. Kuzmic P: **Program DYNAFIT for the analysis of enzyme kinetic data: application to HIV proteinase.** *Analytical Biochemistry* 1996, **237:**260-273.
53. Mendes P, Kell D: **Non-Linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation.** *Bioinformatics* 1998, **14(10):**869-883.
54. Guay M, McClean D: **Optimization And Sensitivity Analysis for Multiresponse Parameter Estimation in Systems of Ordinary Differential Equations.** *Comput Chem Eng* 1995, **19:**1271-1285.