A Backoff Strategy for Model-Based Experiment Design Under Parametric Uncertainty

Federico Galvanin
DIPIC - Dipartimento di Principi e Impianti di Ingegneria Chimica, Università di Padova, via Marzolo 9 I-35131 Padova PD, Italy

Dept. of Chemical Engineering, Imperial College London, South Kensington Campus, London SW7 2AZ, U.K.

Massimiliano Barolo and Fabrizio Bezzo
DIPIC - Dipartimento di Principi e Impianti di Ingegneria Chimica, Università di Padova, via Marzolo 9 I-35131 Padova PD, Italy

Sandro Macchietto
Dept. of Chemical Engineering, Imperial College London, South Kensington Campus, London SW7 2AZ, U.K.

DOI 10.1002/aic.12138
Published online December 23, 2009 in Wiley InterScience (www.interscience.wiley.com).

Model-based experiment design techniques are an effective tool for the rapid development and assessment of dynamic deterministic models, yielding the most informative process data to be used for the estimation of the process model parameters. A particular advantage of the model-based approach is that it permits the definition of a set of constraints on the experiment design variables and on the predicted responses. However, uncertainty in the model parameters can lead the constrained design procedure to predict experiments that turn out to be, in practice, suboptimal, thus decreasing the effectiveness of the experiment design session. Additionally, in the presence of parametric mismatch, the feasibility constraints may well turn out to be violated when that optimally designed experiment is performed, leading in the best case to less informative data sets or, in the worst case, to an infeasible or unsafe experiment. In this article, a general methodology is proposed to formulate and solve the experiment design problem by explicitly taking into account the presence of parametric uncertainty, so as to ensure both feasibility and optimality of the planned experiment. A prediction of the system responses for the given parameter distribution is used to evaluate and update suitable backoffs from the nominal constraints, which are used in the design session to keep the system within a feasible region with specified probability. This approach is particularly useful when designing optimal experiments starting from limited preliminary knowledge of the parameter set, with great improvement in terms of design efficiency and flexibility of the overall iterative model development scheme. The effectiveness of the proposed methodology is demonstrated and discussed by simulation through two illustrative case studies concerning the parameter identification of physiological models related to diabetes and cancer care. © 2009 American Institute of Chemical Engineers *AIChE J,* 56: 2088–2102, 2010

Keywords: model-based design of experiments, feasibility under uncertainty, backoff strategy, mathematical modelling, biomedical engineering

Correspondence concerning this article should be addressed to S. Macchietto at s.macchietto@imperial.ac.uk.

Introduction

Simulation, design, control, and optimization of chemical processes rely on the availability of detailed mathematical models. As a large class of phenomena can be represented by systems of differential and algebraic equations (DAEs), the goal of every model building procedure is to identify both the model structure and the model parameters to represent the underlying phenomena in the most reliable and accurate way. Modern model-based design of experiments (MBDoE) techniques[1] represent a powerful tool for the rapid development and assessment of dynamic deterministic models, yielding the most informative set of data to collect from an experimental session to estimate precisely the parameter set of a given model (MBDoE for parameter estimation) or to find, among several candidates, an adequate model structure (MBDoE for model discrimination). These techniques allow for detection of the ''best'' experimental conditions to maximize the information content of the experimental runs, and their effectiveness has been demonstrated in several areas.[2–8] Specific user-friendly software has also been proposed for model-based experiment design.[9]

The optimal experiment design problem for improving parameter estimation involves the maximization of a measure of the expected information (i.e., the prediction of the information that will be gained from the experiment, usually evaluated from the Fisher information matrix[10]) by acting on the experiment decision variables. The technique is usually embodied in a sequence of three key activities (experiment design, experiment execution, parameter estimation)[11] and is particularly flexible, allowing for the definition of a set of active constraints on both state and design variables during the optimization.[12] The goal of a constrained MBDoE is to achieve both optimality (maximization of the expected information) and feasibility (no constraint violations) during the experimental trials. As the methodology is model based, both model mismatch (i.e., a model structure inadequate to represent the physical systems) and parametric mismatch (i.e., incorrect values of the parameters) may affect the consistency of the whole design procedure.[13] Despite the importance of ensuring optimally informative as well as feasible experiments, relatively little work has been done to develop a model-based experiment design technique capable of overcoming both of the aforementioned issues.

In the topic of process systems design, the problem of constrained optimization under uncertainty, seen as a tradeoff between feasibility and optimality, has long been recognised as a key issue,[14] because the presence of both variations in the operating conditions and uncertainty in the process model (in terms of process model parameters and mathematical structure) deeply affects the optimality of process and equipment design.[15] Several approaches have been proposed to solve the process design problem in the presence of parametric uncertainty, where uncertain parameters are described by probability distribution functions and the design problem is formulated using probabilistic decision criteria.

Several works have appeared where the issue has been tackled through a robust implementation based on the solution of a max-min optimization problem (worst case approach).[16,17] In this way the design solution (formally an ''overdesign'') represents the best decision based on the actual knowledge on the process. Different methodologies have been proposed to relax the worst case assumption, where an expected value approach is used to increase the design feasibility.[18–21] Somehow, a similar route was considered by Monningmann and Marquardt,[22] who proposed a robust optimization approach to guarantee feasibility and stability over the expected range of variation introducing conservatism to handle parametric uncertainty.

In fact, as devised by Chachuat et al.,[23] in the presence of model uncertainty, feasibility is often of greater importance than optimality. To tackle this issue, a more tailored strategy is to enforce feasibility by the presence of backoffs from active constraints. In the backoff approach (which also may be implemented according to a worst-case approach[24] or to an expected value approach,[25] or defining the magnitude of the output variation[26]), the actual operating point is moved away from the nominal operating point to ensure feasibility of the process to compensate for the effect of disturbances.

Other formulations have been proposed to solve specific operational issues like flexibility (i.e., the ability of the process to preserve feasibility in the presence of uncertainties[27,28]), robustness (i.e., the ability to preserve optimality conditions for disturbances in the inputs[29]), controllability (i.e., the ability of the system to recover from process disturbances or dynamic plant behavior[30]), economic performance (i.e., the choice of the compromise between feasibility and optimality in terms of the economy of the process itself[31]) and the integration of some of the aforementioned issues.[32,33]

From an MBDoE perspective, robust techniques for optimal experimental design have been proposed in literature[34,35] to preserve the optimality of the design in the presence of parametric uncertainty, either through a worst case approach or performing a dynamic optimization over all the predicted uncertainty region of model parameters (expected value approach). Rustem and Zakovic[36] proposed a semi-infinite programming algorithm to solve the global optimization design and the feasibility problems in parallel, with great benefit in terms of computational time saving; in this case, the robust-constrained MBDoE problem was solved with constraints on the design variables only. Rojas et al.[37] proposed a min-max approach to solve the robust optimal design problems with simple constraints on the manipulated inputs. Interestingly, the authors also compare different design criteria linking robust control techniques[38] and nominal experimental design procedure. Chu and Hahn[39] proposed a technique to integrate optimal parameters selection with experimental design under parametric uncertainty for nonlinear dynamic systems. The robust design was performed by adopting a hybrid method combining a genetic algorithm and a stochastic approximation technique.

However, to the best of our knowledge a framework for explicitly taking into account the feasibility issue within an MBDoE approach has not been presented so far. In this article, a methodology is illustrated and discussed to address the problem of the constrained optimal experimental design under parametric uncertainty. Similarly to what was successfully proposed in other fields (and discussed in the above), a backoff policy is adopted that allows guaranteeing the feasibility of the optimally designed experiment in the presence of parametric uncertainty. The technique is particularly

suitable for planning experiments in such systems (for example, physiological systems or reactive systems) where the operability is strictly reduced by the presence of active constraints on state variables that are inherently related to the physical system. The proposed technique is illustrated and discussed through two simulated case studies concerning parameter identification in physiological models related to the care of diabetes mellitus and of cancer.

Problem Statement

Let us assume that in an existing system a number of outputs can be measured and the vector $\mathbf{y}(t) \in \Re^{N_y}$ represents the measured values of the outputs. Also, it is assumed that the system can be described by a dynamic deterministic model constituted by a set of index-1 DAEs of the form

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

with the set of initial conditions $\mathbf{x}(0) = \mathbf{x}_0$, subject to

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

where $\mathbf{x}(t) \in \Re^{N_x}$ is the vector of time-dependent state variables, $\mathbf{u}(t) \in \Re^{N_u}$ and $\mathbf{w} \in \Re^{N_w}$ are, respectively, the time-dependent and time-invariant control variables (manipulated inputs), $\hat{\theta} \in \Re^{N_\theta}$ is the set of the parameters values used in the model, $\hat{\mathbf{y}}(t)$ is the vector of the output values estimated by the model, and t is time. $\mathbf{C}$ is an N_c-dimensional set of constraint functions expressed through the set $\mathbf{G}(t) \in \Re^{N_c}$ of (possibly time-varying) active constraints on the state variables.

Model-based experimental design procedures aim at decreasing the model parameter uncertainty region by acting on the design vector $\varphi \in \Re^{n_\varphi}$:

$$\varphi = \left\{\mathbf{y}_0, \mathbf{u}(t), \mathbf{w}, \mathbf{t}^{\mathrm{sp}}, \tau\right\}, \tag{3}$$

or equivalently

$$\varphi = \begin{bmatrix} \varphi_1 & \varphi_2 & \ldots & \varphi_{n_\varphi} \end{bmatrix}^{\mathrm{T}}, \tag{4}$$

where $\mathbf{y}_0$ is the set of initial conditions of the measured variables, and τ is the duration of an experiment. The set of time instants at which the output variables are sampled is a design variable itself, and is expressed through the vector $\mathbf{t}^{\mathrm{sp}} = [t_1 \ \ldots \ t_{n_{\mathrm{sp}}}]^{\mathrm{T}}$ of n_{sp} sampling times. Control vector parameterization techniques[40] are used to discretize the control input $\mathbf{u}$ profiles. These profiles can be approximated as piecewise constant, piecewise linear, or polynomials functions over a predefined number of intervals. The optimal design under constraints problem can be formulated as finding:

$$\varphi = \arg\min \psi\left(\mathbf{V}_\theta\left(\hat{\theta}, \varphi\right)\right) \tag{5}$$

subject to $\mathbf{C}$, where ψ is a metric of the variance-covariance matrix of model parameters $\mathbf{V}^\theta$ expressing the selected design criterion (A-, D-, E-optimal,[10] SV-based[41] or P-based[42]). In addition to (2), a n_φ-dimensional set of constraints on the design variables may be present, too, usually expressed as

$$\varphi_i^l \leq \varphi_i \leq \varphi_i^u \qquad i = 1 \ldots n_\varphi \tag{6}$$

with lower (superscript l) and upper (superscript u) bounds on the elements of φ, constraining the design to a hyper-rectangular sub-space of the overall design space $\Re^{n_\varphi}$. For a single experiment the variance-covariance matrix of the model parameters is

$$\mathbf{V}_\theta\left(\hat{\theta}, \varphi\right) = \left[\Sigma_\theta^{-1} + \sum_{i=1}^{N_y}\sum_{j=1}^{N_y} s_{ij}\mathbf{Q}_i^T\mathbf{Q}_j\right]^{-1} = \left[\Sigma_\theta^{-1} + \mathbf{H}_\theta\left(\hat{\theta}, \varphi\right)\right]^{-1} \tag{7}$$

where s_{ij} is the ij-th element of the inverse of the $(N_y \times N_y)$ estimated variance-covariance matrix Σ of measurement errors, and $\mathbf{Q}_i$ is the matrix of the sensitivity coefficients for the i-th estimated output with respect to the model parameters at each of the n_{sp} sampling points:

$$\mathbf{Q}_i = \left[\frac{\partial \hat{y}_i(t_l)}{\partial \theta_m}\right] \qquad l = 1, \ldots, n_{\mathrm{sp}} \qquad m = 1, \ldots, N_\theta. \tag{8}$$

The matrix $\mathbf{H}_\theta$ is the N_θ-dimensional Fisher information matrix. Prior information on the model parameter uncertainty region in terms of a statistical distribution (for instance, a uniform or Gaussian distribution) can be included through matrix Σ_θ.

The solution to the constrained MBDoE optimization problem is the optimal design vector φ that through model (1) simulataneously satisfies the design optimality condition (5), the feasibility constraints on the state variables (2) and the constraints on the design variables (6).

Note that both the optimality and the feasibility conditions are evaluated at the current estimated values $\hat{\theta}$ of model parameters, which is different from the true (and unknown) value of model parameters θ. The parametric mismatch affects both the optimality condition (5) and the feasibility condition (2) as well as the constraints in (6). Prior knowledge on the physical system (in particular of the sources of uncertainty) and a preliminary analysis of the model around a set of nominal experimental conditions may help to define the boundaries of an "expected" uncertainty region of model parameters. Hyper-rectangular uncertainty regions are frequently used[19,32] but, as suggested by Rooney and Biegler,[43] the adoption of nonlinear confidence regions derived from the likelihood ratio test leads to a more accurate representation of the uncertainty.

To predict the effect of parametric uncertainty on the optimality and feasibility conditions, a stochastic approach may be adopted, taking into account all the possible realizations of the parameter vector elements over all the (expected) uncertainty. In the stochastic approach, the entire set of possible realizations of θ has to be defined through some probabilistic assumptions, concerning the type of distribution and the deviation metrics from the current estimate of the model parameters, $\hat{\theta}$. In this perspective, the set of model parameters can be regarded as a stochastic variable (symbol $\sim$), $\tilde{\theta}$, i.e., a function considering all the possible realizations from

the N_θ-dimensional expected uncertainty of model parameters $\boldsymbol{T}$ to the N_θ-dimensional field of real numbers.

In the presence of parametric uncertainty, the solution of the constrained MBDoE problem is not deterministic and the solution of the optimal design under constraints problem is a stochastic design vector $\tilde{\varphi}$ (i.e., a set of possible realizations of the optimal design vector) satisfying at the same time the model equations, the design optimality condition and the feasibility constraints on both state and design variables. The stochastic design vector represents the field of optimal and feasible solutions of the design problem in the presence of parametric uncertainty where the optimal design problem under constraints is solved over $\boldsymbol{T}$. Considering that, unless a large number of identical experimental facilities is available, only one experiment can be performed at a time, the above general formulation needs simplifying so that a unique feasible optimal solution for the experiment design problem is found. The design objective function (5) can be evaluated, adopting a conservative approach by considering an expected value or worst case metric for $\mathbf{V}_\theta$ (emphasizing robustness, as is done in Asprey and Macchietto[35]), or (as is done in this work) at the actual information point (emphasizing optimality). Thus, the following set of equations has to be solved

$$\varphi = \arg\min \psi\left(\mathbf{V}_\theta\left(\hat{\theta}, \varphi\right)\right) \tag{5}$$

subject to

$$\mathbf{f}(\dot{\tilde{\mathbf{x}}}(t), \tilde{\mathbf{x}}(t), \mathbf{u}(t), \mathbf{w}, \tilde{\theta}, t) = 0 \tag{9}$$

$$\hat{\mathbf{y}}(t) = \mathrm{g}(\mathbf{x}(t)) \tag{10}$$

$$\mathbf{C} = \tilde{\mathbf{x}}(t) - \mathbf{G}(t) + \beta\left(\dot{\tilde{\mathbf{x}}}(t), \tilde{\mathbf{x}}(t), \mathbf{u}(t), \mathbf{w}, \tilde{\theta}, t\right) \leq 0 \tag{11}$$

$$\varphi_i^l \leq \varphi_i \leq \varphi_i^u \quad i = 1...n_\varphi \tag{6}$$

where β is a N_c-dimensional set of time-dependent backoff functions taking into account the effect of parametric uncertainty on the state variables at the designed experimental conditions. Note that β is a function of a subset of the design vector φ (i.e., of $\mathbf{u}$ and $\mathbf{w}$ only). With some abuse of notation the stochastic vector of model outputs is still indicated as $\hat{y}(t)$.

The adoption of a backoff strategy allows to satisfy the stochastic feasibility condition

$$\tilde{\mathbf{C}} = \tilde{\mathbf{x}}(t) - \mathbf{G}(t) \leq 0 \tag{12}$$

where the effect of parametric uncertainty is taken into account exclusively through the backoff vector β. As the backoff vector is a function of stochastic variables, a stochastic simulation approach has been adopted in this article. The stochastic simulation procedure for backoff vector evaluation consists of three key steps:

1. Characterization of the parametric uncertainty: some assumptions have to be made on the multidimensional uncertainty domain $\boldsymbol{T}$ of model parameters and a reliable sampling of $\boldsymbol{T}$ has to be carried out.

2. Mapping the uncertainty region of the state variables: in our approach, several simulations are carried out adopting random values for model parameters and a subsequent statistical analysis of the profiles of state variables is used to provide a probabilistic description of the uncertainty region of the state variables.

3. Backoff formulation and policy: starting from the description of the uncertainty region of the state variables, the user can build the set of backoff functions in (11). Note that the experimenter's decisions may deeply affect the fulfilment of (12) because some constraints could be enforced or relaxed through a backoff policy.

The steps involved in the stochastic simulation approach for the backoff building are analysed in detail in the following subsections.

Characterization of the Parameter Uncertainty

If a probability function is associated with the expected parameter uncertainty domain, then it is possible to define $\boldsymbol{T}$ as

$$T = \left[\tilde{\theta}_{ij} \middle| \tilde{\theta}_{ij} \in p_\theta\left(\hat{\theta}, \xi_\theta\right), i = 1...N_\theta, j = 1...N\right] \tag{13}$$

where ξ_θ is a n_{ξ_θ}-dimensional vector of parameters defining the specific probability distribution p_θ, $\tilde{\theta}_{ij}$ is the realization of the i-th element of the parameters vector in the j-th event and N is the population abundance. Note that the realizations of $\tilde{\theta}$ could be either independently distributed or correlated random variables, coming either from a joint probability distribution or from a set of univariate probability distributions. A sampling of the expected uncertainty domain $\boldsymbol{T}$ needs to be carried out to assess the effect of the possible realizations of the unknown parametric set on the state variables of the model. Different sampling methods can be used at the purpose.[44]

A critical aspect of the sampling procedure is the choice of the number N' of samples of the parameters probability distribution. For one random variable,[45] the following formula can be used, according to the central limit theorem

$$N' = \frac{t'\hat{\sigma}^2}{\varepsilon'^2} \tag{14}$$

with t' the t-value for the selected confidence level α (set by the experimenter), ε' the error that the experimenter is willing to expect and $\hat{\sigma}^2$ the expected variance value. No general formulas for N' to define an appropriate sampling in multivariate distributions are available, and a multivariate statistical analysis of the sampled region is highly recommended.[46] Chao[47] proposed a principal component analysis (PCA) method for sampling from multivariate distributions to summarize most of the variability using the principal components with highest variance. Global sensitivity analysis (GSA) methods involving Fourier amplitude sensitivity test (FAST), Sobol,[48] or the more computationally efficient derivative-based global sensitivity measures (DGSM) techniques[49] could be useful to detect the most relevant subsets of model parameters, allowing to decrease the sampling size of the analysis. The main drawback is that GSA is usually evaluated at some specified experimental conditions and, because of the computational effort, it is difficult to integrate it in a MBDoE optimization framework.

Mapping the uncertainty region of the state variables

After sampling the space of uncertain model parameters $\boldsymbol{T}$, a stochastic simulation is carried out where model (1) is

solved repeatedly for the entire subset of possible realizations of the parametric uncertainty. The goal of the stochastic simulation is to evaluate a set of time-dependent statistical parameters $\xi_x(t)$ describing $p_{x|\varphi}(\xi_x(t),t)$, the probability distribution of the state variables in the presence of parametric mismatch at the experimental settings defined by an assigned φ. As the state variables are usually correlated, $p_{x|\varphi}(\xi_x(t),t)$ generally defines a time-dependent joint confidence region of the state variables. A set of N' simulations based on the N' sampled values of the model parameters around the nominal point $\hat{\theta}$ at the experimental conditions φ is carried out, generating the N'-dimensional set of dynamic responses, which are collected in a $N' \times N_y$ time-dependent matrix $\mathbf{X}$. The problem of mapping the uncertainty region of the state variables can be interpreted as finding the n_{ξ_x}-dimensional set of time-dependent parameters $\xi_x(t)$ that are specific for describing the given distribution (e.g., for a normal distribution, $n_{\xi_x} = 2$ and the distribution parameters are the vector of average values and the variance matrix of model parameters).

It must be pointed out that:

- the number of simulations might be sufficient for a complete description of p_θ, but not of $p_{x|\varphi}(\xi_x(t),t)$ as the model is nonlinear and the two distributions are usually different;
- the evaluation of $\mathbf{X}$ is computationally expensive, involving the repeated numerical integration of a nonlinear differential system.

We define the N_x-dimensional vector of average responses $\bar{\mathbf{x}}$ as

$$\bar{\mathbf{x}}(t) = \frac{\sum_{i=1}^{N'} \mathbf{X}_i(t)}{N'} \tag{15}$$

and the N_x-dimensional variance vector is

$$\sigma_x^2(t) = \frac{\sum_{i=1}^{N'} (\mathbf{X}_i(t) - \bar{\mathbf{x}})^2}{N' - 1}. \tag{16}$$

In the hypothesis of (i) independence and identical distribution of the responses for each x_i trajectory after random sampling on θ during the whole experimental horizon (i.e., each x_i trajectory belongs to the same kind of distribution at a given experimental time and can be treated as a purely random variable), (ii) finite variance of the model responses, (iii) N' being a sufficiently large number of simulations, and (iv) linear correlation between θ and x_i, then it is possible to apply the central limit theorem. Under those assumptions the x_i's can be considered normally distributed with mean $\bar{x}_i$ and standard deviation $\sigma_{x,i}$. The basic idea is to capture the overall uncertainty of state variables through a mean-variance regression model whose responses can be represented by mean profiles and deviations from the mean profiles. This approach resembles the one used in nonlinear optimization under uncertainty (e.g., Darlington et al.[50]) and robust design through metamodeling (e.g., Apley et al.[51]). For a normal distribution the confidence intervals κ for a $(1 - \alpha) = 95\%$ and $(1 - \alpha) = 99.7\%$ confidence levels can be easily approximated by the following expressions

$$\kappa^{95\%} \simeq 2\sigma_x = 2\sqrt{\frac{\sum_{i=1}^{N'} (\mathbf{X}_i - \bar{\mathbf{x}})^2}{N' - 1}} \tag{17}$$

$$\kappa^{99.7\%} \simeq 3\sigma_x = 3\sqrt{\frac{\sum_{i=1}^{N'} (\mathbf{X}_i - \bar{\mathbf{x}})^2}{N' - 1}} \tag{18}$$

The normal distribution usually provides a starting point for the evaluation of the shape of the actual $p_{x|\varphi}(\xi_x(t),t)$ distribution that, in practice, because of a nonlinear correlation between x_i and θ, might present some peculiarities:

1. Different dispersion around the vector of mean values;
2. Asymmetric dispersion around a critic value (skewed distribution) with consequent inconsistency of the standard normality assumption.

To overcome these issues, more complex distributions may be considered (skew normal, Weibull, multivariate normal, Rosin-Rammler, bimodal, etc.).

Backoff formulation and policy

Once a predicted uncertainty region of the state variables is defined, the backoff vector β of (12) can be approximated by

$$\mathbf{C} = \mathbf{x}(t) - \mathbf{G}(t) + \beta\big(p_{x|\varphi}(\xi_x(t),t),\alpha,t\big) \le 0 \tag{19}$$

where the N_c-dimensional time-varying backoff vector is a function of the probability distribution of the state variables at the experimental settings φ, and of a confidence vector α. The confidence vector can be set by the experimenter to tune the backoff from the active constraints $\mathbf{G}$. Prior information on the system and convenience factors (e.g., the constraints on some state variables might be relaxed or enforced according to their relative importance) can guide the choice of the proper confidence vector. One possible backoff formulation is:

$$\beta\big(p_{x|\varphi}(\xi_x(t),t),\alpha,t\big) = \kappa \tag{20}$$

where the backoff takes into account the overall $(1-\alpha)\%$ uncertainty region of state variables at the nominal conditions. To increase the flexibility of the backoff strategy it is also possible to adopt the expression

$$\beta\big(p_{x|\varphi}(\xi_x(t),t),\alpha,\Lambda,t\big) = \Lambda\kappa \tag{21}$$

where Λ is a N_c-dimensional vector of coefficients larger than 1, used to increase conservatism. Through (21), the experimenter could always favor one direction of the variability instead of another, thus guiding the backoff policy. It must be pointed out that both backoff formulations (20) and (21) do not depend on the closeness to the active constraints, but on the predicted uncertainty region of the state variables only. More complex formulations may include the backoff action in the region of possible constraints violation only (i.e., for all the possible realizations of $\hat{\theta} \in \boldsymbol{T}$ at given φ).

Integration of the Stochastic Information: MBDoE with Backoff Algorithm

The stochastic approach for backoff building described in the previous section needs to be integrated into a constrained

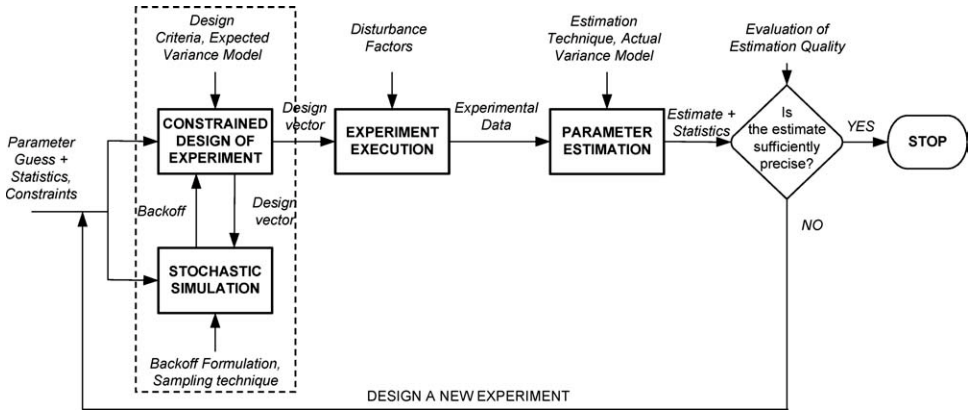

Figure 1. Constrained MBDoE with backoff: iterative scheme: the sequence of operations enclosed in the dashed box is detailed in Figure 2.

MBDoE scheme. The final goal of the whole procedure is to estimate the set of model parameters in the most precise and reliable way by performing a sequence of highly informative experiments within the feasible design region. For a sequence of experiments to be designed, the general scheme is shown in Figure 1. The methodology involves an iterative scheme requiring as initial inputs (i) the definition of the active constraints **C**, (ii) some knowledge on the parameter system (initial value of the model parameters and related statistics), and (iii) some information about the backoff policy and sampling technique.

The key activities are as follows:

1. The design with backoff step including the simultaneous execution of the following tasks: (i) the constrained design of the experiment, with the optimality condition (5) and feasibility condition (12) adjusted for the backoff from the active constraints; (ii) the stochastic simulation providing the backoff vector β, given the nominal value of model parameters.

2. The experiment execution, performed at the designed experimental conditions.

3. The parameter estimation (different estimation techniques can be used: least-squares, maximum likelihood, Bayesian estimation) from the collected experimental data.

4. The assessment of the statistical precision of the parameters.

The sequence of activities can be iterated until a sufficiently precise estimation is achieved. Figure 2 shows the flux of information and tasks occurring in the stochastic simulation (Step 1i) defining the back-offs β (depending on φ), and in the MBDoE (Step 1ii) defining of the design vector φ (depending on β).

The critical steps are the description of the predicted uncertainty region of model parameters and the mapping of the predicted uncertainty region of the state variables. As for the first issue, the focus is on how to exploit the prior information and available knowledge to define the domain $\boldsymbol{T}$ of parametric uncertainty in a reliable way. In particular, it is not trivial to define a probability density function representing the variability of the parametric set. As for the second issue, the problem of mapping the predicted uncertainty region of state variables is an approximation problem solved through a probabilistic approach driven by the experimenter. In fact, the problem can be seen as choosing an optimal trade-off between the accurate mapping of the uncertainty region of the state variables and the computational effort for the stochastic simulation.

Two case studies are examined in this article; they differ in terms of the number of model parameters to be estimated,

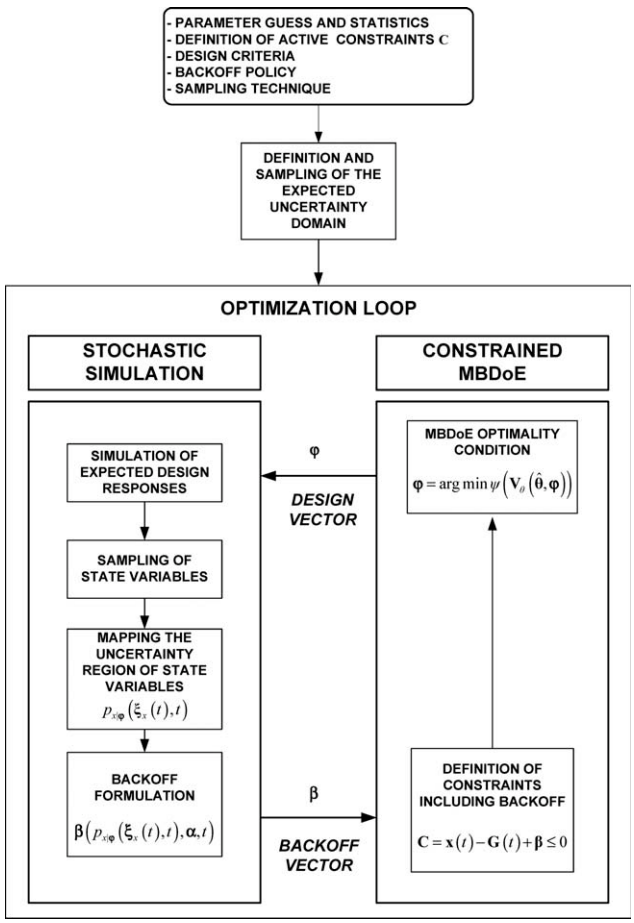

Figure 2. Flux of information in the constrained MBDoE and stochastic simulation coupling.

Table 1. Case Study 1: Value of Model Constants and Description of Basal Parameters. Subject's Body Weight: 78 kg

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_I	Insulin distribution volume (L)	12
n	Disappearance rate of insulin (min^{-1})	5/54
u_b	Basal insulin infusion rate (mU/min)	10.0

and of the type of active constraints on the state variables. Both examples consider physiological models; however, the methodology can be applied without any further extension to generic process models.

In this article, the gPROMS® modelling environment[52] is used for modelling, simulation and optimization purposes, as well as to design the experiments. The SRQPD optimization solver of gPROMS has been coupled with the SIMLAB® software[53] to generate the perturbed set of model parameters for the stochastic simulations. An SQP (sequential quadratic programming) routine was adopted in a two-step multiple shooting technique[54] 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 randomised to provide different initial points for the subsequent multiple shooting optimization.

Case Study 1: Optimal Insulin Infusion Rate in a Subject Affected by Diabetes

Optimal MBDoE techniques have been recently applied to a detailed model of glucose homeostasis to design a set of clinical tests that allow to estimate the model parameters in a statistically sound way for a subject affected by type-1 diabetes mellitus.[8] However, because of the parametric mismatch between the subject and the model (both are represented by the same DAE model, but with different sets of parameters), a design strategy may provide an infeasible solution due to a violation of the existing constraints on the output (blood glucose concentration). Here, the goal is to assess the effectiveness of a backoff-based experiment design strategy aimed at ensuring, in the presence of parametric mismatch, a feasible and optimally informative clinical test for parameter estimation purposes. A simplified model of glucose homeostasis[55] is adopted to describe blood glucose and insulin concentrations dynamics. The model is represented by the following set of differential equations:

$$\frac{dC_g}{dt} = -\theta_1 C_g - X\left(C_g + C_{g,b}\right) + D(t) \tag{22}$$

$$\frac{dX}{dt} = -\theta_2 X + \theta_3 I \tag{23}$$

$$\frac{dI}{dt} = -n(I + I_b) + \frac{u(t)}{V_I} \tag{24}$$

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 meal disturbances model adopted in the study is the one proposed by Hovorka[56]:

$$D(t) = 2.5At\exp(-0.05t) \tag{25}$$

with A the amount of glucose of the meal (g CHO). The basal parameters (considered as constants) are given in Table 1.

The constraints on the system are the upper ($G_1 = 150$ mg/dL) and lower ($G_2 = 60$ mg/dL) thresholds on blood glucose concentration, which is the only state variable being constrained (i.e., $y = C_g = x_1$). In reality, the lower bound only is a hard constraint not to be violated. However, for the sake of example, both constraints will be treated as hard ones. Additional equality constraints are set on the final glucose concentration (which must be equal to the basal value of $C_{g,b} = 81$ mg/dL and with a zero derivative) and on the final insulin infusion rate (which must be equal to u_b).

The test has to be optimally informative and safe for the subject. Accordingly, an MBDoE with backoff is realised, where the design vector is

$$\varphi = [u(t), \mathbf{t}^{sp}]. \tag{26}$$

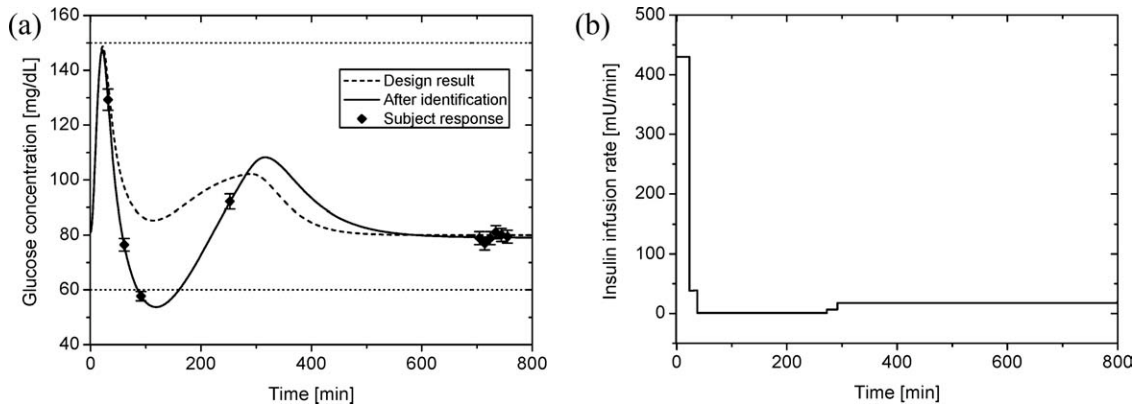

Figure 3. Case study 1, standard MBDoE.

(a) Glucose concentration profiles 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 with measurement error bars. (b) Profiles of the designed insulin infusion rate.

Table 2. Case Study 1, Standard MBDoE

Model Parameter	Final Value	Initial Guess	True Value	Confidence Interval 95%	95% t value	Standard Deviation
θ_1	0.028204	0.028735	0.0250	0.01396	2.021	0.00590
θ_2	0.014063	0.028344	0.0150	0.00241	5.843	0.00102
θ_3	1.241E-5	1.300E-5	1.26E-5	1.189E-6	10.44	5.028E-7

Parameter estimation, initial guess, true values, and statistics as 95% confidence intervals, t-values (reference t value = 1.898) and standard deviations.

The design variables are the insulin infusion rate and the vector of sampling times. The experiment design sessions are carried out by approximating the insulin infusion rate $u(t)$ as a piecewise constant function, with $n_{sw} = 7$ switching times and $n_z = 8$ switching levels. The optimal scheduling of a preset number $n_{sp} = 10$ of samples is also to be optimized, considering a minimum time of 10 min between two consecutive glucose concentration measurements. The glucose amount in the meal A is kept constant and equal to 60 g of carbohydrates. The measured variable is the blood glucose concentration C_g, with an expected relative error on the measurements of 3% of the reading. The chosen design criterion is the E-optimal experiment design for all the design configurations. The two constraints equations including backoff in the form (20) are:

$$C_1 = y + \beta_1 - G_1 \leq 0 \tag{27}$$

$$C_2 = -y + \beta_2 + G_2 \leq 0 \tag{28}$$

with $\beta = [\beta_1\ \beta_2]^T$ depending on the probability distribution of the system response at the nominal conditions $p_{x_1|\varphi}(\xi_{x_1}(t),t)$. The nominal values for the model parameters, valid for a healthy subject, are $\hat{\theta} = [0.02873 \quad 0.02834 \quad 1.30E-5]^T$.

As discussed by Furler et al.,[57] a subject affected by diabetes should have a lower value of the first parameter. Here it is assumed that:

1. The subject is diabetic and his/her condition is defined by the parameter set $\theta = [0.0250\ 0.0150\ 1.26E-5]^T$ (the relative deviations from the healthy subject set are therefore of −13%, −47% and −3% respectively).

2. The experimental design procedure is based on the $\hat{\theta}$ set describing a healthy subject.

To take into account the uncertainty of model parameters in mapping $p_{x_1|\varphi}$ a stochastic approach is followed by running $N' = 500$ simulations; therefore, the expected uncertainty of the model parameters adopted in the study is the following:

$$T = \left[\tilde{\theta}_{ij} \middle| \tilde{\theta}_{ij} \in p_\theta\left(\hat{\theta}_i, \xi_\theta\right), i = 1...3, j = 1...500\right] \tag{29}$$

defining an hypher-rectangular region of uncertainty where,

$$\begin{aligned} \tilde{\theta}_1 &\in R_1\left(\hat{\theta}_1 - \xi_{\theta_1}, \hat{\theta}_1 + \xi_{\theta_1}\right) = p_\theta\left(\hat{\theta}_1, \xi_{\theta_1}\right) \\ \tilde{\theta}_2 &\in R_2\left(\hat{\theta}_2 - \xi_{\theta_2}, \hat{\theta}_2 + \xi_{\theta_2}\right) = p_\theta\left(\hat{\theta}_2, \xi_{\theta_2}\right) \\ \tilde{\theta}_3 &\in R_3\left(\hat{\theta}_3 - \xi_{\theta_3}, \hat{\theta}_3 + \xi_{\theta_3}\right) = p_\theta\left(\hat{\theta}_3, \xi_{\theta_3}\right) \end{aligned} \tag{30}$$

$\mathbf{R} = [R_1\ R_2\ R_3]^T$ is a family of independent uniform distributions defined by a set of upper and lower variability bounds set by $\xi_\theta = [\xi_{\theta_1}\ \xi_{\theta_2}\ \xi_{\theta_3}]^T = [0.006\ 0.015\ 0.1E-5]^T$. These settings for the perturbed values of parameters include a wider uncertainty on the second parameter representing a subject with an altered insulin sensitivity.

If the uncertainty region of the state variables is built assuming a normal distribution and a 99.7% confidence region, it was observed that the distribution of the system responses is skewed. Therefore, the hyphothesis of a normal distribution provides a poorly accurate representation of the distribution of the system responses for this case study. As a consequence, to increase conservatism, the backoff is defined as the maximum variation from the nominal profile, i.e., the uncertainty region of the state variables is described through

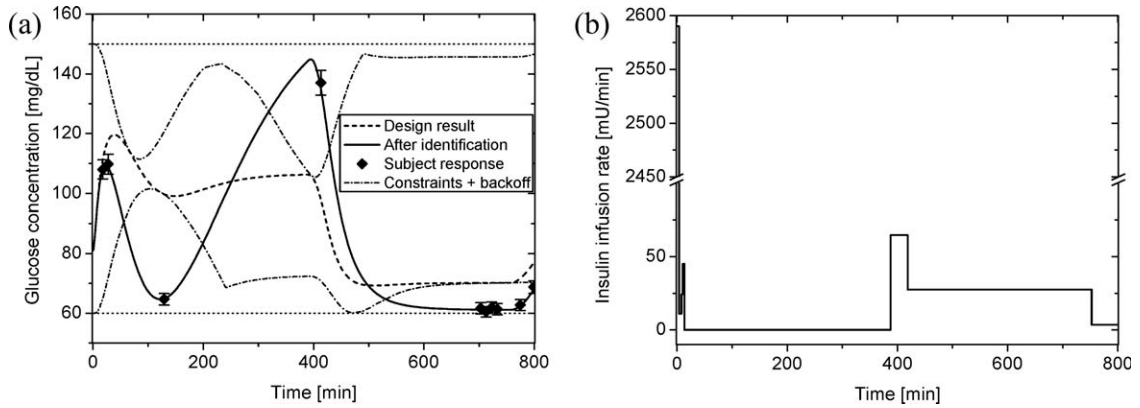

Figure 4. Case study 1, MBDoE with backoff.

(a) Glucose concentration profiles predicted by the model during the experiment design (broken line), after parameter identification (solid line), and effect of backoff on active constraints (dash-dot lines); the subject actual response to the designed experiment is indicated by diamonds with measurement error bars. (b) Profiles of the designed insulin infusion rate.

Table 3. Case Study 1, MBDoE with Backoff

Model Parameter	Final Value	Initial Guess	True Value	Confidence Interval 95%	95% t Value	Standard Deviation
θ_1	0.02517	0.028735	0.0250	0.00336	7.480	0.00142
θ_2	0.01513	0.028344	0.0150	0.00193	7.848	0.00081
θ_3	1.287E-5	1.300E-5	1.26E-5	1.244E-6	10.350	5.260E-7

Parameter estimation, initial guess, true values, and statistics as 95% confidence intervals, t-values (reference t value = 1.898) and standard deviations.

the maximum and minimum blood glucose concentration profiles over the parameter uncertainty domain.

To verify the effectiveness of a backoff-based experimental design, two different configurations were compared:

1. standard MBDoE with simple constraints and no backoff;
2. MBDoE with backoff from constraints.

The results from the simple design (Figure 3) can be seen as a motivating example for the adoption of a backoff-based strategy. The optimal design conditions do not comply with the lower constraint on the glucose concentration when applied to a diabetic subject. The test is unsafe for the subject because hyphoglycaemia occurs at $t \simeq 90$ min. It can be noted that the parameter set is estimated well (Table 2) using the data from the test: as expected, the designed experiment, although infeasible, is optimally informative for parameter estimation purposes.

To avoid constraint violations in the presence of parametric uncertainty, the experiment design procedure is coupled to a stochastic simulation to estimate the necessary backoff solving the (5), (9-11), (6) optimization problem. In this way,

$$\beta = \beta\left(R_{x_1|\varphi}\left(x_1^{\mathrm{MIN}}, x_1^{\mathrm{MAX}}, t\right), \alpha, t\right) = \beta\left(\hat{u}, p_\theta\left(\hat{\theta}, \xi_\theta\right), \alpha, t\right) \quad (31)$$

and the backoff is a function of $R_{x_1|\varphi}$ (x_1^{MIN}, x_1^{MAX}, t), a uniform distribution defined by the highest value and the lowest value of x_1 at the t time. This distribution is function of the parametric uncertainty distribution (30) and the estimated value of the actual manipulated input. The simulation is carried out with $N' = 500$ and is computationally expensive, although the calculations burden could be reduced if an appropriate initial guess profile of the manipulated input is chosen (e.g., by using by the solution of a standard design or, more efficiently, by evaluating the profile of the manipulated inputs that satisfy the constraints of the problem with backoff through a preliminary dynamic optimization).

Figure 4 shows the resulting profiles for the experimental design with backoff, and Table 3 shows the parameter estimation after the designed experiment with backoff. The parameter estimation is again statistically satisfactory (as can be seen from the 95% confidence t-test values and from the narrow confidence intervals) and the parameter values close to those of the diabetic subject. It is interesting to note that the design with backoff defines a test that is now both feasible and optimally informative. As can be seen from Figure 4a, the dynamics of glucose concentration is constrained within a narrow range of operability to take into account the stochastic contribution to the response of the parametric uncertainty. Also note that according to Table 3, the design with backoff allows obtaining a more precise estimation of the model parameters. This may sound counter-intuitive as the design space is restricted by the effect of the backoffs, and indeed it can be verified that the design with backoff does predict a less informative experiment (the final value for the minimised objective function is 0.00547 against 0.00310 for the standard design). Thus, the "unexpected" better estimation of the parameter values can be explained by the fact that the design (and the design objective function) depends on the current value of the model parameters, whereas the experiment and subsequent estimation of the parameter values are based on the actual (unknown) values of the parameters.

Case Study 2: Optimal Chemotherapeutic Drug Administration

A second case study considers the model originally proposed by Martin[58] for the optimal chemotherapeutic drugs administration to people affected by cancer. This model was further analysed by Banga et al.[59] in the topic of robust dynamic optimization, to determine the optimal cancer drug scheduling to decrease the size of a malignant tumor as measured at some particular time in the future (note that more complex models can be found in the literature, e.g., in the optimization study by Dua et al.[60]). Here, the goal is to assess the effectiveness of a backoff-based experiment design strategy on ensuring, in the presence of parametric mismatch, a feasible and optimally informative clinical test for parameter estimation purposes, with the additional constraint of maintaining an effective therapy in terms of a reduction of the number of cancer cells to be observed during the test.

The cell-cycle nonspecific model comprises the following set of equations:

$$\frac{\mathrm{d}x_1}{\mathrm{d}t} = -\theta_1 x_1 + \theta_2 (x_2 - \theta_3) H \quad (32)$$

$$\frac{\mathrm{d}x_2}{\mathrm{dt}} = u_{\mathrm{d}} - \theta_4 x_2 \quad (33)$$

$$\frac{\mathrm{d}x_3}{\mathrm{d}t} = x_2 \quad (34)$$

Table 4. Case Study 2: Nominal Values and Description of Model Parameters

Parameter	Value	Description
θ_1	9.9E-4	Cancer cells proliferation (days)
θ_2	8.4E-3	Drug action in cancer cells elimination ($\mathrm{days}^{-1}\mathrm{D}^{-1}$)
θ_3	10	Drug threshold effect (D^{-1})
θ_4	0.27	Drug elimination from the body (days^{-1})

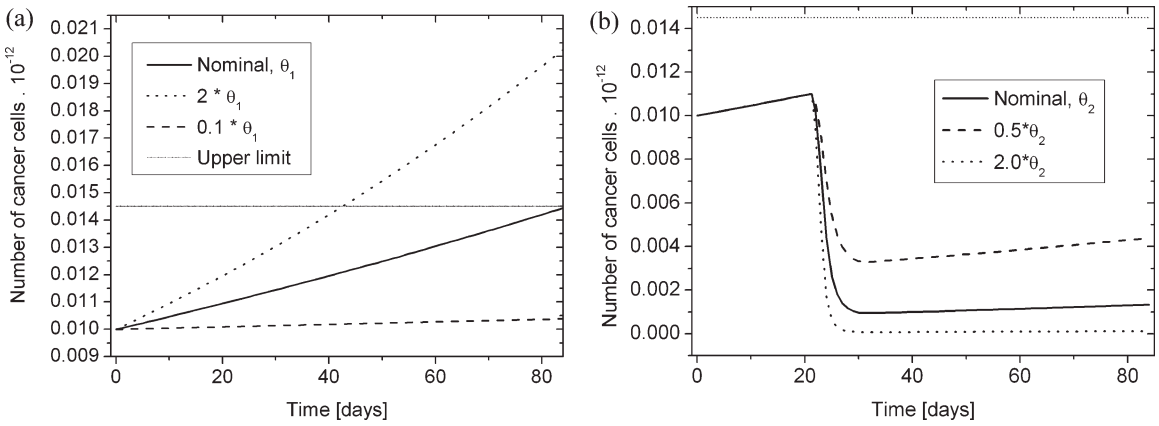

Figure 5. Case study 2: effect of the value of θ_1 and θ_2 on the number of cancer cells over a 12-week period.

(a) No drug administration. (b) Step administration of drug after three weeks.

where x_1 represents the reduction of tumor cells (dimensionless), x_2 is the drug concentration in the body in drug units (D), x_3 is the cumulative (toxic) effect of the drug ($D\times$days); θ_1, θ_2, θ_3, and θ_4 are the model parameters to be estimated, u_d is the drug administration rate (D/days) and t is the time (days). The tumor mass in terms of the number of cancer cells is given by:

$$N_{\text{cells}} = 10^{13}\exp(-x_1). \tag{35}$$

The drug concentration must be kept below an assigned level during the treatment period and the cumulative effect of the drug must be kept below the ultimate tolerance level. Function H depends on x_2 and θ_3 as in the following:

$$H = \begin{cases} 1, & x_2 \geq \theta_3 \\ 0, & x_2 < \theta_3 \end{cases} \tag{36}$$

This takes into account the fact that the drug is effective only if its concentration in the body is above a threshold level.

The dynamic optimization problem consists of finding the optimal $u_d(t)$ over $t \in [0,t_f]$ by maximizing

$$J = x_1\big|_{t_f} \tag{37}$$

subject to (32–34) and to the following path constraints:

$$\begin{aligned} x_2(t) &\leq G_1 = 50 \\ x_3(t) &\leq G_2 = 2100 \times \text{days} \end{aligned} \tag{38}$$

as well as to the following interior point constraints:

$$\begin{aligned} x_1|_{t=21} &\geq G_3 = \ln(200) \\ x_1|_{t=42} &\geq G_4 = \ln(400) \\ x_1|_{t=63} &\geq G_5 = \ln(800) \end{aligned} \tag{39}$$

stating that there must be at least 50% reduction in the size of the tumor every 3 weeks. In this case study, we see that $N_y = 1$ and $N_c = 5$ (i.e., this is a problem with a single-measured response with multiple constraints); in addition, the control input is bounded ($0 \leq u_d(t) \leq 100$). The upper threshold for the cancer biomass is $N_{\text{cells}} = 1.5\text{E} + 10$, which can be seen as a terminal state for the subject. The maximum experiment time acceptable for the optimization is $t_f = 84$ days (12 weeks). The measured variable is $y = \exp(-x_1)$, and the measurements are available with a relative error of 3%. The initial state is taken at $\mathbf{x}^0 = [\ln(100)\ 0\ 0]^T$ and the nominal values of model parameters are shown in Table 4, with a short explanation of their physical meaning.

The model is rather sensitive to the parameter values. Figure 5 shows the effect of a change in one parameter value on the proliferation of cancer cells (Figure 5a) and on the effectiveness of the drug therapy (Figure 5b).

A constrained dynamic optimization adopting (37) as the objective function would provide an efficient test, but a low information level for parameter estimation purposes. More importantly, the parametric uncertainty might lead the test to be infeasible and to return sub-optimal solutions. On the other hand, a simple MBDoE procedure based on (5) would provide an optimally informative but less effective test protocol, with the same problems of feasibility and robustness of the solution in the presence of parametric uncertainty. A backoff strategy can guarantee the feasibility of both an optimally informative (i.e., MBDoE-based) and an optimally efficient (i.e., dynamically optimized) test.

To solve the optimization problems, the model parameters are scaled to unity using the values reported in Table 4 (symbol $\hat{\Theta}$ is used for the estimated normalized set). A model with $\Theta = [1.2\ 0.8\ 1.2\ 0.8]^T$ (i.e., 20% deviation from the nominal) is considered to represent a subject with a greater proliferation of cancer cells and a less effective response to drug delivery. The manipulated input is approximated with a piecewise constant function with $n_{sw} = 10$ and $n_z = 11$. The sampling points ($n_{sp} = 13$) are collected to identify the model parameters with a minimum time of 1 day between two consecutive measurements. An E-optimal criterion is chosen for MBDoE.

The expected uncertainty domain $\boldsymbol{T}$ of model parameters is defined by

$$T = \left[\tilde{\Theta}_{ij}\middle|\tilde{\Theta}_{ij} \in p_{\Theta_i}\left(\hat{\Theta}_i, \xi_{\Theta_i}\right),\quad i = 1...3, j = 1...N'\right] \tag{40}$$

where the normalized parameter vector components are assumed to be independent and normally distributed stochastic variables

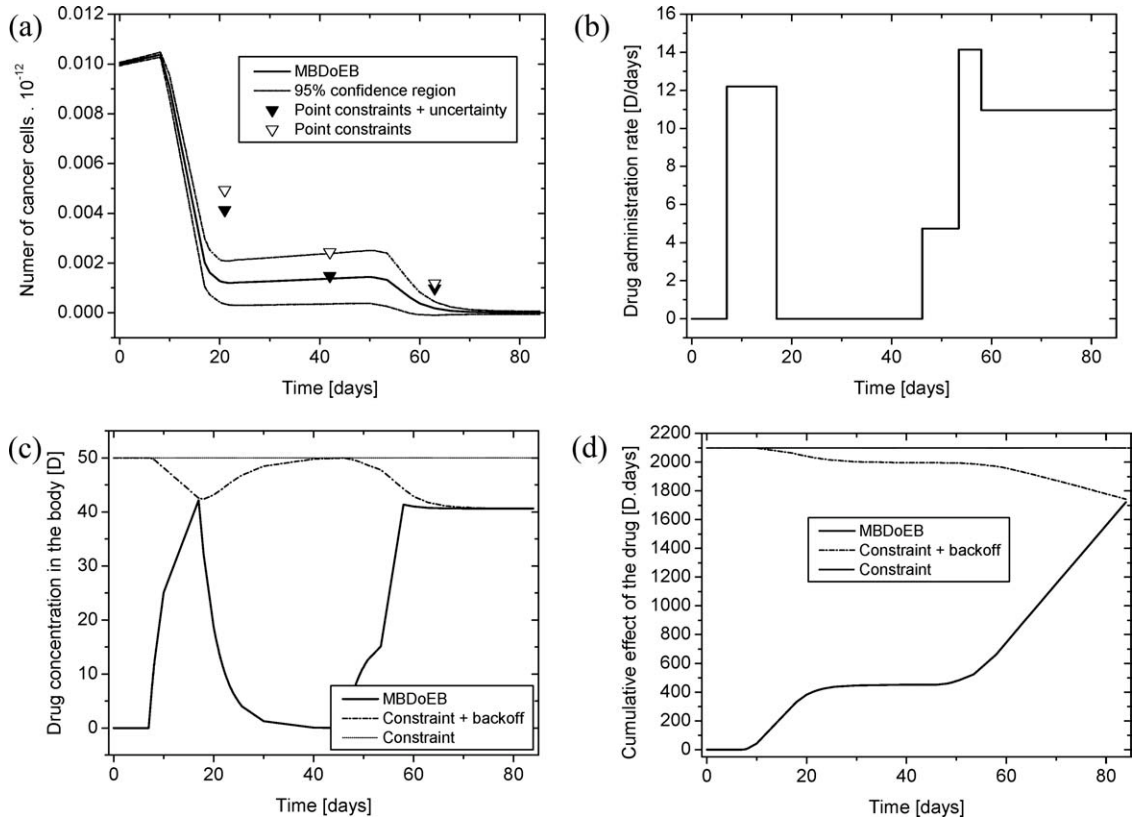

Figure 6. Case study 2, MBDoEB.

(a) Profile of the number of cancer cells N_{cells} (solid line) with 95% confidence region (short dash-dotted lines), point constraints (open triangles) and effect of the backoff on point constraints (closed triangles). (b) Optimal profile for drug administration rate. (c) Drug concentration in the body. (d) Cumulative effect of the drug profiles; path constraints (dot lines) and effect of the backoff on constraints (dash-dotted lines) during the optimization.

$$\tilde{\Theta}_i \in p_{\Theta_i}\left(\hat{\Theta}_i, \xi_{\Theta_i}\right) = N_i(\mu_i, \sigma_i) \quad i = 1...N_\theta \tag{41}$$

with mean $\mu = \hat{\Theta}$ and standard deviation $\boldsymbol{\sigma} = 0.15$. $N' = 100$ simulations were carried out to build the $(1-\alpha) = 99.7\%$ confidence region of system responses. Note that the selected vector of standard deviations defines a wide uncertainty region for the model parameters. The choice of the number of simulations is related to the definition of the predicted uncertainty: the wider the uncertain region the smaller the number of simulations required to represent it (i.e., there is a trade-off between calculation/experimental effort and confidence on the parameters value).

Although results are not reported here for the sake of conciseness, both a standard dynamic optimization (DOPT) as in Banga et al.[59] and a standard MBDoE were carried out. In both cases, the fundamental issue is that the interior point constraints on the number of cancer cells and the upper bound on the cumulative drug concentration in the body were violated. Also a dynamic optimization with backoff (DOPTB) was carried out: in this case no violation of the constraint occurred, but yet an evenly spaced (not optimized) sampling policy was not effective for a sound estimation of the model parameters.

To verify the effectiveness of a backoff approach for optimal drug scheduling and to overcome the limitations of standard optimization and standard experiment design, a constrained design of experiment with backoff (MBDoEB) is used to determine the optimal drug administration rate.

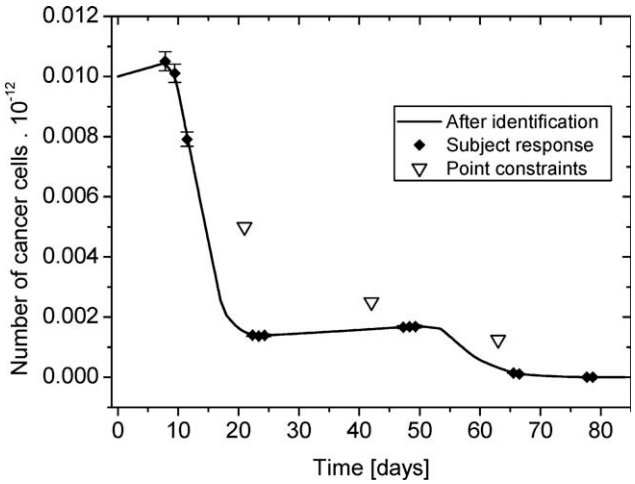

Figure 7. Case study 2, MBDoEB.

Profile of the number of cancer cells N_{cells} predicted by the model after parameter identification (solid line), sample measurements (diamonds) and interior point constraints (open triangles).

Table 5. Case Study 2, MBDoEB

Model Parameter	Final Value	Initial Guess	True Value	Confidence Interval 95%	95% t Value	Standard Deviation
Θ_1	1.2389	1.0	1.2	0.3693	3.355	0.1633
Θ_2	0.7773	1.0	0.8	0.1211	6.417	0.0535
Θ_3	1.1844	1.0	1.2	0.1473	8.038	0.0651
Θ_4	0.7841	1.0	0.8	0.0996	7.871	0.0440

Parameter estimation, initial guess, true values and statistics as 95% confidence intervals, t-values (reference t value = 1.833) and standard deviations.

The results in terms of statistics on parameter estimation, feasibility and effectiveness of the care are analysed in the following section.

Model-based experiment design with backoff

To formulate the necessary backoff from the active constraints, the distribution of the entire set of state variables (y, x_2, x_3) was approximated by a set of independent normal distributions. The backoff vector in the (22) form can be evaluated from the $(1-\alpha)\%$ confidence intervals for the state variables

$$\beta(\mathbf{N}(\xi_x(t),t),\alpha,t) = \kappa \tag{42}$$

where the set of independent normal distributions $\mathbf{N}$ is defined by $\xi_x(t)$ parameters, defining a vector of average profiles and standard deviations from the average profiles.

The vector $\mathbf{C} = [C_1\ C_2\ C_3\ C_4\ C_5]$ of active constraints on state variables including backoff can be expressed for the state variables x_2 and x_3 as

$$C_1 = x_2(t) + \beta_1(t) - G_1 \le 0 \tag{43}$$

$$C_2 = x_3(t) + \beta_2(t) - G_2 \le 0 \tag{44}$$

concerning the path constraints, and

$$C_3 = y|_{t=21} - \beta_3|_{t=21} - G_3 \ge 0 \tag{45}$$

$$C_4 = y|_{t=42} - \beta_4|_{t=42} - G_4 \ge 0 \tag{46}$$

$$C_5 = y|_{t=63} - \beta_5|_{t=63} - G_5 \ge 0 \tag{47}$$

for the interior point constraints. The N_c-dimensional vector of backoff functions coming from the normality assumption is

$$\beta = \begin{bmatrix} 2\sigma_y|_{t=21} \\ 2\sigma_y|_{t=42} \\ 2\sigma_y|_{t=63} \\ 2\sigma_{x_2}(t) \\ 2\sigma_{x_3}(t) \end{bmatrix} \tag{48}$$

for a confidence level of $(1-\alpha) = 95\%$. In this case, the assumption of independence and identical distribution for the states allows describing the uncertainty region of system responses in a satisfactory and adequate way.

A further constraint is added to ensure the effectiveness of the therapy in the worst case (minimum reduction of the size of the tumor over the considered uncertainty domain of model parameters $\boldsymbol{T}$), stating that

$$J|_T \ge 13.8 - \varepsilon \tag{49}$$

where ε is a small nonzero number taking into account the deviation of model-based design of experiments with backoff (MBDoEB) objective function from the optimal conditions derived from DOPTB. The idea underneath this approach is to perform an optimal experiment design ensuring the effectiveness of a dynamic optimization with backoff. The MBDoEB optimization problem consists on finding the optimal $u_d(t)$ profile satisfying through the model (32–36) the design optimality condition and the feasibility constraints (43–47) and (49).

Results are shown in Figure 6. The backoff strategy guarantees the feasibility of the designed experiment with the specified level of uncertainty. We verified that the final reduction of the mass of the tumor is indeed very similar to the one obtained through a DOPTB. The optimal settings provided by the optimization lead the x_2 and x_3 profiles close to the upper path constraints, without crossing them.

The measurements from the designed experiment (MBDoEB) and the estimated profile are shown in Figure 7 (also showing how the interior point constraints on the measured variable y are largely fulfilled). Note that the design optimization chooses a very uneven sampling profile (i.e., sample are taken where the information content is higher).

The parameter estimation is statistically satisfactory (even if the informative content of the test is lower than the one obtained in the standard MBDoE) as summarised in Table 5.

Table 6 summarizes all the simulation results considered in Case Study 2 in terms of proposed technique, feasibility, and design optimality.

One drawback of the proposed methodology is the high computational effort for the stochastic simulation required to build the necessary backoff. As mentioned before, parallel computing and SA-based sampling methods can drastically reduce the computational burden. In any case, the computational burden is not a critical issue in this case study as the calculation time (<5 h in a Pentium D 3Ghz CPU) is small when compared to the duration of the therapy and is to be done before the therapy has commenced.

Table 6. Case Study 2, Summary of the Results Achieved with Different Proposed Techniques

Experiment	Technique	Feasibility	Design Optimality
DOPT	Dynamic optimization	No	No
DOPTB	Dynamic optimization with backoff	Yes	No
MBDoE	Constrained experiment design	No	Yes
MBDoEB	Constrained experiment design with backoff	Yes	Yes

Conclusions

In this article, a methodology for the constrained MBDoE in the presence of parametric uncertainty was proposed and

discussed. The optimal design of an experiment for improving parameter estimation is a particular form of dynamic optimization problem that can be very effective where both optimality and feasibility of the designed experiment are important issues to consider. As parameter uncertainty affects both design optimality and experiment feasibility, a modified methodology exploiting stochastic information about the parametric system was adopted to design the necessary backoffs from the active constraints. The backoff strategy allows moving the optimal point to keep the experiment in the feasible region of the state variables. Two simulated case studies have been proposed to assess the effectiveness of the new technique.

In the first case study, the methodology was applied to a model of the glucose homeostasis to detect the best insulin infusion rate profile to infuse to estimate the parameters of a subject with type-1 diabetes mellitus when no preliminary information about the subject is available. The backoff strategy allows estimating the parametric set describing the diabetic subject in a safe manner, while a standard design (even if optimal) leads the subject to hypoglicaemia.

In the second case study, dealing with the optimal delivery of chemotherapeutic agents in cancer treatment, the problem of estimating the model parameters was faced by considering the effectiveness of the care (i.e., its capability of decreasing the tumor size in a given amount of time) as well as the optimality and feasibility of the optimized test. As a standard model-based experiment design leads to an infeasible test when parametric uncertainty is present, an MBDoE with backoff was analyzed and discussed. This methodology proved to ensure both optimality and feasibility of the planned experiment, overcoming the limitations of the other two.

Future work will aim at further improving the proposed approach by reducing the computational burden (which may hinder the method convenience/applicability in short-lasting experiments or whenever an on-line design update is desired[61]) and by generalizing (and optimizing) the handling of the stochastic representation. Additional benefits may derive by the definition and implementation of global optimization algorithms preventing the risk of incurring into local minima.

Acknowledgments

The financial support of the University of Padova (Italy) under Progetto di Ateneo 2007 (cod. CPDA074133): "A process systems engineering approach to the development of an artificial pancreas for subjects with type-1 diabetes mellitus" is gratefully acknowledged.

Notation

A = glucose amount of the meal
C_g = glucose concentration in the blood
$C_{g,b}$ = basal glucose concentration in the blood
c_{ij} = ij-th element of the correlation matrix for model parameters ($\mathbf{C}$)
D = meal disturbance function
f = differential and algebraic system implicit function
g = measurements selection function
H = threshold-based function for drug effectiveness
I = insulin concentration in the accessible compartment
I_b = insulin basal value
J = objective function of the dynamic optimization
k_i = i-th Bolus release relaxing factor
n = disappearance rate of insulin
n_{sp} = number of samples
N = population abundance
N^{cells} = number of cancer cells
N' = number of samples representing probability p_θ
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
n_{ξ_θ} = number of parameters needed to define probability density function p_θ
n_{ξ_x} = number of parameters needed to define probability density function p_x
p_θ = probability density function of model parameters
p_x = probability density function of model responses
p_Θ = probability density function of normalised model parameters
$p_{\theta|\varphi}$ = probability density function of model parameters conditioned by φ
$p_{x|\varphi}$ = probability density function of model responses conditioned by φ
q_i = i-th element of the dynamic sensitivity matrix ($\mathbf{Q}$)
R = probability density function of a uniform distribution
s_{ij} = ij-th Element of the inverse matrix of measurements errors
t = time
t' = t-value for Bartlett's formula (16)
t_f = maximum experimental time
t_i = i-th t-value
t^{sw} = switching time
u = insulin infusion rate
u_d = drug administration rate
u_b = time-invariant basal insulin infusion rate
u_{bol} = insulin bolus amount
v_{ij} = ij-th element of the variance-covariance matrix $\mathbf{V}_\theta$
V_I = insulin distribution volume
x = generic state variable
X = insulin concentration in the nonaccessible compartment
y = generic measured output
α = statistical significance factor
β_i = i-th element of the backoff vector
ε' = expected error for Bartlett's formula (16)
ε = small nonzero number for formula (51)
φ_i = i-th element of the design vector
Γ = response selection function
κ_i = i-th confidence interval
$\hat{\sigma}$ = expected variance for Bartlett's formula (16)
$\sigma_{x,i}$ = standard deviation of the i-th response
σ_y = standard deviation of the measured variable
θ_i = i-th model parameter
$\tilde{\theta}_i$ = stochastic realization of the i-th model parameter
$\tilde{\theta}_{ij}$ = stochastic realization of the i-th element of the parameters vector in the j-th event
Θ_i = i-th normalised model parameter
$\tilde{\Theta}_i$ = stochastic realization of the i-th normalised model parameter
$\tilde{\Theta}_{ij}$ = stochastic realization of the i-th normalised parameter in the j-th event
τ = test duration
ψ = $\mathbf{V}_\theta$ measurement function

Vectors and Matrices [dimension]

$\mathbf{C}$ = set of constraint functions [N_c]
$\tilde{\mathbf{C}}$ = stochastic vector 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 [$N_\theta \times N_\theta$]
$\mathbf{y}_0$ = vector of initial conditions [N_y]
$\mathbf{y}$ = measurements vector [N_y]
$\hat{\mathbf{y}}$ = vector of estimated responses [N_y]
$\mathbf{Q}$ = sensitivity matrix [$n_{sp} \times N_\theta$]
$\mathbf{t}^{sp}$ = vector of sampling points [n_{sp}]
$\mathbf{t}^{sw}$ = vector of switching times [$n_{sw}+1$]
$\mathbf{u}$ = 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 conditions for state variables [N_x]
$\mathbf{x}^0$ = vector of initial states [N_x]
$\dot{\mathbf{x}}$ = vector of derivatives on state variables [N_x]
$\mathbf{X}$ = matrix of stochastic simulation responses [$N' \times N_y$]
$\bar{\mathbf{x}}$ = vector of average responses [N_x]
$\mathbf{X}^i$ = vector of responses in the i-th simulation run [N_x]
α = vector of significance factors [N_c]
β = vector of backoff functions [N_c]
φ = design vector [n_φ]
$\tilde{\varphi}$ = stochastic design vector [n_φ]
κ = vector of confidence intervals [N_x]
Λ = vector of user-defined coefficients for equation (23) [N_c]
θ = vector of values of model parameters for the subject [N_θ]
$\hat{\theta}$ = vector of estimated values of model parameters [N_θ]
$\tilde{\theta}$ = stochastic vector of model parameters [N_θ]
θ^0 = vector of initial guesses of model parameters [N_θ]
Θ = vector of normalised model parameters for the subject [N_θ]
$\hat{\Theta}$ = vector of estimated values of normalised model parameters [N_θ]
σ_x = vector of standard deviations on model responses [N_x]
Σ = measurement errors variance-covariance matrix [$N_y \times N_y$]
Σ_θ = prior variance-covariance matrix of model parameters [$N_\theta \times N_\theta$]
ξ_θ = vector of parameters needed to define probability density function p_θ [n_{ξ_θ}]
ξ_x = vector of parameters needed to define probability density function p_x [n_{ξ_x}]

Other symbols

$\mathbf{N}$ = set of probability density functions with normal distributions [N_c]
$\mathbf{R}$ = set of probability density functions with uniform distribution [N_c]
$\boldsymbol{T}$ = uncertainty domain of model parameters

Literature Cited

1. Franceschini G, Macchietto S. Model-based design of experiments for parameter precision: state of the Art. *Chem Eng Sci*. 2008;63:4846–4872.
2. Bauer I, Bock HG, Körkel S, Schlöder JP. Numerical methods for optimum experimental design in DAE systems. *J Comput Appl Mathem*. 2000;120:1–25.
3. Chen BH, Bermingham S, Neumann AH, Kramer HJM, Asprey SP. 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 JF. Optimal dynamic experiment design for estimation of microbial growth kinetics at sub-optimal temperatures: Modes of implementation. *Sim Mod Pract Th*. 2005;13:129–138.
5. Gadkar KG, Gunawan R, Doyle FJ. Iterative approach to model identification of biological networks. *BMC Bioinformatics*. 2005;6:155–174.
6. Franceschini G, Macchietto S. Validation of a model for biodiesel production through model-based experiment design. *Ind Eng Chem Res*. 2007;46:220–232.
7. Prasad V, Vlachos DG. Multiscale model and informatics-based optimal design of experiments: application to the catalytic decomposition of ammonium on ruthenium. *Ind Eng Chem Res*. 2008;47:6555–6567.
8. 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.
9. Rasch A, Bücker HM, Bardow A. Software supporting optimal experiment design: a case study of binary diffusion using EFCOSS. *Comput Chem Eng*. 2009;33:838–849.
10. Pukelsheim F. *Optimal Design of Experiments*. New York: Wiley, 1993.
11. Asprey SP, Macchietto S. Statistical tools for optimal dynamic model building. *Comput Chem Eng*. 2000;24:1261–1267.
12. Bruwer MJ, MacGregor JF. Robust multi-variable identification: optimal experimental design with constraints. *J Process Control*. 2006;16:581–600.
13. Ford I, Titterington DM, Kitsos CP. Recent advances in nonlinear experimental design. *Technometrics*. 1989;31:49–60.
14. Grossmann IE, Sargent RWH. Optimum design of chemical plants with uncertain parameters. *AIChE J*. 1978;37:517.
15. Raspanti CG, Bandoni JA, Biegler LT. New strategies for flexibility analysis and design under uncertainty. *Comput Chem Eng*. 2000;24:2193–2209.
16. Halemane KP, Grossmann IE. Optimal process design under uncertainty. *AIChE J*. 1983;29:425.
17. Swaney RE, Grossmann IE. An index for operational flexibility in chemical process design. *AIChE J*. 1985;31:621–629.
18. Pistikopoulos EN. Uncertainty in process design and operations. *Comput Chem Eng*. 1995;19:553.
19. Mohideen MJ, Perkins JD, Pistikopoulos EN. Optimal design of dynamic systems under uncertainty. *AIChE J*. 1996;42:2251–2272.
20. Ierapetritou MG, Pistikopoulos EN. A novel optimization approach of stochastic planning models. *Ind Eng Chem Res*. 1994;33:1930.
21. Li P, Arellano-Garcia H, Wozny G. Chance constrained programming approach to process optimization under uncertainty. *Comput Chem Eng*. 2008;32:25–45.
22. Mönnigmann M, Marquardt W. Steady-state process optimization with guaranteed robust stability and feasibility. *AIChE J*. 2003;49:3110–3126.
23. Chachuat F, Srinivasan B, Bonvin D. Model parameterisation tailored to real time optimization. In: Braunschweig B, Joulia X, editors. *Proceedings of 18th European Symposium on Computer Aided Process Engineering* (ESCAPE), Elsevier, Amsterdam, The Netherlands, 2008;1–13.
24. Bahri PA, Bandoni JA, Barton GW, Romagnoli JA. Backoff calculations on optimising control: a dynamic approach. *Comput Chem Eng*. 1995;19:S699.
25. Loeblein C, Perkins JD, Srinivasan B, Bonvin D. Economic performance analysis in the design of on-line batch optimisation systems. *J Process Control*. 1999;9:61–78.
26. Lear JB, Barton GW, Perkins JD. Interaction between process design and process control: the impact of disturbances and uncertainty of the estimates of achievable economic performance. *J Process Control*. 1995;18:244–257.
27. Grossmann IE, Morari M. *Operability, resiliency and flexibility-process design objectives for a changing world*. In: Westerberg AW, Chien HH, editors. *Proceedings of 2nd International Conference on Foundations of Computer-Aided Process Design (FOCAPD)*. 1984;931–1010.
28. Bansal V, Perkins JD, Pistikopoulos E. Flexibility analysis and design of linear systems by parametric programming. *AIChE J*. 2000;46:335–354.
29. Bernardo FP, Saraiva PM. A robust optimization framework for process parameter and tolerance design. *AIChE J*. 1998;44:2007–2017.
30. Bahri PA, Bandoni JA, Romagnoli JA. Effect of Disturbances in optimizing control: steady-state open-loop back-off calculation. *AIChE J*. 1996;42:983–984.
31. Loeblein C, Perkins JD. Economic analysis of different structures of on-line process optimisation systems. *Comput Chem Eng*. 1998;22:1257–1269.
32. Bahri PA, Bandoni JA, Romagnoli JA. Integrated flexibility and controllability analysis in design of chemical processes. *AIChE J*. 1997;43:997–1015.
33. Bernardo FP, Saraiva PM, Pistikopoulos EN. Inclusion of information costs in process design optimization under uncertainty. *Comput Chem Eng*. 2000;24:1695–1701.
34. Körkel S, Kostina E, Bock HG, Schlöder JP. Numerical methods for optimal control problems in design of robust optimal experiments for nonlinear dynamic processes. *Opt Methods Software*. 2004;19:327–338.
35. Asprey SP, Macchietto S. Designing robust optimal dynamic experiments. *J Process Control*. 2002;12:545–556.
36. Rustem B, Zakovic S. Semi-infinite programming and application to min-max problems. *Ann Oper Res*. 2003;124:81–110.
37. Rojas CR, Welsh JS, Goodwin GC, Feuer A. Robust optimal experiment design for system identification. *Automatica*. 2007;43:993–1008.

38. Hjalmarsson H. From experiment design to closed-loop control. *Automatica*. 2005;41:393–438.
39. Chu Y, Hahn J. Integrating parameter selection with experimental design under uncertainty for nonlinear dynamic systems. *AIChE J*. 2008;54:2310–2320.
40. Vassiliadis VS, Sargent RWH, Pantelides CC. Solution of a class of multistage dynamic optimizations problems. 1-Problems without path constraints. *Ind Eng Chem Res*. 1994;33:2111–2122.
41. Galvanin F, Macchietto S, Bezzo F. Model-based design of parallel experiments. *Ind Eng Chem Res*. 2007;46:871–882.
42. Zhang Y, Edgar TF. PCA combined model-based design of experiments (DOE) criteria for differential and algebraic system parameter identification. *Ind Eng Chem Res*. 2008;47:7772–7783.
43. Rooney WC, Biegler LT. Design for model parameter uncertainty using nonlinear confidence regions. *AIChE J*. 2001;47:1794–1804.
44. Cochran WG. *Sampling Techniques*. New York: Wiley, 1977.
45. Bartlett JE, Kotrlik JW, Higgins C. Organizational research: determining appropriate sample size for survey research. *Inform Technol Learning Perform J*. 2001;19:43–50.
46. Bilodeau M, Brenner D. *Theory of Multivariate Statistics*. Berlin: Springer, 1999.
47. Chao CT. Selection of sampling units under a correlated population based on the eigensystem of the population covariance matrix. *Environmetrics*. 2004;15:757–775.
48. Saltelli A, Tarantola S, Chan KPS. A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics*. 1999;41:39–56.
49. Kucherenko S, Rodriguez-Fernandez M, Pantelides CC, Shah N. Monte Carlo evaluation of derivative-based global sensitivity measures. *Reliab Eng Syst Saf*. 2009;94:1135–1148.
50. Darlington J, Pantelides CC, Rustem B, Tanyi BA. An algorithm for constrained nonlinear optimisation under uncertainty. *Automatica*. 1999;35:217–228.
51. Apley DW, Liu J, Chen W. Understanding the effects of model uncertainty in robust design with computer experiments. *ASME J Mech Des*. 2006;128:945–958.
52. gPROMS. Advanced User Guide (release 3.0). Process Systems Enterprise Ltd., London, United Kingdom, 2008.
53. SimLab User Manual (release 3.2.5), Joint Research Centre (European Commission), Brussels, 2008. Available at: http://simlab.jrc.ec.europa.eu/docs/index.htm; accessed on July 15, 2009.
54. Bock H, Kostina E, Phu HX, Rannacher R. *Modeling, Simulation and optimization of complex processes*. Berlin: Springer, 2003.
55. Lynch SM, Bequette BW. Model predictive control of blood glucose in type I diabetics using subcutaneous glucose measurements. *Proc ACC*. 2002;4039–4043.
56. Hovorka R, Shojaee-Moradie F, Carroll PV, Chassin LJ, Gowrie IJ, Jackson NC, Tudor RS, Umpleby AM, Jones RH. Partitioning glucose distribution/transport, disposal, and endogenous production during IVGTT. *Am J Physiol Endocrinol Metab*. 2002;282:E992–E1007.
57. Furler SM, Kraegen EW, Smallwood RH, Chilsom DJ. Blood glucose control by intermittent loop closure in the basal mode: computer simulation studies with a diabetic model. *Diabetes Care*. 1985;8:553–561.
58. Martin RB. Optimal control drug scheduling of cancer chemotherapy. *Automatica*. 1992;28:1113–1123.
59. Banga JR, Balsa-Canto E, Moles CG, Alonso AA. Dynamic optimisation of bioprocesses: efficient and robust numerical strategies. *J Biotechnol*. 2005;117:407–419.
60. Dua P, Dua V, Pistikopoulos EN. Optimal delivery of chemotherapeutic agents in cancer. *Comput Chem Eng*. 2008;32:99–107.
61. 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.

Manuscript received Mar. 9, 2009, and revision received Oct. 8, 2009.