OPTIMAL EXPERIMENT DESIGN IN BIOPROCESS MODELING: FROM THEORY TO PRACTICE

Astrid M. Cappuyns * **Kristel Bernaerts** *
Ilse Y. Smets * **Ositadinma Ona** ** **Els Prinsen** ***
Jos Vanderleyden ** **Jan F. Van Impe** *

* *BioTeC - Katholieke Universiteit Leuven,*
W. de Croylaan 46, B-3001 Leuven (Belgium)
Fax: +32-16-32.29.91
email: jan.vanimpe@cit.kuleuven.be
** *CMPG - Katholieke Universiteit Leuven,*
Kasteelpark Arenberg 20, B-3001 Leuven (Belgium)
*** *Department of biology, University of Antwerp,*
Universiteitsplein 1, B-2610 Antwerpen (Belgium)

Abstract: In this paper the problem of parameter identification for the Monod model is considered. As known for a long time, noisy batch measurements do not allow unique and accurate estimation of the kinetic parameters of the Monod model. Techniques of optimal experiment design are, therefore, addressed to design informative experiments and improve the parameter estimation accuracy. During the design process, practical feasibility has to be kept in mind. In this paper it is demonstrated how a theoretical optimal design can successfully be translated to a feasible optimal design. Both design and validation of informative fed batch experiments are illustrated with a case study that models the growth of the nitrogen fixing bacteria *Azospirillum brasilense*. Copyright ©2006 IFAC

Keywords: optimal experiment design, parameter identification, Monod kinetics, *Azospirillum brasilense*, bioreactor

1. INTRODUCTION

When modeling (bio)chemical processes, some limitations which will hamper the model identification process, have to be kept in mind (Bernaerts and Van Impe, 2004). To overcome these problems, an accurate design of the experiments is needed. Experimental data should contain sufficient information in order to enable correct model structure characterization, and accurate and unique parameter estimation. It has been demonstrated that the use of optimal experiment design for parameter estimation can contribute to an improvement of the parameter estimation ac-

curacy (e.g., Walter and Pronzato, 1997; Versyck and Van Impe, 1999).

Azospirillum brasilense belongs to a group of bacteria that exert beneficial effects on plant growth. One of the factors responsible for the plant growth promotion is the production of phytohormones, e.g., the auxin indole-3-acetic acid (Baldani *et al.*, 1983). This characteristic opens perspectives to exploit the nitrogen fixing bacteria of the genus *Azospirillum* as alternative for, or supplement to chemical fertilization. Therefore, a quantitative analysis of growth and phytohormone production by *Azospirillum brasilense* is very interesting

(Smets *et al.*, 2004). In this paper the modeling of growth and more specifically the estimation of the Monod kinetic parameters, is addressed.

In this case study the feed rate profile of a fed batch is optimized to enable accurate estimation of the growth model parameters. The designed optimal experiment requires advanced control to be realized in practice. The case study illustrates how, in a second stage, a trade-off can be found between maximum information content and practical feasibility of the experiment.

The structure of this paper is as follows. First, the material and methods and the theoretical background of optimal experiment design are introduced in Section 2. In Section 3 the implementation of optimal experiments is discussed. Finally, Section 4 summarizes the major conclusions.

2. MATERIALS AND METHODS

2.1 Bioreactor experiments

Experiments were performed in a computer controlled BioFlo 3000 benchtop fermentor (New Brunswick Scientific, USA) with an autoclavable vessel of 1.25 to 5L working volume. 100 mL of a preculture containing *Azospirillum brasilense* was transferred to the vessel containing a minimal malate medium (MMAB) (Vanstockem *et al.*, 1987). L-malate is provided as sole carbon source. PID cascade controllers ensure that the fermentation temperature is kept constant at 30°C, pH at 6.3 and the dissolved oxygen concentration at 3% (micro-aerobic range).

Culture media samples were removed at regular intervals. Cell density was obtained through measurement of optical density (OD) at 600 nm (Genesis 10S, Thermo Spectronic). L-malate was measured using test kits from Roche (R-biopharm, Germany).

2.2 Growth model

The evolution of biomass C_X [OD] and substrate concentration C_S [g/L] in a fed batch reactor can be described by following mass balance type equations:

$$\frac{dC_X}{dt} = \mu \cdot C_X - \frac{U}{V} \cdot C_X$$
$$\frac{dC_S}{dt} = -\sigma \cdot C_X + \frac{U}{V} \cdot (C_{S,in} - C_S) \qquad (1)$$
$$\frac{dV}{dt} = U$$

with

$$\mu = \mu_{max} \cdot \frac{C_S}{C_S + K_M} \qquad (2)$$
$$\sigma = \frac{\mu}{Y_{X/S}} + m \qquad (3)$$

with V [L] the volume of the liquid phase and $C_{S,in}$ [g/L] the substrate concentration in the volumetric feed rate U [L/h]. μ [h^{-1}] is the specific growth rate and is specified by the Monod equation (2) with μ_{max} [h^{-1}] the maximum specific growth rate and K_M [g/L] the half-saturation constant. The relation between specific growth rate μ [h^{-1}] and specific consumption rate σ [(g/(OD·L))·h^{-1}] is given by the linear law (3). $Y_{X/S}$ [(OD·L)/g] is a yield coefficient of biomass over substrate and m [(g/(OD·L))·h^{-1}] represents a maintenance factor. In this case study, maintenance is, in a first stage, neglected ($m=0$).

2.3 Parameter estimation

The identification cost imposed for parameter estimation is the sum of squared errors SSE:

$$SSE = \sum_{i=1}^{n} (y_{exp}(t_i) - y_{model}(t_i))^2 \qquad (4)$$

with $y_{model}(t_i)$ the model predictions, $y_{exp}(t_i)$ the experimental observations and n the number of samples.

As m is set equal to zero, the yield coefficient $Y_{X/S}$ can be estimated separately by eliminating the specific growth rate from the growth model:

$$\frac{dZ}{dt} = \frac{U}{V} \cdot (Y_{X/S} \cdot C_{S,in} - Z) \qquad (5)$$

with

$$Z = Y_{X/S} \cdot C_S + C_X$$

This leaves two growth parameters to be estimated, i.e., K_M and μ_{max}, together with the initial conditions $C_X(0)$ and $C_S(0)$.

The implemented identification routines for model parameter identification are the `E04UCF` routine from the NAG library (Numerical Algorithms Group) in Fortran and the `lsqnonlin` routine in Matlab (The Mathworks Inc., Natick). Numerical integration is performed with the NAG-routine `D02EJF` in Fortran.

2.4 Optimal experiment design

The information content of an experiment, with respect to parameter identification, can be evaluated through the Fisher information matrix **F** (e.g., Walter and Pronzato, 1997):

$$\mathbf{F} \triangleq \int_0^{t_f} \left(\frac{\partial \mathbf{y}}{\partial \theta}\right)^T \mathbf{Q} \left(\frac{\partial \mathbf{y}}{\partial \theta}\right) \cdot dt \qquad (6)$$

The main components of the Fisher information matrix $\mathbf{F}$ are the model output sensitivities $\frac{\partial \mathbf{y}}{\partial \theta}$, and the uncertainty of the measurements. The latter is represented by the weighting matrix $\mathbf{Q}$, which is set equal to the inverse of the measurement error covariance matrix. The model output sensitivities reflect the sensitivity of the model output $\mathbf{y}$ to small variations of the parameters θ.

Depending on the requirements imposed by the application, a specific scalar function of $\mathbf{F}$ is used as performance index for optimal experiment design. Different design criteria are available and the choice of the criterion will influence the resulting design (Walter and Pronzato, 1997; Vanrolleghem and Dochain, 1998). In this case study, the modified E-criterion is selected. This criterion aims at the minimization of the condition number of $\mathbf{F}$, i.e., the ratio of the largest to the smallest eigenvalue of $\mathbf{F}$.

$$\Lambda(\mathbf{F}) = \frac{\lambda_{max}(\mathbf{F})}{\lambda_{min}(\mathbf{F})} \qquad (7)$$

The objective is to have eigenvalues as close as possible to each other. When the condition number reaches its minimal value, i.e., $\Lambda(\mathbf{F}) = 1$, the contour lines of the cost surface for a two parameter problem will be circular. Such highly informative experiment allows unique parameter estimation. Values of $\Lambda(\mathbf{F})$ greater than 1 induce ellipsoid contour lines.

Given the nonlinear model structure, the design also depends on the nominal parameter values, i.e., the initial guess for the unknown parameters, used in the optimization. The closer the nominal values approach the true parameter values, the better the obtained design. Optimal experiment design is, therefore, an iterative procedure. After evaluation, the design is implemented. The resulting experimental data and identified parameters are used as a basis for a next round of optimal experiment design.

3. RESULTS AND DISCUSSION

3.1 Parameter identification from batch data

Nihtilä and Virkkunen (1977) showed that the parameters of the Monod kinetics cannot be uniquely identified from noisy batch measurements. This is illustrated in Figure 1. The upper plot depicts the experimental data of a preliminary batch experiment together with one of the many possible solutions of the parameter estimation problem. The identification problem is

illustrated in the lower plot. Joint and individual uncertainty are very large for the kinetic parameters K_M and μ_{max}. The contour plot reveals a valley with different parameter combinations which result in an equally low cost. This means that a change in one of the parameters can be compensated by a change in the other parameter.

To overcome this problem, new experiments need to be designed which are more informative in the sense of accurate parameter estimation. Techniques of optimal experiment design for parameter estimation are addressed to tackle this problem.
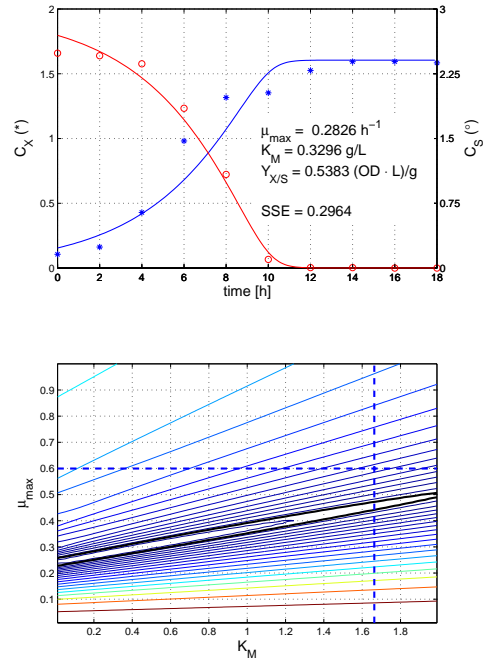

Fig. 1. Model parameter identification from a batch experiment. Upper plot: experimental data $(\ast, \circ)$ and model predictions $(-)$ for biomass and substrate concentrations. Lower plot: contours of equal identification cost (SSE) as function of the model parameters K_M and μ_{max}. The bold line is the 95% joint confidence region, the dashed lines depict the 95% individual confidence intervals on K_M and μ_{max}.

3.2 Feeding profile

The optimal control problem is to find the best possible admissible feed rate profile $U(t)$ with respect to the quality of the estimates for the Monod parameters K_M and μ_{max}. Van Impe et al. (1995) formulated following conjecture:

> *A feed rate strategy which is optimal in the sense of process performance, is an excellent starting point for feed rate optimization with respect to estimation of those parameters with large influence upon process performance.*

A feeding profile optimal for process performance is one in which substrate concentration is kept constant from the beginning. For unique parameter estimation an extra perturbation is required, which can be achieved by preceding the singular feeding phase by an initial batch phase (Versyck and Impe, 1999). The structure of this feeding profile is depicted in Figure 2 (dashed line). In the feeding phase the feed rate $U(t)$ is given by a feed forward control law of the form:

$$U_{sing}(t) = \frac{\sigma C_X V}{C_{S,in} - C_S^*} \qquad (8)$$

with $C_{S,in}$ the limiting substrate concentration in the feeding solution and C_S^* the constant substrate concentration aimed at during the feeding phase. There are two degrees of freedom in this feed rate optimization problem, i.e., the initial substrate concentration $C_S(0)$ and the substrate concentration during the feeding phase C_S^*.
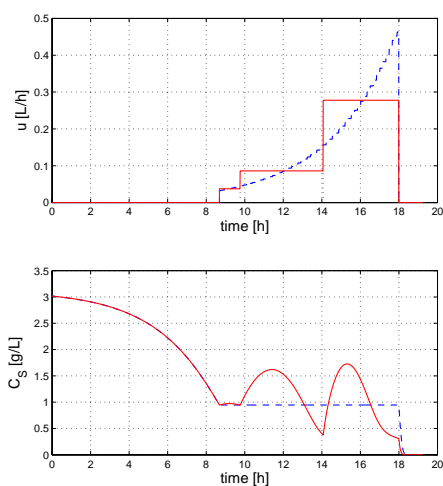

Fig. 2. Singular (- -) and 3-step (—) feed rate profile for optimal growth parameter identification.

3.3 optimization of the feeding profile

The parameters obtained from an initial batch experiment are used as nominal values for the design of a new and more informative experiment. Different combinations of $C_S(0)$ and C_S^* were found which give a condition number equal to 1. These profiles entail, however, some important practical problems.

A first problem are the values found for $C_S(0)$ and C_S^*. The obtained concentrations are very low and hard to realize in practice. For fixed values of $C_S(0)$, the optimal C_S^* and corresponding condition number were computed. A higher concentration for $C_S(0)$ yields a higher C_S^*. This way a range of designs with suboptimal $(C_S(0), C_S^*)$ combinations was defined, which are still informative enough with regard to parameter identification.

As a second step to increase practical feasibility, the time-varying feeding profile was simplified by replacing the singular feeding phase by steps of constant feed rate. This step approximation is done in such a way that the volume added per step of feeding is the same as for the time-varying feeding profile in that period:

$$U_{cte}(t, \theta) = \frac{\int_{t_i}^{t_i+1} u_{sing}(t, \theta) dt}{t_{i+1} - t_i} \qquad (9)$$

Profiles with three as well as with one step were computed. An example of a feed rate profile with three steps is illustrated in Figure 2 (solid line). For this example there are two degrees of freedom to be optimized, i.e., the time points for switching from one feed rate to the next (t_1 and t_2). A profile with one step of constant feeding has only one degree of freedom, i.e., the time instant t_f at which the feeding stops. The resulting condition numbers for the different optimal and suboptimal feeding profiles are listed in Table 1.

Table 1. Overview of different designs.

Feeding profile	Condition number $\Lambda(\mathbf{F})$
Unconstrained singular profile	
$C_S(0)$= 0.4340 g/L	1.00
C_S^*= 0.1702 g/L	
Constrained singular profile	
$C_S(0)$= 3.0167 g/L	50.29
C_S^*= 0.9463 g/L	
Simplifications of the constrained singular profile	
3 steps:	
t_1= 9.76 h	62.13
t_2= 14.07 h	
1 step:	
t_f= 16.35 h	321

with $C_{S,in} = 50$ g/L and nominal parameter values μ_{max}=0.421 h^{-1}, K_M=0.439 g/L, Y_{XS}=0.777 (OD·L)/g

To evaluate the loss of information content through simplification of the feed rate profile, the different optimal and simplified designs were extensively tested through identification of the growth parameters on simulated noisy data. All four designs listed in Table 1 delivered satisfying results concerning accurate parameter identification. Therefore, the most simple design with regard to practical feasibility, i.e., the design with only one step of constant feeding, was selected for implementation.

3.4 Implementation and validation

The results presented in this section were obtained in a second round of optimal experiment design. The results of the performed fed batch experiment are illustrated in the upper plot of Figure 3 and the identified growth parameters are summarized in Table 2. A malate solution of 10 g/L was added to the reactor with a feed rate of 0.07

L/h starting three hours after the start of the experiment. The period of feeding is represented by the vertical dashed lines on the plot. The lower plot depicts the contour plot together with the 95% individual confidence intervals and 95% joint confidence region for μ_{max} and K_M. The 95% joint confidence region now forms a closed ellips. Comparing the confidence intervals with the ones calculated for the batch experiments (see Figure 1, lower plot), confirms that the estimation accuracy of the parameters K_M and μ_{max} is significantly improved.

The predictive quality of the obtained model parameters was subsequently evaluated by comparing simulations with the identified model parameters and experimental data[1]. Hereto, a new fed batch experiment with a different feeding profile was performed. The feed rate for this experiment was also 0.07 L/h, but the period of feeding was shifted. Additionally, the data of the initial batch experiment were used for evaluation. These validation results are depicted in Figure 4.
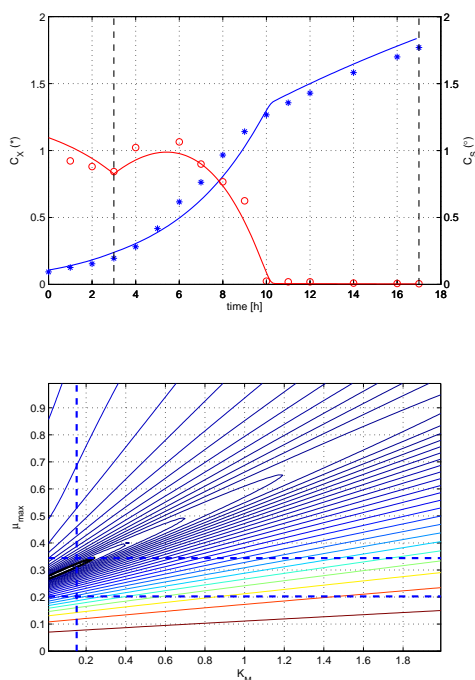

Fig. 3. Model parameter identification from an informative fed batch experiment. Upper plot: experimental data ($*$,$\circ$) and model predictions ($-$) for biomass and substrate concentrations. Lower plot: contours of equal identification cost (SSE) as function of the model parameters K_M and μ_{max}. The bold line is the 95% joint confidence region, the dashed lines depict the 95% individual confidence intervals on K_M and μ_{max}.

---

[1] The initial conditions ($C_X(0)$ and $C_S(0)$) have been reestimated for each simulation.

Table 2. Parameter values for the model (1,2,3) with and without maintenance.

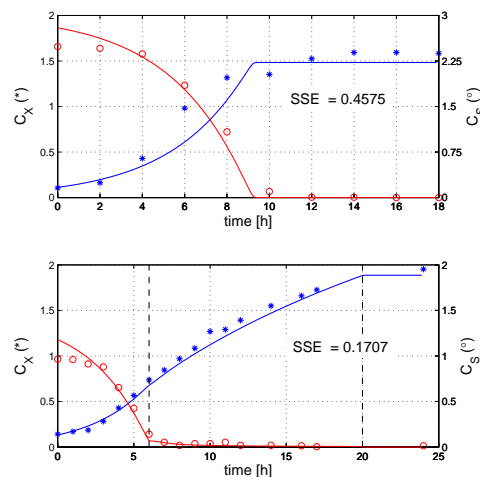|	no maintenance	with maintenance
$Y_{X/S}$	0.4905	0.5468
μ_{max}	0.2733	0.2961
K_M	$1.441 \cdot 10^{-2}$	$4.163 \cdot 10^{-2}$
m_S	-	$1.445 \cdot 10^{-2}$
SSE	0.1626	0.1362

Fig. 4. Experimental data ($*$,$\circ$) and simulation of the growth model ($-$) on batch (upper plot) and fed batch experiments (lower plot). Parameters are taken from Table 2.

3.5 Remarks concerning the growth model

The growth model for *Azospirillum brasilense* presented in this paper, started from the assumption that maintenance can be neglected. The available experimental data, however, do not allow to determine whether maintenance can be omitted or not.

This shortcoming is illustrated by identification of the parameters for the model (1,2,3) taking maintenance into account. Expression (5) cannot be used, in this case, to estimate the yield coefficient $Y_{X/S}$. Here, the four parameters (μ_{max}, K_M, $Y_{X/S}$ and m) have to be identified simultaneously. The results are presented in Table 2 and in Figure 5. The model with maintenance seems to provide a better description of the last hours (10h till 17h) of the fed batch experiment, while the validation results (see Figure 6) are less good for that same period. The model with maintenance predicts a decrease in biomass concentration after depletion of the substrate. This phenomenon is, however, not observed in the data.

Another problem of the model is the overestimation of the initial substrate concentration $C_S(0)$. The estimated values for $C_S(0)$ are consistently higher than the experimental values. The consumption of malate in the first hours of the experiments seems to exhibit a delay or lag which cannot be described by the model.
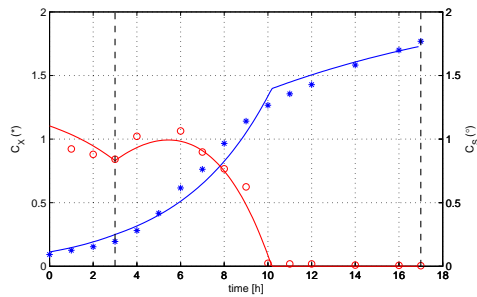
Fig. 5. Identification of parameters for a model including maintenance: experimental data (∗,◦) and model prediction (−) for biomass and substrate concentrations.
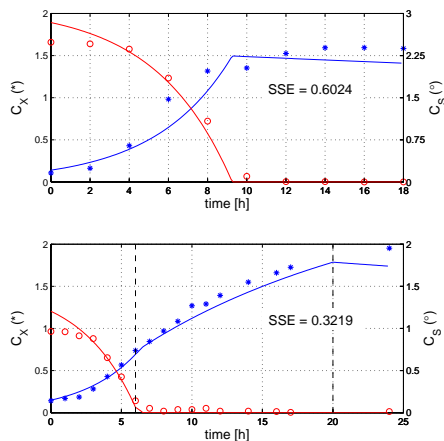

Fig. 6. Experimental data (∗,◦) and simulation of the growth model including maintenance (–) on batch (upper plot) and fed batch experiments (lower plot).

Although the model does have some shortcomings as mentioned above, it provides an accurate description of the transitions of one growth phase to another. The current model is in its most simple form, and shall be extended in a further stage to overcome these problems.

4. CONCLUSIONS

This paper presents a successful validation of optimal experiment design for parameter identification for the Monod kinetics. Due to some practical limitations, a trade-off has to be made between maximum information content and practical feasibility. Theoretical optimal designs have successfully been translated to feasible (sub)optimal designs by imposing constraints on substrate concentrations and simplifying the feeding phase. With only a few additional experiments the accuracy of the kinetic parameters was significantly increased as illustrated by the individual confidence intervals and joint confidence regions.

REFERENCES

Baldani, V.L.D., J.I. Baldani and J. Döbereiner (1983). Effects of *Azospirillum* inoculation on root infection and nitrogen incorporation in wheat. *Canadian Journal of Microbiology* **29**, 924–929.

Bernaerts, K. and J.F. Van Impe (2004). Data-driven approaches to the modelling of bioprocesses. *Transactions of the Institute of Measurement and Control* **26**(5), 349–372.

Nihtilä, M. and J. Virkkunen (1977). Practical identifiability of growth and substrate consumption models. *Biotechnology and Bioengineering* **19**, 1831–1850.

Smets, I., K. Bernaerts, A. Cappuyns, O. Ona, J. Vanderleyden, E. Prinsen and J.F. Van Impe (2004). A prototype model for indole-3-acetic acid (IAA) production by *Azospirillum brasilense* sp 245. In: *Proceedings of the 7th International Conference on DYnamics and COntrol of Process Systems, Dycops 7, CDROM* (S.L. Shah and J.F. MacGregor, Eds.). 6p.

Van Impe, J.F., J.E. Claes and G. Bastin (1995). Optimal feed rate profile for combined bioprocess modeling and optimization. In: *Proceedings of the 1995 European Control Conference* (A. Isidori, S. Bittani, E. Mosca, A. De Luca, M.D. Di Benedetto and G. Oriolo, Eds.). pp. 3510–3515.

Vanrolleghem, P.A. and D. Dochain (1998). Bioprocess model identification. In: *Advanced instrumentation, data interpretation and control of biotechnological processes* (J.F.M. Van Impe, P.A. Vanrolleghem and D.M. Iserentant, Eds.). pp. 251–318. Kluwer Academic Publisher, Dordrecht.

Vanstockem, M., K. Michiels, J. Vanderleyden and A. Van Gool (1987). Transposon mutagenesis of *Azospirillum brasilense* and *Azospirillum lipoferum*: physical analysis of Tn5 and Tn5-mob insertion mutants. *Applied and Environnemetal Microbiology* **53**, 410–415.

Versyck, K.J. and J.F. Van Impe (1999). Feed rate optimization for fed-batch bioreactors: from optimal process performance to optimal parameter estimation. *Chemical Engineering Communications* **172**, 107–124.

Walter, E. and L. Pronzato (1997). *Identification of parametric models from experimental data*. Springer, Masson.