A Kernel-Based Method to Determine Optimal Sampling Times for the Simultaneous Estimation of the Parameters of Rival Mathematical Models

BRECHT M. R. DONCKELS,[1,2] DIRK J. W. DE PAUW,[2] PETER A. VANROLLEGHEM,[3] BERNARD DE BAETS[2]

[1] *BIOMATH, Department of Applied Mathematics, Biometrics and Process Control, Ghent University, Coupure links 653, B-9000 Ghent, Belgium*

[2] *KERMIT, Department of Applied Mathematics, Biometrics and Process Control, Ghent University, Coupure links 653, B-9000 Ghent, Belgium*

[3] *modelEAU, Département de Génie Civil, Pavillon Pouliot, Université Laval, Québec, QC, Canada, G1K 7P4*

Received 9 August 2008; Revised 10 October 2008; Accepted 21 October 2008

DOI 10.1002/jcc.21171

Published online 22 January 2009 in Wiley InterScience (www.interscience.wiley.com).

Abstract: When several models are proposed for one and the same process, experimental design techniques are available to design optimal discriminatory experiments. However, because the experimental design techniques are model-based, it is important that the required model predictions are not too uncertain. This uncertainty is determined by the quality of the already available data, since low-quality data will result in poorly estimated parameters, which on their turn result in uncertain model predictions. Therefore, model discrimination may become more efficient and effective if this uncertainty is reduced first. This can be achieved by performing dedicated experiments, designed to increase the accuracy of the parameter estimates. However, performing such an additional experiment for each rival model may undermine the overall goal of optimal experimental design, which is to minimize the experimental effort. In this article, a kernel-based method is presented to determine optimal sampling times to simultaneously estimate the parameters of rival models in a single experiment. The method is applied in a case study where nine rival models are defined to describe the kinetics of an enzymatic reaction (glucokinase). The results clearly show that the presented method performs well, and that a compromise experiment is found which is sufficiently informative to improve the overall accuracy of the parameters of all rival models, thus allowing subsequent design of an optimal discriminatory experiment.

 J Comput Chem 30: 2064–2077, 2009

Key words: dynamic modeling; mathematical models; optimal experimental design; parameter estimation; model discrimination; biokinetics; kernel functions; compromise experiment

Introduction

Many experimental studies are performed (1) to determine the model structure that adequately describes the process under study (often called model discrimination), or (2) to obtain (more) accurate estimates of the model parameters. For both problems, literature provides experimental design methods that help the experimenter to plan the experiments. For the problem of model discrimination, the methods described in refs. 1–5 can be used, while the ones described in refs. 6–9 can be used to design experiments that result in an increased accuracy of the parameter estimates. Common to these experimental design methods is the overall goal to maximize the information content of the designed experiments and thus to minimize the experimental effort needed.

The most intuitive approach to address the problems of model discrimination and accurate parameter estimation is to deal with them successively.[9,10] First, experiments are designed and performed to choose between the rival model structures, and then, once the most promising model structure has been selected, experiments are designed and performed to accurately estimate its parameters. Alternatively, one could deal with both problems simultaneously. For this purpose, a joint criterion has been described[10] where

Correspondence to: B. M. R. Donckels; e-mail: brecht.donckels@biomath.ugent.be

Contract/grant sponsor: Institute for the Promotion of Innovation by Science and Technology in Flanders (SBO Project); contract/grant number: 040125 (MEMORE)

the basic design strategy is to emphasize model discrimination when there is considerable doubt as to which model is best, and gradually shift the emphasis to parameter estimation as experimentation progresses, and model discrimination becomes possible. Both approaches thus deal with model discrimination first, and then the focus is (gradually) shifted to parameter estimation.

From literature, however, it is known that the uncertainty on the model predictions is of crucial importance for model discrimination. This is because the experimental design methods are model-based, and high model prediction uncertainties obviously hamper the efficacy and efficiency of the model discrimination procedure.[1–3,11–15] These model prediction uncertainties are determined by the quality of the available data, since low-quality data will result in poorly estimated parameters, which on their turn result in uncertain model predictions. The discrimination among several rival models may thus become more efficient and effective if this uncertainty could be reduced prior to the start of the model discrimination procedure.

Reducing the uncertainty on the model predictions can be achieved by designing and performing experiments dedicated to reducing the uncertainty on the parameter estimates. However, performing an additional experiment for each rival model may undermine the overall goal of optimal experimental design, since this would require at least as many experiments as the number of rival models. Therefore, this article presents a method to design a compromise experiment, which is not optimal for one or more of the individual rival models but is sufficiently informative to improve the overall accuracy of the parameters of all rival models.

This article is organized as follows. In the section "Methods," the theory on parameter estimation and optimal experimental design for parameter estimation is explained, as well as a method to design the compromise experiment. This kernel-based method is illustrated on a case study in the section "Results and Discussion," where a number of models is proposed to describe the kinetics of an enzyme, and where a compromise experiment is designed and evaluated. The conclusions drawn from these results are listed in the last section.

Methods

Mathematical Model Representation

In what follows, general deterministic models in the form of a set of (possibly mixed) differential and algebraic equations are considered, using the following notations:

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}(\boldsymbol{x}(t), \boldsymbol{u}(t), \boldsymbol{\theta}, t); \quad \boldsymbol{x}(t_0) = \boldsymbol{x}_0, \tag{1}$$

$$\hat{\boldsymbol{y}}(t) = \boldsymbol{g}(\boldsymbol{x}(t)), \tag{2}$$

where $\boldsymbol{x}(t)$ is an n_s-dimensional vector of time-dependent state variables, $\boldsymbol{u}(t)$ is an n_u-dimensional vector of time-varying inputs to the process, $\boldsymbol{\theta}$ is an n_p-dimensional vector of model parameters taken from a continuous, realizable set $\boldsymbol{\Theta}$, and $\hat{\boldsymbol{y}}(t)$ is an n_m-dimensional vector of measured response variables that are function of the state variables, $\boldsymbol{x}(t)$. An experiment will be denoted as $\boldsymbol{\xi}$ and is determined by the experimental degrees of freedom, such as sampling times, initial conditions, and time-varying or constant process inputs.

Parameter Estimation

The values of the model parameters, which by definition do not change during the course of the simulation, have to be determined from experimental data through parameter estimation. It consists of minimizing the weighted sum of squared errors (WSSE) functional through an optimal choice of the parameters $\boldsymbol{\theta}$. This can be written as

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \mathrm{WSSE}(\boldsymbol{\theta}), \tag{3}$$

where $\mathrm{WSSE}(\boldsymbol{\theta})$ is calculated as

$$\mathrm{WSSE}(\boldsymbol{\theta}) = \sum_{k=1}^{n_\mathrm{e}} \sum_{l=1}^{n_{\mathrm{sp}_k}} \Delta\hat{\boldsymbol{y}}(\boldsymbol{\xi}_k, \boldsymbol{\theta}, t_l)' \cdot \mathbf{Q} \cdot \Delta\hat{\boldsymbol{y}}(\boldsymbol{\xi}_k, \boldsymbol{\theta}, t_l), \tag{4}$$

and

$$\Delta\hat{\boldsymbol{y}}(\boldsymbol{\xi}_k, \boldsymbol{\theta}, t_l) = \boldsymbol{y}(\boldsymbol{\xi}_k, t_l) - \hat{\boldsymbol{y}}(\boldsymbol{\xi}_k, \boldsymbol{\theta}, t_l) \tag{5}$$

represents the difference between the vector of the n_m measured response variables and the model predictions at time t_l ($l = 1, \ldots, n_{\mathrm{sp}_k}$) of experiment $\boldsymbol{\xi}_k$ ($k = 1, \ldots, n_\mathrm{e}$). Furthermore, n_e represents the number of experiments from which data are used to estimate the model parameters, n_{sp_k} represents the number of sampling times in experiment $\boldsymbol{\xi}_k$, which are assumed to be the same for all measured state variables, and $\mathbf{Q}$ is an n_m-dimensional square matrix of user-supplied weighing coefficients. Typically, $\mathbf{Q}$ is chosen as the inverse of the measurement error covariance matrix $\boldsymbol{\Sigma}$ to incorporate the measurement uncertainty in the WSSE.[8,16,17]

Optimal Experimental Design for Parameter Estimation

In this section, the methodology used to design experiments to obtain more accurate parameter estimates, often called optimal experimental design for parameter estimation (OED/PE), is briefly described.

Fisher Information Matrix

As stated earlier, the accuracy of the parameter estimates highly depends on the quality or the information content of the experimental data from which they are determined. The information content of n_e experiments, $\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{n_\mathrm{e}}$, with regard to the model parameters is represented by the so-called Fisher information matrix (**FIM**),[7–9,18] which is calculated as

$$\mathbf{FIM}(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{n_\mathrm{e}}, \hat{\boldsymbol{\theta}}) = \sum_{k=1}^{n_\mathrm{e}} \mathbf{FIM}(\boldsymbol{\xi}_k, \hat{\boldsymbol{\theta}}), \tag{6}$$

where $\mathbf{FIM}(\boldsymbol{\xi}_k, \hat{\boldsymbol{\theta}})$ is calculated as

$$\sum_{l=1}^{n_{\mathrm{sp}_k}} \left(\frac{\partial\hat{\boldsymbol{y}}}{\partial\boldsymbol{\theta}}(\boldsymbol{\xi}_k, \boldsymbol{\theta}, t_l)\bigg|_{\hat{\boldsymbol{\theta}}} \right)' \cdot \boldsymbol{\Sigma}(\boldsymbol{\xi}_k, t_l)^{-1} \cdot \left(\frac{\partial\hat{\boldsymbol{y}}}{\partial\boldsymbol{\theta}}(\boldsymbol{\xi}_k, \boldsymbol{\theta}, t_l)\bigg|_{\hat{\boldsymbol{\theta}}} \right). \tag{7}$$

A closer look at eq. (7) shows that the **FIM** is composed of two components, the parameter sensitivities ($\partial\hat{\mathbf{y}}/\partial\boldsymbol{\theta}$) and the measurement error covariance matrix ($\boldsymbol{\Sigma}$). The parameter sensitivity with respect to a certain state variable expresses how much that state variable will change when a parameter is slightly perturbed. A state variable that is highly sensitive to a certain parameter will therefore contain a lot of information about this parameter, while a variable that is insensitive to the parameter does not contribute to the information content for that parameter. The role of the measurement error covariance matrix in the calculation of the **FIM** is rather straightforward, since it is obvious that a measurement associated with a large measurement error will contribute less to the information content than a measurement with a small measurement error.

Central Rationale Behind Optimal Experimental Design for Parameter Estimation

In general, optimal experimental design is an optimization problem, where the optimum of a well-defined objective function is sought by varying the experimental degrees of freedom. The experimental degrees of freedom, $\boldsymbol{\xi}$, are restricted by a number of constraints that define a set of possible experiments, denoted as $\boldsymbol{\Xi}$. These constraints are determined by the experimental setup and are specified before the start of the experimental design exercise.

The Fisher information matrix described in the previous section, expresses the information content of the n_e experiments with regard to the model parameters, and its maximization is the central rationale behind optimal experimental design for parameter estimation.[7–9,18] The $(n_e + 1)$th experiment, denoted as $\boldsymbol{\xi}^{\star}_{n_e+1}$, is obtained as

$$\boldsymbol{\xi}^{\star}_{n_e+1} = \arg\max_{\boldsymbol{\xi}\in\boldsymbol{\Xi}} \mathbf{FIM}(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{n_e+1}, \hat{\boldsymbol{\theta}}_{n_e}), \tag{8}$$

with

$$\mathbf{FIM}(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{n_e+1}, \hat{\boldsymbol{\theta}}_{n_e}) = \sum_{k=1}^{n_e} \mathbf{FIM}(\boldsymbol{\xi}_k, \hat{\boldsymbol{\theta}}_{n_e}) + \mathbf{FIM}(\boldsymbol{\xi}_{n_e+1}, \hat{\boldsymbol{\theta}}_{n_e}). \tag{9}$$

The information content of the proposed $(n_e + 1)$th experiment, which is represented by $\mathbf{FIM}(\boldsymbol{\xi}_{n_e+1}, \hat{\boldsymbol{\theta}}_{n_e})$, is thus maximized, given the information content of the already performed experiments ($\mathbf{FIM}(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{n_e}, \hat{\boldsymbol{\theta}}_{n_e})$) and the parameter values derived from these experiments ($\hat{\boldsymbol{\theta}}_{n_e}$). For simplicity, $\mathbf{FIM}(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{n_e+1}, \hat{\boldsymbol{\theta}}_{n_e})$ will be denoted as **FIM** in the following.

Experimental Design Criteria Based on the FIM

Since the **FIM** is a matrix, it cannot be maximized as such. Therefore, several design criteria/objective functions have been proposed based on the **FIM**,[7,8,19,20] all of which exploit the inversely proportional relationship between the **FIM** and the parameter estimation error covariance matrix. This relationship is dictated by the Cramér-Rao inequality,[9,18] which states that under certain conditions (that is, uncorrelated white measurement noise), the inverse of the **FIM** gives the lower bound of the parameter estimation error covariance matrix. In this way, properties of the **FIM** determine the size, shape, and orientation of the confidence region of the parameter estimates and thus their accuracy.

In this article, only the so-called D-optimality and modE-optimality design criteria will be discussed and applied. They are briefly discussed below.

D-Optimality Design Criterion: $\max_{\boldsymbol{\xi}\in\boldsymbol{\Xi}} \det(\mathbf{FIM})$. Here, the idea is to maximize the determinant of the **FIM**. The latter is inversely proportional to the volume of the confidence region of the parameter estimates, and this volume is thus minimized when maximizing $\det(\mathbf{FIM})$. In other words, one minimizes the geometric average of the variances of the parameter estimates. Moreover, D-optimal experiments possess the property of being invariant with respect to any rescaling of the parameters.[20,21]

modE-Optimality Design Criterion: $\min_{\boldsymbol{\xi}\in\boldsymbol{\Xi}} \frac{\lambda_{\max}(\mathbf{FIM})}{\lambda_{\min}(\mathbf{FIM})}$. With this criterion, the focus is on the minimization of the condition number, which is the ratio between the largest and the smallest eigenvalue. The minimum of this ratio is one, which corresponds to the case where the shape of the confidence ellipsoid is a (hyper)sphere.

The effect of these criteria on the confidence region is illustrated in Figure 1 for an estimation problem with two parameters (θ_1 and θ_2). The size, shape, and orientation of the confidence region, which is an ellipse in the case of two parameters, are determined by the eigenvalues and eigenvectors of the **FIM**. The largest axis of the confidence ellipse is inversely proportional to the square root of the smallest eigenvalue ($\lambda_{\min}$), while the smallest axis is inversely proportional to the square root of the largest eigenvalue ($\lambda_{\max}$).

Design of a Compromise Experiment

As stated in the "Introduction", this article investigates the possibility to design a compromise experiment, that is, an experiment which may not be optimal for each individual rival model, but sufficiently informative to improve the overall accuracy of the parameters of all rival models. In this section, a method is presented to determine such a compromise experiment. It is inspired by what is called kernel density estimation or the Parzen window approach.[23,24] After briefly explaining the theory of kernel density estimation (the next section), it will be explained how its rationale can be applied to design a compromise experiment (section "Kernel-Based Method for Experimental Design").

Introduction to Kernel Density Estimation

Kernel density estimation[23] is a nonparametric method for estimating the probability density function (pdf) of a random variable from its independent and identically distributed (i.i.d.) samples.[24,25] The term nonparametric refers to the fact that it is not necessary to assume a particular model for the pdf prior to the density estimation exercise.[25]

Suppose n samples of a random variable (x) are drawn i.i.d. according to the (unknown) probability density function $p(x)$. The kernel density estimation of this probability density function, denoted as $\hat{p}(x)$, is given by

$$\hat{p}(x) = \frac{1}{n \cdot h} \cdot \sum_{i=1}^{n} \kappa\left(\frac{x - x_i}{h}\right) \tag{10}$$

where x_i represents the ith sample, κ represents the so-called kernel function (or Parzen window), and h represents the smoothing

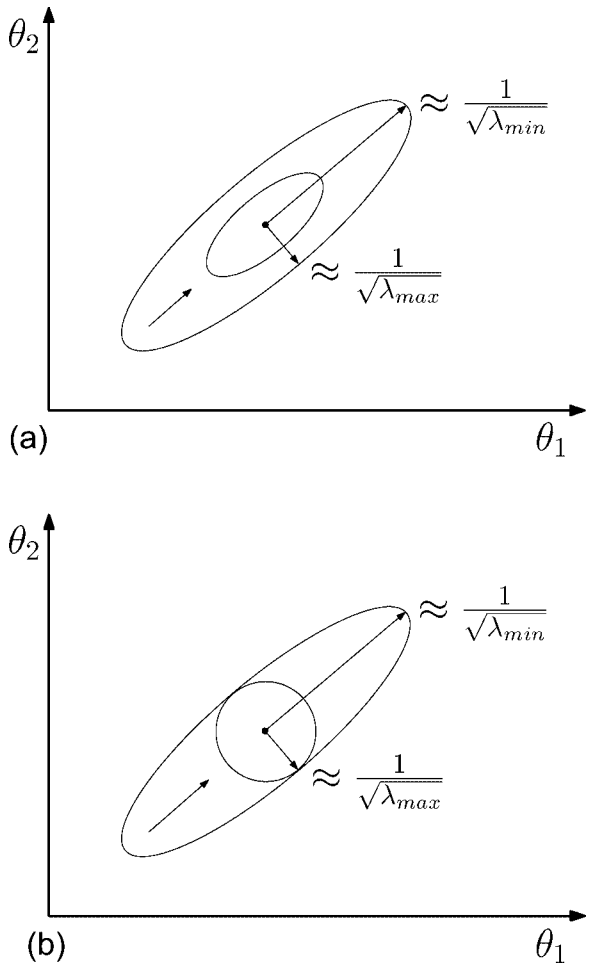

Figure 1. Illustration of (a) the D-optimality design criterion that causes the volume of the confidence region to decrease, and (b) the modified E-optimality design criterion that causes the shape of the confidence region to become as circular as possible. Reprinted from *Water Science and Technology*, 53(1), 117, with permission from the copyright holders, IWA.

parameter (or bandwidth parameter). Quite often, κ is taken to be a standard Gaussian function with zero mean and a variance equal to one, given by

$$\kappa(u) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\cdot u^2}. \tag{11}$$

Apart from this Gaussian kernel function, other kernel functions have been proposed, but the choice of the kernel function seems to be less important than the choice of the smoothing parameter.[24]

The principle of kernel density estimation and the importance of the smoothing parameter are illustrated in Figure 2. It is clear that when the smoothing parameter is too small, an erratic and noisy estimate of $p(x)$ will be found, while too large values will produce a very smooth and out-of-focus estimate of $p(x)$. Techniques are available to determine an optimal smoothing parameter in the context of kernel density estimation, but for this the reader is referred to refs. 24 and 26 and the citations therein. An approach to determine the smoothing parameter in the context of experimental design is presented in the next section.

Kernel-Based Method for Experimental Design

This section explains how the concepts of kernel density estimation described above can be useful in an experimental design context. Although this article focuses on its application for those cases where only the optimal sampling times are determined, the presented method could in principle be extended to applications where experimental degrees of freedom of all types (manipulations, initial conditions, and sampling times) are considered.

Suppose n_{sp} optimal sampling times were determined for each of the m rival models by optimizing one of the optimal design criteria described in the previous section. Then, similarly to what is done in kernel density estimation, a function is defined, given by

$$\hat{p}(t) = \sum_{i=1}^{m}\sum_{j=1}^{n_{sp}} \kappa\left(\frac{t - t_{ij}}{h}\right) \tag{12}$$

where t_{ij} represents the jth sampling time that was found to be optimal for model i, h represents the smoothing parameter, and κ represents the kernel function, for which a Gaussian-like function is chosen, given by

$$\kappa(u) = e^{-u^2}. \tag{13}$$

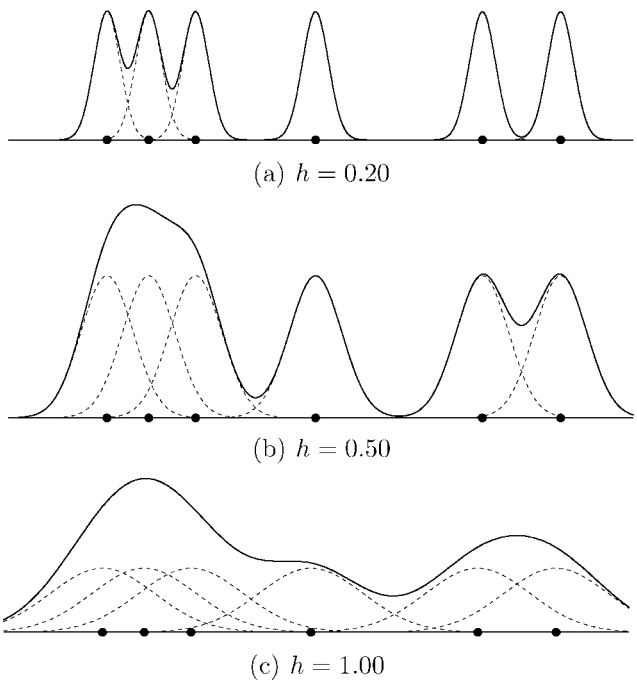

Figure 2. Kernel density estimation applied to an illustrative example for three different values of the smoothing parameter h. The dots represent six samples from an unknown probability density function, the dashed lines represent the individual kernel functions, and the full line represents the estimated probability density function that is calculated as the sum of the individual kernel functions.

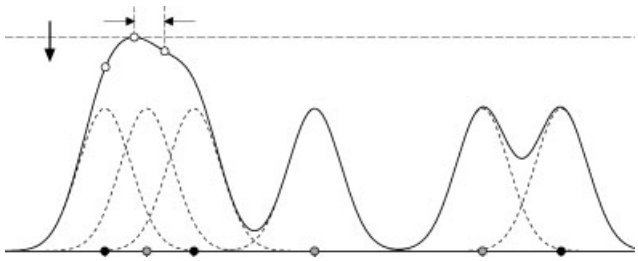

Figure 3. The black dots represent the optimal sampling times for model i, whereas the gray ones represents those for model j. The compromise sampling times correspond to those points that maximize $\hat{p}(t)$ (represented by the full line) under the constraint that a minimum time between two sampling times is required. Their location is indicated by the white dots.

The reason why the factors $1/(n \cdot h)$, $1/\sqrt{2\pi}$, and $1/2$ were omitted from eqs. (10) and (11) to form eqs. (12) and (13) is that in kernel density estimation $\hat{p}(x)$ represents an estimate of a probability density function for which $\int_{-\infty}^{+\infty} p(x)dx$ has to be equal to one. Since this is not required when applying this method for optimal experimental design purposes, these factors were omitted. Note that this does not influence the resulting experiment.

Now, to appoint $n_{sp,c}$ compromise sampling times from the $m \cdot n_{sp}$ optimal sampling times ($n_{sp,c} \leq m \cdot n_{sp}$), the following approach is adopted. The compromise sampling times are those that maximize $\hat{p}(t)$, under the constraint that a minimum time interval between two sampling times is required by the experimental setup. Basically, the compromise sampling times correspond to those points of $\hat{p}(t)$ that a horizontal line through $\max(\hat{p}(t))$ encounters while going down, taking into account the requirement of a minimum time interval. This is illustrated in Figure 3, where three compromise sampling times were determined. Note that, in principle, the presented approach allows the compromise sampling times to coincide.

Choice of the Smoothing Parameter

An interesting feature of the proposed method is the fact that neighboring samples intensify each other, because the associated kernel functions overlap. The extent to which this intensifying effect occurs not only depends on the time interval between the individual sampling times but also on the smoothing parameter h. An illustration of this can be seen in Figure 2, especially when comparing the three samples on the left for the three different values of h. The intensifying effect is hardly present for small values of h (as in Fig. 2a), while it does occur for higher values (Figs. 2b and 2c). The latter can be explained by the fact that higher values of the smoothing parameter result in broader kernel functions (dashed lines in Fig. 2), which overlap with those of other samples. For the sharp kernel functions obtained with the lower value of the smoothing parameter, no significant overlap occurs.

As stated before, methods are described in literature to determine the optimal smoothing parameter in the context of kernel density estimation. However, in this context, we prefer to link this parameter to the minimum time interval, denoted as t_{min}. Since t_{min} is dictated by the experimental setup, the choice of the smoothing parameter is straightforward, general and objective. More specifically, we suggest to define h as $0.25 \cdot t_{min}$.

This suggestion is based on a compromise. On the one hand, a high value for h is desired because the intensifying effect becomes more apparent (as explained above), but on the other hand, the value for h must not be too high. The latter is illustrated in Figure 4, where two sampling times are shown that are located as close to each other as allowed by the minimum time interval (the black dots). Since both sampling times are equally important, they should both be selected as compromise sampling times (indicated by the white dots). This is the case when a value of $0.25 \cdot t_{min}$ is chosen for the smoothing parameter (see Fig. 4a). Higher values of h result in compromise sampling times that are different from the optimal ones, whereas lower values of h will lessen the intensifying effect (see Fig. 4d).

Weighing of the Sampling Times

The contribution of the individual sampling times to the information content of the experiment varies, and some sampling times are thus more important/informative than others. With the presented method, it is possible to give weights to the n_{sp} individual sampling times, representing their importance. The higher their contribution to the

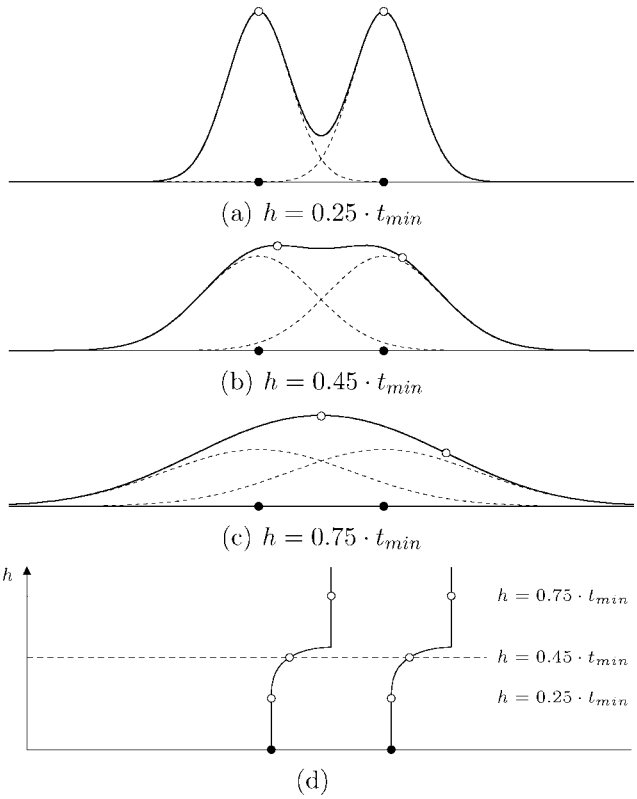

Figure 4. Kernel density estimation applied to an illustrative example for three different values of the smoothing parameter h. The time interval between the two sampling times (black dots) is equal to the minimum time interval (t_{min}), and the location of the compromise sampling times is indicated by the white dots. The lower graph (d) indicates where the compromise sampling times will be located for values of h between 0 and 1 times the minimum time interval.

information content of the experiment, the higher their weight. For this, eq. (12) can be easily modified into

$$\hat{p}(t) = \sum_{i=1}^{m} \sum_{j=1}^{n_{sp}} w_{ij} \cdot \kappa \left(\frac{t - t_{ij}}{h} \right), \tag{14}$$

where w_{ij} represents the weight of the jth sampling time of the optimal experiment for model i (t_{ij}).

The weight of an individual sampling time is defined relative to the extent to which a chosen optimality criterion (see the section "Experimental Design Criteria Based on the FIM") diminishes when that sampling time is removed from the set of optimal sampling times. In the case where the D-optimality design criterion is used (maximized), the weight can be determined as follows. Suppose $\boldsymbol{\xi}_i^\star$ represents the optimal experiment for model i, and $\boldsymbol{\xi}_{i\setminus t_j}$ represents the same experiment, but without sampling time t_j. Then, the weight of this sampling time can be determined as

$$w_{ij} = \frac{\mathrm{D}(\boldsymbol{\xi}_i^\star) - \mathrm{D}(\boldsymbol{\xi}_{i\setminus t_j})}{\mathrm{D}(\boldsymbol{\xi}_i^\star)}, \quad 0 \leq w_{ij} \leq 1, \tag{15}$$

where $\mathrm{D}(\boldsymbol{\xi})$ represents the D-optimality design criterion value associated with experiment $\boldsymbol{\xi}$. When sampling time t_j is not important with regard to the information content of the experiment, $\det(\mathbf{FIM}(\boldsymbol{\xi}_{i\setminus t_j}))$ will only be slightly smaller than $\det(\mathbf{FIM}(\boldsymbol{\xi}_i^\star))$ and w_{ij} will be close to zero. Besides, it would not be realistic that $\det(\mathbf{FIM}(\boldsymbol{\xi}_{i\setminus t_j}))$ is larger than $\det(\mathbf{FIM}(\boldsymbol{\xi}_i^\star))$, since information is lost when a sampling time is discarded. Taking into account that the eigenvalues and thus the determinant of the **FIM** cannot take negative values (the **FIM** is a positive definite matrix[9]), the value of w_{ij} cannot be negative.

Contrary to the D-optimality design criterion, the modE-optimality design criterion has to be minimized. A similar equation for the calculation of the weights can be formulated as follows:

$$w_{ij} = \max \left(0, \frac{\mathrm{modE}\left(\boldsymbol{\xi}_{i\setminus t_j}\right) - \mathrm{modE}\left(\boldsymbol{\xi}_i^\star\right)}{\mathrm{modE}\left(\boldsymbol{\xi}_{i\setminus t_j}\right)} \right), \quad 0 \leq w_{ij} \leq 1, \tag{16}$$

where $\mathrm{modE}(\boldsymbol{\xi})$ represents the modE-optimality design criterion value associated with experiment $\boldsymbol{\xi}$. When sampling time t_j does not have a significant influence on the value of the modE-optimality design criterion, the weight will be close to zero. In this case, however, it cannot be guaranteed that discarding sampling time leads to a worse (larger) value of the modE-optimality design criterion. Therefore, w_{ij} can in principle take negative values. For the case study described in this article, this situation did not occur, but if it occurs one can set the weight of the corresponding sampling time to zero.

Evaluating the Designed Compromise Experiment

To evaluate the presented method on its capability to design a compromise experiment, the following approach was adopted. Since an optimal experiment was designed for each model, each of these experiments could have been performed instead of the compromise experiment. The information that is lost, or gained when doing so, is used for the evaluation.

Because the information content of an experiment is reflected by the value of the design criterion, the basis of the evaluation lies in the comparison of these criterion values. In this respect, it is important to realize that the information content or the quality of an experiment with regard to the parameters of a particular model can be compared with that of another experiment, but it is not meaningful to compare design criterion values from different models.

When the D-optimality criterion is used, the criterion values are calculated for each of the D-optimal experiments ($\boldsymbol{\xi}_j^\star$, with $j = 1, \ldots, m$), and these are compared with the criterion value associated with the compromise experiment ($\boldsymbol{\xi}_c$). The ratio between these criterion values, denoted as $\Gamma_{\mathrm{D}_{ij}}$, is eventually used for the evaluation and is calculated as

$$\Gamma_{\mathrm{D}_{ij}} = \frac{\mathrm{D}(m_i, \boldsymbol{\xi}_c)}{\mathrm{D}\left(m_i, \boldsymbol{\xi}_j^\star\right)}, \tag{17}$$

where $\mathrm{D}(m_i, \boldsymbol{\xi})$ represents the D-optimality criterion value for model m_i associatated with experiment $\boldsymbol{\xi}$.

Since a higher information content is represented by a higher value of the D-optimality design criterion, it holds that $\Gamma_{\mathrm{D}_{ij}} > 1$ when the compromise experiment contains more information with regard to the parameters of model m_i than the optimal experiment for model m_j ($\boldsymbol{\xi}_j^\star$). In other words, when $\Gamma_{\mathrm{D}_{ij}} > 1$, the estimates of the parameters of model m_i should be more accurate when the compromise experiment is performed instead of experiment $\boldsymbol{\xi}_j^\star$.

For the modE-optimality design criterion, which has to be minimized, smaller criterion values are associated with better experiments. The expression used to quantify the designed compromise experiment is therefore given by

$$\Gamma_{\mathrm{modE}_{ij}} = \frac{\mathrm{modE}\left(m_i, \boldsymbol{\xi}_j^\star\right)}{\mathrm{modE}(m_i, \boldsymbol{\xi}_c)}, \tag{18}$$

where $\mathrm{modE}(m_i, \boldsymbol{\xi})$ represents the value of the modE-optimality design criterion for experiment $\boldsymbol{\xi}$ with regard to the parameters of model m_i. Since a lower modE-optimality design criterion corresponds to a better experiment, $\Gamma_{\mathrm{modE}_{ij}} > 1$ when the compromise experiment ($\boldsymbol{\xi}_c$) is preferred to the optimal experiment for model j ($\boldsymbol{\xi}_j^\star$) with regard to the estimation of the parameters of model m_i.

Optimization Algorithms

Both parameter estimation and optimal experimental design are optimization problems. To find the optimum, the use of optimization algorithms is required. In this work, the SIMPSA optimization algorithm proposed by Cardoso et al.[27] was used. This algorithm combines the nonlinear simplex[28] and the simulated annealing algorithm.[29] For more information on these optimization algorithms, the reader is referred to the cited papers.

Results and Discussion

In this section, the experimental design concepts introduced in the previous section will be illustrated in a relatively simple case study,[3] where nine models are proposed to describe the kinetic behavior of the enzyme glucokinase (*glk*, EC: 2.7.1.2). This enzyme catalyzes the conversion of glucose (GLU) and ATP to glucose-6-phosphate (G6P) and ADP, which is the first reaction of the glycolysis pathway.

General Model

Before describing the different kinetics, a general model for the enzymatic conversion process is formulated. For this, it is assumed that the experimental setup allows one to give a pulse of glucose, ATP, and phosphoenolpyruvate (PEP), or a mixture thereof.

The volume of the reaction vessel [L], denoted as V, is determined by the flow rate of the pulse [L/s], denoted as F_p, and by the sampling volume and frequency. However, in this example, the sampling volume will be neglected and the volume can thus be described by

$$\frac{d\mathrm{V}}{dt} = F_\mathrm{p}. \tag{19}$$

For the concentration of glucokinase [mg/L], denoted as GLK, only a dilution effect is considered, and inactivation of the enzyme is thus neglected. Given the fact that a typical experiment ends after 20 min, this is a reasonable assumption. The resulting equation for describing the enzyme concentration is given as

$$\frac{d\mathrm{GLK}}{dt} = -\frac{F_\mathrm{p}}{\mathrm{V}} \cdot \mathrm{GLK}. \tag{20}$$

The equations used to describe the other state variables (all of which are expressed in mM) are given as:

$$\frac{d\mathrm{GLU}}{dt} = \frac{F_\mathrm{p}}{\mathrm{V}} \cdot (\mathrm{GLU_p} - \mathrm{GLU}) - v_\mathrm{glk}, \tag{21}$$

$$\frac{d\mathrm{ATP}}{dt} = \frac{F_\mathrm{p}}{\mathrm{V}} \cdot (\mathrm{ATP_p} - \mathrm{ATP}) - v_\mathrm{glk}, \tag{22}$$

$$\frac{d\mathrm{G6P}}{dt} = -\frac{F_\mathrm{p}}{\mathrm{V}} \cdot \mathrm{G6P} + v_\mathrm{glk}, \tag{23}$$

$$\frac{d\mathrm{ADP}}{dt} = -\frac{F_\mathrm{p}}{\mathrm{V}} \cdot \mathrm{ADP} + v_\mathrm{glk}, \tag{24}$$

$$\frac{d\mathrm{PEP}}{dt} = \frac{F_\mathrm{p}}{\mathrm{V}} \cdot (\mathrm{PEP_p} - \mathrm{PEP}). \tag{25}$$

Here, $\mathrm{GLU_p}$, $\mathrm{ATP_p}$, and $\mathrm{PEP_p}$ represent the concentrations [mM] of glucose, ATP and PEP in the pulse, respectively, and v_glk represents the velocity equation describing the kinetic behavior of glucokinase [mM/s].

Rival Models

The conversion catalyzed by glucokinase is a bi-reactant system.[30] Two reaction mechanisms are possible for such a system: random and ordered. In a random bi-reactant system, the order in which the two substrates bind does not matter, whereas in an ordered bi-reactant system one of the substrates has to bind to the enzyme first, before the second substrate can bind and the reaction can take place. In addition, it was recently suggested that glucokinase may be inhibited by PEP.[31]

Based on these considerations, nine models were defined to describe the enzyme kinetics.[30] The models differ in the equation used to describe the enzyme kinetics, each of which is based on a particular hypothesis of how the enzyme works. Although the kinetic equation is different for each rival model, each one is of the following form:

$$v_\mathrm{glk} = k \cdot \mathrm{GLK} \cdot \frac{\frac{\mathrm{GLU}}{K_\mathrm{GLU}} \cdot \frac{\mathrm{ATP}}{K_\mathrm{ATP}}}{\varphi(\mathrm{GLU}, \mathrm{ATP}, \mathrm{PEP})}, \tag{26}$$

where the parameter k expresses the maximum specific reaction rate [U/mg], where one unit is defined as that amount of enzyme that catalyzes one μmol of substrate in 1 min. The part that is different for each rival model is represented by $\varphi(\mathrm{GLU}, \mathrm{ATP}, \mathrm{PEP})$.

For models m_1, m_2, and m_3, it is assumed that the reaction mechanism is random. With regard to the inhibition by PEP, three scenarios are possible (also for the other models described further on): there is no inhibition by PEP [eq. (27)], PEP inhibits the binding of ATP [eq. (28)], and PEP inhibits the binding of glucose [eq. (29)]. This results in the following equations for $\varphi(\mathrm{GLU}, \mathrm{ATP}, \mathrm{PEP})$:

$$1 + \frac{\mathrm{GLU}}{K_\mathrm{GLU}} + \frac{\mathrm{ATP}}{K_\mathrm{ATP}} + \frac{\mathrm{GLU}}{K_\mathrm{GLU}} \cdot \frac{\mathrm{ATP}}{K_\mathrm{ATP}}, \tag{27}$$

$$1 + \frac{\mathrm{GLU}}{K_\mathrm{GLU}} + \frac{\mathrm{ATP}}{K_\mathrm{ATP}} + \frac{\mathrm{PEP}}{K_\mathrm{PEP}} + \frac{\mathrm{GLU}}{K_\mathrm{GLU}} \cdot \frac{\mathrm{PEP}}{K_\mathrm{PEP}} + \frac{\mathrm{GLU}}{K_\mathrm{GLU}} \cdot \frac{\mathrm{ATP}}{K_\mathrm{ATP}}, \tag{28}$$

$$1 + \frac{\mathrm{GLU}}{K_\mathrm{GLU}} + \frac{\mathrm{ATP}}{K_\mathrm{ATP}} + \frac{\mathrm{PEP}}{K_\mathrm{PEP}} + \frac{\mathrm{ATP}}{K_\mathrm{ATP}} \cdot \frac{\mathrm{PEP}}{K_\mathrm{PEP}} + \frac{\mathrm{GLU}}{K_\mathrm{GLU}} \cdot \frac{\mathrm{ATP}}{K_\mathrm{ATP}}. \tag{29}$$

For the other six models, an ordered reaction mechanism is assumed. For models m_4, m_5, and m_6, it is assumed that glucose is the first binding substrate, which results in the following equations for $\varphi(\mathrm{GLU}, \mathrm{ATP}, \mathrm{PEP})$:

$$1 + \frac{\mathrm{GLU}}{K_\mathrm{GLU}} + \frac{\mathrm{GLU}}{K_\mathrm{GLU}} \cdot \frac{\mathrm{ATP}}{K_\mathrm{ATP}}, \tag{30}$$

$$1 + \frac{\mathrm{GLU}}{K_\mathrm{GLU}} + \frac{\mathrm{GLU}}{K_\mathrm{GLU}} \cdot \frac{\mathrm{PEP}}{K_\mathrm{PEP}} + \frac{\mathrm{GLU}}{K_\mathrm{GLU}} \cdot \frac{\mathrm{ATP}}{K_\mathrm{ATP}}, \tag{31}$$

$$1 + \frac{\mathrm{GLU}}{K_\mathrm{GLU}} + \frac{\mathrm{ATP}}{K_\mathrm{ATP}} \cdot \frac{\mathrm{PEP}}{K_\mathrm{PEP}} + \frac{\mathrm{GLU}}{K_\mathrm{GLU}} \cdot \frac{\mathrm{ATP}}{K_\mathrm{ATP}}. \tag{32}$$

The equations associated with models m_7, m_8, and m_9 are similar, but ATP is assumed to be the first binding substrate. The equations

Table 1. Parameters of the Real Model ($m_5^\star$) That Were Used to Generate Experimental Data, and the Parameter Estimates Obtained After Fitting the Rival Models to the Data From the Preliminary Experiment, Together With the 95% Confidence Intervals and the Corresponding WSSE-Values.

Model	k	K_{GLU}	K_{ATP}	K_{PEP}	WSSE
$m_5^\star$	312	0.15	0.13	0.10	–
m_1	314.13 ± 90.48	0.0173 ± 0.1135	0.1407 ± 0.0694	–	61.5287
m_2	336.14 ± 107.66	0.0451 ± 0.1341	0.1533 ± 0.0772	0.1466 ± 0.2198	57.1080
m_3	317.21 ± 93.38	0.0191 ± 0.1162	0.1412 ± 0.0705	0.0091 ± 0.0544	56.9125
m_4	307.64 ± 49.17	0.1299 ± 0.8481	0.1245 ± 0.0441	–	61.2821
m_5	312.41 ± 51.28	0.2011 ± 0.9320	0.1207 ± 0.0461	0.1261 ± 0.2145	56.9285
m_6	319.87 ± 55.23	0.3112 ± 1.0616	0.1182 ± 0.0491	0.1076 ± 0.3577	57.1491
m_7	412.58 ± 180.30	0.0099 ± 0.1706	27.9603 ± 464.53	–	94.2223
m_8	428.11 ± 236.59	0.0148 ± 0.2146	19.4047 ± 265.68	8.7458 ± 127.33	77.5805
m_9	543.60 ± 438.34	0.1102 ± 0.3893	3.6812 ± 9.2598	0.0127 ± 0.0327	88.2185

for $\varphi(\mathrm{GLU}, \mathrm{ATP}, \mathrm{PEP})$ are given by:

$$1 + \frac{\mathrm{ATP}}{K_{\mathrm{ATP}}} + \frac{\mathrm{GLU}}{K_{\mathrm{GLU}}} \cdot \frac{\mathrm{ATP}}{K_{\mathrm{ATP}}}, \tag{33}$$

$$1 + \frac{\mathrm{ATP}}{K_{\mathrm{ATP}}} + \frac{\mathrm{GLU}}{K_{\mathrm{GLU}}} \cdot \frac{\mathrm{PEP}}{K_{\mathrm{PEP}}} + \frac{\mathrm{GLU}}{K_{\mathrm{GLU}}} \cdot \frac{\mathrm{ATP}}{K_{\mathrm{ATP}}}, \tag{34}$$

$$1 + \frac{\mathrm{ATP}}{K_{\mathrm{ATP}}} + \frac{\mathrm{ATP}}{K_{\mathrm{ATP}}} \cdot \frac{\mathrm{PEP}}{K_{\mathrm{PEP}}} + \frac{\mathrm{GLU}}{K_{\mathrm{GLU}}} \cdot \frac{\mathrm{ATP}}{K_{\mathrm{ATP}}}. \tag{35}$$

Real Model and Data Generation

According to literature,[31,32] the reaction mechanism of glucokinase is ordered, with glucose as first binding substrate, and PEP inhibiting the binding of ATP to the enzyme. Based on these considerations, the fifth model was chosen as the real or *true* model ($m_5^\star$). This model was used to generate experimental data by simulating the experiment using the parameter values tabulated in Table 1, and by adding random noise to mimic the measurement error. The standard deviation of the measurements were calculated in the same way as

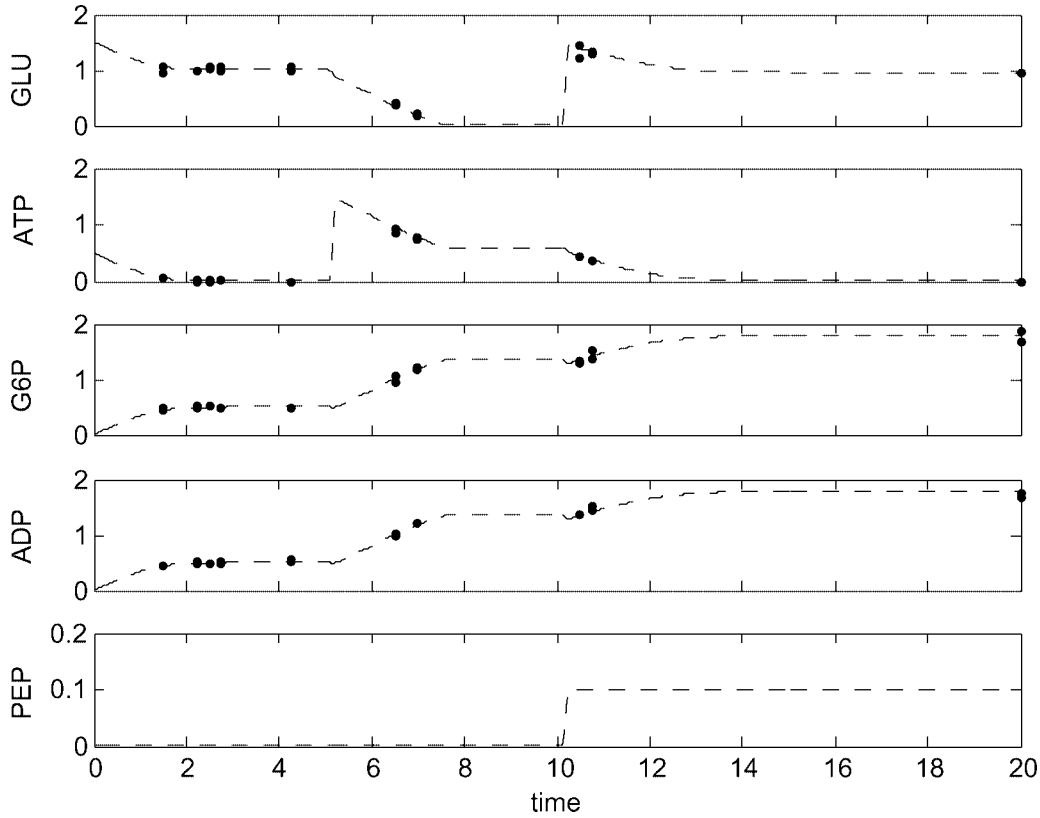

Figure 5. Preliminary experiment simulated with the real model ($m_5^\star$) (- -) and the experimental data (•) obtained from it.

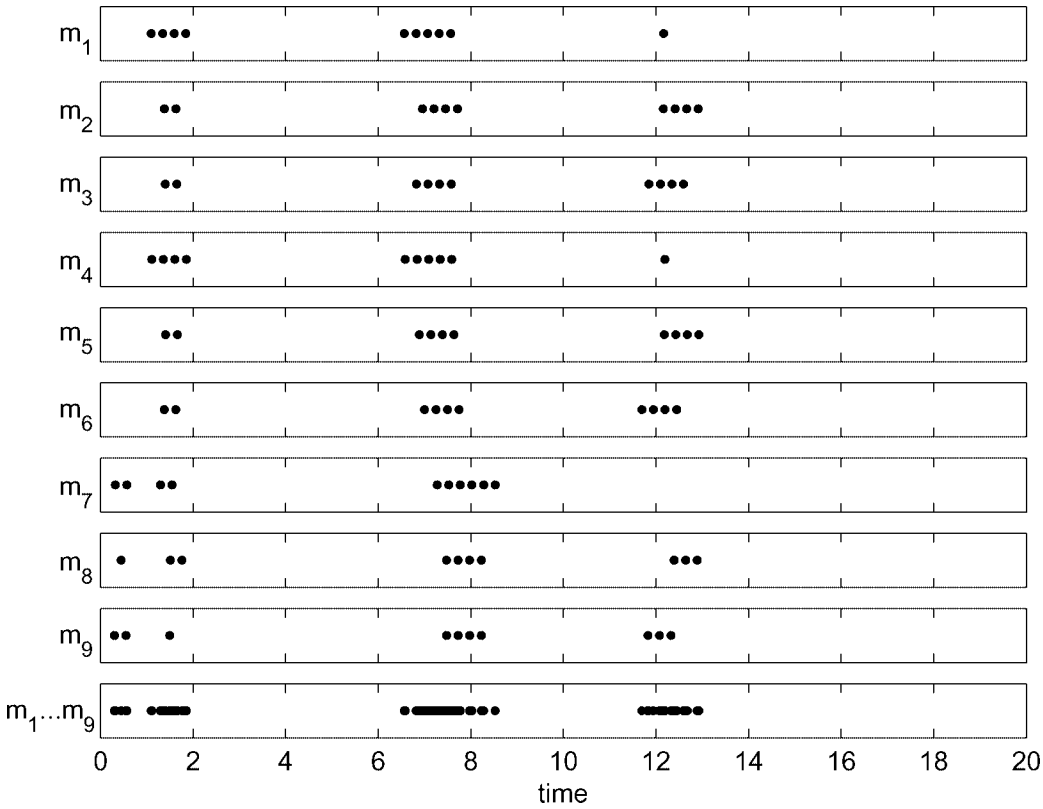

Figure 6. Optimal sampling times (•) found for the case where the D-optimality design criterion was optimized. The graph at the bottom was obtained by plotting the ten optimal sampling times of the nine individual models on the same axis.

suggested by Ternbach et al.:[15]

$$\sigma_y = \hat{y} \cdot \varsigma_y \cdot \left(1 + \frac{1}{\frac{\hat{y}}{lb_y} + \left(\frac{\hat{y}}{lb_y}\right)^2} \right). \tag{36}$$

Here, ς_y represents a constant minimal relative error, and lb_y represents the lower accuracy bound on the measurement of y. In this way, the standard deviation of the measurements is proportional to the value of $\hat{y}$, but increases when the latter approaches the lower accuracy bound on the measurement.

Preliminary Experiment

To initiate the case study, a preliminary experiment was defined and performed *in silico*. For this experiment, the volume of the reaction vessel was set to 10 mL, and the initial glucokinase concentration was set such that 5 units were present in the reaction mixture. Furthermore, it was assumed that no G6P, ADP and PEP were present at the start of the experiment, and the initial concentrations of glucose and ATP were set to 1.5 and 0.5 mM, respectively.

During the experiment, two pulses were given, for both of which the volume was equal to 1 mL. The first pulse was given five minutes after the start of the experiment, and only contained ATP. The ATP concentration was chosen such that the ATP concentration in the reaction mixture was raised to 1.5 mM. The second pulse, given ten minutes after the start of the experiment, contained glucose and PEP, and their concentrations were chosen such that the resulting concentrations were 1.5 and 0.1 mM, respectively.

The experiment stopped after 20 min, and 10 measurements of GLU, ATP, G6P, and ADP were taken in duplicate (see Fig. 5). The minimal relative errors (ς) were arbitrarily set to 0.05 for all measured state variables, and the lower accuracy bounds on the measurements were defined as 0.1 mM.

Parameter Estimation

The parameters of all rival models were estimated using the data from this preliminary experiment (Fig. 5), and using the optimization algorithm described in the section "Optimization Algorithms". Since negative parameter values do not make sense, the lower bounds were set to zero. The upper bounds were set to 1000 U/mg for parameter k, 2 mM for parameter K_{GLU}, 50 mM for parameter K_{ATP}, and 25 mM for parameter K_{PEP}. The results of this parameter estimation exercise are shown in Table 1. From these results, one can conclude that the accuracy of the parameter estimates is quite low, indicating that it may be beneficial to perform a compromise experiment to increase the accuracy of the parameter estimates prior to the start of the model discrimination procedure.

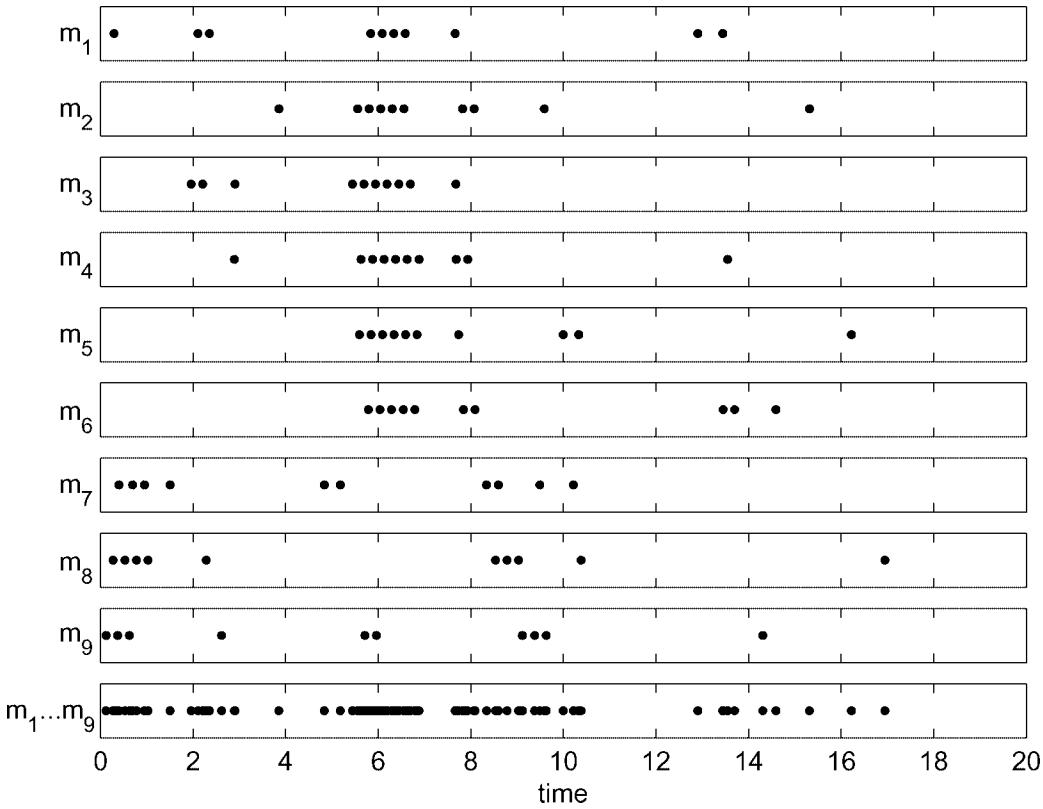

Figure 7. Optimal sampling times (•) found for the case where the modE-optimality design criterion was optimized. The graph at the bottom was obtained by plotting the ten optimal sampling times of the nine individual models on the same axis.

Optimal Experimental Design for Parameter Estimation

For each rival model, an experiment was designed to accurately estimate its parameters. The experimental degrees of freedom were chosen as in the preliminary experiment, except for the ten sampling times, which were optimized. This experimental design exercise was performed both for the case where the D-optimality design criterion was optimized, and for the case where the modE-optimality design criterion was used, the results of which are shown in Figures 6 and 7, respectively. In the following, a compromise experiment with ten sampling times is designed for both cases, based on the corresponding optimal experiments.

Compromise Experiment Obtained From D-Optimal Experimental Designs

This section describes the results for the case where the D-optimality design criterion is optimized (maximized). To clearly illustrate the different steps of the methodology, these results will be discussed in more detail than the results for the case where the modE-optimality design criterion is used (the next section).

Based on the D-optimal experiments corresponding to the individual models (shown in Fig. 6), the weighted kernel functions were determined after weighing the sampling times as explained in the section "Weighing of the Sampling Times". As an example, the kernel functions associated with the D-optimal sampling times for model m_1 are shown in Figure 8. This figure clearly illustrates that the importance of the individual sampling times with regard to the

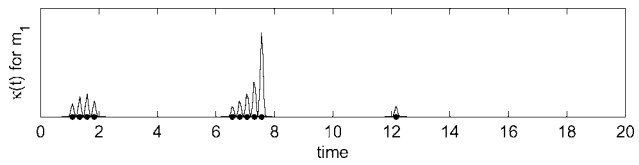

Figure 8. Weighted kernel functions associated with the D-optimal sampling times for model m_1.

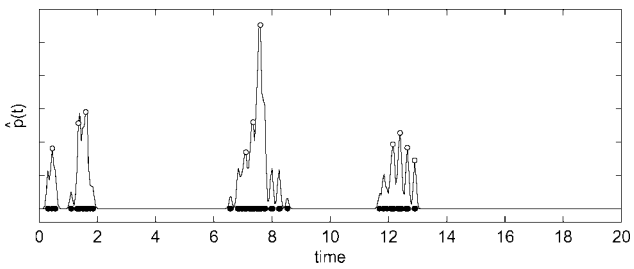

Figure 9. Trajectory of $\hat{p}(t)$ for the case where the D-optimality design criterion is applied, and illustration of how the compromise sampling times (○) were obtained from it. The optimal sampling times for the different models are represented by the black dots (•).

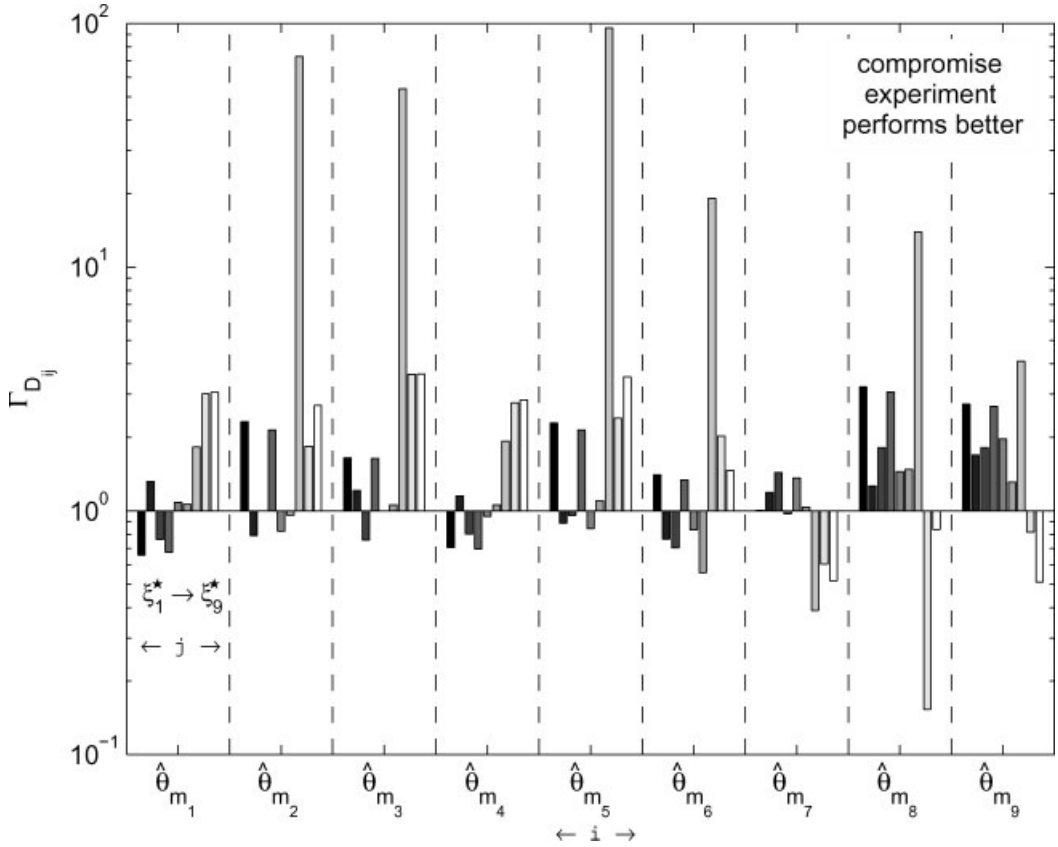

Figure 10. Comparison of the D-optimality design criterion values obtained when performing the compromise experiment, with those that would be obtained when the optimal experiments for the individual models were performed instead. The ratio between these criterion values is shown in the figure for each model, and this for the optimal experiments associated with models m_1 (black bars) to m_9 (white bars).

information content of the experiment differs, and that it makes sense to weigh the different sampling times.

The weighted kernel functions associated with the different models are used to calculate $\hat{p}(t)$ according to eq. (14), and the resulting trajectory of $\hat{p}(t)$ is shown in Figure 9. The compromise sampling times are determined from this trajectory as explained in the section "Kernel-Based Method for Experimental Design", and their location is indicated by the white dots. One can see that the optimal sampling times are located in four groups, and that the compromise sampling times are (not surprisingly) spread over these groups as well. An interesting observation in Figure 6 is that the optimal experiments for models m_7 to m_9 contain sampling times in the first minute of the experiment, while this is not the case for models m_1 to m_6. The fact that this region is important for one third of the models is reflected in the compromise experiment, where one of the sampling times is put in this region. This example clearly illustrates that the presented method considers the optimal sampling times for each of the models when designing the compromise experiment.

The capability of the presented method to design a compromise experiment was evaluated as explained in the section "Evaluating the Designed Compromise Experiment". The results of this evaluation are presented in Figure 10. For each model m_i, the value of $\Gamma_{D_{i1}}$ is represented by the black bar ($\boldsymbol{\xi}_1^\star$), and the bars become increasingly white as the model number increases ($\boldsymbol{\xi}_1^\star \rightarrow \boldsymbol{\xi}_9^\star$). To present the results in a systematic and easily interpretable form, the values of $\Gamma_{D_{ij}}$ are represented on a logarithmic scale. In this way, it is easy to see when $\Gamma_{D_{ij}} > 1$. Some interesting observations are discussed below.

The results for model m_1, for example, show that $\Gamma_{D_{1j}} < 1$ for experiments $\boldsymbol{\xi}_1^\star$, $\boldsymbol{\xi}_3^\star$ and $\boldsymbol{\xi}_4^\star$, which indicates that these experiments contain more information with regard to the parameters of

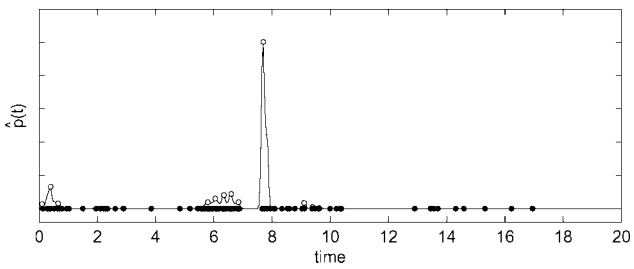

Figure 11. Trajectory of $\hat{p}(t)$ for the case where the modE-optimality design criterion is applied, and illustration of how the compromise sampling times (○) were obtained from it. The optimal sampling times for the different models are represented by the black dots (●).

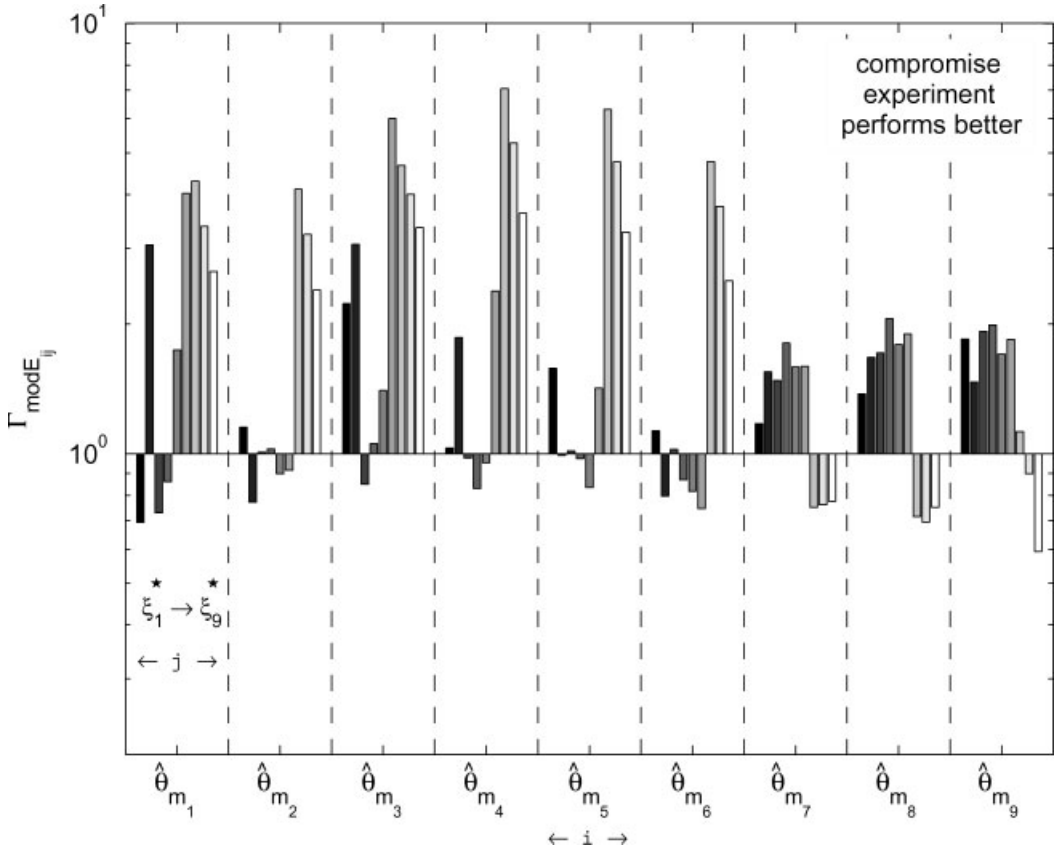

Figure 12. Comparison of the modE-optimality design criterion values obtained when performing the compromise experiment, with those that would be obtained when the optimal experiments for the individual models were performed instead. The ratio between these criterion values is shown in the figure for each model, and this for the optimal experiments associated with models m_1 (black bars) to m_9 (white bars).

model m_1 than the compromise experiment. For the other optimal experiments, this is not the case and the compromise experiment is preferred. If one would perform $\boldsymbol{\xi}_4^\star$ instead of the compromise experiment, the information content would indeed be higher for model m_1, but it would be lower for the other models (except for model m_4, of course). The latter can be seen when comparing the bars corresponding to $\boldsymbol{\xi}_4^\star$ for the different models. Similar observations can be made for the other models/optimal experiments, which clearly shows the ability of the presented method to design an experiment with the characteristics of a compromise experiment. That the compromise experiment is not optimal for the individual models is a direct result of the fact that the timings of the optimal sampling times are different for the individual models (see Fig. 6). Yet, the compromise experiment seems to be sufficiently informative to improve the overall accuracy of the parameter estimates.

Compromise Experiment Obtained From modE-Optimal Experimental Designs

This section describes the results for the case where the modE-optimality design criterion is optimized (minimized). From Figure 7, one can see that the optimal sampling times are not located in distinct groups as was the case for the other criterion (see Fig. 6). This makes the exercise more challenging than the previous one.

The trajectory of $\hat{p}(t)$ and the compromise sampling times derived from it are shown in Figure 11. The results of the evaluation are presented in Figure 12, and show that the capability of the presented method to design a compromise experiment is at least as good as for the case where the D-optimality design criterion was used. Here too, it is clear that the presented method leads to an experiment with the characteristics of a compromise experiment. For each model, some of the optimal experiments are preferred to the compromise experiment ($\Gamma_{\text{modE}_{ij}} < 1$), and vice versa ($\Gamma_{\text{modE}_{ij}} > 1$). The results also show that for some of the models, a significant amount of information is lost when an optimal experiment is performed instead of the compromise experiment. For instance, if experiments $\boldsymbol{\xi}_7^\star$, $\boldsymbol{\xi}_8^\star$, or $\boldsymbol{\xi}_9^\star$ were performed, a substantial amount of information would be lost with regard to the parameters of models m_1 to m_6, while the gain in information for models m_7 to m_9 is not that large.

Dependence of the Experimental Designs on the Parameter Estimates

The values of the D-optimality design criterion shown in Figure 10 represent the expected information content of the designed experiments. As explained in the section "Central Rationale Behind

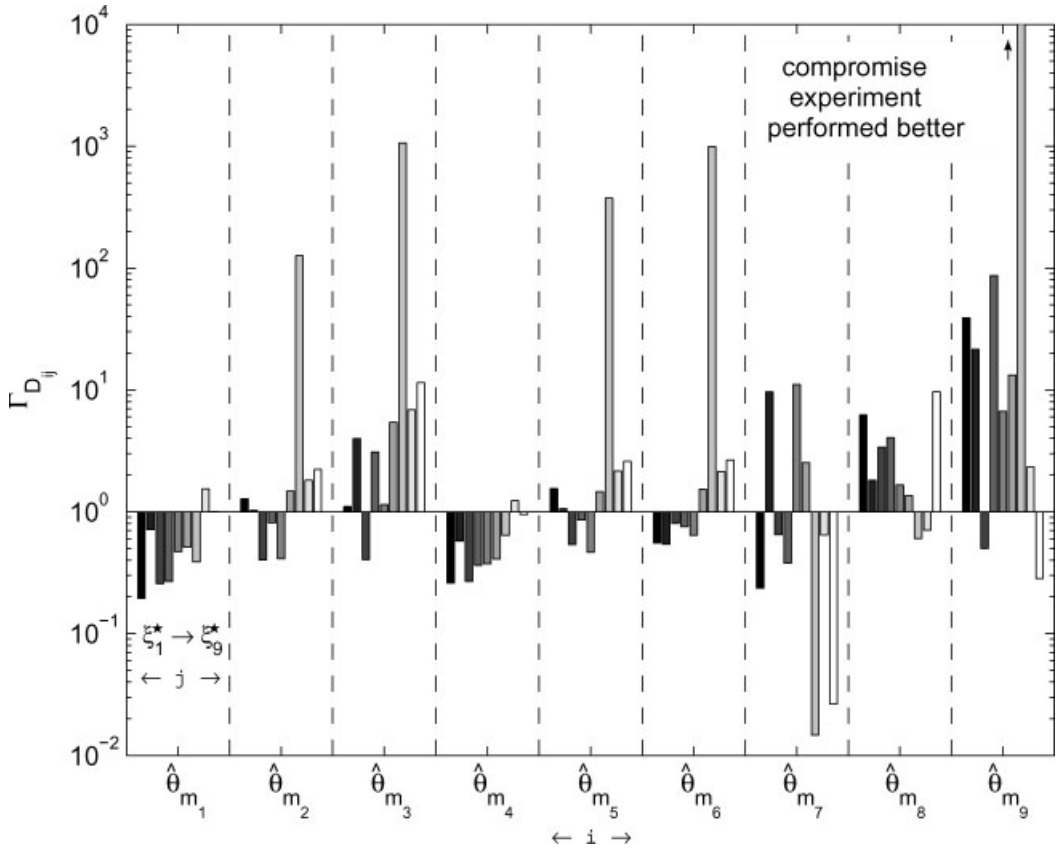

Figure 13. Comparison of the D-optimality design criterion values obtained after performing the compromise experiment and re-estimating the parameters, with those that were obtained when the optimal experiments for the individual models were performed instead. The ratio between these criterion values is shown in the figure for models m_1 (black bars) to m_9 (white bars), and this for each optimal experiment.

Optimal Experimental Design for Parameter Estimation", the latter is assessed based on the **FIM**. For linear models, the parameter sensitivities ($\partial\hat{\boldsymbol{y}}/\partial\boldsymbol{\theta}$) are, by definition, independent of the parameters to be estimated. Hence, the **FIM** is independent of the parameter estimates as well [eq. (7)]. However, for nonlinear models (as the ones used in this article), this is not the case, and the **FIM**, as well as the design criteria derived from it, are dependent on the values of the parameter estimates available at the experimental design step. Thus, at the start of a modelling exercise, when the parameter estimates are still uncertain, the designed experiment may be less informative/optimal than expected, and the resulting parameter estimates may not be that accurate after all. Therefore, the designed experiments are often called locally optimal instead of optimal.[19]

This is illustrated for the case where the D-optimality design criterion is used. A similar figure as Figure 10 was made after performing the ten different experiments in silico ($\boldsymbol{\xi}_c$ and $\boldsymbol{\xi}_j^\star$), and using the data from these experiments to re-estimate the parameters of each of the nine models (Fig. 13). As expected from the discussion above, the results shown in Figure 13 show that the values for $\Gamma_{D_{ij}}$ are not entirely the same as the ones presented in Figure 10, which indicates that the parameter estimates changed after re-estimating them.

Nevertheless, the overall accuracy of the parameter estimates has improved by performing the compromise experiment. This can be concluded from Table 2, in which the parameter estimates and their corresponding 95% confidence intervals are reported. Note, however, that the covariance or correlation between the parameter estimates is not considered in the calculation of the confidence intervals. A better picture of the uncertainty (or accuracy) of the parameter estimates can be obtained from Figure 13. Since the D-optimality criterion values are proportional to the volume of the confidence region of the parameter estimates (as explained in the section "Experimental Design Criteria Based on the FIM"), they can also be interpreted as an overall measure for the accuracy of the parameter estimates.

Conclusions

In this article, a method was presented to design an experiment to simultaneously improve the accuracy of the parameter estimates of several rival models. The method is inspired by kernel density estimation and uses the optimal sampling times for the individual models to design a compromise experiment, which is an experiment that is not optimal for any of the individual models, but sufficiently

Table 2. Parameters of the Real Model ($m_5^\star$) That Were Used to Generate Experimental Data, and the Parameter Estimates Obtained After Fitting the Rival Models to the Data From Both the Preliminary Experiment and the Compromise Experiment, Together With the 95% Confidence Intervals and the Corresponding WSSE-Values.

Model	k	K_{GLU}	K_{ATP}	K_{PEP}	WSSE
$m_5^\star$	312	0.15	0.13	0.10	–
m_1	356.75 ± 26.76	0.0383 ± 0.0191	0.2275 ± 0.0304	–	506.14
m_2	328.35 ± 22.16	0.0332 ± 0.0177	0.1504 ± 0.0245	0.1182 ± 0.0273	128.50
m_3	323.27 ± 20.49	0.0252 ± 0.0163	0.1458 ± 0.0230	0.0059 ± 0.0038	126.58
m_4	341.31 ± 18.64	0.1930 ± 0.0891	0.1894 ± 0.0189	–	503.37
m_5	317.09 ± 14.99	0.2063 ± 0.1097	0.1264 ± 0.0172	0.0945 ± 0.0219	128.78
m_6	322.97 ± 15.62	0.1759 ± 0.1043	0.1351 ± 0.0252	0.0293 ± 0.0141	137.27
m_7	583.38 ± 123.24	0.0095 ± 0.0618	49.9430 ± 315.47	–	1362.41
m_8	358.86 ± 36.15	0.0034 ± 0.0329	40.4657 ± 394.03	17.6839 ± 174.29	630.45
m_9	459.68 ± 66.39	0.0406 ± 0.0489	5.7254 ± 6.5114	0.0026 ± 0.0028	817.53

informative to improve the overall accuracy of the parameters of all rival models. The fact that the contribution of the individual sampling times to the information content of the experiment varies, and some sampling times are thus more important/informative than others, was taken into account. Although the rationale of the presented approach could in principle be applied to all types of experimental degrees of freedom (manipulations, initial conditions and sampling times), this article was restricted to the optimization of sampling times.

The presented method was illustrated by applying it to a case study where nine rival models are defined to describe the kinetics of an enzyme-catalyzed reaction (glucokinase). The capability of the kernel-based method to design compromise experiments was evaluated, and the results of this evaluation clearly showed that the presented method is capable to design compromise experiments.

Acknowledgment

Peter Vanrolleghem holds the Canada Research Chair on Water Quality Modelling.

References

1. Buzzi-Ferraris, G.; Forzatti, P.; Emig, G.; Hofmann, H. Chem Eng Sci 1984, 39, 81.
2. Chen, B. H.; Asprey, S. P. Ind Eng Chem Res 2003, 42, 1379.
3. Donckels, B. M. R.; De Pauw, D. J. W.; De Baets, B.; Maertens, J.; Vanrolleghem, P. A. Chemom Intell Lab Syst (doi: 10.1016/j.chemolab.2008.08.002).
4. Hunter, W. G.; Reiner, A. M. Technometrics 1965, 7, 307.
5. Vanrolleghem, P.; Van Daele, M. Water Sci Technol 1994, 30, 243.
6. Baltes, M.; Schneider, R.; Sturm, C.; Reuss, M. Biotechnol Prog 1994, 10, 480.
7. Munack, A. In Biotechnology: A Multi-Volume Comprehensive Treatise, Vol. 4: Measuring, Modelling and Control; Schügerl, K., Ed.; VCH: Weinheim, 1991; p. 251.
8. Vanrolleghem, P. A.; Dochain, D. In Advanced Instrumentation, Data Interpretation, and Control of Biotechnological Processes; Van Impe, J. F. M.; Vanrolleghem, P. A.; Iserentant, D. M., Eds.; Kluwer Academic: Dordrecht, 1998; p. 251.
9. Walter, E.; Pronzato, L. Identification of Parametric Models from Experimental Data, Springer–Verlag: New York, 1997; 413 p.
10. Hill, W. J.; Hunter, W. G.; Wichern, D. W. Technometrics 1968, 10, 145.
11. Burke, A. L.; Duever, T. A.; Penlidis, A. J Polym Sci Part A: Polym Chem 1996, 34, 2665.
12. Burke, A. L.; Duever, T. A.; Penlidis, A. Ind Eng Chem Res 1997, 36, 1016.
13. Kremling, A.; Fischer, S.; Gadkar, K.; Doyle, F. J.; Sauter, T.; Bullinger, E.; Allgöwer, F.; Gilles, E. D. Genome Res 2004, 14, 1773.
14. Schwaab, M.; Silva, F. M.; Queipo, C. A.; Barreto, A. G., Jr.; Nele, M.; Pinto, J. C. Chem Eng Sci 2006, 61, 5791.
15. Ternbach, M. A. B.; Bollman, C.; Wandrey, C.; Takors, R. Biotechnol Bioeng 2005, 91, 356.
16. Marsili-Libelli, S.; Guerrizio, S.; Checchi, N. Ecol Model 2003, 165, 127.
17. Omlin, M.; Reichert, P. Ecol Model 1999, 115, 45.
18. Ljung, L. System Identification, Theory for the User; Prentice Hall: New Jersey, 1999; 608 p.
19. Atkinson, A. C.; Donev, A. N. Optimum Experimental Design; Oxford University Press, New York, 1992; 328 p.
20. Petersen, B. Calibration, Identifiability and Optimal Experimental Design of Activated Sludge Models, Ph.D. Thesis; Ghent University: Belgium, 2003; 337 p.
21. Seber, G. A. F.; Wild, C. J. Nonlinear Regression; Wiley: New Jersey, 1989; 768 p.
22. De Pauw, D. J. W.; Vanrolleghem, P. A. Water Sci Technol 2006, 53, 117.
23. Rosenblatt, M. Ann Math Stat 1956, 27, 832.
24. Jenssen, R.; Principe, J. C.; Erdogmus, D.; Eltoft, T. J Franklin Inst 2006, 343, 614.
25. Duda, R. O.; Hart, P. E.; Stork, D. G. Pattern Classification; Wiley, 2001; 654 p.
26. Turlach, B. A. Bandwidth Selection in Kernel Density Estimation: A Review, Discussion Paper 9307; Institut für Statistik und Ökonometrie, Humboldt-Universität zu Berlin, 1993.
27. Cardoso, M. F.; Salcedo, R. L.; Feyo de Azevedo, S. Comput Chem Eng 1996, 20, 1065.
28. Nelder, J. A.; Mead, R. Comput J 1965, 7, 308.
29. Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. Sci 1983, 220, 671.
30. Segel, I. H. Enzyme Kinetics—Behaviour and Analysis of Rapid Equilibrium and Steady–State Enzyme Systems; Wiley: New York, 1975; 957 p.
31. Ogawa, T.; Mori, H.; Tomita, M.; Yoshino, M. Res Microbiol 2007, 158, 159.
32. Monasterio, O.; Cárdenas, M. L. Biochem J 2003, 371, 29.