# Incremental Identification of Fluid Multi-Phase Reaction Systems

**Claas Michalik, Marc Brendel, and Wolfgang Marquardt**
AVT—Process Systems Engineering, RWTH Aachen University, D-52064 Aachen, Germany

*DOI 10.1002/aic.11738*

*Despite their importance, rigorous process models are rarely available for reaction and especially multi-phase reaction systems. The high complexity of these systems, which is due to the superposed effects of mass transfer and intrinsic reaction, is the major barrier for the development of process models. A methodology that allows the systematic decomposition of mass transfer and chemical reaction and thus enables the efficient identification of multi-phase reaction systems is proposed in this work. The method is based on the so-called Incremental Identification Method, recently presented by Brendel et al., Chem Eng Sci. 2006;61:5404-5420. The method allows to easily test the identifiability of a system based on the available measurement data. If identifiability is given, the intrinsic reaction kinetics can be identified in a sound and numerically robust manner. These benefits are illustrated using a simulated 2-phase enzyme reaction system.* © 2009 American Institute of Chemical Engineers *AIChE J,* 55: 1009–1022, 2009
*Keywords: reaction modeling, model identification, multi-phase, biocatalysis, parameter estimation, intrinsic reaction kinetics*

## Introduction

Multi-phase catalytic systems play an important role in today's chemical industry because they allow the immobilization of homogeneous catalysts and an efficient product separation.[1] Zimmermann[2] states that ~95% of all chemical processing steps are heterogeneous. Despite the high importance of multi-phase chemical systems, mechanistic modeling of these systems is rarely done, although the major model ingredients, that is, widely accepted constitutive equations describing mass transfer as well as reaction kinetics, are readily available. The most challenging part remains the identification of the structure of reaction kinetic and mass transfer laws along with the model parameter values.[3] In multi-phase systems, mass transfer and reaction kinetics are superposed and therefore the available concentration measurements reflect the sum of both effects, mass transfer and chemical reaction. This situation renders the identification of reaction kinetics in multi-phase systems a particularly complex task. Therefore, mass transfer effects are typically not treated explicitly during the identification of multi-phase reaction systems but the sum of both effects, mass transfer and intrinsic reaction kinetics, is investigated.[4]Although followed frequently, this approach does not generally allow to identify the intrinsic reaction kinetics in a sound manner. Rather, some lumped model is obtained, which reflects reaction kinetics masked by mass transfer effects.[5] However, knowledge of the intrinsic reaction kinetics is indispensable for a valid process model. Only if the effects of both—intrinsic reaction and mass transfer—are known, predictions on the process behavior in other hydrodynamic regimes can be made. Thus, scale-up or a rigorous optimization of the reaction conditions and media can also only be performed based on knowledge of the intrinsic reaction and mass transfer kinetics.[6,7] However, because of the high complexity of the identification of multi-phase reaction systems, it is still common practice to concentrate investigations on lumped models (see for instance, the recent works by Hu et al.,[4] Ganapati and Neesha,[8] and Saito et al.[9]).

Thus, for multi-phase reaction systems, a gap exists between the high value of mechanistic process models on the one hand and the high burden of obtaining these models on

---

Current address of M. Brendel: Evonik Degussa GmbH, Process Technology, 63457 Hanau, Germany

Correspondence concerning this article should be addressed to W. Marquardt at wolfgang.marquardt@avt.rwth-aachen.de.

the other hand. We therefore present a methodology in this work that allows the efficient identification of multi-phase reaction system models and thus bridges the aforementioned gap. It allows to decouple the mass transfer and reaction kinetics and thus leads to a sound identification of the intrinsic reaction kinetics. The method is based on the incremental identification method,[10] which has been successfully applied to the identification of multi-component diffusion[11] as well as a homogeneous enzyme-catalyzed reaction system.[12]

## Established Methods for the Identification of Multi-Phase Reaction Systems

### *Multi-phase reaction systems*

Multi-phase reaction systems are usually distinguished according to their Hatta number (*Ha*), which relates the velocity of the chemical reaction to the velocity of the mass transfer. If $Ha < 0.3$, the reaction is considered being slow compared with the mass transfer, that is, the reaction is the rate limiting step, and the concentrations in both phases can safely be assumed being in equilibrium at all times.[13] If $0.3 \leq Ha \leq 2$, the system is in the so-called intermediate regime. Here, part of the reaction occurs within the film and another part in the bulk phase.[13] For $Ha > 2$, the system is mass transfer limited. The Hatta number and thus the rate limiting step can in general be influenced by the amount of catalyst present in the system and the energy input by stirring as done for instance in the investigations of Quadros et al.[14] Independent of the setting in which the reaction usually occurs, it is strongly recommended to conduct the experiments for the identification of the reaction laws in the reaction limited regime and for the identification of the mass transfer laws in the mass transfer limited regime. In case of a mass transfer limited reaction system, the reaction laws can not be determined, because the system's behavior and thus the measurable concentration profiles solely depend on the mass transfer rates.[15] In case of the intermediate regime, both, reaction and mass transfer rates are in principle identifiable, but the problem becomes unnecessarily complex, because reaction and diffusion laws have to be identified at the same time.[14] In addition, a relevant part of the reaction occurs within the film between the phase interface and the bulk phase. Here, the concentration is not constant and thus in general, a spatially distributed model has to be employed and spatially distributed measurements within the film are required, except the reaction is known to be of first order. In this special case, the concentration profile in the film can be calculated and no spatially distributed measurements need to be available.[16] However, in the general case of an unknown reaction mechanism, this is not possible.[16] In case of the reaction limited regime, the mass transfer does not need to be modeled explicitly, because equilibrium can safely be assumed at all times.[15]

### *Initial rates method*

The identification of multi-phase reaction systems is a challenging task, and the classical identification methods are not well suited for this case. The initial rates method is well known[15] and has frequently been applied for the identification of reaction systems.[17–19] Here, the initial reaction rate is calculated from an estimate of the initial slope of the concentration trajectory, which is typically obtained by manually drawing the tangent to the first measurement data. Various initial conditions with all but one component provided in high surplus are used. This allows to obtain a set of initial conditions and initial reaction rates. This set of rates and corresponding concentrations is used to estimate the parameters in the kinetic rate expression by some sort of nonlinear least-squares approach. It is obvious, that this method requires a comparably high amount of measurement data because for each experiment, just one initial rate value is obtained. Furthermore, the estimation of a time derivative of a measured signal constitutes an ill-posed problem,[20] which can not properly be solved from few (error prone) experimental data points in principle. In addition, the method cannot be used in all settings. As shown by Zavrel et al.,[21] it is for instance not applicable in biocatalytic reactions where two identical substrates bind to one enzyme. Multi-phase systems can usually not be identified with this method, either. The concentration change due to mass transfer usually dominates at the beginning of the process, and thus the initial change in concentration does not allow to calculate the initial reaction rate. The situation is similar in case of fed-batch experiments, where the observable concentration change is due to the reaction(s) and the feed flow. In addition, multi-phase systems are often designed to allow a simple product separation. Thus, it is desirable to have a low solubility of the products in the reaction phase. This, however, renders it infeasible to provide the products in high surplus, such that a potential product inhibition can not be identified. Despite these and other limitations and although more advanced techniques are known,[22] the method is still popular in the (bio)chemists' communities (see for instance, the recent works[17–19,23,24]).

### *Simultaneous identification method*

Another approach to the identification task is the so-called integral or simultaneous method, which is also known as progress curve analysis in the (bio)chemists community.[25] A dynamic model of the process is at the core of this method. The model, which in general consists of differential and algebraic equations (DAE-model), is then used to identify the reaction kinetics using dynamic parameter estimation techniques.[26,27] The method is ubiquitously applicable, and no limitations to a specific problem class exists. Thus, homogenous and heterogeneous systems can be identified in the same manner, provided that identifiability is given.[28] In contrast to the initial rates method, where the estimation of the initial slope inevitably introduces a bias to the final parameter estimates, this method allows (under suitable assumptions) to compute the statistically optimal parameter values. However, the method comprises some drawbacks. First of all, it requires special purpose software for the dynamic process simulation and parameter estimation, which, in addition, is computationally expensive, because either the model has to be integrated repeatedly[29] or a, generally large, nonlinear programming problem has to be solved.[30–32] In case structurally different model alternatives have to be considered for each kinetic phenomenon, an exponentially growing number of models has to be analyzed.[10] This problem becomes obvious considering a simple reaction system comprising four parallel reactions. If the mechanisms of the reactions were unknown, different alternatives had to be considered for each reaction. If, for example, four alternative model candidates are possible for any of the four reactions, $4^4 = 256$ different combinations of submodels have to be considered, because the method does not allow to identify the mechanisms of individual reactions independently of other reactions occurring in the system. Later, we will demonstrate that the method presented here allows the decomposition of the kinetic phenomena and thus reduces the number of models that have to be identified to $4 \times 4 = 16$ in this case. An exemplary homogeneous reaction system comprising four reactions has been identified with the incremental method by Brendel et al.[10]

In addition, using the simultaneous method, a model that is structurally wrong in a single kinetic expression, either a single mass transfer or reaction kinetic law, will generally result in an overall unsatisfying model performance. The source of this misfit, however, is usually hard to determine because the results of the parameter estimation just show a good or a bad fit but do not allow to reason the influence of the individual rate laws. This is in contrast to the method presented here, where each reaction and mass transfer model can be tested individually for suitability.

The most severe problem of the simultaneous approach, however, is the need for good initial parameter guesses. Different problems can arise if no good initial parameter guesses are available. Bad initial guesses might lead to numerical problems during model integration, as process states hitting bounds or reaching infinity. These and other numerical pitfalls are discussed in detail in Chapter 4.2 of the book by Schittkowski.[27] Another problem is the presence of local optima. Dynamic parameter estimation algorithms typically search for local optima, only. Hence, a good model candidate might be deemed inadequate based on a bad fit of the measurement data that is due to the locally, but not globally optimal parameter values. This problem could be avoided by applying global optimization algorithms for the solution of the parameter estimation problem.[33] Unfortunately, it is not yet possible to apply global optimization strategies to relevantly large models.[34,35] Another approach to lower the risk of identifying only locally optimal parameter estimates is to first conduct initial rate measurements to obtain good initial guesses for the model parameters and then apply the simultaneous method in a second step using the available initial guesses.[36] However, as mentioned above, initial rate experiments are not always feasible and can, for instance, hardly be used in case of multi-phase reaction systems.

These remarks clearly show that the identification of multi-phase reaction systems is a complex task and that an efficient identification method is still lacking. Because of these complications, the identification of multi-phase systems is often performed on a rather qualitative level and only the sum of both effects, reaction and mass transfer is identified, or no model identification is performed at all.[37,38] For scale-up, process optimization, and process control, however, exact process models are needed. Thus, in the next section, we will introduce a new method based on the incremental identification method,[10] which allows the efficient identification of multi-phase reaction systems. The method follows a multi-step approach, including steps to obtain initial guesses of the parameters close to the optimal solution and a last step to obtain statistically sound parameter values.

## A Novel Multi-Step Approach to the Identification of Multi-Phase Reaction Systems

The method presented here is an extension of the incremental identification method presented by some of the authors[10] to the general case of multi-phase reaction systems. One main idea of the suggested approach to the identification of multi-phase reaction systems is to replace the integration of the dynamic model by the differentiation of the measurement data, thereby decomposing the identification task into a sequence of fully transparent steps, each of which is of comparably low complexity. The estimation of derivatives based on measurement data and the subsequent use of this information in a parameter estimation framework is also presented elsewhere.[39] The more general idea of using derivatives estimated from measurements to obtain information on a system under investigation has first been presented by Ramsay and Munhall with application to a biomechanics problem.[40] Spline functions have also been used to enhance the quality of parameter estimates in dynamic parameter estimation problems in a nuclear engineering context.[41]

The method presented here will in detail be explained considering a general reaction system with $n_p$ phases, $n_c$ components, and $n_r$ reactions as shown in Figure 1. The identification of this generic multi-phase system can be divided into the following seven steps:

1. System reformulation.
2. Identifiability analysis.
3. Estimation of component flows.
4. Estimation of kinetic rates.
5. Estimation of kinetic model parameters.
6. Identification of the most suitable model structures.
7. Simultaneous correction step.

Some of these steps may have to be carried out repeatedly, depending on the model candidates and the available measurement data. The seven steps will now be described in detail.

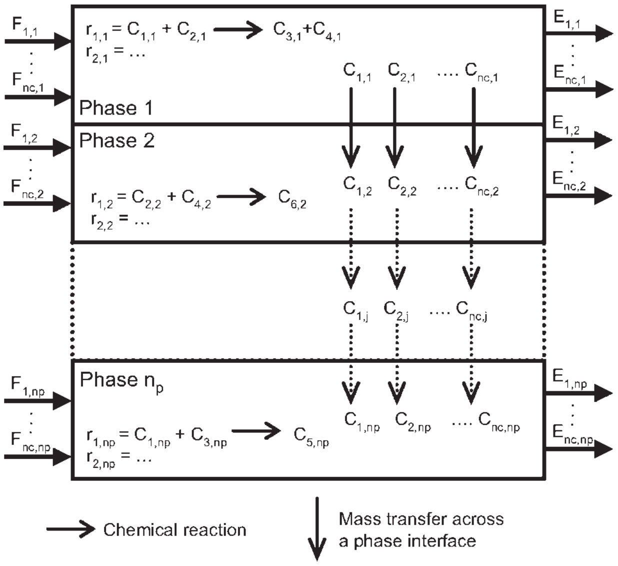

**Figure 1. Scheme of a generic multi-phase reaction system.**

### *Step 1: System reformulation*

In the first step, the system reformulation, certain notation rules are applied to transform the multi-phase system representation into a pseudo-single phase representation. Let us consider a general reaction system consisting of $n_{\mathrm{c}}$ components and $n_{\mathrm{p}}$ perfectly mixed phases. The reformulation starts by indexing the phases by $j \in \mathfrak{S}_{\mathrm{p}}$ and the components by $i \in \mathfrak{S}_{\mathrm{c}}$, with $\mathfrak{S}_{\mathrm{c}}$ being the set of $n_{\mathrm{c}}$ components and $\mathfrak{S}_{\mathrm{p}}$ being the set of $n_{\mathrm{p}}$ phases. The volume of phase $j$ is indicated by $V_j$ and the molar concentration of component $i$ in phase $j$ is denoted as $c_{i,j}$. Furthermore, component feed $F_{i,j}$ and effluent $E_{i,j}$ streams are assumed being potentially present for any component in any phase. For the sake of a simple representation, we assume constant volumes in all phases. The extension to time varying volumes is possible and can easily be worked out.

Assuming also mass transfer of any component across any of the $n_{\mathrm{pi}} = n_{\mathrm{p}} - 1$ phase interfaces, $n_{\mathrm{mt}} = n_{\mathrm{pi}}{*}\ n_{\mathrm{c}}$ unknown mass transfer rates* have to be considered. With $n_{\mathrm{r},j}$ reactions occurring in phase $j$

$$n_{\mathrm{r}} = \sum_{j=1}^{j=n_{\mathrm{p}}} n_{\mathrm{r,j}} \tag{1}$$

reaction rates result. Hence, a model involving $c_{\mathrm{cp}} = n_{\mathrm{c}}{*}\ n_{\mathrm{p}}$ pseudo-components and $n_{\mathrm{k}} = n_{\mathrm{r}} + n_{\mathrm{mt}}$ generally unknown kinetic laws (comprising both, chemical reaction and mass transfer) can be developed. Note that the same chemical component occurring in more than one phase is accounted for as a separate pseudo-component. The pseudo-components can in principle be ordered arbitrarily. However, for later processing, it is advantageous to order them such that the first pseudo-component is the first component in the first phase, the second pseudo-component is the second component in the first phase, and the last pseudo-component is component $n_{\mathrm{c}}$ in phase $n_{\mathrm{p}}$.

The $n_{\mathrm{r}}$ reactions can be described by $n_{\mathrm{p}}$ stoichiometric matrices, whereas $\mathrm{N}_j$ with $j \in \mathfrak{S}_{\mathrm{p}}$ describes the reactions occurring in phase $j$. Each stoichiometric matrix is of size $n_{\mathrm{r},j} \times n_{\mathrm{c}}$. Following the pseudo-component concept, the reactions in all phases can be described by the lumped stoichiometric matrix

$$\mathrm{N}_{\mathrm{ar}} = \begin{bmatrix} \mathrm{N}_1 & 0 & \cdots & 0 \\ 0 & \mathrm{N}_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \mathrm{N}_{n_{\mathrm{p}}} \end{bmatrix}.$$

Each chemical reaction is assumed to be written as a reversible reaction, comprising a forward reaction ($r_+$) and a backward reaction ($r_-$). The resulting total reaction rate is $r = r_+ - r_-$, and the sign of $r$ indicates the direction of the net production or consumption. Note that, although we assume reversible reactions in any case, we use just a single arrow to indicate a reaction. The direction of the arrow is based on the direction of the reaction implicitly assumed by the choice of the sign of the stoichiometric coefficients.

Because mass transfer and chemical reactions result in same terms in the balance equations of the segregated and well-mixed phases, it seems reasonable to treat them in a unifying manner. Hence, mass transfer of chemical component $C_i$ between phases $j$ and $j + 1$ will be described as a set of pseudo-reactions according to

$$C_{i,j} \rightarrow C_{i,j+1}, i \in \mathfrak{S}_c, 1 \leq j \leq n_{pi} \tag{2}$$

giving the pseudo-stoichiometric matrix for the mass transfer kinetics

$$\mathrm{N}_{\mathrm{mt}} = \begin{bmatrix} -1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 0 & \cdots & 0 & 1 & \cdots & \vdots \\ \vdots & 0 & \ddots & 0 & \cdots & 0 & \ddots & \vdots \\ 0 & \cdots & 0 & -1 & 0 & \cdots & 0 & 1 \end{bmatrix}.$$

In general, the size of $\mathbf{N}_{\mathrm{mt}}$ can be reduced drastically, because the solubility of some components will be practically zero in some phases, and thus the number of pseudo-components can be reduced.

$\mathbf{N}_{\mathrm{mt}}$ implicitly contains information on the direction of each mass transfer rate in form of the signs of the stoichiometric coefficients. All mass transfer rates are formally assumed to proceed from phase $i$ to phase $i + 1$, which is mechanistically unreasonable. Therefore, it is important to mention that the direction, postulated for mass transfer, does not need to be known a priori. Parallel to the case of a reaction, a negative mass transfer rate will be calculated, if the mass transfer proceeds in the opposite direction. Consequently, the rate law to be identified should also give negative values in this case, if it is mechanistically correct.

Finally, the whole multi-phase reaction system of $n_{\mathrm{c}}$ components, $n_{\mathrm{p}}$ phases and $n_{\mathrm{r}}$ reactions can be described as a pseudo-single phase system comprising the set $\mathfrak{S}_{\mathrm{pc}}$ of $n_{\mathrm{pc}}$ pseudo-components and the set $\mathfrak{S}_{\mathrm{k}}$ of $n_{\mathrm{k}}$ pseudo-reactions. The system can thus be described introducing the pseudo-stoichiometric matrix

$$\mathrm{N}_{\mathrm{p}} = \begin{bmatrix} \mathrm{N}_{\mathrm{ar}} \\ \mathrm{N}_{\mathrm{mt}} \end{bmatrix}$$

of size $n_{\mathrm{k}} \times n_{\mathrm{pc}}$. This definition of $\mathbf{N}_{\mathrm{p}}$ also orders the pseudo-rates such that the first $n_{\mathrm{r}}$ pseudo-rates represent reactions and the last $n_{\mathrm{mt}}$ pseudo-rates represent mass transfer.

### *Step 2: Identifiability analysis*

The second step, the identifiability analysis, is based on a rank analysis of the so-called measurement matrix $\mathbf{N}_{\mathrm{m}}$. This matrix derives from the pseudo-stoichiometric matrix $\mathbf{N}_{\mathrm{p}}$ by zeroing all columns representing nonmeasured pseudo-components. The complete reaction system can only be identified using the incremental method if $\mathbf{N}_{\mathrm{m}}$ has full rank. If $\mathbf{N}_{\mathrm{m}}$ is rank-deficient, it is still possible to estimate a subset of the reactions which in turn allows to identify a subset of the model parameters. A detailed description of the procedure

---

* Because mass transfer is area specific, the term mass transfer flux rather than mass transfer rate should be used; however, to keep the notation of reaction and mass transfer kinetics identical, we will use the term mass transfer rate throughout the article.

that also explains, which rates can be estimated based on the available measurement data, is given elsewhere.[10] In case of a non or only partly identifiable system, either the system's complexity has to be reduced by posing more restrictive assumptions on the model or the measured data set needs to be extended by measuring more chemical components to achieve identifiability. However, the rank analysis of $\mathbf{N}_{\mathrm{m}}$ is only a necessary but not a sufficient condition for identifiability. A positive result of the rank analysis means that rate estimates for all involved kinetic phenomena (reactions and mass transfer) can be calculated in step 4. Based on the rate estimates and the initial conditions of the involved components, it is also possible to calculate estimates of the concentration trajectories of these components. If the initial conditions are (in part) unknown, they can be included in the rate expressions as unknown parameters. Identifiability is only given if all rates can be estimated and all variables involved in the rate laws except for the parameters are known. Hence, it also has to be assured that those variables that influence the rate law but are not considered in the stoichiometric matrix are measured. Such variables are, for instance, the reaction temperature in case of a temperature dependent kinetic law or the catalyst concentration, if the reaction is investigated on a coarse level and the catalyst-substrate-intermediates are not modeled explicitly.

A positive result of this identifiability test only assures so-called structural identifiability.[28] Hence, if the identifiability test is passed, the model can in principle be identified based on the available measurements. The accuracy of the estimates, however, depends on the type and quality of measurement data and might be insufficient even for a structurally identifiable model.[28] In case of insufficient accuracy, the whole procedure can—without any restrictions—be incorporated into a model-based optimal experimental design framework.[42]

### *Step 3: Estimation of component flows*

The third step in the procedure is the estimation of component flows, defined as

$$f_{i,j}(t) := V_j \cdot \frac{dc_{i,j}}{dt} - F_{i,j}(t) + E_{i,j}(t),\ i \in \mathfrak{S}_c,\ j \in \mathfrak{S}_p. \qquad (3)$$

Please note that the term component flow is not defined in the chemical engineering community. We introduce this term to indicate the change in the amount of a component independently of the cause. Hence, a component flow $f_{i,j}$ refers to the change in the amount of component $i$ in phase $j$ due to the sum of chemical reaction and mass transfer and is thus neither a flux (which is area specific) nor a rate (which is volume specific). In addition, we define $\mathbf{f}(t)$ as the row vector of all component flows of length $n_{\mathrm{pc}}$ ordered according to the order of the pseudo-components. It is obvious that, to estimate $f_{i,j}(t)$, the concentration $c_{i,j}(t)$, the feed $F_{i,j}(t)$, and the effluent $E_{i,j}(t)$ need to be known, that is, either measured or calculated based on, for instance, the mass balance. It is important to mention that the estimation of $f_{i,j}(t)$ is an ill-posed problem in the sense of Hadamard[43] due to the necessary differentiation of the noise-corrupted concentration measurements. This means that small errors in the inputs (the measured concentrations) might lead to large errors in the outputs (the derivatives of the measured concentrations with respect to time). However, regularization techniques allow to obtain meaningful results even for ill-posed problems.[20] A variety of regularization techniques as for instance Tikhonov regularization[44] or filter based approaches[45,46] are known. The aim of regularization is to find the solution of a well-posed problem, which is as close as possible to the original, ill-posed problem. Any regularization method relies on a tradeoff between staying as close as possible to the original problem and obtaining a problem as well-posed as possible. The amount of regularization has thus to be chosen wisely. Different methods to determine a reasonable or even statistically optimal amount of regularization are known. The most popular alternatives are the heuristic L-curve criterion and the statistically funded generalized cross validation method. For an introduction to ill-posed problems and regularization techniques, see, for example, the publication by Hansen.[20] According to our experience, the estimation of a smooth concentration profile based on smoothing splines[47] and subsequent differentiation of the smooth curve leads to good estimates. Tikhonov regularization[44] and generalized cross validation[20] to determine the regularization parameter also proved to be appropriate for the problem class at hand. However, independently of the regularization technique used, the estimated derivatives will be biased. The bias introduced at this step will propagate through step 4 and 5 of the estimation procedure. Hence, only low quality estimates of the model parameters can be expected in step 5, if the measurement quality is low, meaning that either the data are scarce or corrupted by a high amount of noise. However, as shown later, the last step allows to obtain the statistically optimal parameter values, even in the case of low-quality data.

### *Step 4: Estimation of rates*

Assuming identifiability, it is possible to estimate all kinetic rates according to

$$f_{i,j}(t) = \sum_{k=1}^{n_k} S_k \cdot v_{i,k} \cdot r_k(t), i \in \mathfrak{S}_c, j \in \mathfrak{S}_p \cdot \qquad (4)$$

Here $v_{i,k}$ is the stoichiometric coefficient of pseudo-component $i$ in kinetic phenomenon $k$ and $S_k(t)$ is either the volume (in case of a reaction) or the interfacial area (in case of a mass transfer). Sometimes, the interfacial area is not known a priori and can hardly be determined based on measurements.

In such cases, it is common to use a mass transfer model, which contains the product of the mass transfer coefficient and the interfacial area. This approach can also be applied within the incremental identification framework. In this case, an arbitrary interfacial area has to be used. The pseudo mass transfer coefficient comprising both, interfacial area and mass transfer coefficient can then be calculated as the product of the arbitrarily chosen interfacial area and the estimated mass transfer coefficient. Equation 4 can be written in matrix form as

$$\mathrm{f}(t) = \mathrm{r}_{\mathrm{p}}(t) \cdot \mathrm{N}_{\mathrm{ep}}. \qquad (5)$$

Here, $\mathbf{r}_{\mathrm{p}}(t)$ is the row vector of all pseudo-rates and $\mathbf{N}_{\mathrm{ep}}(t)$ is an extended stoichiometric matrix, which can be derived from $\mathbf{N}_{\mathrm{p}}$ by multiplying each element in row $k$ by $S_k$.

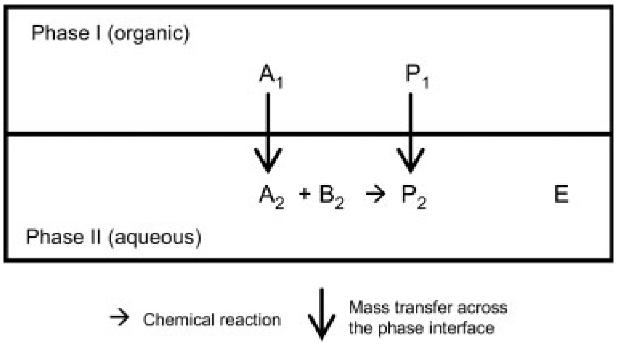

**Figure 2. Scheme of the two phase enzyme-catalyzed reaction system.**

Because identifiability is given, Eq. 5 defines a set of linear equations that is at least determined but generally over-determined. Hence, the row vector of pseudo-reaction rates can be calculated using the pseudo inverse[48] of $\mathbf{N}_{\mathrm{ep}}$ according to

$$\mathrm{r}_{\mathrm{p}}(t) = \mathrm{f}(t) \cdot \mathrm{N}_{\mathrm{ep}}^{\dagger}. \tag{6}$$

A vector of reaction rates $\mathbf{r}_{\mathrm{r}}(t)$ and mass transfer rates $\mathbf{r}_{\mathrm{mt}}(t)$ can be obtained by extracting the first $n_{\mathrm{r}}$ and the last $n_{\mathrm{mt}}$ rates from $\mathbf{r}_{\mathrm{p}}(t)$, respectively.

### *Step 5: Estimation of kinetic parameters*

In the fifth step, $\mathbf{r}_{\mathrm{r}}(t)$ and $\mathbf{r}_{\mathrm{mt}}(t)$ are used to estimate the kinetic parameters of the postulated model candidates using nonlinear regression techniques. Typical model candidates for reaction rates are for example

$$r_{\mathrm{r},1} = k_1 \cdot c_{1,2} \cdot c_{2,2}, \tag{7}$$

$$r_{\mathrm{r},2} = k_2 \cdot c_{1,2}^2 \cdot c_{2,2}, \tag{8}$$

$$r_{\mathrm{r},3} = k_3 \cdot c_{1,2} \cdot e^{\left(\frac{-Ea_3}{R \cdot T}\right)} \tag{9}$$

$$r_{\mathrm{r},4} = \frac{\frac{k_{\mathrm{catf}}}{K_{iA} \cdot K_{mB}} \cdot \left(c_{1,2}^2 - \frac{c_{3,2}}{K_{\mathrm{eq}}}\right)}{1 + \frac{c_{1,2}}{K_{iA}} \cdot \left(1 + \frac{K_{mA}}{K_{mB}}\right) + \frac{c_{1,2}^2}{K_{iA} \cdot K_{mB}} + \frac{c_{3,2}}{K_{mP}} + \frac{c_{1,2} \cdot c_{3,2}}{K_{mP} \cdot K_{iB}}}. \tag{10}$$

The last example is a typical expression for a rate of an enzyme-catalyzed reaction and is obviously more complex than typical rate laws for elementary reactions. The third example shows a typical Arrhenius expression, where the reaction rate is expressed as a function of the temperature $T$, the gas constant $R$, the activation energy $Ea_3$, and the kinetic constant $k_3$. If the temperature is known (either measured or controlled), temperature dependent expressions can be treated by the incremental method in the same manner as nontemperature dependent reactions.

An exemplary, simple mass transfer rate law is given by

$$r_{\mathrm{mt},1} = \beta_1 \cdot \left(\frac{c_{1,1}}{m_{1,1,2}} - c_{1,2}\right) \tag{11}$$

where $\beta_1$ is the mass transfer coefficient of component 1 and $m_{1,1,2}$ the partition coefficient of component 1 between phase 1 and 2. More complex expressions for mass transfer laws can be found elsewhere.[49] Although we assume—for the sake of simplicity—that the partition coefficient is known, this parameter can of course also be estimated using the method presented here.

The estimation of the model parameters requires the solution of a nonlinear least-squares problem. Different methods have been proposed to solve this kind of optimization problem as, for instance, the Levenberg-Marquardt method,[50,51] and are readily available in numerous software packages.[52–54] For nonlinear least-squares problems global optimization codes are available that can in principle be applied to solve the problem at hand.[55] In contrast to the dynamic parameter estimation problem naturally arising in the simultaneous method, the nonlinear problems can be solved to the global optimum in a reasonable amount of time.[56,57] This is especially true because the number of parameters that have to be estimated simultaneously in this step of the incremental method is usually lower than in dynamic parameter estimation using the simultaneous method. This is because of the fact that, at this point, one estimate is available for each reaction or mass transfer rate and thus the model candidates can be tested for each reaction or mass transfer phenomenon individually. This is one of the key advantages when applying the multi-step approach to the identification of multi-phase reaction networks because it allows to decompose the identification of the whole network into the repeated identification of single reaction and mass transfer laws. Each individual subproblem is by nature less complex than the overall problem. Typically the individual subproblems comprise less parameters to be estimated, which is especially advantageous in the context of global optimization, where the computational cost growth exponentially with the number of model parameters.[55] In addition, a wrong model candidate for a single mass transfer or reaction kinetic law can now be identified easily and has no negative influence on the identification of rate laws for other kinetic phenomena. This is a key advantage of the method presented here compared to classical approaches. Following a simultaneous identification procedure, the whole kinetic model has to be identified at once. Because the model equations are typically linked, a model that is structurally wrong in one part—as for instance one reaction kinetic law—will thus generally lead to an overall bad model performance, meaning that the model is not able to fit the experimental data accurately. Hence, in such a framework, it is hard to say whether such an insufficient fit is due to parameter estimates that correspond to a bad local optimum or due to an (in part) structurally wrong model. In addition, if a model is structurally wrong in one part, it is hard to identify

**Table 1. The True Parameter Values of the Kinetic Model**

| Parameter | $k_{\mathrm{cat\ f}}$ | $K_{\mathrm{eq}}$ | $K_{mA}$ | $K_{mB}$ | $K_{mP}$ | $K_{iA}$ | $K_{iB}$ |
|---|---|---|---|---|---|---|---|
| Value (Unit) | 45 ($\mathrm{S}^{-1}$) | 4 ($\mathrm{mM}^{-1}$) | 0.2 (mM) | 0.6 (mM) | 1 (mM) | 1 (mM) | 2.25 (mM) |

Table 2. Experimental Conditions for all Five Experiments

| Experiment No. | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $N_A$ | 14.78 | 1.0 | 7.543 | 14.78 | 14.78 |
| $N_B$ | 3.263 | 20.4044 | 15.18 | 3.263 | 3.263 |
| $N_P$ | 2.435 e −5 | 0.0 | 0 | 0 | 0 |
| $N_E$ | 2.971 e −3 | 2.929 e −3 | 1 e −6 | 2.971 e −3 | 2.971 e −3 |

$N_i$, $i \in \{A, B, P, E\}$ refers to the amount of $i$ in mmol initially present in the system.

the problematic model equations. Besides the advantages, it is also important to remind that the quality of the parameter estimates obtained in this step is directly linked to the quality of the estimated rates. Hence, these estimates are also biased due to the bias originally introduced in the estimation of the time derivatives of the concentration measurements. Therefore, the suitable model candidates (identified in the next step) have to be processed in a final simultaneous correction step, described in more detail below.

### *Step 6: Identification of most suitable model candidate(s)*

If structurally different model candidates are considered, the most suitable candidates have to be identified at this point. The choice should be based on a reasonable model selection criterion, as for instance Akaike's information criterion[58] or the discrimination criterion based on the inverse probability theorem of Bayes proposed by Stewart et al.[59] To decide, which model candidates should further be considered and which ones should be dismissed from the candidate set, error estimates of the estimated rates need to be known. This is necessary, because the model candidates are judged based on the fit of the estimated rate, which deviates from the true rate because its calculation is based on noise corrupted measurements. Bootstrapping or similar methods can be used to obtain error estimates for the estimated reaction rates.[60] The typically high computational cost of this kind of methods is no problem in this framework, because only computationally cheap linear problems have to be solved within the bootstrapping framework.

### *Step 7: Simultanous model correction*

Finally, in a last step, the statistically optimal model along with the statistically optimal parameter values can be identified by solving a dynamic parameter estimation problem, only taking the remaining model candidates into account. This so-called simultaneous model correction is necessary due to the bias, which is inevitably introduced in the estimation of component flows from noise-corrupted measurement data. The computational cost of this last dynamic parameter estimation problem is significantly reduced, because good initial guesses provided by the previous step, are already available. In case of unsatisfying parameter accuracy or if no single model structure can be identified, the procedure can be integrated into an iterative optimal experimental design framework without any restrictions.[61]

## Illustrative Case Study: Two-Phase Enzyme-Catalyzed Reaction System

In this section, we demonstrate the method outlined above using a simulated two-phase enzyme reaction system.

### *The reaction system*

Biocatalysis in nonconventional media, including organic/aqueous two-phase systems, is a promising technology for synthetic chemistry.[1] The perspective of applying highly selective and specific biocatalysts to reactions that are classically performed in organic solvents is very attractive for science and industry. However, in those systems, several limitations exist, for example, low solubility of substrates and products leading to low volumetric productivities, hydrolysis of products or side reactions. To overcome those difficulties, nonconventional media, such as organic solvents or supercritical fluids, have been introduced in bioprocesses.[1] In organic/aqueous two-phase systems, several advantages are foreseen in comparison with traditional aqueous systems. Firstly, the organic phase can serve as a reservoir for substrates and products at high concentrations, whereas the reactive aqueous phase contains the enzymes. Thus, an increase of the volumetric productivity may be achieved.[62] Another advantage is the improvement of product recovery and catalyst recycling as the organic and aqueous phases can easily be separated. Furthermore, a well-designed biphasic system may lead to an improved equilibrium conversion, as for example shown for the reduction of acetophenone with alcohol dehydrogenase from *Lactobacillus brevis* by Eckstein et al.[63] and for the hydrolysis of Penicillin G with penicillin amidase from *Escherichia Coli* by Diender et al.[64] Finally, possible inhibitory effects of the substrate and/or the product can be avoided by reducing the concentration of inhibitory substances in the aqueous phase.[65] Hence, two-phase systems possess desirable properties for biocatalytic reactions. On the other hand, they pose a hard problem for model identification due to the already mentioned coupling of mass transfer and reaction. In addition, the rate expressions for enzyme-catalyzed reactions tend to be complex and highly nonlinear, as exemplarily shown in Eq. 10.

Let us consider the enzyme-catalyzed reaction

$$A + B \xrightarrow{E} P. \qquad (12)$$

This reaction is taking place in the aqueous phase of a two-phase aqueous/organic system. Let us further assume that the solubility of component $B$ and the enzyme ($E$) in the organic phase are practically zero. A system as depicted in Figure 2 results with no effluent and feed. We assume a kinetic law given by

Table 3. Partition Coefficients of Components *A* and *P*

| Component | Partition coefficient |
|---|---|
| $A$ | 0.03 |
| P | 0.95 |

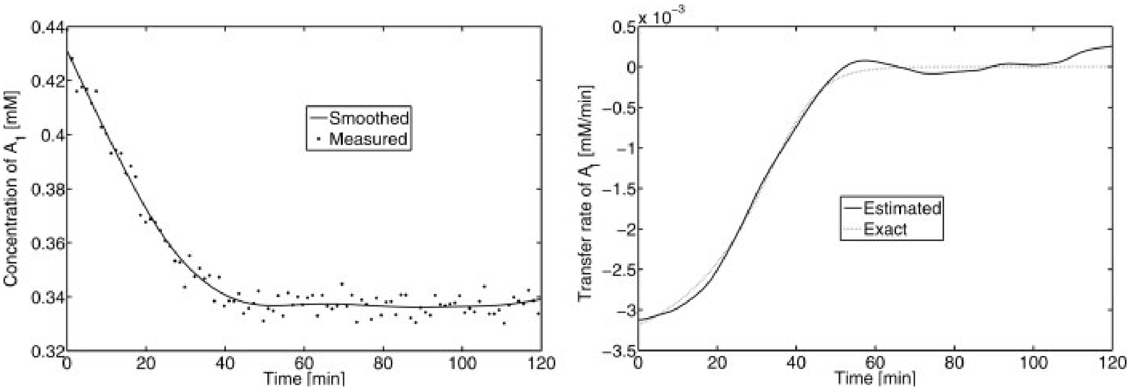

**Figure 3. Measured and smoothed concentration (left) and exact and estimated component flow of $A_1$ (right) in the first experiment.**

$$r = [E] \cdot \frac{\frac{k_{\text{catf}}}{K_{iA} \cdot K_{mB}} \cdot \left([A] \cdot [B] - \frac{[P]}{[K_{\text{eq}}]}\right)}{1 + \frac{[A]}{K_{ia}} + \frac{K_{ma} \cdot [B]}{K_{iA} \cdot K_{mB}} + \frac{[A] \cdot [B]}{K_{iA} \cdot K_{mB}} + \frac{[P]}{K_{mP}} + \frac{[B] \cdot [P]}{K_{mP} \cdot K_{iB}}}, \tag{13}$$

which has for instance been identified by Zavrel et al.[21] for modeling the kinetics of ThDP-dependent enzymes. In Eq. 13, $[A]$, $[B]$, $[P]$, and $[E]$ represent the concentrations in the aqueous phase and $k_{\text{catf}}$, $K_{\text{eq}}$, $K_{mA}$, $K_{mB}$, $K_{mP}$, and $K_{iA}$ are the kinetic parameters to be estimated. $K_{iB}$ is a dependent parameter that can be calculated from estimated parameter values according to

$$K_{iB} = \frac{K_{mB} \cdot K_{iA}}{K_{mA} \cdot \left[1 - \left(\frac{K_{mA}}{K_{iA}} - 1\right) \cdot \frac{K_{mP}}{K_{\text{eq}} \cdot K_{mB} \cdot K_{iA}}\right]}. \tag{14}$$

The true parameter values are given in Table 1.

Let us further assume that both phases comprise a volume of 1 l and that the amounts converted by the reaction and transferred across the phase interfaces are small, such that the phase volumes can be considered constant.

To illustrate the procedure, we create pseudo-measurements by simulating the system and adding uncorrelated white noise with a standard deviation of 3% of the concentration value. We assume that the concentrations of three components, namely $A$ and $P$ in the aqueous phase and $B$ in the organic phase are measurable and that 100 data points are sampled uniformly for each component and each experiment. We simulate five experiments with different initial conditions as given in Table 2. The experiment duration is 120 min for all experiments. We further assume the initial concentrations of all components as well as the constant concentration of the catalyst to be known. The partition coefficients for $A$ and $P$, defined as

$$m_i = \frac{c_{i,1,\text{equ}}}{c_{i,2,\text{equ}}}, \qquad i \in A, P, \tag{15}$$

are known and given in Table 3. $m_{\text{i}}$ is the partition coefficient of component $i$ and $c_{i,1,\text{equ}}$ and $c_{i,2,\text{equ}}$ are the equilibrium concentrations of $i$ in phase 1 and 2, respectively. Note, that any other more realistic phase equilibrium could be employed, instead.

We illustrate the method using the settings outlined above, however, for comparison, we solve the same identification task with four different settings, which differ with respect to the quality of the measurement data. We use two different levels of resolution in time (100 and 25 measurements per component and experiment, corresponding to sampling intervals of 1.2 and 4.8 min) and two different noise levels (3% and 10% relative white noise).

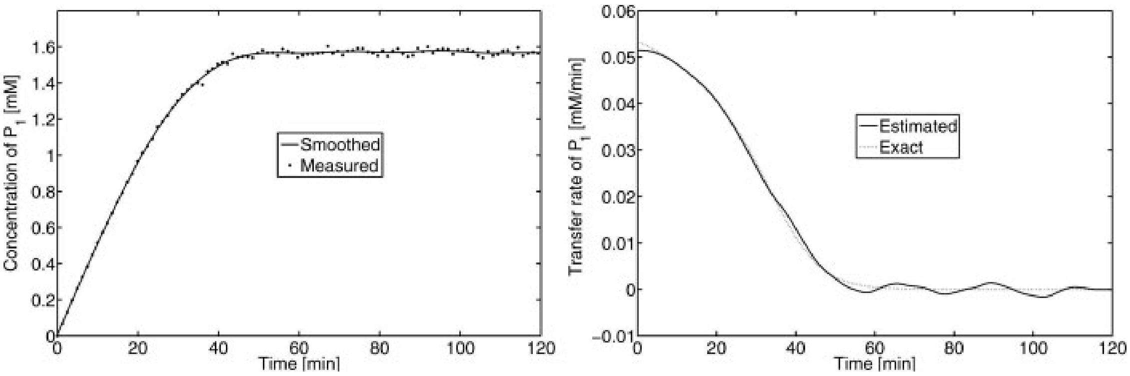

**Figure 4. Measured and smoothed concentration (left) and exact and estimated component flow of $P_1$ (right) in the first experiment.**

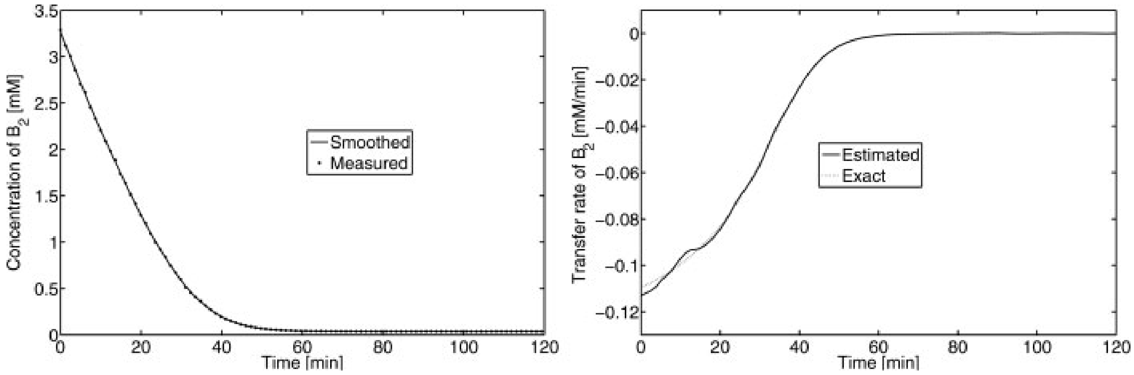

**Figure 5. Measured and smoothed concentration (left) and exact and estimated component flow of $B_2$ (right) in the first experiment.**

### *Identification of the biphasic enzyme-catalyzed reaction system*

To illustrate the procedure outlined above, we demonstrate the seven steps for the simulated two-phase enzyme-catalyzed reaction system.

*Step 1: System Reformulation.* In this case, we have four components ($\mathfrak{S}_c = \{A, B, P, E\}$) and two phases ($\mathfrak{S}_p = \{1, 2\}$). Thus, a set of eight pseudo-components results. However, as components $B$ and $E$ are not soluble in the first phase, only six pseudo-components are considered. Thus, the set $\mathfrak{S}_{pc} = \{A_1, P_1, A_2, B_2, P_2, E_2\}$ results.

Here, $A_1$ is component $A$ in the first phase, $A_2$ component $A$ in the second phase and so on. A single reaction takes place in the second phase, and two mass transfer rates (for A and P) have to be considered. Thus, the pseudo-stoichiometric matrix

$$\mathrm{N_p} = \begin{array}{c} \begin{array}{cccccc} A_1 & P_1 & A_2 & B_2 & P_2 & E_2 \end{array} \\ \left[\begin{array}{cccccc} 0 & 0 & -1 & -1 & 1 & 0 \\ -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 1 & 0 \end{array}\right] \end{array}$$

results.

*Step 2: Identifiabilty Test.* The second step in the procedure is to test the structural identifiability of the system under investigation, and thus $\mathbf{N}_m$ has to be set up as outlined above. Because only $A_1$, $P_1$, and $B_2$ are measurable, $\mathbf{N}_m$ reads as

$$\mathrm{N_m} = \begin{array}{c} \begin{array}{cccccc} A_1 & P_1 & A_2 & B_2 & P_2 & E_2 \end{array} \\ \left[\begin{array}{cccccc} 0 & 0 & 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \end{array}\right] \end{array} \qquad (16)$$

Because $\mathbf{N}_m$ has full rank and the constant concentration of the catalyst is known, all kinetic rates can be estimated and the reaction rate laws can in principle be identified.

*Step 3: Estimation of Component Flows.* The third step in the procedure is the identification of component flows. Because no feed and no effluent are assumed, Eq. 3 simplifies to

$$f_{i,j}(t) := V_j \cdot \frac{dc_{i,j}}{dt}, \quad i \in \mathfrak{S}_c, j \in \mathfrak{S}_p, \qquad (17)$$

here. We use smoothing splines[47] and Tikhonov regularization with generalized cross validation[20] and subsequent differentiation of the smooth estimate to determine the component flows. For the computations, we use the *Spline Toolbox*[66] and the *Hansen Regularization Toolbox*[67] for Matlab. Figures 3–5 exemplarily show the original measurement values and the smooth trajectories as well as the estimated and the exact component flows for the measured components in the first experiment. Each component flow can be calculated for any point in time. However, for later processing, it is advantageous to cal-

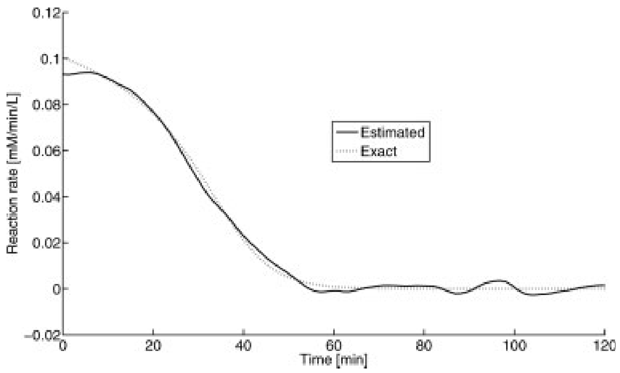

**Figure 6. Estimated and exact reaction rate in the first experiment.**

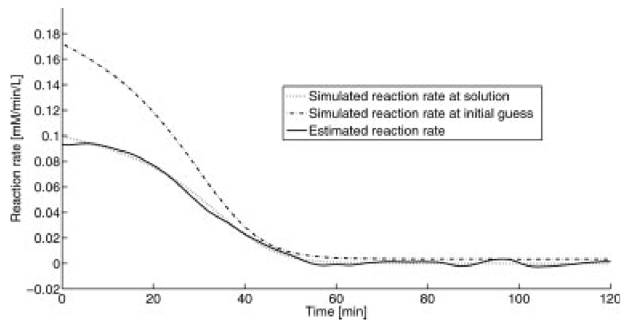

**Figure 7. Model fit after algebraic parameter estimation.**

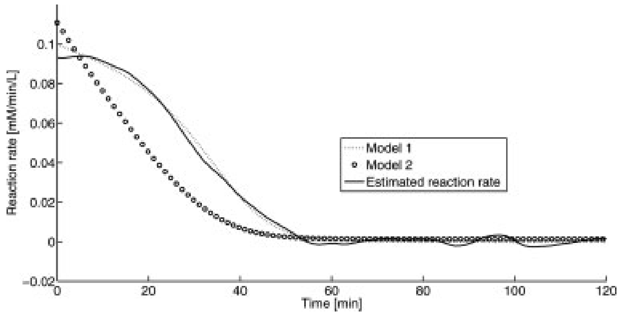

**Figure 8. Comparison of the two competing model candidates.**

culate the component flows at exactly those times, at which measurement information is available.

*Step 4: Estimation of Rates.* Using Eq. 6, the vector of pseudo-reaction rates $\mathbf{r}_p(t)$ can be calculated. This is a row vector of length 3 representing the rate of the chemical reaction and the two mass transfer rates. The vector of reaction rates $\mathbf{r}_r(t)$ (which is a scalar in this case) can be obtained by extracting the first element from $\mathbf{r}_p$ $(t)$. Figure 6 exemplarily shows the estimated and the exact reaction rate for the first experiment.

*Step 5: Estimation of Kinetic Model Parameters.* In this step, the estimated rates are used to estimate the kinetic model parameters by solving a nonlinear least-squares problem. In our implementation, we solve the problem using Ipopt[68] as optimization code by a Matlab-MEX-Interface and a problem reformulation proposed by Schittkowski.[69] In Figure 7, we exemplarily show the result of the algebraic parameter estimation problem for the first experiment[†] and the low initial guess for all parameters values according to Table 5.

*Step 6: Identification of the Most Suitable Model(s).* Let us consider two alternative model candidates here. Candidate one is the correct model given in Eq. 13 and candidate two is

$$r = [E] \cdot \frac{\frac{k_{\mathrm{catf}} \cdot [B]}{K_{mB}}}{1 + \frac{[B]}{K_{mB}} + \frac{[P]}{K_{mP}}}, \tag{18}$$

where $[E]$, $[B]$, and $[P]$ are the concentrations of $A$, $B$, and $P$ in the aqueous phase and $k_{\mathrm{catf}}$, $K_{mB}$, and $K_{mP}$ are the model parameters to be estimated. Figure 8 shows the estimated reaction rate and the model predictions of the two competing model candidates after the parameter estimation for the first experiment. It can be seen, that the second model does not allow to fit the estimated rate well. The residual sum of squares (RSOS) and the value of the Akaike Information Criterion (AIC)[58] are given in Table 4[‡] and clearly indicate that model 1 is more appropriate than model 2.

*Step 7: Simultaneous Model Correction.* Because of the inevitably introduced bias when estimating the time derivatives of the concentration trajectories, the parameter estimates can not be guaranteed to be statistically optimal. To obtain statistically sound parameter values, a simultaneous parameter correction has to be carried out. That is, a dynamic parameter estimation problem is solved using the original concentration measurements and the parameter estimates obtained in step 5 as initial guess. Figure 9 exemplarily shows the model prediction of component $A_1$ in the first experiment with the parameter values obtained after step 5 and those obtained after step 7, the simultaneous model correction. It can be seen, that both curves lie close to each other, which indicates a good quality of the parameter estimates obtained after step 5.

**Table 4. Results of the Model Comparison**

| Model | RSOS | AIC |
|---|---|---|
| 1 | 0.186 e −2 | −6.24 e 3 |
| 2 | 0.435 e 5 | 2.24 e 3 |

## Results and Discussion

As mentioned above, the kinetic models for enzyme-catalyzed reactions tend to be strongly nonlinear and nonconvex and thus good initial guesses for the parameter values are needed but rarely available. To test the robustness of the proposed approach, we solve the parameter estimation problem using a number of different initial guesses according to Table 5 and four different measurement qualities according to Table 6, where $N_{\mathrm{DP}}$ is the number of data points available for each component in each experiment. For each parameter estimation problem, we assume that the same measurement data stemming from the five experiments as given in Table 2 are available.

In case of the multi-step approach, we solve the problem using a MATLAB based in-house software for steps 1–6 and gPROMS (version 3.0.3) for the last step, the simultaneous model correction. For the classical simultaneous approach, we use gPROMS (version 3.0.3). We consider a problem solved correctly, if all final parameter estimates are within a ±10% interval of the true parameter value. We consider a problem solved incorrectly, if the algorithm converges to a final estimate that violates the ±10% limit for at least one parameter value. We apply this relatively high acceptable error level to assure that only estimates corresponding to a wrong local minimum are marked as incorrect solutions. The final error of the estimates for the correctly solved problems

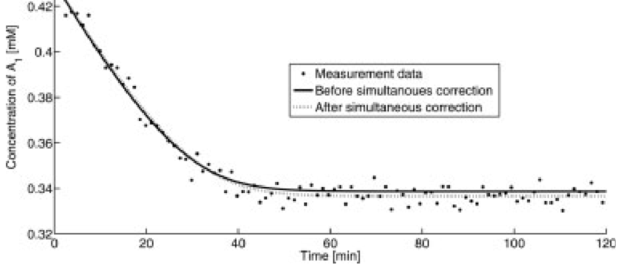

**Figure 9. Model fit for component $A_1$ in experiment No. 1 before and after the simultaneous model correction.**

---

[†] All experiments have been used to estimate the parameters, but only the fit of the first experiment is shown.

[‡] All five experiments have been used to fit the models and to calculate the values in Table 4.

Table 5. True Values, Low, and High Initial Guesses for the Six Parameters that have to be Estimated

| Parameter | Low initial guess | True value | High initial guess | Unit |
|---|---|---|---|---|
| $k_{\text{cat f}}$ | 15 | 45 | 135 | $S^{-1}$ |
| $K_{eq}$ | 1.33 | 4 | 12 | $mM^{-1}$ |
| $K_{mA}$ | 0.07 | 0.2 | 0.6 | mM |
| $K_{mB}$ | 0.2 | 0.6 | 1.8 | mM |
| $K_{mP}$ | 0.33 | 1 | 3 | mM |
| $K_{iA}$ | 0.33 | 1 | 3 | mM |

Table 6. Settings for the Different Simulated Measurement Qualities

| Setting | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $N_{\text{DP}}$ | 100 | 100 | 25 | 25 |
| Noise Level | 3% | 10% | 3% | 10% |

Table 7. Comparison of the Simultaneous and the Multi-Step Approach

| | Simultaneous method | Incremental method |
|---|---|---|
| Setting 1–100 data points and 3% relative noise | | |
| # Problems solved correctly | 30 | 37 |
| # Problems solved incorrectly | 24 | 23 |
| # Problems not solved | 10 | 4 |
| Setting 2–100 data points and 10% relative noise | | |
| # Problems solved correctly | 29 | 35 |
| # Problems solved incorrectly | 25 | 24 |
| # Problems not solved | 10 | 5 |
| Setting 3–25 data points and 3% relative noise | | |
| # Problems solved correctly | 30 | 34 |
| # Problems solved incorrectly | 24 | 27 |
| # Problems not solved | 10 | 3 |
| Setting 4–25 data points and 10% relative noise | | |
| # Problems solved correctly | 29 | 32 |
| # Problems solved incorrectly | 26 | 28 |
| # Problems not solved | 9 | 4 |

is generally significantly lower than ±10%. Finally, we consider a problem not solved, if the algorithm does not converge to a solution within 1 h CPU-time or 500 iterations. The results for both approaches under the aforementioned conditions are given in Table 7.

The mean computation time for all correctly solved problems was 73 sec for the multi-step approach and 108 sec for the simultaneous approach. Because some plots for the measurement quality setting 1—the better measurement quality—have already been shown, we show the same plots here for measurement quality 4—the worse measurement quality—to give a visual impression of the quality of the estimated concentration profiles, component flows and mass transfer/reaction rates in this setting. Figures 10–12 exemplarily show the original measurement values and the smooth trajectories as well as the estimated and the exact component flows for the measured components in the first experiment and the measurement quality of setting 4. Figure 13 shows the estimated and exact reaction rate and the reaction rate simulated with the final parameter estimates.

It can be seen that the quality of the estimated component flows is poor in some cases. The quality of the estimated rate, which results as a linear combination of the estimated component flows, however, is acceptable. In addition, the results show that the method works reliable for high and low quality measurements, although, as for any identification procedure, high quality measurements are preferable. Concerning the computation times, the incremental method performs slightly better than the simultaneous method. This is reasonable, because the repeated integration of the DAE model or the alternative solution of the large scale NLP problem, which is at the core of the simultaneous method, is avoided. The multi-step approach also requires the repeated integration of the DAE model in the last step, the simultaneous model correction, but because initial parameter guesses close to the optimal values are already available the number of iterations and thus the computation time is usually low. It is worth mentioning, that only one reaction is considered in this example and thus the aforementioned combinatorial explosion of model candidates does not occur for the simultaneous method. In case of multiple reactions with different model candidates, the difference in computation times will increase significantly as shown by Brendel et al.[10]

It is also reasonable, that the multi-step approach is more robust. In dynamic parameter estimation problems many numerical pitfalls, as discussed in Chapter 4.2 of the book by

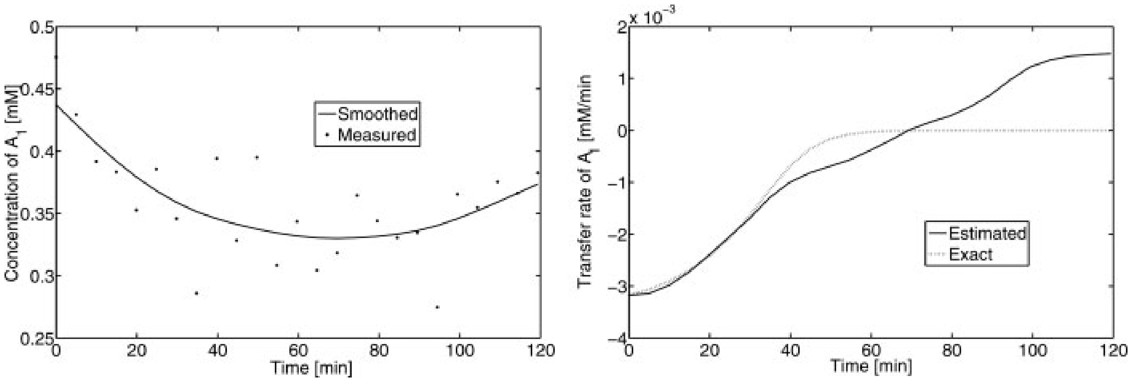

**Figure 10. Measured and smoothed concentration (left) and exact and estimated component flow (right) of $A_1$ in the first experiment for low quality measurements.**

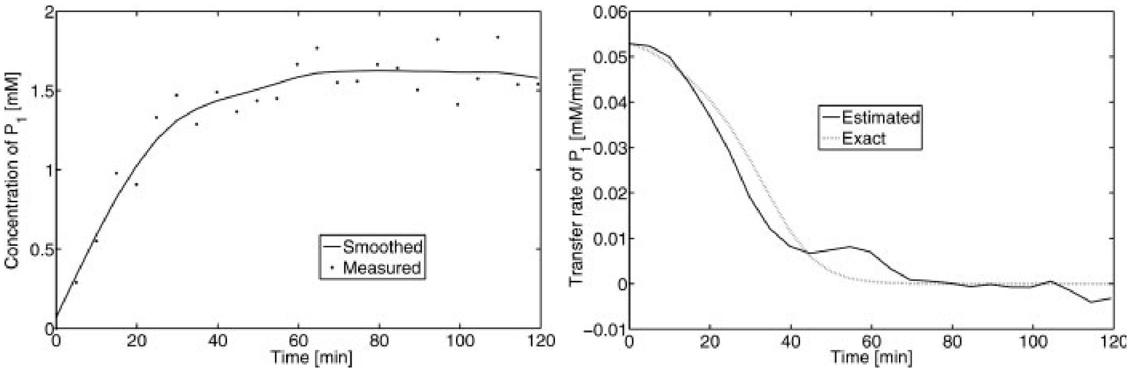

Figure 11. Measured and smoothed concentration (left) and exact and estimated component flow (right) of $P_1$ in the first experiment for low quality measurements.

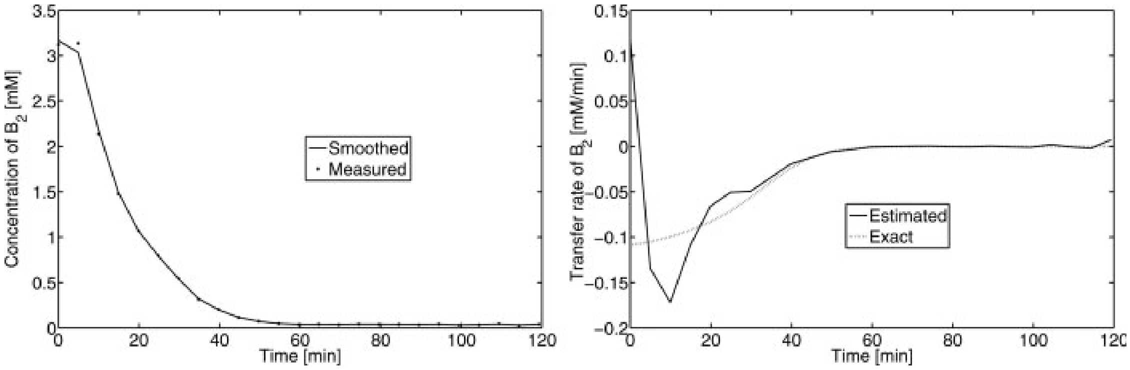

Figure 12. Measured and smoothed concentration (left) and exact and estimated component flow (right) of $B_2$ in the first experiment for low quality measurements.

Schittkowski,[27] exist. These pitfalls are typically more likely to occur if bad initial parameter guesses are provided, because the integration of the DAE system might give unforeseen results (states hitting bounds etc.), if parameter values far off the true ones are assumed. This problem can also be seen in Table 7, where the number of not solved problems is significantly higher for the simultaneous than for the incremental method in any setting considered. However, it is hard to determine in detail why the multi-step approach solves the example considered here more robustly. The reformulation of the problem suggested by Schittkowski[69] proved to have a positive effect on the robustness. Furthermore, Ipopt is known to be a good NLP-solver and finally, the algebraic parameter estimation problem is less complex than the dynamic parameter estimation problem. It is also important to mention that computation times for each method depend on a number of settings as, for instance, the integration tolerances and the optimization tolerance. We use identical tolerances wherever possible, but due to the different nature of the optimization problems, choosing identical or at least comparable tolerances and settings is often tricky and sometimes even impossible.

In summary, the presented multi-step approach proved to be an efficient and robust alternative for the identification of multiphase-reaction systems. The method allows transparent model identification because of the full control on all individual modeling steps and decisions. The difference in computation time, compared to the simultaneous approach, will increase with the number of reactions that have to be identified, because the above described combinatorial explosion is efficiently suppressed.

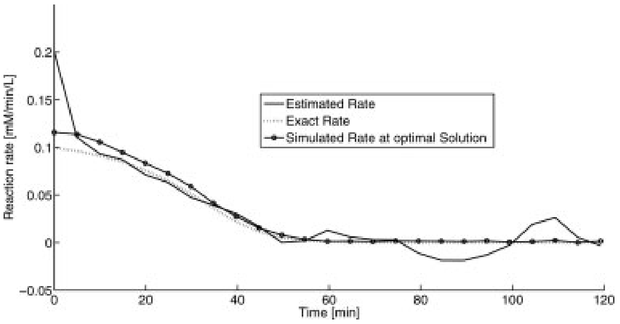

Figure 13. Estimated, simulated, and exact reaction rate for low quality measurements in the first experiment.

## Acknowledgments

The financial support by the *Deutsche Forschungsgemeinschaft* (DFG) via *collaborative research center 540* (SFB 540) is greatly appreciated.

We also would like to thank Michael Zavrel for fruitful discussion on enzyme catalyzed reactions and the reviewer for the thorough review of the original manuscript, which helped improving the quality of the article.

## Notation

$i$ = counting variable for components.
$j$ = counting variable for phases.
$k$ = counting variable for kinetic phenomena.
$n_\mathrm{c}$ = number of components.
$n_\mathrm{p}$ = number of phases.
$n_\mathrm{pc}$ = number of pseudo-components.
$n_\mathrm{pi}$ = number of phase interfaces.
$n_\mathrm{mt}$ = number of considered mass transfer kinetics.
$n_\mathrm{r}$ = number of reaction kinetics.
$n_\mathrm{k}$ = number of kinetic phenomena (reaction and mass transfer).
$n_{r,j}$ = number of reactions in phase $j$.
$f_{i,j}$ = component flow of component $i$ in phase $j$.
$r_k$ = rate of kinetic phenomenon $k$.
$\mathbf{r}_\mathrm{p}$ = row vector of all pseudo-reaction rates.
$\mathbf{r}_\mathrm{r}$ = row vector of all reaction rates.
$\mathbf{r}_\mathrm{mt}$ = row vector of all mass transfer rates.
$\mathbf{f}$ = row vector of all component flows.
$\mathfrak{S}_{\underline{\mathrm{c}}}$ = set of components.
$\mathfrak{S}_\mathrm{p}$ = set of phases.
$\mathfrak{S}_\mathrm{pc}$ = set of pseudo-components.
$\mathfrak{S}_k$ = set of kinetic phenomena.
$v_{i,k}$ = Stoichiometric coefficient of component $i$ in kinetic phenomenon $k$.
$V_j$ = volume of phase $j$.
$S_k$ = interfacial area/volume supporting the $k$'th kinetic phenomenon.
$V_k$ = volume supporting the $k$'th kinetic phenomenon.
$A_k$ = interfacial area supporting the $k$'th kinetic phenomenon.
$c_{i,j}$ = concentration of component $i$ in phase $j$.
$C_i$ = component $i$.
$C_{i,j}$ = component $i$ in phase $j$.
$F_{i,j}$ = feed of component $i$ to phase $j$.
$E_{i,j}$ = effluent of component $i$ from phase $j$.
$\mathbf{N}_j$ = stoichiometric matrix for reactions in phase $j$.
$\mathbf{N}_\mathrm{ar}$ = combined stoichiometric matrix for all reactions.
$\mathbf{N}_\mathrm{p}$ = combined pseudo-stoichiometric matrix of all reactions and all mass transfer kinetics.
$\mathbf{N}_\mathrm{m}$ = measurement matrix.
$\mathbf{N}_\mathrm{ep}$ = extended stoichiometric matrix.
$Ha$ = Hatta number.
$T$ = temperature.
$Ea$ = activation energy.
$R$ = gas constant.
$m_{i,j,j\ +1}$ = partition coefficient for component $i$ between phase $j$ and $j + 1$.
$D_i$ = mass transfer coefficient for component $i$.
$c_{i,j,\ \mathrm{equ}}$ = equilibrium concentration of component $i$ in phase $j$.
$K_{iA}$ = kinetic model parameter.
$K_{iB}$ = kinetic model parameter.
$K_{mA}$ = kinetic model parameter.
$K_{mB}$ = kinetic model parameter.
$K_{mP}$ = kinetic model parameter.
$K_\mathrm{eq}$ = kinetic model parameter.
$k_{\mathrm{cat\ f}}$ = kinetic model parameter.
$A$ = first binding substrate.
$B$ = second binding substrate.
$P$ = product.
$E$ = enzyme.
$N_i$ = amount of component $i$ initially present in the system.
$N_\mathrm{DP}$ = number of data points.
$T$ = transpose of a matrix.
† = pseudo inverse of a matrix.

## Literature Cited

1. Cornils B, Hermann WA, Vogt D, Horvath I, Olivier-Bourbignon H, Leitner W, Mecking S. *Multiphae Homogeneous Catalysis*. Weinheim, Germany: Wiley-VCH, 2005.
2. Zimmermann WB. Nonlinear model reconstruction by frequency and amplitude response for heterogeneous binary reaction in chemostat. *Chem Eng Sci*. 2006;61:605–620.
3. Berger RJ, Stitt EH, Marin GB, Kapteijn F, Moulijn JA. Eurokin—chemical reaction kinetics in practice. *Cattech*. 2001;5:30–60.
4. Hu B, Fishwick RP, Pacek AW, Winterbottom JM, Wood J, Stitt EH, Nienow AW. Simultaneous measurement of in situ bubble size and reaction rates with a heterogeneous catlytic hydrogeneation reaction. *Chem Eng Sci*. 2007;62:5392–5396.
5. Sun P, Grace JR, Lim CJ. Determination of intrinsic rate constants of the CaO-$CO_2$ reaction. *Chem Eng Sci*. 2008;63:5137–5148.
6. Hoorn JAA, Versteeg GF. Modelling of mass transfer in combination with radical reactions. *Chem Eng Sci*. 2006;61:5137–5148.
7. Steensma M, Kuipersa NJM, de Haana AB, Kwant G. Modelling and experimental evaluation of reaction kinetics in reactive extraction for chiral separation of amines, amino acids and amino-alcohols. *Chem Eng Sci*. 2007;62:1395–1407.
8. Ganapati DY, Neesha MD. Phase transfer catalysed alkylation of 2′-hydroxy acetophenone with 1-bromopentane: Kinetics and mechanism of liquid-liquid reaction. *J Mol Catal A: Chem*. 2006;243:278–285.
9. Saito T, Oshima K, Shimogaki Y, Egashira Y, Sugawara K, Takahiro K, Nagata S, Yamaguchi S, Komiyama H. Kinetic modeling of tungsten silicide chemical vapor deposition from $WF_6$ and $Si_2H_6$: Determination of the reaction scheme and the gas-phase reaction rates. *Chem Eng Sci*. 2007;62:6403–6411.
10. Brendel M, Bonvin D, Marquardt W. Incremental identification of complex reaction kinetics in homogeneous systems. *Chem Eng Sci*. 2006;61:5404–5420.
11. Bardow A, Marquardt W. Identification of diffusive transport buy means of an incremental approach. *Comput Chem Eng*. 2004;28: 585–595.
12. Michalik C, Schmidt T, Zavrel M, Ansorge-Schumacher M, Spiess A, Marquardt W. Application of the incremental identification method to the formate oxidation using formate dehydrogenase. *Chem Eng Sci*. 2007;62:5592–5597.
13. Levenspiel O. *Chemical Reaction Engineering*. New York: Wiley, 1972.
14. Quadros PA, Oliveira NMC, Baptista CMSG. Benzene nitration: validation of heterogeneous reaction models. *Chem Eng Sci*. 2004;59: 5449–5454.
15. Fogler HS. *Elements of Chemical Reaction Engineering*. New Jersey: Prentice Hall, 1992.
16. Benbelkacem H, Debellefontaine H. Modeling of a gas-liquid reactor in batch conditions. Study of the intermediate regime when part of the reaction occurs within the film and part within the bulk. *Chem Eng Process*. 2003;42:723–732.
17. Ashby M, Nagy P. Revisiting a proposed kinetic model for the reaction of cysteine and hydrogen peroxide via cysteine sulfenic acid. *Int J Chem Kinetics*. 2007;39:32–38.
18. Bordoloi A, Vinu A, Halligudi SB. Oxyfunctionalisation of adamantane using inorganicorganic hybrid materials based on isopoly and heteropoly anions: Kinetics next term and mechanistic studies. *Appl Catal A: Gen*. 2007;333:143–152.
19. Kadar EP, Darrington RT. Initial rate analysis of zoniporide hydrolysis degradants using high-performance liquid chromatography coupled with mass spectrometric detection. *Pharm Dev Technol*. 2007;12:247–257.
20. Hansen PC. *Rank-Deficient and Discrete Ill-Posed Problems*. Philadelphia: SIAM. 1998.
21. Zavrel M, Schmidt T, Michalik C, Ansorge-Schumacher M, Marquardt W, Büchs J, Spieß A. Mechanistic kinetic model for ThDP-dependent enzymes. *Biotechnol Bioeng*. 2008;101:27–38. Available at http://dx.doi.org/10.1002/bit.21867.
22. Duggleby RG. Quantitative analysis of the time courses of enzyme-catalyzed reactions. *Methods*. 2001;2:168–174.
23. Held B, Wright-Weber B, Grossman SH. Kinetic analysis of two purified forms of arginine kinase: Absence of cooperativity in substrate binding of dimeric phosphagen kinase. *Comp Biochem Physiol B*. 2007;148:6–13.
24. Nikitin D, Mokrishcheva M, Solonin A. 6His-Eco29kI methyltransferase methylation site and kinetic mechanism characterization. *Biochim Biophys Acta*. 2007;1774:1014–1019.
25. Gutiérrez OA, U, Danielson UH. Detection of competitive enzyme inhibition with end point progress curve data. *Anal Biochem*. 2006; 358:11–19.

26. Bard Y. *Nonlinear Parameter Estimation*. New York: Academic Press, 1974.
27. Schittkowski K. *Numerical Data Fitting in Dynamical Systems*. Dordrecht, The Netherlands: Kluwer Academic Publishers, 2002.
28. Walter E, Pronzato L. *Identification of Parametric Models from Experimental Data*. Berlin: Springer, 1997.
29. Bock HG. Numerical treatment of inverse problems in chemical reaction kinetics. In: Ebert K, Deuflhard P, Jäger W, editors. *Modelling of Chemical Reaction Systems*, Heidelberg, Germany: Springer, 1980:102–125.
30. Van Den Bosch B, Hellinckx L. A new method for the estimation of parameters in differential equations. *AIChE J*. 1974;20:250–255.
31. Tjoa IB, Biegler LT. Simultaneous solution and optimization strategies for parameter estimation of differential-algebraic equation systems. *Ind Eng Chem Res*. 1991;30:376–385.
32. Biegler LT, Cervantes AM, Wächter A. Advances in simultaneous strategies for dynamic process optimization. *Chem Eng Sci*. 2002;57: 575–593.
33. Chachuat B, Singer AB, Barton PI. Global solution of mixed-integer dynamic optimization problems. *Comput Aided Chem Eng*. 2005;20: 133–138.
34. Papamichail I, Adjiman CS. Global optimization of dynamic systems. *Comput Chem Eng*. 2004;28:403–415.
35. Singer AB, Taylor JW, Barton PI, Green WH. Global dynamic optimization for parameter estimation in chemical kinetics. *J Phys Chem A*. 2006;110:971–976.
36. Vasic-Racki D, Kragl U, Liese A. Benefits of enzyme kinetics modelling. *Chem Biochem Eng Q*. 2003;17:7–18.
37. Lundström L, Salmi T, Lehtonen J, Lindfors LP, Toppinen S, Aittamaa J. Modelling of specialty chemicals production in liquid-liquid reactors. A case study: synthesis of diols. *Chem Eng Sci*. 1999;54:1–18.
38. Franceschini G, Macchietto S. Validation of a model for biodiesel production through model-based experiment design. *Ind Eng Chem Res*. 2007;46:220–232.
39. Poyton A, Varziri M, McAuley K, McLellan P, Ramsay J. Parameter estimation in continuous-time dynamic models using principal differential analysis. *CCE*. 2006;30:698–708.
40. Ramsay J, Munhall K. Functional data analysis of lip motion. *J Acoust Soc Am*. 1996;99:3718–3727.
41. Reutter BW, Gullberg GT, Huesman RH. Accuracy and precision of compartmental model parameters obtained from directly estimated dynamic SPECT time-activity curves. *IEEE Trans Nucl Sci*. 2004; 51:170–176.
42. Franceschini G, Macchietto S. Model-based design of experiments for parameter precision: state of the art. *Chem Eng Sci*. 2008;Doi: 10.1016/j.ces. 2007.11.034.
43. Hadamard J. *Lectures on Cauchy Problem in Linear Partial Differential Equations*. London: Oxford University Press, 1923.
44. Tikhonov AN. Solution of incorrectly formulated problems and the regularization method. *Soviet Math Dokl*. 1963;4:1035–1038.
45. Kalman RE. A new approach to linear filtering and prediction problems recursive solution to the discrete-data linear filtering. *T-ASME*. 1960;35–45.
46. Mhamdi A. *State and Input Estimation Based On The Theory of Inverse Problems. Dissertation*. Düsseldorf: RWTH Aachen, 2004.
47. Reinsch CH. Smoothing by spline functions. *Numer Math*. 1967;10: 177–183.
48. Engl HW, Hanke M, Neubauer A. *Regularization of Inverse Problems*. Dordrecht: Kluwer, 1996.
49. Taylor R, Krishna R. *Multicomponent Mass Transfer*. New York: Wiley, 1993.
50. Levenberg K. A method for the solution of certain problems in least squares. *Q Appl Math*. 1944;2:164–168.
51. Marquardt D. An algorithm for least-squares estimation of nonlinear parameters. *SIAM J Appl Math*. 1963;11:431–441.
52. The MathWorks I. *Optimization Toolbox 3.0—User's guide*. Natick: The MathWorks, 2007.
53. Holmström K, Göran AO, Edvall MM. *User's Guide for Tomlab Optimization*. San Diego: Tomlab, 2007.
54. Schittkowski K. *DFNLP: A Fortran Implementation of an SQP-Gauss-Newton Algorithm—User's Guide, Version 2.0. 2005*.
55. Tawarmalani M, Sahinidis NV. Global optimization of mixed-integer nonlinear programs: A theoretical and computational study. *Math Program*. 2004;99:563–591.
56. Esposito WR, Floudas CA. Global optimization in parameter estimation of nonlinear algebraic models via the error-in-variables approach. *Ind Eng Chem Res*. 1998;37:1841–1858.
57. Esposito WR, Floudas CA. Parameter estimation in nonlinear algebraic models via global optimization. *Comput Chem Eng*. 1998;22: 213–220.
58. Akaike H. A new look at statistical model identification. *Automat Control*. 1974;19:716–722.
59. Stewart WE, Henson TL, Box GEP. Model discrimination and criticism with single-response data. *AIChE J*. 1996;42:3055–3062.
60. Elfron B. Bootstrap methods: another look at the jackknife. *Ann Stat*. 1979;1:1–26.
61. Marquardt W. Model-based experimental analysis of kinetic phenomena in multi-phase reactive systems. *Trans IChemE*. 2005;83: 561–573.
62. Groger H, Hummel W, Rollmann C, Chamouleau F, Husken H, Werner H, Wunderlich C, Abokitse K, Drauz K, Buchholz S. Preparative asymmetric reduction of ketones in a biphasic medium with an (S)-alcohol dehydrogenase under in situ-cofactor-recycling with a formate dehydrogenase. *Tetrahedron*. 2004;60:633–640.
63. Eckstein MF, Peters M, Lembrecht J, Spiess AC, Greiner L. Maximise equilibrium conversion in biphasic catalysed reactions: mathematical description and practical guideline. *Adv Synth Catal*. 2006; 348:1591–1596.
64. Diender MB, Straathof AJJ, van der Does T, Ras C, Heijnen JJ. Equilibrium modeling of extractive enzymatic hydrolysis of penicillin G with concomitant 6-aminopenicillanic acid crystallization. *Biotechnol Bioeng*. 2002;78:395–402.
65. Rosche B, Breuer M, Hauer B, Rogers PL. Biphasic aqueous/organic biotransformation of acetaldehyde and benzaidehyde by *Zymomonas mobilis* pyruvate decarboxylase. *Biotechnol Bioeng*. 2004;86:788–794.
66. de Boor C. *Spline Toolbox 3.0—User's Guide*. Natick: The MathWorks, 2007.
67. Hansen PC. Regularization tools version 4.0 for Matlab 7.3. *Numer Algorithms*. 2007;46:189–194.
68. Wächter A, Biegler LT. On the Implementation of a primal-dual interior point filter line search algorithm for large-scale nonlinear programming. *Math Program*. 2006;106:25–57.
69. Schittkowski K. Solving nonlinear least squares problems by a general purpose SQP-method. In: Hoffmann K-H, Hiriart-Urruty J-B, Lemarechal C, Zowe J, editors. *Trends in Mathematical Optimization*. Klosterberg: Birkhäuser, 1988:295–309.

*Manuscript received Dec. 22, 2007, and revision received Sept. 16, 2008.*