

METHODOLOGY ARTICLE

Open Access

Parameter estimation of dynamic biological network models using integrated fluxes

Yang Liu and Rudiyanto Gunawan*

Abstract

Background: Parameter estimation is often the bottleneck step in biological system modeling. For ordinary differential equation (ODE) models, the challenge in this estimation has been attributed to not only the lack of parameter identifiability, but also computational issues such as finding globally optimal parameter estimates over highly multidimensional search space. Recent methods using incremental estimation approach could alleviate the computational difficulty by performing the parameter estimation one-reaction-at-a-time. However, incremental estimation strategies usually require data smoothing and are known to produce biased parameter estimates.

Results: In this article, we presented a new parameter estimation method called integrated flux parameter estimation (IFPE). We employed the integral form of the ODE such that we could compute the integral of reaction fluxes from time-series concentration data without data smoothing. Here, we formulated the parameter estimation as a nested optimization problem. In the outer optimization, we performed a minimization of model prediction errors over parameters associated with a subset of reactions labeled as independent. The dimension of the independent reaction subset was equal to the degrees of freedom in the calculation of integrated fluxes (IF) from concentration data. We selected the independent reactions such that given their IF values, the IFs of the remaining (dependent) reactions could be uniquely determined. Meanwhile, in the inner optimization, we estimated the model parameters associated with the dependent reactions, one-reaction-at-a-time, by minimizing the dependent IF prediction errors. We demonstrated the performance of the IFPE method using two case studies: a generalized mass action model of a branched pathway and a lin-log ODE model of *Lactococcus lactis* glycolytic pathway.

Conclusions: The IFPE significantly outperformed standard simultaneous parameter estimation in terms of computational efficiency and scaling. In comparison to incremental parameter estimation (IPE) method, the IFPE produced parameter estimates with significantly lower bias and did not require time-series data smoothing. The advantages of IFPE over the IPE however came at the cost of a small increase in the computational time.

Keywords: Parameter estimation, ODE model, Power-law model, Lin-log model

Background

Mathematical modeling is one of the pillars of systems biology. Here, ODEs have been commonly used to model cellular systems, especially when dynamical behavior is of interest. An ODE model is formulated based on viewing cellular networks as chemical reaction networks, where the equations describe the mass or molar balance as follow:

$$\frac{d\mathbf{X}(t)}{dt} = \mathbf{S}\mathbf{v}(\mathbf{X}(t), \mathbf{p}); \quad \mathbf{X}(0) = \mathbf{X}_0, \quad (1)$$

where $\mathbf{X}(t) \in \mathbf{R}^m$ is the vector of m species concentrations, $\mathbf{S} \in \mathbf{R}^{m \times n}$ is the stoichiometric matrix, $\mathbf{v}(\mathbf{X}(t), \mathbf{p})$ is the vector of n reaction rate equations, $\mathbf{p} \in \mathbf{R}^d$ is the vector of d kinetic parameters, and \mathbf{X}_0 is the vector of initial concentrations. The creation of ODE models in biology is often hampered by imprecise knowledge of the reaction rate equations and kinetic parameters [1]. Therefore, many model parameters have to be estimated from experimental data. Intuitively, time-series concentration data are desirable for estimating kinetic parameters of ODE models.

Model parameter estimation is typically formulated as a global optimization problem, minimizing the difference

*Correspondence: rudi.gunawan@chem.ethz.ch
Institute for Chemical and Bioengineering, ETH Zurich, Vladimir-Perlog-Weg 1,
8093 Zurich, Switzerland

between experimental observations and model prediction. We can classify existing methods for estimating ODE model parameters into two general groups: simultaneous and incremental approach [2]. In the simultaneous approach, we search for the optimal parameter combination that minimizes the deviation of simulated concentration predictions from the experimental concentration data. Unfortunately, the parameter estimation from biological data is often underdetermined, where many parameter combinations could fit the data equally well [3,4]. In addition, other computational factors such as finding the global optimal solution over highly multidimensional parameter space and integrating stiff ODEs, often make the parameter estimation numerically intractable, even for models with 10–20 parameters [5].

The incremental approach has been used recently to alleviate the computational issues mentioned above [6–9]. In this approach, the parameter estimation is performed in several (incremental) steps. First, we smoothen time-series concentration data $\mathbf{X}_M(t)$, and differentiate the resulting smoothing functions to obtain estimates of $d\mathbf{X}_M(t)/dt$. Subsequently, we evaluate the dynamic reaction rate or flux values $\mathbf{v}(t)$ from $d\mathbf{X}_M(t)/dt$ by solving Eq. (1) algebraically. If the stoichiometric matrix \mathbf{S} has a full column rank, then the flux estimates could be obtained by multiplying the (pseudo-)inverse of \mathbf{S} with $d\mathbf{X}_M(t)/dt$. Finally, we perform the kinetic parameter estimation one reaction at a time, by minimizing the sum of squares of the differences between $\mathbf{v}(t)$ and $\mathbf{v}(\mathbf{X}_M(t), \mathbf{p})$. Here, not only the ODE model is not integrated, but also the optimizations involve much smaller parameter space than those in the simultaneous approach. For these reasons, methods based on the incremental approach are usually much faster than those using the simultaneous approach. However, the parameters obtained using incremental estimation strategy are known to be biased and sensitive to data smoothing procedure [2].

In the models of cellular networks, such as metabolic networks, the (pseudo-)inverse of \mathbf{S} often does not exist since cellular species typically participate in more than one reaction (i.e. $m < n$). In addition, experimental measurements are typically taken for only a fraction of the species in the model. In this case, there are degrees of freedom in the flux estimation when applying the incremental estimation approach. Recently, we presented an incremental parameter estimation method that addressed the above issue [10]. We formulated the parameter estimation as a nested optimization problem. Here, we partitioned the reaction rates into independent and dependent subsets, \mathbf{v}_I and \mathbf{v}_D , respectively (i.e. $\mathbf{v} = [\mathbf{v}_I^T \mathbf{v}_D^T]^T$). Because of the degrees of freedom, we could select an appropriate

$\mathbf{v}_I(t)$, such that $\mathbf{v}_D(t)$ can be uniquely obtained from $d\mathbf{X}_M(t)/dt$ and $\mathbf{v}_I(t)$ using Eq. (1). We formulated the outer optimization problem to estimate parameters associated with the independent reactions, referred to as independent parameters \mathbf{p}_I . Meanwhile, the remaining (dependent) parameters \mathbf{p}_D were estimated in the inner optimization using the dependent reaction flux estimates. Henceforth, we refer the aforementioned estimation method as the incremental parameter estimation (IPE). The IPE could offer several orders of magnitude reduction in the computational time in comparison to standard simultaneous and incremental methods. However, the IPE method was affected by the same issues related to parameter bias and sensitivity to data smoothing mentioned above.

A new class of incremental parameter estimation methods has recently been proposed without the need to smoothen and differentiate noisy time series data [11–13]. In these methods, one calculates the overall extents of reactions directly from time-series concentration data. The extent of a reaction gives the cumulative amount of moles produced by the reaction [11]. In contrast to the traditional incremental estimation strategy, the kinetic parameters are estimated from the reaction extents. However, this method again requires that the stoichiometric matrix \mathbf{S} has a full column rank.

In this work, we present a new parameter estimation method, called integrated flux parameter estimation, which does not require the stoichiometric matrix \mathbf{S} to have a full column rank. Similar to the IPE, we formulate the IFPE as a nested optimization problem. However, in contrast to the IPE, the IFPE relies on the calculation of integrated fluxes directly from concentration data, thereby avoiding time-series data smoothing and differentiation. We show using two case studies that the IFPE method can provide much reduced parameter bias and higher reliability in comparison to the IPE method, with only a small increase in the computational time. Nevertheless, for certain ODE models such as those using lin-log rate equations, the IFPE can converge to the optimal parameter solution much faster than the IPE method.

Methods

In developing the IFPE method, we start with the integral form of the ODE model, given by:

$$\mathbf{X}(t) - \mathbf{X}(0) = \mathbf{S} \int_0^t \mathbf{v}(\mathbf{X}(\tau), \mathbf{p}) d\tau = \mathbf{S}\eta(\mathbf{X}, \mathbf{p}) \quad (2)$$

where $\eta(\mathbf{X}, \mathbf{p})$ denotes the vector of IFs. Here, the i -th IF η_i is analogous to the overall extent of the i -th reaction per unit volume of a batch reactor [13]. If the stoichiometric

matrix \mathbf{S} has a full column rank, the IF vector can be obtained directly from the concentration measurements \mathbf{X}_M as follows:

$$\eta(t_k) = \mathbf{S}^\dagger (\mathbf{X}_M(t_k) - \mathbf{X}_M(0)) \quad (3)$$

where t_k denotes the k -th measurement time point, and $\mathbf{S}^\dagger = \mathbf{S}^{-1}$ for a square \mathbf{S} matrix or $\mathbf{S}^\dagger = (\mathbf{S}^T \mathbf{S})^{-1} \mathbf{S}^T$ otherwise.

As mentioned earlier, the matrix \mathbf{S} in cellular network models often does not have a full column rank. Here, there exist degrees of freedom in the calculation of $\eta(t_k)$ from \mathbf{X}_M , which is equal to $n - \text{rank}(\mathbf{S})$, the dimension of the (right) null space of \mathbf{S} . In this case, we can select a subset of (independent) reaction rates such that given their IF values $\eta_I \in \mathbf{R}^{n-\text{rank}(\mathbf{S})}$, the values of the remaining (dependent) IFs $\eta_D \in \mathbf{R}^{\text{rank}(\mathbf{S})}$ can be uniquely determined from the relationship in Eq. (2). By partitioning the vector η into the independent and dependent components $\eta = [\eta_I^T \ \eta_D^T]^T$ and respectively the matrix \mathbf{S} into $\mathbf{S} = [\mathbf{S}_I \ \mathbf{S}_D]$, we derive the following relationship between η_I and η_D :

$$\eta_D(t_k) = \mathbf{S}_D^\dagger (\mathbf{X}_M(t_k) - \mathbf{X}_M(0) - \mathbf{S}_I \eta_{I,k}(\mathbf{X}_M, \mathbf{p}_I)), \quad (4)$$

where

$$\eta_{I,k}(\mathbf{X}_M, \mathbf{p}_I) = \int_0^{t_k} \mathbf{v}_I(\mathbf{X}_M, \mathbf{p}_I) dt \quad (5)$$

and \mathbf{p}_I are the parameters that appear in the independent reaction subset. Note that when \mathbf{S} has a full row rank, we can choose η_I such that the submatrix \mathbf{S}_D is a square non-singular matrix.

Figure 1 shows the schematic diagram of the IFPE method. Here, we consider the scenario where time-series

concentration data of all species in the model are available. However, the IFPE can be extended to a more general scenario where only a subset of species are measured (see Additional file 1: Figure S1). In the IFPE, the parameter estimation comprises a nested optimization problem, where the outer optimization involves the minimization of the error function $\Phi(\mathbf{p}_I, \mathbf{X}_M)$ given by:

$$\Phi(\mathbf{p}_I, \mathbf{X}_M) = \sqrt{\frac{1}{mK} \sum_{k=1}^K (\mathbf{X}(t_k; \mathbf{p}_I) - \mathbf{X}_M(t_k))^T (\mathbf{X}(t_k; \mathbf{p}_I) - \mathbf{X}_M(t_k))} \quad (6)$$

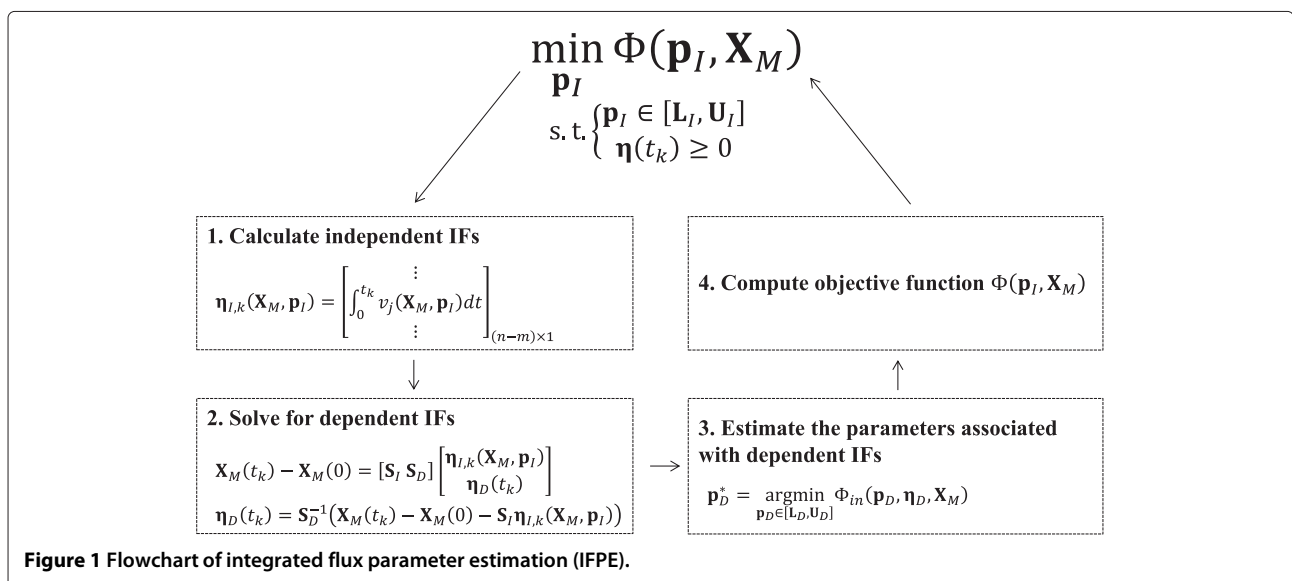
where K denotes the number of measurement time points, and $\mathbf{X}(t_k; \mathbf{p}_I)$ is the simulated concentration prediction \mathbf{X} at time t_k . The calculation of $\Phi(\mathbf{p}_I, \mathbf{X}_M)$ involves several steps. First, given the values of \mathbf{p}_I , we evaluate the independent IF functions $\eta_{I,k}(\mathbf{X}_M, \mathbf{p}_I)$ according to Eq. (5) using a modified Simpson's rule (see Additional file 1: Section 1 and Figure S2 for more detail). Subsequently, we compute the dependent IFs $\eta_D(t_k)$ using Eq. (4) and obtain the dependent parameter estimates as $\mathbf{p}_D^* = \arg \min_{\mathbf{p}_D} \Phi_{in}(\mathbf{p}_D, \eta_D, \mathbf{X}_M)$ with the (inner) error function:

$$\begin{aligned} \Phi_{in}(\mathbf{p}_D, \eta_D, \mathbf{X}_M) &= \sqrt{\frac{1}{mK} \sum_{k=1}^K (\eta_{D,k}(\mathbf{X}_M, \mathbf{p}_D) - \eta_D(t_k))^T (\eta_{D,k}(\mathbf{X}_M, \mathbf{p}_D) - \eta_D(t_k))}. \end{aligned} \quad (7)$$

where

$$\eta_{D,k}(\mathbf{X}_M, \mathbf{p}_D) = \int_0^{t_k} \mathbf{v}_D(\mathbf{X}_M, \mathbf{p}_D) dt. \quad (8)$$

The integration in Eq. (8) is also performed using a modified Simpson's rule. When each of the parameters \mathbf{p}_D



appears only in one reaction rate equation, the minimization of Φ_{in} can be performed one reaction at a time. Finally, we compute $\mathbf{X}(t_k; \mathbf{p}_I)$ either from $\eta(\mathbf{X}_M, \mathbf{p})$ according to the integral form of the ODE model (see Eq. (2)) or by simulating the ODE model. In the case studies, the latter variant of the IFPE is labeled as IFPE-ODE to indicate that the calculation of the objective function requires solving the ODE model.

Finally, there maybe more than one way to appropriately partition the reactions into the independent and dependent subsets. Here, we use a few guidelines in selecting the independent subset. First and foremost, we select η_I such that the (pseudo-)inverse of \mathbf{S}_D exists. As the computational cost of the nested optimization mainly scales with the parameter search space of the outer optimization, we therefore prefer η_I with fewer \mathbf{p}_I . Finally, we also consider prior information regarding the parameter values, and select the set of η_I with smaller ranges of \mathbf{p}_I values.

Results

Below we demonstrate the performance of the IFPE method on two case studies: a branched metabolic pathway [6] and a lin-log model of the glycolytic pathway in *L. lactis* [14]. In the case studies, we used CVODE subroutine from the package SUNDIALS (SUite of Nonlinear and Differential/ALgebraic equation Solvers) for the ODE integrations [15], with the option MaxNumSteps set to 5000. We performed the outer optimization in Eq. (6) using the enhanced Scatter Search (eSS) algorithm from SSmGO (Scatter Search for Matlab Global Optimization) toolbox [16-18], in which we terminated the parameter search when the objective function improved less than a relative tolerance of 10^{-5} for 50 successive iterations. Finally, for the inner optimization in Eq. (7), we employed the MATLAB subroutine *lsqcurvefit* with the trust region reflective option. In the outer and inner optimization, we enforced constraints on the parameter values to be within upper and lower bounds and to produce only positive IF values.

We compared the performance of the IFPE with the IPE [10]. In the IPE implementation, we first smoothed the time series data using piecewise polynomial spline fitting, and subsequently differentiated the spline functions to obtain estimates of $d\mathbf{X}_M(t)/dt$. We also performed the outer optimization using the eSS algorithm with the same convergence criterion as in the IFPE implementation. However, for the inner optimization, we chose MATLAB function *quadprog* using interior-point-convex algorithm because in the two case studies below, the inner optimization problem constituted a quadratic programming problem. We implemented two variants of the IPE using different error functions for the outer optimization: the IPE-ODE method with the error function Φ in

Eq. (6) and the IPE-slope method with the following error function:

$$\Phi_{IPE\ slope}(\mathbf{p}_I, \mathbf{X}_M) = \sqrt{\frac{1}{mK} \sum_{k=1}^K \left(\frac{d\mathbf{X}(t_k; \mathbf{p}_I)}{dt} - \frac{d\mathbf{X}_M(t_k)}{dt} \right)^T \left(\frac{d\mathbf{X}(t_k; \mathbf{p}_I)}{dt} - \frac{d\mathbf{X}_M(t_k)}{dt} \right)}. \quad (9)$$

Note that the IPE-slope method did not require any integration of the ODE model. We also enforced constraints on the parameter values by setting upper and lower bounds, and on the positivity on the reaction fluxes.

In addition to the IPE method, we further compared the IFPE to simultaneous parameter estimation (SPE) method. Here, we estimated the kinetic parameters by minimizing model prediction errors over all unknown parameters simultaneously. We also implemented two versions of the SPE method: the SPE-ODE method with the objective function in Eq. (6) and the SPE-slope method with the objective function in Eq. (9). We again used the eSS to find the global parameter solution using the same convergence criterion and parameter bounds as in the IFPE implementation.

A generic branched pathway

The first case study comes from a generalized mass action (GMA) model of a generic branched pathway shown in Figure 2. Here, the rate equations follow the canonical power-law function:

$$v_j(\mathbf{X}, \mathbf{p}) = a_j \prod_i X_i^{g_{ji}}, \quad (10)$$

where a_j is the rate constant and g_{ji} is the kinetic order associated with the i -th metabolite in the j -th reaction. The generic branched pathway comprises four metabolites and six reactions, described by the ODE model:

$$\frac{d}{dt} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{bmatrix} \quad (11)$$

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{bmatrix} = \begin{bmatrix} a_1 X_3^{-g_{13}} \\ a_2 X_1^{g_{21}} \\ a_3 X_2^{g_{32}} \\ a_4 X_3^{g_{43}} X_4^{g_{44}} \\ a_5 X_1^{g_{51}} \\ a_6 X_4^{g_{64}} \end{bmatrix}. \quad (12)$$

Using the parameters and initial conditions reported previously [6] (see Additional file 1: Section 2), we gen-

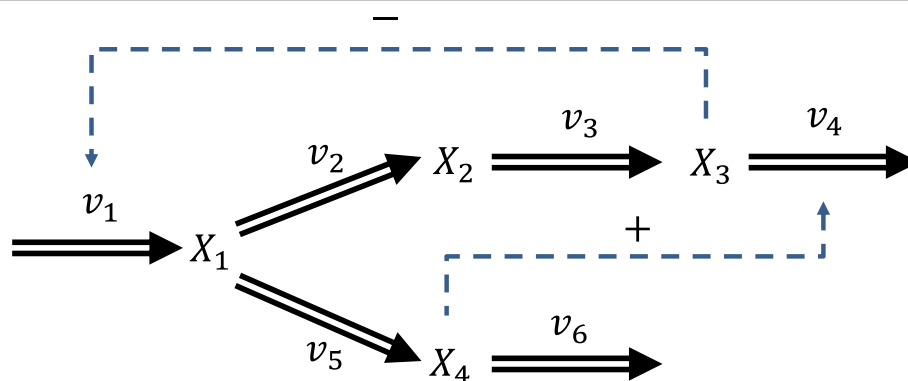


Figure 2 Metabolic network of a generic branched pathway. Double-line arrows indicate metabolic transformations and dashed arrows with plus or minus signs represent activation or inhibition, respectively.

erated noise-free and noisy time-series concentration of all metabolites X_1 to X_4 (see Additional file 1: Figure S3 for the case of missing measurements). For the noisy dataset, we simulated five technical replicates under the same condition, assuming independent additive Gaussian noise with zero mean and 10% coefficient of variation. The time-series concentration data used in this example are available in Additional file 2. Here, we selected v_1 and v_6 as the independent reaction subset, leading to a square invertible S_D and the fewest p_I of 4 parameters. Furthermore, we constrained the rate constants to within $[0, 25]$ and the kinetic orders to within $[0, 2]$.

The median relative errors of the parameter estimates from the IFPE, IPE and SPE methods are given in Table 1. We performed the IPE and SPE-slope methods using several settings of piecewise spline fitting (see Additional file 1: Figure S4), where s is the number of piecewise sections and o is the degree of the polynomials. Comparing the outcomes of the IPE and SPE-slope methods using three different (s, o) combinations showed that the

accuracy of the parameter estimates from these methods sensitively depends on the manner of which the time-series data are smoothen, especially for the IPE methods. We could generally obtain improved parameter accuracy by increasing the number of pieces and the degree of the polynomials. But, we urge caution when using more pieces and higher degree polynomials for spline fitting, as this may lead to data overfitting.

The outcomes of the noise-free data showed that the IFPE methods could provide more accurate parameter estimates than the IPE and SPE methods. For the noisy dataset, the IFPE parameter estimates have similar accuracy to the IPE and SPE methods using the best data smoothing setting. However, in practice the optimal data smoothing setting is not known. Despite the differences in the accuracy of parameter estimates from the SPE, IPE and IFPE methods, Figure 3 shows that except for the SPE-slope, methods considered here could produce parameters with a reasonably good fit to the noisy concentration data (also see Additional file 1: Table S1 for more detail).

Table 1 Comparison of median parameter errors for the branched pathway case study

Median parameter error ^a (%)	Noise-free data ^b			Noisy data ^c		
	$(s, o) = (3, 3)$	$(s, o) = (5, 3)$	$(s, o) = (3, 5)$	$(s, o) = (3, 3)$	$(s, o) = (5, 3)$	$(s, o) = (3, 5)$
SPE-slope	51.9	14.1	46.3	68.8 ± 19.4	81.2 ± 9.6	93.4 ± 7.4
IPE-slope	103	59.4	59.8	87.9 ± 4.3	68.6 ± 27.5	58.0 ± 21.6
IPE-ODE	50.0	19.7	7.74	90.9 ± 14.7	76.6 ± 35.6	71.1 ± 27.1
SPE-ODE		11.4 ^d			37.9 ± 11.5	
IFPE		0.276			66.9 ± 32.5	
IFPE-ODE		0.746			70.0 ± 31.6	

^aThe median is taken over 13 parameters in the branched pathway model.

^bFor noise-free data, five independent runs were carried out. The median parameter error corresponds to the run with the lowest objective function value.

^cFor noisy data, the reported values are the mean \pm standard deviation of five technical replicates of the data.

^dOnly three out of five repeated runs finished within 24 hours. The median parameter error is reported for the parameter estimate corresponding to the lowest objective function value among the three successful runs.

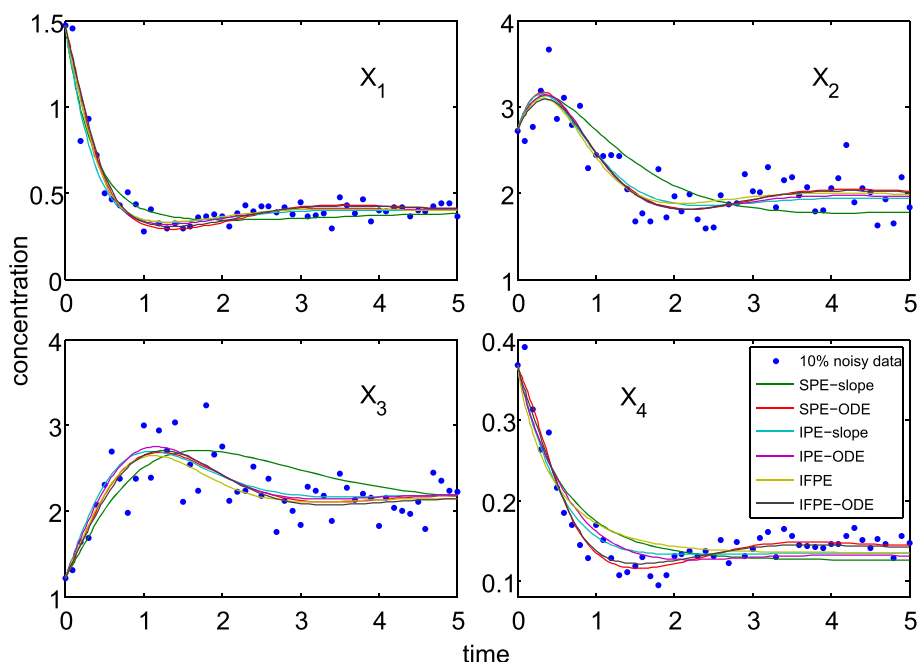


Figure 3 Comparison of model predictions using the IPE, IFPE and SPE parameter estimates for the branched pathway case study. The noisy data correspond to one set of the five technical replicates. For the IPE and SPE-slope estimates, the results correspond to $(s, o) = (5, 3)$.

Tables 2 and 3 give the computational times and the number of eSS iterations for each of the estimation methods, respectively. In general, methods requiring the integration of fluxes and/or ODEs (i.e. IFPE, IFPE-ODE, IPE-ODE and SPE-ODE) took longer to converge than the rest. Here, the SPE-ODE was the slowest method as the optimization involved the entire parameter space \mathbf{p} and the integration of ODEs. For GMA models with power-law rate equations, the inner optimization of the IPE simplified into a (log-)linear least square optimization, which could be solved much more efficiently than those in the IFPE. For this reason, the IPE-ODE method converged about twice faster than the IFPE-ODE despite requiring

more eSS iterations. The IPE-slope method was the fastest among the methods considered as it did not require any integration. Interestingly, the parameter estimations using the noise-free dataset took longer to solve than those using the noisy dataset. In this regard, data noise has a smoothing effect on the objective function surface, and the low amount of noise enhanced the convergence to optimal solution [19]. Finally, we also observed high variability in repeated runs of eSS in the parameter estimation using noise-free data. Hence, for the noise-free results in Tables 1, 2 and 3, we reported the best parameter estimates corresponding to the lowest objective function value out of five repeated runs.

Table 2 Comparison of CPU times for the branched pathway case study

CPU time ^a (sec)	Noise-free data ^b			Noisy data ^c		
	$(s, o) = (3, 3)$	$(s, o) = (5, 3)$	$(s, o) = (3, 5)$	$(s, o) = (3, 3)$	$(s, o) = (5, 3)$	$(s, o) = (3, 5)$
SPE-slope	933.84	1255.4	1182.6	207.7 ± 149.3	111.2 ± 70.5	40.2 ± 13.8
IPE-slope	106.48	104.55	185.38	96.8 ± 17.1	105.4 ± 16.0	102.9 ± 36.5
IPE-ODE	415.32	1380.1	1018.9	433.9 ± 70.6	456.4 ± 141.9	469.3 ± 174.4
SPE-ODE		14.8 hours ^d			9002 ± 4839	
IFPE		1263			655.9 ± 198.5	
IFPE-ODE		2154			1023 ± 315	

^aThe CPU times were recorded using a workstation with Intel Xeon processor 3.33 GHz with 18 GB RAM.

^bFor noise-free data, five independent runs were carried out. The CPU time is reported for the run with the lowest objective function value.

^cFor noisy data, the reported values are the mean ± standard deviation of five technical replicates of the data.

^dOnly three out of five repeated runs finished within 24 hours. The CPU time corresponds to the run with the lowest objective function value among the three successful runs.

Table 3 Comparison of the numbers of eSS iterations for the branched pathway case study

eSS iterations	Noise-free data ^a			Noisy data ^b		
	(s, o) = (3, 3)	(s, o) = (5, 3)	(s, o) = (3, 5)	(s, o) = (3, 3)	(s, o) = (5, 3)	(s, o) = (3, 5)
SPE-slope	6730	9144	8636	1440 ± 1045	753.6 ± 490.7	259.0 ± 94.9
IPE-slope	63	65	143	66.2 ± 7.2	87.0 ± 18.6	92.4 ± 69.7
IPE-ODE	75	305	225	76.6 ± 17.8	83.8 ± 33.2	92.6 ± 43.3
SPE-ODE		3827 ^c			787.8 ± 438.7	
IFPE		112			67.0 ± 13.1	
IFPE-ODE		156			70.2 ± 11.8	

^aFor noise-free data, five independent runs were carried out. The number of eSS iterations corresponds to the run with the lowest objective function value.

^bFor noisy data, the reported values are the mean ± standard deviation of five technical replicates of the data.

^cOnly three out of five repeated runs finished within 24 hours. The number of eSS iterations corresponds to the run with the lowest objective function value among the three successful runs.

In this example, there were more than one way to partition the fluxes into dependent and independent subsets, even when following the guidelines provided in the previous section. In order to investigate the sensitivity of the IFPE and IFPE-ODE performance with respect to the partitioning of the fluxes, we repeated the parameter estimation runs using five different dependent-independent sets, in which four runs involved the same number of independent parameters as above and one run involved a larger number of independent parameters ($\mathbf{v}_I = \{v_1, v_4\}$ had 5 independent parameters). The results are given in Tables S2–S5 in the Additional file 1, showing that the performance of the IFPE and IFPE-ODE is robust with respect to the flux partitioning.

The glycolytic pathway in *Lactococcus lactis*

In the second case study, we consider the parameter estimation involving a lin-log (linear-logarithmic) modeling of the glycolytic pathway in *L. lactis* [14]. Here, the enzymatic reaction rate is expressed as a linear function of the logarithm of normalized concentrations [20] as follows:

$$\frac{v_j(\mathbf{X}, \mathbf{p})}{J_j^0} = \frac{e_j}{e_j^0} \left(1 + \sum_i \varepsilon_{ji} \ln \frac{X_i}{X_i^0} \right), \quad (13)$$

where J_j^0 is the rate of the j -th reaction at the reference state, X_i^0 and e_j^0 denote the reference concentrations of the i -th metabolite and j -th enzyme, respectively, and ε_{ji} denotes the elasticity representing the influence of the i -th metabolite concentration on the j -th reaction rate. Lin-log models can be considered as an extension of metabolic control analysis (MCA) for dynamical systems, and have similar mathematical features to power-law rate equations.

The metabolic pathway is shown in Figure 4, involving nine metabolites: glucose 6-phosphate (G6P) – X_1 , fructose 1, 6-biphosphate (FBP) – X_2 , 3-phosphoglycerate (3-

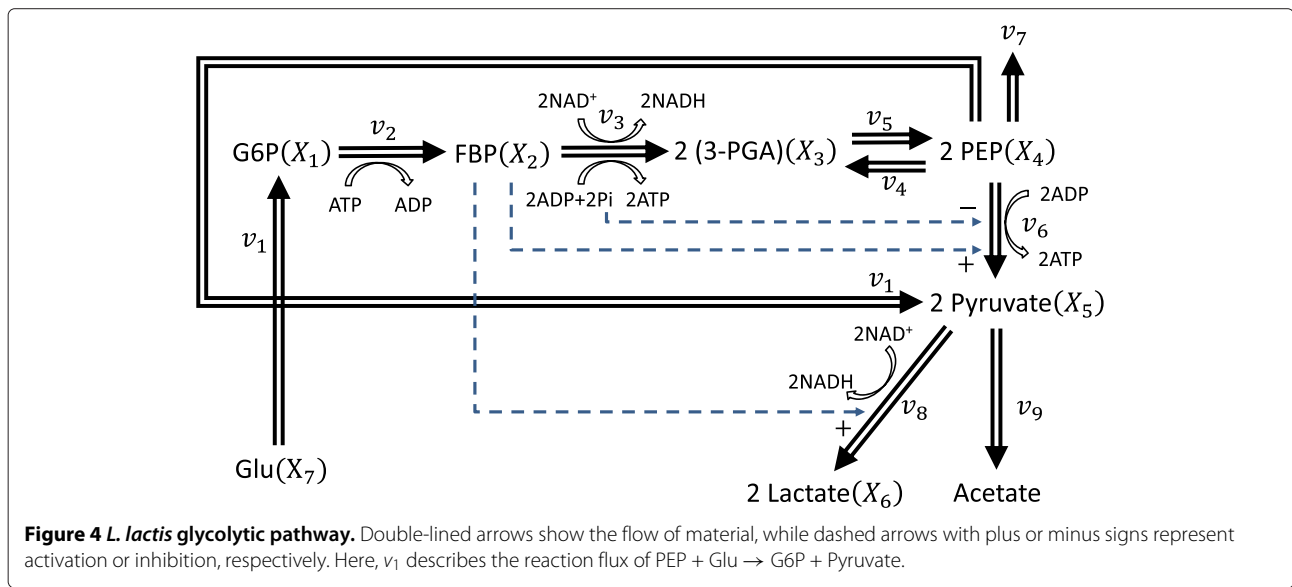
PGA) – X_3 , phosphoenolpyruvate (PEP) – X_4 , Pyruvate – X_5 , Lactate – X_6 , external glucose (Glu) – X_7 , ATP – X_8 and P_i – X_9 ; and nine metabolic fluxes. The corresponding lin-log model is given by:

$$\frac{d}{dt} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 1 & -1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & -1 & 1 & -1 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \\ v_8 \\ v_9 \end{bmatrix} \quad (14)$$

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \\ v_8 \\ v_9 \end{bmatrix} = \begin{bmatrix} a_1 + g_{11} \log X_1 + g_{14} \log X_4 + g_{17} \log X_7 \\ a_2 + g_{21} \log X_1 + g_{28} \log X_8 \\ a_3 + g_{32} \log X_2 + g_{39} \log X_9 \\ a_4 + g_{44} \log X_4 \\ a_5 + g_{53} \log X_3 \\ a_6 + g_{62} \log X_2 + g_{64} \log X_4 + g_{69} \log X_9 \\ a_7 + g_{74} \log X_4 \\ a_8 + g_{82} \log X_2 + g_{85} \log X_5 \\ a_9 + g_{95} \log X_5 \end{bmatrix} \quad (15)$$

The parameters of the lin-log model above have been simplified into a_i 's and g_{ij} 's, and thus do not necessarily have any direct physical interpretation. The experimental data consisted of time-series *in vivo* NMR measurements of the metabolites [21] (data taken from supplementary material of [14]).

Following a previous parameter estimation case study of the above model, we considered only metabolite concentration data up to 10 minutes (see Figure 5), in order to



avoid taking logarithms of zero concentration values [14]. We treated the external glucose, ATP and P_i (i.e. X_7 , X_8 and X_9) as off-line variables, using piecewise spline fitting of time-series concentration data with $(s, o) = (6, 4)$ (missing time-points were first linearly interpolated from the remaining data). We assigned v_5 , v_7 and v_9 as the independent fluxes to ensure an invertible \mathbf{S}_D submatrix and the fewest independent parameters \mathbf{p}_I . Furthermore,

we constrained all the parameters of the lin-log model to within $[-500, 500]$. For lin-log models, the inner optimization in the IFPE methods reduces to a linear least-square problem and thus can be performed very efficiently. In this case, we computed $\int \ln X_i dt$ for each metabolite beforehand.

The parameter estimation of the lin-log model above has been shown to be extremely challenging. Using a

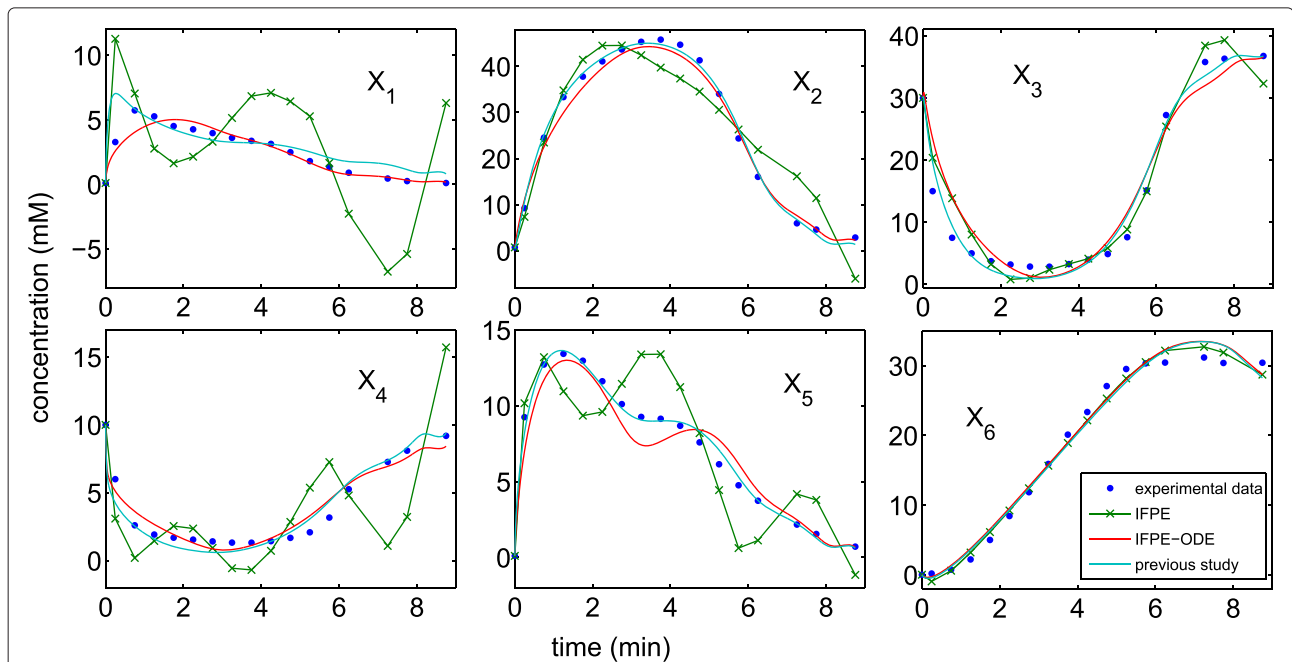


Figure 5 Model prediction of concentration data. For the IFPE without ODE integration, the concentration predictions were calculated from the integrated flux function at the given measurement time points using Eq. (2). For the IFPE-ODE, the concentration predictions were generated by integrating the ODE model. The concentration predictions of the previous study were generated by integrating the ODE model using the *Isqcurvefit* parameters in Table one of [14].

deterministic optimization algorithm, a previous parameter estimation showed that the convergence to the optimal parameters depended strongly on prior information of the parameter values, which were used as initial parameter guess [14]. In particular, the only successful initial guess came from the parameter estimation for the GMA model of the same pathway. Here, the IPE method did not converge within a preset maximum time-limit of 24 hours. On the other hand, the two variants of the IFPE provided parameter estimates within the time-limit, which are summarized in Table 4. The IFPE method without integrating the ODE model was expectedly the quickest between the two IFPE variants, but the resulting fit to the data was rather poor, as shown in Figure 5. The slower IFPE-ODE method could produce parameter estimates that fit the time-series metabolite data as well as those from the previous study [14]. In contrast to the previous parameter estimation, the IFPE-ODE however does not require any prior information on parameter values, aside from the parameter bounds.

Discussion

The estimation of kinetic parameters of ODE models represents an active research area in systems biology. The main challenges in this topic involve the ill-posedness of the estimation problem such that many indistinguishable solutions exist [4], as well as the high computational cost in solving the associated multidimensional global optimization problem. In this work, we present a new parameter estimation method, called integrated flux parameter estimation, based on an incremental approach using integrated reaction fluxes. In the IFPE, we formulate the parameter estimation as a nested optimization problem, in which we decouple the parameter estimation into outer and inner optimization over independent and dependent reaction subspaces, respectively. In comparison to standard estimation strategies, the outer and inner optimizations of the IFPE method involve smaller parameter dimension, translating to faster convergence and computational time. As demonstrated in the case studies, the nested optimization strategy could offer a significant improvement in computational speed over the standard simultaneous estimation.

The IFPE offers a couple advantages over a previous method, the IPE, which uses a similar nested optimization.

One weakness of the IPE method is that the parameter accuracy sensitively depends on the data smoothing procedure, as demonstrated in Table 1. Unfortunately, the optimal data smoothing setting for a given dataset, one that leads to the most accurate parameter estimates, is usually not known. In contrast, the IFPE does not require any time-series data smoothing and differentiation, and is therefore not affected by the issue above. In addition, as the IFs are directly estimated from time-series data, the IFPE can provide much lower parameter bias than the IPE. The advantages of the IFPE over the IPE come at the cost of increased computational time due to the numerical integration of reaction flux equations. We note that in the case studies, the increase in the computational cost was reasonably low, where the IFPE methods were typically 1.5 to 2 times slower than the IPE method with ODE integrations (i.e. IPE-ODE). In some cases, such as in the parameter estimation of lin-log models, the IFPE however could offer better computational performance than the IPE by taking advantage of the structure of the reaction flux functions.

In the first example (the branched pathway model), we tested the performance of the two variants of IFPE using a GMA model with *in silico* noise-free and noisy data. Here, we used the same model equations in the data generation and the parameter estimation. Thus, this example represented an idealized parameter estimation case study, where the only unknown information in the model was the parameter values. The parameter estimates from noise-free data indicated that both IFPE variants could produce nearly unbiased estimates with median parameter errors of less than 1%. The high parameter accuracy using noise-free data suggested that the parameters are *a priori* identifiable. Noise in data expectedly led to less accurate parameter estimates, not only for the IFPE methods, but also for the other estimation methods. In the second example, we applied the IFPE to another popular class of metabolic network models, namely lin-log kinetic model. The dataset in the second case study was less dense than that in the first example. In addition, the resulting data fitting also appeared worse than that in the first example. However, the mismatch between model prediction and concentration data depends not only on the accuracy of the parameter values, but also on how well model equations approximate the metabolic reactions. In this regard, the lin-log model in Eqs. 14-15 has difficulty in describing the transient behavior of *L. lactis* metabolism. This issue is not surprising as the lin-log kinetics are derived from thermodynamics concepts and are in principle valid only around steady state (or reference state) [20]. As shown in Figure 5, the IFPE method could fit the concentration data as well as the simultaneous estimation approach using an optimized initial parameter guess.

Table 4 Performance comparison for the lin-log modeling of *L. lactis* glycolytic pathway

	CPU Time ^a (sec)	Φ	eSS Iterations
IFPE	152.7	3.362	73
IFPE-ODE	3354	1.723	133

^aThe CPU time was recorded using a workstation with Intel Xeon processor 3.33 GHz with 18 GB RAM.

The computational cost of the IFPE is a product of the number of iterations in the outer optimization and the cost of each iteration. In general, the computational complexity of finding global optimal solution(s) is expected to increase exponentially with the dimension of the search space [22]. Assuming that each flux function has roughly a fixed number of unknown parameters, we thus expect that the complexity of the outer optimization of the IFPE scales exponentially with the number of independent reactions. Meanwhile, the cost per iteration is associated with the integration of reaction flux functions, the inner optimization, and in the case of IFPE-ODE, the integration of ODE model. When kinetic parameters are not shared among reaction rate equations, the inner optimization could be performed one-reaction-at-a-time and thus its complexity should increase linearly with the number of dependent fluxes. Meanwhile, the ODE integrations can slow down the parameter estimation significantly, especially when the ODE model is stiff.

Here, we have used the eSS global optimization algorithm, which is a population-based metaheuristic method combining scatter search and local deterministic optimizations. In comparison to deterministic global optimization algorithms, metaheuristic methods have better computational scalability. But, these methods lack rigorous guarantee in convergence to the global optimal solution, and repeated runs of the algorithm may not necessarily converge to the same solution. In the eSS algorithm, we track a population of parameter solutions, which is updated at every iterations. As the recommended size of the population increases linearly with the dimension of the search space, the computational cost per iteration should also increase linearly with the dimension of p_I . Unfortunately, it is difficult to predict how the convergence rate of eSS would scale with the dimension of p_I . This is because the convergence rate of eSS or any numerical optimization algorithm depends not only on the problem size but also on the topology of the search space (J. Banga, private communication).

Conclusions

The estimation of ODE model parameters from time-series data is often the bottleneck step in biological system modeling. In this work, we develop a reliable and efficient parameter estimation method, called integrated flux parameter estimation, for ODE models having more reactions than (measured) species. Such ODE models are common in biology as cellular species often participate in more than one reaction and concentration measurements are available only for a fraction of the species. The IFPE method relies on the integral form of the ODE model, based on which one can compute the

integrated fluxes directly from time-series concentration data. Here, the reactions are partitioned into independent and dependent subsets such that the dependent IFs can be uniquely determined from the independent IFs. We formulate the parameter estimation as a nested optimization problem, where the outer optimization involves the minimization of model prediction errors over the independent parameters (i.e. parameters appearing in the independent reaction rates), and the inner optimization involves the minimization of IF prediction errors over the dependent parameters (i.e. parameters appearing in the dependent reaction rates). Using two case studies comprising a GMA model of branched metabolic pathway and a lin-log model of *L. lactis* glycolytic pathway, we show that the IFPE can produce parameter estimates with a low bias and in much faster computational times than standard parameter estimation method based on simultaneous approach. In comparison to a previously published nested estimation method, the IFPE does not require any smoothing and differentiation of noisy time-series data. These advantages come at the cost of a small increase in the computational times. The IFPE and IPE are available through a MATLAB user interface called REDEMPTION (Reduced Dimension Ensemble Modeling and Parameter Estimation) at <http://www.cabsel.ethz.ch/tools/redemption>, or upon request.

Additional files

Additional file 1: Supplementary to parameter estimation of dynamic biological network models using integrated fluxes. The additional file contains the flowchart of integrated flux parameter estimation (IFPE) methods with unmeasured concentrations, the modified Simpson's rule used in the calculation of IFs, the parameters and initial conditions used to generate time-series data in the branched pathway case study, the smoothed time-series concentration figure using different piecewise spline-fitting parameters, the comparison of concentration errors Φ , the parameter estimation results using noise-free data without the concentration measurements of X_3 , and the IFPE performance with different selections of independent fluxes.

Additional file 2: Time-series Concentration Data for the Branched Pathway Case Study. The additional file contains *in silico* noise-free and noisy dataset used in the case study of branched metabolic pathway.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

YL and RG conceived the IFPE algorithms. YL performed parameter estimation in two case studies. RG conceived and guided the whole study. Both authors wrote, read, edited and approved the final manuscript.

Acknowledgements

The authors would like to thank Prof. Julio Banga for providing an implementation of eSS with a convergence criterion.

Received: 7 August 2014 Accepted: 29 October 2014

Published online: 18 November 2014

References

1. Chou IC, Voit EO: **Recent developments in parameter estimation and structure identification of biochemical and genomic systems.** *Math Biosci* 2009, **219**(2):57–83.
2. Bardow A, Marquardt W: **Incremental and simultaneous identification of reaction kinetics: methods and comparison.** *Chem Eng Sci* 2004, **59**(13):2673–2684.
3. Srinath S, Gunawan R: **Parameter identifiability of power-law biochemical system models.** *J Biotechnol* 2010, **149**(3):132–140.
4. Szederkenyi G, Banga J, Alonso A: **Inference of complex biological networks: distinguishability issues and optimization-based solutions.** *BMC Syst Biol* 2011, **5**(1):177.
5. Voit EO, Almeida J, Marino S, Lall R, Goel G, Neves AR, Santos H: **Regulation of glycolysis in lactococcus lactis: an unfinished systems biological case study.** *Syst Biol* 2006, **153**(4):286–298.
6. Voit EO, Almeida J: **Decoupling dynamical systems for pathway identification from metabolic profiles.** *Bioinformatics* 2004, **20**(11):1670–1681.
7. Goel G, Chou IC, Voit EO: **System estimation from metabolic time-series data.** *Bioinformatics* 2008, **24**(21):2505–2511.
8. Jia G, Stephanopoulos GN, Gunawan R: **Parameter estimation of kinetic models from metabolic profiles: two-phase dynamic decoupling method.** *Bioinformatics* 2011, **27**(14):1964–1970.
9. Nim TH, Luo L, Clément M-V, White JK, Tucker-Kellogg L: **Systematic parameter estimation in data-rich environments for cell signalling dynamics.** *Bioinformatics* 2013, **29**(8):1044–1051.
10. Jia G, Stephanopoulos G, Gunawan R: **Incremental parameter estimation of kinetic metabolic network models.** *BMC Syst Biol* 2012, **6**:142.
11. Amrhein M, Bhatt N, Srinivasan B, Bonvin D: **Extents of reaction and flow for homogeneous reaction systems with inlet and outlet streams.** *AIChE J* 2010, **56**(11):2873–2886.
12. Bhatt N, Amrhein M, Bonvin D: **Incremental identification of reaction and mass-transfer kinetics using the concept of extents.** *Ind Eng Chem Res* 2011, **50**(23):12960–12974.
13. Bhatt N, Kerimoglu N, Amrhein M, Marquardt W, Bonvin D: **Incremental identification of reaction systems—a comparison between rate-based and extent-based approaches.** *Chem Eng Sci* 2012, **83**:24–38.
14. del Rosario RC, Mendoza E, Voit EO: **Challenges in lin-log modelling of glycolysis in lactococcus lactis.** *IET Syst Biol* 2008, **2**(3):136–149.
15. Hindmarsh AC, Brown PN, Grant KE, Lee SL, Serban R, Shumaker DE, Woodward CS: **Sundials: Suite of nonlinear and differential/algebraic equation solvers.** *ACM Trans Math Software* 2005, **31**(3):363–396.
16. Egea JA, Rodríguez-Fernández M, Banga JR, Martí R: **Scatter search for chemical and bio-process optimization.** *J Global Optim* 2007, **37**(3):481–503.
17. Rodríguez-Fernández M, Egea JA, Banga JR: **Novel metaheuristic for parameter estimation in nonlinear dynamic biological systems.** *BMC Bioinformatics* 2006, **7**:483.
18. Egea JA, Martí R, Banga JR: **An evolutionary method for complex-process optimization.** *Comput Oper Res* 2010, **37**(2):315–324.
19. Leander J, Lundh T, Jirstrand M: **Stochastic differential equations as a tool to regularize the parameter estimation problem for continuous time dynamical systems given discrete time measurements.** *Math Biosci* 2014, **251**(0):54–62.
20. Visser D, Heijnen JJ: **Dynamic simulation and metabolic re-design of a branched pathway using linlog kinetics.** *Metab Eng* 2003, **5**(3):164–176.
21. Neves AR, Ramos A, Nunes MC, Kleerebezem M, Hugenholtz J, de Vos WM, Almeida J, Santos H: **In vivo nuclear magnetic resonance studies of glycolytic kinetics in lactococcus lactis.** *Biotechnol Bioeng* 1999, **64**(2):200–212.
22. Pintér JD: **Global optimization: software, test problems, and applications.** In *Handbook of Global Optimization. Volume 2*. Edited by Pardalos PM, Romeijn HE. Dordrecht: Kluwer Academic Publishers; 2002:515–569.

doi:10.1186/s12918-014-0127-x

Cite this article as: Liu and Gunawan: Parameter estimation of dynamic biological network models using integrated fluxes. *BMC Systems Biology* 2014 **8**:127.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

