German Aerospace Center

From the SelectedWorks of René Schenkendorf

2009

Optimal experimental design with the sigma point method

René Schenkendorf, German Aerospace Center
Andreas Kremling
Michael Mangold

Available at: https://works.bepress.com/rene_schenkendorf/4/
Optimal experimental design with the sigma point method

R. Schenkendorf A. Kremling M. Mangold
Max-Planck-Institute for Dynamics of Complex Technical Systems, Sandtorstr. 1, 39106 Magdeburg, Germany
E-mail: rschenke@mpi-magdeburg.mpg.de

Abstract: Using mathematical models for a quantitative description of dynamical systems requires the identification of uncertain parameters by minimising the difference between simulation and measurement. Owing to the measurement noise also, the estimated parameters possess an uncertainty expressed by their variances. To obtain highly predictive models, very precise parameters are needed. The optimal experimental design (OED) as a numerical optimisation method is used to reduce the parameter uncertainty by minimising the parameter variances iteratively. A frequently applied method to define a cost function for OED is based on the inverse of the Fisher information matrix. The application of this traditional method has at least two shortcomings for models that are nonlinear in their parameters: (i) it gives only a lower bound of the parameter variances and (ii) the bias of the estimator is neglected. Here, the authors show that by applying the sigma point (SP) method a better approximation of characteristic values of the parameter statistics can be obtained, which has a direct benefit on OED. An additional advantage of the SP method is that it can also be used to investigate the influence of the parameter uncertainties on the simulation results. The SP method is demonstrated for the example of a widely used biological model.

1 Introduction

The use of mathematical models to analyse complex systems is very common in many research fields. One important step of model development is parameter identification (PI), that is, the determination of unknown model parameters by minimising the difference between simulation results and measurements. Even for the deterministic case with ideal measurement, it is hard to obtain unique estimations of parameters such as the maximum growth rate $m$ and the substrate affinity constant $K_S$ of Michaelis–Menten kinetics [1], which is frequently used for biological models. In practice, the measurement is disturbed with noise leading to some uncertainty of the estimated parameters [2]. This fact has a strong impact on the model quality as only parameters with small variances ensure simulation results with a highly predictive power, which is necessary for process observation and control or to study biological systems. The design of appropriate experiments for PI is known as optimal experimental design (OED) and is an important problem in the field of systems biology [3, 4]. General overviews on OED methods for different fields of application can be found in [5, 6].

The OED mainly consists of two steps: (i) determining the parameter variances from PI with available experimental data and (ii) minimising these variances by stimulating the system ‘optimally’ in a new experiment. These steps can be reiterated until a sufficient accuracy is achieved (Fig. 1). A frequently used method to obtain some information about the statistics of the estimated parameters is based on the inverse of the Fisher information matrix (FIM) [7]. In cases where models are nonlinear with respect to their parameters, the FIM may lead to a poor approximation of the variances. To overcome this problem, various methods have been developed to improve the calculation of parameter uncertainties. In the majority of cases, the approaches are based on Monte Carlo methods as the global sensitivity analysis [3, 8, 9] and the Bootstrap approach [4, 10]. This kind of realisation has an increased computational complexity, tending to prohibit their use in an iterative OED process, as the variances of the parameters have to be determined a lot of times. Here, we show that by applying the SP method a better approximation of the parameter statistics with a workable computational effort can be obtained.
2 Methods

2.1 PI and model quality

Ordinary differential equations (o.d.e.'s) are widely used to describe the dynamics of many systems. The o.d.e. system

\[
\frac{dx(t)}{dt} = f(x(t), u(t), \omega(t), \Theta, t)
\]

and its solution

\[
x(t) = x(t_0) + \int_{t_0}^{t} f(x(\tau), u(\tau), \omega(\tau), \Theta, \tau) d\tau
\]

are functions of the dynamic states \( x(t) \in \mathbb{R}^n \), the system input \( u(t) \in \mathbb{R}^r \), some process noise \( \omega(t) \in \mathbb{R}^m \), parameters \( \Theta \in \mathbb{R}^p \) and time \( t \).

The correlation between the states, inputs and the simulation results \( \hat{y}(t) \in \mathbb{R}^m \) is given by the output equation

\[
\hat{y}(t) = h(x(t), u(t), \omega(t), \Theta, t)
\]

where \( \omega \in \mathbb{R}^m \) denotes an additional measurement noise. If the structure of the model, for example, given by the stoichiometric network and the reaction kinetics, is known, then the parameter vector \( \Theta \) determines the predictive power of the model. This can be expressed by the weighted difference between the simulation results \( \hat{y}(t) \in \mathbb{R}^m \) and measurements \( y(t) \)

\[
J_{ML}(\hat{\Theta}) = \frac{1}{2} (y - \hat{y}(x_0, u, \hat{\Theta}))^T C_p^{-1} (y - \hat{y}(x_0, u, \hat{\Theta}))
\]

The unknown parameters \( \hat{\Theta} \) are estimated by minimising the cost function equation (4) for a pre-defined input vector \( u(t) \). In this work, \( \hat{\Theta} \) denotes the estimated values of the unknown model parameters, whereas \( \Theta \) are the actual parameter values. If the covariance matrix of the measurements is used for \( C_p \), then a maximum likelihood estimator results [2]. In practice, the measurement noise \( \omega(t) \) and the process noise \( \omega(t) \) lead to parameter uncertainties, expressed by the covariance matrix

\[
C_\Theta = E\left[ (\hat{\Theta} - \Theta)(\hat{\Theta} - \Theta)^T \right]
\]

where \( E[\cdot] \) is the expected value of a random variable. In addition, the nonlinearities in the model may lead to a bias in the estimated parameters, that is, the expectation \( M_\Theta = E[\Theta] \) may be different from the actual parameter vector \( \Theta \). As the ‘true’ parameter vector \( \Theta \) is unknown, an approximation of \( C_\Theta \) is needed to assess the estimation and if necessary to plan new experiments with an increased information content. The mean value \( M_\Theta \) and the matrix \( C_\Theta \) give a measure for the accuracy or quality of the model. If the covariances of the identified parameters turn out to be unacceptably high, then one can try to reduce them by choosing other experimental conditions, that is, by making use of the additional degrees of freedom given by the input vector \( u(t) \). This is the key idea of OED.

2.2 Reduction of the parameter uncertainty by OED

The OED as a numerical optimisation problem reduces the parameter uncertainty by the variation of some design variables \( u(t) \). In order to formulate the optimisation problem, it is necessary to define a scalar cost function \( \Phi \), which depends on the parameters’ bias \( M_\Theta = \Theta \) and on the covariance matrix \( C_\Theta \). In general, the bias is neglected in the cost function, that is, \( \Phi = \Phi(C_\Theta) \). Well-known optimality criteria [2] have been proposed for \( \Phi \)

\[
A\text{-optimal design } \Phi_A(C_\Theta) = \text{trace}(C_\Theta) \\
D\text{-optimal design } \Phi_D(C_\Theta) = \text{det}(C_\Theta) \\
M\text{-optimal design } \Phi_M(C_\Theta) = \max\left(\sqrt{C_{\Theta,ii}}\right) \\
E\text{-optimal design } \Phi_E(C_\Theta) = \frac{1}{\lambda_{\text{max}}(C_\Theta)} \\
E^*\text{-optimal design } \Phi_{E^*}(C_\Theta) = \frac{\lambda_{\text{max}}(C_\Theta)}{\lambda_{\text{min}}(C_\Theta)}
\]

with \( \lambda_{\text{max}}(\lambda_{\text{min}}) \) as the maximum (minimum) eigenvalue of \( C_\Theta \). Which of them leads to the best result, depends on

\[
\text{Experiment 1} \quad \text{Experiment 2}
\]

Figure 1 Iterative nature of the OED process for a two-dimensional parameter space

The information about the parameter variances, expressed by confidence regions, is used to design new experiments that minimise the parameter uncertainties, whereby the arrows indicate the interaction between the experiments (measurements) and the PI.
the model and the design parameters, hence no general advice can be given. A challenging problem of OED is how to obtain an accurate estimation of the covariance matrix $C_\theta$ and of the estimated mean $M_\theta$. Three approaches will be discussed in the following. Although the first two are well-known, the third one has not yet been used to identify parameter uncertainties directly.

### 2.3 Fisher information matrix

For the case of additive white measurement noise, the FIM is determined by the following equation

$$\text{FIM} = \sum_{t_i} S^T_{t_i} \cdot C_{y}^{-1} \cdot S_{t_i} \quad (11)$$

with the parameter sensitivity matrix $S$

$$S_{t_i} = \begin{bmatrix} \frac{\partial y_1}{\partial \theta_1}_{t_i} & \frac{\partial y_1}{\partial \theta_2}_{t_i} & \cdots & \frac{\partial y_1}{\partial \theta_p}_{t_i} \\ \frac{\partial y_2}{\partial \theta_1}_{t_i} & \frac{\partial y_2}{\partial \theta_2}_{t_i} & \cdots & \frac{\partial y_2}{\partial \theta_p}_{t_i} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial \theta_1}_{t_i} & \frac{\partial y_m}{\partial \theta_2}_{t_i} & \cdots & \frac{\partial y_m}{\partial \theta_p}_{t_i} \end{bmatrix} \quad (12)$$

the covariance matrix of the measurement data $C_y$ and the measurement time point $t_i$. As the dynamic states $x(t)$ are time dependent, the following matrix differential equation for the sensitivities has to be solved in parallel with the dynamic system given in (1)

$$\dot{S} = \frac{\partial f}{\partial y} \cdot S + \frac{\partial f}{\partial \theta} \cdot S(0) = 0^{p \times p} \quad (13)$$

The inverse of the FIM provides an estimation of the parameter variances based on the Cramer–Ra’o inequality [7]

$$C_\theta \geq \frac{\partial E[\hat{\theta}]}{\partial \theta} \text{FIM}^{-1} \frac{\partial E[\hat{\theta}]}{\partial \theta}^T \quad (14)$$

The equality only holds, if (i) the measurement errors are additive and (ii) the model is linear in its parameters. Furthermore, the FIM does not give any information on $E[\hat{\theta}]$. Therefore in many cases, it is assumed in addition that (iii) the estimate is unbiased, that is, $E[\hat{\theta}] = \theta$. This simplifies (14) to

$$C_\theta \geq \text{FIM}^{-1} \quad (15)$$

The use of (15) is state of the art in the field of OED [2, 11–14]. However, esp. conditions (ii) and (iii) are not met when analysing complex biological models. Consequently, the approximation $\text{FIM}^{-1}$ may strongly underestimate the actual covariances [4]. To make things worse, $\text{FIM}^{-1}$ is not even a reliable lower bound for $C_\theta$. If the gradient $\partial E[\hat{\theta}] / \partial \theta$ is small, the actual covariances can be overestimated by (15). With an increasing measurement uncertainty, these effects become more and more important depending on the nonlinearity. In conclusion, the use of the FIM is unsatisfactory for the OED of nonlinear biological models. Alternative approaches are required.

### 2.4 Bootstrap

The roots of Bootstrap go back to the Monte Carlo method [10]. A set of $B$ fictitious measurement vectors $y'_i$ is created randomly. The set has to represent the statistics of the measurement in the sense that the mean and the covariance of the samples agree with the actual distribution

$$y = \frac{1}{B} \sum_{i=0}^{B} y'_i \quad (16)$$

$$C_y = \frac{1}{B-1} \sum_{i=0}^{B} (y'_i - y)(y'_i - y)^T \quad (17)$$

A parameter estimation is done for each of the samples $y'_i$. This leads to a set of $B$ estimated parameters $\hat{\theta}_i$. To make a statement about the confidence regions of the estimated parameters, the percentile method is mostly preferred [4, 10], as it also enables a determination of higher moments, for example, the skewness of the random variables. However, the first two moments, the mean and the variances, are sufficient for the OED process and can be easily approximated by the parametric bootstrap [10] with the following equations

$$\bar{\theta} = \frac{1}{B} \sum_{i=0}^{B} \hat{\theta}_i \quad (18)$$

$$C_{\theta,\text{Boot}} = \frac{1}{B-1} \sum_{i=0}^{B} (\hat{\theta}_i - \bar{\theta})(\hat{\theta}_i - \bar{\theta})^T \quad (19)$$

As only a sample number of $B \to \infty$ leads to a correct calculation of the parameter variance $C_{\theta,\text{Boot}}$, a compromise between the accuracy and computational effort has to be found. Typically, 1000 to 10 000 samples are necessary to obtain reasonably accurate results. As each sample involves a PI step, and as a Bootstrap only returns a single evaluation of the objective function for OED, the Bootstrap method is prohibitively expensive for the use in the framework of OED.

### 2.5 Sigma points

Julier and Uhlmann [15] suggested the use of the so-called SPs in order to determine the mean and covariance of a random variable $\eta \in \mathbb{R}^l$ from the mean and covariance of a random variable $\xi \in \mathbb{R}^f$, where $\eta$ is related to $\xi$ by the nonlinear mapping

$$\eta = g(\xi) \quad (20)$$

They showed that an accurate estimation of $E[\eta]$ and $C_\eta$ can be obtained from $(2 \cdot f + 1)$ evaluations of $g(\cdot)$ for the
(2 \cdot f + 1) \text{ deliberately chosen samples of } \xi \text{ (details of the method will be given below). Although the method was originally developed for applications in the area of nonlinear filtering, known as the unscented Kalman filtering [16], it can also be useful for OED problems, as will be shown in the following. To simplify the discussion, it is assumed that the measurement noise is additive, that is,}

\[ y = b(x) + v \]  
\[ E[v] = 0 \]  
\[ E[v \cdot v^T] = C_y \]  

although the method is not restricted to this case.

Then the PI problem can be formally written in the form of (20), where

- \( \xi \) is the vector of all available measurement data. If \( K \) is the number of measurement time points, \( \xi \) has the dimension \( f = m \cdot K \) and is given by

\[ \xi = [y_1(t_1), \ldots, y_1(t_K), y_2(t_1), \ldots, y_2(t_K), \ldots, y_m(t_1), \ldots, y_m(t_K)]^T \]

- \( g(\cdot) \) stands for the complete PI process, that is, the determination of a parameter vector \( \hat{\theta} \) that minimises the objective function equation (4) by numerical optimisation;

- \( \eta \) is the resulting vector of estimated model parameters, that is, \( \eta = \hat{\theta} \).

By using this analogy, the SP method by Julier and Uhlmann can be applied directly to calculate the mean and the covariances of \( \hat{\theta} \). For a measurement vector \( y \) with the dimension \( m \) and \( K \) measurement time points, the covariance matrix of the measurement noise \( C_y \) is used to determine \( 2mK + 1 \) vectors of \( \varepsilon_i \)

\[ \varepsilon_0 = \begin{pmatrix} (mK + \lambda) \sqrt{C_y} \end{pmatrix} \quad \text{ or } \begin{pmatrix} y_i = +\sqrt{(mK + \lambda)}(\sqrt{C_y})_i \end{pmatrix} \quad \text{ for } i = 1, \ldots, mK \]
\[ \varepsilon_i = -\sqrt{(mK + \lambda)}(\sqrt{C_y})_i \quad \text{ for } i = mK + 1, \ldots, 2mK \]

where \( (\sqrt{C_y})_i \) is the \( i \)-th column of the matrix square root. The disturbances \( \varepsilon_i \), which represent the statistics of the measurement, generate a well-defined set of measurement samples

\[ y_i = b(x, \theta) + \varepsilon_i \quad \text{ for } i = 0, \ldots, 2mK \]

These samples are used to identify \( 2mK + 1 \) parameter vectors \( \hat{\theta}_i \) with the mean \( \hat{\theta} \) and the covariance matrix \( C_{\theta,SP} \),

\[ \hat{\theta} = \sum_{i=0}^{2mK} \varepsilon_i^T \theta_i \]
\[ C_{\theta,SP} = \sum_{i=0}^{2mK} \varepsilon_i (\theta_i - \theta) (\theta_i - \theta)^T \]

where the weights \( \varepsilon_i \) are given by

\[ \varepsilon_0^T = \frac{\lambda}{mK + \lambda} \]
\[ \varepsilon_i^T = \frac{\lambda}{mK + \lambda} + 1 - \alpha^2 + \beta \]
\[ \varepsilon_i^T = \frac{1}{2} \frac{(mK + \lambda)}{i = 1, \ldots, 2mK} \]

with \( \lambda = (mK + \kappa) - mK \) and \( \alpha, \beta, \kappa \) are scaling parameters. The meaning and the influence of \( \alpha, \beta, \kappa \) are explained in [15]. In addition, the estimated mean \( \hat{\theta} \) and the initial parameter vector \( \theta_0 \) enable the definition of a bias

\[ Bi = \hat{\theta} - \theta_0 \]

that leads to the following approximation of the parameter covariance matrix [11]

\[ E[(\theta - \hat{\theta})(\hat{\theta} - \theta)^T] = BiBi^T + C_{\theta,SP} \]

For the sake of simplicity, the initial parameter vector is set to the actual reference value: \( \theta_0 = \theta \). Then, the bias defined in (33) is identical to the deviation of the estimates’ mean value from the reference value. In the general case, the reference values of the parameters are, of course, unknown a priori. Nevertheless, the SP method can still be used to assess if the parameter estimation method produces a bias or not. In order to do this, \( \theta_0 \) is set to a first estimation \( \hat{\theta}_0 \), which is usually different from \( \theta \). The SPs are grouped around this first estimate, and an expectation \( \hat{\theta} \) of the PI is computed from (28). If \( \hat{\theta} \neq \theta_0 \), then the identification method is found to produce a bias, and the estimate of the bias is

\[ \hat{Bi} = \hat{\theta} - \theta_0 \]
of OED, it becomes possible not only to minimise the covariance but also the bias of the estimated parameters.

- The SP method does not assume some idealised PI procedure, but works with the numerical method that is actually applied to the identification problem. Therefore all kinds of imperfections and special properties of the numerical optimiser are taken into account.

- In contrast to the FIM, the SP method does not require the calculation of gradients or Jacobians for the sensitivities. Therefore it is applicable to a broad class of models, including for example, Monte Carlo models.

- The implementation of the SP method is very easy, the parallelisation is straightforward.

- As the SPs are not chosen randomly but deliberately, the number of required samples will usually be much smaller than for the Bootstrap. In addition, the necessary number of samples is clearly defined and not a matter of guessing, as it is to some extent for the Bootstrap method.

3 Applications

3.1 Substrate uptake in a bio-reactor

The application of OED is demonstrated by a simple unstructured growth model for a continuous stirred tank bio-reactor. Under the assumption that the inlet flow \( q_{\text{in}} \) [\( \text{h}^{-1} \)] and the outlet flow \( q_{\text{out}} \) [\( \text{h}^{-1} \)] of the reactor are equal \( q_{\text{in}} = q_{\text{out}} = q \) [\( \text{h}^{-1} \)], the dilution rate is defined as \( D = q/V \) [\( \text{h}^{-1} \)] and the material balance equations for concentration of biomass \( c_\text{x} \) [\( \text{g}l^{-1} \)] and substrate \( c_\text{s} \) [\( \text{g}l^{-1} \)] can be written as

\[
\dot{c}_\text{x} = \mu \cdot c_\text{x} - D \cdot c_\text{x} \quad (36)
\]

\[
\dot{c}_\text{s} = -\frac{1}{Y_{\text{xy}}} \cdot \mu \cdot c_\text{s} + (c_{\text{s, in}} - c_\text{s}) \cdot D \quad (37)
\]

where the specific growth rate \( \mu \) is determined by a Michaelis–Menten kinetics

\[
\mu = \frac{\mu_m \cdot c_\text{s}}{K_m + c_\text{s}} \quad (38)
\]

The simplicity of this unstructured model does not provide a deeper insight into biological mechanisms, but it seems appropriate to demonstrate the typical problems of determining and minimising the parameter covariance matrix for models that are nonlinear in their parameters. In total, the model has three parameters: (i) \( Y_{\text{xy}} \) – the yield factor describes how much biomass is produced by the uptake of a certain amount of substrate; (ii) \( \mu_m \) – the maximum growth rate is the upper limit of the growth rate \( \mu \) and (iii) \( K_m \) – the substrate affinity constant represents the substrate concentration at which the specific growth rate is half its maximum value. For reasons of simplification, the following assumptions are made: (i) only the concentration of biomass is measurable \( y(t) = c_\text{s}(t) \) and (ii) \( Y_{\text{xy}} \) is known from literature; consequently, only the parameters of the Michaelis–Menten kinetics, \( \mu_m \) and \( K_m \), have to be identified, which results in a two-dimensional parameter space.

3.1.1 Parameter identifiability: Before starting the PI, it seems reasonable to check if the unknown parameters can, in principle, be determined uniquely from the available measurement information. Here, available measurement information means the measurement data \( y(t) \) taken over some finite time interval \( t \in [0, T] \), or – what is equivalent to this – the measurement data \( y(0) \) at time zero plus the time derivatives \( y(0), y(0), \ldots \) up to some arbitrary order.

To simplify the analysis, it is assumed in the following that the initial substrate concentration \( c_\text{s}(0) \) is known. One obtains from the model equations (36) and (37)

\[
y(0) = c_\text{s}(0) \quad (39)
\]

\[
\dot{y}(0) = \dot{c}_\text{s}(0) = (\mu - D) \cdot c_\text{s}(0) = (\mu - D) \cdot y(0) \quad (40)
\]

\[
\dot{y}(0) = (\mu - D) y(0) + \frac{\partial \mu}{\partial c_\text{s}} \cdot \dot{c}_\text{s}(0) \cdot y(0)
\]

\[
\quad - \frac{\dot{y}(0)^2}{y(0)} + y(0) \cdot \frac{\partial \mu}{\partial c_\text{s}} \cdot \left( -\frac{1}{Y_{\text{xy}}} \cdot \mu \cdot y + D \cdot (c_{\text{s, in}} - c_\text{s}) \right)
\]

(41)

Equations (40) and (42) can be rearranged to

\[
\mu \bigg|_{t=0} = D + \frac{\dot{y}(0)}{y(0)}
\]

\[
\frac{\partial \mu}{\partial c_\text{s}} \bigg|_{t=0} = \frac{\dot{y}(0) - (\dot{y}(0)^2/\dot{y}(0))}{\dot{y}(0) \cdot (1/(Y_{\text{xy}}) \cdot (D \cdot y(0) = \mu(0)
\quad + y(0) + D \cdot (c_{\text{s, in}} - c_\text{s}))}
\]

(42)

In other words, it is possible to express \( \mu \bigg|_{t=0} \) and \( \partial \mu/\partial c_\text{s} \bigg|_{t=0} \) by the known quantities \( y(0), \dot{y}(0), y(0) \) and \( c_\text{s}(0) \). From the definition of \( \mu \), it follows immediately that also \( K_m \) and \( \mu_m \) can be expressed uniquely by known quantities in the following way

\[
K_m = \frac{c_\text{s}(0)^2}{(\mu(0)/\mu_m(0)) - c_\text{s}(0)} \quad (43)
\]

\[
\mu_m = \frac{\mu(0) \cdot (K_m + c_\text{s}(0))}{c_\text{s}(0)} \quad (44)
\]

This shows that it is – at least in principle for the case of perfect measurement – possible to identify \( K_m \) and \( \mu_m \) from the available measurement information.

3.1.2 Confidence intervals of estimated parameters:

The results of the previous section are based on a theoretical concept and hence the parameter identifiability is only a necessary condition for successful PI. In reality, one has to
cope with a finite number of measurements taken at discrete time points and corrupted by noise that leads to an uncertainty in the estimated parameters. To assess the quality of the identification process, it is important to determine this uncertainty in terms of the covariance matrix $C_u$. In Section 2, three numerical methods have been discussed, two more traditional ones (FIM and Bootstrap) and the SP method as a novel approach. The three methods will be now compared by applying them to the simple bio-reactor example. The following assumptions are made:

- the measurement of the biomass concentration is taken at three time points $t_k = [0.5, 1.0, 1.5]$ h;
- artificial measurement data are used, which are obtained from a reference simulation and corrupted by normal distributed noise.

The reference values of the parameters are $K_s = 2$ and $\mu_m = 5$. The variance of the measurement noise is taken as $C_y = 10^{-4} [g^2 l^{-2}]$. This extremely small measurement error is chosen deliberately to demonstrate that the used methods provide different results in spite of almost unrealistically precise measurements. The computation is done in MATLAB, using the time integrator ODE15s and the optimiser LSQNONLIN, which uses the ‘Levenberg–Marquardt’ method for optimisation. All optimisations start with the reference values.

In a first step, the Bootstrap method is used. An intrinsic problem of the bootstrap approach is the correct choice of the number of samples, that is, the question which number of samples $B$ is sufficient to obtain accurate results for the mean value $\bar{\mu}$ of the estimated parameters as well as for their covariance matrix $C_u$.

To overcome this, the Bootstrap is done repeatedly with an increasing sample number $B$, so the characteristic values of the parameter statistics are a function of $B$ (Fig. 2). Apparently, the Bootstrap needs more than 2000 samples until a convergence is detectable. Only if the number of samples is increased to $B = 10,000$, the values of the mean and the covariances tend to stabilise. One can also see from Fig. 2 that the PI is slightly biased. The estimated mean values of $\mu_m$ and $K_s$ are slightly different from the reference values.

In a second step, the SP method is tested. For a scalar measurement ($n = 1$) and three measurement time points ($K = 3$), one needs $2mK + 1 = 7$ SPs, that is, 7 samples. Compared with the 10 000 samples of the Bootstrap, this is a tremendous reduction of the computational effort. Nevertheless, both methods produce nearly the same results. This indicates that the SP method works very well and is highly effective in this case.

In a third step, the FIM, evaluated at the reference values, is applied to the problem. As discussed in Section 2, the FIM is based on the assumption of an unbiased parameter estimation. Therefore it only provides information on the covariances but not on the mean values of the estimated parameters. A comparison of the 95% confidence intervals determined by the three methods is given in Fig. 3. It is evident that the FIM underestimates the parameter uncertainties strongly, especially the variance of $K_s$.

The results of the three methods are summarised in Table 1. Obviously, the SP method and the Bootstrap with 10 000 samples agree well, whereas the FIM only gives a poor estimate for the lower bound of the covariances, but no estimate at all of the mean values.
Both effects, the underestimation of the parameter uncertainties and the neglecting of the bias by the FIM increase with a growing measurement error (Fig. 4). The Bootstrap approach and the SP method provide almost the same results for smaller measurement errors ($C_y^0:082 \leq C_0^2$). For larger measurement errors, the Bootstrap approach encounters numerical problems as the optimiser fails for a number of samples. This is another point in favour of the SP method, which works reliably also for large measurement errors.

It is obvious that, at least for this example, the SP method clearly offers by far the best compromise between the accuracy and computational effort. The determined covariances can be used to investigate the correlation between the two parameters. All three methods calculate nearly the same correlation coefficient $r$, defined as

$$r_{ij} = \frac{C_{\theta_i \theta_j}}{\sqrt{C_{\theta_i \theta_i}} \cdot \sqrt{C_{\theta_j \theta_j}}} \quad (45)$$

The correlation coefficient is very close to unity (Table 1) which indicates a very strong correlation between the two parameters. This is clearly an unsatisfactory situation, but agrees with experimental findings that it is very hard to estimate $\mu_m$ and $K_s$ accurately from measurement data [1, 17]. The question arises if it is possible to find more favourable experimental conditions for the parameter estimation. This leads to the problem of OED, which is discussed in the next section.

### 3.1.3 Optimal experimental design:

Once the covariance matrix $C_\theta$ is known, the problem of OED, that is, the reduction of the parameter uncertainties by varying design parameters $u(t)$, can be addressed. The inlet flow $q_in(t)$ is defined as the design variable, so the objective is to find an optimal trajectory of $q_in(t)$, which maximises the measurement information content and minimises the parameter variances, respectively. This numerical optimisation problem can be solved, using one of the above-mentioned optimality criteria. As the correlation of $\mu_m$ and $K_s$ makes the estimation difficult, the $E^*$-criterion (10) is used to reduce both, the

<table>
<thead>
<tr>
<th>Before OED</th>
<th>FIM</th>
<th>SP</th>
<th>Bootstrap</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_{K_s}^2$</td>
<td>$0.1745 \times 10^{-3}$</td>
<td>$0.1694 \times 10^{-1}$</td>
<td>$0.1689 \times 10^{-1}$</td>
</tr>
<tr>
<td>$\sigma_{\mu_m}^2$</td>
<td>$0.2105 \times 10^{-3}$</td>
<td>$0.1076 \times 10^{-1}$</td>
<td>$0.1070 \times 10^{-1}$</td>
</tr>
<tr>
<td>$E[K_s]$</td>
<td>$-2.0028$</td>
<td>$2.0028$</td>
<td>$2.0023$</td>
</tr>
<tr>
<td>$E[\mu_m]$</td>
<td>$-5.0022$</td>
<td>$5.0022$</td>
<td>$5.0018$</td>
</tr>
<tr>
<td>$\rho$</td>
<td>$0.9959$</td>
<td>$0.9966$</td>
<td>$0.9965$</td>
</tr>
<tr>
<td>$E^*$-criterion</td>
<td>$4.9118 \times 10^2$</td>
<td>$6.1317 \times 10^2$</td>
<td>$6.0063 \times 10^2$</td>
</tr>
</tbody>
</table>
correlation and the parameter uncertainties. The theoretical
minimum of $\Phi_{E}(C_0)$ is 1, that is, $\lambda_{\text{max}}(C_0) = \lambda_{\text{min}}(C_0)$,
suggesting that there is no correlation between the parameters.

The inlet flow is defined as the following linear function

$$q(t) = a + b(t - c)$$

which is a very efficient inlet profile for the improvement of the
parameter accuracy of such a simple unstructured growth model
[17]. Now the OED problem consists in the optimal choice of
the parameters $a$, $b$, and $c$, which minimises the objective function
$\Phi_{E}(C_0)$. The covariance $C_u$ is computed either from the FIM
or by the SP method. The Bootstrap is not used for OED, as it
would require extremely long computational times. As the
covariances computed by FIM and SP differ, the two
methods provide different conditions for a new experiment

(Table 2). Consequently, the inlet flow $q_{\text{in}}(t)$ and the
corresponding growth rate $\mu(t)$ (Fig. 5) differ. The designed
experiments start with a higher inlet flow, leading to an
increased substrate concentration $c_{S}$. As a result, the growth
rate $\mu(c_{S}(t))$ is close to $\mu_{m}$, which should provide a better
estimation of the maximum growth rate, but without a benefit

![Figure 4](image-url) **Figure 4** 95% confidence intervals and expected values of the estimated parameters with respect to an increasing measurement error, that is, standard derivation (std)

![Figure 5](image-url) **Figure 5** Inlet flow $q_{\text{in}}$ and growth rate $\mu$ of the unoptimised (dark grey) and the two designed experiments: FIM (light grey) and SP (black)

| Parameter of the inlet flow function $q_{\text{in}}(t)$ before (INI) and after the OED for both methods (FIM and SPs) |
|---|---|---|
| | $a$ | $b$ | $c$ |
| INI | 0.100 | 0.000 | 0.000 |
| FIM | 0.925 | 1.257 | 0.992 |
| SP | 0.666 | 3.089 | 0.867 |
of \( K_S \). The SP method suggests an earlier and steeper increase of the inlet flow, which may allow a more precise estimation of the substrate limiting constant. To judge which of the two experiments suggested by the FIM and by the SP method is really the best, a Bootstrap is made for both experiments. The resulting scatter plot (Fig. 6) with 2000 samples per newly designed experiment shows that the SP method generates a smaller and more roundish point cloud of \( \hat{\theta} \), which is equivalent to a reduction of the parameter uncertainties and their correlation. Especially, the variance of \( K_S \) in the SP-designed experiment is smaller than in the FIM-designed experiment. The main reason for this seems to be that the inverse FIM underestimates the variance of \( K_S \). Therefore in the FIM-based OED, a high accuracy of the \( K_S \) estimate is implied, and no effort is made to increase it further. Furthermore, the scatter plot provides characteristic values of the parameter statistics (Table 3), which correspond very well with the results of the SP method. This demonstrates again the reliability of the SP method.

To illustrate that the process of PI does not end up in a local minimum, the contour plot of the sum of the weighted squared errors (4) is analysed (Fig. 7). In all three cases, that is, the unoptimised, the FIM optimised and the SP optimised experiment, the cost functions have only one global minimum. Therefore the application of a local optimiser does not pose a problem. Furthermore, the contour plots can also be used to assess the results of the OED independent of the used optimiser. As the SP method provides the most appropriate contour, that is, tight and roundish, the SP-designed experiment not only have a benefit on the Levenberg–Marquardt optimiser but also on any other minimum search algorithm.

### 3.1.4 Confidence regions of the dynamic states:

An additional advantage of the SP method is that it can also be used to investigate the influence of the parameter uncertainties on the simulation results. To do this, one has to use for \( \xi \) in (20) the estimated parameter vector \( \hat{\theta} \), \( g(\cdot) \) stands for the solution of the ODE system and \( \eta \) represents the state vector. The possibility to determine a confidence region of the states \( x_i \) is especially interesting in the field of observation and control, and also a coming up in systems biology [18]. Furthermore, the 95% confidence regions of the dynamic states (Fig. 8) obviously explain the benefit of the OED.

After a successful demonstration of the potential of the SP method, the OED is realised for a more complex model in the next section.

### 3.2 Incorporation of intracellular components

To demonstrate that the SP method is also applicable to more complex models, the simple bio-reactor model is extended by three intracellular components (\( c_m,1 \), \( c_m,2 \), \( c_m,3 \) [gl⁻¹]). The corresponding balance equations can be formulated as

\[
\begin{align*}
\dot{c}_x &= \mu \cdot c_x - q \cdot c_x \\
\dot{c}_s &= \frac{1}{Y_x} \cdot r_1 \cdot c_x + (c_{s,in} - c_s) \cdot q \\
\dot{c}_{m,1} &= r_1 - r_2 = \mu \cdot c_{m,1} \\
\dot{c}_{m,2} &= r_2 - r_3 = \mu \cdot c_{m,1} \\
\dot{c}_{m,3} &= r_3 - r_4 = \mu \cdot c_{m,1}
\end{align*}
\]

![Figure 6](image)  
**Figure 6** Scatter diagram of the optimised parameter estimation  
The SP-designed experiment provides a tighter and more roundish point cloud (black dots) than the use of FIM (grey dots), indicating that the SP enable a higher information content and an improved parameter estimation, respectively.
where the rates are a set of mass action and Michaelis–Menten kinetics

\[
\begin{align*}
    r_1 &= \frac{m_{m,1} \cdot c_s}{K_{s,1} + c_s} \\
    r_2 &= \frac{m_{m,2} \cdot c_{m,1}}{K_{s,2} + c_{m,1}} \\
    r_3 &= \frac{m_{m,3} \cdot c_{m,2}}{K_{s,3} + c_{m,2}} \\
    r_4 &= \frac{m_{m,4} \cdot c_{m,3}}{c_m,3}
\end{align*}
\]

(52)

(53)

(54)

(55)

Consequently, the number of unknown parameters increases from two to eight \( \theta = [m_{m,1} \ K_{s,1} \ m_{m,2} \ K_{s,2} \ m_{m,3} \ K_{s,3} \ m_{m,4} \ n] \)

Following assumptions about the measurement information are made:

- the concentration of every component is measured;
- measurement samples are taken at four times \( t_k = [0.2 \ 0.8 \ 1.4 \ 2.0] \ h \);
- artificial measurement data are used, which are obtained from a reference simulation and corrupted by normal distributed noise;
- the variances of the measurement noise are assumed as \( C_{\gamma,ii} = 0.1^2 \) for all components.

For the measurement vector \( y(m = 5) \) and the four measurement time points \( (K = 4) \), the number of SPs increases to \( 2mK + 1 = 41 \). Compared with the at least 1000 samples of a Bootstrap approach, it is still a considerable reduction of computational effort.

### 3.2.1 Determination and reduction of the parameter uncertainties

After a first PI, the statistics of the parameters are determined. As the Bootstrap approach requires too much computational time, only the FIM and the SP method are applied to the extended model. In Fig. 9, the 95% confidence regions of the estimated parameters \( ^\wedge \theta \) are shown. Obviously, the assumption of an unbiased estimator is not met; furthermore, the confidence regions do not agree, that is, the covariance matrix \( C_{\theta} \) calculated by the FIM and SP method are different. The previous sections pointed out clearly that the inverse of the FIM provides a relative unrealistic approximation of the actual parameter statistics, hence the OED is realised with the SP method in the next step. Once again the inlet flow \( q(t) \) is defined as the design variable, to minimise the parameter uncertainties. To show that the accuracy of the parameter estimation can be improved with a minimal effort, \( q(t) \) is assumed as a constant \( q_{opt} = 0.336 \) [lh/h]

Using the A-optimal design criteria (6) for the OED process, the SP method provides \( q_{opt} = 0.336 \) [lh/h] as the optimal choice of the design variable. A new determination of the parameter covariance matrix via the SP method for the optimally designed experiment indicates (Fig. 10) that it is not possible to reduce the parameter uncertainties of

| Table 3 Characteristic values of the parameter statistics after the OED of the FIM-designed and the SP-designed experiment |
|---------------------------------|-----------------|-----------------|-----------------|
|                                 | FIM             | SP              | Scatter plot    |
| **FIM-designed experiment**     |                 |                 |                 |
| \( \sigma^2_{K_s} \)           | 0.2667 \times 10^{-5} | 0.9842 \times 10^{-3} | 0.9891 \times 10^{-3} |
| \( \sigma^2_{\mu_m} \)         | 0.0453 \times 10^{-5} | 0.0609 \times 10^{-3} | 0.0611 \times 10^{-3} |
| \( E[K_s] \)                    | -               | 2.0026          | 2.0020          |
| \( E[\mu_m] \)                  | -               | 5.0007          | 5.0005          |
| \( \rho \)                      | 0.9694          | 0.9967          | 0.9971          |
| \( E^\wedge \)-criterion       | 1.3159 \times 10^2 | 2.9919 \times 10^3 | 3.1406 \times 10^3 |
| **SP-designed experiment**     |                 |                 |                 |
| \( \sigma^2_{K_s} \)           | 0.5122 \times 10^{-5} | 0.2087 \times 10^{-3} | 0.2130 \times 10^{-5} |
| \( \sigma^2_{\mu_m} \)         | 0.1410 \times 10^{-5} | 0.0462 \times 10^{-3} | 0.0452 \times 10^{-5} |
| \( E[K_s] \)                    | -               | 1.9990          | 1.9999          |
| \( E[\mu_m] \)                  | -               | 4.9996          | 4.9999          |
| \( \rho \)                      | 0.9891          | 0.8254          | 0.8584          |
| \( E^\wedge \)-criterion       | 2.7148 \times 10^2 | 24.5355         | 24.2887         |

For both results, the values are determined via the FIM, SP method and the Bootstrap approach (scatter plot)
all parameters in one step considerably, for example, the uncertainty of the parameter $n$ is increased. This result is a well-known problem in the field of OED, which can be solved by estimating only the ‘important’ subset $\theta_i \subseteq \theta$ of the unknown parameters, whereby the $\theta_j = \theta \setminus \theta_i$ parameters are assumed to be known and not part of the estimation [19].

Nevertheless, the global benefit of the OED to the dynamic states $\mathbf{x}(t)$ is much bigger than the confidence regions of the

![Figure 7](image1.png)

**Figure 7** Contour plots of the parameter estimator cost function (sum of the weighted squared errors) for the unoptimised, the FIM-optimised and the SP-optimised experiment

![Figure 8](image2.png)

**Figure 8** 95% confidence regions of the dynamic states: grey for the unoptimised and black for the SP-optimised parameter estimation
estimated parameter (Fig. 9) suggest. The uncertainties of all dynamic states are reduced obviously (Fig. 11). This effect agrees well with the assumption that only a few linear combinations of the unknown parameters determine the qualitative behaviour of the a model [19, 20]. The objective of OED, obtaining models with an increased predictive power, could be achieved with a minimal effort, for example, minimal cost and minimal technical equipment.
4 Conclusions

In this paper, we have presented a general, systematic procedure to calculate and minimise the parameter covariance matrix $C_\theta$ for models that are nonlinear with respect to their parameters. Starting with the determination of $C_\theta$ to assess the quality of a PI, the disadvantages of the traditionally used FIM method become clear. The Bootstrap approach and the SP method are able to approximate $C_\theta$ in a much more realistic way.

If the FIM is used for an OED, the process is designed under assumptions that are not met by many dynamical systems, leading to a sub-optimal choice of the design variables. Owing to the fact that the computational effort of the Bootstrap approach prohibits its application to the OED process, this leaves the SP method as an attractive alternative for determining and minimising the parameter uncertainties. Further, confidence regions of the dynamic states can be obtained via the SP method. This makes the benefit of OED obvious and enables novel approaches of cost functions to improve the predictive power of models.

5 Acknowledgment

We gratefully acknowledge the financial support by Forschungszentrum für Dynamische Systeme in Biomedizin und Prozesstechnik, Sachsen-Anhalt.

6 References


