10th International Symposium on Process Systems Engineering - PSE2009
Rita Maria de Brito Alves, Claudio Augusto Oller do Nascimento and Evaristo Chalbaud Biscaia Jr. (Editors)

Sequential design of dynamic experiments in modeling for optimization of biological processes

Mariano Cristaldi[a], Ricardo Grau[a], Ernesto Martínez[b]

[a]*INTEC (CONICET-UNL), Guemes 3450, Santa Fe 3000, Argentina*
[b]*INGAR (CONICET-UTN), Avellaneda 3657, Santa Fe S3002 GJC, Argentina*

Abstract

Finding optimal operating conditions fast with a scarce budget of experimental runs is a key problem to speed up the development and scaling up of innovative bioprocesses. A methodology for model-based design of dynamic experiments in modeling for optimization is proposed and successfully applied to the optimization of a fed-batch bioreactor related to the production of r-interleukin-11 whose DNA has been cloned in an *E. coli* strain. A library of tendency models is used to increasingly bias bioreactor operating conditions towards an optimum. Parametric uncertainty of tendency models is iteratively reduced using Global Sensitivity Analysis (GSA). At each iteration, the 'most informative' tendency model is used for designining the next dynamic experiment. Model selection is based on minimizing an error measure which separates parametric uncertainty from structural errors to trade-off exploration with exploitation.

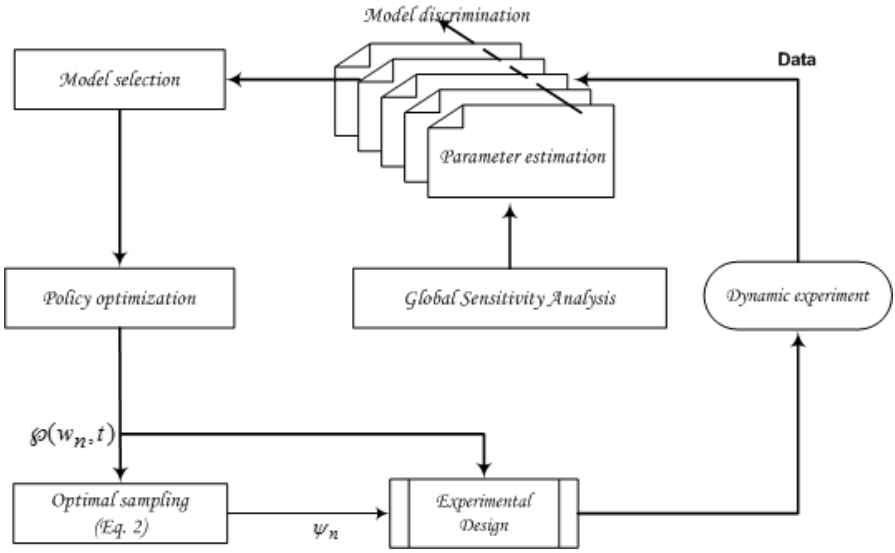
Keywords: Modeling, optimization, experimental design, dynamic experiments.

1. Introduction

Findings in the 1950s that DNA is the molecule that encodes proteins, which in turn controls all the cellular processes including metabolic pathways, have provided the impetus for the biotechnology era (Walsh, 2007) in the development of high valued-added pharmaceuticals such as insulin, humanized antibodies, interferons or interleukins along with biofuels (Fischer *et al*., 2008) and new materials (e.g. biodegradable polymers). The recombinant microorganism is typically grown in a fed-batch bioreactor to high cell concentration and then expression of the heterologous protein is triggered so as to obtain considerable quantities of the target product (Cooney, 1983). In a fed-batch culture, the feed rate of the carbon source, usually glucose, must be manipulated in order to restrict overflow metabolism and glucose repression. To this aim, model-based optimization of a bioreactor operating condition seems to be the safe and economic approach to resort with. Application of Pontryagin's maximum principle for fed-batch bioreactor optimization has been studied by several researchers (Mahadevan and Doyle III, 2003). However, most of these optimization methodologies for bioprocess scaling up and productivity inprovement have not been widely adopted for industrial use since the perfect model assumption is far from realistic and bioreactor behavior is quite often deviant from model predictions. Also, relevant measurement are sparse and delayed. Considering the large uncertainty and poor reproducibility in novel bioprocesses along with metabolic regulation, the development of an accurate mathematical model of bioreactor dynamics is a costly and very difficult undertaking. A better approach in innovative bioprocesses is to improve the operating policy by resorting to tendency models (Visser *et al.*, 2000) for designing optimally informative experiments which iteratively reduce the performance prediction uncertainty (Martínez and Wilson, 2003).

2. Modeling for optimization

At each iteration, the core idea of "modeling for optimization" is to select from a library of tendency models the one which allows computing inputs that increasingly improve the operating policy and bias data gathering accordingly. To this aim, a dynamic experiment is designed around the current policy, and optimal sampling times are calculated so as to maximize information content regarding performance improvement. The experiment is carried out and new data are collected. Based on incoming data the sub-set of model parameters for each tendency model are re-estimated which selectively reduces its parametric uncertainty. Based on total modeling errors a tendency model is selected for policy re-optimization. With the new input policy, a new iteration begins. The identification-optimization cycle is continued until no performance improvement is obtained and the input policy converges. Fig. 1 provides an overall picture of the proposed methodology.



Fig, 1. Modeling for optimization using designed dynamic experiments

2.1. Problem statement

In what follows let's assume that the dynamic behavior of the bioreactor under study may be modeled alternatively by a library of ℓ tendency models

$$\frac{dx_i}{dt} = f(x_i(t), \wp(w,t), \theta, t) \;\; 0 \le t \le t_f, \quad x_i(0): \text{given}, i = 1,2,...,\ell \qquad (1)$$

and the optimization objective to be maximized iteratively via experimental runs is

$$J(w) = h(x(t_f)) + \int_0^{t_f} g(x(t), \wp(w,t), \theta, t)dt \qquad (2)$$

where $x(t)$ is an n_s-dimensional vector of time dependent state variables, w is an m-dimensional vector of parameters for the input policy $\wp$, θ is a p-dimensional vector of model parameters and t_f is the final time of a batch run. The function g is the instantaneous reward function along the state trajectory $x(t)$ defined by a given policy parameterization whereas the function h is the specific reward for the final state of the batch run when using the input policy $\wp(w,t)$.

2.2. Model selection

Model selection is based on distinguishing between parametric uncertainty and structural errors in performance prediction using tendency models (see Fig. 2 for details). For a given tendency model and a plausible realization of its parameters, the corresponding simulated trajectory of the process performance index $\tilde{J}_i, i = 1,2,..., n_{sp}$. At the *ith* sampling point, a sample average $\bar{J}_i$ of n different model parameterizations can be used to characterize the *parametric uncertainty* for the tendency model as follows (Asprey and Machietto, 2002; Chen and Asprey, 2003):

$$\varepsilon_{J_i \to \bar{J}_i} == \frac{1}{n_{sp}\, n} tr\left[\left(\tilde{J}_{ij} - \bar{J}_i\right) W_{ij} \left(\tilde{J}_{ij} - \bar{J}_i\right)^T \right]; i = 1,2,..., n_{sp}; j = 1,2,..., n \qquad (3)$$

As a measure of *structural errors* inherent to a given tendency model, the average performance trajectory $\bar{J}_i$ is compared to the actual (observed) trajectory J_i

$$\varepsilon_{\bar{J} \to J} = \frac{1}{n_{sp}} tr\left[\left(\bar{J}_i - J_i\right)^T W_{ii} \left(\bar{J}_i - J_i\right) \right]; i = 1,2,..., n_{sp}; W_{ii} : \text{ weighting matrix} \qquad (4)$$

where n_{sp} is the number of sampling points. The total error of a model is expressed as the sum of parametric uncertainty in Eq. (3) plus its structural error defined as it is shown in Eq. (4). For policy optimization, model selection in each iteration is based on choosing the tendency model from the model library whose total error is the lowest. More elaborated strategies for model selection can also be developed. For example, initially model selection may emphasize reducing parametric uncertainty and as more data are gathered model selection is more based on structural errors.

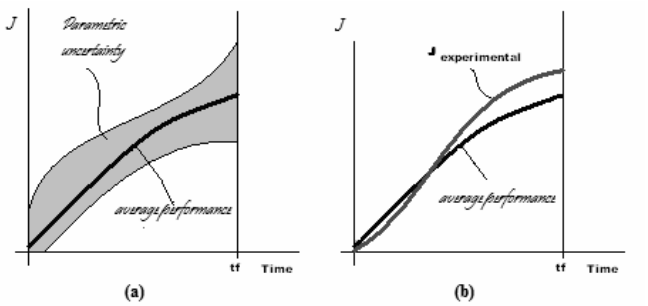


Fig. 2. Model selection based on total error: (a) Parametric uncertainty; (b) structural errors.

2.3. Optimal sampling

For a given policy in the current policy iteration of Fig. 1 optimal sampling times ψ^{opt} along a batch run must be calculated so as to bring new information to selectively reduce parametric uncertainty which affect the most the value estimation of the performance index J trajectory. Assuming model parameters are set to $\hat{\theta}$ and the

current version of the optimal policy is $\wp(w,t)$, the issue of optimal sampling is related to calculating at which sampling times $\psi^{opt} \in \Psi$ in a dynamic experiment the values of measured process variables are most informative in modeling for optimization assuming that the policy evaluation step should narrow down the uncertainty about the optimal input. To this end, the following optimization problem is solved:

$$\psi^{opt} = \max_{\psi \in \Psi} \det\left|M(\hat{\theta}, \wp(w_n,t), \psi)\right|, \; M = Q^T Q, \; Q = \begin{pmatrix} Si_{11} & \cdots & Si_{1n} \\ \vdots & \ddots & \vdots \\ Si_{m1} & \cdots & Si_{mn} \end{pmatrix} \tag{5}$$

where each entry of the matrix Q, Si_{ij}, measures the sensitivity of the performance index $J(w)$ at the *i-th* sampling time with respect to *j-th* parameter of the operating policy. To calculate Si_{ij}, Global Sensitivity Analysis (GSA) is proposed (Saltelli et al., 2006). GSA is a variance-based technique that decomposes model outputs variability as a combination of uncertainty intervals for each independent input factor and its interactions with other factors using Monte Carlo sampling techniques.

3. Case study

To illustrate the proposed methodology results obtained in the optimization of fed-batch fermentation process for the recombinant protein **rIL-11** using a genetically modified *E. coli* strain are presented. Production of recombinant proteins in *E. coli* has been widely applied in both laboratory research and bioproduct manufacturing since this microorganism is considered a reliable source of proteins. This method may achieve profitable mass productivity due to high density cell growth and fast product formation. A structured kinetic model proposed by Tang *et al.* (2007) which describes state variables trajectories such as: biomass (X), substrate (S), intracellular recombinant protein concentration (P) will be used as an *in silico* bioreactor to generate the required data in the modeling for optimization approach. Four unstructured (tendency) models which differ in their biomass growth kinetics are used as guidelines for policy optimization so that the mismatch between the "real" bioprocess and alternative models of the fed-batch bioreactor is accounted for by increasingly biasing data gathering. Also, the operation policy has been defined based on the substrate feeding rate and induction time t_{ind} as the main components subject to optimization, including the initial culture condition. The performance index $J(t)$ is related to the amount of recombinant protein obtained at the final time of production runs. Tendency model equations and their alternative biomass growth kinetics are:

$$\frac{dX}{dt} = \mu X; \qquad \frac{dS}{dt} = -\frac{\mu}{Y_{xs}} X - f(X,t); \; \frac{dP}{dt} = r_P - \mu P \tag{6}$$

$$r_P = \begin{cases} 0, & t < t_{ind} \\ K_P^{\max}\left(\frac{S}{K_S + S}\right)\left[\frac{1}{1+\left(\frac{P}{KI_P}\right)^5}\right], & t \geq t_{ind} \end{cases}$$

$$\begin{aligned} &First\ order: \mu = \mu_{\max} S, f(X,t) = 0 \\ &Monod: \mu = \mu_{\max} \frac{S}{K_s + S}, f(X,t) = 0 \\ &Contois: \mu = \mu_{\max} \frac{S}{K_X X + S}, f(X,t) = 0 \\ &Monod: \mu = \mu_{\max} \frac{S}{K_s + S}, f(X,t) = m.X \end{aligned}$$

Based on experimental data provided by Tang et al. (2007), a rather rough estimation of each tendency model parameters was made and referred to as "initial values" in Table 1.

Due to the significant level of parametric uncertainty a ±50% confidence interval around these initial values for each parameter is assumed in the first policy optimization iteration. A uniform distribution over its confidence interval is assumed initially for each model parameter. Experimental data have been conveniently pre-treated to reduce significantly the signal-to-noise ratio and outliers are not present.

Table 1. Initial parameterizations of tendency models based on experimental data

Parameter	Unit	Model			
		1st Order	Monod	Contois	Maintenance
μ_{max}	h^{-1}	0.2000	0.6301	0.5607	0.5261
K_s	$g\ L^{-1}$	2.0184	1.4956	-	0.7190
Y_{xs}	$g_{biomass}\ g_{substrate}^{-1}$	0.3982	0.4506	0.4826	0.4464
K_P^{max}	$g\ L^{-1}$	0.0759	0.0629	0.0557	0.0536
KI_p	$g\ (L\ h)^{-1}$	0.0877	0.0609	0.0627	0.0600
K_x	$g_{substrate}\ g_{biomass}^{-1}$	-	-	1.7291	-
m	h^{-1}	-	-	-	0.0100

At any time *t*, the input policy is defined by a vector *w* of parameters corresponding to two different degrees of freedom for process optimization. A subset of the policy parameters corresponds to inputs that can be modified only from run-to-run but are time-invariant in a given run such as the substrate feeding concentration, run duration or induction time. The remaining entries are parameters which are used here for describing the profile of time-varying controls such as the feeding rate. In the latter case, a key issue is the mathematical description to be used so as to provide ample room for different variability patterns within economic and safety constraints with a minimum number of independent parameters. Even though other profile functions (high-order polymonials, Gaussian Processes, etc.) can be used for shape flexibility with a small number of parameters, the following quadratic inverse polynomial is used hereafter:

Table 2. Sequential optimization of a E. coli culture for rIL-11

Parameter	Units	Initial Condition	1st iter	2nd iter	3rd iter	4th iter	5th iter	6th iter
Z	$L\ h^{-1}$	1	-	-	-	-	-	-
A	$L\ h^{-2}$	0	0.0544	0.0121	0.0911	0.1431	0.2385	0.2389
B	h^{-1}	-	$3\ 10^{-4}$	$3\ 10^{-4}$	$3\ 10^{-4}$	$3\ 10^{-4}$	$3\ 10^{-4}$	$3\ 10^{-4}$
C	h^{-2}	-	$3\ 10^{-4}$	$4\ 10^{-4}$	$8\ 10^{-4}$	0.0156	0.0222	0.0223
S_f	$g\ L^{-1}$	10	30	30	30	30	30	30
t_{feed}	h	6	0	0	3.19	4.05	5	5
t_{ind}	h	4	4	4	4	4	4	4
t_f	h	12	16	16	16	16	16	16
V_0	L	6	5.31	10	5	6.61	5	5
X_0	$g\ L^{-1}$	0.05	0.1	0.1	0.1	0.1	0.1	0.1
S_0	$g\ L^{-1}$	6	3.93	7	7	7	7	7
$J.V_f$	g	**1.60**	**6.06**	**3.75**	**7.15**	**6.40**	**7.22**	**7.22**

$$F_{in} = \begin{Bmatrix} 0 & t < t_0 \\ \dfrac{At}{1 + Bt + Ct^2} & t \geq t_0 \end{Bmatrix} \tag{7}$$

As it is shown in Table 2, despite the rough approximation of tendency models, they provide a very valuable guideline for fast optimization with a handfull of experiments. Based on their total errors, *Monod* and *Contois* kinetics are the most infomative for re-calculating the input policy for dynamic experiments (see Table 3 and Table 4 below).

Table 3. Total error in 1st iter.

Model	$E_{\hat{J}_i \to \langle \hat{J} \rangle}$	$E_{\langle \hat{J} \rangle \to J}$	E_{total}
1st Order	0.0044	0.0209	0.0254
Monod	$7.13\ 10^{-12}$	$6.23\ 10^{-5}$	**$6.23\ 10^{-5}$**
Contois	$1\ 10^{-4}$	$5.78\ 10^{-5}$	$1.58\ 10^{-4}$
Manteinance	1.1403	$6.35\ 10^{-5}$	1.1404

Table 4.Total model errors in 4th iter

$E_{\hat{J}_i \to \langle \hat{J} \rangle}$	$E_{\langle \hat{J} \rangle \to J}$	E_{total}
$6.41\ 10^{-7}$	0.0279	0.0279
$6.92\ 10^{-6}$	0.0011	0.0011
$1.65\ 10^{-5}$	$9.85\ 10^{-4}$	**0.0010**
$2.01\ 10^{-4}$	0.0045	0.0047

4. Final remarks

A systematic procedure is proposed for sequential design of dynamic experiments in modeling for optimization using a library of tendency models for safe exploration of alternative parameterizations of the input policy to improve operating conditions. At each iteration, model selection is based on the total model error which accounts separately for parametric uncertainty and structural errors. Since tendency models are initially plagued with uncertainty model selection using poorly estimated total errors makes possible to trade off exploitation with exploration which is instrumental for model-based optimization with imperfect models. Global sensitivity analysis has been used to formulate the optimal sampling in each experiment as an optimization problem.

References

S. Asprey, S. Machietto, 2002, Designing robust optimal dynamic experiments, *J. of Process Control,* 12, 545-556.

B. Chen, S. Asprey 2003, On the design of optimally informative dynamic experiments for model discrimination in multiresponse nonlinear situations*, Ind. Eng. Chem. Res*., 42, 1379-1390.

C. Cooney, 1983, Bioreactors: design and operation, *Science* 19, 728-740.

C. Fischer, D. Klein-Marcuschamer, G., Stephanopoulos, 2008, Selection and optimization of microbial hosts for biofuels production. *Metabolic Engineering*, (**in press**).

R. Mahadevan, F. Doyle III, 2003, On-Line optimization of recombinant product in a fed-batch bioreactor, *Biotechnol. Prog*., 19, 639-646.

E. Martínez, J. Wilson, 2003, Evolutionary optimization of batch process systems using imperfect models, *Proc. Indian Natn. Sci. Acad*., 403-428.

A. Saltelli., M. Ratto, S. Tarantola, F. Campolongo, 2006, Sensitivity Analysis Practices. Strategies for Model-Based Inference, *Reliability Engineering & System Safety*, 1109-1125.

S. Tang, J. Chen, Z. Zhang, 2007, Structured models for recombinant human interleukin-11 fermentation, *Biochemical Engineering J*., 35, 210-217.

D. Visser, et al., (2000). Tendency Modeling:A New Approach to Obtain Simplified Kinetic Models of Metabolism, *Metabolic Engineering*, 252-275.

G. Walsh, 2007*, Pharmaceutical biotechnology: concepts and applications*, John Wiley & Sons Ltd, Chichester, England.