Eur Biophys J (2010) 39:1019–1039
DOI 10.1007/s00249-009-0520-3

ORIGINAL PAPER

Calibration of dynamic models of biological systems with KInfer

Paola Lecca · Alida Palmisano · Adaoha Ihekwaba · Corrado Priami

Received: 16 February 2009 / Revised: 2 July 2009 / Accepted: 5 July 2009 / Published online: 11 August 2009

Abstract Methods for parameter estimation that are robust to experimental uncertainties and to stochastic and biological noise and that require a minimum of a priori input knowledge are of key importance in computational systems biology. The new method presented in this paper aims to ensure an inference model that deduces the rate constants of a system of biochemical reactions from experimentally measured time courses of reactants. This new method was applied to some challenging parameter estimation problems of nonlinear dynamic biological systems and was tested both on synthetic and real data. The synthetic case studies are the 12-state model of the SERCA pump and a model of a genetic network containing feedback loops of interaction between regulator and effector genes. The real case studies consist of a model of the reaction between the inhibitor κB kinase enzyme and its substrate in the signal transduction pathway of NF-κB, and a stiff model of the fermentation pathway of *Lactococcus lactis*.

Keywords Parameter estimation · Biochemical networks · Dynamical models · Systems biology

Proceedings of the XIX Congress of the Italian Society of Pure and Applied Biophysics (SIBPA), Rome, September 2008.

P. Lecca (✉) · A. Palmisano · A. Ihekwaba · C. Priami
The Microsoft Research—University of Trento,
Centre for Computational and Systems Biology,
Piazza Manci 17, 38123 Povo, Trento, Italy
e-mail: lecca@cosbi.eu

A. Palmisano
e-mail: palmisano@cosbi.eu

A. Ihekwaba
e-mail: ihekwaba@cosbi.eu

C. Priami
e-mail: priami@cosbi.eu

A. Palmisano · C. Priami
DISI, University of Trento, via Sommarive 18,
38100 Povo, Trento, Italy

Introduction

Mathematical modeling and dynamic simulation of biochemical networks are central in systems biology, as they provide new methods for the analysis of omics data and lead to a greater understanding of the language of cells and organisms (Cho et al. 2003). Models and simulations are systematic strategies for key issues in medicine and the pharmaceutical and biotechnological industries. For example, model-based approaches and in silico experiments on a computer provide a rational framework to guide drug development, taking into account the effects of possible new drugs on biochemical pathways and physiology. Once a mathematical model of even a small part of a biochemical network is established, then the potential benefits are noticeable: generating new hypotheses, suggesting experiments to test them, and, more generally, supporting experimental design.

The construction of a mathematical model of a network consists of two tasks: (1) deciding on the model structure and (2) estimating the involved parameter values. This work is focused on the key step of parameter estimation, assuming the structure of the model as given. Parameter estimation (also known as model calibration) from experimental data is a bottleneck for a major breakthrough in computational systems biology in the present post-genomic era (Cho et al. 1998). Parameter inference aims to find the

parameters of the model that give the best fit to a set of experimental data.

Estimating the parameters of a dynamic model of a biochemical network is difficult, because often the model is nonlinear, and thus no general analytic result exists. 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 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. When estimating parameters of dynamical systems, a number of difficulties may arise, 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 nondifferentiable terms in the systems dynamics. Due to the nonlinear and constrained nature of the systems dynamics, these problems are very often multimodal. Thus, traditional gradient-based methods, such as 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.

The recent literature reports many examples of new effective methods for parameter estimation both in deterministic and stochastic models. Here we briefly mention the most recent ones and refer the reader to an extensive review of the state of the art reported by Chou and Voit (2009). Goel et al. (2008) and Voit et al. (2006) discussed the problem of lack of convergence of search algorithms on generalized mass action law and proposed a novel methodological framework called dynamic flux estimation for estimating parameters for models of metabolic systems from time-series data. Polisetty and Voit (2006) suggested global optimization techniques as an alternative to traditional local methods. Rodrigez-Fernandez et al. (2006) developed a hybrid stochastic-deterministic global optimization method. Moles et al. (2003) explored several state-of-the-art deterministic and stochastic global optimization techniques and compared their accuracy and effectiveness on nonlinear biochemical dynamic models. Tian et al. (2007) presented a simulated maximum likelihood method to evaluate parameters in stochastic models described by stochastic differential equations. They proposed different types of transitional probability and a genetic optimization algorithm to search for optimal reaction rates. Chou et al. (2006) developed an alternate regression method that dissects the parameter inference problem into iterative steps of linear regression. Sugimoto et al. (2005) developed a computational technique based on genetic programming that simultaneously generates biochemical equations and their parameters from time-series data. Reinker et al. (2006) proposed the approximate maximum likelihood method and the singular value decomposition likelihood method that estimate stochastic reaction constants from molecule count data measured with errors at discrete time points.

Tools for parameter fitting through regression or maximum likelihood methods can be found as an integral part of simulation tools (e.g., Hoops et al. 2006), but there exist also stand alones, such as PET (Zwolak et al. (2001) and BioBayes (Vyshemirsky and Girolami (2008). BioBayes was recently developed by Vyshemirsky and Girolami (2008a, b); it is a software package that provides a framework for Bayesian parameter estimation and model ranking over models of biochemical systems defined using ordinary differential equations. Finally, Boys et al. (2008), Golightly and Wilkinson (2008), and Wilkinson (2006, 2007) developed Bayesian model-based inference techniques for discrete models. Bayesian schemes offer some advantages over the maximum likelihood methods when the volume of data is limited or the analytic form of the kinetic model makes the maximization of the likelihood difficult. Nevertheless, Bayesian approaches require specifying a prior distribution for all unknown parameters, but in problems of calibration of biochemical models, prior knowledge is either vague or nonexistent and that makes it very difficult to specify a unique prior distribution.

We note that most of the current tools for parameter estimation lack robustness to noise and lack any estimate of experimental error in their outcome. Experimental uncertainties on parameters propagate from the measurements of the concentrations of the species. Inferring the parameters with an estimate of their uncertainty is essential if we want to use these tools in the context of optimal experimental design. Furthermore, most of the current tools based on optimization techniques do not univocally find the global optimal solution, and ask the user to provide a priori the optimization algorithm with the region of parameter space in which to perform the search for the global maximum/minimum.

In this paper we present a new approach to parameter estimation, whose accuracy is robust to the experimental noise. The method is based on a probabilistic, generative model of the variations in reactant concentrations. We observe time series of concentrations for all the reactant species, gathered in N state vectors $\mathbf{X}_1,\ldots,\mathbf{X}_N$. Our method approximates the rate of change of the reactant concentration by finite differences and provides a tool to predict the values of the variables X_i at time t, conditioned on their values at the previous time point. The variations in the concentration of the species at different time points are conditionally independent owing to the Markov nature of this approximated model of the rate equation. Assuming

the observation noise to be Gaussian with variance σ^2, the probability of observing a variation D_i for the concentration $[X]_i$ of species i between time t_{k-1} and t_k is a Gaussian with variance depending on σ and with mean the expectation value of the law of mass action under the noise distribution. The likelihood of the observed increments/decrements D_i can be optimized with respect to the kinetic rate constants of the biochemical network under consideration and with regard to the level of noise σ affecting the time course of the reactant concentration.

The approximation of the time derivative of reactant concentration by finite differences provides a model for the variations in the species concentration rather than a model for the time trajectory of the species concentrations. This makes the evaluation of the expectation value of the law mass action function (the integral of the transitional probability) simpler and analytically tractable. The rate coefficients and the level of noise are then obtained by maximizing the likelihood function defined by the observed variations. Our method infers the rate coefficients, the level of noise σ, and an error range on the estimates of rate constants. Its probabilistic formulation handles the noise in the data due to experimental uncertainties, and it enables further extensions, such as a fully Bayesian treatment of the parameter inference and automated model selection strategies based on the comparison between marginal likelihoods of different models. Finally, the implementation of this method may be used as an interface tool, connecting the outcomes of the wet-lab activity for the concentration measurements and any software for the simulation of chemical kinetics.

The paper proceeds as follows: The next section presents the mathematical model of parameter inference. The third section describes the software tool KInfer (*K*nowledge *I*nference) that implements the inference procedure presented in the section "The model for inference." KInfer (2008, 2007, 2009) is a freeware software available at http://www.cosbi.eu/index.php/research/prototypes/kinfer. Sections "Synthetic case studies" and "Real case studies" illustrate the results of the application of the KInfer parameter estimation method to challenging simulated and real case studies, respectively. The artificial case studies are the Maclennan-Higgins SERCA pump model (MacLennan et al. (1997) and a small-scale genetic network with feedback loops (Kikuchi et al. 2003). The real case studies are the binding kinetics of the enzyme inhibitor kappa B kinase to its substrate inhibitor kappa B alpha, whose interaction is an integral part of the transduction of signals in the NF-kappa B signalling pathway (Ihekwaba et al. 2004, 2005, 2007; Nelson et al. 2004), and the fermentation pathways of *Lactococcus lactis* (Goel et al. 2008; Voit et al. 2006). Finally, the paper ends with some conclusive remarks and a brief description of the future directions of this study.

The model for inference

Consider N reactant species, S_1, S_2,…,S_N, with concentrations X_1, X_2,…,X_N, that evolve according to a system of rate equations

$$\frac{\mathrm{dX_i}}{\mathrm{d}t} = f_i(\mathbf{X}^{(i)}(t); \theta_i) \tag{1}$$

where $\boldsymbol{\theta}_i$, $i = 1, 2, \ldots, N$, is the vector of the rate coefficients, which are present in the expression of the function f_i. We wish to estimate the set of parameters $\Theta = \cup\boldsymbol{\theta}_i$ (i = 1, 2,…, N), whose element $\boldsymbol{\theta}_i$ is the set of rate coefficients appearing in the rate equations of ith species, therefore

$$\boldsymbol{\theta}_1 = \{\theta_{11}, \theta_{12}, \ldots, \theta_{1N_1}\}, \ldots, \boldsymbol{\theta}_N = \{\theta_{N1}, \theta_{N2}, \ldots, \theta_{NN_N}\}$$

$\mathbf{X}^{(i)}$ is the vector of concentrations of chemicals that are present in the expression of the function f_i for the species i. According to the law of mass action, the functions f_i have the general form

$$\begin{aligned} & f_i(\mathbf{X}^{(i)}(t); \boldsymbol{\theta}_i) \\ & \quad = \theta_{i1} \prod_{w \in S_1 \subseteq [1,N]} X_w^{\alpha_w} + \cdots + \theta_{iN_i} \prod_{w \in S_{N_i} \subseteq [1,N]} X_w^{\alpha_w} \\ & \quad = \sum_{h=1}^{N_i} \left(\theta_{ih} \prod_{w \in S_h} X_w^{\alpha_w} \right) \end{aligned} \tag{2}$$

where $\alpha_w \in \mathbf{R}$ and N_i is the number of parameters in the f_i rate equation. The rate equations in Eq. 2 form the so-called generalized mass action law (Marin-Sanguino et al. 2007).

We assume we have noisy observations $\hat{X}_i = X_i + \epsilon$ at times t_0,…, t_M, where $\epsilon \sim \mathcal{N}(0, \sigma^2)$ is a Gaussian noise term with mean zero and variance σ. With this choice, we are assuming that the concentration measurements are not significantly affected by systematic errors but by uncontrolled random errors and that an error is equally likely to occur in either the positive or negative direction with respect to the symmetry axis of the distribution.

We also assume a number M of concentration measurements for each considered species. Approximating the rate equation (Eq. 1) as a finite difference equation between the observation times gives

$$X_i(t_k) = X_i(t_{k-1}) + (t_k - t_{k-1}) f_i(\mathbf{X}^{(i)}(t_{k-1}); \boldsymbol{\theta}_i) \tag{3}$$

where k = 1,…, M. In Eq. 3, the rate equation is viewed as a model of increments/decrements of reactant concentrations, i.e., given a value of the variables at time $t_{k\text{-}1}$, the model can be used to predict the value at the next time point t_k. Increments/decrements between different time points are conditionally independent owing to the Markov nature of the model (3). Therefore, given the Gaussian model for the noise, it is possible to estimate the probability to observe the

value $\hat{X}_i(t_k)$ given the model at time t_{k-1}, $X_i(t_{k-1})$, and the set of parameters $\boldsymbol{\theta}_i$, as

$$p\left(\hat{X}_i(t_{k-1})|X_i(t_{k-1})\right) = \mathcal{N}\left(X_i(t_{k-1}) + (t_k - t_{k-1})f_i(X_i(t_{k-1}, \boldsymbol{\theta}_i)), \sigma^2\right) \quad (4)$$

We then also have that the true value of $X_i(t_k)$ is normally distributed around the observed value $\hat{X}_i(t_k)$, so that

$$p\left(X_i(t_{k-1})|\hat{X}_i(t_{k-1}))\right) = \mathcal{N}\left(\hat{X}_i(t_{k-1}), \sigma^2\right) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(X_i(t_{k-1}) - \hat{X}_i(t_{k-1}))^2}{2\sigma^2}\right] \quad (5)$$

Therefore, the probability of observing a variation $D_i(t_k) = X_i(t_k) - X_i(t_{k-1})$ for the concentration of the ith species between the time t_{k-1} and t_k, given the parameter vector $\boldsymbol{\theta}_i$ is

$$p(D_i(t_k)|\boldsymbol{\theta}_i, \sigma) = \mathcal{N}\left(E\left[f_i(\mathbf{X}^{(i)}(t_{k-1}), \boldsymbol{\theta}_i)\right], 2\sigma^2\right) \quad (6)$$

and

$$E\left[f_i(\mathbf{X}^{(i)}(t_{k-1}, \boldsymbol{\theta}_i))\right] = \int_{\Omega_{\mathbf{X}^{(i)}}} f_i(\mathbf{X}^{(i)}(t_{k-1}), \boldsymbol{\theta}_i) \prod_{i=1}^{K_i}\left[p_i\left(X_i(t_{k-1})|\hat{X}_i(t_{k-1})\right)\right]\mathrm{d}\mathbf{X}^{(i)} \quad (7)$$

where $\Omega_{\mathbf{X}^{(i)}}$ is the sample space of $\mathbf{X}^{(i)}$, and K_i is the number of chemical species in the expression for f_i. While the increments/decrements are conditionally independent given the starting point $X_i(t_k)$, the random variables $D_i(t_k)$ are not independent of each other. Intuitively, if $X_i(t_k)$ happens to be below its expected value because of random fluctuations, then the following increment $D_i(t_{k+1})$ can be expected to be bigger as a result, while the previous one $D_i(t_k)$ will be smaller. A simple calculation allows us to obtain the covariance matrix of the vector of increments for the ith species. This is a banded matrix $\mathbf{C}_i \equiv \mathbf{C} = \mathrm{Cov}\,(\mathbf{D}_i)$ with diagonal elements given by

$$E\left[D_i^2(t_k) - E[D_i^2(t_k)]\right] = 2\sigma^2$$

and a non-zero band above and below the diagonal given by

$$E[(D_i(t_k) - E[D_i(t_k)])(D_i(t_{k-1}) - E[D_i(t_{k-1})])] = -\sigma^2$$

with all other entries zero. The likelihood for the observed increments/decrements therefore will be

$$p(\mathbf{D}|\Theta) = \prod_{i=1}^{N}\mathcal{N}(\mathbf{D}_i|\mathbf{m}_i(\Theta), \mathbf{C}) = \left(\frac{1}{\sqrt{2\pi\det(\mathbf{C})}}\right)^N e^{\sum_{i=1}^{N} -\frac{1}{2}(\mathbf{D}_i - \mathbf{m}_i)^T\mathbf{C}^{-1}(\mathbf{D}_i - \mathbf{m}_i)} \quad (8)$$

where $\mathbf{D} = \{\mathbf{D}_1, \ldots, \mathbf{D}_N\}$, $\mathbf{D}_i = D_i(t_1), D_i(t_2), \ldots D_i(t_M)$ $(i = 1, 2, \ldots, N)$, and $\mathbf{m}_i(t_{k-1}) \equiv E[f_i(\mathbf{X}(t_{k-1}), \boldsymbol{\theta}_i)]$.

Equation 8 can be optimized with respect to the parameters $\Theta = (\theta_1, \theta_2, \ldots, \theta_N)$ of the model to yield estimates of the parameters themselves and of the noise level. The chief numerical problem of this approach is the computation of the expectations of the rate functions given by Eq. 7. Noninteger values of the coefficient α can make estimating the integral analytically difficult. We propose an approximate method in which the Gaussian noise is replaced by an approximate uniform (white) noise, with the amplitude of the uniform noise being obtained as a sample from the Gaussian cumulative distribution function. At the first order, for small σ, we can approximate the Gaussian with zero mean and variance σ with a uniform distribution defined on the interval $[-\frac{\sqrt{2\pi}\sigma}{4}, \frac{\sqrt{2\pi}\sigma}{4}]$, so that

$$\prod_{i=1}^{K_i} p_i = \prod_{i=1}^{K_i}\chi_i \quad (9)$$

where

$$\chi_i(X_i) = \begin{cases} \frac{2}{\sqrt{2\pi}\sigma} & \text{if } -\frac{\sqrt{2\pi}\sigma}{4} \le X_i \le \frac{\sqrt{2\pi}\sigma}{4} \\ 0 & \text{otherwise.} \end{cases}$$

This approximation makes the calculation of the expectation value of the rate equation (Eq. 7) simpler and reduces the computational time of the procedure. Moreover, experiments not illustrated in this paper demonstrate that it does not influence the accuracy of the parameter estimates until σ is less than 30% of the concentration measurement.

Substituting Eq. 9 in Eq. 7 gives

$$E[f_i(\mathbf{X}^{(i)}(t_{k-1}), \theta)] = \left(\frac{2}{\sqrt{2\pi}\sigma}\right)^{K_i}\int_{\hat{X}-\frac{\sqrt{2\pi}\sigma}{4}}^{\hat{X}+\frac{\sqrt{2\pi}\sigma}{4}} f_i(\mathbf{X}^{(i)}(t_{k-1}), \theta_i)\mathrm{d}\mathbf{X}^{(i)} \quad (10)$$

Now, substituting Eq. 2 in Eq. 10 leads to

$$E[f_i(\mathbf{X}^{(i)}(t_{k-1}), \theta_i)] = \left(\frac{2}{\sqrt{2\pi}\sigma}\right)^{K_i}\left\{\sum_{h=1}^{N_i}\theta_{ih}\left[\left(\frac{\sqrt{2\pi}\sigma}{2}\right)^{\#(S-S_h)} \times \prod_{w\in S_h}\frac{1}{\alpha_w + 1}\left(\left(\hat{X}_w + \frac{\sqrt{2\pi}\sigma}{4}\right)^{\alpha_w+1} - \left(\hat{X}_w - \frac{\sqrt{2\pi}\sigma}{4}\right)^{\alpha_w+1}\right)\right]\right\} \quad (11)$$

where S is the set containing the indexes referring to all the K_i species appearing in f_i, and $\alpha_w \neq -1$. In case some orders are equal to -1, Eq. 11 takes the following form

$$E[f_i(\mathbf{X}^{(i)}(t_{k-1}),\theta_i)] = \left(\frac{2}{\sqrt{2\pi}\sigma}\right)^{K_i}\sum_{h=1}^{N_i}\theta_{ih}\left\{\left(\frac{\sqrt{2\pi}\sigma}{2}\right)^{\#(S-S_h)} \times\left[\prod_{w\in S_h'}\frac{1}{\alpha_w+1}\left(\left(\hat{X}_w+\frac{\sqrt{2\pi}\sigma}{4}\right)^{\alpha_w+1} -\left(\hat{X}_w-\frac{\sqrt{2\pi}\sigma}{4}\right)^{\alpha_w+1}\right)\right]\cdot\left[\prod_{w\in S_h''}\ln\frac{\hat{X}_w+\frac{\sqrt{2\pi}\sigma}{4}}{\hat{X}_w-\frac{\sqrt{2\pi}\sigma}{4}}\right]\right\} \quad (12)$$

where S_h' is the set of indexes $\{h_1', h_2',\ldots, h_s'\}$ such that $\alpha_{h'} \neq -1 \forall h' \in S_h'$, and S_h'' is the set of indexes $\{h_1'', h_2'',\ldots, h_s''\}$ such that $\alpha_{h''} = -1 \forall h'' \in S_h''$.

If in Eq. 8, $\mathbf{m}_i$ is substituted with the expression Eq. 11 or 12, Eq. 8 becomes more tractable and can be optimized with respect to the parameters $\Theta = (\theta_1, \theta_2, \ldots, \theta_N)$ and σ. The values of the model's parameters for which $p(\mathbf{D}|\Theta)$ has a maximum are the most likely values giving the observed kinetics.

Variance of the estimated parameters

To seek the parameter matrix Θ that maximizes the function in Eq. 8 is equivalent to seeking the parameter matrix Θ that maximizes the log-likelihood function given by

$$\ln p(\mathbf{D}|\Theta) = -\frac{N}{2}(\ln(2\pi)\ln(\det(\mathbf{C}))) - \frac{1}{2}\sum_{i=1}^{N}\left((\mathbf{D}_i-\mathbf{m}_i)^T\mathbf{C}^{-1}(\mathbf{D}_i-\mathbf{m}_i)\right) \quad (13)$$

Maximizing the log-likelihood function amounts to minimizing the last term of Eq. 13 since the other terms do not depend on Θ. The estimation problem is therefore reformulated as follows:

$$\Theta^{\mathrm{MLE}} = \arg\min_{\Theta}\sum_{i=1}^{N}\left((\mathbf{D}_i-\mathbf{m}_i)^T\mathbf{C}^{-1}(\mathbf{D}_i-\mathbf{m}_i)\right) \quad (14)$$

The maximum likelihood estimate Θ^{MLE} has the following appealing asymptotic properties: it is asymptotically unbiased (i.e., $E(\Theta^{\mathrm{MLE}}) = \Theta^*$, where Θ^* denotes the vector of the true values of Θ); consistent; asymptotically efficient; and asymptotically Gaussian (Casella and Berger 2002). The latter implies that the distribution of the Θ^{MLE} converges to a normal distribution with a covariance matrix given by the Cramér-Rao bound that is also the inverse of the Fisher information matrix

$$\mathbf{F}_{\Theta^{\mathrm{MLE}}} = \sum_{i=1}^{N}\mathbf{G}_i^T\mathbf{C}^{-1}\mathbf{G}_i \quad (15)$$

where $\mathbf{G}_i$ is called the *sensitivity matrix*.

All the matrices $\mathbf{G}_i$ can be obtained from the sensitivity matrix $\mathbf{S}(t)$ evaluated at the sampling instants. The sensitivity matrix is a $N \times P$ time-dependent matrix, where N is the number of species and p is the length of the parameter vector Θ. It is defined as follows

$$\mathbf{S}(t) = \frac{\partial \ln \mathbf{m}(t,\Theta)}{\partial \ln \Theta} = \left(\frac{1}{\mathbf{m}(t,\Theta)}\frac{\partial \mathbf{m}(t,\Theta)}{\partial \Theta}\right)\Big|_{\Theta=\Theta^{\mathrm{MLE}}} \quad (16)$$

The $\mathbf{G}_i$ matrices are obtained from $\mathbf{S}(t)$ as

$$\mathbf{G}_i = [s_i(t_1)^T,\ldots,s_i(t_M)^T]^T$$

where $s_i(t_k)$, (k = 1,…, M), is the ith row of $\mathbf{S}(t_k)$ (i = 1,…, N), where

$$\mathbf{S}(t_k) = \begin{pmatrix} \frac{\partial \ln m_1(t_k)}{\partial \ln \theta_{i1}} & \cdots & \frac{\partial \ln m_1(t_k)}{\partial \ln \theta_{iN_i}} & \cdots & \frac{\partial \ln m_1(t_k)}{\partial \ln \theta_{NP}} \\ \frac{\partial \ln m_2(t_k)}{\partial \ln \theta_{i1}} & \cdots & \frac{\partial \ln m_2(t_k)}{\partial \ln \theta_{iN_i}} & \cdots & \frac{\partial \ln m_2(t_k)}{\partial \ln \theta_{NP}} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial \ln m_N(t_k)}{\partial \ln \theta_{i1}} & \cdots & \frac{\partial \ln m_N(t_k)}{\partial \ln \theta_{iN_i}} & \cdots & \frac{\partial \ln m_N(t_k)}{\partial \ln \theta_{NP}}\theta_{NP} \end{pmatrix}$$

and thus

$$\mathbf{G}_i^T = \begin{pmatrix} \frac{\partial \ln m_i(t_1)}{\partial \ln \theta_{i1}} & \frac{\partial \ln m_i(t_2)}{\partial \ln \theta_{i1}} & \cdots & \frac{\partial \ln m_i(t_M)}{\partial \ln \theta_{i1}} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial \ln m_i(t_1)}{\partial \ln \theta_{iN_i}} & \frac{\partial \ln m_i(t_2)}{\partial \ln \theta_{iN_i}} & \cdots & \frac{\partial \ln m_i(t_M)}{\partial \ln \theta_{iNi}} \\ \frac{\partial \ln m_i(t_1)}{\partial \ln \theta_{i+1\,1}} & \frac{\partial \ln m_i(t_2)}{\partial \ln \theta_{i+1\,1}} & \cdots & \frac{\partial \ln m_i(t_M)}{\partial \ln \theta_{i+1\,1}} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial \ln m_i(t_1)}{\partial \ln \theta_{NP}} & \frac{\partial \ln m_i(t_2)}{\partial \ln \theta_{NP}} & \cdots & \frac{\partial \ln m_i(t_M)}{\partial \ln \theta_{NP}} \end{pmatrix}$$

An element $\gamma_{ab}^{(i)}(t)$ of the $\mathbf{G}_i$ (a = 1, 2,…, M and b = 1, 2,…, N), is

$$\gamma_{ab}^{(i)}(t) = \left(\frac{1}{\mathbf{m}_i}\frac{\partial \mathbf{m}_i}{\partial \theta_{ib}}\right)_{\Theta=\Theta^{\mathrm{MLE}}} = \left(\frac{1}{\mathbf{m}_i}\right)_{\Theta=\Theta^{\mathrm{MLE}}}\left(\frac{\sqrt{2\pi}\sigma}{2}\right)^{-\#(S_b)} \times\prod_{w\in S_b}\left[\left(\hat{X}_w(t)+\frac{\sqrt{2\pi}\sigma}{4}\right)^{\alpha_w+1} -\left(\hat{X}_w(t)-\frac{\sqrt{2\pi}\sigma}{4}\right)^{\alpha_w+1}\right]\frac{1}{\alpha_w+1} \quad (17)$$

if $\alpha_w \neq -1 \forall w \in S_h$. The square root of the pth diagonal element of $\mathbf{F}_{\Theta^{\mathrm{MLE}}}^{-1}$ gives an estimate of the standard deviation of the pth component of Θ (David and Bastin 1999; Lecca et al. 2009).

KInfer: a prototype for parameter inference

We developed the prototype KInfer (Knowledge Inference) that implements the procedure described in the previous section. KInfer consists of four main blocks: (1) the input interface, (2) the model generator, (3) the maximization

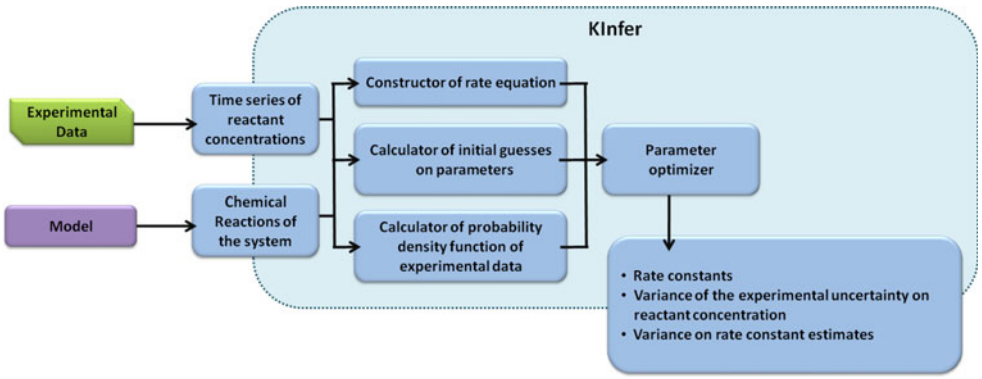

Fig. 1 Main modules of KInfer

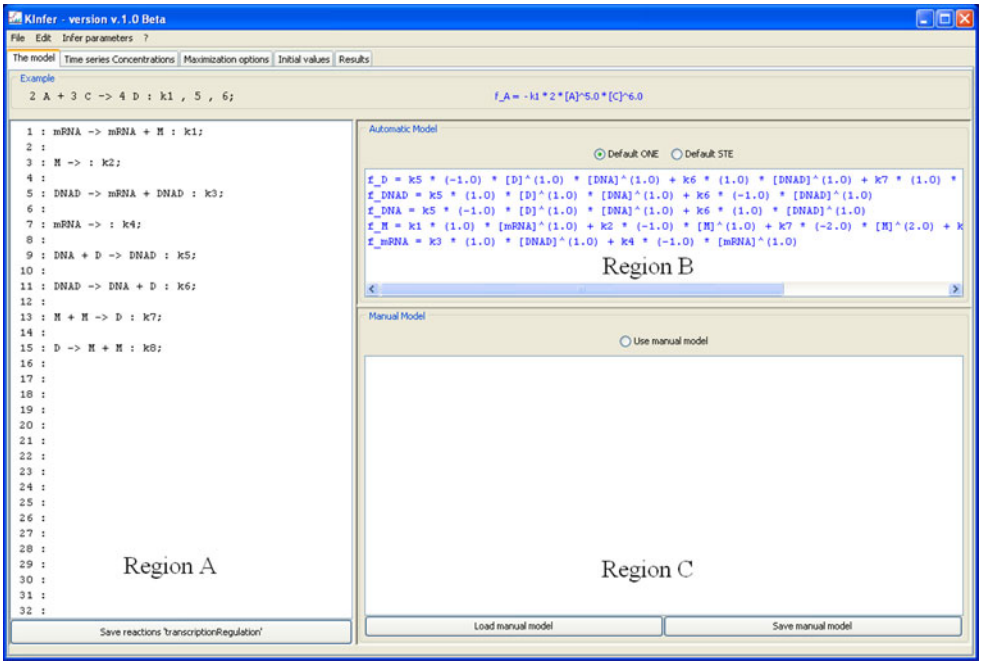

Fig. 2 Front-end of KInfer. It is divided into three regions: region A on the left, region B on the upper right, and region C on the bottom right. In region A, the user can write the reactions of the system. In region B, the rate equation model is automatically generated. The user is allowed to change the automatically generated equations in region C

algorithm, and (4) the output interface (Fig. 1). A view of the screenshots of the tool is shown in Figs. 2, 3, and 4.)

The syntax of the input reactions is

$$aA + bB + \cdots \rightarrow a'A' + b'B' + \cdots : \ k, \ \alpha, \ \beta, \ldots;$$

The reactants (A, B,…) and their stoichiometric coefficients (a, b,…) are indicated on the left-hand side of the arrow, whereas the products (A', B',…) and the product stoichiometric coefficients (a', b',…) are indicated on the right-hand side. The reaction specification also contains the indication of the name of the rate constant (k) and the partial orders of reaction (α, β,…). The reaction orders are supposed to be given and do not undergo any inference procedure. If the user does not specify any reaction order, the assigned default value is 1. Along with the specification of the set of reactions involved in the system, KInfer requires the experimental time-series data of the concentration (or number of molecules) of the species in the system. From the set of chemical reactions the tool automatically generates the ordinary differential equations model, consisting of a system of equations in the form of Eq. 2 and displays them in region B of the front-end (see Fig. 2). In region C of the front-end, the user is also allowed to insert an ordinary differential equation model that is not in the form of the generalized mass action law (e.g., Hill functions, Michaelis-Menten kinetics). In this case inserting the list of chemical reactions is superfluous, as the tool will ignore them.

KInfer derives the form of the probability density function in Eq. 8 to maximize the initial guesses for the parameters from the data set of the time series of concentration and from the model of rate equation. Although

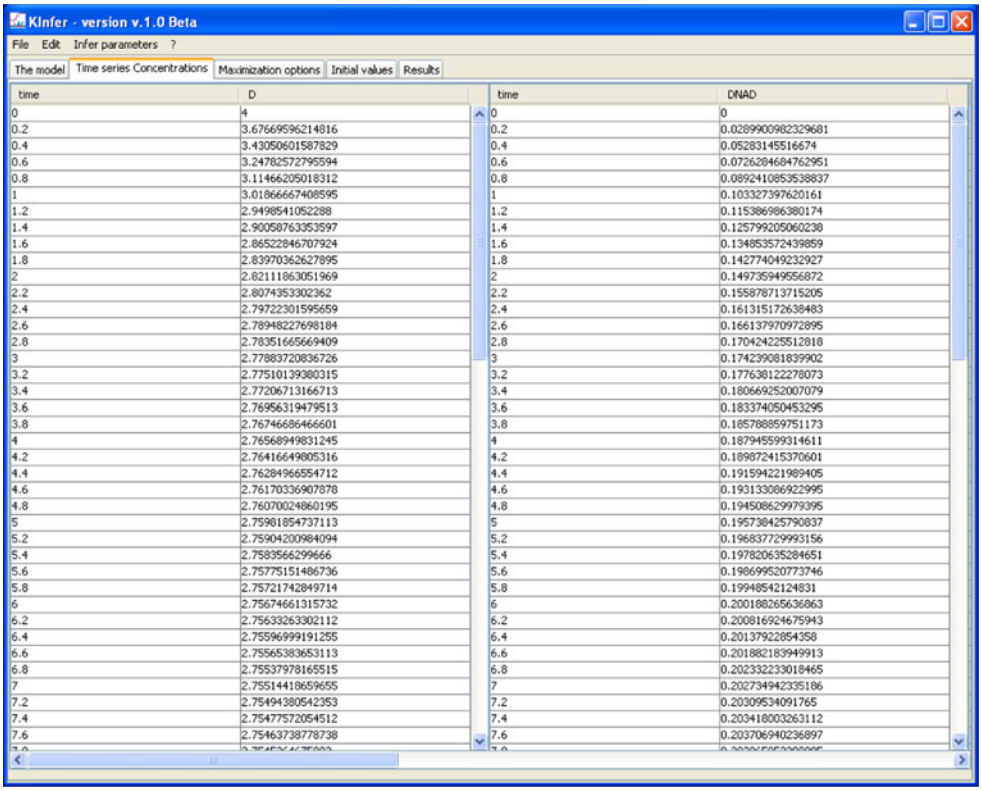

Fig. 3 The table of uploaded experimental data

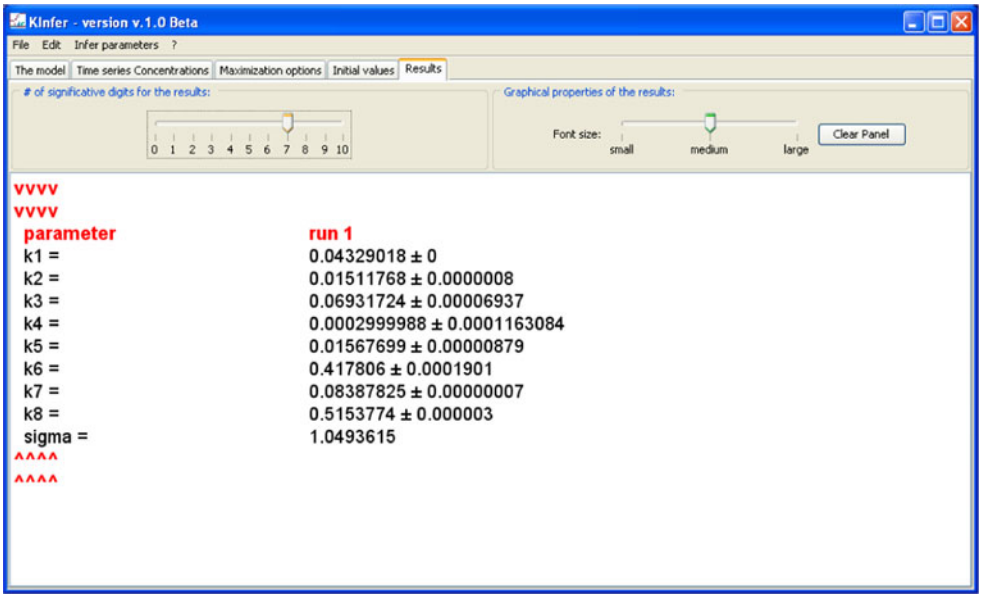

Fig. 4 The screenshot showing the results

KInfer automatically calculates the initial guesses of the parameters, the user is allowed to change and directly insert different prediction intervals for the parameters.

The optimization algorithm of KInfer is a genetic algorithm [GA; see Goldberg (1989) for technical details]. This choice was driven by the fact that a biological model of realistic size and complexity presents a high number of parameters with possible nonlinear relations among them. Here we briefly recall that a genetic algorithm is a population-based stochastic optimization technique, which,

starting from a set of initial guesses about the solution, determines the next set of possible solutions to the optimization problem on the basis of the results obtained from the preceding set and approaching the best solution step by step.

In particular, in the GA approach, the evolution starts from a population of randomly generated individuals. Then in each generation the fitness of every individual in the population is evaluated, and multiple individuals are stochastically selected from the current population (based on their fitness). The chosen individuals are modified (recombined and possibly randomly mutated) to form a new population. The new population will be used in the next iteration of the algorithm. The algorithm terminates when either a maximum number of generations has been produced or a satisfactory fitness level has been reached for the population. The selection operation involves the evaluation of each possible solution with respect to the target assigned; the lower the log-likelihood value is, the better the solution is considered to be. The next step is to select the solutions for the next generation in such a way that those with higher fitness have higher probability of selection; each guessed solution will be assigned a selection probability derived from the ratio of its square fitness and the sum of the squared fitness of all the solutions. The selected solutions are then subjected to cross-over, mutation, and innovation operators.

To realize cross-over, every two parents create two children in the following way: The algorithm selects randomly from the first parent how many and which variables will have to be kept in the first child. Then from the second parent the algorithm takes the complementary number of variables and uses these values to complete the first child. The second child is then built with the remaining variables of the two parents. The mutation operation, with a low probability (in our examples $p = 0.1$), randomly selects one variable to be mutated. After the selection, the value of the variable is changed, selecting (again randomly) from the possible values it can take excluding the currently one. Finally, the innovation operator randomly selects new solutions never tested to be performed. Usually this operator is kept at a low rate (here at 5%), trying to optimize the trade-off between exploration and exploitation.

Once the new population of experiments is derived from the algorithm, it is then proposed as a new generation for the next algorithm iteration. The size of each population of solution in each generation is held constant. These methods have been designed primarily to address problems that cannot be tackled through traditional optimization algorithms. Such problems are characterized by discontinuities, lack of derivative information, noisy function values, and disjoint search spaces.

Parameter space restriction

The search for the optimal values of rate constants can be made more efficient if we provide the algorithm of optimization of Eq. 8 with initial guesses for these constants. In this way the algorithm does not waste time in exploring large regions of the parameter space or regions in which the model in Eq. 3 is not valid. For this purpose, we also developed and included in KInfer a procedure for the automatic calculation of the initial guesses of the parameters. Therefore, the task of directing the inference method to efficiently explore the parameter space is not left to the user, who often does not have a precise idea about a reasonable value of the parameters.

The derivatives dX/dt at all measured time points t_k can be interpreted as slopes (Savageau and Voit 1982; Voit and Almeida 2004). Given the species i (with $i = 1,\ldots, N$), we can estimate these slopes from the data as $s_i(t_k)$ and approximate the differential equations as

$$s_i(t_k) \approx \left.\frac{dX_i}{dt}\right|_{t=t_k} \tag{18}$$

If the data consist of N species and the concentration of each species i is measured at M time points $[X_i(t_1), X_i(t_2),\ldots, X_i(t_M)]$, we estimate $M \times N$ slopes $s_i(t_k)$ $(k = 1,\ldots, M)$. In fact, for each species we have M differential equations of the form

$$s_i(t_k) \approx f_i(X_1(t_k), X_2(t_k), \ldots, X_N(t_k); \theta_{i1}, \theta_{i2}, \ldots, \theta_{iN_i}) \tag{19}$$

that form a system of M algebraic equations with $M \times N_i$ unknown variables θs, as the slopes s are measurable from the data. In general $M \neq N_i$ and more often $M \gg N_i$ so that the system of $M \times N_i$ equation results overdetermined. The prediction intervals for the parameter can thus be obtained by computing the solution of the system (Eq. 18) with the least-squares procedure for overdetermined systems. Note that at this stage we are not interested in a very precise estimate of the rate constants but only in an approximate guess. Note also that the least-squares method should be considered only as a method of fitting a line to a set of data, not as a method of statistical inference. The parameters calculated with the method of least squares should be called least-squares *solutions* rather than least-squares *estimates* because they are the solutions of the mathematical problem of minimizing the residual sum of squares rather then estimates derived from a statistical model. However, if we make the assumption that the residuals are normally distributed and independent with the same variance σ^2, then the maximum likelihood approach to the estimation of model parameters from data yields the classical formulae for least squares.

A system of equations similar to system (18) can also be written for the experimental uncertainties Δs_i affecting the slopes s_i:

$$\Delta s_i(t_k) \approx \Delta f_i(X_1(t_k), X_2(t_k), \ldots, X_N(t_k); \theta_{i1}, \theta_{i2}, \ldots, \theta_{iN_i}) \tag{20}$$

where

$$\Delta s_i = \Delta\left[\theta_{i1} \prod_{j \in S_1 \subseteq [1,N]} X_j^{\alpha_j}\right] + \Delta\left[\theta_{i2} \prod_{j \in S_2 \subseteq [1,N]} X_j^{\alpha_j}\right] + \cdots + \Delta\left[\theta_{iN_i} \prod_{j \in S_{N_i} \subseteq [1,N]} X_j^{\alpha_j}\right] \tag{21}$$

By using the standard formulas of the error propagation, a single term for the sum on the right-hand side of Eq. 21 is

$$\frac{\Delta\left[\theta_{i1} \prod_{j \in S_1 \subseteq [1,N]} X_j^{\alpha_j}\right]}{\left|\theta_{i1} \prod_{j \in S_1 \subseteq [1,N]} X_j^{\alpha_j}\right|} = \frac{\Delta\theta_{i1}}{\theta_{i1}} + \frac{\Delta\left[\prod_{j \in S_1 \subseteq [1,N]} X_j^{\alpha_j}\right]}{\left|\prod_{j \in S_1 \subseteq [1,N]} X_j^{\alpha_j}\right|} = \frac{\Delta\theta_{i1}}{\theta_{i1}} + \sum_{h=1}^{\#S_1} |\alpha_h| \frac{\Delta X_h}{|X_h|}$$

where $\#S_1$ is the cardinality of the set S_1. Therefore, Eq. 21 becomes

$$\Delta s_i = \sum_{v=1}^{N_i} \left\{ \left(\frac{\Delta\theta_{iv}}{\theta_{i\mu}} + \sum_{h=1}^{\#S_v} |\alpha_h| \frac{\Delta X_h}{|X_h|} \right) \cdot \left| \theta_{i\mu} \prod_{j \in S_v \subseteq [1,N]} X_j^{\alpha_j} \right| \right\} \tag{22}$$

By assuming that the measurements of times are not affected by errors, the error Δs_i is calculated from Eq. 3 as follows

$$\Delta s_i(t_k) = \frac{1}{t_k - t_{k-1}} (\Delta X_i(t_k) - \Delta X_i(t_{k-1}))$$

where $\Delta X_i(t_k)$ is the experimental error for the measurement of concentration of species i at time t_k. Therefore $\Delta s_i(t_k)$ can be obtained from the data and system (22) can be solved to find the size of the prediction intervals $\Delta\theta$ of θs with the same procedure used for system (18). These intervals are also approximate measures of the errors that propagate to the rate constants from the concentration measurements.

We end this section by noting that KInfer is integrated with a design and simulation environment (Dematté et al. 2008a) based on the BlenX language (Dematté et al. 2008b) developed by our resarch group. KInfer can feed the BlenX models with rate constants for running simulations due to the clear de-coupling of qualitative and quantitative descriptions of the programming language approach.

Synthetic case studies

In this section we report and discuss the performance of KInfer in inferring the rate constants of two challenging biochemical networks on simulated data: a model of a buffering SERCA (sarcoplasmic/endoplasmic reticulum calcium ATPase) pump (Higgins et al. 2006; MacLennan et al. 1997) and a small-scale genetic network with feedback loops. Both of these examples are commonly considered as benchmarks for testing the performances of the methods and software that infer the dynamic of a network from time-course data of concentration.

A buffering SERCA pump

Calcium oscillation in nonexcitable cells acts as a messenger between extracellular stimulations and cell function, such as secretion of enzymes. The oscillations are the result of an influx of calcium into the cytosol from the endoplasmic reticulum (ER) through the inositol triphosphate receptors (IP_3R) and the ryanodine receptors (RyR) followed by reuptake of calcium into the ER through the SERCA pumps. Many models have been constructed to reproduce calcium oscillations, and all these models contain a model of the SERCA pump. The SERCA pump uses the chemical energy produced from the conversion of adenosine triphosphate (ATP) into adenosine diphosphate (ADP) to transport calcium ions across the membrane from the cytosol to the ER, against a concentration gradient.

When calcium ions are transported into the ER through the SERCA pump, they are bound to pump proteins on the cytosolic side of the membrane. The protein undergoes a conformational change, which is powered by the energy released from the conversion of ATP in ADP, and the calcium ions are then released on the ER side of the membrane. Although the calcium ions are bound to the pump protein, they do not contribute to the calcium concentration inside the cytosol or to the calcium concentration inside the ER. For this reason the calcium is said to be *buffered* by the SERCA pump. Because there is a large amount of pump protein present [15–75 μM/L in a cardiac ventricular cell (Bers 2001)], the pump is able to bind a large amount of calcium and so the buffering effect is significant (Higgins et al. 2006). Moreover, when the SERCA pump transports calcium, the amount of calcium bound on the cytosolic side of the membrane may not always be equal to the amount released on the ER side, as some calcium remains bound to the pump proteins.

Many different models have been proposed for describing the dynamics of the SERCA pumps. Maclennan et al. (1997) propose a reaction cycle involving four transitions, including the binding of cytosolic calcium, the change of conformation of the pump proteins, the release of

calcium on the ER side of the membrane, and the return to the original conformation. Others suggest a larger number of reactions in the cycle. For example, Stockes and Green (2003) give a reaction diagram with eight reactions. Dode et al. (2002) give a model with six transitions. Lauger suggests a 12-state model with 24 reactions (Lauger 1991).

The SERCA model we consider here is based on the four-state diagram given in Fig. 5 and proposed by Higgins et al. (2006). It contains the same transitions as the reaction cycle given by Maclennan et al. (1997). Table 1 reports the reactions illustrated in the cycle of Fig. 5. X_1 denotes the pump protein on the cytosolic side with no calcium bound, X_2 denotes the pump protein on the cytosolic side with two calcium ions bound, and Y_1 and Y_2 are the analogues on the ER side. C is the calcium on the cytosolic side and C_E is the calcium on the ER side. Using the state diagram, we can write down the set of reactions and convert them into a system of rate equations through the law of mass action.

The parameters used in the model are given in Table 2. The values of k_2, k_4, k_{-2}, and k_{-4} have been determined by Higgins et al. (2006), whereas the values of k_1, k_3, k_{-1} and k_{-3} are assumed to be one order of magnitude greater than the others. Higgins et al. (2006) assume that these rate constants are fast and that $k_1/k_{-1} = 0.7$ $(\mu M/L)^2$ and $k_3/k_{-3} = 1.11 \times 10^{-5}$ $(\mu M/L)^2$.

We report the results of our inference procedure applied to two synthetic datasets. In order to be able to compare our results with those published by Higgins et al. (2006), Dode et al. (2002), and Yano et al. (2004), the first configuration has been generated for $C_E(0)$ = 10 µmol/L ER (configuration 1) and the second has been generated for $C_E(0)$ = 150 µmol/L ER (configuration 2). In both cases $C(0)$ = 5 µmol/L Cyt. The time courses obtained from these two

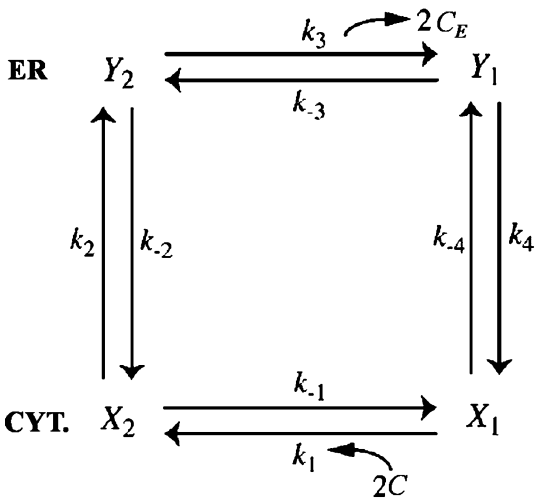

Fig. 5 State diagram of the SERCA pumps. Adapted from Higgins et al. (2006)

Table 1 Set of reactions derived from the four-state SERCA pump diagram in Fig. 5

$X_1 + C + C \xrightarrow{k_1} X_2$
$X_2 \xrightarrow{k_2} Y_2$
$Y_2 \xrightarrow{k_3} Y_1 + C_E + C_E$
$Y_1 \xrightarrow{k_4} X_1$
$X_2 \xrightarrow{k_{-1}} X_1$
$Y_2 \xrightarrow{k_{-2}} X_2$
$Y_1 \xrightarrow{k_{-3}} Y_2$
$X_1 \xrightarrow{k_{-4}} Y_1$

Table 2 Parameters used to generate the synthetic time course of the reagents ofSERCA pumps (Higgins et al. 2006). Note that X_1, X_2, and C are in the units µmoles per liter cytosol(µmol/L Cyt) and Y_1, Y_2, and C_E are in the units µmoles per liter ER (µmol/L ER).Two initial configurations of the system are considered: the first with $C_E(0)$ = 10µmol/L ER and the second with $C_E(0)$ = 150 µmol/L ER. The first configuration isnot realistic but is used here so that the results can be compared with those of Higginset al. (2006), Dode et al. (2002), and Yano et al. (2004)

Parameter	Value
k_1	0.1 s^{-1}
k_{-1}	1 s^{-1}
k_2	2 s^{-1}
k_{-2}	0.97 s^{-1}
k_3	0.5 s^{-1}
k_{-3}	0.1 s^{-1}
k_4	0.4 s^{-1}
k_{-4}	1.2×10^{-3} s^{-1}
Reactant	**Initial concentration**
X_1	5 µmol/L Cyt
X_2	5 µmol/L Cyt
Y_1	15 µmol/L ER
Y_2	15 µmol/L ER
C	5 µmol/L Cyt
C_E	10 and 150 µmol/L ER

initial configurations contain 40 time points sampled in the interval from 0 to 0.00125 s every 3.2 µs. They have been perturbed by a noise with variance equal to 5% of the measurements (this value of variance reflects the typical precision achieved in the experiments).

For the first configuration, the average noise variances (in µmol/L) are as follows: $\sigma_C = 0.0083$, $\sigma_{CE} = 0.0083$, $\sigma_{X1} = 1.66$, $\sigma_{X2} = 0.04$, $\sigma_{Y1} = 0.19$, and $\sigma_{Y2} = 0.1$. For the second configuration, we obtained $\sigma_C = 8.85$, $\sigma_{CE} = 0.67$, $\sigma_{X1} = 0.04$, $\sigma_{X2} = 0.19$, $\sigma_{Y1} = 0.2$, and $\sigma_{Y2} = 0.1$. We considered these values as belonging to a possible range of values for σ, and thus with the simulator Dizzy (Ramsey et al. 2005), we generated noisy time-

course datasets perturbed by these values of σ and applied the parameter inference procedure to all of them. The estimates of the parameters at different levels of noise in the input data are reported in Tables 3 and 5, and the variances of the estimates are shown in Tables 4 and 6 for the datasets from configuration 1 and configuration 2, respectively. In this experiment σ is a priori fixed and is not treated as a parameter to be estimated.

Figures 6, 7, 8, 9, 10, and 11 show the comparisons between the simulations obtained with the "true" parameters listed in Table 2, which are assumed to be the "experimental" time courses, and those obtained with the estimated values for different levels of noise in the input data. The simulations confirm the robustness of the inference accuracy to the noise in the data and consequently the agreement with the parameters estimated by Higgins et al. (2006) and Dode et al. (2002). In particular, for the system in configuration 1, Figs. 6 and 7 demonstrate that the endoplasmic calcium concentration and the concentration of the pump protein on the ER side (Y_1) time course are the most affected by the inaccuracies that the input data propagate to the rate constants. The kinetics of the unperturbed and the perturbed systems are the same in the first instants of reaction and then stabilize on different equilibrium values. For the system in configuration 2, we obtained strong agreement between estimated and "experimental" behaviors of the time course of the endoplasmic calcium concentration (Fig. 9), and a bad agreement, due to the significant deviation of k_2, k_{-3}, and k_{-4} from the "true" values, for the X_1 and Y_1 time courses for $\sigma = 0.67$ μmol/L (Figs. 10, 11). We finally note that the estimated time courses of some of the variables are shifted up or down in their entirety. This is a typical outcome of slope-based estimation of the initial guesses of the parameters; it is caused by the fact that small inaccuracies in fitting initial slopes propagate and accumulate throughout the time course.

Table 3 Estimated parameters for the simulation generated with the following initial conditions $C_E(0) = 10$ μmol/L ER and $C(0) = 5$ μmol/L Cyt

Parameter	Values				
	$\sigma = 0.008$	$\sigma = 0.01$	$\sigma = 0.04$	$\sigma = 0.19$	$\sigma = 1.66$
k_1	0.0945	0.0954	0.0955	1.104	0.012
k_{-1}	0.164	0.146	0.140	0.024	0.726
k_2	3.669	3.640	3.0556	1.176	1.0526
k_{-2}	0.499	0.439	0.442	0.129	1.281
k_3	0.498	0.497	0.498	0.549	0.194
k_{-3}	0.127	0.123	0.144	0.254	0.074
k_4	0.554	0.557	0.577	0.552	0.444
k_{-4}	0.00183	0.00186	0.00256	0.0287	0.00668

In these experiments, σ does not undergo the estimation procedure, and it is fixed at the values indicated in the header of the table. The parameters of the noise-free model are given in Table 2

Table 4 Variance of the estimated parameters from the time courses simulated with the following initial conditions: $C_E(0) = 10$ μmol/L ER and $C(0) = 5$ μmol/L Cyt

Parameter	$\sqrt{\gamma}\times 10^2$ $\sigma = 0.008$	$\sqrt{\gamma}\times 10^2$ $\sigma = 0.01$	$\sqrt{\gamma}\times 10^2$ $\sigma = 0.04$	$\sqrt{\gamma}\times 10^2$ $\sigma = 0.19$	$\sqrt{\gamma}\times 10^2$ $\sigma = 1.66$
k_1	0.382	0.424	2.494	23.771	99.429
k_{-1}	20.756	18.470	18.342	12.065	404.882
k_2	884	899.497	850.718	547.226	235.165
k_{-2}	465.14	641.145	448.693	70.356	453.39
k_3	333.662	324.004	328.261	541.59	114.914
k_{-3}	233.069	230.589	308.925	646.497	100.528
k_4	480.749	483.655	502.737	568.489	193.075
k_{-4}	49.432	50.891	85.66	193.197	415.83

Table 5 Estimated parameters for the simulation generated with the following initial conditions $C_E(0) = 150$ μmol/L ER and $C(0) = 5$ μmol/L Cyt

Parameter	Values				
	$\sigma = 0.04$	$\sigma = 0.1$	$\sigma = 0.20$	$\sigma = 0.67$	$\sigma = 8.85$
k_1	0.0953	0.0982	0.099	0.124	0.000109
k_{-1}	0.241	0.453	0.450	1.095	0.134
k_2	2.908	0.473	0.0541	1.386	0.606
k_{-2}	0.45	0.385	0.201	0.0129	0.639
k_3	0.491	0.488	0.446	0.0977	0.448
k_{-3}	0.131	0.151	0.117	0.0198	0.0325
k_4	0.541	0.478	0.426	0.142	1.094
k_{-4}	0.00248	0.00477	0.00566	0.00381	0.0106

In these experiments σ does not undergo the estimation procedure, and it is fixed at the values indicated in the header of the table. The parameters of the noise-free model are given in Table 2

Table 6 Variance of the estimated parameters from the time courses simulated with the following initial conditions: $C_E(0) = 150$ μmol/L ER and $C(0) = 5$ μmol/L Cyt

Parameter	$\sqrt{\gamma}\times 10^2$ $\sigma = 0.04$	$\sqrt{\gamma}\times 10^2$ $\sigma = 0.1$	$\sqrt{\gamma}\times 10^2$ $\sigma = 0.20$	$\sqrt{\gamma}\times 10^2$ $\sigma = 0.67$	$\sqrt{\gamma}$ $\sigma = 8.85$
k_1	2.096	6.499	30.962	125.249	2,090.852
k_{-1}	31.345	68.353	76.958	147.044	145.544
k_2	730.32	227.941	24.572	131.256	131.256
k_{-2}	469.598	302.94	324.685	477.676	1,027.918
k_3	327.871	349.081	471.648	36.996	36.996
k_{-3}	248.41	524.356	344.394	104.64	104.64
k_4	488.769	479.545	477.328	866.457	860.835
k_{-4}	78.505	144.568	110.221	621.524	621.288

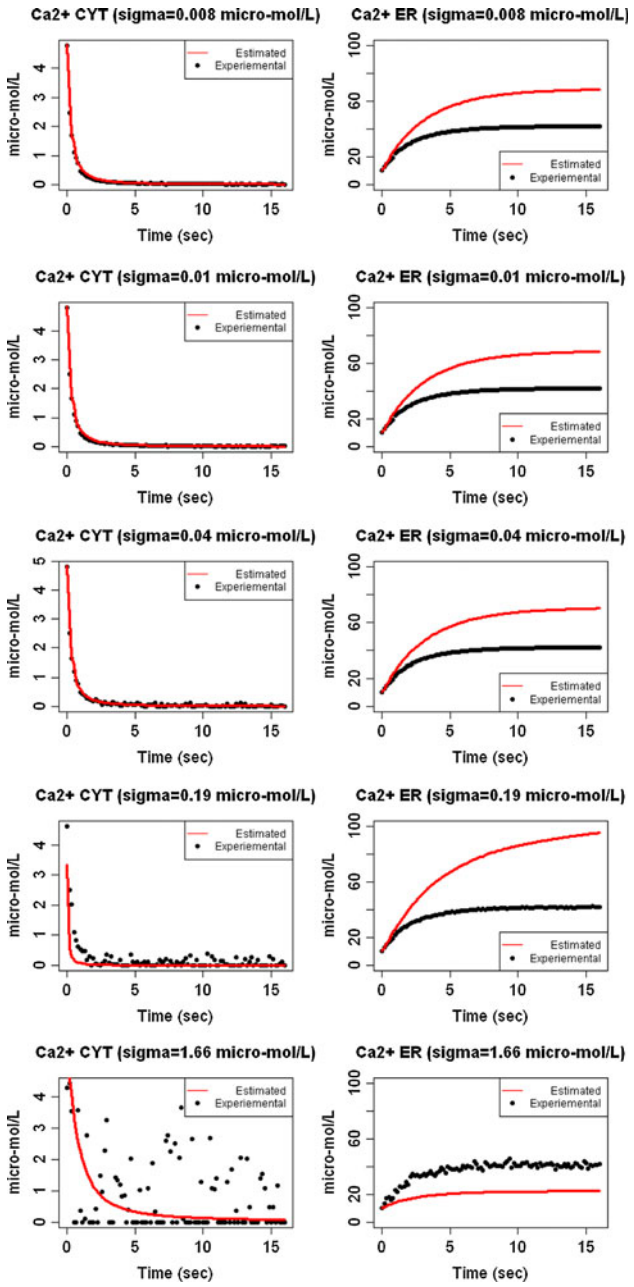

Fig. 6 Synthetic smoothed and noisy dynamics of cytosolic and ER concentrations of Ca^{2+} for $C_E(0) = 10$ μmol/L ER and $C(0) = 5$ μmol/L

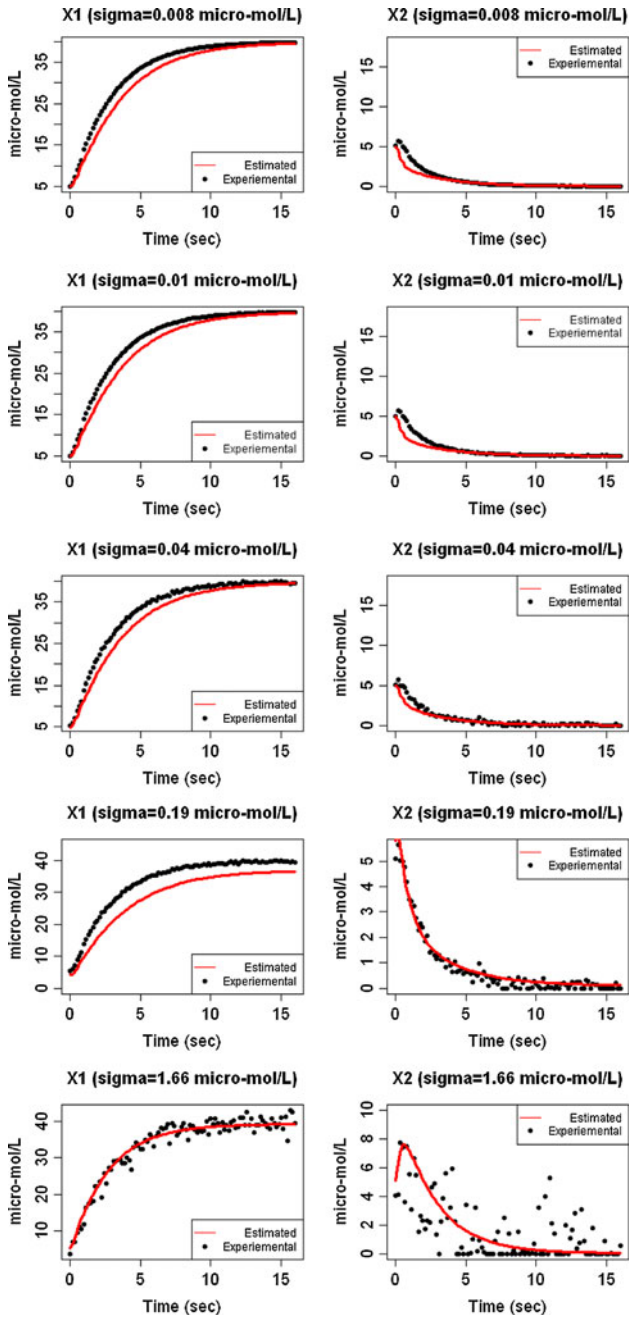

Fig. 7 Synthetic smoothed and noisy dynamics of cytosolic and ER concentrations of X_1 and X_2 for $C_E(0) = 10$ μmol/L ER and $C(0) = 5$ μmol/L

A regulatory genetic network

To confirm the validity of our inference model we also considered the calibration of a typical model of a small-scale gene network shown in Fig. 12. Because of the presence of positive and negative loops and its nonlinear dynamics, this model is considered a challenging benchmark for comparative tests and analysis of inference methods. Here we report a brief description of the model, and we refer the reader to the studies of Hlavacek and Savageau (1996), Savageau (1985), Savageau and Sands (1985), and Kikuchi et al. (2003) for the details of the biological and mathematical aspects of this regulatory network.

Figure 12 is a diagrammatic representation of the gene interaction system originally proposed by Hlavacek and Savageau (1996) to analyze the interaction of regulator and effector genes. The regulator gene encodes a protein that acts at the level of transcription to bring about induction,

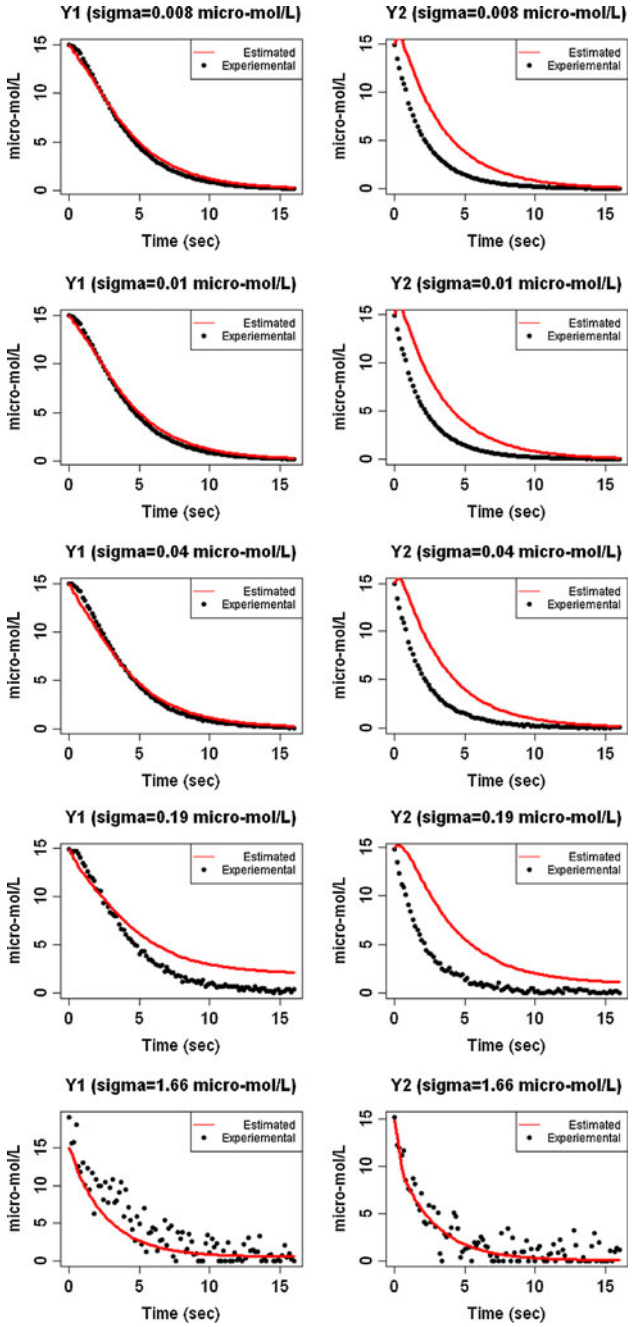

Fig. 8 Synthetic smoothed and noisy dynamics of cytosolic and ER concentrations of Y_1 and Y_2 for $C_E(0) = 10$ μmol/L ER and $C(0) = 5$ μmol/L

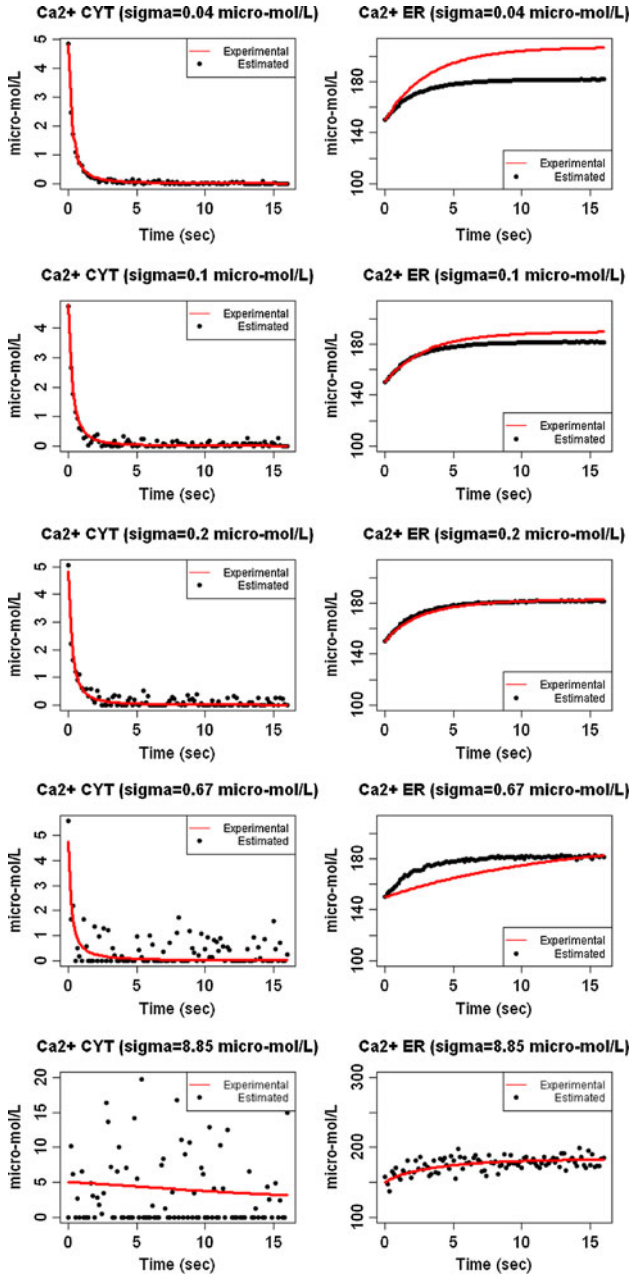

Fig. 9 Synthetic smoothed and noisy dynamics of cytosolic and ER concentrations of Ca^{2+} for $C_E(0) = 150$ μmol/L ER and $C(0) = 5$ μmol/L

and the effector gene encodes an enzyme that catalyzes a pathway in which the inducer of the system is an actual or functional intermediate (Hlavacek and Savageau 1996). The regulator can negatively or positively influence transcription at the promoter of each gene, and these influences can be respectively facilitated or antagonized by the inducer. This gene network model is an idealized regulatory system. Nevertheless it captures the essential features of many actual systems (Hlavacek and Savageau 1996; Savageau 1985; Savageau and Sands 1985).

The network consists of two genes (gene 1, called the *effector* gene, and gene 4, called the *regulator gene*). The state of the system is described by five variables X_1, X_2, X_3, X_4, and X_5. X_1 is an mRNA produced from gene 1, X_2 is an enzyme protein it produces, and X_3 is an inducer protein catalyzed by X_2. X_4 is an mRNA produced from gene 4, and X_5 is a regulator protein it produces. Positive feedback from the inducer protein X_3 and negative feedback from the

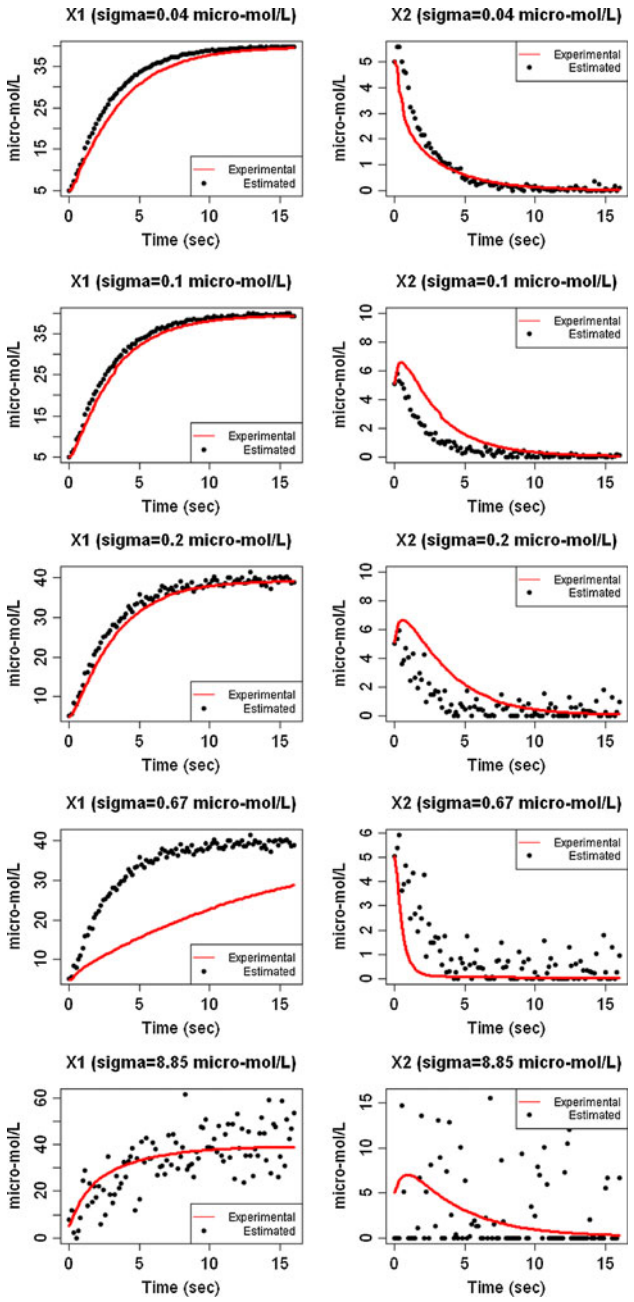

Fig. 10 Synthetic smoothed and noisy dynamics of cytosolic and ER concentrations of X_1 and X_2 for $C_E(0) = 150\mu$mol/L ER and $C(0) =$ 5 μmol/L

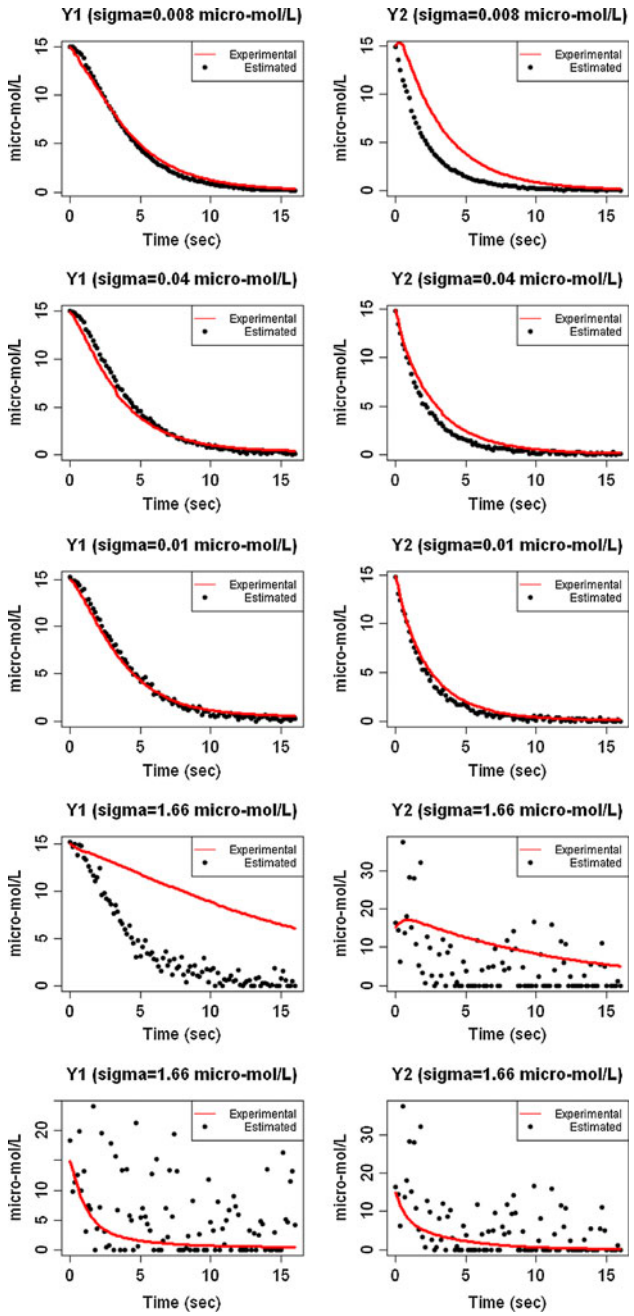

Fig. 11 Synthetic smoothed and noisy dynamics of cytosolic and ER concentrations of Y_1 and Y_2 for $C_E(0) = 150$ μmol/L ER and $C(0) =$ 5 μmol/L

regulator protein X_5 are assumed in the mRNA production of genes 1 and 4. These five variables represent quantities that vary during induction and are determined by the processes of transcription, translation, specific degradation, dilution, and metabolism. The variables X_6 and X_7 denote the precursor pools for mRNA and protein biosynthesis, respectively. In our model, as in the study of Kikuchi et al. (2003) they are held constant with respect to the time. Finally, X_8 denotes a substrate concentration representing an independently determined environmental signal to which the system responds. Our computational experiment used five time courses $X_1(t)$, $X_2(t)$, $X_3(t)$, $X_4(t)$, and $X_5(t)$. They were artificially prepared by solving with the numerical integrator XPPAUT (2008) the S-system in (23) with the parameter values in Table 7 ($n = 5$) and the following initial values $X_1(0) = 0.1$, $X_2(0) = 0.12$, $X_3(0) = X_4(0) = 0.70$, $X_5(0) = 0.18$. These values have also been

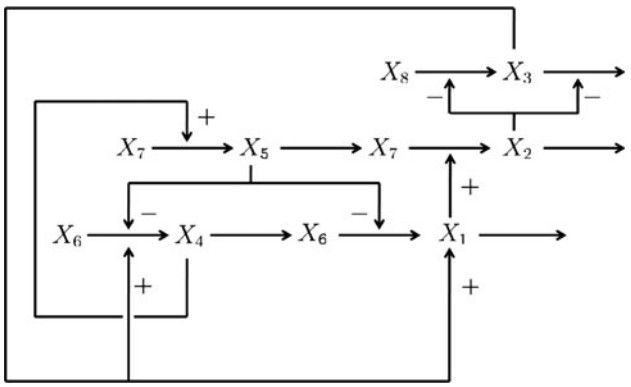

Fig. 12 Genetic network model of Hlavacek and Savageau (1996)

used by Kikuchi et al. (2003) and are expressed in arbitrary units.

$$\frac{dX_i}{dt} = \alpha_i \prod_{j=1}^{n} X_j^{g_{ij}} - \beta_i \prod_{j=1}^{n} X_j^{h_{ij}} \tag{23}$$

Substituting the values of the orders of reaction g_{ij} and h_{ij} reported in Table 7 in system (23) leads to the following set of equations describing the dynamics of the regulatory network.

$$\begin{aligned} \frac{dX_1}{dt} &= \alpha_1 X_3 X_5^{-1} - \beta_1 X_1^2, & \frac{dX_2}{dt} &= \alpha_2 X_1^2 - \beta_2 X_2^2 \\ \frac{dX_3}{dt} &= \alpha_3 X_2^{-1} - \beta_3 X_2^{-1} X_3^2, & \frac{dX_4}{dt} &= \alpha_4 X_3^2 X_5^{-1} - \beta_4 X_4^2 \\ \frac{dX_5}{dt} &= \alpha_5 X_4^2 - \beta_5 X_5^2 \end{aligned} \tag{24}$$

The equation system (24) has been solved with the software XPPAUT (2008), setting the integration tolerance to 10^{-4}. The value of tolerance measures the error affecting the numerical solution of the differential equations. Since in this experiment we did not add any extra noise to the data, the only contribution to the level of noise in the input data is just the integration tolerance, i.e., $\sigma = 10^{-4}$.

Table 8 presents the parameter estimates obtained by KInfer, and the graphs in Fig. 13 show the simulation curves (solid lines) obtained by solving system (24) with the estimated parameters. We maintained the partial reaction orders given by the model of Kikuchi et al. (2003), since KInfer requires them as an input. The inferred rate constants are comparable to those used to generate the input data in Table 7 and reproduce the expected time behavior (black points in graphs of Fig. 13).

Table 8 Estimates of the kinetic parameters of model (24) and the level of noise σ in the input data

Parameter	Value	$\Delta\alpha$	σ_α
α_1	5.41	0.01	0.31
α_2	7.91	0.01	0.76
α_3	2.77	0.01	~ 0
α_4	9.61	0.01	0.27
α_5	10.68	0.01	0.92
Parameter	Value	$\Delta\beta$	σ_β
β_1	11.12	0.01	0.28
β_2	7.691	0.005	0.4
β_3	2.75	0.005	~ 0
β_4	11.84	0.01	0.42
β_5	10.992	0.002	0.81
Parameter	Value		
σ	0.0001		

Real case studies

The analysis of in vivo time series for inferring kinetic parameters is a worthwhile challenge. Since in vivo data have not undergone any artificial process of isolation and purification, they reflect how cells and organisms really behave, how they respond to signals and stimuli, and how they orchestrate functions such as gene expression and regulation and the time evolution of protein concentration.

In this section we report the results obtained with KInfer for the estimate of the rate constants of two biologically

Table 7 Parameters that determine the dynamic action of the S-system in Eq. 23

n	α_i	g_{i1}	g_{i2}	g_{i3}	g_{i4}	g_{i5}	β_i	h_{i1}	h_{i2}	h_{i3}	h_{i4}	h_{i5}
1	5	0	0	1	0	−1	10	2	0	0	0	0
2	10	2	0	0	0	0	10	0	2	0	0	0
3	10	0	−1	0	0	0	10	0	−1	2	0	0
4	8	0	0	2	0	−1	10	0	0	0	2	0
5	10	0	0	0	2	0	10	0	0	0	0	2

These values are expressed in arbitrary units and were determined artificially by Kikuchi et al. (2003) in order to realize the network in Fig. 12. The index i ranges over the interval [1, n], where $n = 5$

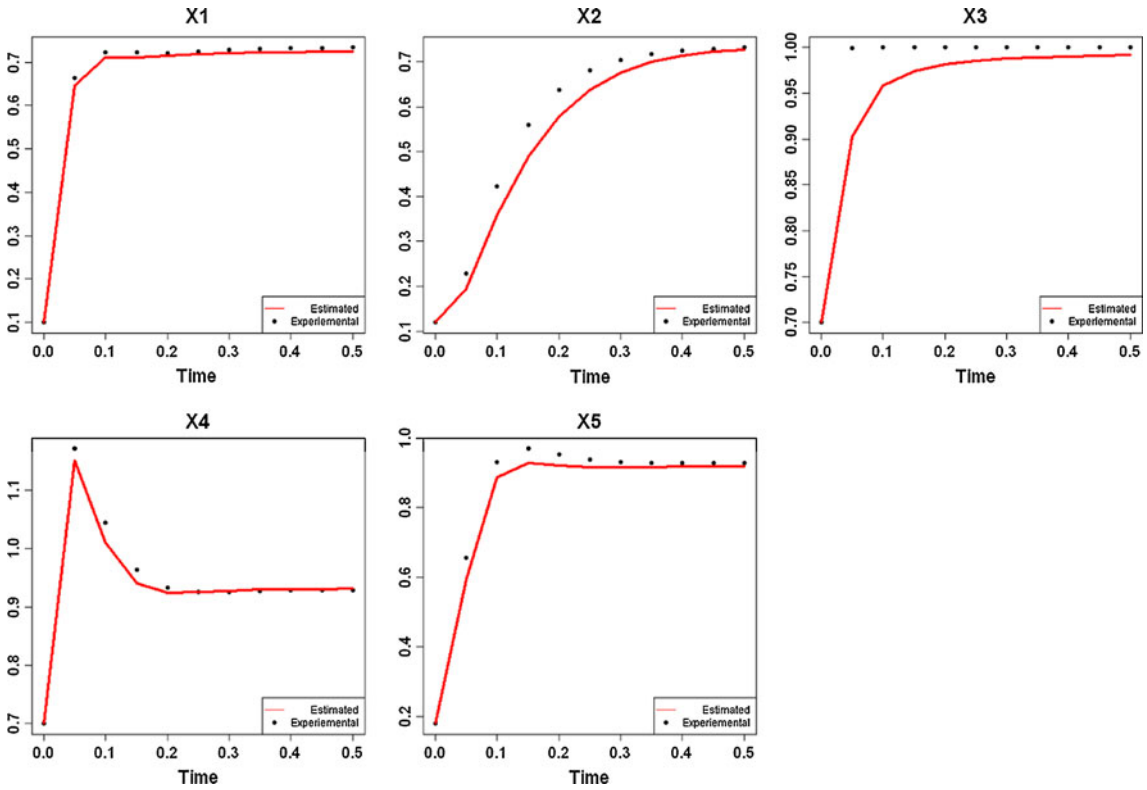

Fig. 13 Comparison between expected behavior, i.e., solution of equation system (24) with parameters given in Table 7 (*black circles*) and estimated behavior obtained as a solution of equation system (24) with the parameters inferred by KInfer (Table 8)

relevant real case studies. The first case study is glycolysis and lactate production in the bacterium *L. lactis*. The relative simplicity of the *L. lactis* metabolism, which converts sugars via the Embden-Meyerhalf-Parnas pathway to pyruvate (Voit et al. 2006), makes the metabolic machinery of this bacterium an attractive case study for testing systemic approaches to modelling biochemical networks. Moreover, although the regulation of glycolysis in *L. lactis* has been the subject of intensive research over the past three decades, a comprehensive understanding of sugar metabolism and regulatory mechanism of the glycolytic pathway in *L. lactis* has not yet been achieved.

The second case study is the subnetwork involving the IκB phosphorylation in the NF-κB pathway. NF-κB is a collective name for the complexes formed by the multigene family that functions as DNA-binding proteins and transcription factors. They are regulators of gene expression in eukaryotic cells but are held in an inactive state by a family of inhibitors (IκB). The biological relevance of this case study is due to the crucial role that NF-κB plays as the central mediator of inflammation with roles in cell death; it has also been implicated in a myriad of common diseases, such as cancer, arthritis, asthma, diabetes, atherosclerosis, and septic shock, to name but a few, and in the regulation of immune responses to infection. Recent detailed theoretical studies and experimental data about the NF-κB pathway can be found in Ihekwaba et al. (2004, 2005, 2007) and Nelson et al. (2004).

Glucose metabolism of *Lactococcus lactis*

We applied our method to infer the rate constants of the biochemical pathway that converts glucose into lactate in the bacterium *L. lactis* (Goel et al. 2008; Voit et al. 2006) (see Fig. 14). The experimental data provided by Prof. Eberhard O. Voit consist of the time series of glucose (X_1), glucose-6-phosphate G6P (X_2), total fructose 1,6-biphosphate FBP (X_3), 3-phosphoglycerate 3-PGA (X_4), phosphoenolpyruvate PEP (X_5), pyruvate (X_6), lactate (X_7), acetate (X_8), ATP, and inorganic phosphate P_i. The mathematical model of this pathway has been formulated by Goel et al. (2008) and Voit et al. (2006) as in the equation system (25).

$$\begin{aligned} \frac{dX_1}{dt} &= -v_1, & \frac{dX_2}{dt} &= v_1 - v_2 \\ \frac{dX_3}{dt} &= v_2 - v_3, & \frac{dX_4}{dt} &= 2v_3 + v_4 \\ \frac{dX_5}{dt} &= -v_4 - v_1 - v_5, & \frac{dX_6}{dt} &= v_1 + v_5 - v_6 - v_7 \\ \frac{dX_7}{dt} &= v_6, & \frac{dX_8}{dt} &= v_7 \end{aligned} \tag{25}$$

where

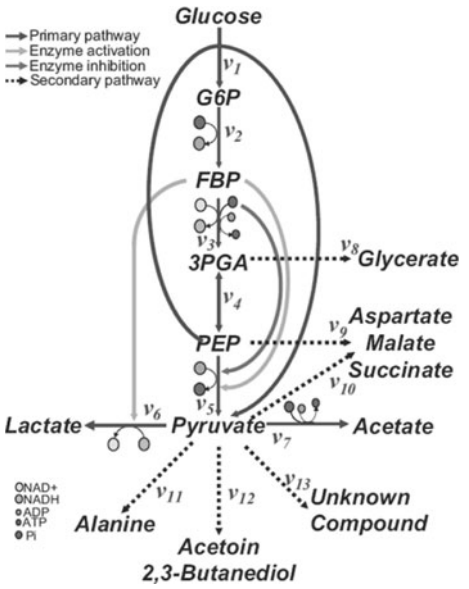

Fig. 14 Pathway of glycolysis and lactate production in *L. lactis*. *Black arrows* indicate flow of material, *gray arrows* enzyme activation and inhibition, *dashed arrows* leakage of material into secondary pathways, which are not considered in the model presented in this paper. This figure has been adapted from Goel et al. (2008)

$$
\begin{aligned}
v_1 &= k_1 X_1^{a_1} X_5^{a_2} \\
v_2 &= k_2 X_2^{a_3} \mathrm{ATP}^{a_4} \\
v_3 &= k_3 X_3^{a_5} \mathrm{Pi}^{a_6} \\
v_4 &= k_4 X_5^{a_7} - k_5 X_4^{a_8} \\
v_5 &= k_6 X_5^{a_9} X_3^{a_{10}} \mathrm{Pi}^{a_{11}} + k_7 X_5^{a_{12}} \\
v_6 &= k_8 X_6^{a_{13}} X_3^{a_{14}} \\
v_7 &= k_9 X_6^{a_{15}} \mathrm{Pi}^{a_{16}}
\end{aligned}
\tag{26}
$$

and the values of the orders of reaction are given in Table 9.

The experimental time-course data were collected with the nuclear magnetic resonance technique by Voit and coworkers (2006, 2008). They monitored the time behavior of the pools of labelled intermediates and end products of the pathway with a resolution of 30 s in nongrowing *L. lactis* bacteria suspension following a pulse of ^{13}C-labelled glucose (Neves et al. 2002; Voit et al. 2006). All the time series consist of 85 data points.

The parameter inference in this model suffers from difficulties of a technical nature due to the peculiarities of the data. In fact, the time behavior of 3-PGA and PEP dips down very quickly, then recovers, overshoots, and slowly degrades (see Figs. 15 and 16). Even if it is possible to model such dynamics with generalized mass action law, the search algorithm might find a set of parameters causing the time course to cross over into the negative domain. In this case, noninteger reaction orders force the integration of the equation system to produce results with imaginary values.

Table 9 Values of the partial orders of reaction in model (26)–(27)

Parameter	Value	Parameter	Value
a_1	0.4	a_9	0.53
a_2	0.81	a_{10}	1.33
a_3	0.74	a_{11}	−0.0001
a_4	0.4	a_{12}	2.30
a_5	0.88	a_{13}	0.46
a_6	0.01	a_{14}	1.04
a_7	0.43	a_{15}	1.0
a_8	0.32	a_{16}	0.46

The values were proposed by Goel et al. (2008)

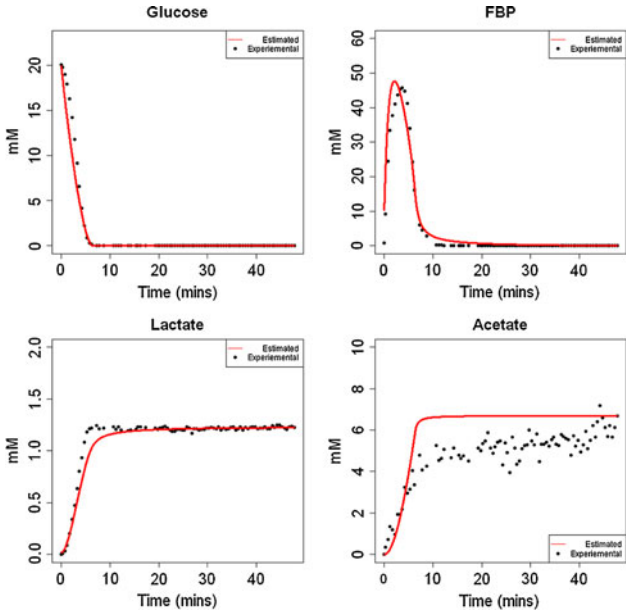

Fig. 15 Comparison between experimental behavior (*black circles*) and estimated behavior obtained as a solution of equation system (25) with the parameters inferred by KInfer (Table 10)

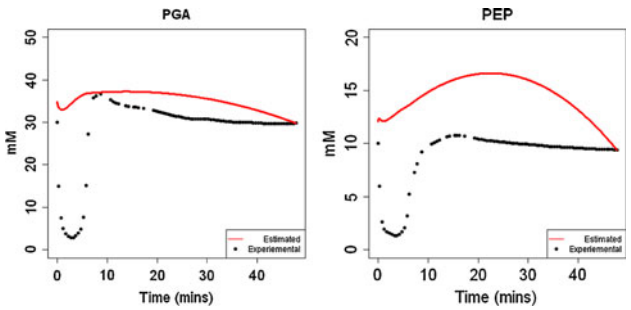

Fig. 16 Comparison between experimental behavior (*black circles*) and estimated behavior obtained as a solution of equation system (25) with the parameters inferred by KInfer (Table 10) for 3-PGA and PEP

Moreover, some variables approach zero toward the end of the experiment. Due to the numerical inaccuracies of any integration software, these variables may become negative.

Also in our specific case, the search algorithm selects a parameter combination such that the simulation of the generalized mass action model with those parameters does not identify a global fit of the experimental data over their entire time domain, namely, the integration with the XPPAUT software stops at $t \approx 6.4$ min. Therefore, to avoid termination of integration, we artificially stopped the simulation of the dynamics of the pathway at $t \approx 6.4$ min, and with the Stineman interpolation algorithm, extrapolated the behavior from $t \approx 6.4$ min until $t \approx 40$ min. We maintained the values of the orders of reaction as in Table 9, and we estimated the kinetic rate constants of the model. Table 10 shows the KInfer estimates of these parameters. The rate constant estimates reproduce the experimental time behavior of the involved species, except for 3-PGA and PEP.

Binding affinity of IκB kinase

Activation of the NF-κB transcription factor can be triggered by exposing cells to a multitude of external stimuli such as tumour necrosis factor (TNF-α) and interleukin 1 (IL-1α). These cytokines initiate numerous and diverse intracellular signalling cascades, most of which activate the IKK complex. This IKK complex regulates the activity of the NF-κB transcription factor positively by phosphorylating the inhibitor I-κB. The IKK catalyzes the transfer of the terminal phosphoryl group of ATP to the I-κB protein substrates, thereby tagging the inhibitor protein for ubiquitination and then degradation. The previously inactive NF-κB is thus activated and available for gene expression. This crucial component in the NF-κB activation cascade typically consists of two catalytic subunits, IKKα (IKK1) and IKKβ (IKK2), and a regulatory unit NEMO (IKKγ). The cytoplasmic inhibitors of NF-κB (the IκBs) are phosphorylated by activated IKK at specific N-terminal residues, tagging them for poly-ubiquitination and rapid proteasomal degradation. Since recombinant human IKK2 (rhIKK2) phosphorylates GST-IκB in vitro, we examined the activity of the synthesized GST-IκB followed by the association and dissociation reaction, as shown in the following chemical reaction system (Ihekwaba et al. 2007):

$$\mathrm{IKK2} + \mathrm{ATP} \xrightarrow{k_1} \mathrm{IKK2} \cdot \mathrm{ATP}$$

$$\mathrm{IKK2} \cdot \mathrm{ATP} \xrightarrow{k_2} \mathrm{IKK2} + \mathrm{ATP}$$

$$\mathrm{IKK2} \cdot \mathrm{ATP} + \text{GST-I}\kappa\mathrm{B}\alpha \xrightarrow{k_3} \mathrm{IKK2} \cdot \mathrm{ATP} \cdot \text{GST-I}\kappa\mathrm{B}\alpha$$

$$\mathrm{IKK2} \cdot \mathrm{ATP} \cdot \text{GST-I}\kappa\mathrm{B}\alpha \xrightarrow{k_4} \mathrm{IKK2} \cdot \mathrm{ATP} + \text{GST-I}\kappa\mathrm{B}\alpha$$

$$\mathrm{IKK2} \cdot \mathrm{ATP} \cdot \text{GST-I}\kappa\mathrm{B}\alpha \xrightarrow{k_5} \mathrm{IKK2} + \text{GST-I}\kappa\mathrm{B}\alpha\text{- P} + \mathrm{ADP}$$

$$\mathrm{IKK2} \cdot \mathrm{ATP} \cdot \text{GST-I}\kappa\mathrm{B}\alpha \xrightarrow{k_6} \mathrm{IKK2} \cdot \text{GST-I}\kappa\mathrm{B}\alpha + \mathrm{ATP}$$

$$\mathrm{IKK2} \cdot \text{GST-I}\kappa\mathrm{B}\alpha + \mathrm{ATP} \xrightarrow{k_7} \mathrm{IKK2} \cdot \mathrm{ATP} \cdot \text{GST-I}\kappa\mathrm{B}\alpha$$

$$\mathrm{IKK2} \cdot \text{GST -I}\kappa\mathrm{B}\alpha + \mathrm{ATP} \xrightarrow{k_8} \mathrm{IKK2} + \mathrm{ATP} + \text{GST-I}\kappa\mathrm{B}\alpha$$

$$\mathrm{IKK2} + \mathrm{ATP} + \text{GST-I}\kappa\mathrm{B}\alpha \xrightarrow{k_9} \mathrm{IKK2} \cdot \text{GST-I } \kappa\mathrm{B}\alpha + \mathrm{ATP}$$

where GST-IκBα-P is the phosphorylated substrate GST-IκBα. A time-course plot of the increase in GST-IκBα-P was recorded to monitor the reaction advance (filled circles in Fig. 18). The measured initial concentrations of the system's components are as follows: $[\mathrm{IKK2}]_0 = 50$ nM, $[\mathrm{ATP}]_0 = 200$ nM, [GST-IκBα] = 1,000 nM, and $[\mathrm{IKK2 \cdot ATP}]_0 = [\mathrm{IKK2 \cdot ATP + GST\text{-}I}\kappa\,\mathrm{B}\,\alpha\text{-P}]_0 = [\text{GST-I}\,\kappa\,\mathrm{B}\,\alpha\text{-P}]_0 = [\mathrm{ADP}]_0 \approx 0$ nM.

In this experiment, the only measured time course is the one of GST-IκBα-P. Nevertheless, from the law of mass action and assuming that ATP is in excess and thus its concentration does not change significantly over time, the time course of IKK2 and GST-IκBα can be derived respectively as:

$$[\mathrm{IKK2}] \approx [\mathrm{IKK2}]_0 - \frac{1}{k_{\mathrm{cat}}}[\text{GST-I}\kappa\mathrm{B}\alpha\text{-P}]'$$

and

$$[\text{GST-I}\kappa\mathrm{B}\alpha] \approx [\text{GST-I}\kappa\mathrm{B}\alpha]_0 - \frac{c_2[\text{GST} - \mathrm{I}\kappa\mathrm{B}\alpha - \mathrm{P}]'}{c_1 - c_3[\text{GST-I}\kappa\mathrm{B}\alpha\text{-P}]'}$$

where the second equation is the best fit of the experimental data reported in Fig. 17, and [GST-I κB α-P]′ is the time derivative of the [GST-I κB α-P]. The coefficients of the fit are $c_2 = 2996.03$ nM min and $c_3 = 0.67078$ min. The inferred values of the rate constants are listed in Table 11. Figure 18 shows a good agreement within the error range, between the estimated time course and the measured time course of [GST-Iκ Bα-P].

Table 10 Estimates of the kinetic rate constants of the pathway of regulation of glycolysis in *L. lactis*. In this experiment, the variances of the estimated parameters are two orders of magnitudes bigger than the estimates and indicate the large scale of the parameters of being spread out

Parameter	Value	Δk	σ_k
k_1	0.388	0.001	16.9
k_2	10.35	0.011	153.25
k_3	1.300496	0.000012	192.78
k_4	89.16	0.12	1,602.93
k_5	87.41	0.10	1,699.55
k_6	0.0538	0.0011	11.78
k_7	0.0050	0.0008	39.18
k_8	0.00824	0.00007	7.91
k_9	0.012458	0.000006	1.88

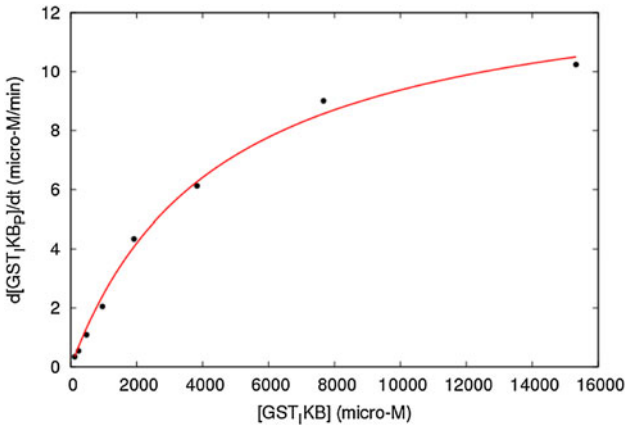

Fig. 17 Reaction velocity (d[GST-IκBα-P]/dt) versus substrate concentration ([GST-IκBα])

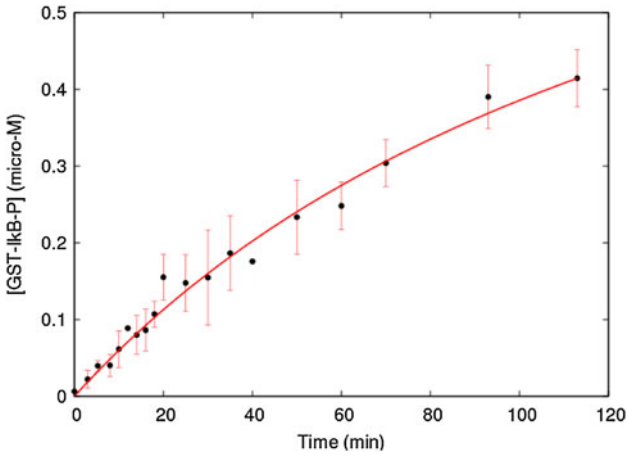

Fig. 18 Time behavior of GST-IκBα-P. The *error bars* show the standard deviation associated with each measurement. IKK modifies IκBα by chemically adding phosphate groups to them (phosphorylation). This causes a functional change. IκBα is recognized and tagged for ubiquitination and degradation. It also affects association (or binding) with other proteins (this is why binding to ATP is also measured). IKK works by removing a phosphate group from ATP and covalently attaching it to Ser 32 and Ser 36 on IκBα site. Since ATP is known to transport chemical energy, radio-labelled ATP is added into the assay to track the transfer of the phosphate group at each time point by liquid scintillation counting. Therefore, the amount of IKK-catalyzed incorporation of ^{33}P into each peptide was quantified by liquid scintillation counting

Table 11 Estimates of the model parameters. Δk is the experimental error propagating from the measures of concentration to the parameters, and σ_k is the variance of the parameter estimate; σ is the variance of the experimental uncertainty on the concentration measurements

Parameter	Value	Δk	σ_k	Units
k_1	1.2308	0.0008	9.6	$(\text{nM min})^{-1}$
k_2	0.93060	0.00018	14.4	min^{-1}
k_3	2.8506	0.0003	12.6	$(\text{nM min})^{-1}$
k_4	0.148654	0.000012	43.2	min^{-1}
k_5	0.692538	0.000018	42.6	min^{-1}
k_6	0.043230	0.000006	42.6	min^{-1}
k_7	12.247	0.005	0.54	$(\text{nM min})^{-1}$
k_8	1.351	0.004	0.54	$(\text{nM min})^{-1}$
k_9	0.4345	0.0004	8.4	$(\text{nM min})^{-1}$
σ	0.8006856	–	–	nM

Discussion

The results of the application of our inference procedure to the calibration of models of synthetic and real biochemical networks show that the method converges to the expected solutions within the bounds of the experimental errors that propagate from concentration measurements to the kinetic rate constants. The good estimates obtained confirm the validity of the procedure applied to any kind of reaction and the validity of the approximated model of mass action law for the rate equation. Moreover, some important features missing from the existing methods for parameter inference are present in our method. The first is the automatic computation of the initial guesses of the parameters. In this way, the user is not forced to insert any a priori knowledge about the system, which often is quite hard to find. At the same time, the method is equipped with a rigorous procedure referring only to the experimental concentration measurements to identify a region of the parameter space where the optimization of the probability density function takes place. The second feature is the implementation of the experimental error propagation. The evaluation of the experimental uncertainty on the rate constant estimates is particularly useful if the parameter inference is incorporated in projects of experimental design. The size of the errors for the kinetic constants is indicative of the optimality of the experimental setup. Thus, any procedure devoted to the reduction of this error is definitely part of a methodology aiming to optimize the design of the experimental configuration. Finally, the software prototype KInfer, implementing the mathematical model of inference, can be used for interfacing the outcomes of the wet-lab activity for the concentration measurements with software for modelling and simulation of biochemical networks.

KInfer is part of the CoSBiLab platform (http://www.cosbi.eu) and interfaces with Beta Workbench, which is a biochemical kinetics simulator based on the programming language BlenX. At present, we are translating the ordinary differential equation models presented in this paper into the BlenX programming language using the KInfer estimated kinetic parameters, to obtain the dynamics simulation in a way that does not rely on the numerical integration of differential equations, and consequently that might not suffer

from the technical difficulties related to the integration of stiff dynamics.

Acknowledgements The authors wish to thank Guido Sanguinetti of the Dept. of Computer Science of University of Sheffield, UK, for valuable discussions and suggestions useful in the development of the theory and Eberhard O. Voit of Systems Biology, Georgia Institute of Technology, USA, for providing data on the glucose metabolism for *Lactococcus lactis*.

References

Bers DM (2001) Excitation-contraction coupling and cardiac contractile force, 2nd ed. Kluwer Academic, Amsterdam

Boys RJ, Wilkinson DJ, Kirkwood TB (2008) Bayesian inference for a discretely observed stochastic kinetic model. Statistics and computing. Springer, Amsterdam

Casella G, Berger RL (2002) Statistical inference. Duxbury, North Scituate

Cho RJ, Campbell MJ, Winzeler EA, Steinmetz L, Conway A, Wodicka L, Wolfsberg TG, Garielian AE, Landsman D, Lockhart DJ, Davis RW (1998) A genome-wide transcriptional analysis of the mitotic cell cycle. Mol Cell 2:65–73

Chou I-C, Voit EO (2009) Recent developments in parameter estimation and structure identification of biochemical and genomic systems. Math Biosci 219(2):57–83

Cho KH, Koch W, Wolkenhauer W (2003) Experimental design in systems biology, based on paramter sensitivity analysis using a Monte Carlo method: a case study for the tnfα-mediated nf-κb signal transduction pathway. Simulation 79(12):726–739

Chou IC, Martens H, Voit EO (2006) Parameter estimation in biochemical systems models with alternating regression. Theor Biol Med Model 3:25

David B, Bastin G (1999) A maximum likelihood parameter estimation method for nonlinear dynamics systems. In: 38th conference on decision & control. Phoenix, AZ, USA, December 1999

Dematté L, Priami C, Romanel A (2008a) The Beta Workbench: a computational tool to study the dynamics of biological systems. Brief Bioinform 9(5):437–449

Dematté L, Priami C, Romanel A (2008b) The BlenX language: a tutorial. LNCS (5016):313–365

Dode LB, Vilsen K, van Baelen F, Wuytack F, Clausen JD, Andersen JP (2002) Dissection of the functional differences between sarco(endo)plasmic reticulum ca^{2+}-ATPase (SERCA) 1 and 3 isoforms by steady state and transient kinetic analyses. J Biol Chem 277:45579–45591

Goel G, Chou I-C, Voit EO (2008) System estimation from metabolic time-series data. Bioinformatics 24(21):2505–2511

Goldberg DE (1989) Genetic algorithms in search, optimization and machine learning. Addison-Wesley, Massachusetts

Golightly A, Wilkinson DJ (2008) Bayesian inference for nonlinear multivariate diffusion models observed with error. Comput Stat Data Anal 52(3):1674–1693

Higgins ER, Cannel MB, Sneyd J (2006) A buffering SERCA pump in models of calcium dynamics. Biophys J 91:151–163

Hlavacek WS, Savageau MA (1996) Rules for coupled expression of regulator and effector genes in inducible circuits. J Mol Biol 255:121–139

Hoops S, Sahle S, Gauges R, Lee C, Pahle J, Simus N (2006) Copasi—a complex pathway simulator. Bioinformatics 22: 3067–3074

Ihekwaba AEC, Broomhead DS, Grimley R, Benson N, Kell DB (2004) Iκbα. Syst Biol 1:99–103

Ihekwaba AEC, Broomhead DS, Grimley R, Benson NN, White MRH, Kell DB (2005) Iκbα. IEE Proc Syst Biol 152:153–160

Ihekwaba AEC, Wilkinson SJ, Broomhead DS, Waithe D, Grimley R, Benson N, Kell DB (2007) Bridging the gap between in silico and cell based analysis of the nf-κb signalling pathway by in vitro studies of ikk2. FEBS J 27:1678–1690

Kikuchi S, Tominaga D, Arita M, Takahashi K, Tomita M (2003) Dynamic modeling of genetic networks using genetic algorithm and S-system. Bioinformatics 10(5):643–650

Lauger P (1991) In: Nadels L, Stein D (eds) Electrogenic ions pumps. Addison-Wesley, Reading

Lecca P, Palmisano A, Priami C, Sanguinetti G (2007) A new method for inferring rate coefficients from experimental time-consecutive measurements of reactant concentrations. In: International conference on systems biology, Long Beach, California. http://www.icsb07.org

Lecca P, Palmisano A, Priami C (2008) Inferring rate coefficients of biochemical reactions from noisy data with KInfer. Technical report 17. The Microsoft Research, University of Trento, Centre for Computational and Systems Biology. http://www.cosbi.eu/Rpty_Tech.php

Lecca P, Palmisano A, Priami C, Sanguinetti G (2009) A new probabilistic generative model of parameter inference in biochemical networks. In: Proceedings of the 2009 ACM symposium on applied computing '09. Hawaii, USA

Lecca P, Palmisano A, Priami C (2009) Deducing chemical reaction rate constants and their regions of confidence from noisy measurements of time series of concentration. In: 11th International conference on computer modelling and simulation (UKSim 2009), Cambridge, England

MacLennan DH, Rice WJ, Green NM (1997) The mechanism of the Ca^{2+} transport by sarco(endo)plasmic reticulum Ca^{2+}-ATPase. J Biol Chem 272:28815–28818

Marin-Sanguino A, Voit EO, Gonzalez-Alcon C, Torres NV (2007) Optimization of biotechnological systems through geometric programming. Theor Biol Med Model 4:38

Moles GC, Mendes P, Banga JR (2003) Parameter estimation in biochemical pathways: a comparison of global optimization methods. Genome Res 13:2467–2474

Nelson DE, Ihekwaba AEC, Elliott M, Johnson JR, Gibney CA, Foreman BE, Nelson G, See V, Horton CA, Spiller DG, Edwards SW, McDowell HP, Unitt JF, Sullivan E, Grimley R, Benson N, Broomhead D, Kell DB, White MRH (2004) Oscillations in nf-κb signaling control the dynamics of gene expression. Science 306:704–708

Neves AR, Ramos A, Costa and H, van Swam II, Hugenholtz J, de Vos M, Kleerebezem VW, Santos H (2002) Effect of different NADH oxidase levels on glucose metabolism by *Lactococcus lactis*: kinetics of intracellular metabolite pools determined by in vivo nuclear magnetic resonance. Appl Environ Microbiol 68:6332–6342

Polisetty PK, Voit EO (2006) Identification of metabolic system parameters using global optimization methods. Theor Biol Med Model 3:4

Ramsey S, Orrell D, Bolouri H (2005) Dizzy: stochastic simulation of large-scale genetic regulatory networks. J Bioinform Comput Biol 3(2):415–436

Reinker S, Altman RM, Timmer J (2006) Parameter estimation in stochastic biochemical reactions. IEEE Proc Syst Biol 153: 168–178

Rodrigez-Fernandez M, Mendes P, Banga J (2006) A hybrid approach for efficient and robust parameter estimation in biochemical pathways. Biosystems 83:248–265

Savageau M (1985) Coupled circuits of gene regulation. In: Calendar R, Gold L (eds) Sequence specificity in transcription and translation. Liss, New York

Savageau MA, Sands PJ (1985) Completely uncoupled and perfectly coupled circuits for inducible gene regulation. In: Voit EO (ed) Canonical non-linear modeling: S-system approach to understanding complexity. Van Nostrand Reinhold, New York

Savageau MA, Voit EO (1982) Power-law approach to modeling biological-systems theory. J Ferment Technol 60:221–228

Stokes DL, Green NM (2003) Structure and function of the calcium pumps. Annu Rev Biophys Biomol Struct 32:445–668

Sugimoto M, Kikuchi S, Tomita M (2005) Reverse engineering of biochemical equations from time-course data by means of genetic programming. Biosystems 80:155–164

Tian T, Xu S, Burrage K (2007) Simulated maximum likelihood method for estimating kinetic rates in gene expression. Bioinformatics 23(1):84–91

Voit EO, Almeida J (2004) Decoupling dynamical systems for pathway identification from metabolic profiles. Bioinformatics 20:1670–1681

Voit EO, Almeida J, Marino S, Lall R, Goel G, Neves AR, Santos H (2006) Regulation of glycolysis in *Lactococcus lactis*: an unfinished systems biological case study. IEE Proc Syst Biol 153(4):286–298

Vyshemirsky V, Girolami MA (2008a) Bayesian ranking of biochemical system models. Bioinformatics 24(6):833–839

Vyshemirsky V, Girolami MA (2008b) Biobayes: a software package for bayesian inference in systems biology. Bioinformatics 24(17):1933–1934

Vyshemirsky V, Girolami MA (2008c) Biobayes: a software package for Bayesian inference for systems biology. Bioinformatics 24:1933–1934

XPPAUT (2008) X-windows phaseplane plus auto. http://www.math.pitt.edu/~bard/xpp/xpp.html

Wilkinson D (2006) Stochastic modelling for systems biology. Chapman and Hall, London

Wilkinson DJ (2007) Bayesian methods in bioinformatics and computational systems biology. Brief Bioinform 8:109–116

Yano K, Petersen OH, Tepikin AV (2004) Dual sensitivity of sarcoplasmic/endoplasmic Ca^{2+}-ATPase to cytosolic and endoplasmic reticulum Ca^{2+} as a mechanism of modulating cytosolic Ca^{2+} oscillations. Biochem J 383:353–360

Zwolak JW, Tyson JJ, Watson LT (2001) Estimating rate constants in cell cycle models. In: Tentner A (ed) Proceedings of high performance constants in cell cycle models, San Diego, CA, pp 53–57