Model-Based Design of Experiments in the Presence of Continuous Measurement Systems

ARTICLE *in* INDUSTRIAL & ENGINEERING CHEMISTRY RESEARCH · JANUARY 2011

Impact Factor: 2.59 · DOI: 10.1021/ie1019062

4 AUTHORS, INCLUDING:

Federico Galvanin
University College London
32 PUBLICATIONS **142** CITATIONS
SEE PROFILE

Fabrizio Bezzo
University of Padova
104 PUBLICATIONS **1,334** CITATIONS
SEE PROFILE

I&EC research
Industrial & Engineering Chemistry Research

ARTICLE

pubs.acs.org/IECR

Model-Based Design of Experiments in the Presence of Continuous Measurement Systems

Federico Galvanin, Andrea Boschiero, Massimiliano Barolo, and Fabrizio Bezzo*

Computer-Aided Process Engineering Laboratory (CAPE-Lab), Dipartimento di Principi e Impianti di Ingegneria Chimica, Università di Padova, via Marzolo 9, I-35131 Padova PD, Italy

ABSTRACT: Model-based design of experiments (MBDoE) techniques are a useful tool to maximize the information content of experimental trials when the purpose is identifying the set of parameters of a deterministic model in a statistically sound way. Traditionally, the problem of MBDoE has been addressed for discrete measurement systems. In this case, formulation of the optimal design problem is based on maximization of the expected information, usually calculated from discrete forms of the Fisher information matrix. However, current measurement technology allows measurements to be taken at a much higher frequency than in the past, to a point that measurements may be assumed to be obtained in a continuous way. A novel design criterion allowing for the continuous model-based design of the experiments (CMBDoE) is formulated in this paper by optimizing a continuous measurement function of the Fisher information matrix, with the purpose of reaching a statistically satisfactory estimation of model parameters in a computationally efficient way. The benefits of the proposed strategy are discussed by means of two simulated case studies, where the effectiveness of the design is assessed by comparison to a standard MBDoE approach.

1. INTRODUCTION

Simulation, design, control, and optimization of chemical processes are based on the availability of reliable mathematical models. A large class of physical phenomena can be described by dynamic deterministic models through the statement of laws and correlations in the form of a system of differential and algebraic equations (DAEs). Once a dynamic model structure is found adequate to represent a physical system, experiments need carrying out to estimate the set of parameters of the model in the most precise and accurate way. Model-based design of experiments (MBDoE) techniques for parameter identification have been developed and applied to linear dynamic models since the 1970s,[1,2] but their practical extension to complex nonlinear dynamic systems has been a relatively recent achievement.[3−8] MBDoE techniques represent a consolidated tool for the rapid assessment and identification of dynamic deterministic models, allowing for maximization of the information content of the experimental trials needed to perform the parameter identification task.[9]

In experimental practice, information is usually acquired from the experiment through discrete collection (sampling) of data. Consequently, when planning an experiment through a conventional MBDoE technique, the mathematical formulation of the expected information being maximized (i.e., the prediction of the information that will be gained from the experiment) is also expressed in a discrete way, i.e., through a discrete form of the Fisher information matrix.[10] A significant benefit provided by this design formulation is that it allows for the optimal allocation of the sampling points,[11] calling for more samples within the time windows where the experiment is expected to be more informative.

Thanks to the progress of sensor technology, nowadays several system outputs can be measured almost continuously and at a cost much lower than was possible just one decade ago. This is true not only for standard process measurements (e.g., temperature, flow, level, pressure) but also for quality measurements (e.g., density, composition, pH), which usually are critical for biological processes and biomedical systems. To quote just a few examples, continuous measurement systems have been developed for monitoring concentration in biological processes through near-infrared spectroscopy[12] and online respirometry techniques;[13] continuous glucose monitoring systems (CGMSs) now allow for the continuous recording of glycaemia in diabetic subjects over extended periods of time.[14] System monitoring and control can undoubtedly benefit from the features of these modern measurement technologies. Investigating to which extent these features can be exploited to design effective experiments for the model identification task is the intent of this paper.

If the samples are collected very frequently, the measure of the actual information gained from the experiment can be approximated by a continuous profile over the experimental horizon. Following this premise, a novel experiment design criterion, which significantly reduces the computational burden related to the MBDoE exercise, is presented and discussed in this paper. The proposed technique is suitable for systems in which continuous (or highly frequent) measurements are available. In this case, the optimal design problem can be formulated and solved by optimizing a continuous measurement function of the Fisher information matrix with the purpose of reaching a statistically satisfactory estimation of model parameters in a computationally efficient way.

The applicability to nonlinear dynamic systems and effectiveness of the proposed continuous MBDoE (CMBDoE) approach are illustrated using two simulated case studies: the first one concerns identification of a fermentation bioreactor model, and the second one is related to identification of a physiological model of glucose homeostasis.

Received: September 16, 2010
Accepted: December 22, 2010
Revised: December 22, 2010
Published: January 19, 2011

dx.doi.org/10.1021/ie1019062 | *Ind. Eng. Chem. Res.* 2011, 50, 2167–2175

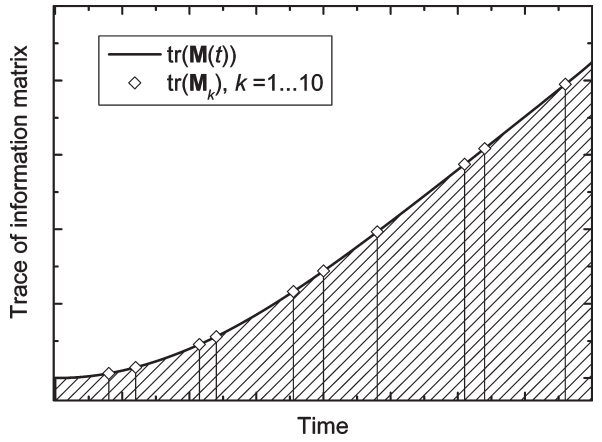

Figure 1. Evaluation of expected information: standard A-optimal design aims at maximizing the sum of the contributions (open diamonds) to the overall information, while CMBDoE maximizes the area underneath the trace function.

2. EXPERIMENT DESIGN METHODOLOGY

Conventional MBDoE procedures involve the sequential interaction between three key activities:

1 design of the dynamic experiment;
2 experiment execution;
3 parameter estimation.

Within the sequence of key activities, an information mismatch is introduced between the expected information (i.e., information as predicted by the model, usually scarcely accurate at the beginning of the procedure) and the actual information provided by the experiment. In order to ensure the effectiveness of the designed experiment, a reliable mathematical representation of information during each key activity is required. Let us consider a process described by a set of DAEs in the form

$$\begin{cases} \mathbf{f}(\dot{\mathbf{x}}(t), \mathbf{x}(t), \mathbf{u}(t), \mathbf{w}, \boldsymbol{\theta}, t) = 0 \\ \hat{\mathbf{y}}(t) = \mathbf{g}(\mathbf{x}(t)) \end{cases} \tag{1}$$

subject to

$$\mathbf{C} = \mathbf{x}(t) - \mathbf{G}(t) \leq 0 \tag{2}$$

with the set of initial conditions $\mathbf{x}(0) = \mathbf{x}_0$, where $\mathbf{x}(t)$ is the N_x-dimensional vector of time-dependent state variables, $\mathbf{u}(t)$ and $\mathbf{w}$ are the time-dependent and time-invariant control variables (of dimensions N_u and N_w), respectively, $\boldsymbol{\theta}$ is the N_θ-dimensional set of unknown model parameters to be estimated, and t is time. The symbol ^ is used to indicate the estimate of a variable (or of a set of variables): thus, $\mathbf{y}(t)$ is the vector of measured values of the outputs, while $\hat{\mathbf{y}}$ is the vector of the corresponding values estimated by the model. $\mathbf{C}$ is an N_c-dimensional set of constraint functions expressed through the set $\mathbf{G}(t) \in \boldsymbol{R}^{N_c}$ of active constraints on the state variables. Design optimization is carried out by acting on the n_φ-dimensional experiment design vector $\boldsymbol{\varphi}$

$$\boldsymbol{\varphi} = [\mathbf{y}_0, \mathbf{u}(t), \mathbf{w}, \mathbf{t}^{sp}, \tau]^{\mathrm{T}} \tag{3}$$

which includes the N_y-dimensional set of initial conditions $\mathbf{y}_0$ on the measured variables, the duration of the experiment τ, the continuously manipulated inputs $\mathbf{u}(t)$, and the set of time-invariant manipulated inputs $\mathbf{w}$. The set of time instants at which the output variables are sampled is also a design variable and is expressed through the n_{sp}-dimensional vector $\mathbf{t}^{sp}$ of sampling times.

Conventional MBDoE techniques aim at decreasing the model parameter uncertainty region predicted by model 1 as the solution of the optimization problem

$$\begin{aligned} \boldsymbol{\varphi}^{\mathrm{opt}} &= \arg \min_{\boldsymbol{\varphi}} \{\psi[\mathbf{V}_\theta(\boldsymbol{\theta}, \boldsymbol{\varphi})]\} \\ &= \arg \min_{\boldsymbol{\varphi}} \{\psi[\mathbf{H}_\theta^{-1}(\boldsymbol{\theta}, \boldsymbol{\varphi})]\} \end{aligned} \tag{4}$$

subject to eq 2 and to a n_φ-dimensional set of constraints on design variables. $\mathbf{V}_\theta$ and $\mathbf{H}_\theta$ are the variance−covariance matrix of model parameters and the dynamic information matrix, respectively. As stated in eq 4, the experiment is designed so as to minimize a measurement function ψ of $\mathbf{V}_\theta$. The particular form of the measurement function represents the design criterion selected in order to maximize the expected information content of the experiment as predicted by the model. The most common design criteria are the so-called alphabetical ones, i.e. A-, D-, and E-optimal criteria (minimizing the trace, the determinant, and the maximum eigenvalue of $\mathbf{V}_\theta$, respectively[15]) or they are based on singular values decomposition.[16,17] Uncertainty on the design variables can also be considered in the design formulation to take into account possible deviations from planned conditions during experiment execution.[18]

The dynamic information matrix is usually expressed by a discrete dynamic form of the Fisher information matrix as proposed by Zullo[10]

$$\begin{aligned} \mathbf{H}_\theta(\boldsymbol{\theta}, \boldsymbol{\varphi}) &= \mathbf{H}_\theta^0 + \sum_{k=1}^{n_{sp}} \sum_{i=1}^{N_y} \sum_{j=1}^{N_y} s_{ij} \left[\frac{\partial \hat{y}_i(t_k)}{\partial \theta_l} \frac{\partial \hat{y}_j(t_k)}{\partial \theta_m}^T \right]_{l,\, m = 1 \ldots N_\theta} \\ &= \mathbf{H}_\theta^0 + \sum_{k=1}^{n_{sp}} \mathbf{M}_k \end{aligned} \tag{5}$$

In eq 5 s_{ij} is the ijth element of the $N_y \times N_y$ inverse matrix of measurements error, $\mathbf{M}_k$ represents the amount of information that can be recovered from the kth sample, and $\mathbf{H}_\theta^0$ is the prior dynamic information matrix, taking into account the preliminary statistical information about the parametric system before each trial is carried out. Here, the measurement errors are assumed to be Gaussian and uncorrelated, which, in general, may be an approximation. However, note that a correct description of the measurement errors would depend on the sensor system; furthermore, assuming that the errors be uncorrelated is a widely adopted simplification,[19] even for continuous systems.[20]

2.1. Continuous Model-Based Design of Experiments. Let us consider a constant sampling interval Δt. Also, let us consider the following operation on matrix $\mathbf{H}_\theta$

$$\begin{aligned} \lim_{n_{sp} \to \infty} [\mathbf{H}_\theta(\boldsymbol{\theta}, \boldsymbol{\varphi}) \Delta t] &= \lim_{n_{sp} \to \infty} \left(\sum_{k=1}^{n_{sp}} [\mathbf{M}_k \Delta t] + [\mathbf{H}_\theta^0 \Delta t] \right) \\ &= \int_0^\tau [\mathbf{M}(t)] \mathrm{d}t \end{aligned} \tag{6}$$

where $\mathbf{M}(t)$ allows for the dynamic evaluation of the expected information.

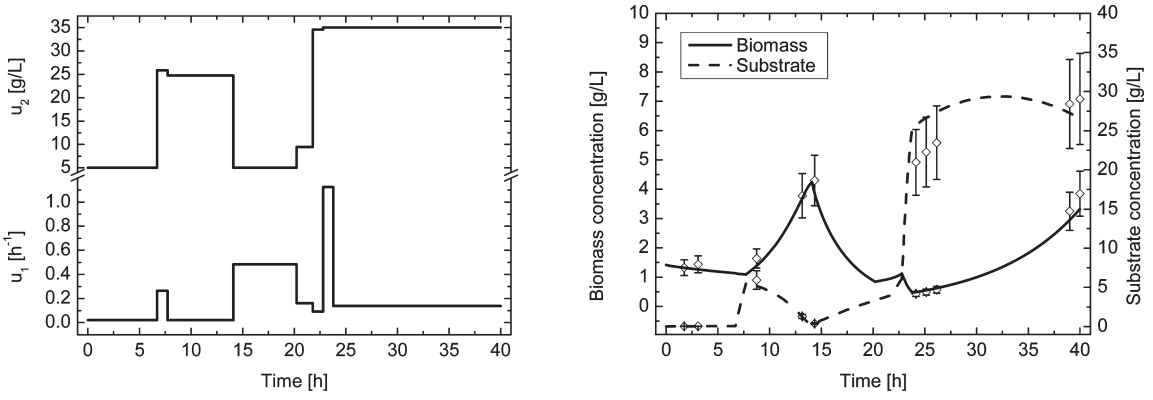

Figure 2. Case study 1: standard experiment design. (a) Profiles of the manipulated inputs as optimized by a conventional MBDoE. (b) Profiles for biomass and substrate concentration after execution of the MBDoE-planned experiment (full lines); measurements are indicated by diamonds. The 95% confidence intervals on measurements is represented by the error bars.

Table 1. Case Study 1: Standard Experiment Design—Parameter Estimation and Related a Posteriori Statistics as 95% Confidence Intervals and *t* Values

	MBDoE		
model parameter	final value	confidence interval 95%	95% *t* value
Θ_1	1.0132	0.2190	4.62
Θ_2	1.1361	0.3800	2.99
Θ_3	0.9633	0.0993	9.70
Θ_4	1.0600	0.9460	1.12[a]

[a] Superscript denotes *t* values failing the test (the reference *t* value is 1.74).

Now, let us consider a generic metric based on the above continuous formulation

$$\lim_{n_{sp}\to\infty} \psi\left[\mathbf{H}_\theta(\boldsymbol{\theta},\boldsymbol{\varphi})\Delta t\right] = \psi\left[\int_0^\tau \mathbf{M}(t)\mathrm{d}t\right] = \psi[\mathbf{L}(\boldsymbol{\theta},\boldsymbol{\varphi})] \quad (7)$$

where the $N_\theta \times N_\theta$ symmetrical positive definite matrix $\mathbf{L}$ of integral functions is defined as

$$\mathbf{L}(\theta,\boldsymbol{\varphi}) = \begin{bmatrix} \int_0^\tau m_{11}(t)\mathrm{d}t & \cdots & \int_0^\tau m_{1N_\theta}(t)\mathrm{d}t \\ \vdots & \ddots & \vdots \\ \int_0^\tau m_{N_\theta 1}(t)\mathrm{d}t & \cdots & \int_0^\tau m_{N_\theta N_\theta}(t)\mathrm{d}t \end{bmatrix} \quad (8)$$

In eq 8 m_{ij} represents the ijth element of $\mathbf{M}$, the overall expected information being maximized by design.

A novel design criterion for the continuous model-based design of experiments is introduced as follows

$$\boldsymbol{\varphi}^{\mathrm{opt}} = \arg\max_{\boldsymbol{\varphi}} \left\{ \psi\left[\int_0^\tau \mathbf{M}(t)\mathrm{d}t\right] \right\} \quad (9)$$

or similarly to eq 4

$$\boldsymbol{\varphi}^{\mathrm{opt}} = \arg\min_{\boldsymbol{\varphi}} \left\{ \psi\left[\int_0^\tau \mathbf{M}(t)\mathrm{d}t\right]^{-1} \right\} \quad (10)$$

The benefit of adopting eq 9 or 10 as the design objective function comes from the fact that the metric of the expected information becomes continuous as the measurement system itself, thus allowing for a continuous exploitation of the available information. In particular, the information can be exploited from the very beginning of the experiment, which allows to steer the information dynamics according to a convenient profile.

Evaluation of eq 8 requires evaluating $N_\theta(N_\theta+1)/2$ integral functions, and this may be computationally expensive even for a limited number of model parameters. Consequently, in this work a simplified approach is adopted. If an A-optimal criterion is chosen, then the CMBDoE optimality condition is expressed as

$$\begin{aligned} \boldsymbol{\varphi}^{\mathrm{opt}} &= \arg\max_{\boldsymbol{\varphi}} \left\{ tr \int_0^\tau \left[\mathbf{M}(t)\right]\mathrm{d}t \right\} \\ &= \arg\max_{\boldsymbol{\varphi}} \left\{ \int_0^\tau tr\left[\mathbf{M}(t)\right]\mathrm{d}t \right\} \end{aligned} \quad (11)$$

Criterion 11 is illustrated in Figure 1. While a standard A-optimal design criterion on matrix $\mathbf{H}_\theta$ minimizes the sum of n_{sp} trace values, the proposed criterion aims at minimizing the whole area underneath the trace function.

Unlike for the solution of eq 10, thanks to the linearity of the trace operator only one evaluation of the integral function is required to solve eq 11, with great benefit in terms of computational robustness and efficiency during design optimization. On the other hand, also note that, unlike eq 4 or 10, no matrix inversion is carried out (i.e., optimization is carried out on $\mathbf{H}_\theta$, not on $\mathbf{V}_\theta$) and only the diagonal contributions to information are considered. Although neglecting the off-diagonal correlation terms is generally a disadvantage, criterion 11 aims at direct optimization of the dynamic sensitivity of $\mathbf{y}$ on $\boldsymbol{\theta}$, with the benefit of avoiding numerical instability problems related to inversion of the Fisher information matrix when the system is affected by structural identifiability issues.[21]

The integral in eq 11 is estimated numerically by using a quadrature method, whose computational cost usually depends on the number of adopted quadrature nodes n_q used in the evaluations of the interpolating functions. Note, however, that in general the number of nodes is significantly smaller than the number of collected samples ($n_q \ll n_{sp}$) when a continuous

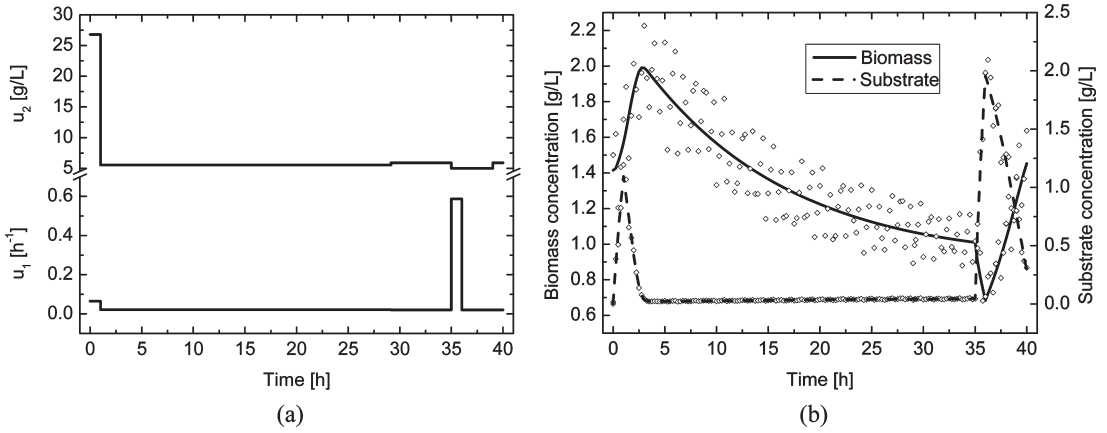

Figure 3. Case study 1: CMBDoE. (a) Profiles of the optimal manipulated inputs. (b) Estimated profiles for biomass and substrate concentration after execution of the CMBDoE planned experiment (full lines); measurements are indicated by diamonds (confidence intervals on measurements are not shown for clarity).

Table 2. Case Study 1: CMBDoE—Parameter Estimation and Related a Posteriori Statistics as 95% Confidence Intervals and *t* Values for the Nominal (τ = 40 h) and Minimal Length (τ = 14.5 h) Experiment[a]

	CMBDoE (τ = 40 h)			CMBDoE (τ = 14.5 h)		
model parameter	final value	confidence interval 95%	95% *t* value	final value	confidence interval 95%	95% *t* value
Θ_1	0.9806	0.0942	10.40	0.9264	0.2074	4.46
Θ_2	0.9761	0.0761	12.85	0.9789	0.1206	8.12
Θ_3	0.9685	0.1354	7.15	0.9089	0.2511	3.62
Θ_4	0.9820	0.1740	5.62	0.9140	0.3840	2.38

[a] The reference *t* values for the two tests are, respectively, 1.65 and 1.70.

measurement system is adopted. Adaptive quadrature methods[22] and adaptive integration algorithms[23] can be usefully adopted to decrease the number of nodes required in the integral evaluation, with great benefit in terms of computational time saving. In this paper, an implicit backward differentiation formula (BDF) has been used to evaluate integral 11 by enclosing it in the solution of the DAEs system in eq 1.

Two different case studies are examined in order to compare a standard MBDoE approach with the proposed CMBDoE methodology; they differ in terms of the number of model parameters to be estimated and for the absence/presence of active constraints on the state variables. The gPROMS modeling environment has been used for modeling, simulation, and optimization purposes, as well as to design the experiments. As in previous works, an SQP (sequential quadratic programming) routine was adopted in the two-step multiple shooting technique[23] to solve the nonlinear optimization problem. In the first step, the optimal design problem is solved as a maximization of the trace of the dynamic information matrix over the experimental horizon. In the second step, the preliminary optimal design vector evaluated in the first step is randomized to provide different initial points for the subsequent multiple shooting optimization. Results of the parameter estimation are given in terms of the a posteriori statistics obtained after performing a maximum likelihood parameter estimation. The quality of the final estimates is assessed by analyzing both the precision and the accuracy of the estimate. Precision is assessed observing for each parameter the interval of estimation confidence and the *t*-value statistic obtained after the optimally

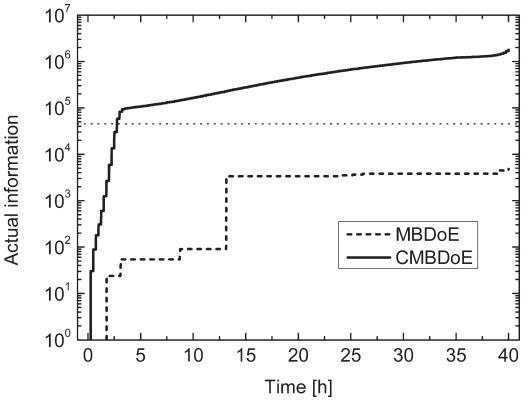

Figure 4. Case study 1: MBDoE vs CMBDoE. Profiles of actual information as given by the summation term of eq 6; the dotted line represents the A-optimal information limit for a 10% deviation on the final estimate.

designed experiments have been executed and model parameters re-estimated with the new data. For a reliable parameter estimation the *t* value must be greater than a computed reference value derived from a Student *t* distribution (*t*-test) with $n_{sp} - N_\theta$ degrees of freedom. Accuracy is assessed by comparing the closeness of the estimate to the true set of model parameters defining the system. In practice, only precision can be assessed in an exhaustive way by a posteriori statistics, since the true set of model parameters is obviously unknown.

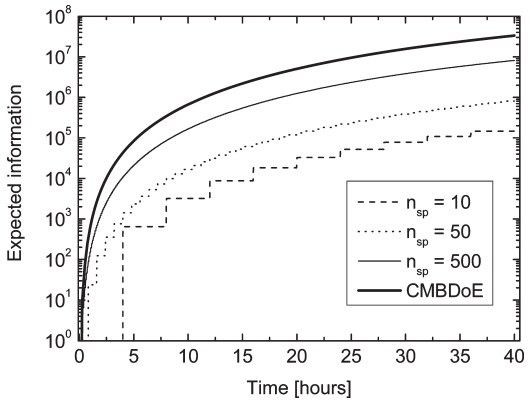

Figure 5. Effect of the number of samples on evaluation of the expected information.

Table 3. Comparison between CPU Times (Pentium D, 3 GHz) in Case Study 1[a]

design formulation	no. of design variables	time ratio
MBDoE	33	1.8
CMBDoE	23	1.0
Discrete CMBDoE	23	12.2

[a] Time ratio is expressed with reference to CMBDoE computational time.

Finally, note that although the CMBDoE criterion is inherently continuous (and therefore an infinite sampling frequency is assumed), estimation of the model parameters must be based on the actual measurements of the outputs, which are always discrete. Accordingly, the parameter estimation procedure retains its standard discrete formulation.

3. CASE STUDY 1: FERMENTATION BIOREACTOR

The MBDoE and CMBDoE methodologies illustrated in the previous sections are here compared and applied to a simulated biomass fermentation process that appeared in several papers on the subject.[24,25] Assuming Monod-type kinetics for biomass growth and substrate consumption, the system is described by the following set of DAEs

$$\frac{dx_1}{dt} = (y - u_1 - \theta_4)x_1 \qquad \frac{dx_2}{dt} = -\frac{yx_1}{\theta_3} + u_1(u_2 - x_2)$$
$$y = \frac{\theta_1 x_2}{\theta_2} + x_2 \qquad (12)$$

where x_1 is the biomass concentration (g/L), x_2 is the substrate concentration (g/L), u_1 is the dilution factor (h^{-1}), and u_2 is the substrate concentration in the feed (g/L). The experimental conditions that characterize an experiment are the dilution factor u_1 (range 0.05−0.20 h^{-1}) and the substrate concentration in the feed u_2 (range 5−35 g/L). These conditions are approximated by piecewise constant profiles over 8 switching intervals (the duration of each interval is allowed to be between 1 and 20 h). The initial biomass and substrate concentration $x_1(0)$ and $x_2(0)$ are set to 1.4 and 0 g/L, respectively. It is assumed that both x_1 and x_2 can be measured during the experiment. The final objective is to design a single experiment (lasting τ = 40 h) to yield the best possible information for estimation of the four parameters θ_i.

Table 4. Case Study 2: Constant Values and Description of Basal Parameters

basal parameters	description	value
$C_{g,b}$	basal glucose concentration in the blood [mg/dL]	81
I_b	basal insulin concentration [mU/L]	15
V_g	glucose distribution volume [dL]	112
V_I	insulin distribution volume [L]	12
R_{ut}	tissue rate of utilization [mg/dL min]	0.75
n	disappearance rate of insulin [min^{-1}]	5/54
u_b	basal insulin infusion rate [mU/min]	16.66

The following settings are considered in order to compare the two different design approaches:

1 MBDoE: standard A-optimal experiment design optimizing the time allocation of n_{sp} = 10 samples (the minimum allowed time between any two consecutive sampling points is 1 h);
2 CMBDoE: continuous experiment design using eq 11 as the objective function; measurements are supposed to be available with Δt = 15 min.

Synthetic "experimental" data are obtained by simulation with $\boldsymbol{\theta}$ = [0.310, 0.180, 0.550, 0.050]^T as the "true" parameters by adding a Gaussian noise to the measurement with a constant relative variance of 0.20. The initial guess for the model parameters values is set to $\boldsymbol{\theta}^0$ = [1.000, 1.000, 1.000, 1.000]^T. For numerical reasons, all parameters have been normalized by dividing them by their true values; from now on, we will always refer to the parameter-normalized values (indicated by symbol $\boldsymbol{\Theta}$).

3.1. Results. In the MBDoE design protocol the system is excited after 7 h from the beginning of the experiment (Figure 2a). The optimal excitation patterns provided by MBDoE are shown in Figure 2b. Results in terms of parameter estimation are reported in Table 1. The estimation is not statistically satisfactory, as MBDoE fails on estimating Θ_4 in a precise way.

When a CMBDoE approach is followed, the system is excited from the very beginning of the experiment (Figure 3a), and this generates a significant difference in the dynamics of the measured outputs (Figure 3b). Results from parameter estimation tasks are reported in Table 2. CMBDoE performs considerably better than the standard design technique, providing a statistically satisfactory estimation for all parameters. Interestingly, the different excitation policy has a major impact on the information dynamics (Figure 4), showing that a significant amount of information can be extracted from the data during the first few hours from the beginning of the experiment. Thus, a different design was performed to reduce the experiment duration and yet to obtain a statistically satisfactory estimation of the model parameters. As reported in Table 2, steering the information dynamics allows concluding the experiment very early (a statistically satisfactory parameter estimation is reached in less than 15 h).

Note that the fact that by increasing the number of measurements the amount of collected information also increases is quite an intuitive concept and is illustrated in Figure 5, where an A-optimal design referred to the information matrix $\mathbf{H}_\theta$ is carried out by increasing the number of regularly spaced sampling points up to the continuous limit given by the CMBDoE. In fact, it can be verified that if the number of sampling points is high enough then the resulting MBDoE design matches the one obtained through a CMBDoE.

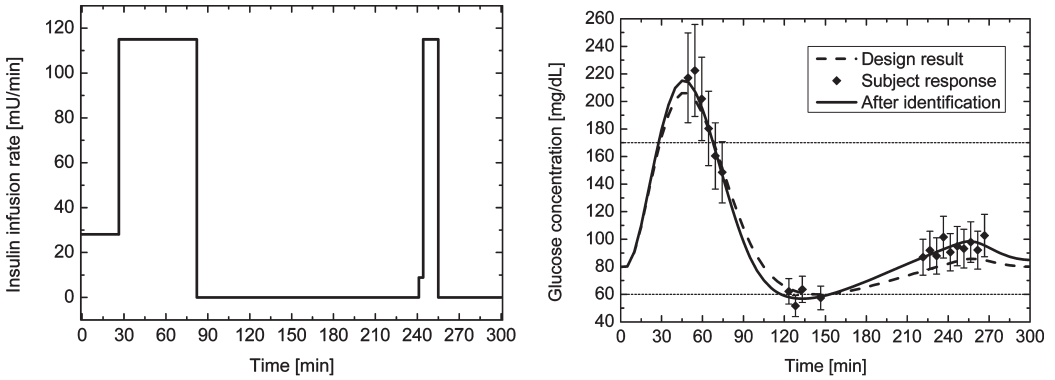

Figure 6. Case study 2: standard design. (a) Optimal profile of the insulin infusion rate. (b) Glucose concentration profiles as predicted by the model during the experiment design (broken line) and after parameter identification (solid line); the subject actual response to the designed experiment is indicated by diamonds. The 95% confidence intervals on measurements are represented by the error bars.

Table 5. Case Study 2: Standard Design—Parameter Estimation and Related a Posteriori Statistics as 95% Confidence Intervals and *t* Values[a]

	MBDoE		
model parameter	final value	confidence interval 95%	95% *t* value
Θ_1	1.0199	0.5414	1.884
Θ_2	0.9612	0.3565	2.696
Θ_3	0.9942	0.3307	3.006

[a] The reference *t* value is 1.74.

One advantage of the CMBDoE formulation is that it expresses this limiting condition (i.e., the maximum obtainable information at the design conditions) without the need of representing a very large (infinite) number of measurements explicitly. Furthermore, one additional advantage is related to computational time. Since in CMBDoE the allocation of sampling times does not need optimizing, the computational burden is reduced by one-half with respect to MBDoE. Additionally, calculations are considerably less expensive if compared to a discrete approach to CMBDoE (i.e., an MBDoE based on an A-criterion on $\mathbf{H}_\theta$ and fixed Δt). Using MBDoE with $\Delta t = 15$ min (i.e., 160 samples) results in a computational time more than 12 times longer than using CMBDoE. Results about computational time are summarized in Table 3.

4. CASE STUDY 2: OPTIMAL DESIGN OF A CLINICAL TEST FOR THE IDENTIFICATION OF A PHYSIOLOGICAL MODEL OF TYPE 1 DIABETES MELLITUS

A simple model of glucose homeostasis[26] is adopted here to describe the blood glucose and insulin concentrations dynamics

$$\frac{dC_g}{dt} = -\theta_1 C_g - X(C_g + C_{g,b}) + D(t) \quad (13)$$

$$\frac{dX}{dt} = -\theta_2 X + \theta_3 I \quad (14)$$

$$\frac{dI}{dt} = -n(I + I_b) + \frac{u(t)}{V_I} \quad (15)$$

where C_g is the blood glucose concentration (mg/dL), X the insulin concentration (mU/L) in the nonaccessible compartment, I the insulin concentration (mU/L), and $u(t)$ the rate of infusion of exogenous insulin (mU/min). The measured response is the subcutaneous glucose concentration C_{gsc}, modeled by the following relationship representing a first-order 5 min lag between C_{gsc} and C_g

$$\frac{dC_{gsc}}{dt} = \frac{(C_g - C_{gsc})}{5} - R_{ut} \quad (16)$$

where R_{ut} is the tissue rate of utilization (mg/dL/min). The meal disturbances model is the one proposed by Hovorka and coworkers[27]

$$D(t) = \frac{2.5At\exp(-0.05t)}{V_g} \quad (17)$$

with A being the amount of carbohydrates of the meal (g_{CHO}); here set to 40 g_{CHO} (fixed). The constant basal parameters are given in Table 4.

The inequality constraint equations imposed on the system are

$$C_1 = y - G_1 \le \gamma_1 \quad (18)$$

$$C_2 = -y + G_2 \le 0 \quad (19)$$

where G_1 and G_2 are the upper (G_1 = 170 mg/dL) and lower (G_2 = 60 mg/dL) thresholds on the subcutaneous glucose concentration C_{gsc}, which is the only variable being constrained (i.e., $y = C_{gsc}$). The lower bound only is a hard constraint not to be violated (it corresponds to hypoglycaemic conditions), while the upper bound has been treated as a "soft" constraint where the relaxation function $\gamma_1(t)$ is chosen in such a way that

$$\int_0^\tau \gamma_1 \, dt \le 165 \text{ mg/dL where}$$

$$\gamma_1 = \begin{cases} 0 & \text{if} \quad C_1 < 0 \\ C_1 & \text{if} \quad C_1 \ge 0 \end{cases} \quad (20)$$

in order to tolerate slight (and short) hyperglycaemic conditions possibly occurring during the first postprandial phase of the test.

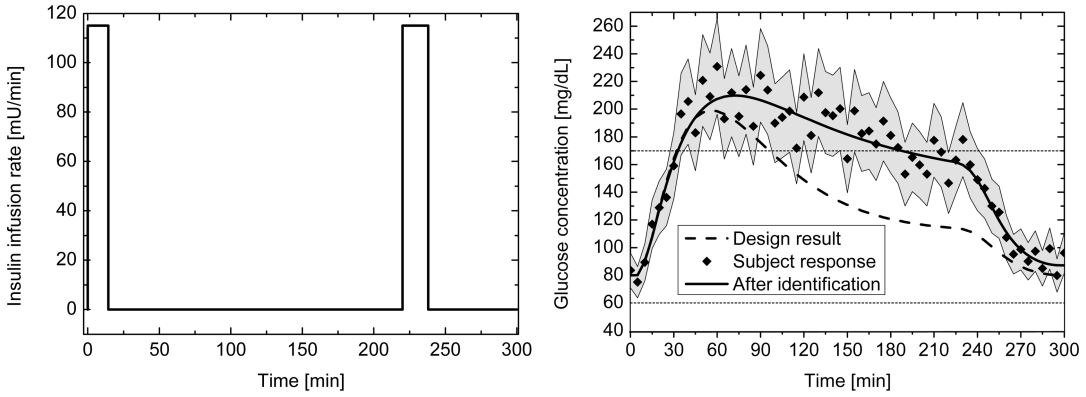

Figure 7. Case study 2: CMBDoE. (a) Optimal profile of the insulin infusion rate and (b) glucose profiles as predicted by the model during the experiment design (broken line) and after parameter identification (solid line); the subject actual response to the designed experiment is indicated by diamonds. The 95% confidence intervals on measurements are represented by the gray shaded area.

Table 6. Case Study 2: CMBDoE—Parameter Estimation and Related a Posteriori Statistics as 95% Confidence Intervals and *t* Values for the Nominal (τ = 5 h) and Minimal Length (τ = 2.8 h) test[a]

	CMBDoE (τ = 5 h)			CMBDoE (τ = 2.8 h)		
model parameter	final value	confidence interval 95%	95% *t* value	final value	confidence interval 95%	95% *t* value
Θ_1	1.0402	0.0381	27.28	0.9563	0.1232	7.76
Θ_2	1.0363	0.1215	8.52	0.6797	0.3971	1.71
Θ_3	1.1285	0.1381	8.17	0.9414	0.2572	3.66

[a] Reference *t* values are 1.66 and 1.70. respectively.

Table 7. Comparison between CPU Times (Pentium D, 3 GHz) in Case Study 2[a]

design formulation	no. of design variables	time ratio
MBDoE	37	6.4
CMBDoE	17	1.0
Discrete CMBDoE	17	10.2

[a] Time ratio is expressed with reference to CMBDoE computational time.

Additional equality constraints are enforced on the final glucose concentration (which must be equal to the basal value of $C_{g,b}$ = 81 mg/dL) and on the final insulin infusion rate (which must be equal to u_b).

The design vector is constituted by the profile of the insulin infusion rate $u(t)$, approximated as a piecewise constant function, with n_{sw} = 8 switching times and n_z = 9 switching levels. Constraints on the design variables are also present: the maximum infusion rate is set to 115 mU/min (and is related to the technical limitation of the insulin infusion pump), and the maximum total amount of insulin that can be administered is set to 8.96 U, a value representing a physiological threshold for the given amount of ingested carbohydrates.

The goal of the study is to assess and compare the effectiveness of two experiment (i.e., clinical test) design strategies in terms of precision of the final $\boldsymbol{\theta}$ estimates:

1 MBDoE: standard A-optimal design of a clinical test optimizing the time allocation of n_{sp} = 20 blood samples available infrequently (Δt = 5−100 min);
2 CMBDoE: continuous design of a clinical test adopting eq 11 as the objective function; frequent measurements are supposed to be available from a CGMS (Δt = 5 min = const.).

Synthetic experimental data are obtained by simulation with $\boldsymbol{\theta}$ = [0.0168 0.0315 1.54*E*-5]^T as the "true" parametric set defining a diabetic subject. Measurements are available with a maximum relative error of 15% on the readings. This high noise level is consistent with the accuracy of recent CGMSs.[28] Uncertainty on the prior estimation of the metabolic parameters is taken into account during the design phase assuming an initial guess on model parameters of $\boldsymbol{\theta}^0$ = [0.0235 0.0189 0.92*E*-5]^T, corresponding to a relative deviation of 40% from the true metabolic parameters $\boldsymbol{\theta}$ describing the physiology of the diabetic subject. Analogously to Case study 1, all parameters have been normalized and indicated by symbol $\boldsymbol{\Theta}$.

4.1. Results. The results for the MBDoE design configuration in terms of optimal insulin infusion rate is illustrated in Figure 6a, while Figure 6b shows the resulting glucose profile. Note that the optimal design does not comply with the constraints on the glucose concentration (in particular, the subject is driven to unsafe hypoglycaemic conditions), and this is due to the parametric mismatch between the subject and the model being identified. The results in terms of parameter estimation are illustrated in Table 5. In this case, a satisfactory parameter estimation is achieved. Even if the estimations of Θ_2 and Θ_3 are far from the true values defining the diabetic subject, the insulin sensitivity S_I (defined by the ratio between Θ_2 and Θ_3 and representing the ability of insulin to enhance plasma glucose disappearance and inhibit hepatic glucose production) is only 3% overestimated.

The test designed by CMBDoE suggests a different profile for the insulin infusion rate (Figure 7a), and a significantly lower

amount of total insulin is administered to the subject (3.71 U vs 7.2 U in MBDoE). As illustrated in Figure 7b, no health-threatening hypoglycaemic event occurs. Although a different design approach should be adopted to avoid the occurrence of the hyperglycaemic transient,[29] it seems that the possibility to exploit the experimental information continuously reduces the need for large variations in the glucose concentration profile and therefore makes CMBDoE more conservative from a safety point of view.

Parameter estimation (Table 6) shows how a significant improvement of the precision in the estimate can be realized when CMBDoE is carried out. As in Case study 1, the different exploitation of information dynamics realized by CMBDoE allows for a significant reduction (to 2.8 h) of the test duration yet ensuring a statistically sound parameter estimation.

In terms of computational burden, the calculation time is reduced 6 times with respect to MBDoE. If a discrete approach to CMBDoE is considered (an MBDoE with Δt = 5 min, i.e., 60 samples), a 10-fold increase in the computational time is obtained. Results about computational time are summarized in Table 7.

5. CONCLUSIONS

A novel design criterion (CMBDoE) for the parametric identification of dynamic systems, suitable for systems where continuous measurements are available, has been proposed and analyzed in this paper. The distinguished feature of CMBDoE is that it designs an experiment by maximizing the experiment information content from the very beginning of the experiment and throughout its entire duration. The problem formulation is such that significant computational time savings can be obtained in execution of the experiment design task. Two simulated case studies have been used to assess the effectiveness of the new technique and compare it with a conventional MBDoE approach where discrete measurements are available: one relates to a bioreactor system and the other to a physiological system. Results have shown a substantial reduction of the overall computational burden. This feature may be particularly useful if CMBDoE is embedded within an online experiment redesign scheme.[30]

AUTHOR INFORMATION

Corresponding Author

*Tel: +39 049 827 5468. Fax: +39 049 827 5460. E-mail: fabrizio.bezzo@unipd.it.

ACKNOWLEDGMENT

The authors gratefully acknowledge the financial support granted to this work by the University of Padova under Project CPDR095313-2009 (Assegni di ricerca 2009) on "Towards the development of an artificial pancreas for diabetes mellitus care: optimal model-based design of experiments for parameter identification of physiological models".

NOMENCLATURE

General symbols

A = glucose amount of the meal
C_g = glucose concentration in the blood
$C_{g,b}$ = basal glucose concentration in the blood
C_{gsc} = subcutaneous glucose concentration
D = meal disturbance function
f = differential and algebraic system implicit function
g = measurements selection function
I = insulin concentration in the accessible compartment
I_b = insulin basal value
m_{ij} = ijth element of $\mathbf{M}$
n = disappearance rate of insulin
n_q = number of quadrature points
n_{sp} = number of samples
N_u = number of manipulated inputs
N_x = number of state variables
N_w = number of time invariant controls
N_y = number of measured variables
N_θ = number of model parameters
N_c = number of constraints
N_g = number of inequality constraints
n_{sw} = number of switching levels
n_φ = number of design variables
R_{ut} = rate of glucose utilization
S_I = insulin sensitivity
s_{ij} = ijth element of the inverse matrix of measurements errors
t = time
t_i = ith t value
u = insulin infusion rate
u_b = time-invariant basal insulin infusion rate
V_I = insulin distribution volume
x = generic state variable
X = insulin concentration in the nonaccessible compartment
y = generic measured output
α = statistical significance factor
φ_i = ith element of the design vector
γ = relaxation function
θ_i = ith model parameter
Θ_i = ith normalized model parameter
τ = test duration
ψ = $\mathbf{V}_\theta$ measurement function

Vectors and Matrices [dimension]

$\mathbf{C}$ = set of constraint functions $[N_c]$
$\mathbf{G}$ = set of active constraints $[N_c]$
$\mathbf{H}_\theta$ = dynamic information matrix $[N_\theta \times N_\theta]$
$\mathbf{H}_\theta^0$ = preliminary information matrix $[N_\theta \times N_\theta]$
$\mathbf{L}$ = time-dependent information matrix of $[N_\theta \times N_\theta]$
$\mathbf{M}$ = time-dependent information matrix $[N_\theta \times N_\theta]$
$\mathbf{M}_k$ = information matrix of the kth sample $[N_\theta \times N_\theta]$
$\mathbf{y}_0$ = vector of initial conditions $[N_y]$
$\mathbf{y}$ = measurements vector $[N_y]$
$\mathbf{y}$ = vector of estimated responses $[N_y]$
$\mathbf{t}^{sp}$ = vector of sampling points $[n_{sp}]$
$\mathbf{u}$ = vector of manipulated inputs $[N_u]$
$\mathbf{V}_\theta$ = variance-covariance matrix of model parameters $[N_\theta \times N_\theta]$
$\mathbf{w}$ = vector of time-invariant control $[N_w]$
$\mathbf{x}$ = vector of state variables $[N_x]$
$\mathbf{x}^0$ = vector of initial states $[N_x]$
$\mathbf{x}$ = vector of derivatives on state variables $[N_x]$
$\boldsymbol{\varphi}$ = design vector $[n_\varphi]$
$\boldsymbol{\theta}$ = vector of values of true model parameters for the subject/system $[N_\theta]$
$\hat{\boldsymbol{\theta}}$ = vector of estimated values of model parameters $[N_\theta]$
$\boldsymbol{\theta}^0$ = vector of initial guesses of model parameters $[N_\theta]$
$\boldsymbol{\Theta}$ = vector of normalized model parameters for the subject $[N_\theta]$

Acronyms
CGMS = continuous glucose monitoring system
CHO = carbohydrates
CMBDoE = continuous model-based design of experiments
DAE = differential and algebraic equations
MBDoE = model-based design of experiments

REFERENCES

(1) Mehra, R. K. Optimal input signals for parameter estimation in dynamic systems: survey and new results. *IEEE Trans. Autom. Control* **1974**, *19*, 753–768.

(2) Goodwin, G. C.; Payne, R. L. *Dynamic system identification: experiment design and data analysis*; Academic Press: New York, 1977.

(3) Chen, B. H.; Bermingham, S.; Neumann, A. H.; Kramer, H. J. M.; Asprey, S. P. On the design of optimally informative experiments for dynamic crystallization process modeling. *Ind. Eng. Chem. Res.* **2004**, *43*, 4889–4902.

(4) Bernaerts, K.; Van Impe, J. F. Optimal dynamic experiment design for estimation of microbial growth kinetics at sub-optimal temperatures: modes of implementation. *Simul. Modell. Pract. Theory* **2005**, *13*, 129–138.

(5) Franceschini, G.; Macchietto, S. Validation of a model for biodiesel production through model-based experiment design. *Ind. Eng. Chem. Res.* **2007**, *46*, 220–232.

(6) Prasad, V.; Vlachos, D. G. Multiscale model and informatics-based optimal design of experiments: application to the catalytic decomposition of ammonia on ruthenium. *Ind. Eng. Chem. Res.* **2008**, *47*, 6555–6567.

(7) Galvanin, F.; Barolo, M.; Macchietto, S.; Bezzo, F. Optimal design of clinical tests for the identification of physiological models of type 1 diabetes mellitus. *Ind. Eng. Chem. Res.* **2009**, *48*, 1989–2002.

(8) Surisetty, K.; De la Hoz Siegler, H.; McCaffrey, W. C.; Ben-Zvi, A. Model re-parameterization and output prediction for a bioreactor system. *Chem. Eng. Sci.* **2010**, *65*, 4535–4547.

(9) Franceschini, G.; Macchietto, S. Model-based design of experiments for parameter precision: state of the art. *Chem. Eng. Sci.* **2008**, *63*, 4846–4872.

(10) Zullo, L. *Computer aided design of experiments. An engineering approach*; The University of London: U.K., 1991.

(11) Jacques, J. A.; Greif, P. Numerical parameter identifiability and estimability: integrating identifiability, estimability, and optimal sampling design. *Math. Biosci.* **1985**, *11*, 20l–227.

(12) Tosi, S.; Rossi, M.; Tamburini, E.; Vaccari, G.; Amaretti, A.; Matteuzzi, D. Assessment of in-line near-infrared spectroscopy for continuous monitoring of fermentation processes. *Biotechnol. Prog.* **2008**, *19*, 1816–1821.

(13) Dias, J. M. L.; Pardelha, F.; Eusebio, M.; Reis, M. A. M.; Oliveira, R. On-line monitoring of PHB production by mixed microbial cultures using respirometry, titrimetry and chemometric modelling. *Process Biochem.* **2009**, *44*, 419–427.

(14) Mazze, R. S.; Strock, E.; Borgman, S.; Wesley, D.; Stout, P.; Racchini, J. Evaluating the accuracy, reliability and clinical applicability of continuous glucose monitoring (CGM): is CGM ready for real time? *Diabetes Technol. Ther.* **2009**, *11*, 11–18.

(15) Pukelsheim, F. *Optimal Design of Experiments*; J. Wiley & Sons, New York, 1993.

(16) Galvanin, F.; Macchietto, S.; Bezzo, F. Model-based design of parallel experiments. *Ind. Eng. Chem. Res.* **2007**, *46*, 871–882.

(17) Zhang, Y.; Edgar, T. F. PCA combined model-based design of experiments (DOE) criteria for differential and algebraic system param eter identification. *Ind. Eng. Chem. Res.* **2008**, *47*, 7772–7783.

(18) Barz, T.; Arellano-Garcia, H.; Wozny, G. Handling uncertainty in model-based optimal experimental design. *Ind. Eng. Chem. Res.* **2010**, *49*, 5702–5713.

(19) Fuller, W. A. *Measurement error models*; J. Wiley & Sons, New York, 2006.

(20) Chase, J. G.; Shaw, G. M.; Wong, X. W.; Lotz, T.; Lin, J.; Hann, C. E. Model-based control in critical care —A review of the state of the possible. *Biomed. Signal Process. Contr.* **2006**, *1*, 3–21.

(21) Hochwald, B.; Nehorai, A. On identifiability and information-regularity in parameterized normal distributions. *Circuits Syst. Signal Process.* **1997**, *16*, 83–89.

(22) Rice, J. R. A Metalgorithm for adaptive quadrature. *J. ACM* **1975**, *22*, 61–82.

(23) Bock, H.; Kostina, E.; Phu, H. X.; Rannacher, R. *Modeling, simulation and optimization of complex processes*; Springer: Berlin, 2003.

(24) Espie, D.; Macchietto, S. The optimal design of dynamic experiments. *AIChE J.* **1989**, *35*, 223–229.

(25) Asprey, S. P.; Macchietto, S. Statistical tools for optimal dynamic model building. *Comput. Chem. Eng.* **2000**, *24*, 1261–1267.

(26) Lynch, S. M.; Bequette, B. W. *Model predictive control of blood glucose in type I diabetics using subcutaneous glucose measurements. Proceedings of the 2002 American Control Conference*, Anchorage, AK, 2002; pp 4039—4043.

(27) Hovorka, R.; Canonico, V.; Chassin, L. J.; Haueter, U.; Massi-Benedetti, M.; Federico, M. O.; Pieber, T. R.; Schaller, H. C.; Schaupp, L.; Vering, T.; Wilinska, M. E. Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes. *Physiol. Meas.* **2004**, *25*, 905–920.

(28) Hann, C. E.; Chase, J. G.; Desaive, T.; Ypma, M. F., Elfring, J.; Shaw, G. M. The impact of model-based therapeutics on glucose control in an intensive care unit. *IFMBE Proceedings*, Antwerp, Belgium, 2009; pp 1570—1573.

(29) Galvanin, F.; Barolo M.; Macchietto S.; Bezzo F. Optimal design of clinical tests for the identification of physiological models of type 1 diabetes mellitus in the presence of model mismatch. *Med. Biol. Eng. Comput.*, **2010**, in press (DOI: 10.1007/s11517-010-0717-8).

(30) Galvanin, F.; Barolo, M.; Bezzo, F. Online model-based re-design of experiments for parameter estimation in dynamic systems. *Ind. Eng. Chem. Res.* **2009**, *48*, 4415–4427.