Global Sensitivity Analysis Challenges in Biological Systems Modeling

A. Kiparissides, S. S. Kucherenko, A. Mantalaris,* and E. N. Pistikopoulos

Biological Systems Engineering Laboratory, Centre for Process Systems Engineering, Department of Chemical Engineering, Imperial College London, London SW7 2AZ, U.K.

Mammalian cell culture systems produce high-value biologics, such as monoclonal antibodies, which are increasingly being used clinically. A complete framework that interlinks model-based design of experiments (DOE) and model-based control and optimization to the actual industrial bioprocess could assist experimentation, hence reducing costs. However, high fidelity models have the inherent characteristic of containing a large number of parameters, which is further complicated by limitations in the current analytical techniques, thus resulting in the experimental validation of merely a small number of parameters. Sensitivity analysis techniques can provide valuable insight into model characteristics. Traditionally, the application of sensitivity analysis on models of biological systems has been treated more or less as a black box operation. In the present work, we elucidate the aspects of sensitivity analysis and identify, with reasoning, the most suitable group of sensitivity analysis methods for application to highly nonlinear dynamic models in the context of biological systems. Specifically, we perform computational experiments on antibody-producing mammalian cell culture models of different complexities and identify, as well as address, problems associated with such "real life" models. In conclusion, a novel global screening method (derivative based global sensitivity measures, DGSM) is proven to be the most time-efficient and robust alternative to the established variance-based Monte Carlo methods.

Introduction

Mammalian cell bioprocesses produce high-value biologics, which represent the fastest growing section of pharmaceuticals, including monoclonal antibodies, vaccines and recombinant proteins.[1] However, production of such products on an industrial scale still remains a challenging task in terms of overcoming the complications associated with optimal culture growth and productivity.[2] Implementation of modern model-based process control and optimization techniques by the biologics industry lags far behind other process industries.[3] As a result, currently, control and optimization of mammalian cell culture systems is still fundamentally manual, which inevitably increases costs and time to market.

The establishment of a complete and systematic framework that interlinks model-based design of experiments (DOE) to model-based control and optimization of the actual industrial process is yet to be achieved. One of the limitations of the process is the development of high fidelity models able to capture the required biological functions involved in the generation of the end product while remaining computationally tractable in order to be viable options for model-based control and optimization. However, high fidelity models, inherently, contain a large number of parameters. Furthermore, the inability of the current analytical techniques to measure most intracellular parameters results in the experimental validation of merely a small number of parameters.[4]

Model analysis techniques and sensitivity analysis, in particular, can provide valuable insight regarding the dependence of the model output on its parameters. Parameters identified as significant can be experimentally estimated, while insignificant parameters can be fixed at their literature values (if available) or approximated. Up to now, the application of sensitivity analysis on models of biological systems has been treated, in general, as a black box operation. In the present work, we elucidate the aspects of sensitivity analysis and identify, with reasoning, global sensitivity analysis (GSA) as the most suitable group of sensitivity analysis methods for application in highly nonlinear dynamic models. Specifically, GSA methods reported in the literature were evaluated with two methods being selected as the most suitable for application in the context of biological models, i.e., highly nonlinear systems of coupled differential equations. The first method is the established Sobol' global sensitivity indices[5] method, which is a variance-based method providing high quality information; alas it is excessively computationally demanding. The second method is the novel derivative based global sensitivity measures (DGSM) method developed by Kucherenko et al.,[6] which provides the same quality of information as the Sobol' indices while being significantly less demanding computationally.

A significant part of the proposed model development framework for biological systems presented in Figure 1 (see the Appendix) is highlighted as parameter significance ranking. This is the part where model analysis and GSA in particular come into play. Herein, we aim to elucidate the technical and theoretical aspects associated with GSA and how it can be used to minimize experimental labor and costs. In order to achieve the said aim, we have set the following objectives:

(1) Examine the applicability of GSA techniques to highly nonlinear models of biological systems.
(2) Compare the computational efficiency of GSA techniques in "real life" models of biological systems.
(3) Examine the quality of the produced results and examine any possible biological insight gained as has been shown by previous studies.[7,8]

Challenges in Model Analysis of Biological Systems

Dynamic models describing complex biological functions involve highly nonlinear terms and include a large number of parameters with varying orders of magnitude. Thus, commonly used sensitivity analysis techniques are not able to provide results with any practical value for such models. Sensitivity

* To whom correspondence should be addressed. E-mail: a.mantalaris@imperial.ac.uk.

10.1021/ie900139x CCC: $40.75
Published on Web 07/07/2009

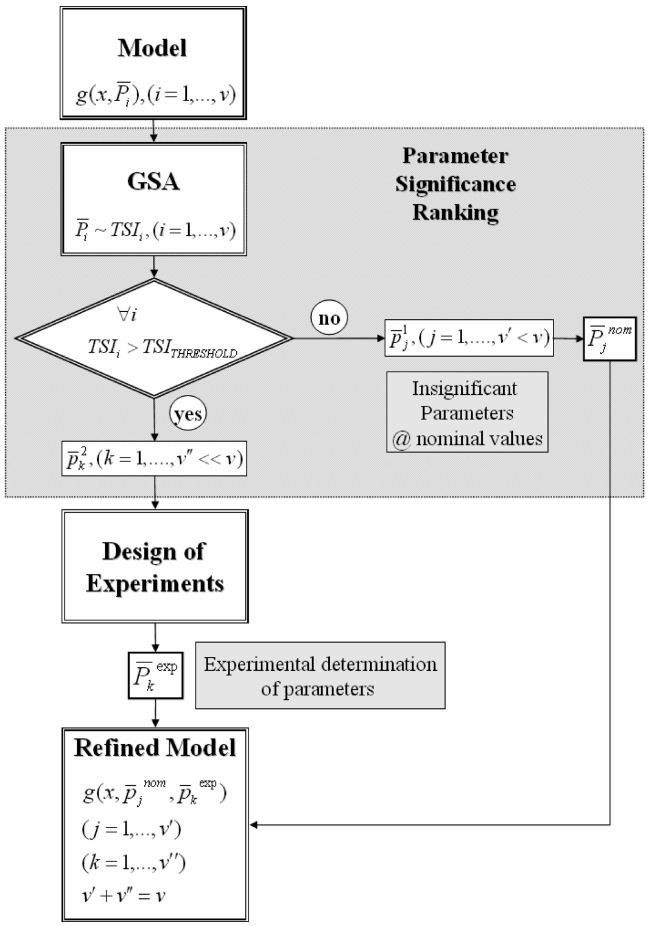

Figure 1. Model development framework for biological systems.

analysis methods are commonly grouped in three main categories, namely screening, local, and global methods.

Screening methods are randomized, one-at-a-time numerical experiments, which aim to indicate the most important factors among the totality of model parameters. While screening methods involve computationally efficient algorithms, their use is limited to only preliminary results due to calculation of only first order effects (i.e., effects the input factors have on the model output, without including their mutual interactions) and inherently lack precision, especially when used on nonlinear models.[9] Efforts to calculate higher order effects, through screening methods, have been recorded in the literature,[10,11] though these methods fall short in terms of either accuracy or computational time.

Local methods derive measures of importance by estimating the effects infinitesimal variations of each factor have on the model output, in the area of a predetermined nominal point. Local methods are commonly used on steady-state models, or on studies dealing with the stability of a nominal point. Consequently, local methods fail to capture large variations in the parameter set and can only account for small variations from the parameter nominal values.

Global methods have the unique advantage of performing a full search of the parameter space, hence providing data independent of nominal points, and are applicable to the whole range of the model's existence. Moreover, global methods apportion the total uncertainty in the model output to the various sources of variation, while all parameters are varied at the same time. GSA provides the most complete set of results and mapping of the system, being able to cope with nonlinearities and identify parameter interaction effects.[9] The main drawback of GSA methods is their extensive computational cost for large models. GSA methods are commonly grouped in two categories, namely methods that utilize a model approximation in order to generate measures of importance and methods that study the total output variance of the model. Model approximation methods, such as regression analysis, correlation ratios, and rank transformation, cannot account for higher order effects.

Variance-based methods provide measures of importance, i.e., sensitivity indices that apportion the total output variance to its contributors, namely the model parameters. In order to estimate the total output variance and its fractions, model parameters are treated as random variables within the parameter space. In the present context, randomness refers to the statistical independence of the generated samples. Since the models' parameters are treated as random variables, the resulting model output will be a random variable itself. The model output can thus be decomposed into summands of increasing dimensionality, a procedure also known as analysis of variance (ANOVA) decomposition.

$$f(X_1,...,X_n) = f_0 + \sum_{i=1}^{n} f_i(X_i) + \sum_{1\le i<j\le n} f_{ij}(X_i,X_j) + ... + f_{1,2,...,n}(X_1,...,X_n) \quad (1)$$

Under the assumption that each of the terms in eq 1 is orthogonal,[5,12] the decomposition is unique and, therefore, integration of any term over any of the variables it may contain results to zero. This unique decomposition enables variance-based methods to discriminate between the first order and higher order effects. First order information refers to the significance of merely the first summand with respect to the model output, while higher order information explores the effect of parameter interactions and their contribution to the total output variance.

Main effects can be used to generate a significance—with respect to the model output—ranking of the model parameters. While rankings based solely on main effects are quite efficient in the case of linear models, the effect of the remaining summands cannot be neglected for nonlinear models. Chan et al.[13] have illustrated the significance of higher order sensitivity indices in understanding the behavior of the model parameters and how the uncertainty associated with them propagates through the model. Significance ranking for the model parameters should be based on the calculation of the total sensitivity index (TSI).[12,14] The TSI for parameter i is estimated as the sum of all higher order terms in eq 1 which include parameter i. The vast majority of sensitivity analysis techniques do not include a decomposition similar to the one presented in eq 1; therefore, it is not possible to discriminate whether the measure of importance they estimate refers to first or higher order information. Therefore, in order to obtain a realistic insight into the model's affecting parameters, obtaining information in the form of TSI is required. For a more comprehensive description of ANOVA decomposition and the TSI, refer to the work of Sobol',[5] Saltelli,[9] and Chan et al.[13]

The most commonly used and established variance-based GSA methods currently are the Sobol' global sensitivity indices (Ilya M. Sobol' first published his work on global sensitivity indices in 1990 in Russian) and the extended Fourier amplitude sensitivity test (FAST[15]). Both of these methods derive the same results, although from totally different contexts. The key idea behind FAST is the evaluation of a multidimensional integral via a monodimensional integral, exploiting Weyl's ergodic theorem.[16] On the other hand, the Sobol' algorithm involves direct evaluation of the multidimensional integral via Monte

Carlo (MC) integration. For all intents and purposes, the results for both the main effect indices and the TSIs are considered identical between the two methods. A general consensus as to which method is superior to the other does not seem to exist in the literature. Homma and Saltelli[10] favor the Sobol' method for its clarity in the derivation of the higher order indices, while Chan et al.[13] conclude that FAST is computationally more efficient and accurate. The review of Frey and Patil[17] refers to FAST as a method that suffers from issues of computational efficiency and mentions the method's weakness when dealing with discrete inputs, whereas Saltelli[9] concluded that FAST is computationally more efficient than the Sobol' method in terms of the required number of model evaluations. However, more recent developments presented by Saltelli[18] and later by Sobol' et al.[19] provide an upgrade to Sobol's original algorithm both in terms of required number of model evaluations[18] and in terms of accuracy with the use of new improved formulas.[19] Consequently, the ease in the algorithmic implementation of the Sobol' method, as well as the transparency of the calculation of the higher order effects, offers the Sobol' method an advantage over FAST.

DGSM is a novel method that can be classified as a global screening method. Since it is based on a concept originally developed by Morris,[20] it remains a one-factor-a-time method, hence its classification as a screening method. In addition, it maintains the computational efficiency of screening methods, making it an affordable option, especially for models with a large number of parameters. However, DGSM provides information superior to that of the Morris method in terms of accuracy especially in the case of nonlinear models. Furthermore, work in progress[21] has shown that the Sobol' total index can be correlated to the G measure of DGSM via an inequality. The main scope of DGSM is to provide high quality global information through a computationally affordable algorithm.

Various attempts have been made over the years to introduce new sensitivity analysis methods, or to apply the existing methods to specific scientific areas—the most popular of which have been chemical reactions,[22−25] financial applications,[26] ecosystem biology,[27] and process systems engineering.[28−30] However, most of the new methods have been crude local sensitivity methods. Furthermore, to our knowledge, application of GSA methods to biological systems has been limited to the work of Kontoravdi et al.,[7] Feng et al.,[31] and a Monte Carlo approach for biological models.[32] Researchers in the field have to look to the reviews of Rabitz et al.[23] and Frey and Patil,[17] which although rich in information are not concerned with the complexities involved in models of biological systems. The work by Gunawan et al.[33] on discrete stochastic models is not applicable in the case of dynamic continuous models of cell cultures. Leis and Kramer[34] addressed only first order sensitivity indices for models containing ordinary differential equations (ODEs), while Fennel et al.[35] presented essentially an error analysis rather than a global sensitivity analysis approach to marine ecosystems. Borgonovo and Apostolakis[36] introduce an interesting differential importance measure (DIM) in the field of probabilistic safety assessment, which has the valuable property of being additive, though it suffers from being limited to local analysis. The review of Sidoli et al.[37] on modeling of mammalian cell cultures discusses the significance of model analysis techniques and how valuable biological insight can be gained from such techniques; however, it does not contain a lot of information on sensitivity analysis.

Global Sensitivity Analysis Methods

Having clarified the scope and use of GSA in the context of model analysis of biological systems, we provide a brief overview of the Sobol' global sensitivity indices and DGSM methods. Consequently we examine the behavior and performance of the presented GSA techniques on two "real life" models of mammalian cell cultures. A presentation of the two GSA methods, namely the Sobol' global indices and DGSM, is provided in this section. For a more detailed introduction to the two methods, refer to the original work of Sobol'[5] and Kucherenko et al.[6]

Sobol' Global Sensitivity Indices. The basis of the Sobol' global indices is an ANOVA-like decomposition of the total variability of the model into summands of increasing dimensionality. This decomposition takes place within the boundaries of the n-dimensional unit cube, where n is the number of input factors. Let us define the output variable as $Y = f(\mathbf{X})$, where $\mathbf{X}$ is the vector of input factors. It is possible to decompose $f(\mathbf{X})$ into summands of increasing dimensionality:

$$f(X_1, ..., X_n) = f_0 + \sum_{i=1}^{n} f_i(X_i) + \sum_{1 \le i < j \le n} f_{ij}(X_i, X_j) + ... + f_{1,2,...,n}(X_1, ..., X_n) \quad (2)$$

The decomposition presented in eq 2 is termed ANOVA if f_0 is constant and the integral of every summand over any of the variables it contains is equal to zero.

$$\int f(x)\, dx = f_0 \quad (3)$$

and

$$\int_0^1 f_{i1,...,in}(X_{i1_1}, ..., X_{in})\, dX_{ik_k} = 0, \quad 1 \le k \le s \quad (4)$$

It can be easily proven that eqs 3 and 4 uniquely define all the terms in eq 2. In order to define the one-dimensional terms, eq 2 is integrated over all variables except x_i.

$$\int f(x) \prod_{k \ne i} dx_k = f_0 + f_i(x_i) \quad (5)$$

This can be extended to include all higher order terms as well. Assuming $f(x)$ to be square integrable implies that all the $f_{i1,...,is}$ are square integrable as well. Squaring and integrating eq 2 we obtain

$$\int f^2(x)\, dx - f_0^2 = \sum_{s=1}^{n} \sum_{i1<...<is}^{n} \int f_{i1,...,is}^2\, dx_{i1} ... dx_{is} \quad (6)$$

If x were a random point uniformly distributed within the unit hypercube I^n, then $f(x)$ and $f_{i1,...,is}(x_{i1},..., x_{is})$ would become random variables with variances D and $D_{i1,...,is}$, respectively. Hence, the following constants from eq 6 are named variances:

$$D = \int f^2\, dx - f_0^2 \qquad D_{i1,...,is} = \int f_{i1,...,is}^2\, dx_{i1} ... dx_{is} \quad (7)$$

A measure of the main effect of each factor on the model output can be now defined, namely the Sobol' global sensitivity indices, as

$$S_{i1,\ldots,is} = \frac{D_{i1,\ldots,is}}{D} \tag{8}$$

It is obvious that

$$\sum_{s=1}^{n} \sum_{i1<\ldots<is}^{n} S_{i1,\ldots,is} = 1 \tag{9}$$

Apart from sensitivity indices for individual input factors, the Sobol method allows indices to be evaluated for subsets of factors. Therefore, after an initial screening of the model parameters, those indicated as less important can be grouped into subsets, thus gaining in computational time. Input factors can also be grouped and studied according to their physical or biological nature.[31]

Let us now consider a subset of m input factors, where $1 \leq m \leq n - 1$, namely $y = (x_{k1},\ldots, x_{km})$. Let z denote the set of $n - m$ remaining input factors. The variance corresponding to the subset y can be defined as

$$D_y = \sum_{s=1}^{m} \sum_{(i1<\ldots<is)\in K}^{n} D_{i1,\ldots,is}; \quad K = (k_1, \ldots, k_m) \tag{10}$$

Equation 10 can be rewritten for the variance of the second subgroup, D_z. The total variance corresponding to the group of input factors y is

$$D_y^{\text{tot}} = D - D_z \tag{11}$$

It is now possible to define the two global sensitivity indices for the subset of input factors y. The S_y term represents the main effect of the subset on the output variance, whereas the term S_y^{tot} corresponds to the TSI. The latter includes both the individual effect of the subset y and the effects originating from its interactions with other subsets or individual factors.

$$S_y = \frac{D_y}{D} \quad \text{and} \quad S_y^{\text{tot}} = \frac{D_y^{\text{tot}}}{D} \tag{12}$$

From eqs 11 and 12 one can easily derive that $S_y^{\text{tot}} = 1 - S_z$ and that $0 \leq S_y \leq S_y^{\text{tot}} \leq 1$. Two extreme cases exist for the values of S_y and S_y^{tot}. Either they are both 0, in which case they do not have any effect on the model output, or they are both equal to 1, in which case the model output variance is a direct result of the variance of this specific subset of input factors.

Derivative Based Global Sensitivity Measures (DGSM). Consider a differentiable function $f(x)$, where $x = \{x_i\}$ is a vector of input variables defined within the unit hypercube I^n. Local sensitivity measures, which are based on partial derivatives, are of the form

$$E_i(x^*) = \frac{\partial f}{\partial x_i} = \frac{f(x_i + \Delta) - f(x_i)}{\Delta} \tag{13}$$

Sensitivity measure $E_i(x^*)$ depends on a nominal point x^*, and its value varies according to the value of $\int x^*$. This deficiency can be overcome by averaging $E_i(x^*)$ over the parameter space I^n. Such a measure can be defined as

$$\bar{M}_i = \int_{I^n} E_i \, \mathrm{d}x \tag{14}$$

Non-monotonic functions consist of regions with both positive and negative values of partial derivatives $E_i(x^*)$. Consequently, due to the effect of averaging, values of $\bar{M}_i$ can be very small or even zero. To avoid such situations, measures based on the absolute value of $|E_i(x^*)|$ can be used: Another informative measure to consider is the variance of $\bar{M}_i$, which can be estimated by

$$\bar{\Sigma}_i^2 = \int_{I^n} \bar{E}_i^2 \, \mathrm{d}x - \bar{M}_i^2 \tag{15}$$

By combining the presented measures $\bar{M}_i$ and $\bar{\Sigma}_i$, a new measure can be derived.

$$\bar{G}_i = \bar{\Sigma}_i^2 + \bar{M}_i^2 = \int_{I^n} \bar{E}_i^2 \, \mathrm{d}x \tag{16}$$

where $1 \leq s \leq n$. The $\bar{G}$ measure can account for the fractional significance of a particular parameter with respect to the total variance, or for a group of parameters with respect to the total variance.

From eq 16 we can obtain the ratio

$$\frac{\bar{M}_i}{\bar{\Sigma}_i} = \left(\frac{1}{\bar{\Sigma}_i^2} \int_{I^n} \bar{E}_i^2 \, \mathrm{d}x - 1 \right)^{1/2} \tag{17}$$

Calculation of DGSM is based on the evaluation of integrals from eqs 14−17, which can be presented in the following generic form:

$$I[f] = \int_{I^n} f(x) \, \mathrm{d}x \tag{18}$$

A quasi-Monte Carlo algorithm can be used for the evaluation of integrals of the general form (18):

$$I_N = \frac{1}{N} \sum_{i=1}^{N} f(q_i) \tag{19}$$

Here, $\{q_i\}$ is a set of quasi-random points, also known as low discrepancy points, uniformly distributed within the unit hypercube. The total number of function evaluations required for the calculation of a full set of $\{\bar{M}_i\}$ and $\{\bar{\Sigma}_i\}$ is $N_F = N(k + 1)$. Throughout this work we have used Sobol' (LP-τ) quasi-random sequences in order to generate our sets of quasi-random points.

Biological Models

The performance of the Sobol' global indices and DGSM methods was evaluated on two biological models of differing size and complexity. Sensitivity analysis was conducted at different phases of the cell culture, and significance rankings as well as the required computational times were obtained. The time points were selected based on the growth phases of a batch cell culture. Specifically, time points from the lag phase, the exponential growth phase, and the decline phase were evaluated. Each one of these phases exhibits different complexities with respect to the mathematics involved. During the growth phase, steep gradients are observed, whereas during the decline phase, near-zero values are obtained due to the depletion of various metabolites, characteristics that tend to complicate the function of numerical solvers. All simulations were performed on an Intel Core 2 Duo (E4600 - 2.4, 2.39) personal computer with 3.24 GB of RAM memory, while all mathematical codes and GSA algorithms were implemented using the Mathworks MATLAB mathematical suite.

Unstructured Growth Model of Hybridoma Cell Cultures (16 Parameters). The first model is an unstructured model describing cell proliferation in batch and fed batch cultures based on the consumption of two basic nutrients (glucose and glutamine) and the production of the two corre-

sponding byproducts of the cell's metabolism (lactate and ammonia). The model has been adapted from Kontoravdi[38] based on the work of Jang and Barford[39] and consists of a total of six ordinary differential equations and 16 parameters in total. For the purposes of our study, the model has been operated in its batch mode.

A material balance for viable cells within the bioreactor is given by the following equation:

$$\frac{\mathrm{d}VX_\mathrm{u}}{\mathrm{d}t} = \mu VX_\mathrm{u} - \mu_\mathrm{d} VX_\mathrm{u} \tag{20}$$

where X_u is the concentration of viable cells in the bioreactor measured in cells per liter and μ and μ_d are the specific growth and death rates, respectively, measured in h^{-1}. Detailed formulas for the estimation of the specific growth and death rates will be presented at a later stage. The material balance for the total cell concentration (the sum of both dead and viable cells within the bioreactor) is

$$\frac{\mathrm{d}VX_\mathrm{t}}{\mathrm{d}t} = \mu VX_\mathrm{u} \tag{21}$$

where X_t denotes the total cell concentration and is measured in cells per liter.

The specific growth rate that appears in eqs 20 and 21 is estimated through the following formula:

$$\mu = \mu_\mathrm{max} f_\mathrm{lim} f_\mathrm{inh} \tag{22}$$

where μ_max is the maximum possible growth rate for the specific cell line (h^{-1}) and the terms f_lim and f_inh represent, respectively, the nutrient limitation and product inhibition. These can be defined through the following equations:

$$f_\mathrm{lim} = \left(\frac{[\mathrm{GLC}]}{K_\mathrm{glc} + [\mathrm{GLC}]}\right)\left(\frac{[\mathrm{GLN}]}{K_\mathrm{gln} + [\mathrm{GLN}]}\right),$$
$$f_\mathrm{inh} = \left(\frac{\mathrm{KI}_\mathrm{amm}}{\mathrm{KI}_\mathrm{amm} + [\mathrm{AMM}]}\right)\left(\frac{\mathrm{KI}_\mathrm{lac}}{\mathrm{KI}_\mathrm{lac} + [\mathrm{LAC}]}\right) \tag{23}$$

where the K_i parameters are the Monod constants for the primary nutrients, namely glucose and glutamine. Similarly, the KI_i parameters are the inhibition constants of the primary products of metabolism, namely lactate and ammonia. [GLC], [GLN], [LAC], and [AMM] represent the extracellular concentrations of the aforementioned nutrients and products and are measured in mM.

The term μ_d represents the specific death rate of the cells within the bioreactor and can be defined in a way similar to the specific growth rate.

$$\mu_\mathrm{d} = \frac{\mu_\mathrm{d,max}}{1 + \left(\frac{K_\mathrm{d,amm}}{[\mathrm{AMM}]}\right)^n}, \quad \text{with } n > 1 \tag{24}$$

where $\mu_\mathrm{d,max}$ represents the maximum specific death rate (h^{-1}) and $K_\mathrm{d,amm}$ describes the rate of cell death by ammonia.

The presented differential equations along with the accompanying algebraic equations describe the growth and proliferation of the mammalian cell culture within the bioreactor. Since the model is unsegregated, it only represents the overall concentrations of nutrients and byproduct of cellular metabolism within the bioreactor. Therefore, by performing material balances on each biological compound, four ordinary differential equations yielding the temporal concentration of nutrients/metabolites are obtained. Specifically, the material balance for the concentration of glucose can be formulated as shown:

$$\frac{\mathrm{d}(V[\mathrm{GLC}])}{\mathrm{d}t} = -Q_\mathrm{glc} VX_\mathrm{u} \tag{25}$$

where Q_glc is the specific glucose consumption rate (mmol/cell/h) and is defined as

$$Q_\mathrm{glc} = \frac{\mu}{Y_\mathrm{x,glc}} + m_\mathrm{glc} \tag{26}$$

The parameters $Y_\mathrm{x,glc}$ and m_glc which appear in eq 26 are the cell yield on glucose (cells/mmol) and maintenance energy of glucose (mmol/cell/h), respectively. Equation 25 was originally presented[39] with an additional term for glucose consumption by glucokinase, which as Kontoravdi[40] later argued, based on evidence by Tatiraju et al.,[41] has negligible effects. The material balance for glutamine is similar as described by the following equation:

$$\frac{\mathrm{d}(V[\mathrm{GLN}])}{\mathrm{d}t} = -Q_\mathrm{gln} VX_\mathrm{u} - K_\mathrm{d,gln} V[\mathrm{GLN}] \tag{27}$$

The only obvious difference is the term containing glutamine degradation. Glutamine is known to be spontaneously converted into pyrrolidonecarboxylic acid at high temperatures and when in weakly acidic or alkaline solutions.[42] Bray et al.[43] showed that even in medium temperatures, around 37 °C, glutamine degrades in the presence of weakly acidic or alkaline solutions. The degradation is more pronounced when the solution contains phosphate buffer. The specific consumption rate for glutamine is calculated, as was the case with glucose, through a formulation containing the cell yield on glutamine, $Y_\mathrm{x,gln}$, and the maintenance energy of glutamine, m_gln.

$$Q_\mathrm{gln} = \frac{\mu}{Y_\mathrm{x,gln}} + m_\mathrm{gln} \tag{28}$$

where

$$m_\mathrm{gln} = \frac{a_1[\mathrm{GLN}]}{a_2 + [\mathrm{GLN}]} \tag{29}$$

a_1 and a_2 are the relevant kinetic constants. Equation 28 is presented in the updated version[41] and not as originally presented.[40]

Similarly, mass balances can be formulated to describe the temporal evolution of the concentrations of the primary byproduct of cell metabolism. More specifically, for ammonia:

$$\frac{\mathrm{d}(V[\mathrm{AMM}])}{\mathrm{d}t} = Q_\mathrm{amm} VX_\mathrm{u} + K_\mathrm{d,gln} V[\mathrm{GLN}] \tag{30}$$

with

$$Q_\mathrm{amm} = Y_\mathrm{amm,gln} Q_\mathrm{gln} \tag{31}$$

Similarly, for lactate:

$$\frac{\mathrm{d}(V[\mathrm{LAC}])}{\mathrm{d}t} = Q_\mathrm{lac} VX_\mathrm{u} \tag{32}$$

with

$$Q_\mathrm{lac} = Y_\mathrm{lac,glc} Q_\mathrm{glc} \tag{33}$$

Q_lac and Q_amm represent the specific production rates (mmol/cell/h), while $Y_\mathrm{lac,glc}$ and $Y_\mathrm{amm,gln}$ represent the yield of a particular product on its primary nutrient (mmol/mmol). Table 1 sum-

Table 1. Nominal Values of Model Parameters[38]

parameter symbol	units	nominal value
μ_{max}	h^{-1}	0.05801
K_{glc}	mM	0.75
K_{gln}	mM	0.075
KI_{lac}	mM	171.756
KI_{amm}	mM	28.484
μ_d	h^{-1}	0.003
$K_{d,gln}$	h^{-1}	9.6×10^{-3}
$K_{d,amm}$	mM	1.759
$Y_{lac,glc}$	dimensionless	1.399
n	real integer >1	2
m_{glc}	mmol/cell/h	4.853×10^{-14}
$Y_{x,gln}$	cells/mmol	5.565×10^{8}
$Y_{x,glc}$	cells/mmol	1.061×10^{8}
$Y_{amm,gln}$	dimensionless	0.4269
a_1	mM·L/cell/h	3.4×10^{-13}
a_2	mM	4
$[GLC]_{in}$	mM	0
$[GLN]_{in}$	mM	0

marizes all parameters contained in the model and provides details of their nominal values as they have been recorded in the work of Kontoravdi et al.[38] Table 1 shows that the nominal values of the parameters vary orders of magnitude with a range as small as 10^{-14} and as large as 10^{8}. This is an inherent characteristic of most biological models, as biological functions occur at varying scales and levels of complexity while always maintaining an overall "macroscopic" effect. Since the simple model (16 parameters) was utilized, varying time scales, multiple interactions, feedback loops, and numerous other complex biological activities partaking in the cell cycle were not considered. Nonetheless, the computational challenge of performing sensitivity remains high. The presence of varying orders of magnitude necessitates the use of a global method, while the high dimensionality narrows the choice of available methods to a few.

Previously,[44] we had explored the computational efficiency of DGSM as well as identifying a significant weakness, namely its dependence on the parameter Δ. Employing the automatic differentiation algorithm[45] addresses the issue. The algorithm is an extension to the standard Matlab ODE15 solver, which while performing the numerical integration calculates at each time step the derivatives of the solution with respect to the parameter values. Herein, the numerical simulations were repeated with the inclusion of the automatic differentiation algorithm, resulting in a significant improvement in the performance of DGSM. The model was evaluated for 150 h of cell culture and sensitivity analysis was performed at three key time points, as discussed earlier. The chosen time points were 20, 50, and 120 h. The uncertainty range associated with each parameter was set to $\pm 100\%$ of its nominal value. Since the main objective of the present work is to compare the results and computational efficiency of the above presented GSA methods, we have worked under the assumption that there is no a priori knowledge available regarding the behavior of the model parameters and therefore all parameters were studied. The simulations involved scanning of all model parameters with respect to each possible model output variable.

The results for the viable cell and glutamine concentrations displayed in Figure 2 indicate a good agreement between the two methods. Certain discrepancies still occur between the two methods; however, the top five parameters in terms of significance are successfully identified by both methods. It is worthwhile mentioning that DGSM is known for stressing as more important parameters involved in nonlinear terms, whereas the Sobol' total effect index seems to favor parameter interactions.[44] Therefore, the observed discrepancies are both expected and justified. This is in agreement with the fact that the two methods measure different quantities in order to provide the sensitivity indices. While the Sobol' indices are based on the fractional variance held by a parameter or group of parameters, DGSM is a global screening technique, which averages derivative information via a quasi-random sampling technique, hence deriving global information.

Table 2 summarizes the computational time that each method required in order to generate the presented results. DGSM provides a significant upgrade in computational time when compared to the Sobol' indices while maintaining the quality of information. Therefore we present an upgrade to the already established framework for the modeling of mammalian systems,[4,8,38] albeit this time by the reduction of the computational time associated with GSA. Moreover, the significantly reduced computational burden of DGSM allows for its utilization on standard desktop computers without any special memory requirements, thus extending its applicability.

Unstructured Growth Model of Hybridoma Cell Cultures (16 Parameters), a View in the Complexities Involved with Studying the Sensitivities of a Group of Model Parameters. The dimensionality of the sensitivity analysis problem is defined by the number of model parameters; therefore, a feasibility constraint regarding the maximum possible number of individually scanned parameters is imposed implicitly in terms of computational time. This constraint is unavoidable due to the—increasing with dimension—number of model evaluations required for the Monte Carlo integrals to converge. Researchers in the field of GSA often resolve to parameter grouping in order to reduce the dimensionality of the problem, thus solving a more tractable version of the original problem.[40]

Grouping in the Sobol' method is considered both straightforward and well-defined from the original work of Sobol'.[5] However, considering the fact that DGSM is a derivative based method, grouping parameters is not as straightforward. DGSM involves the estimation of local derivatives with respect to each parameter; hence, the combined effect of two or more parameters would be a resulting "directional" derivative. Figure 3 is a schematic representation of the notion of directional derivatives for the case of two parameters. Instead of incrementing only one parameter value, all parameters contained in the group are incremented and the direction of the model output in response to this combined incrementation is studied.

This idea was adapted from the work of Campolongo et al.,[46] where an extension to the Morris[20] screening design capable of handling groups of parameters was introduced. Since DGSM is based on the Morris method and involves the estimation of the same elementary effect, the encouraging results presented by Campolongo et al.[46] should be applicable to DGSM with proper adaptation.

As far as the algorithms of the two methods are concerned, there are no limitations or guidelines defining the formation of parameter groups. Moreover there is, to the extent of our knowledge, no available information in the literature concerning the significance of the structure of parameter groups. The grouping can of course be random or arbitrary, although one can extract valuable information from an analysis if the grouping is appropriately structured. Kontoravdi et al.[38] proposed to group parameters according to their biological function and thus obtain information as to which part of the model, corresponding to a specific cellular function, is more significant. Another common practice[9,40] is to perform a fast screening of the model parameters and group together parameters identified as insig-

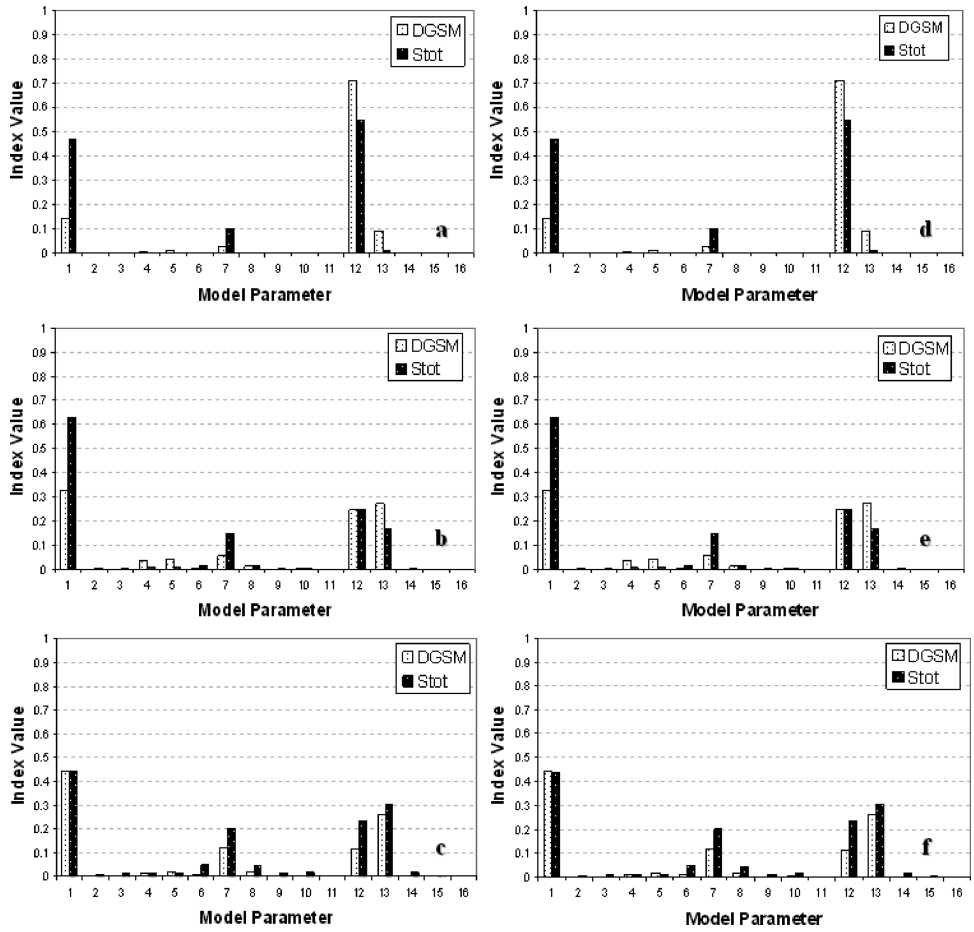

Figure 2. Calculated sensitivity indices for various time points and two different output variables. (a−c) Sensitivity indices for viable cell concentration at time points 20, 50, and 120 h, respectively. (d−f) Sensitivity indices for glutamine concentration at time points 20, 50, and 120 h, respectively. Model parameters: 1, μ_{max}; 2, k_{glc}; 3, k_{gln}; 4, KI_{lac}; 5, KI_{amm}; 6, $\mu_{d,max}$; 7, $k_{d,gln}$; 8, $k_{d,amm}$; 9, $Y_{lac,glc}$; 10, n; 11, m_{glc}; 12, $Y_{x,gln}$; 13, $Y_{x,glc}$; 14, $Y_{amm,gln}$; 15, a_1; 16, a_2.

Table 2. Computational Time for the Unstructured Model of Hybridoma Growth

	computational time (s)	
time points	Sobol' global indices	DGSM
growth model		
20 h	17 878	2 685
50 h	43 000	6 046
120 h	268 505	17 432
total	329 383 (3.8 days)	26 164 (0.30 days)

nificant, thereby reducing the dimensionality of the model. Herein we consider another interesting possibility, namely to group parameters according to their structural role in the mathematical equations describing the system, what we term as "functional" grouping. Imagine, for example, a metabolic pathway described by a set of ODEs, utilizing first order Michaelis−Menten kinetics to describe enzyme kinetics. According to the proposed "functional" grouping one would form two groups of parameters, one containing all the affinity or dissociation constants (K_i) and the other containing all the theoretical maximum reaction velocities ($V_{max,i}$).

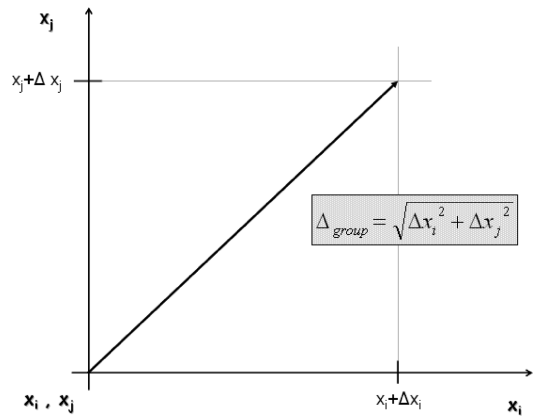

Figure 3. Measuring the combined elementary effect of a group of parameters.

From a theoretical standpoint, no method of grouping holds an advantage over the other. However, from a practical point of view there are several aspects to take into consideration. Obviously, the dimensionality of the problem is directly linked to the number of parameter groups formed. Therefore, the fewer groups we form, the faster our problem will converge. However, the smaller the number of groups, the lower the resolution of information we receive will be. It should be made clear that the size (i.e., number of parameters per group) of the parameter groups does not affect in any way the computational efficiency. Moreover, it is not necessary for the groups to contain an even number of parameters. Another important aspect is the scope of the analysis being carried out. If GSA is part of a model development framework leading to a DOE step, like the one presented in Figure 1, the biological grouping would be favorable over other choices. Conversely, if the analysis is part of a model reduction process, the "functional" grouping would be the most appropriate choice.

Before examining a more complex biological model, we will demonstrate the use of parameter grouping in the model

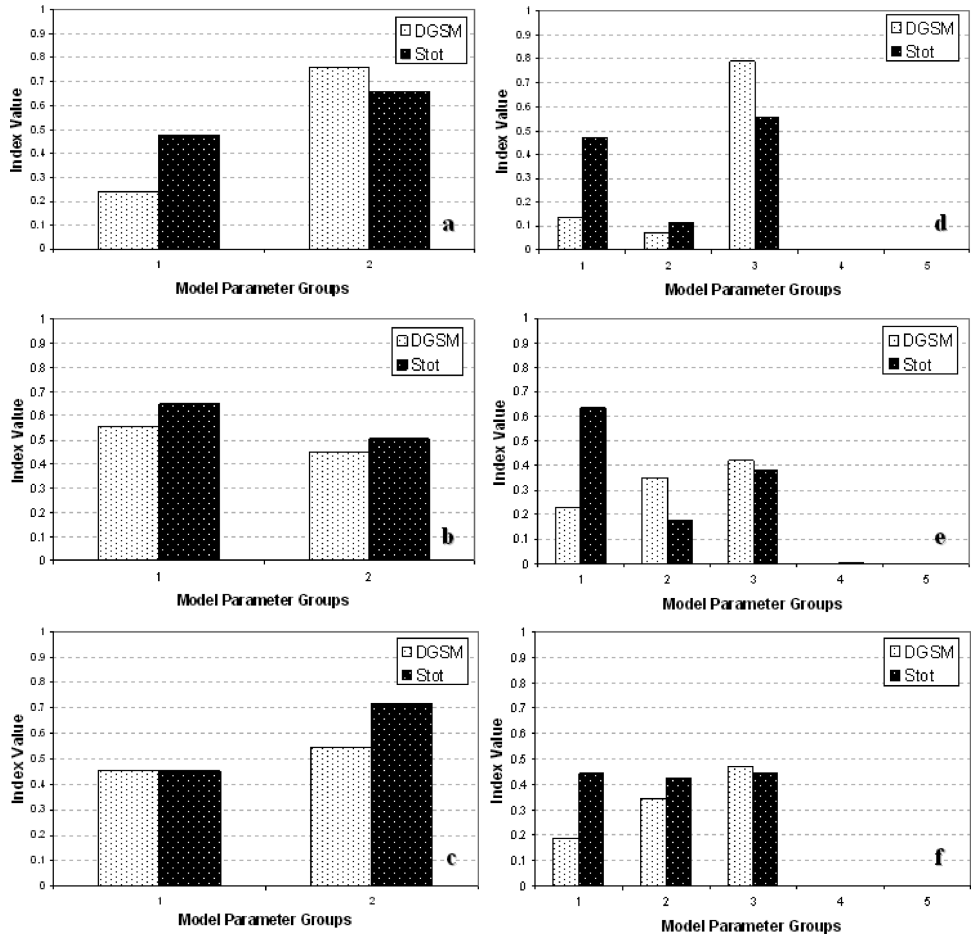

Figure 4. Calculated sensitivity indices for various time points and two different grouping methods. (a−c) Sensitivity indices for glutamine with biological grouping at time points 20, 50, and 120 h. (d−f) Sensitivity indices for glutamine with "functional" grouping at time points 20, 50, and 120 h.

presented in the previous section as a proof of concept. We present results obtained with all three possible groupings and discuss practical issues that have arisen. Figure 4 contains results obtained by grouping according to biological relevance and "functional" grouping. Details on the exact formulation of the two groups for the presented model can be found in Table 3. Table 4 summarizes the computational time required for the generation of these results. The simulations were carried out as described in the previous section, with the only difference being that the sensitivity indices were calculated for parameter groups as discussed above.

Note that the primary purpose of these simulations is to compare the two methods of grouping and not to compare the efficiency of the applied GSA algorithms. Nonetheless, the agreement between the two methods is very good, especially for the biological grouping. From the results presented in Figure 2d−f, we observe the tendency of the Sobol' method to stress out the maximal growth rate, μ_{max}, while DGSM seems to favor the cell yield on glutamine, $Y_{x,gln}$, at least for the initial stages of the culture (20 h). As expected, this difference is also present when studying groups of parameters and is clear in the case of the functional grouping. Group 1, which contains the maximal growth and death rates, is stressed out by the Sobol' method, while group 3 which contains the cell yields is stressed out by DGSM.

The TSI assigned to each group in Figure 4 refers to the combined effect the parameters in the said group have on the model output. However, it yields no information regarding the contribution of each parameter in the group toward the observed TSI value. In order to obtain information about the significance of individual parameters, the modeler would have to perform a second analysis, this time scanning only the parameters of the most significant group individually. Unfortunately, that is an unavoidable computational burden unless a priori knowledge regarding the significance of certain parameters is available.

The computational time required for any GSA method to converge is a direct function of the required model evaluations.

Table 3. Groups of Parameters for the Model Describing Hybridoma Growth[38]

Biological Grouping

group 1: growth/death related	group 2: metabolism related
μ_{max}	$Y_{lac,glc}$
KI_{amm}	m_{glc}
KI_{lac}	$Y_{x,glc}$
K_{glc}	$Y_{x,gln}$
K_{gln}	$K_{d,gln}$
$\mu_{d,max}$	a_1
$K_{d,amm}$	a_2
N	$Y_{amm,gln}$

"Functional" Grouping

group 1: maximal rates	group 2: K_i values	group 3: cell yields ($Y_{i,j}$)	group 4: exponential terms	group 5: maintenance terms
μ_{max}	K_{glc}	$Y_{lac,glc}$	N	m_{glc}
$\mu_{d,max}$	K_{gln}	$Y_{x,glc}$		a_1
	KI_{lac}	$Y_{x,gln}$		a_2
	KI_{amm}	$Y_{amm,gln}$		
	$K_{d,gln}$			
	$K_{d,amm}$			

Table 4. Computational Time for the Calculation of Sensitivity Indices with Various Grouping Methods

GSA methods	computational time (s)	
	biological	functional
DGSM	17 246	32 505
Sobol' global indices	26 669	38 063

The implementation of the Sobol' method presented by Saltelli[18] and Sobol' et al.[19] requires a total of $N(n + 2)$ model evaluations, whereas DGSM requires $N(n + 1)$ model evaluations[6] in the case of numerical evaluation of derivatives and N in the case of automatic differentiation. N stands for the number of quasi-random points used, and n stands for the number of parameters under investigation. Theoretically the accelerated convergence of the QMC method sets at $N \sim \exp(n)$,[48] which can be unrealistic at high n; however, practical studies show that the QMC method outperforms MC even at small N. It is very important to stress the efficiency of DGSM in the case of usage of automatic differentiation. It increases accuracy and solves the problem of choosing correct Δ values, but more importantly it makes the computational time not dependent on n. One can easily deduce that by forming parameter groups the dimensionality of the problem is reduced and therefore the required number of model evaluations is dramatically reduced. Our results are in agreement with the above as can be seen when comparing Tables 2 and 4.

In terms of computational efficiency, as expected, the biological grouping converged faster, since it is of smaller dimensionality. Another notable difference is that the structural grouping allows us to exclude from further analysis a certain number of model parameters, namely those belonging to groups 4 and 5. This provides us with additional information regarding the behavior of the utilized kinetics model and whether it contains redundant parameters. On the other hand, due to the biological simplicity of the utilized model, the biological grouping cannot provide any valuable information allowing us to exclude any parameter groups. It does however inform us that the predominant element of the modeled culture shifts with time. At the early stages (20 h) the depletion of glutamine is metabolism driven, while at 50 h the trend shifts to growth driven, and during the decline phase (120 h) the depletion of glutamine is again metabolism driven. This information can be verified from the ungrouped version of the analysis presented in the previous section, albeit in a more detailed manner. The choice of grouping method ultimately depends on the scope of the analysis conducted.

Finally, random grouping was applied, arbitrarily grouping parameters in four uneven groups. The results are shown in Figure 5 and as expected are consistent both with the other grouping methods and with the ungrouped results. However, it is worthwhile to note that no information on the structure of the model or the biology of the modeled process can be derived from these results. From a theoretical point of view, the results obtained with all three grouping methods are equal and identical, albeit from a practical standpoint the biological and "functional" grouping methods provide valuable insight into the modeled system. Table 5 summarizes the grouping step of the proposed model development algorithm presented in Figure 1.

Unstructured Model of Monoclonal Antibody (mAb) Producing Hybridoma Cell Cultures (33 Parameters). Even though the 16-parameter model that was used in the analysis above is far more complex than the typical test function traditionally utilized, it still remains a rather small and intuitive model in terms of biological significance. Therefore, the model's

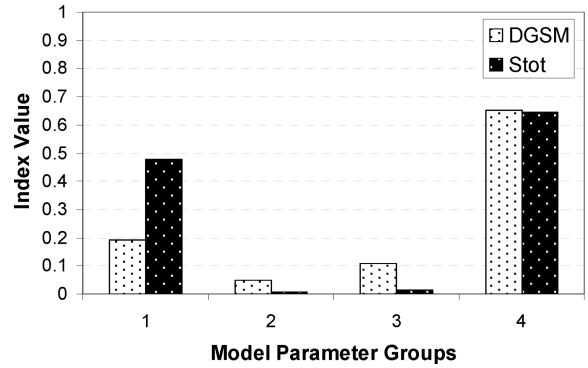

Figure 5. Sensitivity indices for glutamine with random grouping at time point 20 h. Group 1: μ_{max}; k_{glc}; k_{gln}; KI_{lac}; KI_{amm}. Group 2: $\mu_{d,max}$; $k_{d,amm}$; n. Group 3: $Y_{lac,glc}$; m_{glc}; $Y_{x,glc}$. Group 4: $k_{d,gln}$; $Y_{x,gln}$; $Y_{amm,gln}$; a_1; a_2.

Table 5. Overview of Grouping Parameters in GSA

consideration	scope of the analysis
number of groups	
1. few groups	computationally efficient, low resolution
2. many groups	computationally expensive, high resolution
grouping method	
1. random/arbitrary	parameter significance ranking
2. biological significance related	model analysis leading to DOE
3. functional	model analysis/reduction

size and biological relevance were enhanced using a hybrid model,[40] which describes monoclonal antibody production by hybridoma cell cultures. Specifically, the unstructured model describing cellular growth (described above) is coupled to a structured model describing the process of antibody formation in the cell. The hybrid model assumes no intracellular accumulation of the species involved in the mAb production pathway and no proteolytic or other degradation of the antibody chains in the cell.

The model consists of an intracellular heavy- and light-chain mRNA balance:

$$\frac{dm_H}{dt} = N_H S_H - K m_H \tag{34}$$

and

$$\frac{dm_L}{dt} = N_L S_L - K m_L \tag{35}$$

where m_H and m_L are the intracellular heavy- and light-chain mRNA concentrations (mRNAs/cell), N_H and N_L are the heavy- and light-chain gene copy numbers (genes/cell), S_H and S_L are the heavy- and light-chain gene specific transcription rates (mRNAs/gene/h), and, finally, K is the heavy- and light-chain mRNA decay rate (h^{-1}).

The intra endoplasmic reticulum (ER) heavy- and light-chain balances are

$$\frac{d[H]}{dt} = T_H m_H - R_H \tag{36}$$

and

$$\frac{d[L]}{dt} = T_L m_L - R_L \tag{37}$$

where [H] and [L] are the free heavy- and light-chain concentrations in the ER (chains/cell), T_H and T_L are the heavy- and light-

chain specific translation rates (chains/mRNA/h), and R_H and R_L are the rates of heavy- and light-chain consumption in assembly (chains/cell/h). mAb's consist of two heavy (H) and two light (L) amino acid chains. Each molecule is synthesized in the ER following the sequence described below:

$$H + H \leftrightarrow H_2 \qquad H_2 + L \leftrightarrow H_2L \qquad H_2L + L \leftrightarrow H_2L_2 \tag{38}$$

Assuming that the rates of heavy- and light-chain consumption in the assembly stage are given by

$$R_H = \frac{2}{3}K_A[H]^2, \qquad R_L = 2K_A[H_2][L] + K_A[H_2L][L] \tag{39}$$

an intra-ER balance can be performed for each of the assembly intermediates:

$$\frac{d[H_2]}{dt} = \frac{1}{3}K_A[H]^2 - 2K_A[H_2][L] \tag{40}$$

$$\frac{d[H_2L]}{dt} = 2K_A[H_2][L] - K_A[H_2L][L] \tag{41}$$

where $[H_2]$ and $[H_2L]$ are the concentrations of the assembly intermediates in the ER (molecules/cell), and K_A is the assembly rate constant ((molecules/cell) h^{-1}).

A balance can then be performed on the assembled mAb structure ($[H_2L_2]_{ER}$) in the ER:

$$\frac{d[H_2L_2]_{ER}}{dt} = K_A[H_2L][L] - K_{ER}[H_2L_2]_{ER} \tag{42}$$

where $[H_2L_2]_{ER}$ is the mAb concentration in the ER (molecules/cell) and K_{ER} is the rate constant for ER-to-Golgi antibody transport (h^{-1}). Once the mAb is assembled in the ER, it proceeds to the Golgi apparatus, where the main part of its glycosylation process takes place. An intra-Golgi mAb balance yields

$$\frac{d[H_2L_2]_G}{dt} = \varepsilon_1 K_{ER}[H_2L_2]_{ER} - K_G[H_2L_2]_G \tag{43}$$

where $[H_2L_2]_G$ is the mAb concentration in the Golgi (molecules/cell), ε_1 is the ER glycosylation efficiency factor, and K_G is the rate constant for Golgi-to-extracellular medium antibody transport (h^{-1}). Finally, the expression for antibody secretion (production) is

$$\frac{d(V[mAb])}{dt} = (\gamma_2 - \gamma_1\mu)Q_{mAb}VX_u \tag{44}$$

where

$$Q_{mAb} = \varepsilon_2\lambda K_G[H_2L_2]_G \tag{45}$$

where Q_{mAb} is the specific mAb production rate (mg/cell/h), λ is the molecular weight of IgG_1 (g/mol), and ε_2 is the Golgi glycosylation efficiency factor. In eq 44, [mAb] is the mAb concentration in the culture, and $\gamma_1$and γ_2 are constants.

The model consists of 15 differential equations and has a total of 33 parameters. As stated earlier, a batch operation mode was considered, the model was simulated for 150 h of culture time, and the sensitivity analysis was performed at three characteristic time points (20, 50, and 120 h). The simulations involved scanning of all model parameters with respect to each possible model output variable. Results for viable cell concentration and mAb concentration are presented in Figure 6. The uncertainty range associated with each of the 33 model parameters was set to ±100% from the parameter nominal value. Following the discussion in the previous section, using parameter grouping was imposed by the dimensionality of the problem and the model's 33 parameters were grouped into four unequal groups, which were chosen based on their biological function. For the purposes of our computational investigation concerning the efficiency of the two methods under study, any form of parameter grouping that would reduce the dimensionality of the said problem would yield analogous results. Therefore it is adequate to study the computational efficiency of the two methods on one type of grouping. Moreover, parameter grouping is a step in the GSA algorithm that cannot be formalized as it essentially depends on the scope of the analysis, the a priori knowledge available, and the modeler's personal judgment. Furthermore, as discussed in the previous section, parameter grouping does not affect the results of the analysis and merely affects the interpretability of the results. The list of model parameters, their nominal values, and the formulated groups can be found in Table 6.

Figure 6 presents the sensitivity indices for the four groups of parameters that were formed and how these affect two of the most significant model variables, namely viable cell concentration and mAb concentration. Again both methods agree in their predictions of the significance ranking for all groups. In the case of viable cell concentration the agreement is close to perfect, whereas in the case of mAb concentration discrepancies are somewhat more noticeable. However, one must bear in mind the different context from which indices are derived for each method as well as the fact that eq 44, which gives mAb concentration, is a coupling point for the unstructured part of the model with the structured part and therefore a source of nonlinearity. In terms of computational time, as we had previously predicted,[44] the gap between the two methods increases dramatically as the complexity of the model increases. The data in Table 7 show a 14-fold reduction in computational time.

Discussion

Following the discussion in the Introduction, one objective that has been set and hopefully met was to elucidate the application and use of GSA in the context of biological models. Avoiding the use of test functions, we compared GSA techniques on actual models of real biological systems, thus identifying and overcoming real concerns and computational problems, for example, the need for grouping parameters. While DGSM[6] did not originally include a grouping algorithm, we have successfully presented a technique to group parameters with DGSM, based on an adaptation of the results reported by Campolongo et al.[46]

We have already shown[44] the potential of the new proposed global screening technique presented, namely DGSM.[6] As presented in the section Biological Models, the implementation of the method has been improved, so its performance can be directly compared to that of the computationally expensive and established variance-based methods. However, even though the quality and more importantly the robustness of DGSM have been successfully improved, the computational times still remain significantly smaller that those of any competing global method. Tables 2, 4, and 7 summarize the computational costs required for the collection of the presented data. DGSM obviously saves impressive amounts of time, while proving to be a robust screening technique. While the second model we have used is not a large one, considering the model sizes met in biological

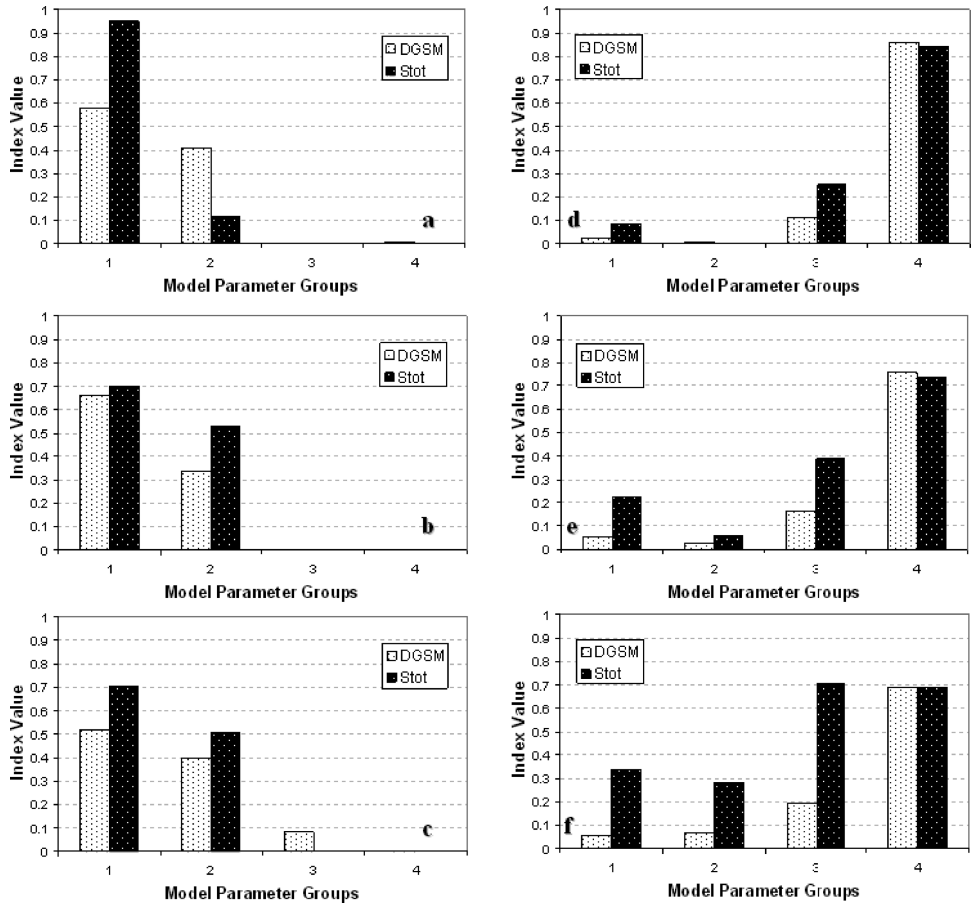

Figure 6. Calculated sensitivity indices for various time points and two different output variables. (a−c) Sensitivity indices for viable cell concentration at time points 20, 50, and 120 h. (d−f) Sensitivity indices for mAb concentration at time points 20, 50, and 120 h.

Table 6. Model Parameters: Nominal Values[40] and Grouping

symbol	units	nominal value	symbol	units	nominal value
	Group 1: Growth/Death Related			Group 2: Metabolism Related	
μ_{max}	h^{-1}	5.8×10^{-3}	$Y_{lac,glc}$	dimensionless	1.399
KI_{amm}	mM	28.484	m_{glc}	mmol/cell/h	4.853×10^{-14}
KI_{lac}	mM	171.756	$Y_{x,glc}$	cells/mmol	1.061×10^{8}
K_{glc}	mM	0.75	$Y_{x,gln}$	cells/mmol	5.565×10^{8}
K_{gln}	mM	0.075	$K_{d,gln}$	h^{-1}	9.6×10^{-3}
$m_{d,max}$	h^{-1}	0.03	a_1	mM·L/cell/h	3.4×10^{-13}
$K_{d,amm}$	mM	0.1386	a_2	mM	4
N	real integer	0.995	$Y_{amm,gln}$	dimensionless	0.4269
	Group 3: mAb Synthesis Related			Group 4: mAb Secretion Related	
K	h^{-1}	0.1	γ_1	dimensionless	0.1
N_H	genes/cell	139.8	γ_2	dimensionless	2
S_H	mRNAs/gene/h	300	ε_2	dimensionless	1
N_L	genes/cell	117.5	λ	g/mol	146
S_L	mRNAs/gene/h	4500			
T_H	chains/mRNA/h	17			
T_L	chains/mRNA/h	11.5			
K_A	(molecules/cell) h^{-1}	10^{-6}			
K_{ER}	h^{-1}	0.693			
K_G	h^{-1}	0.1386			
ε_1	dimensionless	0.995			

systems, we feel that it is adequately sized to point out any possible problems or computational issues related to model size. We are therefore confident that the results drawn from this computational comparison can be readily applied to larger scaled biological models.

While parameter grouping is not an inherent part of sensitivity analysis algorithms, we have designed and examined the efficiency of two proposed methodologies for grouping parameters in larger scale models. First, grouping based on the

Table 7. Computational Time for the Model Describing mAb Production by a Hybridoma Cell Line

	computational time (s)	
time points	Sobol' global indices	DGSM
mAb producing model		
20 h	185 283	8 937
50 h	213 232	10 475
120 h	233 384	24 242
total	631 900 (7.31 days)	43 654 (0.50 day)

biological role of the parameters has been successfully employed previously.[24,32] Herein, the concept of "functional" grouping was evaluated, proving to be an efficient alternative. A detailed analysis of the nature of information derived by GSA on parameter groups was performed in order to verify that the results were in agreement with the ungrouped problem. Interestingly, superimposing the results obtained by the two different grouping approaches may allow the identification of significant parameters more efficiently (data not shown). This has significant implications in large scale models, where "multilayer" sensitivity analysis must be performed traditionally in order to identify significant individual parameters. Thus, by superimposing the results derived from the two different grouping strategies, the resolution of the results can be enhanced while still remaining on the smallest possible dimensionality.

Sensitivity analysis techniques are an essential part of the model building process[48] (Figure 1), as they associate the uncertainty inherent in every model with the various sources of uncertainty, namely the model parameters. Therefore, parameters with high influence on the model output need to be validated experimentally with greater accuracy, thus reducing the model uncertainty. Parameters that do not influence the model output can be set to their nominal value. All in all, sensitivity analysis can aid the modeler by identifying the sources of uncertainty in his model and at the same time aid the experimentalist by indicating which experimental measures would be the most informative. With the present work we have successfully clarified the need and utility of GSA in the context of biological modeling, and we introduced a new sensitivity analysis technique (DGSM) which greatly reduces computational costs, making GSA a more widely affordable option. Finally, we have presented an extension to DGSM, making it possible to study the effects of parameter groups on the model output.

We have shown, in two separate cases, that a mere few parameters out of the total usually affect the model output significantly. This result alone can provide valuable information to the experimentalist, guiding the main focus and purpose of the experiments to be performed. Moreover, as previous studies have already shown,[4,24,32] GSA can be the first step in a design of experiments framework (Figure 1). The efficiency of this framework has been further increased by a significant reduction in the computational time that must be invested in GSA.

Acknowledgment

Financial support from the PROBACTYS (CT-2006-029104) and SYSMO (BBF0035601) grants is gratefully acknowledged.

Appendix: Model Development Framework

Figure 1 is a schematic representation of the model development framework for biological systems established by previous work in the field.[4,7,8] Let us assume a first-principle mathematical model, $g(x,\bar{P})$, formulated to describe a real life process, where x denotes the input vector and $\bar{P}_i$ (where $i = 1,..., \nu$) denotes the parameter vector.

The first step of the model development algorithm is to apportion the uncertainty in the model output to the sources of variation, namely perform GSA. The output of GSA will be a vector of size ν, containing the total sensitivity indices of the model parameters. As a rule of thumb, if $\nu > 20$, the use of parameter grouping will become necessary.[44] This merely affects the notation of Figure 1 and in no way affects the algorithm itself. Therefore, in the case of parameter grouping, $\bar{P}_i$ is the parameter vector and ν is the number of partitions it contains, corresponding to the number of parameter groups formed.

Consequently an empirical criterion, determined by the modeler, is applied in order to discriminate the significant from the insignificant model parameters. The criterion is a threshold value for the TSI, usually set between 0 and 0.2.[9] In the present work the threshold value for TSI was set equal to 0, since the main objective was to compare the computational efficiency of the two presented algorithms and to study the results both methods give even for parameters with small TSI values. Any parameters with values below the set threshold are considered insignificant to the model output and are allocated in a partition of the parameter vector termed $\bar{p}_j^1$ ($j = 1,..., \nu'$). The remaining parameters whose TSI is above the threshold value are allocated in a second partition of the parameter vector termed $\bar{p}_k^2$ ($k = 1,..., \nu''$). The sum of ν' and ν'' should of course always equal the size of the parameter vector $\bar{P}_i$, ν. The values of the parameters in partition $\bar{p}_j^1$ are set to the nominal values, which can be derived either from existing literature or from a parameter estimation algorithm, hence yielding the parameter vector $\bar{p}_j^{\text{nom}}$.

The values of the model affecting parameters in partition $\bar{p}_k^2$ need to be determined experimentally with accuracy in order to reduce the uncertainty in the model output. Therefore, experiments are specifically designed for the determination of the parameters in vector $\bar{p}_k^2$, and once the experimental data are available, the values for parameters $\bar{p}_k^2$ are determined explicitly, yielding vector $\bar{p}_k^{\text{exp}}$. Finally, by substituting the initial parameter vector $\bar{P}_i$, with the newly derived $\bar{p}_j^{\text{nom}}$ and $\bar{p}_k^{\text{exp}}$, we derive a refined version of the original model, $g(x,\bar{p}_j^{\text{nom}},\bar{p}_k^{\text{exp}})$. The refined model has a significantly lower level of uncertainty compared to the initial model.

Literature Cited

(1) Baker, M. Upping the ante on antibodies. *Nat. Biotechnol.* **2005**, *23*, 1065–1072.

(2) Frahm, B.; Lane, P.; Märkl, H.; Pörtner, R. Improvement of a mammalian cell culutre process by adaptive, model-based dialysis fed-batch cultivation and suspenssion of apoptosis. *Bioprocess Biosyst. Eng.* **2003**, *26*, 1–10.

(3) De Tremblay, M.; Perrier, M.; Chavarie, C.; Achambault, J. Fed-batch culture of hybridoma cells: comparison of optimal control approach and closed loop strategies. *Bioprocess Eng.* **1993**, *9*, 13–21.

(4) Sidoli, F. R.; Mantalaris, A.; Asprey, S. P. Toward Global Parametric Estimability of a Large-Scale Kinetic Single-Cell Model for Mammalian Cell Culture. *Ind. Eng. Chem. Res.* **2005**, *44* (4), 868–878.

(5) Sobol', I. M. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math. Comput. Simul.* **2001**, *55*, 271–280.

(6) Kucherenko, S.; Rodriguez-Fernandez, M.; Pantelides, C.; Shah, N. Monte Carlo evaluation of Derivative based Global Sensitivity Indices. *Reliab. Eng. Syst. Saf.* **2007**, in press.

(7) Kontoravdi, C.; Asprey, S. P.; Pistikopoulos, E. N.; Mantalaris, A. Application of Global Sensitivity Analysis to Determine Goals for Design of Experiments: An Example Study on Antibody-Producing Cell Cultures. *Biotechnol. Prog.* **2005**, *21* (4), 1128–1135.

(8) Ho, Y.; Varley, J.; Mantalaris, A. Development and analysis of a mathematical model for antibody-producing GS-NS0 cells under normal and hyperosmotic culture conditions. *Biotechnol. Prog.* **2006**, *22* (6), 1560–9.

(9) Saltelli, A.; Chan, K.; Scott, E. M. *Sensitivity Analysis*; Wiley: New York, 2000.

(10) Box, G. E. P.; Hunter, W. G.; Hunter, J. S. *Statistics for Experimenters. An introduction to Design, Data Analysis and Model Building*; Wiley: New York, 1978.

(11) Cotter, S. C. A screening design for factorial experiments with interactions. *Biometrika* **1979**, *66*, 317–320.

(12) Homma, T.; Saltelli, A. Importance measures in global sensitivity analysis of nonlinear models. *Reliab. Eng. Syst. Saf.* **1996**, *52*, 1–17.

(13) Chan, K.; Saltelli, A.; Tarantola, S. Sensitivity Analysis of the model output: variance-based methods make the difference. *Proceedings of the 1997 Winter Simulation Conference*.

(14) Sobol', I. M. On sensitivity estimation for nonlinear mathematical models. *Matem. Mod.* **1990**, *2*, 112–118.

(15) Saltelli, A.; Tarantola, S.; Chan, K.P.-S. A Quantitative Model-Independent Method for Global Sensitivity Analysis of Model Output. *Technometrics* **1999**.

(16) Weyl, H. Mean motion. *Am. J. Math.* **1938**, *60*, 889–896.

(17) Frey, C. H.; Patil, S. R. Identification and Review of Sensitivity Analysis Methods. *Risk Anal.* **2002**, *22* (3), 553–578.

(18) Saltelli, A. Making best use of model evaluations to compute sensitivity indices. *Comput. Phys. Commun.* **2002**, *145*, 280–297.

(19) Sobol, I. M.; Tarantola, S.; Gatelli, D.; Kucherenko, S.; Mauntz, W. Estimating the Approximation Error when fixing Unessential Factors in Global Sensitivity Analysis. *Reliab. Eng. Syst. Saf.* **2007**, *92* (7), 957–960.

(20) Morris, M. D. Factorial Sampling plans for preliminary computational experiments. *Technometrics* **1991**, *33*, 161–174.

(21) Sobol', I. M.; Kucherenko, S. Derivative based Global Sensitivity Measures and their link with global sensitivity indices. *Math. Comput. Simul.* **2008**, in press.

(22) Thomas, I. M.; Kiparissides, C. Sensitivity Analysis of a Batch Polymerization Reactor. *J. Appl. Polym. Sci.* **1984**, *29*, 2195–2204.

(23) Rabitz, H.; Krammer, M.; Dacol, D. Sensitivity Analysis in Chemical Kinetics. *Annu. Rev. Phys. Chem.* **1983**, *34*, 419–461.

(24) Norva, G. W.; Phillips, M. J.; Missen, R. W.; Smith, W. R. Applications of Equilibrium Sensitivity Analysis to Aromatization Processes. *Ind. Eng. Chem. Res.* **1989**, *28* (12), 1884–1887.

(25) Meadows, E. S.; Crowley, T. J.; Immanuel, C. D.; Doyle, F. J., III. Nonisothermal Modeling and Sensitivity Studies for Batch and Semibatch Emulsion Polymerization of Styrene. *Ind. Eng. Chem. Res.* **2003**, *42* (3), 555–567.

(26) Scaillet, O. Nonparametric Estimation and Sensitivity Analysis of Expected Shortfall. *Math. Finance* **2003**, *14*, 115–129.

(27) Wood, S. N.; Thomas, M. B. Super-sensitivity to structure in biological models. *Proc. R. Soc. London, Ser. B* **1999**, *266*, 565–570.

(28) Yi, C. K.; Luyben, W. L. Evaluation of Plant-Wide Control Structures by Steady-State Disturbance Sensitivity Analysis. *Ind. Eng. Chem. Res.* **1995**, *34* (7), 2393–2405.

(29) Jia, Z.; Ierapetritou, M. G. Short-Term Scheduling under Uncertainty Using MILP Sensitivity Analysis. *Ind. Eng. Chem. Res.* **2004**, *43* (14), 3782–3791.

(30) Wang, C. J.; Wong, D. S. H.; Chien, I.-L.; Shih, R. F.; Liu, W. T.; Tsai, C. S. Critical Reflux, Parametric Sensitivity, and Hysteresis in Azeotropic Distillation of Isopropyl Alcohol + Water + Cyclohexane. *Ind. Eng. Chem. Res.* **1998**, *37* (7), 2835–2843.

(31) Feng, X. J.; Hooshangi, S.; Chen, D.; Li, G.; Weiss, R.; Rabitz, H. Optimising genetic circuit by global sensitivity analysis. *Biophys. J.* **2004**, *87*, 2195–2202.

(32) Cho, K. W.; Shin, S. Y.; Kolch, W.; Wolkenhauer, O. Experimental Design in Systems Biology, Based on Parameter Sensitivity Analysis Using a Monte Carlo Method: A Case Study for the TNFα-Mediated NF-κB Signal Transduction Pathway. *SIMULATION* **2003**, *79* (12), 726–739.

(33) Gunawan, R.; Cao, Y.; Petzold, L.; Doyle, F. J., III. Sensitivity Analysis of discrete stochastic systems. *Biophys. J.* **2005**, *88*, 2530–2540.

(34) Leis, J. R.; Kramer, M. A. The simultaneous solution and sensitivity analysis of systems described by ordinary differential equations. *ACM Trans. Math. Software* **1988**, *14* (1), 45–60.

(35) Fennel, K.; Losch, M.; Schroter, J.; Wenzel, M. Testing a marine ecosystem model: sensitivity analysis and parameter optimisation. *J. Mar. Syst.* **2001**, *28*, 45–63.

(36) Borgonovo, E.; Apostolakis, G. E. A new importance measure for risk-informed decision making. *Reliab. Eng. Syst. Saf.* **2001**, *72*, 193–212.

(37) Sidoli, F. R.; Mantalaris, A.; Asprey, S. P. Modelling of mammalian cells and cell culture processes. *Cytotechnology* **2004**, *44*, 27–46.

(38) Kontoravdi, C.; Asprey, S. P.; Pistikopoulos, E. N.; Mantalaris, A. Development of a dynamic model of monoclonal antibody production and glycosylation for product quality monitoring. *Comput. Chem. Eng.* **2007**, *31*, 392–400.

(39) Jang, J. D.; Barford, J. P. An unstructured kinetic model of macromolecular metabolsim in batch and fed-batch cultures of hybridoma cells producing monoclonal antibodies. *Biochem. Eng. J.* **2000**, *4*, 153–168.

(40) Kontoravdi, C. Development of a combined mathematical and experimental framework model for modelling mammalian cell cultures. Ph.D. Thesis, Department of Chemical Engineering and Chemical Technology Imperial College, London, 2006.

(41) Tatiraju, S.; Soroush, M.; Mutharasan, R. Multi-rate nonlinear state and parameter estimation in a bioreactor. *Biotechnol. Bioeng.* **1998**, *63* (1), 22–32.

(42) Chibnall, A. C.; Westall, R. G. The estimation of glutamine in the presence of aspargine. *Biochem. J.* **1932**, *26* (1).

(43) Bray, H. G.; James, S. P.; Raffan, I. M.; Thorpe, W. V. The enzymic hydrolysis of Glutamine and its spontaneous decomposion in Buffer Solutions. *Biochem. J.* **1948**, *44*, 625–627.

(44) Kiparissides, A.; Rodriguez-Fernandez, M.; Kucherenko, S.; Mantalaris, A.; Pisitikopoulos, E. N. Application of Global Sensitivity Analysis to Biological Models. *Comput.-Aided Chem. Eng.* **2008**, *25*, 689–694.

(45) Mollá, G. V. M.; Padilla, G. R. *Description of the MATLAB functions SENS_SYS and SENS_IND*; 2002; http://www.mathworks.com/matlabcentral/.

(46) Campolongo, F.; Cariboni, J.; Saltelli, A. An effective screening design for sensitivity analysis of large models. *Environ. Modell. Software* **2007**, *22*, 1509–1518.

(47) Morokoff, W. J.; Caflisch, R. E. Quasi-Random Sequences and their Discrepancies. *SIAM J. Sci. Comput.* **1994**, *15*, 1251–1279.

(48) Saltelli, A.; Tarantola, S.; Campolongo, F. Sensitivity Analysis as an Ingredient of Modeling. *Stat. Sci.* **2000**, *15*, 377–395.

Received for review January 27, 2009
Revised manuscript received April 22, 2009
Accepted June 18, 2009

IE900139X