

Nesting genetic algorithms to solve a robust optimal experimental design problem

D.J.W. De Pauw^a, P.A. Vanrolleghem^{a b}

^aBIOMATH, Department of Applied Mathematics, Biometrics and Process Control
Ghent University, Coupure Links 653, 9000 Gent, Belgium
e-mail: dirk.depauw@ugent.be

^bmodelEAU, Département de Génie Civil
Pavillon Pouliot, Université Laval, Québec, G1K 7P4, QC, Canada

Abstract: When calibrating a (dynamic) model, one is often faced with a lack of information-rich data. Without such data, there is little hope in obtaining accurate parameter estimates. In order to improve the situation, optimal experimental design for parameter estimation (OED-PE) can be employed. The main drawback of the classical OED-PE methodology is that values for the model parameters need to be provided in order to obtain an optimal design. If the values of the model parameters are highly uncertain, robust OED-PE should be preferred, yielding a design which guarantees a certain information content given the parameter uncertainty. This approach adds another level of optimization to the design problem. For each proposed experiment (optimization of the experimental degrees of freedom) an additional optimization covering the whole parameter domain needs to be performed. In this work the maximin robust OED-PE technique will be illustrated with a simple model describing substrate consumption based on Monod kinetics. The optimization problem consists of two nested real-value genetic algorithms in which each fitness evaluation for the optimization of the experimental degrees of freedom requires a full genetic algorithm optimization over the parameter domain.

Keywords: Genetic algorithms, real-coded, robust experimental design.

1 INTRODUCTION

When calibrating a (dynamic) model, one is often faced with a lack of information-rich data. Without such data, there is little hope in obtaining accurate parameter estimates and a model with sufficient predicting power. In order to improve on the situation, a methodology called optimal experimental design for parameter estimation (OED-PE) can be employed. This technique allows searching for an optimal experiment that will yield information-rich data when actually performed. Finding this optimal experiment involves an optimization of the available experimental degrees of freedom (e.g. measurement locations, experimental manipulations, ...). The objective surface for this type of optimization problems is known to contain many local minima. Therefore it is advisable to use global optimization algorithms (like genetic algorithms) in order to increase the chance of finding the global minimum, and thus the optimal experiment. However, employing genetic algorithms alone will not guarantee

the optimal experiment to be truly optimal. This is caused by the fact that the design itself is dependent on the parameter values of the model. This poses a problem, since the final goal of optimal experimental design for parameter estimation is to find the true system parameters, or at least approximate them. This issue is a fundamental one for non-linear design problems and is the basis for the need for robust experimental design which tries to deal with this parameter uncertainty.

2 MODEL DESCRIPTION

Throughout this paper a simple model of a fed-batch reactor will be used for illustrative purposes. This model describes substrate (S) consumption and biomass (X) growth based on the well known Michaelis-Menten or Monod kinetics:

$$\mu(S) = \frac{\mu_{max} \times S}{K_S + S} \quad (1)$$

in which the microbial growth rate μ (h^{-1}) is related to the substrate concentration S ($g.l^{-1}$) and is characterized by a maximum growth rate μ_{max} (h^{-1}) and a saturation constant K_S ($g.l^{-1}$) at which the growth rate is half of the maximum growth rate.

The equations describing the substrate, biomass and volume changes in the fed-batch reactor are given by:

$$\frac{dX}{dt} = (\mu - K_d) \times X - \frac{Q_f}{V} \times X \quad (2)$$

$$\frac{dS}{dt} = -\frac{\mu}{Y} \times X + \frac{Q_f}{V} \times (S_f - S) \quad (3)$$

$$\frac{dV}{dt} = Q_f \quad (4)$$

where K_d (h^{-1}) is a constant decay rate, V (l) the reactor volume and Y (gX/gS) a yield coefficient. Q_f ($l.h^{-1}$) is the flow rate of a feed pump which is switched on at a certain time t_f (h) and adds substrate with concentration S_f ($g.l^{-1}$) to the reactor.

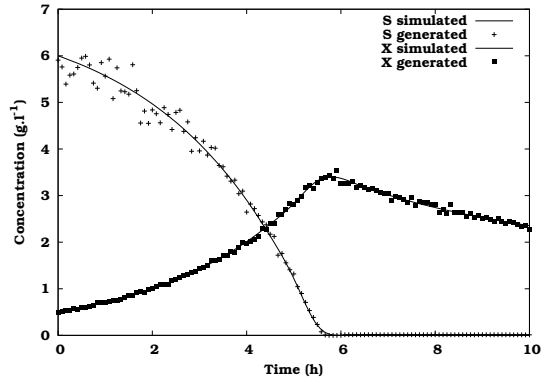


Figure 1. Model simulation results and artificially generated measurements for substrate (S) and biomass (X).

Figure 1 illustrates the dynamic behavior of the substrate and biomass concentration in the reactor. The reactor volume and initial concentrations of substrate and biomass were taken to be 15 l, 6 $g.l^{-1}$ and 0.5 $g.l^{-1}$ respectively. The reactor is operated in batch mode until 6 h at which time a feed pump with flowrate 2.4 $l.h^{-1}$ and feed concentration 4 $g.l^{-1}$ is switched on. The values of the other, biomass related, parameters can be found in Table 1. This system and its parameters will be considered the "true" system and based on it, artificial data was generated with a measurement interval of 5 minutes and assuming a measurement error of 5 and 2 % for biomass and substrate respectively [Baltes et al., 1994]. These data are also shown in Figure 1.

Table 1. True parameter values used for data generation, lower and upper bounds and estimated parameter values resulting from a maximin robust experimental design. The 95 % confidence intervals of the parameter estimates are also shown.

	μ_{max}	K_S	Y	K_d
True	0.39	0.06	0.54	0.037
Lower bound	0.2835	0.0466	0.2313	0
Upper bound	0.4241	0.0788	0.6167	0.0675
Estimated	0.3857	0.0597	0.5327	0.0336
95% conf. int.	0.0025	0.0029	0.0054	0.0036

3 OPTIMAL EXPERIMENTAL DESIGN

3.1 Classical non-linear design

A successful model calibration can only be achieved when information rich data is available. It would therefore be beneficial to design experiments in such a way that the data which will be collected from these experiments are as information-rich as possible. In order to design an experiment, different choices have to be made. It has to be decided whether, where and how the system under study will be manipulated and where, how and when measurements will be performed on this system. The classical scheme of non-linear optimal experimental design is shown in Figure 2. An iterative search for the optimal experiment is started based on an initial and insufficiently calibrated model. Using this model, several different experiments are simulated by varying the degrees of freedom within the applicable constraints and calculating an objective value or design criterion. The optimal experiment is found when the selected design criterion is maximized or minimized, depending on the chosen objective. The final experiment is said to be locally optimal because its design was based on a model using certain nominal parameter values. Once the optimal experiment is found, it can be performed in reality, resulting in new data. Based on these data the model can be recalibrated and the quality of the parameter estimates evaluated. If required, another iteration of the design loop can be performed, potentially leading to an even better experiment. It is beyond the scope of this paper to provide a detailed description of each step in the procedure and the reader is therefore referred to Dochain and Vanrolleghem [2001] for more details.

Probably, the most important aspect of the procedure is the evaluation of the objective. For each of the proposed experiments the information content needs to be quantified. This can be achieved by cal-

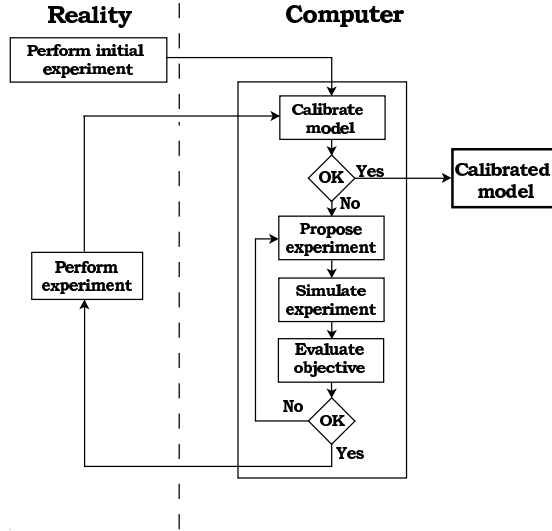


Figure 2. Schematic representation of the optimal experimental design procedure (adapted from Dochain and Vanrolleghem [2001]).

culating a scalar property of the Fisher Information Matrix (FIM):

$$FIM = \sum_{i=1}^N \left(\frac{\partial y(\theta)_i}{\partial \theta} \right)^T \times Q_i^{-1} \times \left(\frac{\partial y(\theta)_i}{\partial \theta} \right) \quad (5)$$

This matrix quantifies the information content of an experiment related to the model parameters in the sense that it combines several aspects: on the one hand the quantity (sum over N measurement points) and quality (measurement error covariance matrix Q_i) of the data and on the other hand the sensitivity of the model variables to the parameters ($\partial y(\theta)_i / \partial \theta$). In our case, these sensitivities are calculated using a finite difference approximation, requiring two model simulations for each considered parameter. Since the FIM is inversely proportional to the parameter estimation error covariance matrix, its eigenvalues and eigenvectors directly relate to the eigenvalues of the parameter estimation error covariance matrix and thus to the shape, size and orientation of the confidence region of the parameter estimates. Different scalar properties based on FIM eigenvalues are used in order to quantify the information content of a certain experiment. The most often used FIM property is its determinant or product of its eigenvalues (D-optimal design, $\psi_D(\theta)$). This design criterion needs to be maximized in order to minimize the overall volume of the confidence region of the estimated parameters. Maximization of the design criterion can be achieved by an appropriate choice of the experimental degrees of freedom

(Ψ):

$$\psi_D(\theta) = \arg \left\{ \max_{\psi \in \Psi} \det [FIM(\theta, \psi)] \right\} \quad (6)$$

3.2 Robust experimental design

As was already mentioned above, the experimental design procedure is an iterative one. Each iteration and parameter estimation improves the knowledge of the system parameters and this knowledge can then be used to improve the quality of the next experiment to be performed. Many authors acknowledge the usefulness of this approach (e.g. Ford et al., 1989; Walter and Pronzato, 1990; Atkinson, 2003). However, some drawbacks are associated with this technique. Firstly, it might not be possible to perform multiple (sequential) experiments on the same system due to limitations in time or resources. Secondly, it is not guaranteed that the parameters will converge to the "true" values. These problems lead to the development of robust experimental design which tries to deal with the uncertainty associated with the model parameters.

A very often used robust experimental design method is the maximin design (MMD) approach. It aims at determining the experiment ψ_{MMD} which optimizes the worst possible performance for any value of θ belonging to the parameter domain Θ [Pronzato and Walter, 1988]:

$$\psi_{MMD}(\theta) = \arg \left\{ \max_{\psi \in \Psi} \left[\min_{\theta \in \Theta} [\det [FIM(\theta, \psi)]] \right] \right\} \quad (7)$$

In other words, for a proposed design, find the parameters for which the D-optimal criterion value is the lowest, i.e. this D-criterion value determines the worst possible obtainable information content for this specific design. Next, find the design which maximizes this worst D-optimal criterion value. For this technique, the prior information on the parameters is limited to the knowledge of the parameter domain Θ , i.e. the upper and lower bounds of the parameters. This approach is only recently being applied to dynamic non-linear models due to the computational burden introduced by the nested optimization.

In our case we propose to implement this algorithm by nesting two genetic algorithms. A schematic overview of the implemented algorithm is shown in Figure 3.

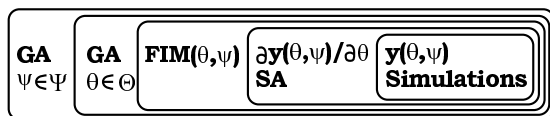


Figure 3. Schematic overview of the implemented maximin algorithm for robust experimental design.

The inner loop GA is used to find the parameter set which minimizes the D-optimal criterion (for a certain design). Each of the generated GA individuals (representing different parameter combinations) requires the calculation of a FIM D-optimal criterion value and thus also a sensitivity analysis requiring several model simulations.

The minimal D-optimal criterion value which is found using the inner GA optimization can then be used as the objective for the outer loop GA which optimizes the experimental degrees of freedom. In this way, the outer optimization loop finds the design which maximizes the worst performing experiment.

4 GENETIC ALGORITHMS

A genetic algorithm (GA) can be considered as a global optimization technique based on the genetic processes of life. Over many generations, natural populations evolve according to the principles of natural selection and "survival of the fittest". The mathematical analogon of these natural processes was first developed by Holland [1975] and is described in detail in many other texts (e.g. Goldberg [1989]; Michalewicz [1992]). In contrast to the "classical" optimization algorithms, GAs work with populations of individuals and not single individuals, each representing a possible solution to a given problem. Using processes like selection, mutation and crossing-over, a population of individuals is evolved until convergence. One way to describe convergence is when the average fitness of all individuals in the population closely approaches or becomes equal to the fitness of the best individual.

The often used binary representation of GAs exhibits some drawbacks when applied to multidimensional, continuous problems requiring solutions with high numerical precision. Therefore a real-coded GA will be used here. The basic idea behind real-coded GAs is that each variable of the optimization problem is represented by one real-value gene in the GA chromosome. This representation is particularly natural to describe optimization problems with variables in continuous domains. Examples of applications and in-depth information on real-coded GAs can be found in the excellent review

of Herrera et al. [1998]. The use of real-coded GAs makes it possible to use large or even unknown domains for the variables while the precision is only restricted to that of the computer on which the algorithm is executed. Since real-coded GAs are based on continuous variables, slight changes in the variables also cause slight changes in the objective function. This gives real-coded GAs the ability to locally "fine-tune" the solution.

5 EXPERIMENTAL DESIGN RESULTS

5.1 Algorithm settings

Based on the above provided description of the fed-batch reactor a robust experimental design using the MMD approach was performed.

The experimental degrees of freedom considered in this case are the time instances of 10 measurements for both substrate and biomass (20 outer loop optimization variables). For these measurements, no repetitions are allowed and a minimal period of 5 minutes between two consecutive measurements should be maintained. Further, it is assumed that all measurements are conducted within a timespan of 10 h.

The target parameters for this design were: μ_{max} , K_S , Y and K_d (4 inner loop optimization variables). The parameter search space was defined by the lower and upper bounds shown in Table 1. These bounds represent the 99 % confidence limits obtained after model calibration based on a classical D-optimal design (results not shown). This classical design was performed on the model with parameters values far from the true values. This step could have been skipped in which case the parameter bounds should have been derived from literature.

Both optimization loops of the MMD approach were optimized using (nested) real-coded GAs. The outer loop was responsible for the optimization of the measurement time instances while the inner loop dealt with the optimization over the parameter space. Beside the population sizes, the settings of both GAs were identical and are shown in Table 2. More information about the specific GA operators used can be found in Herrera et al. [1998]. Population sizes were chosen to be 5 and 10 times the number of optimization variables for the outer and inner loop respectively.

5.2 Performing the experimental design

Once the algorithm settings were fixed, the actual optimization was started. Figure 4 shows a

Table 2. GA settings used for the minimax robust experimental design.

Attribute	Value
Outer loop pop. size	100
Inner loop pop. size	40
Replacement %	10%
Selection scheme	Roulette wheel
Scaling	Linear (multiplier: 1.2)
Crossover probability	0.8
Crossover operator	BLX- α ($\alpha = 0.5$)
Mutation probability	0.01
Mutation operator	Non-uniform

small extract of the evolution of the objective value ($\det[FIM]$) as a function of the evaluation number. This figure shows the evaluation of 4 individuals from the outer loop optimization (measurement schemes). Each of these individuals represents a certain measurement scheme for which the fitness is calculated based on the endresult of a full inner loop GA evaluation (parameter space). For one outer loop individual, the corresponding inner loop optimization is indicated by a two-way arrow. It can clearly be seen from this section of the figure that the inner loop GA is able to minimize the objective according to the MMD requirements without convergence problems. A factor which strongly influences the convergence behavior was found to be the population size. It is therefore important to make an appropriate choice for this value, in this case 10 times the number of optimization variables proved to be useful. However, further increasing this value might be advisable but would result in excessive calculation times.

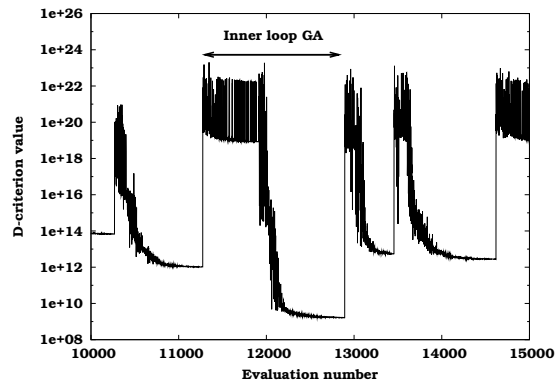


Figure 4. Extract of the evolution of the D-criterion value during the nested GA optimization.

The optimization converged after approximately 3000 outer loop individual evaluations, each requiring approximately 1000 inner loop individual evaluations (i.e. 1000 FIM evaluations). At the optimal

measurement instances found after optimization (D-optimal criterion value: 1.87×10^{24}), the virtual data was sub-sampled and used for calibration. The optimal measurement instances are shown in Figures 5 and 6 for substrate and biomass respectively together with model simulations using the "true" parameter values. It can be seen that the measurements are spread out over a broad region, as was expected for a robust design. In this way, data with high information content is guaranteed for all systems described by parameters within the investigated parameter space.

The results of the calibration based on the optimal measurement time instances are shown in Table 1. This shows that the calibrated parameter values are close to the true values and that the confidence regions are quite small.

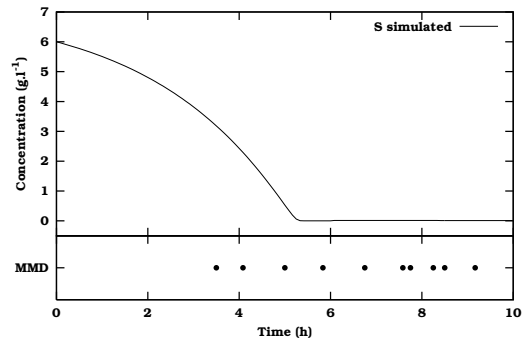


Figure 5. Overview of the optimal measurement times for substrate for the tested design criteria including the model simulations using the "true" values.

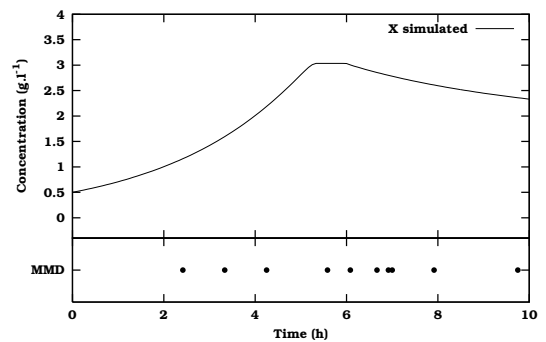


Figure 6. Overview of the optimal measurement times for biomass for the tested design criteria including the model simulations using the "true" values.

5.3 Computational burden

As already mentioned, the optimization algorithm converged after approximately 1.0×10^6 objective evaluations (3000 outer loop evaluations, each requiring approximately 1000 inner loop individual evaluations). Each of these objective evaluations corresponds to one FIM calculation and sensitivity analysis requiring two model simulations for each considered parameter. In this case, 4 parameters were considered. Therefore each FIM calculation required 8 model simulations. Taking this into account, the entire optimization required approximately 8.0×10^6 model evaluations. On a 3.0 GHz Pentium 4 machine this resulted in a calculation time of approximately 2.5 days. Therefore, applications of these techniques will remain, for now, limited to relatively simple models. Larger models can only be considered when the calculations, especially related to the inner loop, could be accelerated. One possibility would be to replace the numerical finite difference sensitivity analysis approximation by its analytical solution. Another solution would be to use or develop more efficient global optimization techniques.

6 CONCLUSION

This paper has shown that the difficult optimization problem related to robust optimal experimental design could be solved by nesting two real-coded genetic algorithms. An outer loop GA is responsible for the optimization of the experimental degrees of freedom while an inner loop GA deals with the optimization over the parameter space. An illustration with a model describing substrate consumption based on Monod kinetics was given. Even with this simple model, computational requirements were large. It was found that the GA population size plays a crucial role in finding the right balance between convergence behavior of the GA and computational demand. Future applications to more complicated models will only be feasible when the calculations related to the inner loop are accelerated. This could be accomplished by decreasing the integration time of the differential equations by using more efficient integrators. Another way to decrease the computational burden of the inner loop would be to replace the finite difference approximation of the sensitivity calculations by a more efficient technique, e.g. the direct differential method [Atherton et al., 1975]. When these improvements are still not adequate, it might be appropriate to use hybrid optimization algorithms in which a genetic algorithm is only used to pinpoint regions that are likely to

contain a minimum. These regions could then be explored by a faster converging local optimization algorithm to quickly determine the minimum.

REFERENCES

- Atherton, R., R. Schainker, and E. Ducot. Statistical sensitivity analysis of models for chemical kinetics. *AIChE Journal*, 21(3):441–448, 1975.
- Atkinson, A. Horwitz's rule, transforming both sides and the design of experiments for mechanistic models. *Journal Of The Royal Statistical Society Series C - Applied Statistics*, 52(3):261–278, 2003.
- Baltes, M., R. Schneider, C. Sturm, and M. Reuss. Optimal experimental design for parameter estimation in unstructured growth models. *Biotechnology Progress*, 10(5):480–488, 1994.
- Dochain, D. and P. Vanrolleghem. *Dynamical Modelling and Estimation in Wastewater Treatment Processes*. IWA Publishing, London, UK, 2001.
- Ford, I., D. Titterton, and C. Kitsos. Recent advances in nonlinear experimental design. *Technometrics*, 31(1):49–60, 1989.
- Goldberg, D. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, Massachusetts, USA, 1989.
- Herrera, F., M. Lozano, and J. Verdegay. Tackling real-coded genetic algorithms: Operators and tools for behavioural analysis. *Artificial Intelligence Review*, 12(4):265–319, 1998.
- Holland, J. *Adaption in Natural and Artificial Systems*. MIT Press, Michigan, USA, 1975. EXT.
- Michalewicz, Z. *Genetic Algorithms + Data Structures = Evolution Programs*. Springer-Verlag, Berlin, Germany, 1992.
- Pronzato, L. and E. Walter. Robust experiment design via maximin optimization. *Mathematical Biosciences*, 89(2):161–176, 1988.
- Walter, E. and L. Pronzato. Qualitative and quantitative experiment design for phenomenological models - a survey. *Automatica*, 26(2):195–213, 1990.