Simultaneous Solution Approach to Model-Based Experimental Design

M. D. Hoang, T. Barz, and V. A. Merchan
Chair of Process Dynamics and Operation, Sekr. KWT-9, Berlin Institute of Technology, Str. des 17 Juni 135, D-10623 Berlin, Germany

L. T. Biegler
Dept. of Chemical Engineering, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213

H. Arellano-Garcia
School of Design, Engineering and Technology, University of Bradford, Richmond Road, Bradford, BD7 1DP, U.K.

DOI 10.1002/aic.14145
Published online July 1, 2013 in Wiley Online Library (wileyonlinelibrary.com)

A model-based experimental design is formulated and solved as a large-scale NLP problem. The key idea of the proposed approach is the extension of model equations with sensitivity equations forming an extended sensitivities-state equation system. The resulting system is then totally discretized and simultaneously solved as constraints of the NLP problem. Thereby, higher derivatives of the parameter sensitivities with respect to the control variables are directly calculated and exact. This is an advantage in comparison with proposed sequential approaches to model-based experimental design so far, where these derivatives have to be additionally integrated throughout the optimization steps. This can end up in a high-computational load especially for systems with many control variables. Furthermore, an advanced sampling strategy is proposed which combines the strength of the optimal sampling design and the variation of the collocation element lengths while fully using the entire optimization space of the simultaneous formulation. © 2013 American Institute of Chemical Engineers *AIChE J*, 59: 4169–4183, 2013
Keywords: model development, experimental analysis, simultaneous approach, dynamic optimization

Introduction

In the last two decades model-based experimental design has significantly gained on importance in the fields of chemical, biochemical, and process engineering.[1–4] By reducing the experimental effort required to develop high-fidelity predictive models, model-based experimental design plays a major role in the propagation of advanced model-based methodologies, which enable the optimization of different aspects related to process design and operation.

The central part of a reliable predictive model is the set of accurate parameter values. It can become very expensive in terms of time and resources depending on the system complexity and the difficulty of data collection, in particular regarding processes where the prediction of the system dynamic behavior is most important. Here, statistical design of an experiment reaches its limits since it cannot describe dynamical processes properly and is especially not capable of considering process constraints, which are crucial in maintaining process stability. This is also why model-based experimental design has decisively gained in importance in process and chemical engineering. Here, the experimental design task is formulated as an optimal control problem using dynamic control trajectories so as to perturb the process with the goal of enabling or improving estimability of model parameters.[5–7]

For the solution of the experimental design as an optimal-control problem, the sequential optimization method[8,9] and multiple shooting methods[10,11] have widely been used as state of the art approaches. Efficient DAE solvers are the backbone of both approaches, which solve the system model with the associated sensitivity equations providing time dependent states and gradient information.[12–14] The subsequent connection of the integration results to an optimization algorithm can be straightforwardly carried out. Disadvantages of the sequential method appear in connection with unstable systems, since computed control outputs of the optimization algorithm may lead the connected DAE-solver into instability. On the other hand, when using multiple shooting the functions of the state variables are divided into small time periods, in which the models are then also solved by DAE-solvers. In contrast to the sequential approach, instability and poor conditioning of the problem can then be avoided[15] since additional inequality constraints can be formulated so as to set bounds on state variables. However, this feature only yields for the endpoints of each element. Both sequential optimization and the multiple shooting approaches are based on DAE solvers, and, thus, the solution of the extended state-sensitivity equation system has to be reintegrated for each optimization step. This can sometimes require an extensive

Correspondence concerning this article should be addressed to H. Arellano-Garcia at h.arellano-garcia@bradford.ac.uk.

computational load since the integration of state variables and target sensitivities constitute the main costs of the optimization process, in particular if the system model includes many states, parameters, and especially control variables. Another common disadvantage of both methods relates to constraints on state variables as mentioned, which therein can only be considered indirectly by approximations.

These bottlenecks can be avoided by using the simultaneous optimization approach.[16,17] At our best knowledge, the first application of experimental design in the sense of large-scale optimization was carried out regarding precise parameter estimation and model discrimination for temperature programmed reduction experiments.[18] In particular, the D-criterion was used with an equidistant sampling strategy and fixed optimization end-time highlighting the potential of the simultaneous formulation in solving model-based experimental design problems. Nevertheless, this application example does not show the whole strength of the simultaneous optimization methodology. The crucial point of the simultaneous formulation is the total discretization of the extended sensitivities-state equation system as equality constraints of a NLP-optimization problem. It is thereby possible to calculate the derivatives of the experimental design criteria directly as functions of the discretized sensitivity variables and to exploit the sparse-structure of the constraint derivatives to a full capacity. Inequality conditions for state and control variables can directly be embedded into the formulation over the whole domain. A further advantage is the implementation of control functions of flexible order, which can be treated straightforwardly due to the same discretization scheme as the state variables. Furthermore, the fact that the number of samplings in practice are much smaller than the number of discretization elements has to be taken into account.

This work presents a detailed analysis of the aforementioned aspects, and additionally provides an adaptive optimal sampling strategy, which combines the strength of the optimal sampling design and the variation of the collocation element lengths, so as to exploit the entire optimization space of the simultaneous formulation. The applicability of the approach on dynamic processes and its effectiveness regarding the results of the experimental design is demonstrated based on the application to an illustrative example of a biomass reactor.

Solution Strategy

In contrast to statistical design of experiments, mechanistic models are used here to describe thermodynamic and chemical phenomena in chemical and process engineering applications. In general, a DAE-system consisting of equations, which describe mass and energy balance, thermodynamic and mass-transport phenomena, is formulated and solved. In model-based experimental design objective functions are optimized under the consideration of model equations as equality constraints. The objective functions of the experimental design usually depend on the sensitivities of the state variables with respect to the parameters.

Extension of the system model with sensitivity equations

In chemical engineering process models can generally be formulated as DAE-systems in a linear-implicit form

$$B\left(\bar{z}(t),\bar{u}(t),\bar{\theta}\right)\dot{\bar{z}}(t)=\bar{f}\left(\bar{z}(t),\bar{u}(t),\bar{\theta},t\right),\bar{z}(t=0)=\bar{z}_0 \quad (1)$$

In this case $\bar{z}\in R^{N_z}$ represents state variables $\bar{u}\in R^{N_u}$, control variables, and $\bar{\theta}\in R^{N_\theta}$ the parameters. The partial derivative with respect to the differential variable $\dot{\bar{z}}$ is usually called the mass matrix $B(\cdot)\in R^{N_z\times N_z}$ It is independent from $\dot{\bar{z}}$ for linear-implicit systems and might be singular in the presence of algebraic equations. To derive the parameter sensitivities, Eq. 1 is reformulated to

$$\bar{g}\left(\dot{\bar{z}},\bar{z},\bar{u},\bar{\theta},t\right)=B(\cdot)\dot{\bar{z}}(t)-\bar{f}\left(\bar{z}(t),\bar{u}(t),\bar{\theta},t\right)=0_{N_z\times 1} \quad (2)$$

and the total differential of Eq. 2 is then formed with respect to $\bar{\theta}$

$$\frac{d\bar{g}}{d\bar{\theta}}=\frac{\partial\bar{g}}{\partial\dot{\bar{z}}}\frac{\partial\dot{\bar{z}}}{\partial\bar{\theta}}+\frac{\partial\bar{g}}{\partial\bar{z}}\frac{\partial\bar{z}}{\partial\bar{\theta}}+\frac{\partial\bar{g}}{\partial\bar{\theta}}=0_{N_z\times N_\theta} \quad (3)$$

Finally, Eq. 3 results in

$$0_{N_z\times N_\theta}=B\frac{\partial\dot{\bar{z}}}{\partial\bar{\theta}}+\frac{\partial(B\dot{\bar{z}}-\bar{f})}{\partial\bar{z}}\frac{\partial\bar{z}}{\partial\bar{\theta}}+\frac{\partial(B\dot{\bar{z}}-\bar{f})}{\partial\bar{\theta}} \quad (4)$$

The partial derivative with respect to the state variables is defined as J_z

$$J_z\left(\dot{\bar{z}}(t),\bar{z}(t),\bar{u}(t),\bar{\theta},t\right):=\left(\frac{\partial(B\dot{\bar{z}}(t))}{\partial\bar{z}}-\frac{\partial\bar{f}}{\partial\bar{z}}\right)\in R^{N_z\times N_z} \quad (5)$$

The partial derivative with respect to the parameters is defined as J_θ

$$J_\theta\left(\dot{\bar{z}}(t),\bar{z}(t),\bar{u}(t),\bar{\theta},t\right):=\left(\frac{\partial(B\dot{\bar{z}}(t))}{\partial\bar{\theta}}-\frac{\partial\bar{f}}{\partial\bar{\theta}}\right)\in R^{N_z\times N_\theta} \quad (6)$$

The sensitivity matrix S is then defined as

$$S(t):=(\bar{s}_1,\cdots,\bar{s}_{N_\theta})\in R^{N_z\times N_\theta} \quad (7)$$

where each column $\bar{s}_i$ corresponds to the sensitivities with respect to the parameter θ_i

$$\bar{s}_i:=\frac{\partial\bar{z}}{\partial\theta_i}=\begin{pmatrix}\frac{\partial z_1}{\partial\theta_i}\\ \vdots\\ \frac{\partial z_{N_z}}{\partial\theta_i}\end{pmatrix}\in R^{N_z\times 1},i=1,\ldots,N_\theta \quad (8)$$

By using the definitions of S, J_z and J_θ one can rewrite Eq. 4 to

$$0_{N_z\times N_\theta}=B\dot{S}(t)+J_zS(t)+J_\theta \quad (9)$$

By stacking the columns of $S(t)$, Eq. 9 can be reformulated to a form, which is more suitable for optimization tasks

$$0_{N_z\cdot N_\theta\times 1}=\begin{pmatrix}\bar{g}_{\theta_1}\\ \vdots\\ \bar{g}_{\theta_{N_\theta}}\end{pmatrix}:=\begin{pmatrix}B\cdot\dot{\bar{s}}_1\\ \vdots\\ B\cdot\dot{\bar{s}}_{N_\theta}\end{pmatrix}+\begin{pmatrix}J_z\cdot\bar{s}_1\\ \vdots\\ J_z\cdot\bar{s}_{N_\theta}\end{pmatrix}+\begin{pmatrix}J_{\theta_1}\\ \vdots\\ J_{\theta_{N_\theta}}\end{pmatrix} \quad (10)$$

with

$$\begin{pmatrix}\bar{f}_{\theta_1}\\ \vdots\\ \bar{f}_{\theta_{N_\theta}}\end{pmatrix}:=\begin{pmatrix}J_z\cdot\bar{s}_1\\ \vdots\\ J_z\cdot\bar{s}_{N_\theta}\end{pmatrix}+\begin{pmatrix}J_{\theta_1}\\ \vdots\\ J_{\theta_{N_\theta}}\end{pmatrix} \quad (11)$$

and

$$\bar{g}_\theta := \begin{pmatrix} \bar{g}_{\theta_1} \\ \vdots \\ \bar{g}_{\theta_{N_\theta}} \end{pmatrix} = 0_{N_z \cdot N_\theta \times 1} \tag{12}$$

Now by combining Eqs. 12 and 2, an extended system can be formed to

$$\begin{aligned} 0_{N_z(1+N_\theta)\times 1} &= \bar{c}\left(\dot{\bar{z}}(t), \bar{z}(t), \dot{\bar{s}}_1, \ldots, \dot{\bar{s}}_{N_\theta}, \bar{s}_1, \ldots, \bar{s}_{N_\theta}, \bar{u}(t), \bar{\theta}, t\right) \\ &:= \begin{bmatrix} \bar{g}\left(\dot{\bar{z}}(t), \bar{z}(t), \bar{u}(t), \bar{\theta}, t\right) \\ \\ \bar{g}_\theta\left(\dot{\bar{s}}_1, \ldots, \dot{\bar{s}}_{N_\theta}, \bar{s}_1, \ldots, \bar{s}_{N_\theta}, \dot{\bar{z}}(t), \bar{z}(t), \bar{u}(t), \bar{\theta}, t\right) \end{bmatrix} \\ \bar{z}(t=0) &= \bar{z}_0, \ \bar{s}_i(t=0) = \frac{\partial \bar{z}_0}{\partial \theta_i}, \ i=1, \ldots, N_\theta \end{aligned} \tag{13}$$

Formulation of the Model-Based Experimental Design as an NLP Problem

Formulation as an optimal control problem

In model-based experimental design, a vector of predicted response variables $\bar{y}(t) \in R^{N_y}$ is considered, whose elements usually are a subset of the state variables $\bar{z}(t)$. More generally, predicted response variables $\bar{y}(t)$ are formulated as nonlinear functions of the state variables

$$\bar{y}(t) = \bar{m}(\bar{z}(t)) \tag{14}$$

Their parameter sensitivities $\bar{S}_y(t) \in R^{N_y \times N_\theta}$ can be calculated via the chain rule and written down directly as nonlinear function of the existing state variables $\bar{z}(t)$, and sensitivities $S(t)$ with respect to $\bar{z}(t)$

$$\bar{S}_y(t) = \begin{bmatrix} \frac{\partial m_1}{\partial z_1} & \cdots & \frac{\partial m_1}{\partial z_{N_z}} \\ \vdots & \ddots & \vdots \\ \frac{\partial m_{N_y}}{\partial z_1} & \cdots & \frac{\partial m_{N_y}}{\partial z_{N_z}} \end{bmatrix} \cdot [\bar{s}_1(t) \cdots \bar{s}_{N_\theta}(t)] = J_z^h \cdot S(t) \tag{15}$$

Both Eqs. 14 and 15 can be straightforwardly included into the existing differential-algebraic formulation DAE2 in (13). For the sake of simplicity and without loss of generality it is assumed in this article that $\bar{y}(t)$ is a subset of $\bar{z}(t)$, and in particular $N_y = N_z$. Therefore, $\bar{z}(t)$ is used when referring to the predicted response variables. Moreover, the predicted responses are generally collected at discrete points in time $t_{sp} \in R^{N_{sp}}$ according to sampling points in the experiments

$$\bar{z}_{sp,k} = \bar{z}(t_k), \quad k=1, \ldots, N_{sp} \tag{16}$$

The information content of an experiment can be represented by the Fischer information matrix $M_F \in R^{N_\theta \times N_\theta}$. For time-dependent systems with discrete measurements, the definition is given as follows

$$M_F(S_{sp}) = (S_{sp})^T \cdot \sum_z \cdot S_{sp} \tag{17}$$

Where $\sum_z \in R^{N_z \cdot N_{sp} \times N_z \cdot N_{sp}}$ denotes the measurement covariance matrix, and $S_{sp} \in R^{N_z \cdot N_{sp} \times N_\theta}$ represents the dynamic sensitivity sampling matrix

$$S_{sp} = \begin{bmatrix} \omega_1 S(t_1) \\ \vdots \\ \omega_{N_{sp}} S(t_{N_{sp}}) \end{bmatrix} = \begin{bmatrix} \omega_1 \bar{s}_{1,t_1} & \omega_1 \bar{s}_{2,t_1} & \cdots & \omega_1 \bar{s}_{N_p,t_1} \\ \omega_2 \bar{s}_{1,t_2} & \omega_2 \bar{s}_{2,t_2} & \cdots & \omega_2 \bar{s}_{N_p,t_2} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{N_{sp}} \bar{s}_{1,t_{N_{sp}}} & \omega_{N_{sp}} \bar{s}_{2,t_{N_{sp}}} & \cdots & \omega_{N_{sp}} \bar{s}_{N_p,t_{N_{sp}}} \end{bmatrix} \tag{18}$$

Furthermore, a proper sampling strategy is a crucial issue in model-based experimental design. For this purpose, the optimal sampling strategy[19] is applied for this simultaneous optimization approach. Therefore, each block $S(t_i)$ of the dynamic sensitivity matrix S_{sp} is weighted with a continuous and bounded control variable $\omega \in R, 0 \le \omega \le 1$.

A series of optimal criteria exists in the literature for model-based experimental design, which aims at maximizing system information content or minimizing the parameter correlations. The optimization task is mostly equivalent to the maximization of an appropriate measure of the Fischer information matrix M_F or minimization of an appropriate measure of the covariance matrix V. Moreover, the following relation exists between the two measures

$$V \cong \left[S_{sp}^T \sum_z S_{sp}\right]^{-1} = M_F^{-1} \tag{19}$$

The meaning of Eq. 17 is that according to the Cramer-Rao theorem, the lower bound of the covariance matrix can be estimated with the inverse of the Fisher information matrix.[20] It has been shown[19] based on the variational principle that the optimal sampling strategy always results in a physically meaningful solution, namely 0 or 1 for the weighting control variables ω, if they enter linearly into the variational formulation. Regarding the simultaneous optimization approach, this idea is equivalent with a strict monotonicity of the weighting control variables in the objective functions (22). It can be shown by a simple symbolic calculation of the objective functions that this can be guaranteed for the formulation with the Fisher information matrix, but not with the covariance matrix. Thus, the formulation with the Fisher information matrix is strongly suggested regarding the simultaneous optimization approach. Moreover, a limited sampling number $N_{sp,U}$ is introduced so as to consider the fact that the sampling number is in general much smaller in practice than the theoretically maximum one N_{sp}. Therefore, due to the optimal sampling strategy theory and as a result of the optimization, exact $N_{sp,U}$ of the total weighting control N_{sp} variables ω become 1 and all the others 0.

Scaling

The Fisher information matrix depends strongly on a scaling of the parameter sensitivities. Since its formulation is based on the absolute parameter values, the influence of parameters with high values is much bigger than those with small values. Therefore, parameter sensitivities have to be scaled so as to take into account the dimension gap of their different nature. A common way to scale sensitivities regarding a parameter ω_i is to multiply with the parameter itself [7]

$$\tilde{s}_i := \theta_i s_i = \theta_i \frac{\partial z}{\partial \theta_i} \tag{20}$$

Thus, the Fisher information matrix can be reformulated to

$$M_F\left(S_{sp}(\tilde{s})\right)=\left(S_{sp}(\tilde{s})\right)^T\cdot\sum\nolimits_z\cdot S_{sp}(\tilde{s}) \tag{21}$$

For the sake of clarity, the scaled Fisher information matrix is also referred as M_F. The reader is asked to keep in mind that all the following equations and calculations are based on scaled sensitivities. The best known and most frequently used objective functions regarding the Fischer information matrix are

$$\begin{aligned}\Phi_A\left(S_{sp}\right)&=tr\left(M_F\left(S_{sp}\right)\right)\\ \Phi_D\left(S_{sp}\right)&=\det\left(M_F\left(S_{sp}\right)\right)\\ \Phi_E\left(S_{sp}\right)&=\lambda_{\min}\left(M_F\left(S_{sp}\right)\right)\end{aligned} \tag{22}$$

The model-based experimental design is formulated as an optimal control problem, where the objective function is a function of the Fischer information matrix and the extended state-sensitivity equation system DAE2 represents the constraints

$$\max_{\bar{u}(t),\bar{\omega},\bar{Z}_0}\Phi(M_F(\bar{s},\bar{\omega})) \tag{23}$$

Subject to:

$$\begin{aligned}&0_{(N_z+1)\cdot N_\theta\times 1}=\bar{c}\left(\dot{\bar{z}}(t),\bar{z}(t),\dot{\bar{s}}_1,\ldots,\dot{\bar{s}}_{N_\theta},\bar{s}_1,\ldots,\bar{s}_{N_\theta},\bar{u}(t),\bar{\theta},t\right)\\ &\bar{z}(t=0)=\bar{z}_0,\bar{s}_i(t=0)=\frac{\partial\bar{z}_0}{\partial\theta_i},i=1,\ldots,N_\theta\\ &\sum\omega_k=N_{sp,U},0\le\omega_k\le 1,k=1,\ldots,N_{sp}\\ &\bar{z}_{0,L}\le\bar{z}(t)\le\bar{z}_{0,U},\bar{z}_L\le\bar{z}(t)\le\bar{z}_U,\bar{u}_L\le\bar{u}(t)\le\bar{u}_U\end{aligned} \tag{24}$$

Discretization procedure

All equations of DAE2 in Eq. 13 are discretized using the orthogonal collocation on FEM (OCFEM), and written in the form of residuals with normalized time τ_j (see Appendix A for further reference)

$$\begin{aligned}&0_{N_z\cdot N_e\cdot K\times 1}=\bar{g}_l(\tau_j)=\sum_{j=0}^{k}B\bar{z}_{lj}\dot{\varphi}_j(\tau_j)-h_l\bar{f}\left(\bar{z}_{lj},\bar{u}_{lj},\bar{\theta},\tau_j\right)\\ &0_{N_z\cdot N_\theta\cdot N_e\cdot K\times 1}=\bar{g}_{\theta\cdot l}(\tau_j)=\begin{pmatrix}\sum_{j=0}^{k}B\bar{s}_{1,lj}\dot{\varphi}_j(\tau_j)+h_l\bar{f}_\theta\left(\bar{z}_{lj},\bar{s}_{1,lj},\bar{u}_{lj},\bar{\theta},\tau_j\right)\\ \vdots\\ \sum_{j=0}^{k}B\bar{s}_{N_\theta,lj}\dot{\varphi}_j(\tau_j)+h_l\bar{f}_\theta\left(\bar{z}_{lj},\bar{s}_{N_\theta,lj},\bar{u}_{lj},\bar{\theta},\tau_j\right)\end{pmatrix}\\ &j=1,\ldots,k;l=1,\ldots,N_e\end{aligned} \tag{25}$$

Furthermore, continuity conditions ensure that no discontinuities arise between the elements of the state functions. For the Radau collocation method,[21] the following continuity conditions are without loss of generality

$$\begin{aligned}&\bar{z}_{l+1j=0}=\bar{z}_{lj=k}\\ &\bar{s}_{i,l+1j=0}=\bar{s}_{i,lj=k}\\ &i=1,\ldots,N_\theta;l=1,\ldots,N_e-1\end{aligned} \tag{26}$$

From Eq. 25, the fully discretized system model results in

$$\begin{aligned}&\begin{pmatrix}0_{N_z\cdot N_e\cdot K\times 1}\\ 0_{N_z\cdot N_\theta\cdot N_e\cdot K\times 1}\end{pmatrix}=\begin{pmatrix}\bar{g}_l(\tau_j,h_l)\\ \bar{g}_{\theta,l}(\tau_j,h_l)\end{pmatrix}=\bar{c}\left(\bar{Z}_{lj},\bar{S}_{1,lj},\ldots,\bar{S}_{N_\theta,lj},\bar{u}_{lj},\bar{\theta},\tau_j h_l\right)\\ &\bar{c}\in R^{(N_z+N_z\cdot N_\theta)\cdot N_e\cdot K\times 1}\\ &l=1,\ldots,N_e;\ j=1,\ldots,K\end{aligned} \tag{27}$$

$$\bar{z}_{11}=\bar{z}_0,\bar{s}_{i,11}=\frac{\partial\bar{z}_0}{\partial\theta_i},i=1,\ldots,N_\theta \tag{28}$$

$$\bar{z}_{l+1j=0}=\bar{z}_{lj=K},\bar{s}_{i,l+1j=0}=\bar{s}_{i,lj=K},l=1,\ldots,N_e-1 \tag{29}$$

By converting all time-dependent functions into discrete variables, the optimization problem OP1 can now be formulated as a NLP optimization problem with DAE3 as constraints, and, thus, resulting in

$$\max_{\bar{z}_{lj},\bar{s}_{1,lj},\ldots,\bar{s}_{N_\theta,lj},\bar{u}_{lj},h_l,\bar{\omega},\bar{z}_0}\Phi(M_F(\bar{s},\bar{\omega})) \tag{30}$$

Subject to:

$$\begin{aligned}&0=\bar{c}\left(\bar{z}_{lj},\bar{s}_{1,lj},\ldots,\bar{s}_{N_\theta,lj},\bar{u}_{lj},\bar{\theta},\tau_j,h_l\right)\\ &l=1,\ldots,N_e;j=1,\ldots,K\\ &\bar{z}_{11}=\bar{z}_0,\bar{s}_{i,11}=\frac{\partial\bar{z}_0}{\partial\theta_i},i=1,\ldots,N_\theta\\ &\bar{z}_{l+1j=0}=\bar{z}_{lj=K},\bar{s}_{i,l+1j=0}=\bar{s}_{i,lj=K},l=1,\ldots,N_e-1\\ &\sum h_l=t_f,t_f\le t_{f,\max},h_L\le h_l<h_U,l=1,\ldots,N_e\\ &\sum\omega_k=N_{sp,U},0\le\omega_k\le 1,k=1,\ldots,N_{sp}\\ &\bar{z}_{0,L}\le\bar{z}_0\le\bar{z}_{0,U},\bar{z}_L\le\bar{z}_{lj}\le\bar{z}_U,\bar{u}_L\le\bar{u}_{lj}\le\bar{u}_U\end{aligned} \tag{31}$$

It should be noted that in the standard optimization tasks no continuity conditions on the control variables are demanded. This generally leads to a bang-bang characteristic of the optimal control trajectories. However, in some cases those solutions are not wanted because of their characteristically instantaneous jumps. There are many reasons why experimenters prefer smooth control profiles, e.g., because the control element cannot carry out instantaneous jumps, for instance, in temperature- and flow-control, or because there are risks that the process can become instable with instantaneous and large changes of the control outputs. To overcome this, the simultaneous approach allows a flexible and convenient way to use high order and continuous control trajectories. Since the control functions are also discretized with OCFEM, one is free to choose different orders for each control variable.

Another crucial point is a flexible sampling strategy which has been realized with a variable collocation element length h_l as additional degree of freedoms. One way to treat this problem is the method of superelement structure.[22] In this work, the collocation element length h_l is rather treated directly as optimization variables. An optimization of the end time is thereby directly embedded over an algebraic equation with respect to the element lengths h_l and the end time t_f and a given upper bound to the end time $t_{f,\max}$, respectively.

Furthermore, the handling of the optimization of the initial states is straightforward since the optimization problem is fully discretized. They are treated in the same manner as the discretized dynamic controls.

Exploiting the structure of the constraints derivatives

For the sake of clarity, the derivation of the constraints derivatives is applied on Eq. 20 with respect to the optimization variables $[\bar{z}(t), \bar{s}_1(t), \ldots, \bar{s}_{N_\theta}(t), \bar{u}(t)]$. Since only the objective function depends on $\bar{\omega}$ but not the constraints, all constraint derivatives with respect to $\bar{\omega}$ become zero and will not be included in the following derivation.

In order to obtain the derivatives with respect to the fully discretized optimization variables $\left[\bar{z}_{lj}, \bar{s}_{1,lj}, \ldots, \bar{s}_{N_\theta,lj}, \bar{u}_{lj}\right]$, one more derivation step over the discretization Eqs. A2–A5 is necessary and is reserved for the interested reader (see equations in Appendix A). More useful is through the information about the extremely sparse structure of the derivatives, and the fact that most of them can be reused from previous calculations, which is illustrated in the following. The optimization variables of OP1 are represented by

$$\bar{x} := \begin{pmatrix} \bar{Z} \\ \bar{S}_1 \\ \vdots \\ \bar{S}_{N_\theta} \\ \bar{u} \end{pmatrix} \in R^{(N_z+N_z\cdot N_\theta+N_u)\times 1} \tag{32}$$

First-order constraint-derivatives

The first-order derivative of the constraints with respect to the iteration variables $\bar{x}$ is then

$$\frac{\partial \bar{c}}{\partial \bar{x}} = \frac{\partial}{\partial \bar{x}}\begin{pmatrix} \bar{g} \\ \bar{g}_\theta \end{pmatrix} = \begin{bmatrix} J_z & 0 & 0 & \cdots & 0 & J_u \\ J_z^{\theta_1} & J_s^{\theta_1} & 0 & \vdots & 0 & J_u^{\theta_1} \\ J_z^{\theta_2} & 0 & J_s^{\theta_{N_2}} & \ddots & 0 & J_u^{\theta_2} \\ \vdots & \vdots & \cdots & \ddots & \vdots & \vdots \\ \underbrace{J_z^{\theta_{N_\theta}}}_{w.r.t.\ \bar{z}} & \underbrace{0 \quad \ldots \quad \ldots \quad J_s^{\theta_{N_2}}}_{w.r.t.\ \bar{s}} & & & & \underbrace{J_u^{\theta_{N_\theta}}}_{w.r.t.\ \bar{u}} \end{bmatrix}$$
$$\frac{\partial \bar{c}}{\partial \bar{x}} \in R^{(N_z+N_z\cdot N_\theta)\times(N_z+N_z\cdot N_\theta+N_u)} \tag{33}$$

The first-order derivative of the system equations with respect to the state variables $\bar{z}$ is the Jacobian J_z defined in Eq. 5, which has already been calculated to formulate the sensitivity equations $\bar{g}_\theta$, and, thus, it can be fully reused. The derivatives of $\bar{g}$ regarding the control variables $\bar{u}$, are represented by

$$J_u\left(\dot{\bar{z}}(t), \bar{z}(t), \bar{u}(t), \bar{\theta}, t\right) := \frac{\partial \bar{g}}{\partial \bar{u}} = \left(\partial\left(\frac{B\dot{\bar{z}}(t)}{\partial \bar{u}} - \frac{\partial \bar{f}}{\partial \bar{u}}\right)J_u \in R^{N_z\times N_u}\right. \tag{34}$$

Furthermore, the derivative of the i-th part of $\bar{f}_\theta$ in Eq. 11 with respect to the sensitivities $\bar{s}_i$ is also J_z since $\bar{s}_i$ is linear in $\bar{f}_{\theta_i}$ leading to

$$J_s^{\theta_i}\left(\dot{\bar{z}}(t), \bar{z}(t), \bar{u}(t), \bar{\theta}, t\right) := \frac{\partial \bar{g}_{\theta_i}}{\partial \bar{s}_i} = \left(\frac{\partial\left(B\dot{\bar{s}}_i\right)}{\partial \bar{s}_i} + J_z\right)$$
$$J_s^{\theta_i} \in R^{N_z\times N_z} \tag{35}$$

The derivatives $\bar{g}_{\theta_i}$ with respect to $\bar{z}$ are defined as

$$J_z^{\theta_i} := \frac{\partial \bar{g}_{\theta_i}}{\partial \bar{z}} = \frac{\partial}{\partial \bar{z}}\left[B\cdot\dot{\bar{s}}_i + J_z\cdot\bar{s}_i + \frac{\partial B}{\partial \theta_i}\dot{\bar{z}}(t) - \frac{\partial \bar{f}}{\partial \theta_i}\right]$$
$$= \left[\frac{\partial\left(B\dot{\bar{s}}_i\right)}{\partial \bar{z}} + \left[\frac{\partial J_z}{\partial z_1}\bar{s}_i \quad \cdots \quad \frac{\partial J_z}{\partial z_{N_z}}\bar{s}_i\right] + \frac{\partial}{\partial \bar{z}}\left(\frac{\partial B}{\partial \theta_i}\dot{\bar{z}}(t)\right) - \frac{\partial}{\partial \bar{z}}\left(\frac{\partial \bar{f}}{\partial \theta_i}\right)\right]$$
$$J_z^{\theta_i} \in R^{N_z\times N_z}; i=1,\ldots,N_\theta \tag{36}$$

In the same way with respect to $\bar{u}$

$$J_u^{\theta_i} := \frac{\partial \bar{g}_{\theta_i}}{\partial \bar{u}} = \frac{\partial}{\partial \bar{u}}\left[B\cdot\dot{\bar{s}}_i + J_z\cdot\bar{s}_i + \frac{\partial B}{\partial \theta_i}\dot{\bar{z}}(t) - \frac{\partial \bar{f}}{\partial \theta_i}\right]$$
$$= \left[\frac{\partial\left(B\dot{\bar{s}}_i\right)}{\partial \bar{u}} + \left[\frac{\partial J_z}{\partial u_1}\bar{s}_i \quad \cdots \quad \frac{\partial J_z}{\partial u_{N_u}}\bar{s}_i\right] + \frac{\partial}{\partial \bar{u}}\left(\frac{\partial B}{\partial \theta_i}\dot{\bar{z}}(t)\right) - \frac{\partial}{\partial \bar{u}}\left(\frac{\partial \bar{f}}{\partial \theta_i}\right)\right]$$
$$J_u^{\theta_i} \in R^{N_z\times N_u}; i=1,\ldots,N_\theta \tag{37}$$

Second-order constraint-derivatives

Based on the derivation of the second-order constraint derivatives, one can conclude that most of the derived partial Jacobian can be reused. Therefore, many of the derivative calculation steps can be reduced. The first part of the second-order constraints derivatives can be achieved by differentiating the system equations $\bar{g}$. For the k-th part the following hessian can be formulated

$$H_{g^k} := \begin{bmatrix} J_{zz}^k & 0_{N_z\times N_z\cdot N_\theta} & J_{zu}^k \\ 0_{N_z\cdot N_\theta\times N_z} & 0_{N_z\cdot N_\theta\times N_z\cdot N_\theta} & 0_{N_z\cdot N_\theta\times N_u} \\ J_{uz}^k & 0_{N_u\times N_z\cdot N_\theta} & J_{uu}^k \end{bmatrix}$$
$$H_{g^k} \in R^{(N_z+N_z\cdot N_\theta+N_u)\times(N_z+N_z\cdot N_\theta+N_u)}$$
$$k=1,\ldots,N_z \tag{38}$$

with

$$J_{zz,k} := \left(\frac{\partial^2\left(B\dot{\bar{z}}(t)\right)^k}{\partial \bar{z}^2} - \frac{\partial^2 f^k}{\partial \bar{z}^2}\right) \in R^{N_z\times N_z} \tag{39}$$

and

$$J_{zu,k} := \left(\frac{\partial^2\left(B\dot{\bar{z}}(t)\right)^k}{\partial \bar{z}\partial \bar{u}} - \frac{\partial^2 f^k}{\partial \bar{z}\partial \bar{u}}\right) \in R^{N_z\times N_u}$$
$$J_{uz,k} = J_{zu,k} \tag{40}$$

where the k-th row of a matrix or the k-th entry of a column is shortly written as $(\cdot)^k$. The partial hessian $J_{zz,k}$ can be assembled from the k-th rows of the terms $\partial J_z/\partial z_1, \ldots, \partial J_z/\partial z_{N_z}$ in Eq. 36. The same counts for the partial hessian $J_{zu,k}$, regarding the terms $\partial J_z/\partial u_1, \ldots, \partial J_z/\partial u_{N_z}$ in Eq. 37. Only $\partial J_{uu,k}$ has to be additionally calculated. Furthermore, the structure of H_{g^k} is extremely sparse regarding its dimension, which can in fact be favorably integrated to advanced optimization routines. In analogy to obtaining the hessians for the k-th parts of $\bar{g}$, one also obtains for the k-th part of each partial sensitivity equations $\bar{g}_{\theta_i}$

$$H_{g_{\theta_i}^k} := \begin{bmatrix} J_{zz,k}^{\theta_i} & 0_{N_z\times(i-1)\cdot N_z} & J_{zs_i,k}^{\theta_i} & 0_{N_z\times(N_\theta-i+1)\cdot N_z} & J_{zu,k}^{\theta_i} \\ 0_{(i-1)\cdot N_z\times .N_z} & \vdots & \cdots & \vdots & 0_{(i-1)\cdot N_z\times .N_u} \\ J_{s_iz,k}^{\theta_i} & \vdots & 0_{N_z\times .N_\theta} & \vdots & J_{s_iu,k}^{\theta_i} \\ 0_{(N_\theta-i+1)\cdot N_z\times .N_z} & \vdots & \cdots & \vdots & 0_{(N_\theta-i+1)\cdot N_z\times .N_u} \\ J_{uz,k}^{\theta_i} & 0_{N_u\times(i-1)\cdot N_z} & J_{us_i,k}^{\theta_i} & 0_{N_u\times(N_\theta-i+1)\cdot N_z} & J_{uu,k}^{\theta_i} \end{bmatrix} \tag{41}$$

$$H_{k_{g_{\theta_i}}} \in R^{(N_z+N_z.N_\theta+N_u)\times(N_z+N_z.N_\theta+N_u)}$$

$$k=1,\ldots,N_z, i=1,\ldots,N_\theta$$

The structure of $H_{k_{g_{\theta_i}}}$ is also extremely sparse with its partial hessian regarding the states

$$J_{zz,k}^{\theta_i} := \frac{\partial\left(J_z^{\theta_i}\right)^k}{\partial\bar{z}} = \frac{\partial^2(J_z\cdot\bar{s}_i)^k}{\partial\bar{z}^2} + \frac{\partial^2}{\partial\bar{z}^2}\left(\frac{\partial B}{\partial\theta_i}\dot{\bar{z}}(t)\right)^k - \frac{\partial^2}{\partial\bar{z}^2}\left(\frac{\partial f^k}{\partial\theta_i}\right)$$
$$J_{zz,k}^{\theta_i} \in R^{N_z\times N_z} \tag{42}$$

and the partial hessian regarding states and sensitivities

$$J_{sz,k}^{\theta_i} := \frac{\partial\left(J_s^{\theta_i}\right)^k}{\partial\bar{z}} = \frac{\partial}{\partial\bar{z}}\left(\frac{\partial\left(B\dot{\bar{s}}_i\right)}{\partial\bar{s}_i}\right)^k + J_{zz,k}$$
$$J_{sz,k}^{\theta_i} \in R^{N_z\times N_s}, J_{zs,k}^{\theta_i} = J_{sz,k}^{\theta_i} \tag{43}$$

where the term $J_{zz,k}$ can be carried over directly from Eq. 39. It is analogous to the partial hessian regarding controls and sensitivities

$$J_{su,k}^{\theta_i} := \frac{\partial\left(J_s^{\theta_i}\right)^k}{\partial\bar{u}} = \frac{\partial}{\partial\bar{u}}\left(\frac{\partial\left(B\dot{\bar{s}}_i\right)}{\partial\bar{s}_i}\right)^k + J_{zu,k}$$
$$J_{su,k}^{\theta_i} \in R^{N_z\times N_u}, J_{us,k}^{\theta_i} = J_{su,k}^{\theta_i} \tag{44}$$

where the term $J_{zu,k}$ can be carried over from Eq. 40. Furthermore, the partial hessian regarding controls is now defined as

$$J_{uu,k}^{\theta_i} := \frac{\partial\left(J_s^{\theta_i}\right)^k}{\partial\bar{u}} = \frac{\partial^2\left(B\cdot\dot{\bar{s}}_i\right)}{\partial\bar{u}^2} + \frac{\partial^2\left(J_z\cdot\dot{\bar{s}}_i\right)}{\partial\bar{u}^2} + \frac{\partial^2}{\partial\bar{u}^2}\left(\frac{\partial B}{\partial\theta_i}\dot{\bar{z}}(t)\right)^k$$
$$- \frac{\partial^2}{\partial\bar{u}^2}\left(\frac{\partial f^k}{\partial\theta_i}\right) J_{uu,k}^{\theta_i} \in R^{N_u\times N_u} \tag{45}$$

and with respect to the controls and states

$$J_{uz,k}^{\theta_i} := \frac{\partial\left(J_u^{\theta_i}\right)^k}{\partial\bar{z}}$$
$$= \frac{\partial^2\left(B\cdot\dot{\bar{s}}_i\right)^k}{\partial\bar{u}\partial\bar{z}} + \frac{\partial^2\left(J_z\cdot\dot{\bar{s}}_i\right)^k}{\partial\bar{u}\partial\bar{z}} + \frac{\partial^2}{\partial\bar{u}\partial\bar{z}}\left(\frac{\partial B}{\partial\theta_i}\dot{\bar{z}}(t)\right)^k$$
$$- \frac{\partial^2}{\partial\bar{u}\partial\bar{z}}\left(\frac{\partial f^k}{\partial\theta_i}\right)$$
$$J_{uz,k}^{\theta_i} \in R^{N_u\times N_z}, J_{zu,k}^{\theta_i} = J_{uz,k}^{\theta_i} \tag{46}$$

Direct derivative calculation of the objective function

The derivatives of the optimal criteria can be calculated directly by a complete discretization of the model-based experimental design into a NLP-problem, which then enables an efficient solution of the problem. Thus, the optimization variables of OP2 are all discretized system variables $\left[\bar{z}_{lj}, \bar{s}_{1,lj}, \ldots, \bar{s}_{N_\theta,lj}, \bar{u}_{lj}, \bar{\omega}, \bar{z}_0, \bar{h}\right]$. However, since the optimal criterion only depends on the sensitivities at the sampling time points $\left[\bar{s}_{1,lj}, \ldots, \bar{s}_{N_\theta,lj}\right]_{t_{lj}\equiv t_{sp}}$, and the weighting controls $\bar{\omega}$, its derivative becomes zero for all $\left[\bar{z}_{lj}, \bar{u}_{lj}, \bar{z}_0, \bar{h}\right]$ and $\left[\bar{s}_{1,lj}, \ldots, \bar{s}_{N_\theta,lj}\right]_{t_{lj}\neq t_{sp}}$. For the following derivation, the dependent variables are summarized to a variable vector X with

$$X := \left[\left[\bar{s}_{1,lj}, \ldots, \bar{s}_{N_\theta,lj}\right]_{t_{lj}\equiv t_{sp}}, \bar{\omega}\right] \tag{47}$$

The derivatives of the optimal criteria are developed in the following way via the chain rule

$$\frac{\partial\Phi\left(M_F\left(S_{sp}\right)\right)}{\partial(X)} = \frac{\partial\Phi(M_F)}{\partial(M_F)}\cdot\frac{\partial M_F\left(S_{sp}\right)}{\partial(X)} \tag{48}$$

The derivatives of the A, D, and E-criteria $\in R^{1\times N_\theta.N_\theta}$ only differ from each other in the first term of Eq. 48

$$\frac{\partial\Phi_A(M_F)}{\partial M_F} = \frac{\partial(tr(M_F))}{\partial M_F} = (vec(I_{N_\theta}))^T \tag{49}$$

$$\frac{\partial\Phi_D(M_F)}{\partial M_F} = \frac{\partial(\det(M_F))}{\partial M_F} = \det(M_F).\left(vec\left(M_F^{-T}\right)\right)^T \tag{50}$$

$$\frac{\partial\Phi_E(M_F)}{\partial M_F} = \frac{\partial(\lambda_{\min}(M_F))}{\partial M_F}$$
$$= \left(v_{\lambda_{\min}}{}^T.v_{\lambda_{\min}}\right)^{-1}.v_{\lambda_{\min}}{}^T.\left(I_{N_\theta}\otimes v_{\lambda_{\min}}{}^T\right) \tag{51}$$

Here $v_{\lambda_{\min}}$ denotes the eigenvector corresponding to the smallest eigenvalue of M_F. For Eq. 51, the following matrix calculus rule is used. The derivative of the determinant of the matrix $A\in R^{n\times n}$ with respect to its elements is given as

$$\frac{d|A|}{dA} = \frac{d|A|}{d(vec(A))} = |A|\left(^{vec(A^{-T})}\right)T \in R^{1\times n\cdot n}A^{-T} = \left(^{A^{-1}}\right)T \tag{52}$$

The second term of Eq. 48 represents the Jacobian of M_F with respect to the elements of X, which is the same for all criteria

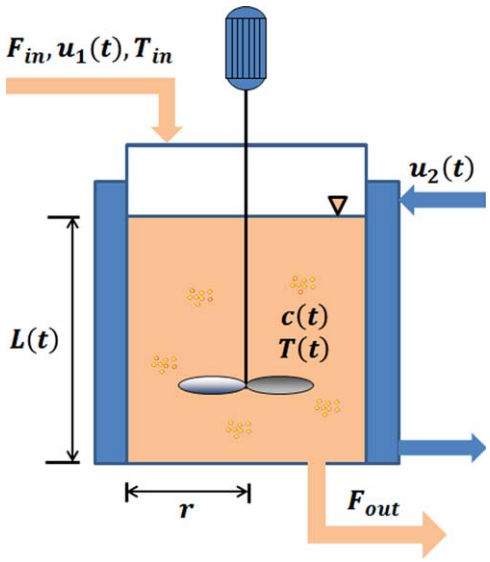

Figure 1. Semibatch bioreactor.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com]

$$\frac{\partial(M_F)}{\partial(X)} = \frac{\partial(vec(M_F))}{\partial(X)} \in R^{N_\theta \cdot N_\theta \times N_z \cdot N_\theta \cdot N_{sp}} \tag{53}$$

In order to obtain the second-order derivatives analytically, it is required to use Kronecker and box matrix products.[23] To avoid the more complicated derivation, symbolic differentiation was used to calculate the second derivatives. For this purpose, the results from Eq. 48 are directly differentiated one more time with respect to X.

An Illustrative Example

The efficiency of the proposed approach to model-based experimental design is demonstrated by using an illustrative example of a bioreactor[24,25] referring to as optimal experimental design (Figure 1). Since this model has been well studied in the past, it is quite suitable to highlight the results of the presented simultaneous approach in terms of comparability and theoretical issues. First, a comparison of the optimization with equidistant sampling and optimal sampling is shown. Then, the adaptive optimal sampling strategy is presented which combines the strength of the optimal sampling design and the variation of the collocation element lengths.

Reactor model

The reactor model consists of two differential equations, which describe a biomass population and its consumption of substrate. The system state variables are the biomass concentration and the substrate concentration c_s. The reactor is fed by a substrate stream which provides two control possibilities. The first one is addressed as the dilution factor u_1, and the second one as the substrate concentration of the feed stream u_2. It is assumed that there are four unknown model parameters $\theta_{1,\ldots,4}$, which all have an initial guess of 0.5

Table 1. Initial Design Settings

States		controls	
$c_B(t=0)$	$2\frac{g}{l}$	$u_1(t)=const.$	$0.1\mathrm{h}^{-1}$
$c_s(t=0)$	$0.1\frac{g}{l}$	$u_2(t)=const.$	$15\frac{g}{l}$
Sampling vector [h]	3, 6, 9, 12		

Table 2. Collocation Settings

Element length [h]	$h=const.=1$
collocation order	$K=3$
number of elements	$N_e=12$

$$\begin{aligned}\frac{dc_B}{dt} &= \dot{r} - (u_1+\theta_4)c_B \\ \frac{dc_S}{dt} &= -\frac{\dot{r}}{\theta_3} + (u_2 - c_S)u_1 \\ \dot{r} &= \theta_1 \frac{c_B c_S}{\theta_2 + c_S}\end{aligned} \tag{54}$$

The aim of the experimental design is to increase the information content by running experiments appropriately, such that a best possible identification of the parameters $\theta_{1,\ldots,4}$ can be achieved afterward. The experimental design criteria as functions of the Fisher information matrix or rather of parameter sensitivities have been presented in Eq. 22. In the remainder, the A-, D- and E-criterion are used for the optimization of the presented system. Moreover, technical limitations of the system have to be adhered to which are the following constraints on state and control variables

$$\begin{aligned}1\frac{g}{l} \le c_B \le 25\frac{g}{l}, 0.1\frac{g}{l} \le c_S \le 25\frac{g}{l} \\ 0.05\frac{1}{h} \le u_1 \le 5\frac{1}{h}, 0.2\frac{g}{l} \le u_2 \le 35\frac{g}{l}\end{aligned} \tag{55}$$

The lower bounds on the state variables guarantee a minimum population of the biomass and a minimum substrate amount so as to feed the biomass. The bounds on the control variables are inherited from the reference literature.[25] For the initial design, the controls are chosen to be constant and the initial conditions of the state variables are fixed as shown in Table 1.

The design end time is fixed to 12 h. Therefore, with a uniform collocation element length of 1 h the element number results in 12. Furthermore, it should be emphasized that

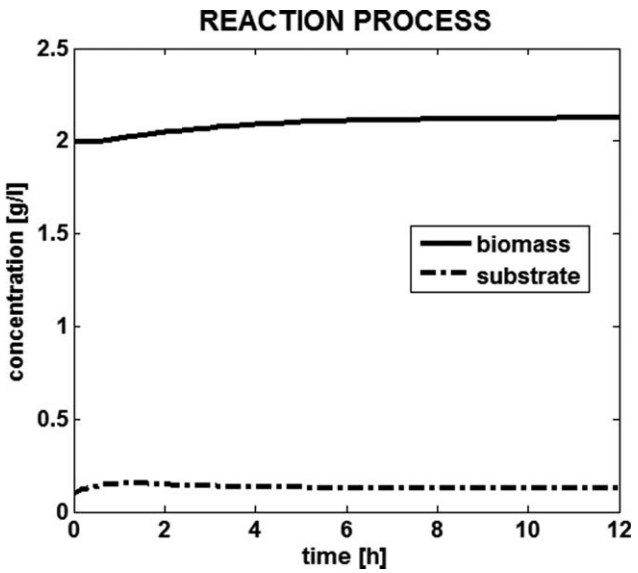

Figure 2. Simulation results with initial design settings.

Table 3. Collocation Settings

Element length [h]	$h=const.=1$
Collocation order	
For states and sensitivities	$K=3$
For controls	$K_{u1}=1, K_{u2}=0$
Number of elements	$N_e=12$
Number of samplings	$N_{sp,U}=4$
possible sampling times [h]	$1,2,3,4,5,6,7,8,9,10,11,12$

because of convenience and due to the characteristic of the OCFEM the sampling times are always taken at the end of a collocation element. The collocation settings are summarized in Table 2.

See Figure 2 for the given initial design settings, along with process results in the concentration profiles.

The resulting nominal values for the A-, D- and E-criterion are 8.39, 1.55×10^{-8} and, respectively, 7.29×10^{-10}, which are relatively concise regarding the information content of the system and far away from an optimum. Because of the constant levels of control inputs, this design is not able to address the dynamic behavior of the system at all. Thus, to overcome the disadvantages of the statistical design in case of dynamic processes the strategy of model-based or also known as optimal experimental design is applied to the system equations in the next section.

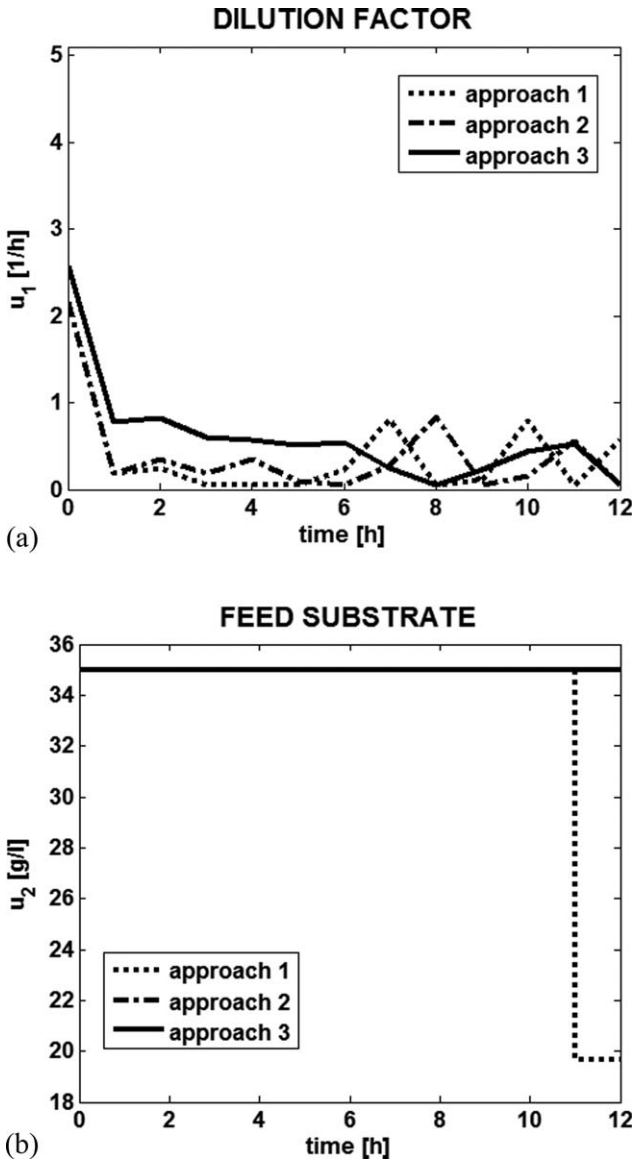

Figure 3. A-criterion— optimized control trajectories.

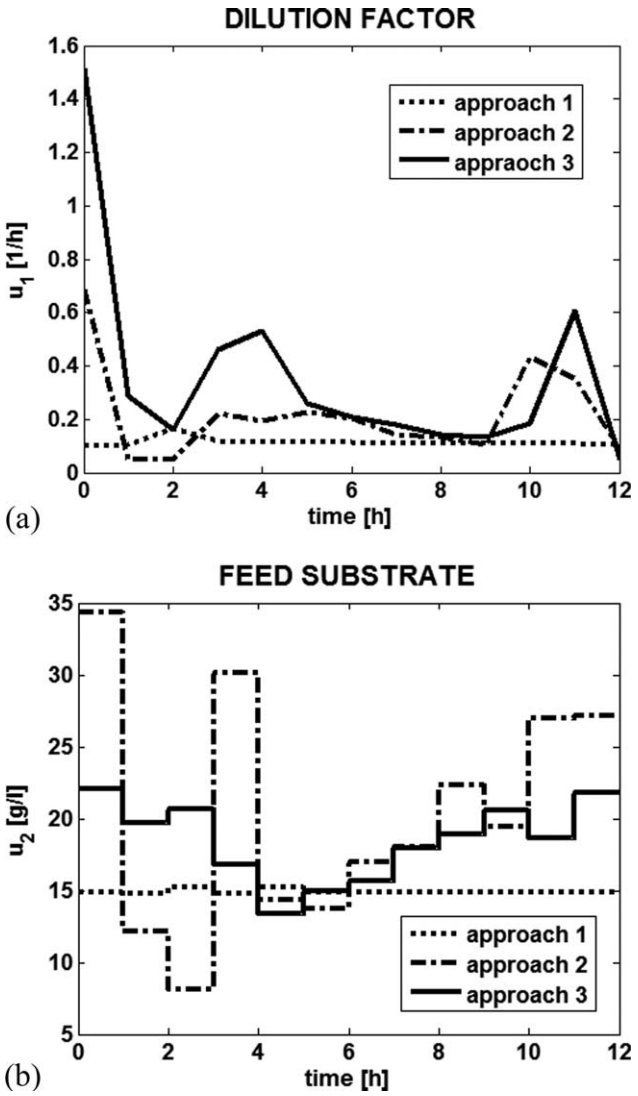

Figure 4. D-criterion—optimized control trajectories.

Optimal sampling strategy

Three approaches are presented in this section to discuss the advantages of the optimal sampling strategy and the results of its implementation to the simultaneous optimization approach. Approach 1 represents a standard optimal experimental design setting, where the sampling intervals are chosen to be equidistant. Approach 2 includes the optimal sampling strategy, where the optimization result converges to the best sampling points of the given possible set. For the sake of clarity regarding the optimal sampling strategy, the optimization of the initial sates is not taken into account in approach 1 and 2. Instead, approach 3 provides the same settings as approach 2, but additionally with the optimization of the initial states including the following constraints

$$1\frac{g}{\mathrm{l}} < c_{B,0} < 25\frac{g}{\mathrm{l}}, 0.1\frac{g}{\mathrm{l}} < c_{S,0} < 25\frac{g}{\mathrm{l}} \tag{56}$$

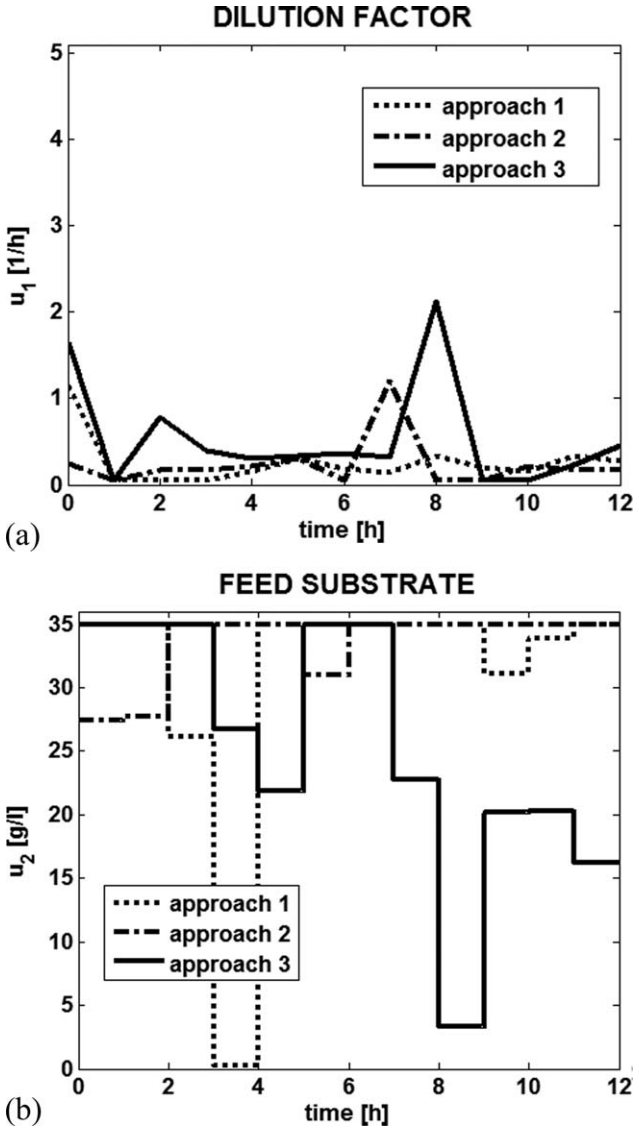

Figure 5. E-criterion—optimized control trajectories.

The collocation settings in this section are similar to those of Table 2 with the additional variation of the control types. The dilution factor, which is addressed as u_1, is set as first-order type (ramps), and the feed substrate u_2 as zero-order type (steps). Furthermore, the amount of samplings $N_{sp,U}$ is set to 4 (Table 3).

All three approaches have been optimized with respect to the A-, D- and E-criterion. The resulting optimized dynamic control trajectories are shown in Figures 3, 4, and 5, respectively. The corresponding optimized values are listed in Tables 4, 5 and 6.

A comparison between approach 2 and 1 clearly shows that the optimal sampling strategy is superior to the standard strategy. The nominal values of all three criteria are considerably improved, and the dynamic characteristic of the system are much better exploited, which are reflected in the optimized control trajectories plots (Figures 3–5), and the corresponding optimized concentration profiles plots (Appendix B, Figures B1 to B3). A more detailed look over the sampling time vector shows that an optimization of the sampling decisions should always be implemented since in general optimal samplings strongly differ from an equidistant one. Moreover, its implementation in the simultaneous optimization approach has been shown as straightforward (OP2), and the convergence behavior of the sampling weighting factors is reliable.

Approach 3 gives the best results of this section regarding all three criteria. In comparison to the settings of approach 2, the initial states are additionally handled as optimization variables leading to an enlarged optimization space. Therefore, in general a better optimum can be found. More important is the fact that this also has an effect to the optimized sampling vector, which can be seen from the optimization results with the A- and especially with the E-criterion. The optimized sampling vector completely changed in the latter case.

The drawback of the three approaches so far is the fixed collocation element length, which constitutes a big disadvantage since the samplings are thereby only allowed at the collocation points (generally at the end points). The consequence is a restricted optimization space. However, there is indeed no need to fix the collocation element length in a large-scale optimization approach, where the problem formulation is fully discretized. This idea leads to the following variable element length approach.

Adaptive optimal sampling strategy

Approach 3 is now enhanced to handle the element length as optimization variables. The settings of Table 3 are extended with the following constraints

$$\begin{aligned} &1\text{h} \le h_l \le 2\text{h}, l=1,\ldots,N_e \\ &\sum h_l = t_f, t_f \le 12 \end{aligned} \tag{57}$$

Furthermore, the experimental end time of the design is limited to 12 h. The total number of element N_e has been varied from 4 to 12 collocation elements. Here, the minimum number results from the limited amount of samplings $N_{sp,U}$ which has been set to 4, and the maximum number results from the end time restriction and the minimum element length of 1 h. Moreover, the latter case with 12 collocation elements represents exactly the results of approach 3 in the previous section.

Table 4. A-Criterion

	Approach 1	Approach 2	Approach 3
Optimized value [-]	1.60E+04	5.29E+04	6.09E+04
Initial state vector [g/l]	2, 0.1	2, 0.1	22.96, 25.0
Sampling vector [h]	3, 6, 9, 12	7, 10, 11, 12	9, 10, 11, 12

Table 5. D-Criterion

	Approach 1	Approach 2	Approach 3
Optimized value [-]	1.04E-03	3.06E+04	1.36E+05
Initial state vector [g/l]	2, 0.1	2, 0.1	7.66, 5.97
Sampling vector [h]	3, 6, 9, 12	2, 3, 4, 12	2, 3, 4, 12

Table 6. E-Criterion

	Approach 1	Approach 2	Approach 3
Optimized value [-]	2.10	2.70	3.09
Initial state vector [g/l]	2, 0.1	2, 0.1	25.0, 25.0
Sampling vector [h]	3, 6, 9, 12	4, 10, 11, 12	1, 2, 3, 7

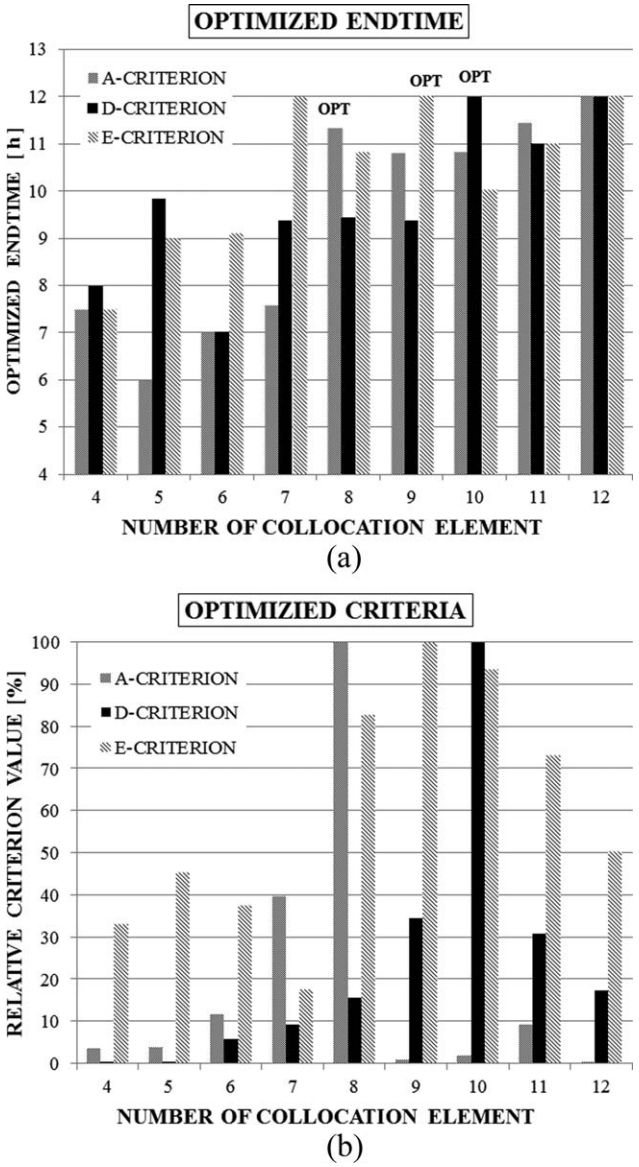

Figure 6. (a) Optimized criteria. (b) Optimized endtime.

The variation of the element number has been carried out with respect to the optimization with the A-, D- and E-criterion. The resulting optimized criteria values and the corresponding optimized end times are shown in Figure 6a and 6b (see also Table C1 and C2). Table 7 shows the criterion with the corresponding best cases in more detail.

In comparison with the results of approach 3, all three criterions have been further improved. Especially the different element lengths and the corresponding sampling vectors of the adaptive optimal sampling strategy are noticeable.

The results reveal that replacing a fixed structure of the sampling intervals by an adaptive one promises the biggest improvement with respect to the standard case. Therefore, the latter approach is more favorable and clearly the most recommended one. Nevertheless, there are two open issues which are worth discussing in the following for future investigations.

A proper handling of the discretization error should be included to the optimization formulation. It is obvious that the upper bound of the collocation element length h_l cannot be as large as possible. Accordingly, it also cannot be answer just by heuristics, whether the lower bounds should be at least as small as the minimum possible sampling interval or smaller. This issue has to be rather investigated in connection with the approximation of the global discretization error $e(t)=z(t)-z^K(t)$, which holds for the OCFEM with Legendre roots[26]

$$\max_{t\in[0,t_f]}||e(t)|| \le C \max_{l\in[1,\ldots,N_e]}\left(h_l^k||T_l(t)||\right)+O\left(h_l^{K+1}\right) \tag{58}$$

Here C denotes a constant which is more of mathematical interest, and $T_l(t)$ depends only on the solution $z(t)$ and independent of the choice of element length $h_l, l=1,\ldots,N_e$. A formulation which allows variable collocation element lengths in the context of tracking and adapting to steep profiles has been proposed in literature.[27] A theory to extend OP2 regarding an adaptive sampling strategy could be derived based from this approach

$$\begin{aligned}&\sum h_l=t_f, h_l \ge 0, l=1,\ldots,N_e\\ &\tilde{C}||T_l(\tau)|| \le \varepsilon\end{aligned} \tag{59}$$

Second, as seen in Figure 6a and especially in 6b, the optimization results can be very nonlinear with respect to the number of collocation elements. This indirectly indicate a strongly tendency to local minima so that gradient-based optimization algorithm generally cannot guarantee finding a global optimum.[28] Nevertheless, regarding global optimum algorithm, the simultaneous approach is compare to sequential approaches more favorable because of its fully discretized formulation and the fact that the model nonlinearities are not increasing from the discretization formula. In particular global optimum strategy like the convex envelops approach[29] is very promising to be applied to the simultaneous solution approach to model-based experimental design.

Conclusions

The analytical investigation of the simultaneous solution approach to model-based experimental design reveals diverse advantages regarding the bottlenecks of the sequential approach and multiple shooting approaches. In particular, the integration of the mix-derivatives with respect to parameter sensitivities and

Table 7. Adaptive Sampling Strategy

	A-CRITERION	D-CRITERION	E-CRITERION
N_θ of the best case	8	10	9
Criterion value [-]	1.04E+07	7.85E+05	6.12
Initial state vector $\begin{bmatrix} g \\ 1 \end{bmatrix}$	6.11, 8.95	1.11, 0.83	10.23, 8.59
Sampling vector [h]	3.59, 5.38, 9.18, 10.18	3.24, 4.32, 5.40, 12.0	1.42, 2.83, 5.67, 12.0
Collocation elementvector [h]	1.79,1.79,1.79,1.79, 1.0, 1.0, 1.0, 1.14	1.08, 1.08, 1.08, 1.08, 1.08, 1.08, 1.47, 1.47, 1.57, 1.0	1.41, 1.42, 1.42, 1.07, 1.42, 1.42, 1.42, 1.42, 1.0

control variables in addition to the integration of the model equations can totally be saved. These derivatives are instead represented in the simultaneous formulation as constraints derivatives. Here they can directly be calculated and their sparse-structure can fully be exploited for an efficient implementation. Furthermore, the derivation of the objective function derivatives shows that they can be separated into two parts. The first part can directly be given as analytic expressions with respect to the A, D- and E-criterion. The second part has the same structure for all three criteria and can be a priori calculated by symbolic derivation allowing an efficient call during the optimization iterations. Another important aspect represents the process constraints on state as well as control variables, which are directly given on every collocation point. Thus, the process stability is handled over the entire domain rather than just at the end points of each interval. The most noticeable aspect is the sampling strategy. For this purpose, the adaptive optimal sampling strategy has been proposed, which shows the best results among the investigated approaches. Further suggested investigations on the proposed simultaneous solution approach to model-based experimental design are first, the extension of the adaptive optimal sampling strategy with a proper error minimization approach, and second, the application of global optimization strategies.

Acknowledgments

This work is part of the Collaborative Research Center InPROMPT SFB/Transregio63 'Integrated Chemical Processes in Liquid Multiphase Systems' and coordinated by the Berlin Institute of Technology. The authors thank the German Research Foundation (DFG Deutsche Forschungsgesellschaft) for the financial support.

Notation

z = state variables
u = control variable
B = mass matrix
M_F = Fisher information matrix
N_θ = number of parameter
N_{sp} = number of measuring points
N_e = number of elements (OCFEM)
Δh = length of an element
j = collocation point index
s = sensitivity of states regarding the parameters
θ = process parameter
N_{sp} = dynamic sensitivity matrix
N_z = number of state variables
N_u = number of controls
K = order of Lagrange polynomials for states
q = order of Lagrange polynomials for controls
i = parameter index
l = element index

Greek letters

Φ = optimal criteria
φ,&phis; = Lagrange basis polynomials collocation time
Σ_z = measurement covariance matrix
τ = normalized
FEM = finite element method
OCFEM = orthogonal collocation on FEM

Literature Cited

1. Schöneberger JC, Arellano-Garcia H, Wozny G, Körkel S, Thielert H. Model-based experimental analysis of a fixed-bed reactor for catalytic SO2 Oxidation. *Ind Eng Chem Res.* 2009;48: 5165–5176.
2. Sidoli FR, Mantalaris A, Macchietto S. Modelling of mammalian cells and cell culture processes. *Cytotechnol.* 2004;44: 27–46.
3. Galvanin F, Boschiero A, Barolo M, Bezzo F. Model-based design of experiments in the presence of continuous measurement systems. 9th International Symposium on Dynamics and Control of Process Systems: Leuven, Belgium; 5–7 July, 2010.
4. Hunter WG, Steinberg DM. Experimental design: review and comment. *Technometrics.* 1984;26:71–97.
5. Mehra KR. Optimal input signals for parameter estimation in dynamic systems: survey and new results. *IEEE Trans Automat Contr.* 1974;19:753–768.
6. Espie D, Macchietto S. The optimal design of dynamic experiments. *AIChE J.* 1989;35:223–229.
7. Franceschini G, Macchietto S. Model-based design of experiments for parameter precision: state of the art. *Chem Eng Sci.* 2008;63: 4846–4872.
8. Körkel S, Bauer I, Bock HG, Schlöder JP. A sequential approach for nonlinear optimum experimental design in DAE systems. *Sci Comput Chem Eng.* 1999;2:338–345.
9. Arellano-Garcia H, Schöneberger J, Körkel S, Optimale Versuchsplanung in der chemischen Verfahrenstechnik. *Chem Ing Tech.* 2007; 79:1625–1638.
10. Bock HG, Pitt KJ. A multiple shooting algorithm for direct solution of optimal control problems. Presented at: Ninth IFAC World Congress, 1984; Budapest, Hungry.
11. Bauer I, Bock HG, Körkel S, Schlöder JP. Numerical methods for optimum experimental design in DAE systems. *J Comput Appl Math.* 2000;120:1–25.
12. Feehery W, Tolsma J, Barton P. Efficient sensitivity analysis of large-scale differential-algebraic system. *App Numer Math.* 1997;25: 41–54.
13. Shengtai Li, Petzold L. Design of New Daspk for Sensitivity Analysis. Santa Barbara, CA: University of California: Technical Report; 1999.
14. Barz T, Kuntsche S, Wozny G, Arellano-Garcia H. An efficient sparse approach to sensitivity generation for large-scale dynamic optimization. *Comput Chem Eng.* 2011;35:2053–2065.
15. England R, Lamour R. Multiple shooting using a dichotomically stable integrator for solving differential-algebraic equations. *Appl Numer Math.* 2002;42:117–131.
16. Biegler LT. Nonlinear Programming: MOS-SIAM Series on Optimization. Philadelphia, PA: SIAM; 2010.
17. Biegler LT, Grossmann IE. Retrospective on optimization. *Comput Chem Eng.* 2004;28:1169–1192.
18. Heidebrecht P, Sundkacher K, Biegler LT. Optimal design of nonlinear temperature programmed reduction experiments. *AIChE J.* 2011: 57: 2888–2901.
19. Sager S. Sampling decisions in optimum experimental design in the light of pontryagin's maximum principle. *SIAM J Control Optim.* 2012. In press.
20. Bard Y. Nonlinear Parameter Estimation. New York: Academic Press; 1974.
21. Karneswaran S, Biegler LT. Convergence rates for direct transcription of optimal control problems with final-time equality constraints using collocation at Radau points. In: Proceedings from the American Control Conference; June 14–16, 2006; Minneapolis, MN.
22. Cuthrell JE, Biegler LT. On the optimization of differential-algebraic process systems. *Comput Chem Eng.* 1987;13:49–62.
23. Olsen A. Peder, Rennie J. Steven, Goel V. Recent advances in algorithmic differentiation: lecture notes. *Comput Sci Eng.* 2012;87: 71–81.
24. Espie D, Macchietto S. The optimal design of dynamic experiments. *AIChE J.* 1989;35: 223–229.
25. Asprey SP, Macchietto S. Designing robust optimal dynamic experiments. *J Process Control.* 2002;12:545–556.
26. Russell RD, Christiansen J. Adaptive mesh selection strategies for solving boundary value problems. *SIAM J Numer Anal.* 1978;15: 59–80.
27. Vasantharajan S, Biegler LT. Simultaneous strategies for parameter optimization and accurate solution of differential-algebraic systems. *Comput Chem Eng.* 1990;14:1083.
28. Schöneberger J, Arellano-Garcia H, Wozny G. Local optima in model-based optimal experimental design. *Ind Eng Chem Res.* 2010; 49(20):10059–10073.

29. Jach M, Michaels D, Weismantel R. Convex envelopes for (n-1)-convex functions, *SIAM J Optim*. 2008;19:1451–1466.

Appendix A: *Discretization of an ODE-system by using the OCFEM-method*

Consider following linear implicit DAE-system

$$B\left(\bar{z}(t),\bar{u}(t),\bar{\theta},t\right)\cdot\dot{\bar{z}}(t)=\bar{f}\left(\bar{z}(t),\bar{u}(t),\bar{\theta},t\right)),\bar{z}(t=0)=\bar{z}_0 \quad \text{(A1)}$$

The pure algebraic equations of the DAE-system are taken into account with a singular mass matrix B, and, thus, are not written explicitly.

Time-dependent functions of the state and control variables are discretized by the orthogonal collocation on finite element method (OCFEM) with N_e-finite elements (see Figure B1).

The functional section in each element is approximated by polynomials of order K. Within each element the overall discrete time is projected to the local normalized collocation time τ_j

$$\begin{aligned} &t_{lj}=h_{l-1}+\tau_j(h_l-h_{l-1})=h_{l-1}+\tau_j\Delta h_l,\tau_j\in[0,1]\\ &l=1,\ldots,N_e,j=0,\ldots,K,h_0=t_0 \end{aligned} \quad \text{(A2)}$$

Lagrange-polynomials are chosen as base functions. The advantage of the Lagrange polynomials is that the coefficients directly represent the desired physical and chemical values, what simplifies the interpretation of the results.

The function of the state variables $\bar{z}_l$ in element is represented by the Lagrange-polynomials φ_{lj}

$$\begin{aligned} \bar{z}_l(t)&=\sum_{j=0}^{K}\bar{z}_{lj}\varphi_{lj}(t)\\ \varphi_{lj}(t)&=\prod_{\substack{k=0\\k\neq j}}^{K}\frac{(t-t_{lk})}{(t_{lj}-t_{lk})} \end{aligned} \quad \text{(A3)}$$

The controls are discretized in the same way but using polynomial approximations of a possibly different order q

$$\begin{aligned} \bar{u}_l(t)&=\sum_{m=0}^{q}\bar{u}_{lm}\psi_{lm}(t)\\ \psi_{lm}(t)&=\prod_{n=0}^{q}\frac{(t-t_{\ln})}{(t_{lm}-t_{\ln})} \end{aligned} \quad \text{(A4)}$$

Since the discretized functions are only time dependent in the base functions the time derivatives of state variables can be written as

$$\dot{\bar{z}}_l(t)=\sum_{j=0}^{K}\bar{z}_{lj}\dot{\varphi}_{lj}(t) \quad \text{(A5)}$$

By applying the chain rule for every $\dot{\varphi}_{lj}(t_{lj})$ on Eq. A2, the following simplification can be obtained

$$\begin{aligned} &\dot{\varphi}_{lj}(t)=\frac{d\varphi_{lj}(t)}{dt}=\frac{\partial\varphi_{lj}(\tau)}{\partial\tau}\frac{\partial\tau}{\partial t}=\frac{d\varphi_{lj}(t)}{\partial\tau}\frac{1}{\Delta h_l}\\ &l=1,\ldots,N_e;\ j=0,\ldots,K \end{aligned} \quad \text{(A6)}$$

Since the same collocation method and the same polynomial order are used in each element

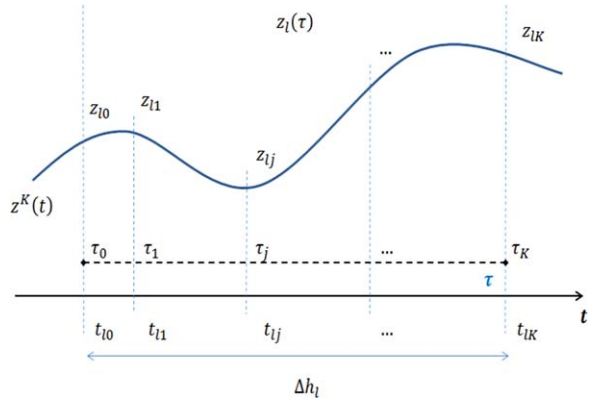

Figure A1. Element *l* with collocation points.

$$\varphi_{lj}(t)=\varphi_j(\tau)=\prod_{\substack{k=0\\k\neq 0}}^{K}\frac{(\tau-\tau_k)}{(\tau_j-\tau_k)} \quad \text{(A7)}$$

one obtains

$$\frac{\partial\varphi_{lj}(\tau)}{\partial\tau}=\frac{\partial\varphi_j(\tau)}{\partial\tau} \quad \text{(A8)}$$

Thus, Eq. A5 can be rewritten as

$$\begin{aligned} &\dot{\bar{z}}_l(t_{lj})=\sum_{j=0}^{K}\bar{z}_{lj}\frac{1}{\Delta h_l}\frac{\partial\varphi_j(\tau_j)}{\partial\tau}\\ &l=1,\ldots,N_e;j=1,\ldots,K \end{aligned} \quad \text{(A9)}$$

Furthermore, continuity conditions ensure that no discontinuities arise between the elements of the discretized functions. They are generally formulated for the OCFEM as

$$\tilde{z}_{l+1,j=0}=\sum_{j=1}^{K}\tilde{z}_{lj}\varphi_j(\tau=1),l=1,\ldots,N_e-1 \quad \text{(A10)}$$

where $\tilde{z}$ represents the system state variables, which only belong to the differential equations of A1. The meaning of Eq. A10 is that the polynomial function of $\tilde{z}_l(t)$ the l-th element is extrapolated to its end point and specifies thereby the initial value for the following $(l+1)$-th element. For the Radau-collocation method,[16] the continuity conditions reduce to

$$\begin{aligned} &\tilde{z}_{l+1,j=0}=\tilde{z}_{l,j=k}\\ &l=1,\ldots,N_e-1 \end{aligned} \quad \text{(A11)}$$

Applying the OCFEM method on Eq. A1, one gets

$$\begin{aligned} &\sum_{j=0}^{K}B\left(\bar{z}_{lj},\bar{u}_{lj},\bar{\theta},\tau_j\right)\bar{z}_{lj}\frac{\dot{\varphi}_j(\tau_j)}{\Delta h_l}=\bar{f}\left(\bar{z}_{lj},\bar{u}_{lj},\bar{\theta},\tau_j\right)\\ &l=1,\ldots,N_e;j=1,\ldots,K\\ &\tilde{z}_{11}=\tilde{z}_0,\\ &\tilde{z}_{l+1j=0}=\tilde{z}_{lj=K},l=1,\ldots,N_e-1 \end{aligned} \quad \text{(A12)}$$

In order to solve Eq. A12, $\dot{\varphi}_j(\tau_k)$ is previously calculated once offline because it only depends on the known selected collocation points τ_j.

Appendix B:

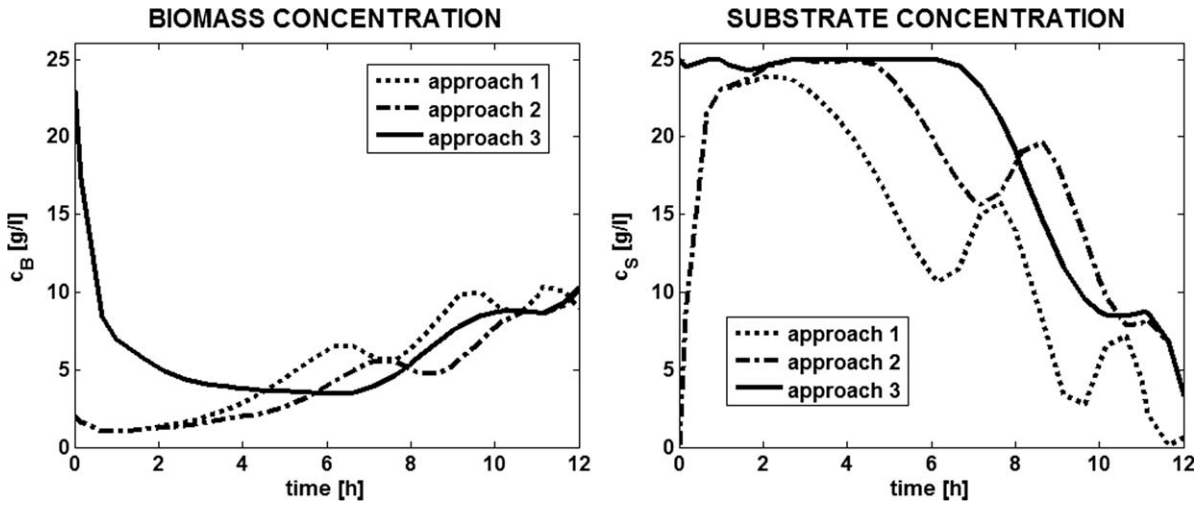

Figure B1. Optimization results with A-criterion—states.

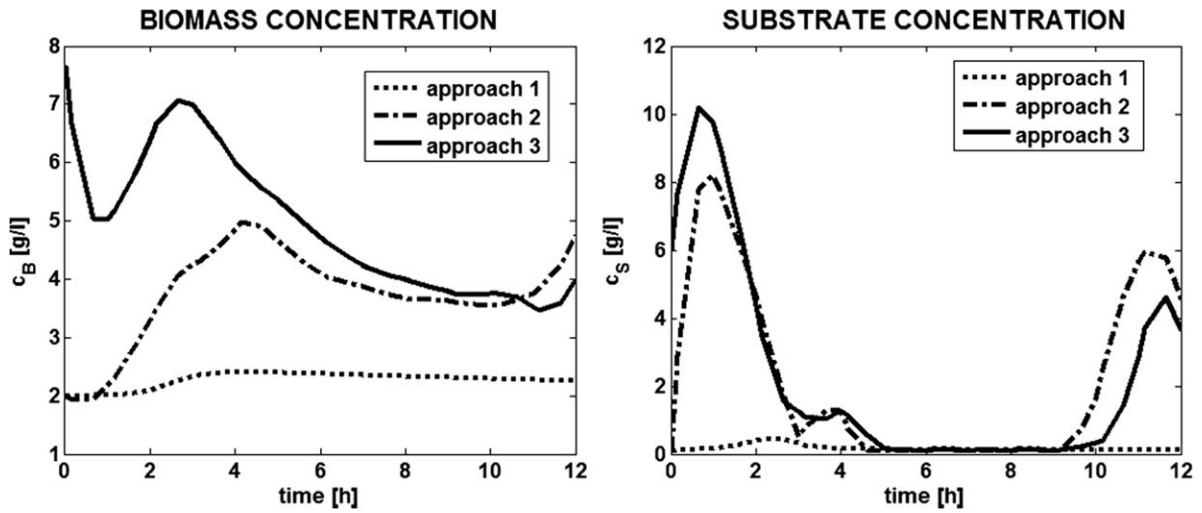

Figure B2. Optimization results with D-criterion—states.

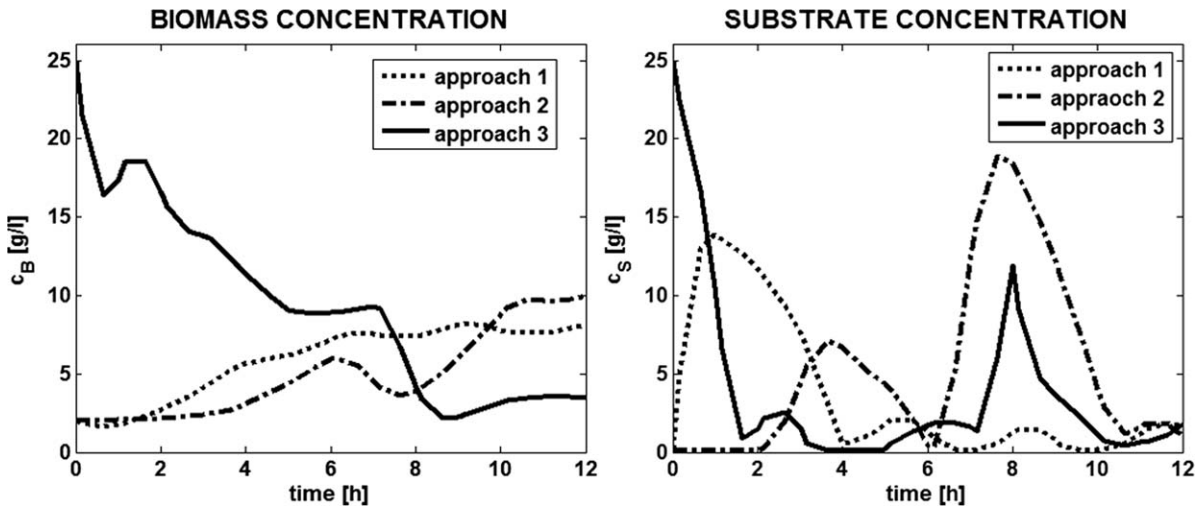

Figure B3. Optimization results with E-criterion—states.

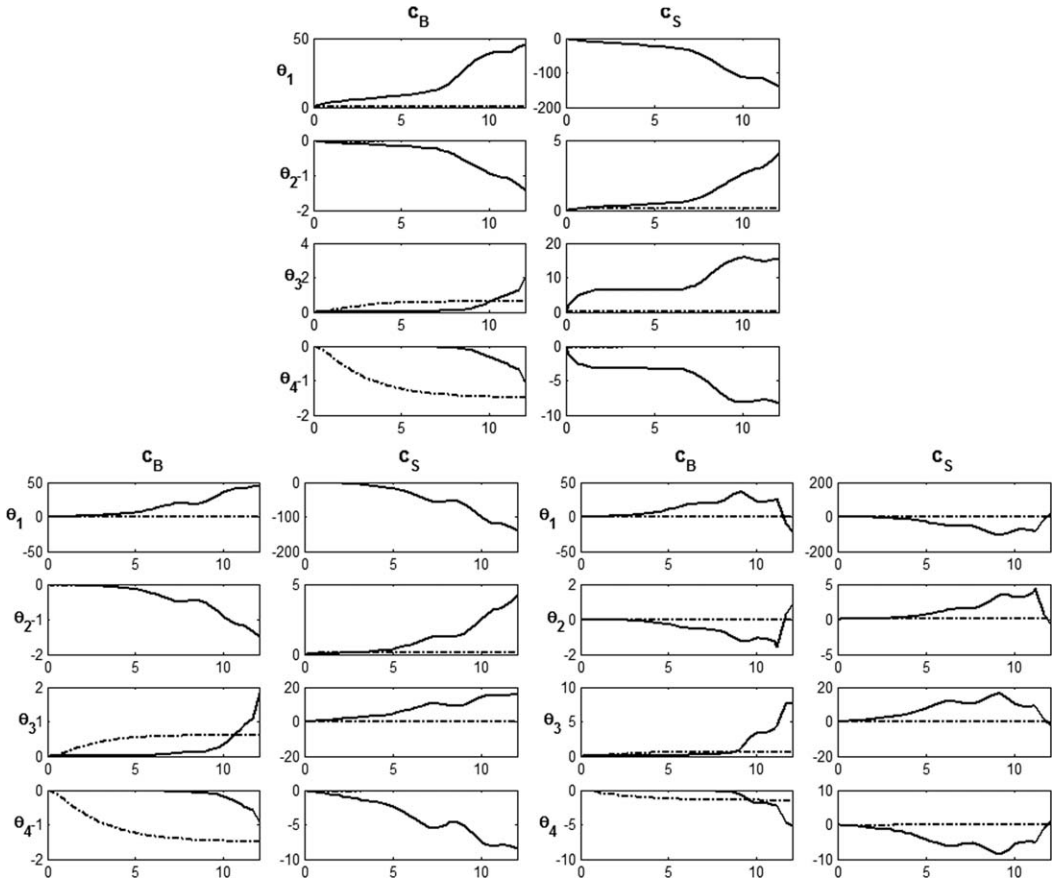

Figure B4. Optimization results with A-criterion—sensitivities, approaches 1, 2, 3.

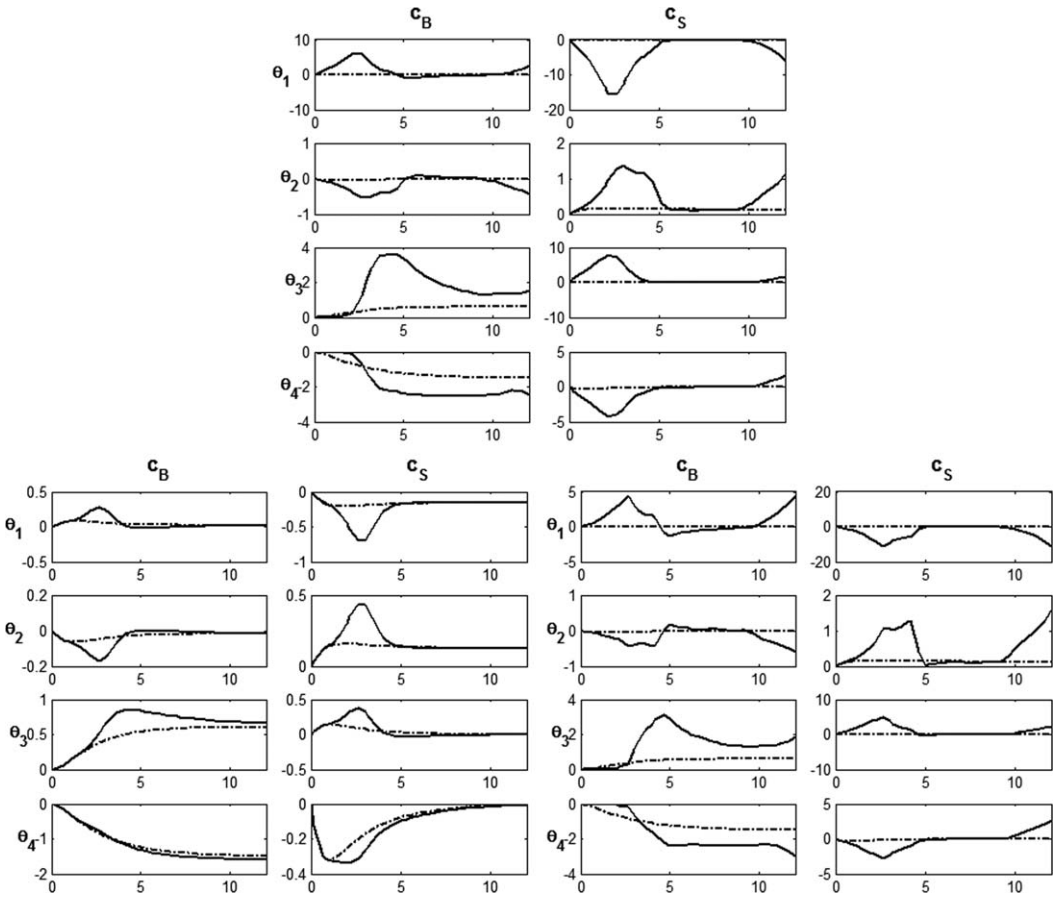

Figure B5. Optimization results with D-criterion— sensitivities, approaches 1, 2, 3.

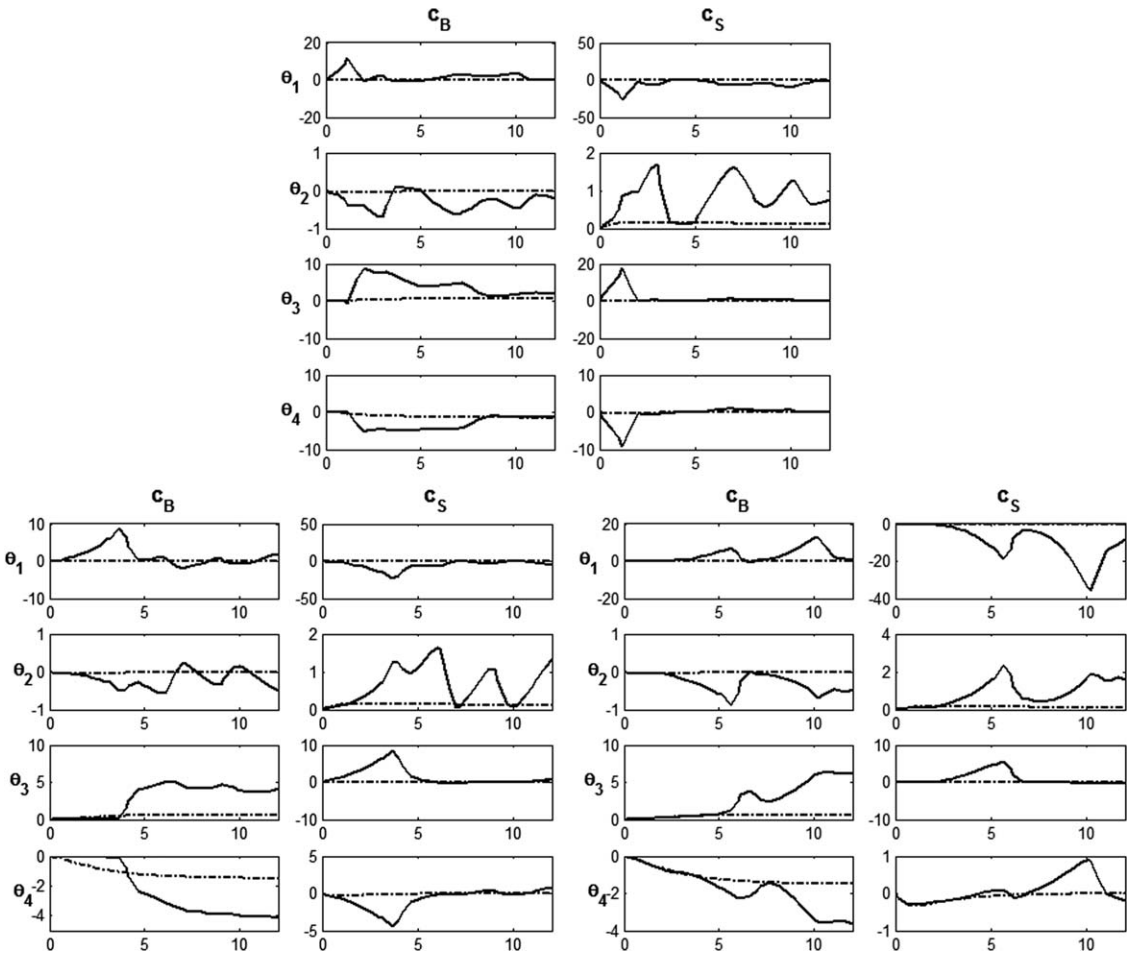

Figure B6. Optimization results with E-criterion— sensitivities, approaches 1, 2, 3.

APPENDIX C:

Table C1. Details from Figure 5a

NUMBER OF ELEMENTS	4	5	6	7	8	9	10	11	12
A-crit opt. result	4.1E+05	4.3E+05	1.3E+06	4.2E+06	1.0E+07	1.4E+05	2.2E+05	9.8E+05	6.1E+04
D-crit opt. result	5.0E-01	1.3E+02	4.5E+04	7.1E+04	1.2E+05	2.7E+05	7.8E+05	2.4E+05	1.4E+05
E-crit opt. result	2.0E+00	2.8E+00	2.3E+00	1.1E+00	5.1E+00	6.1E+00	5.7E+00	4.5E+00	3.1E+00

Table C2. Details from Figure 5b

NUMBER OF ELEMENTS	4	5	6	7	8	9	10	11	12
A-crit opt. endtime [h]	7.5	6.0	7.0	7.6	11.3	10.8	10.8	11.4	12.0
D-crit opt. endtime [h]	8.0	9.8	7.0	9.4	9.4	9.4	12.0	11.0	12.0
E-crit opt. endtime [h]	7.5	9.0	9.1	12.0	10.8	12.0	10.0	11.0	12.0

Manuscript received Sept. 26, 2012, and revision received Feb. 17, 2013.