17th European Symposium on Computer Aided Process Engineering – ESCAPE17
V. Plesu and P.S. Agachi (Editors)
© 2007 Elsevier B.V. All rights reserved.

Optimal experimental design based on global sensitivity analysis

Maria Rodriguez-Fernandez, Sergei Kucherenko, Costas Pantelides, Nilay Shah

Centre for Process Systems Engineering, Department of Chemical Engineering
Imperial College London, London SW7 2AZ, UK
m.rodriguez@imperial.ac.uk

Abstract

The starting values considered for the model parameters strongly affect standard techniques for experimental design. When these values are far from the optimal ones, poor quality experiments are achieved or several steps are required resulting in a large experimental burden. Here, a novel criterion based on global sensitivity analysis, and therefore independent of the parameters values, is presented. In order to illustrate the performance of this methodology, a semi-continuous bioreactor is considered as a case study.

Keywords

Optimal experimental design, parameter estimation, nonlinear dynamic models, global sensitivity analysis, Sobol' sensitivity indices.

1. Introduction

In the last decade, the use of optimal experimental design has gained world-wide acceptance as an essential tool for improving the quality and efficiency of model identification and parameter estimation. Optimal experimental design (OED) allows the identification of a set of experiments with conditions and sampling schemes that deliver data from the measured variables that are the most sensitive to the unknown parameters. One of the common ways to design experiments is to use scalar functions of the Fisher's Information Matrix (FIM)

evaluated at the nominal values of the parameters. This information matrix is founded on the local sensitivity coefficients of the response variables computed for each of the sampling points. Therefore, the results of a FIM based OED depend on the nominal values used for the parameters and the behaviour of the response function is described only locally in the input space. Moreover, preliminary experiments and model calibration tests should be carried out in order to obtain a first guess for the parameter values and an iterative scheme involving both steps is required [1]. In addition, these linear methods are not sufficient for dealing with complex OED problems, especially those in which there are nonlinear interactions between parameters.

In this work, a methodology based on global sensitivity analysis (SA) for increasing the parameters' precision and/or decorrelation is presented. It consists of the minimisation of a criterion based on the Sobol' global sensitivity indices [2] using a stochastic global optimisation method leading to an optimal vector of input variables. This novel approach is here applied to a case study concerning a fed-batch reactor where a first order reaction takes place and it is used for devising the experimental conditions (dilution factor and feed substrate concentration) which allow the estimation of the reaction rate and the activation energy with the best possible statistical quality. The results obtained demonstrate its ability to reduce the quantity of experimental work required and illustrate the need for global SA techniques for the design of reliable dynamic experiments.

2. Problem Statement

Mathematically, the optimal experimental design problem can be formulated as a general dynamic optimisation problem where the objective is to find a set of input variables (controls) for the dynamic experiments that optimise some functional related with the efficacy of the experiments with respect to the parameter identifiability and the estimation accuracy expected from the collected data. This is subject to the dynamics of the system (e.g. state-space model), initial and boundary conditions and possible other algebraic constraints related to experimental limitations. To mathematically represent the time-varying controls, the control vector parametrisation (CVP) technique that approximates them using low order polynomials has been widely used [1].

Regarding the cost function, several scalar functions of the FIM evaluated at the nominal parameters have been suggested in the literature [3]. The limitations of this method arising from its local and linear nature provide an obvious motivation to devise an approach non-dependent on the value of the parameters avoiding the iterative procedure and decreasing considerably the experimental burden.

3. Methodology

In this work, the Sobol' sensitivity indices method has been implemented in Matlab. This method allows for the computation of the global sensitivity indices (SI) that consider the effects of changing parameters simultaneously while spanning the entire admissible region of the parameter space. They are defined as the ratio:

$$SI_{1,\ldots,s} = \frac{D_{1,\ldots,s}}{D}, \tag{1}$$

where $D_{1,\ldots,s}$ is the partial variance in the model output associated with simultaneous changes in parameters *1-s* and D is the total variance [2]. Fist-order sensitivity indices (SI_p) account for the sensitivity of the individual parameters and second (SI_{rp}) and higher-order ($SI_{r,\ldots,p}$) sensitivities account for the effects of interactions of two or more parameters.

To increase the precision of the parameter estimates we should seek for increasing the variance associated with the individual parameters while reducing the interactions (higher order sensitivities) [4]. Analogous to the well known FIM, we suggest the use of a matrix that we call GSIM (Global Sensitivity Information Matrix) based on these first-order SI defined as:

$$GSIM = \sum_{i=1}^{N} \left[Q(t_i)^T W_i Q_i(t_i) \right], \tag{2}$$

$$Q(t_i) = \begin{bmatrix} SI_{1,1}(t_i) & SI_{1,2}(t_i) & \cdots & SI_{1,p}(t_i) \\ SI_{2,1}(t_i) & SI_{2,2}(t_i) & \cdots & SI_{2,p}(t_i) \\ \vdots & \vdots & \ddots & \vdots \\ SI_{s,1}(t_i) & SI_{s,2}(t_i) & \cdots & SI_{s,p}(t_i) \end{bmatrix}, \tag{3}$$

where $SI_{s,p}(t_i)$ measures the sensitivity of the state y_s with respect to the parameter θ_p at the time t_i and W_i is a weighting matrix usually chosen as the measurement error covariance matrix. The optimisation can not usually be carried out in a matrix sense, but a scalar measure must be employed. For that reason, the maximisation of the determinant of the GSIM was considered as the performance index:

$$J_{OED} = \det(GSIM) \tag{4}$$

Due to the non-smoothness of the cost functions, the use of gradient-based methods to solve this NLP might lead to local solutions. Stochastic methods of

global optimisation were presented as robust alternatives for OED. For this study we have selected SRES [5] which is one of the most competitive algorithms available, with the additional advantage of being able to handle arbitrary constraints if needed [6].
In order to asses the precision of the estimated parameters, a Monte-Carlo based method was used for the calculation of their confidence intervals. To perform the analysis, a large number of pseudo-experimental data are generated. Due to measurement noise, each simulation leads to a slightly different set of data and therefore to a different set of estimated parameters. The statistical properties of the resulting distribution are then studied providing a better approximation of the confidence intervals than the standard FIM method [7]

4. Case study

To illustrate the performance of the proposed approach, we have considered a fed-batch fermentation process. Assuming Monod-type kinetics for biomass growth and substrate consumption, the system is described by the following set of differential and algebraic equations (DAEs) [1]:

$$\frac{dy_1}{dt} = (r_m - u_1)y_1 - \theta_2 y_1, \tag{5}$$

$$\frac{dy_2}{dt} = \frac{r_m y_1}{\theta_1} + u_1(u_2 - y_2), \tag{6}$$

$$r_m = \frac{0.5\ y_2}{0.5 + y_2}, \tag{7}$$

where:

- θ_i, $i = 1,2$: parameters to be estimated, $\theta_i \in [0.05,0.98]\ \forall i$,
- u_1 : dilution factor allowed to vary between 0.05 and $0.5\ h^{-1}$,
- u_2 : feed substrate concentration allowed to vary between 5 and $50\ g/l$,
- y_1 : biomass concentration. Initial value: $y_1(t_0) = 10\ g/l$,
- y_2 : substrate concentration. Initial value: $y_2(t_0) = 0.1\ g/l$,
- r_m : reaction rate.

In this study, the goal of the experimental design is to devise a 10 hours duration experiment, take up to 10 samples at equidistant times and measure the biomass concentration, y_1, and the substrate concentration, y_2, in each sample

using a High Pressure Liquid Chromatograph (HPLC). The dilution factor and the feed substrate concentration may be varied every 2 hours.
The OED problem seeks to provide the vector of values for the piecewise-constant time varying controls (feed flow rate for the dilution factor, u_1, and the feed substrate concentration, u_2) at each control step that yield the best possible information for the subsequent estimation of the two parameters. Other statements involving several experiments and/or the time horizon and sampling times as control variables could also be considered.

5. Results

The input profiles for u_1 and u_2 obtained from the maximisation of the determinant of the GSIM using SRES are shown in Figure 1. The required computational time on a 3.2 GHz Pentium 4, 2.00 GB RAM was 67 h.
From this design and taking arbitrary values for the parameters ($\theta_1 = 0.5$, $\theta_2 = 0.5$), pseudo-experimental points were generated with relative normally distributed error of 5 % (see Figure 2).

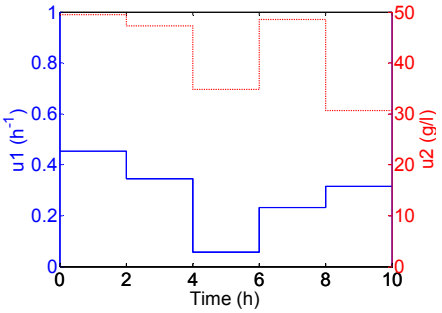

Figure 1.- Optimal input profiles.

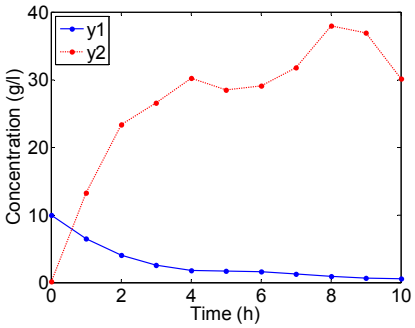

Figure 2.- Pseudo-experimental data.

The data were then considered for the estimation of the parameters by optimising the maximum likelihood function and the 95% confidence intervals were evaluated using a Monte-Carlo method. That led to $\theta_1 = 0.5 \pm 0.05$, $\theta_2 = 0.5 \pm 0.08$. The histograms obtained are presented in Figures 3-4 showing that both parameters θ_1 and θ_2 were estimated accurately.
Note that, the values of the parameters have been taken randomly and were not used for the experimental design. We have verified that equivalent results are obtained with this design for any combination of parameter values.

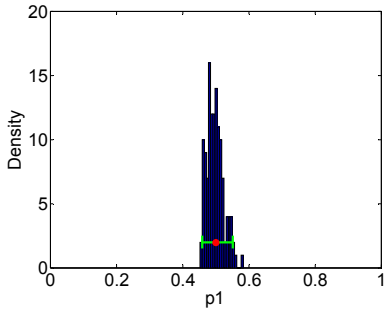

Figure 3.- Confidence interval for θ_1

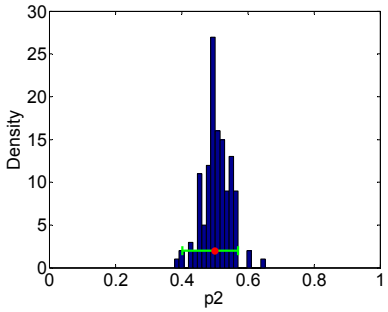

Figure 4.- Confidence interval for θ_2

6. Conclusions and future work

In this work, a novel technique for computing optimal inputs for experiments has been presented. The method makes use of the matrix of global SI for each individual parameter in order to increase parameters' precision for any value within defined bounds. The capabilities of this methodology have been illustrated with a typical semi-continuous bioreactor model. Although the computational time required for each function evaluation is higher than for the traditional approach, this technique does not involve extra experimental effort needed by the traditional one. In order to reduce the computational time, the use of the Morris method [8] for computing the SI will be considered in the near future. Moreover, the application of different scalar functions of the GSIM and their influence on the experimental design is being studied.

Acknowledgements

Financial support by the EPSRC (EP/D506743/1) is gratefully acknowledged.

References

1. S.P. Asprey and S. Macchietto, Comput. Chem. Eng., 24 (2000) 1261.
2. I. Sobol', Math. Comput. Simulat., 55 (2001) 271.
3. E. Walter and L. Pronzato. Identification of Parametric Models from Experimental Data. Springer, 1997.
4. C. Kontoravdi, S.P. Asprey, E.N. Pistikopoulos and A. Mantalaris, Biotechnol. Prog., 21 (2005) 1128.
5. T.P. Runarsson and X. Yao, IEEE Trans. Evol. Comput. 4 (2000) 284.
6. M. Rodriguez-Fernandez, P. Mendes and J.R. Banga, BioSystems, 83 (2006) 248.
7. M. Joshi, A. Seidel-Morgenstern and A. Kremling, Metab. Eng., 8 (2006) 447.
8. M.D. Morris, Technometrics, 33-2 (1991) 161.