BIOINFORMATICS ORIGINAL PAPER

Vol. 26 no. 7 2010, pages 939–945
doi:10.1093/bioinformatics/btq074

Systems biology

Advance Access publication February 21, 2010

An optimal experimental design approach to model discrimination in dynamic biochemical systems

Dominik Skanda[1] and Dirk Lebiedz[1,2,*]

[1]Center for Analysis of Biological Systems (ZBSA), University of Freiburg, Habsburgerstr. 49, 79104 Freiburg and [2]Interdisciplinary Center for Scientific Computing, Im Neuenheimer Feld 368, 69120 Heidelberg, Germany

Associate Editor: Olga Troyanskaya

ABSTRACT

Motivation: Finding suitable models of dynamic biochemical systems is an important task in systems biology approaches to the biosciences. On the one hand, a correct model helps to understand the underlying mechanisms and on the other hand, one can use the model to predict the behavior of a biological system under various circumstances. Typically, before the correct model of a biochemical system is found, different hypothetical models might be reasonable and consistent with previous knowledge and available data. The main goal now is to find the best suited model out of different hypotheses. The process of falsifying inappropriate candidate models is called model discrimination.

Results: We have developed a new computational tool to compute optimal experiments for biochemical kinetic systems with underlying ordinary differential equation (ODE) models for the purpose of model discrimination. We were inspired by the demands of biological experimentalists which perform one run measurement where perturbations to the system are possible. We provide a criterion which calculates the number and location of time points of optimal measurements as well as optimal initial conditions and optimal perturbations to the system.

Availability: The model discrimination algorithm described here is implemented in C++ in the package ModelDiscriminationToolkit. The source code can be downloaded from http://omnibus.uni-freiburg.de/~ds500/_software.html

Contact: dirk.lebiedz@biologie.uni-freiburg.de

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on November 4, 2009; revised and accepted on February 17, 2010

1 INTRODUCTION

Typically, before the correct model of a biochemical system is found different hypothetical models might be reasonable and consistent with previous knowledge and available data. Let us assume that all the models have in common that they fit one initial series of measurements. The main goal now is to find the best suited model out of different hypotheses. This is usually done by iterative measurements and successive fitting of the different models to the collectivity of all series of measurements. This is repeated until all 'wrong' models do not fit to the collectivity of all series of measurements any more. Thus, the 'wrong' models are iteratively falsified. The whole process is called model discrimination.

To discriminate a set of candidate models against a given set of experimental data often likelihood ratio tests based on bootstrap methods are preformed (see e.g. Horn, 1987; Stricker *et al.*, 1994; Timmer *et al.*, 2004). Ranking methods like Stewart's method (Stewart *et al.*, 1998) or the well-known Akaike information criterion (see e.g. Burnham and Anderson, 2002) are popular as well in the field of biological modeling. Applications can be found, for example, in Jain *et al.* (2006) or Bernacki and Murphy (2009).

This article deals with the problem of designing experiments so that these methods can be applied in an optimal sense. This fundamentally differs from the case of finding an experimental design to best estimate the parameters of a model for a given experimental system (see e.g. Balsa-Canto *et al.*, 2008; Bauer *et al.*, 2000; Körkel *et al.*, 1999). An overview of different experimental design techniques can be found in Kreutz and Timmer (2009). Different approaches to design experiments for model discrimination exist. Beside optimization methods (see e.g. Cooney and McDonald, 1995; Kremling *et al.*, 2004; Lacey and Dunne, 1984; Takors *et al.*, 1997), a model-based feedback controller (see e.g. Apgar *et al.*, 2008) and Markov chain Monte Carlo sampling methods (Myung and Pitt, 2009) are used to construct an appropriate design.

We have developed a new computational tool to compute optimal experiments for biochemical kinetic systems with underlying ordinary differential equation (ODE) models for the purpose of model discrimination. We were inspired by the demands of biological experimentalists who perform one run measurement where perturbations to the system are possible. We provide a criterion which calculates optimal measurement time points instead of a continuous design (see e.g. Atkinson and Fedorov, 1975; Cooney and McDonald, 1995; Kremling *et al.*, 2004). The theory of the new criterion is presented in Section 2. In Section 2.1, we give a brief overview of Kullback–Leibler (KL)-optimality introduced by López-Fidalgo *et al.* (2007). In Section 2.2, we derive our optimality criterion by use of KL-optimality. In Section 2.3, we present the numerical implementation of our optimal design algorithm. Numerical results on allosteric models for glycolytic oscillations are presented in Section 3. We provide Supplementary Material for more details on theory and applications.

*To whom correspondence should be addressed.

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 31, 2016

© The Author 2010. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org

2 DISCRIMINATION BETWEEN DYNAMIC BIOCHEMICAL MODELS

2.1 KL-optimal design

In this section, a model discrimination criterion based on the KL distance called KL-optimality introduced by López-Fidalgo *et al.* (2007) is briefly presented.

López-Fidalgo *et al.* (2007) demonstrate that KL-optimality is consistent with T-optimality (Atkinson and Fedorov, 1975) and generalized T-optimality (Uciński and Bogacka, 2004), which are well-known model discrimination criteria.

The advantage of the KL-optimality criterion is that it does not depend on the general non-linear regression models y_1 and y_2 with corresponding candidate Models 1 and 2

$$y_{1,2} = \eta_{1,2}(x,\theta_{1,2}) + \epsilon, \quad x \in \chi \tag{1}$$

where the random variables ϵ are independent and normally distributed with zero mean and constant variance σ^2 and $\eta_{1,2}(x,\theta_{1,2})$ are the model functions with unknown parameters $\theta_1 \in \Omega_1 \subset \mathbb{R}^{m_1}$ and $\theta_2 \in \Omega_2 \subset \mathbb{R}^{m_2}$, but treats the more general case of two rival density functions $f_1(y,x,\theta_1,\tau)$ and $f_2(y,x,\theta_2,\tau)$, where τ is a parameter, e.g. the variances of the models.

The KL distance (Eguchi and Copas, 2006; Lindsey, 1996) of Model 2 with density function $f_2(y,x,\theta_2,\tau)$ from Model 1 with density function $f_1(y,x,\theta_1,\tau)$ is defined by

$$\mathcal{I}(f_1,f_2,x,\theta_1,\theta_2,\tau) = \int f_1(y,x,\theta_1,\tau)\log\left\{\frac{f_1(y,x,\theta_1,\tau)}{f_2(y,x,\theta_2,\tau)}\right\}\mathrm{d}y, \quad \forall x \in \chi. \tag{2}$$

The KL distance is a non-symmetric measure of the difference between two probability distributions. The integral is computed over the sample space of the possible observations. If the probability space is discrete the integral changes to a sum.

As reviewed by Eguchi and Copas (2006), the KL distance is closely related to likelihood ratio tests. Considering $f_1(y,x,\theta_1,\tau)$ as being a null hypothesis H and $f_2(y,x,\theta_2,\tau)$ being the alternative hypothesis A, the log likelihood ratio is defined as

$$\lambda = \lambda(y) = \log\left\{\frac{f_2(y,x,\theta_2,\tau)}{f_1(y,x,\theta_1,\tau)}\right\}. \tag{3}$$

Intuitively, one expects $\lambda(y)$ to have a high value if A is true and both probability distributions are reasonably well separated.

The Neyman–Pearson lemma [a proof is given in Lindsey (1996)] justifies the optimality of the likelihood ratio critical region

$$W = \{y : \lambda(y) \geq u\}, \tag{4}$$

with respect to Type I and Type II errors. We denote $P_H(W)$ as the probability of the region W under the assumption that the null hypothesis is true. By setting u such that $P_H(W) = \alpha$, one can use W as a critical region to reject H by applying the likelihood ratio test. The Neyman–Pearson lemma states that this critical region is superior to any other W^* with respect to the Type II error, i.e.

$$P_A(W) \geq P_A(W^*), \tag{5}$$

where $P_A(W)$ is the probability of W if the alternative hypothesis A is true. For this critical region W, one can see that the Type II error is as much smaller as the mean with respect to A of the likelihood ratio $\lambda(y)$ is bigger.

The mean with respect to A of the likelihood ratio $\lambda(y)$ is given by

$$E_A\{\lambda(y)\} = \int f_2(y,x,\theta_2,\tau)\log\left\{\frac{f_2(y,x,\theta_2,\tau)}{f_1(y,x,\theta_1,\tau)}\right\}\mathrm{d}y. \tag{6}$$

Therefore, $E_A\{\lambda(y)\} = \mathcal{I}(f_2,f_1,x,\theta_2,\theta_1,\tau)$ which directly leads to the KL-optimality criterion

$$I_{21}(\xi) = \max_{\xi} \int f_2(y,x(\xi),\theta_2,\tau,\xi)\log\left\{\frac{f_2(y,x(\xi),\theta_2,\tau,\xi)}{f_1(y,x(\xi),\theta_1,\tau,\xi)}\right\}\mathrm{d}y, \tag{7}$$

where ξ denotes the set of all experimental conditions subject to optimal design.

2.2 Mathematical construction of the optimal experimental design

In the following section, we use the framework of the KL-optimal design to derive an optimization criterion which is constructed to best fit the needs to discriminate between models of kinetic time series experiments often performed by biologist, e.g *in vitro* experiments of enzyme kinetics.

In most situations such experiments are time and cost consuming. Therefore, one wants to get the most information out of a single experiment taking place within a given fixed time span. This means that in an optimal experimental design the most informative measurement time points for one measurement run have to be calculated in such a way that only one measurement at one time point can be performed.

Often, an experiment cannot produce measurements in a time continuous way. Therefore, we assume that there has to be a minimal time Δt for the separation of measurement time points.

Additionally, the initial species concentrations of the participating species should be chosen in a most discriminating way.

A commonly used *in vitro* practice is to disturb time series measurements by external adding of species quantities. In the context of model discrimination, one wants to know the optimal time point of perturbation and the optimal species quantities, which should be added to the experimental system. We further assume that a measurement cannot be done at the same time as a perturbation.

Given the measurement time-vector $t \in \mathbb{R}_+^N$ with entries for the N measurement time points $t_i, i \in 1,\ldots,N$ such that $t_{i+1} \geq t_i$, the model response vectors $y_{1,2}^i := y_{1,2}(t_{i-1},t_i,y_{1,2}^{i-1}+c_i,\theta_{1,2})$, $i \in \{1,\ldots,N\}$ are the solution of the ODEs

$$\frac{\mathrm{d}y_{1,2}}{\mathrm{d}t} = f_{1,2}^{\mathrm{rhs}}(y_{1,2}(t_{i-1},t,y_{1,2}^{i-1}+c_i,\theta_{1,2}),\theta_{1,2}), \quad i \in 1,\ldots,N. \tag{8}$$

The vectors c_i denote species quantities the experimental system gets perturbed with. $f_{1,2}^{\mathrm{rhs}}$ are the right hand side functions of the ODEs of both models. $y_{1,2}^0 := y_I$ denotes the initial species concentration y_I which is the same for both models.

The vector y_{t_i} denotes the species concentration of a measurement at measurement time point t_i.

By assuming normally distributed error vectors ϵ_1^i and ϵ_2^i with zero mean and variance functions $(v_1(y_1^i,t_i,\theta_1))^2$ and $(v_2(y_2^i,t_i,\theta_2))^2$ according to models 1 and 2 one gets for the regression models

$$\begin{aligned} y_{t_i} &= y_1^i + \epsilon_1^i, \\ y_{t_i} &= y_2^i + \epsilon_2^i \end{aligned} \tag{9}$$

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 31, 2016

the two model probability distributions f_1 for model 1 and f_2 for model 2 for one measurement time point

$$f_{1,2}(y_{t_i}|t_{i-1},t_i,y_{1,2}^{i-1},\theta_{1,2},c_i)= \frac{1}{\sqrt{2\pi}|v_{1,2}^i|}e^{-\frac{1}{2}(y_{1,2}^i-y_{t_i})^T V_{1,2}^i(y_{1,2}^i-y_{t_i})}, \quad (10)$$

with $|v_{1,2}^i|=\prod_{j=1}^M v_{1,2}^j(y_{1,2}^i,t_i,\theta_{1,2})$ and diagonal matrices $V_{1,2}^i$ with diagonal entries $\left[V_{1,2}^i\right]_{jj}=(1/v_{1,2}^j(y_{1,2}^i,t_i,\theta_{1,2}))^2$ and the number of species concentrations M.

We allow for different error models for both candidate models.

The error model is dependent on the species concentrations $y_{1,2}^i$, the time t_i and possibly on parameters $\theta_{1,2}$.

For the sake of simplicity we define

$$f_{1,2}(y_{t_i}):=f_{1,2}(y_{t_i}|t_{i-1},t_i,y_{1,2}^{i-1},\theta_{1,2},c_i). \quad (11)$$

With this notation one gets both the probability distribution models for all N measurement time points

$$f_{1,2}(y)=\prod_{i=1}^N f_{1,2}(y_{t_i}). \quad (12)$$

However, by assuming such a model probability distribution one does not prohibit that at one time point several measurements can be performed.

To overcome this problem, we first extend the probability space $\mathbb{R}_i^M$ (where M is the number of species) of the i-th measurement by a one state set $\mathcal{N}_i$ to

$$\mathbb{R}^M \cup \mathcal{N}_i. \quad (13)$$

The state of the set $\mathcal{N}_i$ with $P_{\mathcal{N}_i}\in[0,1]$ denotes the probability that no measurement is performed.

Then, we introduce the Heaviside functions

$$\Theta,\Theta^*:\mathbb{R}_+ \longrightarrow [0,1] \quad (14)$$

with

$$\Theta(t_i)=\begin{cases}1 \text{ if } t_i\geq\Delta t\\ 0 \text{ if } t_i<\Delta t\end{cases}, \quad (15)$$

and

$$\Theta^*(t_i)=\begin{cases}0 \text{ if } t_i\geq\Delta t\\ 1 \text{ if } t_i<\Delta t\end{cases}. \quad (16)$$

By use of these Heaviside functions, we define the new model probabilities

$$\tilde{f}_{1,2}(y)=\prod_{i=1}^N \Theta(t_i) f_{1,2}(y_{t_i}),$$

$$\tilde{f}_{1,2}{}^i(y_{t_k};k\in\{1,\ldots,N\}\setminus\{i\}) = \Theta^*(t_i)\prod_{j=1}^{N,i\neq j}\Theta(t_j)f_{1,2}(y_{t_j}), \quad i\in 1,\ldots,N$$

$$\tilde{f}_{1,2}{}^{i_1,i_2}(y_{t_k};k\in\{1,\ldots,N\}\setminus\{i_1,i_2\})= \Theta^*(t_{i_1})\Theta^*(t_{i_2})\prod_{j=1}^{N,i_1\neq j,i_2\neq j}\Theta(t_j)f_{1,2}(y_{t_j}),$$

$$\text{with } i_{1,2}\in 1,\ldots,N, i_1<i_2, \quad (17)$$

$$\vdots$$

$$\tilde{f}_{1,2}{}^{1,\ldots,N}=\prod_{i=1}^N\Theta^*(t_i).$$

In this case $\Theta^*(t_i)$ is the probability $P_{\mathcal{N}_i}$ that no measurement is performed at the i-th measurement time. The sum over all $\tilde{f}$ now gives the probability density of one experimental outcome. By integrating the density one has to take care that the different $\tilde{f}$ are defined on different spaces. The non-disjunct conjunction of these different spaces now defines the probability space used.

By introducing these modifications, the reduced probability distribution functions $\tilde{f}_{1,2}(y)$ do not dependent on the measurements that are performed in less than Δt time after the previous measurement anymore.

To take into account that a species concentration perturbation to the system can only be performed if no measurement is done, the same procedure is repeated with the Heaviside functions

$$\Theta(t_i)=\begin{cases}1 \text{ if } t_i\geq\Delta t\\ 0 \text{ if } t_i<\Delta t\end{cases}, \quad (18)$$

$$\Theta^*(t_i)=\begin{cases}0 \text{ if } t_i\geq\Delta t\\ 1 \text{ if } t_i<\Delta t\end{cases}, \quad (19)$$

and

$$\Theta(c_i)=\begin{cases}0 \text{ if } c_i>0\\ 1 \text{ if } c_i=0\end{cases}, \quad (20)$$

$$\Theta^*(c_i)=\begin{cases}1 \text{ if } c_i>0\\ 0 \text{ if } c_i=0\end{cases}. \quad (21)$$

The model probabilities now are defined in the same way as above by replacing $\Theta(t_i)$ with $\Theta(t_i)\Theta(c_i)$. One further has to exchange $\Theta^*(t_1)$

$$\Theta^*(t_i)\rightarrow(\Theta(t_i)\Theta^*(c_i)+\Theta^*(t_i)\Theta(c_i)+\Theta^*(t_i)\Theta^*(c_i)). \quad (22)$$

Inserting the two probability models into the KL-optimality criterion

$$\mathcal{I}=\int f_1(y)\log\left\{\frac{f_1(y)}{f_2(y)}\right\}\mathrm{d}y, \quad (23)$$

one gets

$$\mathcal{I}=\sum_{i=1}^N\Bigg[\int\Theta(t_i)\Theta(c_i)f_2(y_{t_i})\log\left\{\frac{\Theta(t_i)\Theta(c_i)f_2(y_{t_i})}{\Theta(t_i)\Theta(c_i)f_1(y_{t_i})}\right\}\mathrm{d}y_{t_i}+ \left(\Theta(t_i)\Theta^*(c_i)+\Theta^*(t_i)\Theta(c_i)+\Theta^*(t_i)\Theta^*(c_i)\right)\cdot \log\left\{\frac{\Theta(t_i)\Theta^*(c_i)+\Theta^*(t_i)\Theta(c_i)+\Theta^*(t_i)\Theta^*(c_i)}{\Theta(t_i)\Theta^*(c_i)+\Theta^*(t_i)\Theta(c_i)+\Theta^*(t_i)\Theta^*(c_i)}\right\}\Bigg]. \quad (24)$$

Due to the fact that $\log(1)=0$ this simplifies to

$$\mathcal{I}=\sum_{i=1}^N\Theta(t_i)\Theta(c_i)\int f_2(y_{t_i})\cdot\log\left\{\frac{f_2(y_{t_i})}{f_1(y_{t_i})}\right\}\mathrm{d}y_{t_i}. \quad (25)$$

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 31, 2016

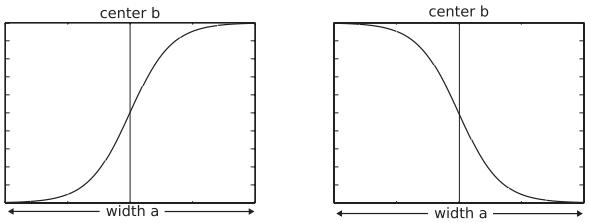

Fig. 1. Switching functions: the left switching function is used to guarantee that only one measurement is done at a time point, the right one is used to guarantee that if a perturbation is done at a time point no measurement is done at the same time point.

By inserting the normal distribution this further simplifies to

$$\mathcal{I}=\frac{1}{2}\sum_{i=1}^{N}\Theta(t_i)\Theta(c_i)\cdot \left(\sum_{j=1}^{M}\left[\frac{(v_2^j(y_2^i,t_i,\theta_2))^2+\left((y_2^i)_j-(y_1^i)_j\right)^2}{(v_1^j(y_1^i,t_i,\theta_1))^2}-2\log\left(\frac{v_2^j(y_2^i,t_i,\theta_2)}{v_1^j(y_1^i,t_i,\theta_1)}\right)\right]-M\right). \tag{26}$$

A detailed derivation can be found in the Supplementary Material.

This criterion has to be maximized with respect to the initial concentrations y_I, the measurement time points t and the system perturbations c.

For our optimal experimental design, we generally start with more measurement time points than needed. By use of the step functions, the number of measurement time points gets reduced in the sense that for $t_i < \Delta t$ the corresponding measurement time point is 'turned off'.

It should be noted that the Heaviside functions can be replaced by any appropriate switching functions i.e $\Theta+\Theta^*=1$ and Θ, $\Theta^*\in[0,1]$, e.g. continuous functions.

2.3 Numerical implementation of the optimal experimental design

To solve the maximization problem (7) numerically by applying efficient derivative-based optimizers, we replace the Heaviside functions $\Theta(t_i)$ and $\Theta(c_i)$ by continuous approximations, parameterized hyperbolic tangent functions

$$\Theta(t_i)=\frac{\left[\tanh\left(\frac{6(\Delta t_i-b_1)}{a_1}\right)+1\right]}{2} \tag{27}$$

and

$$\Theta(c_i)=\frac{\left[\tanh\left(-\frac{6(\Delta c_i-b_2)}{a_2}\right)+1\right]}{2}. \tag{28}$$

The parameters $a_{1,2}$ characterize the width of the transition region between 0 and 1. The parameters $b_{1,2}$ determine the center of the transition region (Fig. 1). By setting the parameters in an adequate way, one can construct arbitrarily close approximations of the Heaviside functions.

The optimization problem is formulated in a multiple shooting setup (see, e.g. Bock and Plitt, 1984; Bock, 1987; Stoer and Bulirsch, 2002). The idea of the multiple shooting method is to subdivide the interval $[0,T_{\text{end}}]$ into several subintervals on each of which an independent initial value problem is solved. In our implementation, each end point of a subinterval corresponds to one measurement time point. Matching conditions which enter the optimization problem as additional equality constraints assure continuity of the optimal state trajectory from one subinterval to the next one.

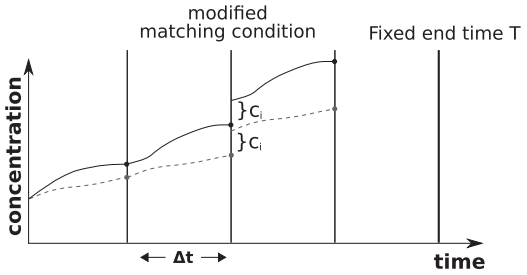

Fig. 2. Scheme of the multiple shooting setup for numerical computation of the experimental design. One dot denotes one measurement time point. The black line denotes Model 1 and the dashed gray one Model 2.

To incorporate the perturbations c, these matching conditions

$$y_{1,2}(t_{i-1},t_i,s_{\{1,2\},i-1},\theta_{1,2})-s_{\{1,2\},i}=0 \tag{29}$$

where $s_{\{1,2\},i}$ denotes the multiple shooting nodes are modified to

$$y_{1,2}(t_{i-1},t_i,s_{\{1,2\},i-1},\theta_{1,2})-s_{\{1,2\},i}=c_i. \tag{30}$$

A graphical scheme of the multiple shooting setup is shown in Figure 2.

For simplicity, we consider the homoscedastic case with equal variances, i.e. $v_1=v_2=\sigma^2$.

The overall optimization problem can be stated as

$$\max_{x_{\text{initial}},\Delta t,c,s_1,s_2}\mathcal{I}$$

$$\mathcal{I}=\sum_{i=1}^{N}\left[y_1(0,\Delta t_i;s_{1,i},\theta_1)-y_2(0,\Delta t_i;s_{2,i},\theta_2)\right]^2\Theta(\Delta t_i)\Theta(c_i)$$

subject to

$$\begin{aligned}
&\frac{\mathrm{d}y_1(t)}{\mathrm{d}t}=f_1^{rhs}(y_1(t)),\qquad \frac{\mathrm{d}y_2(t)}{\mathrm{d}t}=f_2^{rhs}(y_2(t)),\\
&s_{1,0}=y_0,\qquad s_{2,0}=y_0,\\
&y_1(0,\Delta t_i;s_{1,i},\theta_1)-s_{1,(i+1)}=c_i,\quad i\in[0,\ldots,N-2],\\
&y_2(0,\Delta t_i;s_{2,i},\theta_2)-s_{2,(i+1)}=c_i,\quad i\in[0,\ldots,N-2],\\
&y_{0_{\min}}\le y_0\le y_{0_{\max}},\\
&0\le\Delta t\le t_{\max},\\
&0\le c\le c_{\max},\\
&s_{1_{\min}}\le s_1\le s_{1_{\max}},\\
&s_{2_{\min}}\le s_2\le s_{2_{\max}},\\
&\sum_{i=1}^{N}\Delta t_i=T_{\text{end}}.
\end{aligned} \tag{31}$$

Numerical optimization is performed by the interior point package IPOPT (Wächter, 2002; Wächter and Biegler, 2006) using the linear solver MA27 (HSL, 2007). The numerical integration as well as the

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 31, 2016

sensitivity generation within a multiple shooting discretization is done by the CVODES integrator package (Serban and Hindmarsh, 2005). All derivatives are calculated by automatic differentiation using CppAD (Bell and Burke, 2008).

3 NUMERICAL RESULTS

3.1 Allosteric models for glycolytic oscillations

As a test case for model discrimination, we implemented the following models describing glycolytic oscillations (Goldbeter, 1996).

- Model 1 describes an allosteric model with positive feedback and linear product sink.
  The differential equations are given by

$$\frac{d\alpha}{dt} = \nu - \sigma\phi(\alpha,\gamma),$$
$$\frac{d\gamma}{dt} = q\sigma\phi(\alpha,\gamma) - k_s\gamma,$$
$$\phi(\alpha,\gamma) = \frac{\alpha(1+\alpha)(1+\gamma)^2}{L+(1+\alpha)^2(1+\gamma)^2}.$$

- Model 2 describes an allosteric model with positive feedback in the absence of cooperativity when the product sink is represented by Michaelis–Menten kinetics.
  The differential equations are given by

$$\frac{d\alpha}{dt} = \nu - \phi(\alpha,\gamma),$$
$$\frac{d\gamma}{dt} = q\phi(\alpha,\gamma) - \frac{r_s\gamma}{\mu+\gamma},$$
$$\phi(\alpha,\gamma) = \frac{\alpha(1+\gamma)}{L+(1+\alpha)(1+\gamma)}.$$

The species concentration of the substrate is denoted by α and the concentration of the product is denoted by γ. For both models the inflow parameter ν is the same. It is regarded as known and was set to $\nu = 0.22$. It represents the inflow of the substrate to the experimental system, here a tank reactor. The parameter sets (σ, q, k_s, L) of Model 1 and (q, r_s, μ, L) of Model 2 are independent of each other and are regarded as unknown.

Due to the non-availability of experimental data, we simulate data by use of Model 1. Note that σ of the parameter set of Model 1 and the variance σ^2 are not the same. To generate a measurement we add a normally distributed error term with zero mean and variance $\sigma^2 = 1$ to the simulated data. The used parameter set is $(8.89 \times 10^{-1}, 2.16, 1.11 \times 10^{-1}, 1.76 \times 10^4)$. We simulated in a time span of $\Delta T = 400$ with initial concentrations of $\alpha = 1.47 \times 10^1$ and $\gamma = 2.21$. After 10 numerical integration time steps, a simulated measurement is generated. Both models were fitted to the simulated data to determine the model parameters as well as the initial species concentrations. The fitting was performed by standard least squares minimization implemented in the same numerical optimization framework as the model discrimination. The results of the fitting procedure are shown in Figure 3. The result after the fitting procedure for Model 1 is the parameter set $(8.79 \times 10^{-1}, 2.03, 1.09 \times 10^{-1}, 1.61 \times 10^4)$ and initial concentrations $\alpha = 1.50 \times 10^1$ and $\gamma = 1.92$. The result of Model 2 is the parameter set $(6.47, 3.33, 5.11, 2.44 \times 10^2)$ and initial concentrations $\alpha = 1.47 \times 10^1$ and $\gamma = 1.16$.

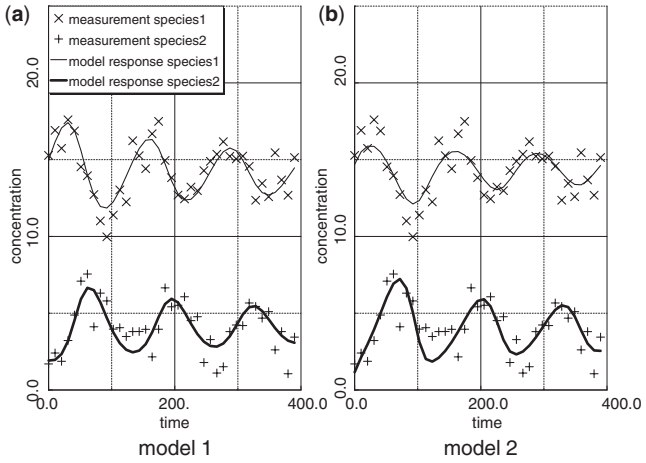

Fig. 3. The parameter fit of Model 1 (**a**) and Model 2 (**b**) to the generated data are shown.

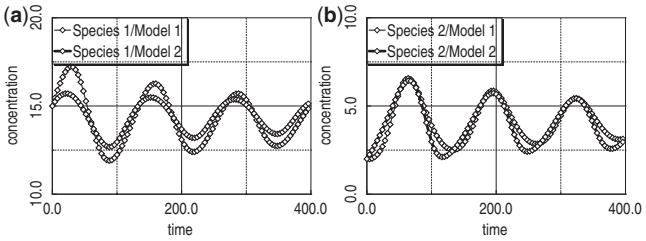

Fig. 4. The model responses are shown before the optimization procedure, (**a**) species 1, (**b**) species 2. One square represents one measurement time point.

3.2 Results of model discrimination procedure

With the so generated parameter estimates of Models 1 and 2, we start the model discrimination algorithm on an overall time span $T_{\text{end}} = 400$. We insert 100 equally spaced possible measurement points. The initial species concentrations to start the procedure are set to $\alpha = 15$ and $\gamma = 2$ and are restricted to be less than 25. The parameters of the switching function $\Theta(\Delta t)$ are chosen as $a_1 = 20$ and $b_1 = 10$. With these parameters we expect the minimal time span between two measurement points to be bigger than 10 time steps.

The parameters of the switching function $\Theta(c)$ are chosen as $a_2 = 0.05$ and $b_2 = 0.025$. We first calculate an experimental design as the solution of problem 31 where no perturbations on the system are allowed. Therefore, the design is optimal with respect to the initial concentrations and the measurement time points. In Figure 4, the situation before the optimization procedure is shown. One can easily see that the model response of Models 1 and 2 are very similar. The result after the optimization procedure is shown in Figure 5a. The measurement time points were reduced from $N = 100$ to $N = 31$. Both model responses are well separated. The computational time to calculate the experimental design was 52 s on one core of a Intel(R) Xeon(R) X5460 CPU with 3.16 GHz. In a second run we add at five measurement points the ability to perturb the system by adding species quantities less or equal to 10 for each species. We further restricted the sum of all initial species concentrations and species perturbations to be less than 100. The results are shown in 5b.

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 31, 2016

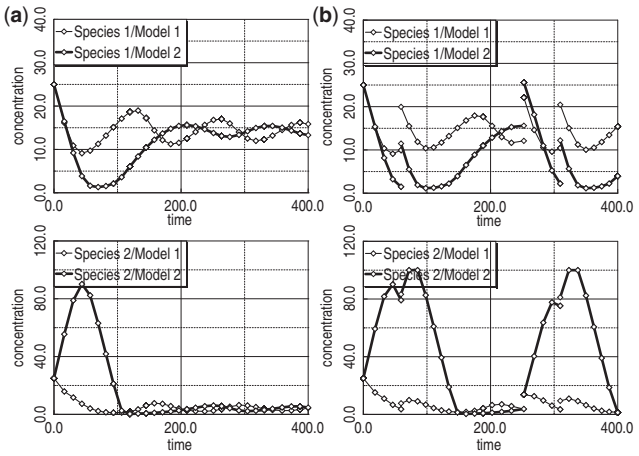

Fig. 5. The discriminating optimal design is shown, (**a**) with respect to the initial species concentrations and optimal time points and (**b**) additionally with optimal perturbations to the system. One square represents one measurement time point.

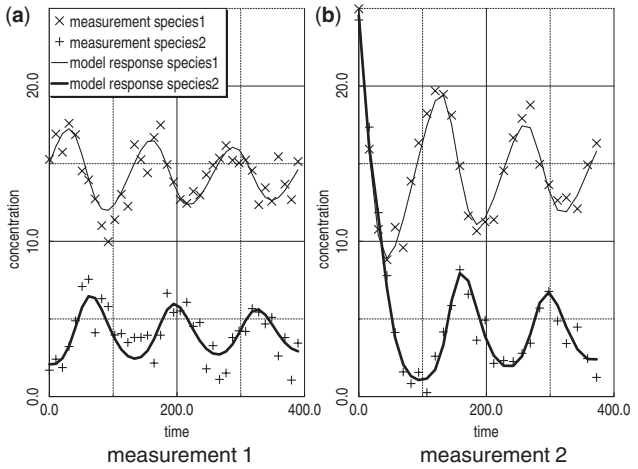

Fig. 6. The coincidental parameter fit of Model 1 to the first generated dataset (**a**) and second generated dataset (**b**) is shown.

Now both model responses are well separated over the entire time span. The number of measurement time points is reduced to $N = 27$. The computational time to calculate the experimental design was 118 s on one core of a Intel(R) Xeon(R) X5460 CPU with 3.16 GHz.

By use of the optimal design calculated with respect to the initial species and optimal time points, we again generate data with the correct Model 1 with correct parameters according to the experimental design. After that we fitted both models coincidentally to the first generated samples and the second generated samples. In Figure 6, one can see that Model 1 naturally fits both generated samples as expected. To test for a local optimum Model 2 was fitted 50 times with uniform random distributed initial values in the range of $[0, \ldots, 100000]$ for each parameter. The best fit is shown in Figure 7. As one can see Model 2 does not coincidentally fit all measurements and therefore Model 1 should be favored. In that sense we are able to clearly discriminate between Model 1 and Model 2 by the use of the optimal design. An effective way to search for a global

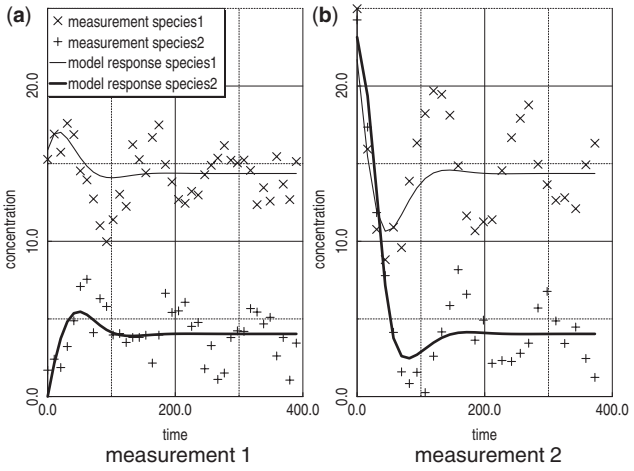

Fig. 7. The coincidental parameter fit of Model 2 to the first generated dataset (**a**) and second generated dataset (**b**) is shown.

optimum is the global scatter search method, see for example, Egea *et al.* (2007). A further numerical example for the situation that the model to simulate the data is different from both candidate models can be found in the Supplementary Material.

4 CONCLUSION

By use of the KL-optimality we derived a model discrimination criterion that might be more appropriate for dynamic time series experiments performed by biologist than the existing approaches, e.g the continuous time design (Atkinson and Fedorov, 1975; Cooney and McDonald, 1995; Kremling *et al.*, 2004). We included optimal initial species concentrations, optimal measurement time points for a one run experiment where the time span between two measurements have to be larger than a given Δt and optimal perturbations to the system with respect to the perturbation time and the amount of perturbing species quantities. We further exemplified our method on a simulated example for allosteric models for glycolytic oscillations by performing the sequential run of making an experiment (a simulated one in our test case) and afterwards calculating an optimally discriminating design. After only one iteration of an experimental design, we can clearly discriminate between Model 1 and Model 2 which shows the efficiency of our method.

Funding: Freiburg Initiative for Systems Biology (FRISYS); the Freiburg part of the BMBF FORSYS systems biology initiative; Freiburg excellence cluster Centre for Biological Signalling Studies (BIOSS).

Conflict of Interest: none declared.

REFERENCES

Apgar,J.F. *et al.* (2008) Stimulus design for model selection and validation in cell signaling. *PLoS Comput. Biol.*, **4**, e30.

Atkinson,A.C. and Fedorov,V.V. (1975) The design of experiments for discriminating between two rival models. *Biometrika*, **62**, 57–70.

Balsa-Canto,E. *et al.* (2008) Computational procedures for optimal experimental design in biological systems. *IET Syst. Biol.*, **2**, 163–172.

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 31, 2016

Bauer,I. *et al.* (2000). Numerical methods for optimum experimental design in DAE systems. *J. Comput. Appl. Math.*, **120**, 1–25.

Bell,B.M. and Burke,J.V. (2008) Algorithmic differentiation of implicit functions and optimal values. In Bischof,C.H. *et al.* (eds) *Advances in Automatic Differentiation*. Springer, pp. 67–77.

Bernacki,J.P. and Murphy,R.M. (2009) Model discrimination and mechanistic interpretation of kinetic data in protein aggregation studies. *Biophys. J.*, **96**, 2871–2887.

Bock,H.G. and Plitt,K.J. (1984) A multiple shooting algorithm for direct solution of optimal control problems. In *Proceedings of the Ninth IFAC World Congress, Budapest*. Pergamon, Oxford.

Bock,H.G. (1987) Randwertproblemmethoden zur Parameteridentifizierung in Systemen nichtlinearer Differentialgleichungen. In *Bonner Mathematische Schriften*, Vol. 183. University of Bonn, Bonn, Germany.

Burnham,K.P. and Anderson,D.R. (2002) *Model Selection and Multimodel Inference: A Practical Information-theoretic Approach*, 2nd edn., Springer, New York, USA.

Cooney,M.J. and McDonald,K.A. (1995) Optimal dynamic experiments for bioreactor model discrimination. *Appl. Microbiol. Biot.*, **43**, 826–837.

Egea,J.A. *et al.* (2007) Scatter search for chemical and bio-process optimization. *J. Global Optim.*, **37**, 481–503.

Eguchi,S. and Copas,J. (2006) Interpreting Kullback-Leibler divergence with the Neyman-Pearson lemma. *J. Multivar. Anal.*, **97**, 2034–2040.

Goldbeter,A. (1996) *Biochemical Oscillations and Cellular Rhythms: The Molecular Bases of Periodic and Chaotic Behaviour*. Cambridge University Press, Cambridge, UK.

Horn,R. (1987) Statistical methods for model discrimination. applications to gating kinetics and permeation of the acetylcholine receptor channel. *Biophys. J.*, **51**, 255–263.

HSL (2007) A collection of fortran codes for large-scale scientific computation. Available at http://www.hsl.rl.ac.uk.

Jain,R. *et al.* (2006) Investigation of bacteriophage ms2 viral dynamics using model discrimination analysis and the implications for phage therapy. *Biotechnol. Progr.*, **22**, 1650–1658.

Körkel,S. *et al.* (1999) A sequential approach for nonlinear optimum experimental design in DAE systems. In Keil,F. *et al.* (eds) *Scientific Computing in Chemical Engineering II*, Vol. 2. Springer, Berlin, pp. 338–345.

Kremling,A. *et al.* (2004) A benchmark for methods in reverse engineering and model discrimination: problem formulation and solutions. *Genome. Res.*, **14**, 1773–1785.

Kreutz,C. and Timmer,J. (2009) Systems biology: experimental design. *FEBS J.*, **276**, 923–942.

Lacey,L. and Dunne,A. (1984) The design of pharmacokinetic experiments for model discrimination. *J. Pharmacokinet. Phar.*, **12**, 351–365.

Lindsey,J.K. (1996) *Parametric Statistical Inference*. Oxford University Press, New York, USA.

López-Fidalgo,J. *et al.* (2007) An optimal experimental design criterion for discriminating between non-normal models. *J. R. Stat. Soc. Ser. B*, **69**, 231–242.

Myung,J.I. and Pitt,M.A. (2009) Optimal experimental design for model discrimination. *Psychol. Rev.*, **116**, 499–518.

Serban,R. and Hindmarsh,A. (2005) CVODES: the sensitivity-enabled ODE solver in SUNDIALS. In *Proceedings of IDETC/CIE 2005*. Long Beach, CA.

Stewart,W.E. *et al.* (1998) Discrimination and goodness of fit of multiresponse mechanistic models. *AIChE. J.*, **44**, 1404–1412.

Stoer,J. and Bulirsch,R. (2002) *Introduction to Numerical Analysis*. Number 12 in *Texts in Applied Mathematics*, 3rd edn. Springer, New York.

Stricker,C. *et al.* (1994) Statistical analysis of synaptic transmission: model discrimination and confidence limits. *Biophys. J. R. Stat. Soc. Ser. B*, **67**, 532–547.

Takors,R. *et al.* (1997) Experimental design for the identification of macrokinetic models and model discrimination. *Biotechnol. Bioeng.*, **56**, 564–576.

Timmer,J. *et al.* (2004) Modeling the nonlinear dynamics of cellular signal transduction. *Int. J. Bifurcat. Chaos*, **14**, 2069–2079.

Uciński,D. and Bogacka,B. (2004) Heteroscedastic T-optimum designs for multiresponse dynamic models. In Bucchianico,A.D. and Lauter,H. (eds) *Proceedings of the 7th International Workshop on Model-Oriented Design and Analysis*, Physica Verlag Germany, Heidelberg, pp. 191–199.

Wächter,A. (2002) An interior point algorithm for large-scale nonlinear optimization with applications in process engineering. PhD Thesis, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA.

Wächter,A. and Biegler,L.T. (2006) On the implementation of a primal-dual interior point filter line search algorithm for large-scale nonlinear programming. *Math. Program.*, **106**, 25–57.

Downloaded from http://bioinformatics.oxfordjournals.org/ at :: on August 31, 2016