Optimal design of stimulus experiments for robust discrimination of biochemical reaction networks

R. J. Flassig[1,*] and K. Sundmacher[1,2]

[1]Otto-von-Guericke University, Process Systems Engineering, Universitätsplatz 2, [2]Max Planck Institute for Dynamics of Complex Technical Systems, Process Systems Engineering Sandtorstr. 1, D-39106 Magdeburg, Germany

Associate Editor: Jonathan Wren

ABSTRACT

Motivation: Biochemical reaction networks in the form of coupled ordinary differential equations (ODEs) provide a powerful modeling tool for understanding the dynamics of biochemical processes. During the early phase of modeling, scientists have to deal with a large pool of competing nonlinear models. At this point, discrimination experiments can be designed and conducted to obtain optimal data for selecting the most plausible model. Since biological ODE models have widely distributed parameters due to, e.g. biologic variability or experimental variations, model responses become distributed. Therefore, a robust optimal experimental design (OED) for model discrimination can be used to discriminate models based on their response probability distribution functions (PDFs).

Results: In this work, we present an optimal control-based methodology for designing optimal stimulus experiments aimed at robust model discrimination. For estimating the time-varying model response PDF, which results from the nonlinear propagation of the parameter PDF under the ODE dynamics, we suggest using the sigma-point approach. Using the model overlap (expected likelihood) as a robust discrimination criterion to measure dissimilarities between expected model response PDFs, we benchmark the proposed nonlinear design approach against linearization with respect to prediction accuracy and design quality for two nonlinear biological reaction networks. As shown, the sigma-point outperforms the linearization approach in the case of widely distributed parameter sets and/or existing multiple steady states. Since the sigma-point approach scales linearly with the number of model parameter, it can be applied to large systems for robust experimental planning.

Availability: An implementation of the method in MATLAB/AMPL is available at http://www.uni-magdeburg.de/ivt/svt/person/rf/roed.html.

Contact: flassig@mpi-magdeburg.mpg.de

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

1 INTRODUCTION

Mathematical models of complex biological processes provide the basis for systems understanding. They are of vital importance for generating predictions of systems behavior based on hypothesized mechanisms. In the case of limited experimental access to biological systems, models can help to find missing links and provide a tool to aggregate existing knowledge and data. In this work, we focus on mathematical models in the form of coupled ordinary differential equations (ODEs), which are being used to describe dynamics of biochemical reaction networks, e.g. signal transduction, metabolic or genetic regulation, on a deterministic, (semi-)mechanistic basis. Here, scientists are often facing limited or even contradicting knowledge about the underlying mechanisms, confined experimental possibilities, large biological variability as well as measurement noise. This leads to largely distributed model parameter sets, which in combination with several plausible alternative model structures renders model-based prediction highly uncertain. Therefore, model-based experimental design (e.g. optimal stimulus or additional experimental readout selection) is used to generate experiments that yield optimal experimental data to (i) reduce the spread in the model parameters (=optimal parameter identification) and/or (ii) reduce the pool of plausible models (=optimal model discrimination).

Much work on optimal experimental stimulus design (OESD) for biological systems focuses on information maximization with respect to parameter identification (Balsa-Canto *et al.*, 2008; Bandara *et al.*, 2009; Heine *et al.*, 2008; Raue *et al.*, 2010; Schenkendorf *et al.*, 2009a). Here, for a given pool of plausible ODE models, OESD is used to find experimental conditions that reduce the parameter uncertainties and thus model response variabilities. The methods used to quantify parameter uncertainties and model response variabilities include linearization, sigma-points (Julier *et al.*, 2000), profile likelihood (Raue *et al.*, 2009) and Markov chain Monte Carlo (MCMC, Geyer, 1992; Vanlier *et al.*, 2012). An OESD aiming at model discrimination for biological systems has been addressed by few authors (Apgar *et al.*, 2008; Melykuti *et al.*, 2010; Skanda and Lebiedz, 2010). Although measurement noise has been included, a rigorous consideration of model response variabilities due to distributed parameters has been missing so far. For chemical reaction kinetics and biotechnology, there exists some work on designing dynamic stimuli for the purpose of model discrimination (Asprey and Macchietto, 2002; Box and Hill, 1967; Kremling *et al.*, 2004). As has been illustrated, the consideration of model response variabilities strongly improves the designed experiments and experimental data quality (Chen and Asprey, 2003; Donckels, 2012; Michalik *et al.*, 2010). Therefore, linearization of the system's parameter mapping has been used. However, by using the so-called Sigma-Point method (Julier *et al.*, 2000), (Heine *et al.*, 2008), (Schenkendorf *et al.*, 2009a) showed that the performance

*To whom correspondence should be addressed.

of a linear OED for best parameter estimation is rather poor for nonlinear systems having widely distributed parameters.

An experimental design aimed at model discrimination is typically generated at a point where existing data do not provide further discriminative information for a pool of competing, validated and identifiable models. With these models, robust experimental designs can be generated, which presumably yield data with optimal discrimination information. Since predictions with ODE models depend on the model parameters, identifiability is important as it ensures the existence of a unique solution to the parameter estimation problem and, consequently, unique model predictions (Raue *et al.*, 2009). A non-identifiable model would yield non-unique model predictions under altered experimental design conditions, as there exists a set of several solutions to the parameter estimation problem. Robustness of the experimental design is achieved by considering (i) pure uncertainty about the model itself, (ii) distributed model predictions that arise from distributed model parameters and (iii) design variabilities (e.g. variations of the applied stimulus) during the conduction of the experiment.

In this work, part (ii) of design robustification for nonlinear models is considered, focusing on computational efficient and accurate estimation of the nonlinear propagated parameter probability distribution function (PDF) in an optimal control framework. We suggest using the sigma-point method as an alternative to the classical linearization approach. Therefore, both approaches are presented, applied and compared in the light of OESD for robust model discrimination, assuming perfect experimental conduction. In the following section, we describe the essential parts of our design approach, including (i) ODE modeling of biological reaction networks with and extended interpretation of distributed determinism, (ii) the model overlap as a PDF-based discrimination criterion, (iii) linearization and sigma-point methods and (iv) two numerical approaches to solve optimal control problems. Then, using two numerical examples, we formulate the optimization problem and benchmark the linearization approach against the sigma-point approach (Section 3).

2 METHODS

2.1 Dynamic modeling of biochemical reaction systems

ODEs provide the modeling basis to describe the dynamics of biochemical reaction networks. The dynamics of the internal states $\mathbf{x}(t, \mathbf{u}(t), \theta_x) \in \mathbb{A}_\mathbf{x} \subset \mathbb{R}^{n_x}$, e.g. protein concentrations, is determined by the solution of an initial value problem of the form

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \theta_\mathbf{x}) \tag{1}$$

with initial system states $\mathbf{x}(t_0) = \mathbf{x}_0$ and right-hand side function $\mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \theta_\mathbf{x})$ describing biologic interaction mechanisms, which depends on the system states $\mathbf{x}(t)$, (multiple) inputs $\mathbf{u}(t)$ (=stimulus) and kinetic parameter set $\theta_\mathbf{x}$. Assuming $\mathbf{f}$ to be Lipschitz in $\mathbf{x}(t)$, $\mathbf{u}(t)$ and continuous in t, the readout variables are determined by

$$\mathbf{y}_{\mathrm{sim}}(t, \theta) = \mathbf{g}(\mathbf{x}(t, \theta_\mathbf{x}), \theta_\mathbf{y}), \tag{2}$$

where the function $\mathbf{g}$—assumed to be sufficiently smooth—relates the internal system states to the readouts of the experiment with corresponding readout parameters $\theta_\mathbf{y}$, which together with dynamic parameters and initial conditions are merged into the model parameter vector $\theta = [\theta_\mathbf{x}, \theta_\mathbf{y}]^\mathrm{T}$, with redefined dynamic parameter vector $\theta_\mathbf{x} \equiv [\theta_\mathbf{x}, \mathbf{x}_0]^\mathrm{T}$.

The dynamic model defined by Equations (1) and (2) can be understood as a time-dependent mapping from the model parameter space $\mathbb{A}_{\theta_\mathbf{x}} \times \mathbb{A}_{\theta_\mathbf{y}} = \mathbb{A}_\theta \subset \mathbb{R}^{n_\theta}$ to the model output space $\mathbb{A}_\mathbf{y} \subset \mathbb{R}^{n_y}$,

$$\mathbf{h} : \mathbb{R} \times \mathbb{A}_\theta \to \mathbb{A}_\mathbf{y} \tag{3}$$

$$(t, \theta) \mapsto \mathbf{h}(t, \theta) = \mathbf{y}_{\mathrm{sim}}(t, \theta). \tag{4}$$

2.2 Distributed determinism

Although biological systems might follow deterministic rules, repeated measurements, even though with very accurate measurement techniques, will yield different results. The reasons for that are manyfold. Additionally to unavoidable measurement errors, biologic variability, i.e. systems with intrinsically distributed parameters, can induce a large spread in the transient dynamics and stationary behavior. In the case of the existence of multiple steady states, this spreading effect might even be more pronounced. Varying parameters during the measuring procedure and local parameter perturbations by non-stationary noise also contribute to a distributed measurement signal (Lorenz *et al.*, 2007). Complex, nonlinear models of biological systems might also behave chaotic, further contributing to distributed response measurements. Thus, the conventional sharp, deterministic system representation needs to be extended by the notion of distributed determinism, i.e. although the system might completely be deterministic, its perceived signals are distributed realizations of the underlying deterministic mechanisms. This can be done by considering the parameters, and hence, the model responses as random variables Θ and $\mathbf{Y}$, respectively, each characterized by a PDF. We point out that, within this interpretation, the system and hence the model is assumed to naturally possess distributed parameters. Consequently, a distributed response is not solely explained by additive measurement noise but also by other sources of variations, which may be represented as distributed parameter sets.

Let the model parameters be distributed according to some well-defined PDF $\rho_\Theta(\theta)$, with $\theta \in \mathbb{A}_\theta$ being a realization of Θ. The PDF of the random model response $\mathbf{Y}$ at time t can be derived from the normalized integral over all possible parameter and corresponding response realizations, weighted with the parameter PDF, i.e.

$$\rho_\mathbf{Y}(\mathbf{y}, t) = \frac{\varrho_\mathbf{Y}(\mathbf{y}, t)}{||\varrho_\mathbf{Y}(\mathbf{y}, t)||_1}, \tag{5}$$

with

$$\varrho_\mathbf{Y}(\mathbf{y}, t) = \int_{\mathbb{A}_\theta} \mathbf{1}_\mathbf{y}[\mathbf{h}(t, \theta)]\rho_\Theta(\theta)\mathrm{d}\theta, \tag{6}$$

where $\mathbf{1}_\mathbf{y}[\mathbf{h}(t, \theta)]$ represents the indicator function

$$\mathbf{1}_\mathbf{y}[\mathbf{h}(t, \theta)] = \begin{cases} 1, & \text{if } \mathbf{y} = \mathbf{h}(t, \theta) \\ 0, & \text{else.} \end{cases} \tag{7}$$

The normalization employs the L_1-norm with respect to $\mathbf{y} \in \mathbb{A}_\mathbf{y}$. Note that $\mathbf{y}$ represents an arbitrary possible realization of $\mathbf{Y}$ in $\mathbb{A}_\mathbf{y}$, whereas $\mathbf{h}(t, \theta)$ describes the model response at time t for a fixed stimulus time course $\mathbf{u}(t_0 \to t)$ given parameter realization θ. Consequently, for every single point in time, the shape of Equation (5) is determined by the parameter PDF, the choice of model [Equations (1) and (2)] and stimulus time course (=experimental design). Consequently, the discrimination process can be robustified by discriminating competing models based on their model response PDFs [Equation (5)], accounting for variabilities in the parameters and model-specific mapping to the response space.

2.3 Robust design criteria for model discrimination

Here, we present the model overlap as a robust discrimination criterion, measuring dissimilarities of model response PDFs used to rate the discriminative power of a design during optimization. We define the *general*

model overlap as the probability product kernel used in vector machine learning to measure statistical distances for the sake of discriminative learning (Jebara *et al.*, 2004). Following Jebara *et al.*, (2004), the probability product kernel of two multivariate PDFs $\rho^p_{\mathbf{Y}|A}, \rho^p_{\mathbf{Y}|B} \in L_2(\mathbb{A}_\mathbf{y})$ is defined as the scalar product

$$\Phi(t, \mathbf{u}(t)) = \int_{\mathbb{A}_\mathbf{y}} \rho_{\mathbf{Y}|A}(\mathbf{y}, t, \mathbf{u}(t))^p \rho_{\mathbf{Y}|B}(\mathbf{y}, t, \mathbf{u}(t))^p \mathrm{d}\mathbf{y} \qquad (8)$$

with the densities being raised to some power $p \in \mathbb{R}^+ \backslash \{0\}$. It provides a bounded, positive-definite measure (Jebara *et al.*, 2004) of similarity between distributions on the set $\mathbb{A}_\mathbf{y}$, whereas the parameter p controls the weighting of regions with small versus large densities.

From this, the average overlap of the time course is

$$\langle \Phi(\mathcal{D}) \rangle_t = \frac{1}{n_t} \sum_{k=1}^{n_t} \Phi(t_k, \mathbf{u}(t_k)), \qquad (9)$$

where $\mathcal{D} \in \mathbb{D} = \mathbb{T} \times \mathbb{U} \times \mathbb{Y}$ represents the experimental design within the feasible design region $\mathbb{D}$, encompassing for instance selection of discrete measurement time points $t_k \in \mathbb{T}$, stimulus design $\mathbf{u}(t) \in \mathbb{U}$ and readout design $\mathbf{g} \in \mathbb{Y}$. Using the model response PDFs $\rho_{\mathbf{Y}|m}(\mathbf{y}, t, \mathbf{u}(t))$ from Equation (5) for two competing models, $m \in \{A, B\}$ in Equation (9) provides us with a measure of average model dissimilarities. For $p = 1$, the overlap is the expected model response probability of model A under model B and vice versa [Equation (8)]. In this case, assuming one of the models to be true, the overlap yields the expected likelihood of the other model depending on the experimental design $\mathcal{D}$. Consequently, an optimal model discrimination design $\mathcal{D}_†$ minimizes Equation (9). Singh (1998) proposed to use Equation (8) with $p = 1/2$—known as Bhattacharyya's affinity between distributions—for discriminating nonlinear regression models. It is closely related to Hellinger's distance, which represents a symmetrized approximation to the Kullback–Leibler divergence (Topsoe, 2000). In this work, we use the overlap as defined in Equation (9) with $p = 1$, which we refer to as *model overlap*, as this directly represent the time-averaged, expected likelihood of one model under the other. In this form, it has been applied by Lorenz *et al.*, (2007) to discriminate dynamic models based on the distributional fit performance. In Schenkendorf *et al.*, (2009b), it is used to optimize an initial condition for a substrate uptake model to discriminate two competing kinetic approaches.

2.4 Estimation of nonlinear PDF mapping

If the solution $\mathbf{h}(t, \boldsymbol{\theta})$ can be obtained in closed form, it is straightforward to derive the model response PDF for a given parameter PDF using Equation (5). However, in most of the cases, the model response $\mathbf{h}(t, \boldsymbol{\theta})$ for a specific parameter realization is obtained by numerical integration. Here, besides random sampling techniques based on Monte-Carlo simulations, the approximate model response PDF may also be obtained by deterministic sampling, e.g. by simple enumeration, optimized Latin hypercubes of the parameter space and application of Equation (5). For an infinite number of samples, the true model response PDF can be constructed from these samples, which can be used for a subsequent evaluation of Equation (9) to judge the quality of a given design. Such procedures become computational inefficient for increasing number of model parameters and cannot be used in an optimization framework. Therefore, an approximation has to be made. From initial data, one may obtain accurate estimates for the true parameter PDF, e.g. by MCMC sampling, which can be approximated by unimodal normal PDFs $\rho_\Theta(\boldsymbol{\theta}) \simeq \mathcal{N}(\mu_\Theta, \Sigma_\Theta)$, possibly multivariate. The model response PDF can also be approximated by $\rho_{\mathbf{Y}|m}(\mathbf{y}, t, \mathbf{u}(t)) \simeq \mathcal{N}(\mu_{\mathbf{Y}|m}(t, \mathbf{u}(t)), \Sigma_{\mathbf{Y}|m}(t, \mathbf{u}(t)))$. Then, the integration in Equation (8) can be performed to yield the approximate overlap

$$\Phi^\mathcal{N}(t, \mathbf{u}(t)) = \left(|\Sigma_{\mathbf{Y}|A}| |\Sigma_{\mathbf{Y}|B}| \left| \Sigma_{\mathbf{Y}|AB}^{-1} \right| e^{I_{\mathbf{Y}|A} + I_{\mathbf{Y}|B}} \right)^{-1/2}, \qquad (10)$$

where

$$I_{\mathbf{Y}|m} = \mu_{\mathbf{Y}|m}^T \Sigma_{\mathbf{Y}|m}^{-1} \mu_{\mathbf{Y}|m} - \mu_{\mathbf{Y}|m}^T \Sigma_{\mathbf{Y}|m}^{-1} \Sigma_{\mathbf{Y}|AB} \Sigma_{\mathbf{Y}|m}^{-1} \mu_{\mathbf{Y}|m},$$

$$\Sigma_{\mathbf{Y}|AB} = \left[\Sigma_{\mathbf{Y}|A}^{-1} + \Sigma_{\mathbf{Y}|B}^{-1} \right]^{-1}; \; |\Sigma_{\mathbf{Y}|m}| = (2\pi)^{n_\mathbf{Y}} \det(\Sigma_{\mathbf{Y}|m})/2.$$

Here, $m = \{A, B\}$ represents the model index, $\mu_{\mathbf{Y}|m} \equiv \mu_{\mathbf{Y}|m}(t, \mathbf{u}(t))$ the $n_\mathbf{Y}$-dimensional true mean and $\Sigma_{\mathbf{Y}|m} \equiv \Sigma_{\mathbf{Y}|m}(t, \mathbf{u}(t))$ the true variance-covariance matrix of the corresponding model response PDFs, which both depend on the model structure m, measurement time point t, stimulus $\mathbf{u}(t)$ and readout selection. For details about the derivation, see for instance, Jebara *et al.* (2004). For n_t discrete measurement time points t_k, the approximate mean overlap is then

$$\langle \Phi^\mathcal{N}(\mathcal{D}) \rangle_t = \frac{1}{n_t} \sum_{k=1}^{n_t} \Phi^\mathcal{N}(t_k, \mathbf{u}(t_k)). \qquad (11)$$

This approximation dramatically reduces the computational costs as only the two first statistical moments (i.e. expectation and variance-covariance) need to be estimated. The task of solving Equation (5) to obtain model-response PDFs and subsequent integration of Equation (11) to evaluate the discriminative power of a given design in an optimization framework is then reduced to estimate the time course of mean vector $\mu_{\mathbf{Y}|m}(t, \mathbf{u}(t))$ and variance–covariance matrix $\Sigma_{\mathbf{Y}|m}(t, \mathbf{u}(t))$ of two model response PDFs for given parameter expectation $\mu_{\Theta|m}$ and variance–covariance $\Sigma_{\Theta|m}$ (Fig. 1). As true mean and variance–covariance of the parameters are unknown, these are replaced by their estimates, i.e. $\mu_\Theta \to \mathbf{E}[\Theta]$ and $\Sigma_\Theta \to \mathbf{C}[\Theta]$, which we will do in the following. Note that for skewed PDFs one should apply a transformation, e.g. Box–Cox or inverse hyperbolic sine transformation to achieve normality of the PDF (Box, 1964; Johnson, 1949). In this way, our presented robust OESD method is not restricted to normal PDFs only.

Estimates of response expectation and variance–covariance can be obtained by linearizing the system at additional computational costs that scale linear with the number of parameters using forward-sensitivity analysis. But this approach can become suboptimal or yield even misleading designs (Section 3.2). On the additional expense of $\mathcal{O}(n_\theta^2)$ estimates may be improved by a quadratic approximation of the system, which may become infeasible for larger systems, as do Monte Carlo-based approaches. Worst-case approaches yielding a minimax design have also been proposed to take model response variabilities into account (Walter and Pronzato, 1997; Skanda and Lebiedz, 2012). However, these are in general NP-hard (Du and Pardalos, 1995). The sigma-point method has an additional computational expense of $\mathcal{O}(n_\theta)$, which compares to linearization.

2.4.1 Estimation based on linearization
The classical approach to estimate model response variabilities is linearization of the nonlinear model mapping $\mathbf{h}(t, \boldsymbol{\theta})$ with respect to the parameters. The linearization of the model response is given by applying the chain rule to Equation (2)

$$\mathbf{y}^\mathcal{L}(t, \boldsymbol{\theta}) = \mathbf{h}(t, \mathbf{E}[\Theta]) + \mathbf{S}(t, \mathbf{y})^T \big|_{\boldsymbol{\theta} = \mathbf{E}[\Theta]} (\boldsymbol{\theta} - \mathbf{E}[\Theta]), \qquad (12)$$

with response sensitivity matrix $\mathbf{S}(t, \mathbf{y}) = \frac{\partial \mathbf{h}(t, \mathbf{x})}{\partial \mathbf{x}} \mathbf{S}_\mathbf{x}(t, \mathbf{x}) + \frac{\partial \mathbf{h}(t, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ and state-sensitivity matrix $\mathbf{S}_\mathbf{x}(t, \mathbf{x}) = \frac{\partial \mathbf{x}}{\partial \boldsymbol{\theta}}$, which can be obtained by solving

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathbf{S}_\mathbf{x}(t, \mathbf{x}) \bigg|_{\boldsymbol{\theta} = \mathbf{E}[\Theta]} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \mathbf{S}_\mathbf{x}(t, \mathbf{x}) \bigg|_{\boldsymbol{\theta} = \mathbf{E}[\Theta]} + \frac{\partial \mathbf{f}}{\partial \boldsymbol{\theta}} \bigg|_{\boldsymbol{\theta} = \mathbf{E}[\Theta]} \qquad (13)$$

with initial condition $\mathbf{S}_\mathbf{x}(t_0, \mathbf{x}_0)$ along the systems dynamics, which is known as the forward-sensitivity matrix equation. The additional computational effort is of order $\mathcal{O}(n_{\theta_\mathbf{x}})$, as only $n_{\theta_\mathbf{x}}$ additional ODEs have to be solved in Equation (13), since $\frac{\partial \mathbf{x}}{\partial \boldsymbol{\theta}_\mathbf{y}} = \mathbf{0}^{n_\mathbf{x} \times n_{\theta_\mathbf{y}}}$. One may also formulate an adjoint system to derive the state sensitivities in a backward manner or use numerical differentiation.
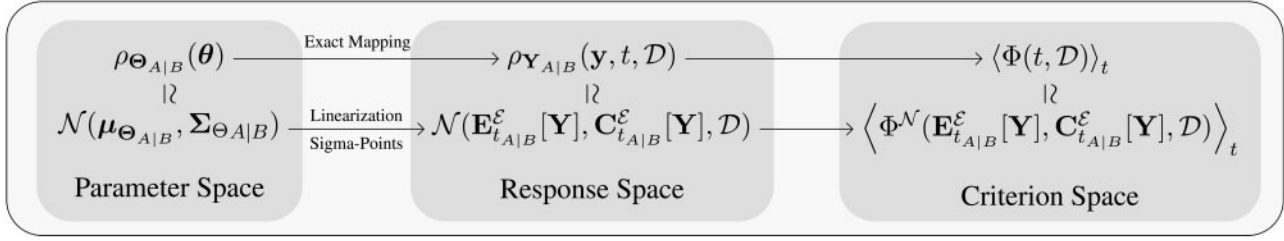
Fig. 1. Approximation of nonlinear PDF mapping

Having determined the parameter sensitivities of the system, the linear estimates of expectation and variance–covariances of the model response PDF can be calculated to yield

$$\mathbf{E}_t^{\mathcal{L}}[\mathbf{Y}] = \mathbf{h}(t, \mathbf{E}[\Theta]) \tag{14}$$

$$\mathbf{C}_t^{\mathcal{L}}[\mathbf{Y}] = \mathbf{S}(t, \mathbf{y}) \, \mathbf{C}[\Theta] \, \mathbf{S}(t, \mathbf{y})^{\mathrm{T}}. \tag{15}$$

For nonlinear models, the estimate of the expectation is typically biased, i.e. $\mathrm{Bi} = \mathbf{E}_t^{\mathcal{L}}[\mathbf{Y}] - \mu_{\mathbf{Y}} \neq \mathbf{0}$, and errors are introduced at second and higher orders. The quality of the predicted variance–covariance cannot readily be judged as the errors are of fourth and higher order, whereas the contributions depend on the system. Notice that the linear design approach yields a local estimate in the parameter space, i.e. parameter-dependent coexisting stable states will be missed, resulting in significant estimation errors in both moments (Section 3.2). The estimators are exact for linear systems, as higher-order terms vanish.

2.4.2 Estimation based on sigma-points Julier *et al.* (2000) introduced the sigma-point method for advanced Kalman filtering and state estimation. It is based on the idea that with a fixed set of parameters (sigma-points), it is easier to approximate a nonlinearly transformed PDF by a Gaussian distribution than the nonlinear transformation itself. Julier *et al.* (2000) show that expectation and variance–covariance of a random variable $\mathbf{Y}$, given by a transformation $\mathbf{Y} = \mathbf{h}(t, \Theta)$, possibly nonlinear, of a random variable Θ with expectation $\mathbf{E}[\theta]$ and variance–covariance $\mathbf{C}[\theta]$ can be estimated according to the following procedure:

(1) Select $2n_\theta + 1$ sigma-points in the original domain according to

$$\theta^{(0)} = \mathbf{E}[\Theta]; \quad \theta^{(i)} = \theta^{(0)} \pm \sqrt{n_\theta + \lambda} \sqrt{\mathbf{C}[\theta]}^{(i)},$$

where $\sqrt{\mathbf{C}[\Theta]}^{(i)}$ is the ith column of the square root of the variance–covariance matrix.

(2) Propagate these points through the model

$$\mathbf{y}_t^{(i)} = \mathbf{h}(t, \theta^{(i)}).$$

(3) Estimated expectation and variance–covariance of the transformed variable based on the sigma-points are given by the linearly weighted sums

$$\mathbf{E}_t^{\mathcal{S}}[\mathbf{Y}] = \sum_{i=-n_\theta}^{n_\theta} w^{(i)} \mathbf{y}_t^{(i)} \tag{16}$$

$$\mathbf{C}_t^{\mathcal{S}}[\mathbf{Y}] = (1 - \alpha^2 + \beta)\Big(\mathbf{y}_t^{(0)} - \mathbf{E}_t[\mathbf{Y}]\Big)\Big(\mathbf{y}_t^{(0)} - \mathbf{E}_t[\mathbf{Y}]\Big)^{\mathrm{T}} + \tag{17}$$

$$+ \sum_{i=-n_\theta}^{n_\theta} w^{(i)} \Big(\mathbf{y}_t^{(i)} - \mathbf{E}_t[\mathbf{Y}]\Big)\Big(\mathbf{y}_t^{(i)} - \mathbf{E}_t[\mathbf{Y}]\Big)^{\mathrm{T}}$$

with weights $w^{(0)} = \frac{\lambda}{n_\theta + \lambda}$, $w^{(\pm i)} = \frac{1}{2(n_\theta + \lambda)}$ and $\lambda = \alpha^2(n_\theta + \kappa) - n_\theta$.

According to Julier *et al.* (2000), the errors of the expectation estimate is of fourth and higher order, whereas the variance–covariance estimates have an error of fourth and higher order. This, however, only holds for scalars, i.e. $n_\theta = 1$, as pointed out by Gustafsson and Hendeby (2008). For $n_\theta > 1$, the sigma-point parameters (α, β, κ) can be used to tune the estimated moments by including a priori knowledge about the PDFs, i.e. β and κ allow to account for higher-order moments of the parameter PDF and should be set to $\beta = 2$ for an initial Gaussian, whereas for $n_\theta > 3$ one should choose $\kappa = 0$. Further, α controls the sigma-point spread and should lie within $0 < \alpha \leq 1$ (Julier *et al.*, 2000). The sigma-point has several advantages:

- no need to calculate derivative information (neither Jacobian nor Hessian have to be available or need to exist), which makes this method numerically robust and applicable to a wide range of system classes,
- use of curvature information of the system,
- deterministic sampling method with computational effort that scales linearly with the number of distributed variables, i.e. $\mathcal{O}(n_\theta)$,
- since each sigma-point is independently propagated, parallelization can easily be applied to speed up estimate calculation of the transformed expectation and variance–covariance.

2.5 Robust optimal stimulus design

The problem of finding an optimal stimulus design can be stated as an optimal control problem. Given a nonlinear dynamic system of the form Equations (1) and (2) and corresponding parameter set (expectation and variance–covariance), an optimal stimulus is an admissible control defined over an interval $[t_0, t_f]$, say experimental time window, at which a cost function assumes its infimum (or supremum) with the set of all admissible controls. Robustness of such a control with respect to distributed model responses can be achieved by incorporating expectation and variance–covariance into a robust design criterion (e.g. model overlap). Within the sigma-point approach, variabilities in the stimulus conductions can also be accounted by interpreting a design $\mathbf{u}(t)$ as a time-dependent mean of a distributed variable $\mathbf{U}$. Then, for a design $\mathbf{u}(t)$, the spread in the model response PDF is determined by the propagation of sigma-points given by mean and variance–covariances of (i) model parameters and (ii) stimulus. The problem of finding an optimal control may be solved by (i) Hamilton-Jacobi-Bellman, (ii) variational or (iii) NLP-based approaches (Nevistic nonlinear programming, 1997). We use the following two direct NLP-based approaches (Biegler, 2007), which can easily be combined with the methods discussed in Sections 2.4 and 2.4.1 for mapping distributed parameters onto the design criterion:

- *Direct sequential approach:* A control vector parameterization in combination with numerical integration of the model equations. This approach is suited for design problems without nonlinear

path constraints and stable behavior with respect to variations in the control and parameters.

- *Direct simultaneous:* A full discretization of the problem, e.g. control vector and state/response vector parameterization, is based on orthogonal collocation on finite elements. If the design problem includes nonlinear path constraints, this solution approach can be beneficial, because feasibility of the solution is ensured at the collocation points of each finite element.

Both NLP approaches are typically non-convex, i.e. there exist several local and possibly one global optimal design solution. Therefore, resulting solutions to the NLP problem are local optima. Global optimality of the design can be achieved—but is not ensured—by (i) performing local optimizations from different initial starting points and/or (ii) deterministic/stochastic/heuristic global optimizers (Floudas and Gounaris, 2009; Horst *et al.*, 2000; Zabinsky, 2003). We point out that optimal design solutions need not necessarily be global in real-life applications. Local optimal solutions can be very close to the global solution with respect to the design criterion. Therefore, non-convexity allows to account for further experimental constraints—restricting the degrees of freedom in the design space—without losing, e.g. discriminative power. In the applications, we use global stochastic and multistart optimizations to avoid a biased comparison between linearization and sigma-point approach.

3 APPLICATION

3.1 Application I: signaling cascade

The highly conserved mitogen-activated protein kinase signaling cascade (Pearson *et al.*, 2001) with two different hypothesized negative feedbacks is used as a nonlinear test system for benchmarking the two design approaches with respect to estimation accuracy and design quality. Multistep signaling cascades are enrolled in many signal-transduction processes of cells to sense and react to external signals. Upon an external stimulus, e.g. growth factor, hormones or stress signals, cascades transduce the signal from the cell membrane to the nucleus to start different cell programs. The respective ODE systems—adapted from Behar *et al.* (2007)—of two model candidates $m \in \{A, B\}$ that describe the change in protein concentration of the phosphorylated forms are

$$\frac{\mathrm{d}}{\mathrm{d}t}x_{1m}^*(t) = \frac{k_{1m}u(t)x_{1m}(t)}{K_{1m}+x_{1m}(t)} - \frac{v_{2m}x_{1m}^*(t)}{K_{2m}+x_{1m}^*} - r_{1m}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}x_{2m}^*(t) = \frac{k_{3m}x_{1m}^*(t)x_{2m}(t)}{K_{3m}+x_{2m}(t)} - \frac{v_{4m}x_{2m}^*(t)}{K_{4m}+x_{2m}^*(t)} - r_{2m}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}x_{3m}^*(t) = \frac{k_{5m}x_{2m}^*(t)x_{3m}(t)}{K_{5m}+x_{3m}(t)} - \frac{v_{6m}x_{3m}^*(t)}{K_{6m}+x_{3m}^*(t)}$$

with model *A*:

$$r_{1A} = k_{9A}x_4^*(t)x_{1A}^*(t)\,;\; r_{2A} = k_{10A}x_{3A}^*(t)x_{2A}^*(t)$$

$$\frac{\mathrm{d}}{\mathrm{d}t}x_{4A}^*(t) = \frac{k_{7A}x_{3A}^*(t)x_{4A}(t)}{K_{7A}+x_{3A}(t)} - \frac{v_{8A}x_{4A}^*(t)}{K_{8A}+x_{4A}^*(t)}$$

and model *B*:

$$r_{1B} = \frac{k_{9B}x_{3B}^*(t)x_{1B}^*(t)}{K_{9B}+x_{1B}^*(t)}\,;\; r_{2B} = \frac{k_{10B}x_{3B}^*(t)x_{2B}^*(t)}{K_{10B}+x_{2B}^*(t)}$$

no $x_{4B}(t)$, $x_{4B}^*(t)$.

For both models, we assume

$$x_{im}^{\mathrm{tot}}(t) = x_{im}(t) + x_{im}^*(t)$$

$$x_{im}^*(t_0) = 0$$

with the total concentration of each species x_{im}^{tot} as an additional model parameter and $i \in \{1, 2, 3, (4)_B\}$.

The measurement response signals are defined as

$$y_{1m}(t) = x_{2m}^*(t) + \boldsymbol{\theta}_\epsilon \quad \text{and} \quad y_{2m}(t) = x_{3m}^*(t) + \boldsymbol{\theta}_\epsilon, \qquad (18)$$

where $\boldsymbol{\theta}_\epsilon$ represents additive measurement noise, which we assume to be normally distributed with zero mean and variance σ_ϵ^2. We assume that the response signals can be measured at n_t specific time points. Based on an initial stimulus design, we adjust the model parameters so that both model responses match up to a small error, which does not allow to prefer one over the other model, to mimic the starting point of an OESD for model discrimination. Identifiability of the models has been checked with the software tool DAISY in combination with global sensitivity analysis (see Supplementary Material, Bellu *et al.*, 2007; Sobol', 1993). Because biological systems often follow a log-normal distribution, we apply a log-normal transformation to the response to improve the normal approximation used in the estimation approaches for the response PDF (Sections 2.4 and 2.4.1). Therefore, we redefine the response signal Equation (18) used for the overlap calculation as

$$\tilde{y}_{im}(t) = \log(y_{im}(t) + \lambda), \qquad (19)$$

with $i = 1, 2, \lambda > 0$. For each model, we assume that all dynamic parameters are log-normally distributed, with nominal value being the expectation $\mathbf{E}_{\log}[\Theta]=\boldsymbol{\theta}$ and diagonal covariance matrix $\sqrt{\mathbf{C}_{\log}[\Theta]} = \mathrm{diag}(\eta\mathbf{E}_{\mathrm{Log}}[\Theta])$, with scaling parameter η. The measurement noise is typically independent on the stimulus design and thus held constant at $\sigma_\epsilon = 0.01$. The sigma-points for the log-normal parameter PDF are obtained in the following way: In the parameter space, we derive the normal equivalents of log-normal expectation and covariance to calculate the normal sigma-points, which we then exponentiate (see Supplementary Material). The log-normal sigma-points are propagated through the model, including Equation (19), to obtain the normal estimates via Equations (16) and (17). In the following, we drop the tilde from the redefined response signal in Equation (19).

Following the direct sequential approach, the stimulus (single input) is parameterized as

$$u(\mathbf{U}, t) = u_k \quad \text{for} \quad t_k \le t \le t_{k+1},$$

with $[\mathbf{U}]_k = (u_k, \mathrm{d}t_k)^{\mathrm{T}}, k \in \{0, \ldots, n_u\}$, whereas $\mathrm{d}t_{nu} \equiv 0$. Here, u_k represents the amount of stimulus between the time point t_k and $t_{k+1} = \mathrm{d}t_k + \sum_{j=0}^{k} t_j$. If for the last time point, we have $t_{n_u} < t_f$, we put $u(\mathbf{U}, t_{n_u} < t \le t_f) = u_{n_u}$. On the other hand, if $t_{n_u} > t_f$, the design is given a penalty. Depending on the estimation method $\mathcal{E} \in \{\mathcal{L}, \mathcal{S}\}$, the resulting optimization problem for discriminating between models *A* and *B* is formulated as an NLP problem (see Supplementary Material for details), i.e.

$$\mathbf{U}_\dagger^\mathcal{E} = \arg\min_{\mathbf{U}\in\mathbb{U}\subset\mathbb{D}} \mathrm{O}^\mathcal{E}(\mathbf{U}) \equiv \left\langle \phi^\mathcal{N}\Big(\mathbf{E}_{t_{A|B}}^\mathcal{E}[\mathbf{Y}], \mathbf{C}_{t_{A|B}}^\mathcal{E}[\mathbf{Y}], \mathbf{U}\Big)\right\rangle_t \qquad (20)$$

subject to systems dynamic, additional constraints and method to estimate $\mathbf{E}_{t_{A|B}}^{\mathcal{E}}[\mathbf{Y}]$ and $\mathbf{C}_{t_{A|B}}^{\mathcal{E}}[\mathbf{Y}]$.

The number of optimization parameters is $n_{u_{\text{tot}}} = 39$, which allows 20 stimulations u_k with 19 stimulus durations dt_k. Because the problem is non-convex, we use a hybrid optimization strategy, consisting of the evolutionary-based CMA-ES algorithm (Covariance Matrix Adaptation Evolution Strategy Hansen and Ostermeier, 2001), in combination with a subsequent gradient-based optimizer. Because of the stochastic nature, the hybrid optimization is performed 40 times for each parameter variance level, which we derive from the scaling parameter η. The benchmark is based on a Monte Carlo verification of the resulting optimal stimulus designs. For each optimal design, the overlap, including expectation and variance–covariance of the model responses, is calculated based on sampling the parameter space 10^4 times for each model and corresponding optimal stimulus design (see Supplementary Material for details). The relative mean-squared error (MSE) of the moment estimates are given by

$$\langle e_{\mathbf{M}} \rangle^{\mathcal{E}} = \frac{1}{2n_t n_y} \sum_{m=A,B} \sum_i^{n_y} \sum_t^{n_t} \left(\frac{\mathbf{M}^{\mathcal{E}} - \mathbf{M}^{\mathcal{M}}}{\mathbf{M}^{\mathcal{M}}} \right)^2$$

with $\mathbf{M}^{\mathcal{E}}$ being the moment estimates of the best designs (expectation $\mathbf{E}_{t_{A|B}}^{\mathcal{E}}[Y_i]$, variance–covariance split into variance $\mathbf{VAR}_{t_{A|B}}^{\mathcal{E}}[Y_i]$ and covariance terms $\mathbf{COV}_{t_{A|B}}^{\mathcal{E}}[Y_i]$). The Monte Carlo reference is represented by $\mathbf{M}^{\mathcal{M}}$.

In Table 1, we see that for all parameter variance levels, both methods have negligible relative MSE in the mean response estimates (maximal MSE: $\langle e_{\mathbf{E}} \rangle^{\mathcal{L}} < 10^{-7}$; $\langle e_{\mathbf{E}} \rangle^{\mathcal{S}} < 10^{-9}$). In contrast, the relative MSEs for linearization increases with parameter variance levels up to 0.03 for the variance and 0.18 for the covariance estimates. Here, the sigma-point approach performs better with maximal relative MSE of the variance 0.007 and covariance 0.096. In this application, both approaches estimate mean responses of the models very well, although the maximal MSE of the sigma-points is still two orders of magnitudes smaller than the maximal MSE for linearization. For the (co)variance estimates, the sigma-point approach consistently outperforms linearization approach for increasing parameter variance level.

In the lower part of Table 1, we compare the discriminative power of the resulting designs for different parameter variance levels. Comparing the Monte Carlo verifications, we see that, for small variances, both methods yield designs that have the same discriminative power ($O_{\dagger}^{\mathcal{LM}} \equiv O^{\mathcal{M}}(\mathbf{U}_{\dagger}^{\mathcal{L}})$ versus $O_{\dagger}^{\mathcal{SM}} \equiv O^{\mathcal{M}}(\mathbf{U}_{\dagger}^{\mathcal{L}})$). However, for widely distributed parameters (starting at $\eta = 0.3$) sigma-point based designs perform up to

1.3 times better than linearization-based designs and their estimates coincide with the MC validation, which is not the case for linearization. For both methods, optimization time for one design is $1.3 \pm 0.1h$ on a standard desktop computer (4 GB RAM, 3 GHz quad core processor) and mainly determined by the optimizer itself, whereas the validation time (10^4 MC samples) of a single optimal design is $0.4 \pm 0.1h$.

In Figure 2, we show the best designs based on the estimates for two levels of parameter variances and the first response component. For small parameter variances, both designs yield the same discriminative power. For widely distributed parameters, we see that both methods tend to minimize the amount of stimulation, as this results into little response variances at maximal distances between expected model responses. Although the linear design is characterized by a strong stimulation right at the beginning, with a subsequent plateau of little stimulation, the sigma-point design starts with an even stronger initial pulse, followed by two small pulses.

3.2 Application II: Schlögl model

In this section, we compare the performance of the nonlinear design based on the sigma-point approach to the linear design in the presence of multiple steady states. The Schlögl model is a canonical example of a biochemical reaction system exhibiting bistability (Schlögl, 1972). It describes an autocatalytic,
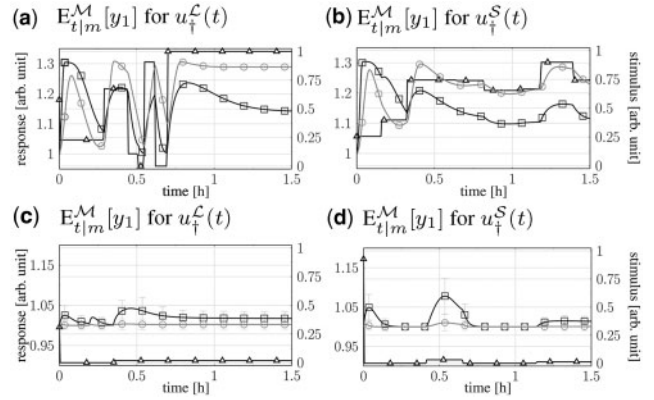
Fig. 2. Expected responses of models *A* (crossed square), *B* (crossed circle) on log-scale for optimized stimuli $u_{\dagger}^{\mathcal{E}}(t)$ (crossed triangle) and corresponding $\sqrt{\mathrm{VAR}_{t|m}[Y_1]}$ bands [first component only, **(a)/(b)** $\eta = 0.01$, **(c)/(d)** $\eta = 0.4$]

Table 1. Relative MSEs of moment estimates and overlap (scaled 10^5) of the best designs based on linearization/sigma-point estimation and corresponding Monte Carlo verification

$\eta = \sigma_i/\mathrm{E}[\theta_i]$		0.01		0.1		0.2		0.3		0.4	
$\langle e_{\mathbf{E}} \rangle^{\mathcal{L}}$	$\langle e_{\mathbf{E}} \rangle^{\mathcal{S}}$	0	0	0	0	0	0	0	0	0	0
$\langle e_{\mathbf{VAR}} \rangle^{\mathcal{L}}$	$\langle e_{\mathbf{VAR}} \rangle^{\mathcal{S}}$	0	0	0.002	0	0.001	0	0.01	0.007	0.03	0.007
$\langle e_{\mathbf{COV}} \rangle^{\mathcal{L}}$	$\langle e_{\mathbf{COV}} \rangle^{\mathcal{S}}$	0	0	0.02	0.002	0.059	0.015	0.106	0.046	0.181	0.096
$O_{\dagger}^{\mathcal{L}}$	$O_{\dagger}^{\mathcal{S}}$	0.2	0.2	2	2	2	2	2	3	3	3
$O_{\dagger}^{\mathcal{LM}}$	$O_{\dagger}^{\mathcal{SM}}$	0.2	0.2	2	2	2	2	3	3	4	3
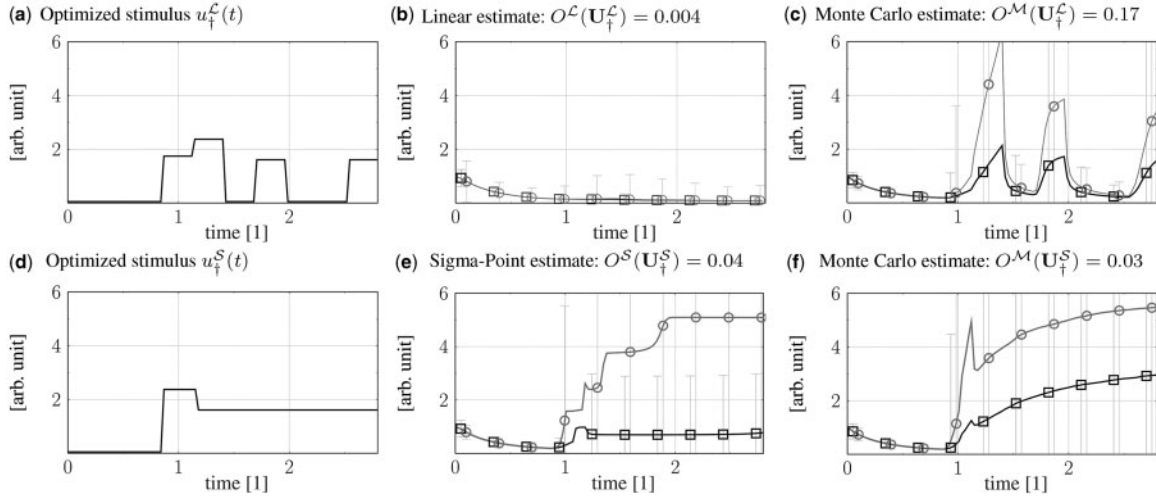
Fig. 3. Upper part: **(a)** Optimal stimulus $u_{\dagger}^{\mathcal{L}}(t)$ (bold line) and corresponding $\sqrt{\mathrm{VAR}_{t|m}[\mathbf{Y}]}$ bands based on linearization and corresponding linear **(b)** and MC **(c)** estimates of expected responses of models A (crossed square) and B (crossed circle) for $\eta = 0.35$. Lower part **(d)-(f)**: sigma-point based results

trimolecular reaction that, for instance, occurs in post-translational modifications of signaling proteins. Two model alternatives for the rate of concentration change of species x are given by

$$\frac{\mathrm{d}}{\mathrm{d}t} x_m(t) = k_1\, a\, s_m(u(t)) x_m^2(t) - k_2 x_m^3(t) - k_4 x_m(t) + k_3 b,$$

$$s_A(u(t)) = u(t) \quad \text{or} \quad s_B(u(t)) = \frac{1}{2}(u(t) + u^2(t))$$

where we assume the four kinetic parameters $k_1 - k_4$ as well as the system parameters a and b to be distributed $\Theta \propto \mathcal{N}(\mathbf{E}[\Theta], \mathrm{diag}(\eta\mathbf{E}[\Theta])^2)$, the initial condition as $X_0 \propto \mathcal{N}(\mathrm{E}[X_0], (\eta\mathrm{E}[X_0])^2)$. The mean parameter values are taken from Vellela and Qian (2009), which also discuss the consequences of bi-(multi)stability for biological modeling in the light of non-equilibrium thermo- and stochastic dynamics. The model alternatives simply differ in the input layer $s_m(u(t))$. Parameters a and b represent the concentration of two reaction partners a and b of species x, which both are in constant exchange with a material reservoir. For an initial, suboptimal experiment with stimulus $u(t) = 1$, models A and B cannot be distinguished, given $y_m(t) = x_m(t) + \theta_\epsilon$ to be the response signal, where θ_ϵ represents additive measurement noise with zero mean and σ_ϵ^2. The stimulus is thought to control the concentration in the reservoir of species a to find an optimal discriminative stimulus, whereas subsequent stimulations can only be applied after a minimal time period has passed to account for possible control limitations (see Supplementary Material for details). Such nonlinear constraint optimization can efficiently be solved within the direct simultaneous approach, which we apply using orthogonal collocation on 100 finite elements (each with three collocation points) to discretize control and system states. The objective of the resulting non-convex NLP problem is the same as in Equation (20), however, subjected to different constraints, i.e. system dynamics in form of a nonlinear algebraic equation system and additional constraints (details see Supplementary Material). For the linear design strategy, we implement the sensitivity Equation (13) and corresponding constraints. For the sigma-point design, the constraints have to

simultaneously hold for all $(2n_\theta + 1)$ sigma–points. The solver AMPL in combination with the optimizer CONOPT is used to solve the aforementioned NLP problem (Drud, 1994). For a given optimization setup (η and estimation method), the solution takes about 2 min on a standard desktop computer. Because CONOPT yields local solutions, the optimization is performed for 1000 different randomized initial designs for a given optimization setup, from which the best solution is selected.

In Figure 3, we show the resulting stimuli designs for $\eta = 0.35$ based on linearization and sigma-point estimation. Reexamination of the optimized linear design with MC simulations reveals a large underestimation of the estimated overlap: $\mathcal{O}^{\mathcal{L}}(\mathbf{U}_{\dagger}^{\mathcal{L}}) = 0.004$ versus $\mathcal{O}^{\mathcal{M}}(\mathbf{U}_{\dagger}^{\mathcal{L}}) = 0.17$, i.e. misleading discriminative power by two orders of magnitude. The local estimation property of the linear approach yields a highly biased expectation and underestimation of the variance with relative MSE of 0.44 for the expected response and 6.66 for the variance [see Fig. 3, estimated response of model B, (b) versus (c)]. The sigma-point based design (d) in Figure 3 has a relative MSE of 0.15 for the expected response and 0.38 for the variance. The overlap estimate of the sigma-point design closely matches the MC validation $[\mathcal{O}^{\mathcal{S}}(\mathbf{U}_{\dagger}^{\mathcal{S}}) = 0.04$ versus $\mathcal{O}^{\mathcal{M}}(\mathbf{U}_{\dagger}^{\mathcal{S}}) = 0.03]$. Further, the sigma-point design performs 5.7 better than the linear design $(\mathcal{O}^{\mathcal{M}}(\mathbf{U}_{\dagger}^{\mathcal{S}}) = 0.03$ versus $\mathcal{O}^{\mathcal{M}}(\mathbf{U}_{\dagger}^{\mathcal{L}}) = 0.17)$. As can be seen in Figure 3 (e,f), the non-local propagation property of the sigma-points enables the optimizer to find a stimulus that mostly drives model B to the upper steady state.

4 CONCLUSION

Biological variability in combination with experimental measurement noise results into widely distributed response signals, which is one of the main challenges when modeling biological system deterministically with nonlinear ODEs. The parameter set needs to be extended to a parameter distribution. In this way, natural variability in the dynamic parameters as well as measurement noise can be readily accounted for. However, an exact quantification is computationally expensive and infeasible in an

optimization framework for large systems. Therefore, approximate descriptions of the PDFs and nonlinear mapping process between parameter and model response space have to be used. Here, we have presented a nonlinear design approach based on the sigma-points within the application of model-based OESD aimed at model discrimination. We illustrate the application and performance in combination with two numerical approaches from optimal control for two nonlinear model examples. Using the overlap as a robust design criterion based on the model response PDFs, we show that in the case of nonlinear models with widely distributed parameter PDFs, the sigma-point predictions and designs consistently outperform a linear design approach. In the case of bi-(multi)stability, we further illustrate the benefit of the nonlocal propagation property. Finally, the sigma-points come with several numerical advantages, including linear scaling of the numerical costs with respect to distributed parameters and derivative free estimation of nonlinearly mapped expectation and variance–covariance. The latter property allows applying a robust OESD to dynamic models that have non-smooth right-hand side functions, e.g. cybernetic models of cellular metabolism (Ramkrishna, 1982).

ACKNOWLEDGEMENTS

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.

Asprey,S.P. and Macchietto,S. (2002) Designing robust optimal dynamic experiments. *J. Process Control*, **12**, 545–556.

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

Bandara,S. *et al.* (2009) Optimal experimental design for parameter estimation of a cell signaling model. *PLoS Comput. Biol.*, **5**, e1000558.

Behar,M. *et al.* (2007) Mathematical and computational analysis of adaptation via feedback inhibition in signal transduction pathways. *Biophys. J.*, **93**, 806–821.

Bellu,G. *et al.* (2007) DAISY: a new software tool to test global identifiability of biological and physiological Systems. *Comput. Methods Programs Biomed.*, **88**, 52–61.

Biegler,L.T. (2007) An overview of simultaneous strategies for dynamic optimization. *Chem. Eng. Process.*, **46**, 1043–1053.

Box,G.E.P. and Cox,D.R. (1964) An analysis of transformations. *J. Roy. Stat. Soc. B*, **26**, 211–252.

Box,G.E.P. and Hill,W.J. (1967) Discrimination among mechanistic models. *Technometrics*, **9**, 57–71.

Chen,H. and Asprey,S.P. (2003) On the design of optimally informative dynamic experiments for model discrimination in multiresponse nonlinear situations. *Ind. Eng. Chem. Res.*, **42**, 1379–1390.

Donckels,B.M.R. *et al.* (2012) Performance assessment of the anticipatory approach to optimal experimental design for model discrimination. *Chemometr. Intell. Lab.*, **110**, 20–31.

Drud,A.S. (1994) CONOPT—a large-scale grg code. *INFORMS J. Comput.*, **6**, 207–216.

Du,D.-Z. and Pardalos,P.M. (1995) *Minimax and Applications (Nonconvex Optimization and Its Applications)*. Kluwer Academic Publishers, The Netherlands.

Floudas,C.A. and Gounaris,C.E. (2009) A review of recent advances in global optimization. *J. Global Optim.*, **45**, 3–38.

Geyer,C.J. (1992) Practical Markov chain Monte Carlo. *Stat. Sci.*, **7**, 473–483.

Gustafsson,F. and Hendeby,G. (2008) On nonlinear transformations of stochastic variables and its application to nonlinear filtering. In *IEEE International Conference on Acoustics, Speech and Signal Processing*. Las Vegas, NV, March 31–April 4.

Hansen,N. and Ostermeier,A. (2001) Completely derandomized self-adaptation in evolution strategies. *Evol. Comput.*, **9**, 159–195.

Heine,Th. *et al.* (2008) Derivative-free optimal experimental design. *Chem. Eng. Sci.*, **63**, 4873–4880.

Horst,R. *et al.* (2000) *Introduction to Global Optimization. Nonconvex Optimization and its Applications*. Kluwer Academic Publishers, The Netherlands.

Jebara,T. *et al.* (2004) Probability product kernels. *J. Mach. Learn. Res.*, **5**, 819–844.

Johnson,N.L. (1949) Systems of frequency curves generated by methods of translation. *Biometrika*, **36**, 149–176.

Julier,S. *et al.* (2000) A new method for the nonlinear transformation of means and covariances in filters and estimators. *IEEE Trans. Automat. Contr.*, **45**, 477–482.

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

Lorenz,S. *et al.* (2007) Discrimination of dynamical system models for biological and chemical processes. *J. Comput. Chem.*, **28**, 1384–1399.

Melykuti,B. *et al.* (2010) Discriminating between rival biochemical network models: three approaches to optimal experiment design. *BMC Syst. Biol.*, **4**, 38.

Michalik,C. *et al.* (2010) Optimal experimental design for discriminating numerous model candidates: the AWDC criterion. *Ind. Eng. Chem. Res.*, **49**, 913–919.

Nevistic,V. (1997) *Constrained control of nonlinear systems*. PhD Thesis, ETH No. 12042, ETH Zurich, Switzerland.

Pearson,G. *et al.* (2001) Mitogen-activated protein (MAP) kinase pathways: regulation and physiological functions. *Endocr. Rev.*, **22**, 153–183.

Ramkrishna,D. (1982) A cybernetic perspective of microbial growth. In Blanch,H.W., Papoutsakis,E.T. and Stephanopoulos,G.N. (eds.) *Foundations of Biochemical Engineering: Kinetics and Thermodynamics in Biological Systems*. American Chemical Society, Washington, DC, pp. 161–178.

Raue,A. *et al.* (2009) Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, **25**, 1923–1929.

Raue,A. *et al.* (2010) Identifiability and observability analysis for experimental design in nonlinear dynamical models. *Chaos*, **20**, 045105.

Schenkendorf,R. *et al.* (2009a) Optimal experimental design with the sigma point method. *IET Syst. Biol.*, **3**, 10–23.

Schenkendorf,R. *et al.* (2009b) Optimal experimental design and model selection by a sigma point approach. In *Proceedings MATHMOD,* Vienna, Austria. February 11–13.

Schlögl,F. (1972) Chemical reaction models for non-equilibrium phase transitions. *Z.f.Physik*, **253**, 147–161.

Singh,S. (1998) On establishing the credibility of a model for a system. *Chem. Eng. Res. Des.*, **76**, 657–668.

Skanda,D. and Lebiedz,D. (2010) An optimal experimental design approach to model discrimination in dynamic biochemical systems. *Bioinformatics*, **26**, 939–945.

Skanda,D. and Lebiedz,D. (2012) A robust optimization approach to experimental design for model discrimination of dynamical systems. *Math. Program*, doi:10.1007/s10107-012-0532-0.

Sobol',I.M. (1993) Sensitivity analysis for nonlinear mathematical models. *Math. Mod. Com. Exp.*, **1**, 407–414.

Topsoe,F. (2000) Some inequalities for information divergence and related measures of discrimination. *IEEE Trans. Inform. Theory*, **46**, 1602–1609.

Vellela,M. and Qian,H. (2009) Stochastic dynamics and non-equilibrium thermodynamics of a bistable chemical system: the Schlögl model revisited. *J. R. Soc. Interface*, **6**, 925–940.

Vanlier,J. *et al.* (2012) A Bayesian approach to targeted experiment design. *Bioinformatics*, **28**, 1136–1142.

Walter,E. and Pronzato,L. (1997) *Identification of Parametric Models from Experimental Data. Communications and Control Engineering*. Springer, Berlin.

Zabinsky,Z.B. (2003) *Stochastic Adaptive Search for Global Optimization. Nonconvex Optimization and Its Applications*. Kluwer Academic Publishers.