Automatica 47 (2011) 1147–1155

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

'Closing the loop' in biological systems modeling — From the *in silico* to the *in vitro*☆

Alexandros Kiparissides, Michalis Koutinas, Cleo Kontoravdi, Athanasios Mantalaris, Efstratios N. Pistikopoulos*

Biological Systems Engineering Laboratory, Centre for Process Systems Engineering, Imperial College London, SW7 2AZ, UK

ARTICLE INFO

Article history:
Received 1 February 2010
Received in revised form
14 September 2010
Accepted 22 October 2010
Available online 26 February 2011

Keywords:
Model analysis
Biological systems model development
Sensitivity analysis
Design of experiments

ABSTRACT

This work presents a holistic 'closed loop' approach for the development of models of biological systems. The ever-increasing availability of experimental information necessitates the advancement of a systematic methodology to organise and utilise these data. Herein, we present a biological model building framework that maps the treatment of the information from the initial conception of the model, through its experimental validation and finally to its application in model-based optimisation studies. We highlight and discuss current issues associated with the development of mathematical models of biological systems and share our perspective towards a holistic 'closed loop' approach that will facilitate the control of the *in vitro* through the *in silico*.

1. Introduction

Mathematical models of biological systems developed over the last decades incorporate various degrees of structure and mathematical complexity. Models of single cells, cell populations and microbial cultures have been central in the understanding and improvement of biological systems, as well as in the optimisation and control of bioprocesses (Thilakavathi, Basak, & Panda, 2007). The large-scale generation of biological data obtained with the development of a variety of high-throughput experimental technologies demand for mathematical model building to become a centre of importance in biology (Covert et al., 2001). Alas, as Bailey (1998) argued the development of mathematically and computationally orientated research has failed to catch up with the recent developments in biology. Furthermore, he concluded that the little attention that mathematical modelling of biological systems receives from experimentalists could be partly attributed to the lack of effective communication of the benefits of formulating and using a mathematical model.

Even relatively simple microorgansisms, which have been extensively studied, are hosts to a complex network of interconnected processes occurring on diverse time scales within a confined volume. The multilevel nature of the regulatory network of cells and the interactions occurring at the intracellular level further augment this complexity (Yokobayashi, Collins, Leadbetter, Weiss, & Arnold, 2003). Therefore, attempts to wholly model the function of even a single cell are currently non-trivial, if not impossible. The amount of delicate intracellular measurements required to validate such a model is exhaustive both in terms of labour as well as cost. Uncertainties introduced on the parameter identifiability level (Sidoli, Mantalaris, & Asprey, 2004) and on the mechanistic level further complicate this task.

Borrowing research principles from the Process Systems Engineering paradigm, mathematical modelling of biological systems can provide a systematic means to quantitatively study the characteristics of the multilevel interactions that occur in cell bio-processing. The literature around mathematical modelling of biological systems, be they prokaryotic or eukaryotic, is arguably too vast to summarise within the limited space of a paper. Indicatively, mathematical models have been successfully used to design optimal media (Xie & Wang, 1994), identify previously ignored growth limiting factors (deZengotita, Miller, & Aunins, 2000), optimise culture growth and productivity (De Tremblay, Perrier, Chavarie, & Archambault, 1992; Dhir, Morrow, Rhinehart, & Wiesner, 1999; San & Stephanopoulos, 1989), and apply control principles to cell culture processes (Frahm et al., 2002).

Pörtner and Schäfer (1996) compared a selection of models that existed in the literature at that time and carried out an analytic error and range of validity analysis. They found significant

☆ This work was supported by the European Union with the following projects: (a) PROBACTYS (FP6 — NEST-PATHFINDER EU call on Synthetic Biology, Project Number 029104), (b) PSYSMO in the framework of the SYSMO initiative (BBSRC — ERA-NET program on the Systems Biology of Microorganisms, Project Number 0133980) (c) TARPOL (FP7 EU — KBBE Coordination Action for SynBio in Environmental Sciences), (d) European Research Council (ERC) Mobile Project (No: 226462) and (e) EP/E047017/1. The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Juergen Hahn under the direction of Editor Francis J. Doyle III.

* Corresponding author. Tel.: +44 0 20 7594 6620; fax: +44 0 20 7594 6606.
E-mail address: e.pistikopoulos@imperial.ac.uk (E.N. Pistikopoulos).

0005-1098/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2011.01.013

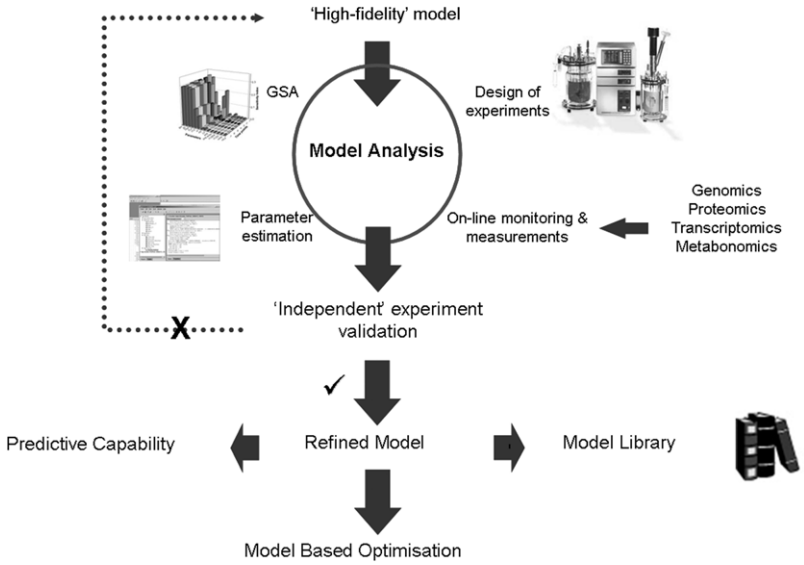

Fig. 1. Model development framework for biological systems.

variations in the values of maximum growth rate, yields and nutrient Monod constants used by researchers. They came to the conclusion that the models' predictions involved significant errors, particularly due to the lack of understanding of cellular metabolism and the limited data ranges within which the model was valid. They further pointed out that the majority of studies presented either utilise literature data to validate the models or generate their own experimental data without any form of systematic design of experiments. In order to maximise the gains from the ever increasing influx of biological information, the approach to modelling needs to shift towards a systematic framework from conception to optimisation. Herein we attempt to formalise a structure upon which experimental and mathematical biology can interact seamlessly. The use of model-based techniques can facilitate the reduction of unnecessary experimentation hence reducing operating labour and cost by indicating the most informative experiments and providing strategies to optimise and automate the process at hand.

2. Biological systems model development framework

One of the challenges in biological systems engineering is the development of high fidelity models able to capture the required biological functions while remaining computationally tractable in order to be viable candidates for model-based control and optimisation. However, high fidelity models, inherently, contain a large number of parameters. Use of a framework that designs experiments in a way that minimises cost and labour whilst simultaneously maximising information obtainable from the data, is the first step towards achieving a uniquely validated model (Sidoli et al., 2004). The work presented by Asprey and Mantalaris (2002), Ho, Varley, and Mantalaris (2006), Kiparissides, Kucherenko, Mantalaris, and Pistikopoulos (2009), Kontoravdi, Asprey, Pistikopoulos, and Mantalaris (2007), Kontoravdi, Pistikopoulos, and Mantalaris (2010), Lam et al. (2008), Sidoli et al. (2004), Sidoli, Mantalaris, and Asprey (2005) defines a systematic approach to modelling biological systems which is depicted in Fig. 1.

The presented framework sets a scientific platform of communication between modeller and experimentalist, thus bridging the communication gap. Each step of the framework organises and directs the flow of experimental information in an effort to alleviate uncertainty where possible. More specifically:

- *Step* 1. *Model development*.

  Define the aim of the model and choose an appropriate model type (i.e. structured vs. unstructured). Define the model equations through 'First principles' relationships.

- *Step* 2. *Parameter identifiability test* (Asprey & Mantalaris, 2002; Sidoli et al., 2004).

  Test how many parameters are identifiable from the available model inputs. If an adequate number of parameters cannot be identified the model structure needs to change.

- *Step* 3. *Model analysis* (Kiparissides et al., 2009).

  Study how the uncertainty introduced through the parameter values affects the model's outputs and define parameters crucial to the model's output. Set insignificant parameters to their nominal values.

- *Step* 4. *Optimal experimental design* (Kontoravdi et al., 2010).

  Based on the available experimental measurements design optimal experiments that minimise the uncertainty in the significant parameters. Estimate parameter values from tailor-made experiments.

- *Step* 5. *Range of validity* (Kontoravdi et al., 2010).

  Test the validity of the model against a set of independent data and against experiments with varying environmental conditions.

- *Step* 6. *Model based optimisation–automation* (Lam et al., 2008).

Optimise the inputs of the modelled bioprocess towards a desired goal (i.e. end product maximisation). Increase reproducibility and stability by introducing process control & automation where applicable.

3. Closing the loop: a holistic approach to monoclonal antibody production

This section aims to highlight the potential of the framework through its application on a "real life" process. An example of an industrial process for the production of monoclonal antibodies (mAbs) harvested from cultures of hybridoma cells is employed. The aim is to model this process and maximise the final antibody titre in the culture through *in silico* experimentation. Batch and fed-batch cultures are currently the culture methods of choice by the biologics industry for the large scale production of mAbs, due to their operational simplicity, reliability, and flexibility for implementation in multipurpose facilities (Bibila & Robinson, 1995).

Table 1
Table of model equations.

$$\frac{dVX_u}{dt} = \mu VX_u - \mu_d VX_u \quad (1)$$

$$\frac{dVX_t}{dt} = \mu VX_u - K_{lysis} V(X_t - X_u) \quad (2)$$

$$\mu = \mu_{max}\left(\frac{[GLC]}{K_{glc}+[GLC]}\right)\left(\frac{[GLN]}{K_{gln}+[GLN]}\right)\left(\frac{KI_{amm}}{KI_{amm}+[AMM]}\right)\left(\frac{KI_{lac}}{KI_{lac}+[LAC]}\right) \quad (3)$$

$$\mu_d = \frac{\mu_{d,max}}{1+\left(\frac{K_{d,amm}}{[AMM]}\right)^n} \quad (4)$$

$$\frac{d(V[GLC])}{dt} = -\left(\frac{\mu}{Y_{x,glc}} + m_{glc}\right) VX_u \quad (5)$$

$$\frac{d(V[GLN])}{dt} = -\left(\frac{\mu}{Y_{x,gln}} + m_{gln}\right) VX_u - K_{d,gln} V[GLN] \quad (6)$$

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

$$\frac{d(V[AMM])}{dt} = Y_{amm,gln}\left(\frac{\mu}{Y_{x,gln}} + m_{gln}\right) VX_u + K_{d,gln} V[GLN] \quad (7)$$

$$\frac{d(V[LAC])}{dt} = Y_{lac,glc}\left(\frac{\mu}{Y_{x,glc}} + m_{glc}\right) VX_u \quad (8)$$

$$\frac{dm_H}{dt} = N_H S_H - Km_H \quad (9)$$

$$\frac{dm_L}{dt} = N_L S_L - Km_L \quad (10)$$

$$\frac{d[H]}{dt} = T_H m_H - R_H \quad (11)$$

$$\frac{d[L]}{dt} = T_L m_L - R_L \quad (12)$$

$$H + H \longleftrightarrow H_2$$
$$H_2 + L \longleftrightarrow H_2L \quad (13)$$
$$H_2L + L \longleftrightarrow H_2L_2$$

$$R_H = \tfrac{2}{3} K_A [H]^2 R_L = 2K_A[H_2][L] + K_A[H_2L][L] \quad (14)$$

$$\frac{d[H_2]}{dt} = \tfrac{1}{3} K_A[H]^2 - 2K_A[H_2][L] \quad (15)$$

$$\frac{d[H_2L]}{dt} = 2K_A[H_2][L] - K_A[H_2L][L] \quad (16)$$

$$\frac{d[H_2L_2]_{ER}}{dt} = K_A[H_2L][L] - K_{ER}[H_2L_2]_{ER} \quad (17)$$

$$\frac{d[H_2L_2]_G}{dt} = \varepsilon_1 K_{ER}[H_2L_2]_{ER} - K_G[H_2L_2]_G \quad (18)$$

$$\frac{d(V[MAb])}{dt} = (\gamma_2 - \gamma_1\mu) Q_{MAb} VX_V \quad (19)$$

$$Q_{MAb} = \varepsilon_2 \lambda K_G [H_2L_2]_G$$

Therefore, for the purposes of our example a model capable of describing both batch and fed-batch cell cultures is required. Since the model will ultimately be utilised for optimisation studies, which are inherently computationally intense techniques, detailed structured models become less applicable. As Sidoli et al. (2005) have argued, overparameterised models lead to parameter identifiability issues which in turn reduce confidence in the model output. Balancing the trade-off between tractability and fidelity is addressed by hybrid models (Ho et al., 2006; Kontoravdi et al., 2007; Lam et al., 2008). A step-by-step overview of the algorithm is provided using experimental and modelling results from the work of Kontoravdi (2006), Kontoravdi et al. (2010) and Lam (2009) who have successfully coupled a structured model of mAb synthesis to an unstructured growth model.

3.1. First principles model derivation

The complete list of model equations can be found in Table 1. A material balance for viable cells within the bioreactor is given by (1) where X_u is the concentration of viable cells in the bioreactor measured in cells per litre and μ, μ_d are the specific growth and death rates respectively (h^{-1}). The material balance for the total cell concentration (the sum of both dead and viable cells) is given by (2) where X_t denotes the total cell concentration and is measured in cells l^{-1}. The specific growth rate is estimated through (3) where μ_{max} is the maximum possible growth rate for the specific cell line (h^{-1}) and K_i's are the Monod constants for the primary nutrients, glucose and glutamine. Similarly, KI_i's are the inhibition constants of the primary metabolites, lactate and ammonia. [GLC], [GLN], [LAC] and [AMM] represent extracellular concentrations (mM). The specific death rate (μ_d) is given by (4) where, $\mu_{d,max}$ represents the maximum specific death rate (h^{-1}) and $K_{d,amm}$ describes the rate of cell death by ammonia.

Since the model is unsegregated, it only represents the overall concentrations of nutrients and by-products of cellular metabolism within the bioreactor. The material balance for glucose is given by (5) where parameters $Y_{x,glc}$ and m_{glc} are the cell yield on glucose (cell $mmol^{-1}$) and maintenance energy of glucose (mmol $cell^{-1}$ h^{-1}), respectively. Eq. (5) was originally presented (Jang & Barford, 2000) with an additional term for glucose consumption by glucokinase, which as Kontoravdi (2006) later argued, based on evidence by Tatiraju, Soroush, and Mutharassan (1999) has negligible effects. The material balance for glutamine is similarly described by (6) where $Y_{x,gln}$ and m_{gln} are the cell yield on glutamine (cell $mmol^{-1}$) and maintenance energy of glutamine (mmol $cell^{-1}$ h^{-1}) respectively, with a_1 and a_2 being the constants. The additional term describes glutamine degradation. Glutamine is known to be spontaneously converted into pyrolidonecarboxylic acid at high temperatures and weakly acidic or alkaline solutions (Chibnall & Westall, 1932). Bray, James, Raffan, and Thorpe (1948) 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, which is often the case with media used for mammalian cell cultures. Eq. (6) is presented in the updated version (Kontoravdi, 2006) and not as originally presented (Tatiraju et al., 1999). The mass balances for ammonia and lactate are given by (7) and (8) respectively $Y_{lac,glc}$ and $Y_{amm,gln}$ represent the yields of the particular product on its primary nutrient (mmol of metabolite/mmol of nutrient).

The structured model describing antibody formation and secretion, as presented by Bibila and Flickenger (1992) consists of an intracellular heavy-(9) and light-(10) chain mRNA balance, where m_H and m_L are the intracellular heavy- and light-chain mRNA concentrations (mRNAs $cell^{-1}$), N_H and N are the heavy- and light-chain gene copy numbers (gene $cell^{-1}$), S_H and S_L are the heavy- and light-chain gene specific transcription rates (mRNAs $gene^{-1}$ h^{-1}), 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 given by (11) and (12) respectively, where [H] and [L] are the free heavy and light chain concentrations in the ER (chain $cell^{-1}$), T_H and T_L are the heavy- and light-chain specific translation rates (chain $mRNA^{-1}$ h^{-1}), and R_H and R_L are the rates of heavy- and light-chain consumption in assembly (chain $cell^{-1}$ h^{-1}). MAbs consist of two heavy (H) and two light (L) amino acid chains. Each molecule is synthesised in the ER according to the mechanism shown in (13) (Percy, 1975). Following the assumption presented by Bibila and Flickenger (1992) that the rates of heavy and light chain consumption in the assembly stage are given by (14) an intra-ER balance can be performed for each of the assembly intermediates (15)–(16) where, $[H_2]$, $[H_2L]$ are the concentrations of the assembly intermediates in the ER (molecule $cell^{-1}$), and K_A is the assembly rate constant (molecule $cell^{-1}$ h^{-1}). A balance can then be performed on the assembled mAb structure ($[H_2L_2]_{ER}$) in the ER (17) where $[H_2L_2]_{ER}$ is the mAb concentration in the ER (molecule $cell^{-1}$), 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 glycosylation process place. An intraGolgi mAb balance is given by (18) where $[H_2L_2]_G$ is the mAb concentration in the Golgi (molecule $cell^{-1}$), ε_1 is the ER glycosylation efficiency factor and K_G is the rate constant for Golgi-to-extracellular medium mAb transport (h^{-1}). Finally, the expression for antibody secretion (production) is given in (19) where Q_{MAb} is the specific mAb production rate (mg $cell^{-1}$ h^{-1}), λ is the molecular weight of IgG_1 (146,000 g mol^{-1}), and ε_2 is the Golgi glycosylation efficiency factor. [MAb] denotes the

Table 2
Model parameters: biological significance grouping.

Symbol	Units	Value
1-growth/death related		
μ_{max}	h^{-1}	$5.8 * 10^{-3}$
$K_{I,Amm}$	mM	28.484
$K_{I,Lac}$	mM	171.756
K_{Glc}	mM	0.75
K_{Gln}	mM	0.075
$m_{d,max}$	h^{-1}	0.03
$K_{d,Amm}$	mM	1.759
n	n/a	2
K_{lysis}	h^{-1}	0.05511
2-metabolism related		
$Y_{Lac,Glc}$	n/a	1.399
m_{Glc}	mmol $cell^{-1}$ h^{-1}	$4.853 * 10^{-14}$
$Y_{x,Glc}$	Cell $mmol^{-1}$	$1.061 * 10^{8}$
$Y_{x,Gln}$	Cell $mmol^{-1}$	$5.565 * 10^{8}$
$K_{d,gln}$	h^{-1}	$9.6 * 10^{-3}$
a_1	mM L $cell^{-1}$ h^{-1}	$3.4 * 10^{-13}$
a_2	mM	4
$Y_{Amm,Gln}$	n/a	0.4269
3-mab synthesis related		
K	h^{-1}	0.1
N_H	gene $cell^{-1}$	139.8
S_H	mRNAs $gene^{-1}$ h^{-1}	3000
N_L	gene $cell^{-1}$	117.5
S_L	mRNAs $gene^{-1}$ h^{-1}	4500
T_H	chain $mRNA^{-1}$ h^{-1}	17
T_L	chain $mRNA^{-1}$ h^{-1}	11.5
K_A	molecule $cell^{-1}$ h^{-1}	10^{-6}
K_{ER}	h^{-1}	0.693
K_G	h^{-1}	0.1386
ε_1	n/a	0.995
4-mab secretion related		
γ_1	n/a	0.1
γ_2	n/a	2
ε_2	n/a	1

mAb concentration in the culture (mg l^{-1}), and γ_1, γ_2 are constants.

Eqs. (1)–(19) form a first principles model consisting of a total of 16 differential equations and 31 model parameters. Prior to further analysis initial estimates for the model parameters need to be derived from relevant experimental data. In case the model already exists, parameter values obtained from relevant literature can be utilised. The derivation of estimates for the presented. model's parameters was performed using experimental data of batch hybridoma cultures from the work of Kontoravdi et al. (2010). All parameter estimation experiments and model simulations were implemented in the advanced process modelling environment gPROMS® (Process Systems Enterprise, 1997–2010). Table 2 summarises the list of model parameter estimates while Fig. 2 presents an overview of experimental data and model simulations. The model is in good agreement with the experimental data and successfully captures the trends of nutrient consumption and metabolite accumulation. This indicates a model capable of describing the process under study.

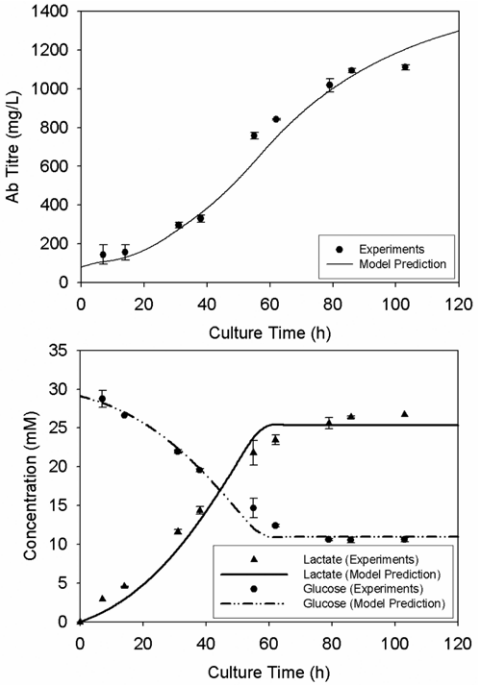

Fig. 2. Experimental data from batcg hybridoma cultures and model predictions. *Source:* Adapted from Kontoravdi et al. (2010).

3.2. Model analysis

Model analysis techniques serve the purpose of increasing the confidence in the chosen structure and resulting output of the mathematical formulation. Application of such techniques prior to experimentation yields valuable information regarding the suitability of the model to describe the studied process.

3.2.1. Parameter identifiability

Parameter identifiability can refer to either structural or numerical identifiability. Structural identifiability studies the form of the model equations and thus requires symbolic manipulation. For relatively simple models where $M + P \leq 10$, where M is the dimensionality of the response vector and P the dimensionality of the parameter vector, identifiability theory is well developed and rigorous structural identifiability can be applied (Walter, 1987). For small scale non-linear models techniques exist that make them amenable to analysis (Walter & Pronzato, 1996). In contrast, very few techniques for arbitrarily large, non-linear models exist (Ljung & Glad, 1994). In these cases, we must resort to numerical identifiability, which studies the numerical behaviour of a model's input–output structure. Asprey and Machietto (2000) have developed an optimisation-based method for global identifiability. A detailed example on the application of this method on models of biological systems is presented in the work of Asprey and Mantalaris (2002). The studied model was a modified version of the unstructured model developed by Jang and Barford (2000). Identifiability was posed as an optimisation problem that sought to maximise the variability of each parameter in the model subject to the output trajectories being invariant. At the end of the optimisation parameters that could take values within a range greater than the predefined tolerance without any effect on the output prediction, were classified as unidentifiable. Thus prior to the performance of any experiments, Asprey and Mantalaris (2002) were able to determine, that the model could not be uniquely identified based on its parameters, and moreover which were the problematic parameters. Campolongo, Tarantola, and Saltelli (2000) mention that the use of global quantitative sensitivity analysis (SA) methods can be used prior to and within the context of parameter identification rendering the two techniques fairly complementary. Therefore in the context of the present study only SA results will be utilised.

3.2.2. Sensitivity analysis

SA allocates model uncertainty to the various sources of uncertainty (i.e. model parameters) facilitating the targeted reduction of output uncertainty by accurate parameter estimation

through tailor-made experiments indicated by a model based design of experiments (DOE) algorithm. On the other hand parameters indicated as insignificant, with respect to the model's output, can be fixed at their literature values (if available) or approximated.

The majority of SA methods met in engineering are derivative based and hence local in nature. SA methods have provided numerous interesting results in a wide variety of fields such as chemical reactions (Meadows, Crowley, Immanuel, & Doyle, 2003; Rabitz, Krammer, & Dacol, 1983; Thomas & Kiparissides, 1984), financial applications (Scaillet, 2003), eco-system biology (Wood & Thomas, 1999) and process systems engineering (Jia & Ierapetritou, 2004). However according to Saltelli, Ratto, Tarantola, and Campolongo (2005) use of derivative or otherwise one-factor-a time techniques requires the model to be linear in all its parameters unless some form of averaging over the parameter space is made. Chan, Saltelli, and Tarantola (1997) illustrate the benefits of using global sensitivity analysis (GSA) methods when studying non-linear systems. In a nutshell, the ability to estimate higher level indices, which quantify the effect on the output of parameter-parameter interactions gives global methods the edge over their local counterparts.

The highly non-linear nature of models of biological systems favours the use of global methods. Herein we have utilised the Sobol Global Sensitivity indices (Sobol', 2001) as the GSA method of choice due to their ability to distinguish between first order sensitivity and non-liner effects. As a result the Sobol' sensitivity indices can be used to exclude parameters with a high level of non-linear interactions from DOE on the basis of singularity. However computational efficiency issues impose the use of parameter grouping for medium and large scale models (Kiparissides et al., 2009). Herein grouping model parameters according to their biological function was the method of choice. The model was simulated for 120 h of batch culture time and SA was performed at three characteristic time points (20, 50 and 120 h). Sensitivity Indices (SI's) change dynamically along the time trajectory of the model output. As the culture progresses and nutrients start being depleted, the model output will become more sensitive towards parameters affecting nutrient uptake and metabolism. SA was conducted at different phases of the cell culture in order to capture the dynamics of the various phases of a batch culture. Specifically time points from the lag, exponential growth and decline phases were evaluated. The simulations involved scanning of all model parameters with respect to the output variables of interest. The uncertainty range associated with the parameters was set to $\pm 100\%$ from the nominal value. The model's parameters were divided into 4 groups, which can be found in Table 2. Fig. 3 presents the results of the GSA, while parameters identified as significant are presented in Table 3. Changes in the value of the SIs can be observed as the culture progresses. Interestingly a change in the ratio Sind/Stot at various time points for a given parameter can be observed. Time points when this ratio is closer to the unit and additionally display a relatively high SI are the optimal time points to extract the value for the said parameter. Groups 1 and 2 generally display similar trends for both output variables. The parameters of group 3 have either small SIs or high Stot but low Sind values. This indicates a high degree of uncertainty in the structured part of the model. Alas, even if a parameter can mathematically guarantee unique identification, it may still not be feasible to conduct the necessary experimental measurements that would allow the precise estimation of its value. Glycosylation efficiency measurements for example are rather complex and cumbersome and require equipment not readily available in every analytical laboratory. Therefore, whilst aware of the uncertainty associated with the parameters (ε_1, ε_2) of groups 3 and 4, we have no choice but to omit them from the DOE algorithm. Similarly, parameter (γ_2) is closely linked to the cells' position in the cell cycle, making its experimental estimation particularly difficult. The difficulty of obtaining experimental measurements for certain parameters is a common "real life" problem. The fact that a parameter cannot be experimentally estimated does not invalidate the sensitivity analysis and on the contrary raises the awareness of the modeller to possible weaknesses of the developed model.

Table 3
Summary of GSA results.

Group 1	Group 2	Group 3	Group 4
μ_{max}	$Y_{x,glc}$	ε_1	γ_2
$\mu_{d,max}$	$Y_{x,gln}$		ε_2
	$Y_{amm,gln}$		
	$Y_{lac,glc}$		

The use of SA effectively reduced the number of parameters that need to be experimentally validated from a total of 31 to a mere 6. Moreover, we have gained valuable information regarding the time points that would yield the most informative experiments.

3.3. Design of experiments and model validation

Fed-batch operation is the most common choice for industrial applications as it can prolong culture longevity and thus the final titre. Consequently, the aim is to extend the model's predictive capabilities to fed-batch conditions. Instead of re-estimating the whole set of model parameters in order to account for the different metabolic characteristics of fed-batch cultivation, the accurate estimation of only the significant parameters is targeted through D-optimal model-based experimental design (Kontoravdi et al., 2010; Process Systems Enterprise, 1997–2010). The latter dictates the appropriate feeding strategy for a fed-batch experiment so that the information content of the collected data is maximised. The data of the designed culture are used to validate the model under fed-batch conditions. As previous studies have discussed (Munack & Posten, 1989; Versyck, Claes, & Impe, 1997), DOE uses the model to design sufficiently informative experiments for this purpose. The most significant parameters (Table 3) can be readily input to the DOE algorithm. The use of a CAD tool, such as gPROMS, where a DOE utility is already implemented, significantly simplifies this step.

The concentrations of glucose and glutamine in the feed were set at 500 mM and 100 mM, respectively. The maximum total volume of feed was fixed at 8.75 ml, which represents nearly 5% of the total culture volume (200 ml), so as to avoid dilution effects. Sampling times, at which measurements were conducted, were determined *a priori*. The output of the algorithm provided the optimal amount of feed supplied at each feeding interval as well as the optimal timing of the intervals and the optimal duration of the experiment (168 h) (Kontoravdi et al., 2010) and enabled the re-estimation of the significant parameters (Table 4).

Fresh media was added to the culture at predefined volumes of 1.25 mL every 24 h starting from 12 h of culture time. Sampling initiated after the first addition of fresh media. Additional samples were taken 0, 6, 12 and 24 h after the addition of media. The "refined" version of the model is simulated for fed-batch operation and is plotted against relevant experimental data (Kontoravdi et al., 2010). The model is found to be in good agreement with the experimental data and can successfully capture the dynamics of a fed-batch culture as shown in Fig. 4. Moreover the model's performance while using the parameter values of Table 4 is compared to the performance while using the parameter values of Table 2. Quality of fit calculations for the

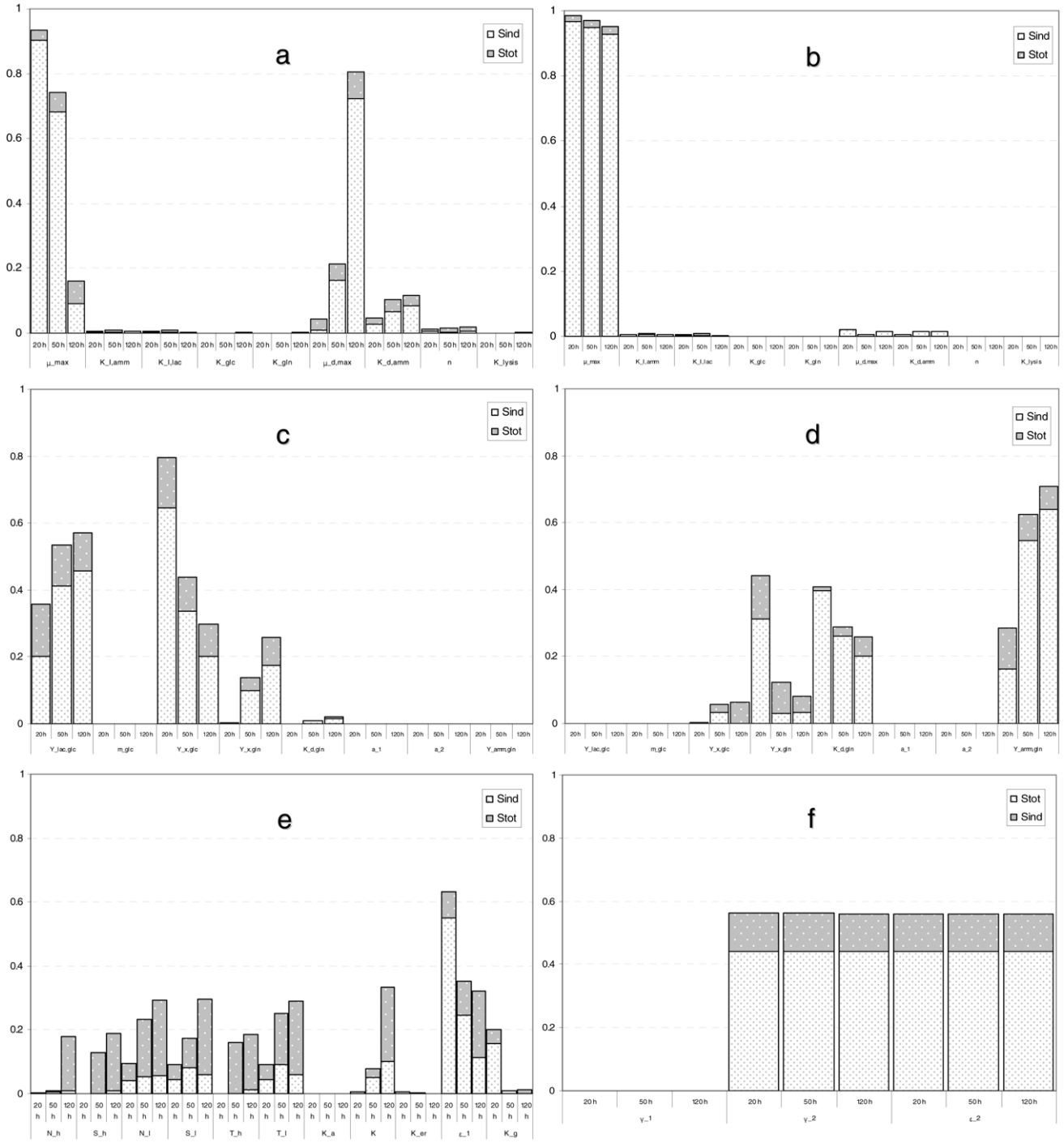

Fig. 3. GSA results; (a): SIs for parameters of group 1 with respect to viable cell concentration; (b): SIs for parameters of group 1 with respect to glucose concentration; (c): SIs for parameters of group 2 with respect to Lactate concentration; (d): SIs for parameters of group 2 with respect to Ammonia concentration; (e): SIs for parameters of group 3 with respect to mAb concentration; (f): SIs for parameters of group 4 with respect to mAb concentration.

predictions shown in Fig. 4 are shown in Table 5. Finally the model's range of applicability should be examined against an independent set of experimental data as shown in Kontoravdi (2006) and Kontoravdi et al. (2010).

3.4. Model based optimisation & control

From a process engineering point of view the ultimate goal of any modelling attempt is to "close the loop" through model-based optimisation. Several studies presented in the literature showcase the potential of the application model-based optimisation and control strategies on bioprocesses (De Tremblay et al., 1992; De Tremblay, Perrier, Chavarie, & Achambault, 1993; Frahm et al., 2002; Frahm, Lane, Märkl, & Pörtner, 2003; Zhou, Rehm, Europa, & Hu, 1997). However, a holistic approach that facilitates optimisation of the *in vitro* through the *in silico* is still lacking. Consequently, the response of the culture to changes in the feeding strategy is usually monitored implicitly through the oxygen uptake rate and the culture pH. The limited number of readily available online measurements in turn limits the complexity of the utilised

Table 4
Results of the parameter estimation algorithm.

Parameter	New Value	Confidence intervals[a]		
		90%	95%	99%
μ_{max}	0.054305	1.001×10^{-3}	1.194×10^{-3}	1.572×10^{-3}
$\mu_{d,max}$	0.031519	1.801×10^{-3}	2.149×10^{-3}	2.829×10^{-3}
$Y_{x,glc}$	2.6×10^{8}	2.275×10^{7}	2.715×10^{7}	3.494×10^{7}
$Y_{x,gln}$	8×10^{8}	3.203×10^{7}	3.822×10^{7}	3.575×10^{7}
$Y_{amm,gln}$	0.381159	2.393×10^{-3}	2.856×10^{-3}	5.032×10^{-3}
$Y_{lac,glc}$	2.05356	1.751×10^{-1}	2.089×10^{-1}	2.751×10^{-1}

[a] The confidence ellipsoid is a linear approximation of the non-linear confidence region (Process Systems Enterprise, 1997–2010).

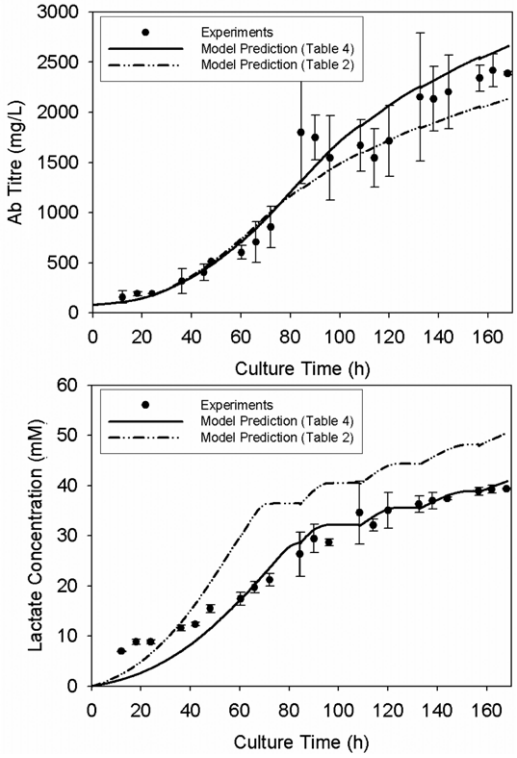

Fig. 4. Fed-batch cultures of HFN 7.1 hybridma cells.
Source: Adapted from Kontoravdi et al. (2010).

Table 5
Quality of fit.

Variable	R^2 using parameters from Table 4	R^2 using parameters from Table 2
[mAb]	0.93177	0.92328
[LAC]	0.93026	0.38679

model. Therefore, the common practice when estimating optimal feeding profiles is to base calculations on the cells' need for glucose (and/or glutamine) alone.

Lam (2009) following the model development algorithm of Fig. 1, performed model-based optimisation studies on CHO-IFNγ cultures. The resulting optimal feeding profile is shown in Fig. 5(b). These results indicate a potential for the fed-batch culture to produce IFNγ at approximately 60 mg L^{-1} at the end of the culture time whereas the highest yield achieved experimentally (Fig. 5(a)) was about 35 mg L^{-1}. The higher product yield in the optimised result was achieved due to a low concentration of toxic ammonium as a result of a better controlled supply of nutrients. Although the model utilised in that study (Lam, 2009) is different than the one presented thus far, the modelling steps were identical.

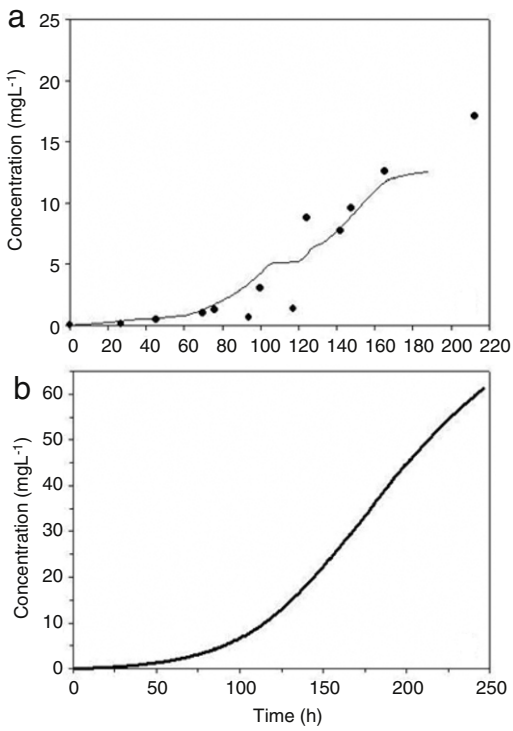

Fig. 5. IFNγ concentration profile of CHO-IFNγ fed-batch culture (a) Simulation (solid line) and experiment data (dots). (b) Optimal profile as a result of dynamic optimisation.
Source: Adapted from Lam (2009).

Consequently, the results of Fig. 5 represent the logical evolution of the algorithm and ultimately 'close the loop' by optimisation of the *in vitro* (Fig. 5) through *in silico* experimentation. Details on the formulation and solution of the optimisation problem can be found in Lam (2009).

4. Concluding remarks

Interestingly, there appear to be many differences between the mathematical models that describe mammalian cell growth. Cell growth has been mathematically related to glucose concentration alone (Frahm et al., 2002), glucose and glutamine (De Tremblay et al., 1992), glucose and lactate (Kurokawa, Park, Iijima, & Kobayashi, 1994) and to all four nutrients and metabolites (Jang & Barford, 2000). Similarly cell death has been related to glutamine, lactate and ammonia (De Tremblay et al., 1992), glucose (Frame & Hu, 1991), glutamine (Dalili, Sayles, & Ollis, 1990), ammonia and lactate (Batt & Kompala, 1989), or ammonia (Jang & Barford, 2000). This apparent lack of consistency, illustrates the need for a framework that formalises model development.

Optimisation of secreting cultures depends on the balance between prolonged culture viability and increased productivity. Alas, conditions that seem to prolong culture viability usually reduce productivity and vice versa. Optimal feeding profiles based on the provision of glucose and glutamine alone might yield an increase in the final product titre without being the global optimum. Moreover, the work of Xie and Wang (1994) illustrated that excessive feeding of glucose is not the best means towards higher product titres, since it shifts metabolism towards energy inefficient pathways. Energy metabolism is a significant element of cell culture that has thus far been ignored from a modelling point of view. Furthermore, deZengotita et al. (2000) has shown that apart from the primary nutrients, other components might be limiting growth. A truly optimal feeding

profile should be derived based on the provision of adequate yet not excessive amounts of energy through controlled quantities of nutrients.

Bailey (1998) predicted the need to shift modelling focus upstream towards the genetic level where the kernel of the cell's control mechanism lies however little work has been done since. The advancements in analytical and theoretical biology will increasingly provide more information in the future, especially with the increasing popularity and availability of the — omics techniques. Paving the way towards a 'closed-loop' approach for bioprocess automation, the work presented herein presents a biological model development framework (Fig. 1) in a step-by-step fashion, highlighting challenges and "real life" problems associated with each stage of model development. By organising available information in a systematic way, unnecessary experimentation is avoided and models with *a priori* established aims able to guide the *in vitro* through the *in silico* are developed.

References

Asprey, S. P., & Machietto, S. (2000). Statistical tools for optimal dynamic model building. *Computers and Chemical Engineering*, *24*, 1261–1267.

Asprey, S. P., & Mantalaris, A. (2002). Global parametric identifiability of a dynamic unstructured model of hybridoma cell culture. In *8th IFAC international conference on computer applications in biotechnology, CAB8* (pp. 25–30).

Bailey, J. E (1998). Mathematical modeling and analysis in biochemical engineering: past accomplishments and future opportunities. *Biotechnology Progress*, *14*, 8–20.

Batt, B. C., & Kompala, D. S. (1989). A structured kinetic modelling framework for the dynamics of hybridoma growth and monoclonal antibody production in continuous suspension cultures. *Biotechnology and Bioengineering*, *34*, 515–531.

Bibila, T. A., & Flickenger, M. C. (1992). Use of a structured kinetic model of antibody synthesis and secretion for the optimisation of antibody product synthesis part 1. Steady-state analysis. *Biotechnology and Bioengineering*, *39*, 251–261.

Bibila, T. A., & Robinson, D. K. (1995). In pursuit of the optimal fed-batch process for monoclonal antibody production. *Biotechnology Progress*, *11*, 1–13.

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

Campolongo, F., Tarantola, S., & Saltelli, A. (2000). Sensitivity analysis as an igredient of modelling. *Statistical Science*, *v4*(15).

Chan, K., Saltelli, A., & Tarantola, S. (1997). Sensitivity analysis of the model output: variance-based methods make the difference. In *Proceedings of the 1997 winter simulation conference*.

Chibnall, A. C., & Westall, R. G. (1932). The estimation of glutamine in the presence of aspargine. *Biochemical Journal*, *26*(1).

Covert, M. W., Schilling, C. H., Famili, I., Edwards, J. S., Goryanin, I. I., Selkov, E., & Palsson, B. O. (2001). Metabolic modeling of microbial strains in silico. *Trends in Biochemical Science*, *26*, 179–186.

Dalili, M., Sayles, G. D., & Ollis, D. F. (1990). Glutamine-limited vatch hybridoma cell growth and antibody production: experiment and model. *Biotechnology and Bioengineering*, *36*, 74–82.

De Tremblay, M., Perrier, M., Chavarie, C., & Archambault, J. (1992). Optimization of fed-batch culture of hybridoma cells using dynamic programming: single and multi feed cases. *Bioprocess and Biosystems Engineering*, *7*, 229–234.

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

deZengotita, M. V., Miller, W. M., & Aunins, J. G. (2000). Phosphate feeding improves high-cell-concentration NS0 myeloma culture performance for monoclonal antibody production. *Biotechnology and Bioengineering*, *69*, 566–676.

Dhir, S., Morrow, K. J., Rhinehart, R. R., & Wiesner, T. (1999). Dynamic optimisation of hybridoma growth in a fed-batch bioreactor. *Biotechnology and Bioengineering*, *67*(2), 197–205.

Frahm, B., Lane, P., Atzert, H., Munack, A., Hoffmann, M., Hass, V. C., & Pörtner, R. (2002). Adaptive, model-based control by the open-loop-feedback-optimal (OLFO) controller for the effective fed-batch cultivation of hybridoma cells. *Biotechnology Progress*, *18*, 1095–1103.

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

Frame, K. K., & Hu, W. S. (1991). Kinetic study of hybridoma cell growth in continuous culture: I. A model for nonproducing cells. *Biotechnology and Bioengineering*, *37*, 55–64.

Ho, Y., Varley, J., & Mantalaris, A. (2006). Development and analysis of a mathematical model for antibody-producing GS-NS0 cells under normal and hyperosmotic culture conditions. *Biotechnology Progress*, *22*(6), 1560–1569.

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

Jia, Z., & Ierapetritou, M. G. (2004). Short-term scheduling under uncertainty using MILP sensitivity analysis. *Industrial & Engineering Chemistry Research*, *43*(14), 3782–3791.

Kiparissides, A., Kucherenko, S., Mantalaris, A., & Pistikopoulos, E. N. (2009). Global sensitivity analysis challenges in biological systems modelling. *Industrial & Engineering Chemistry Research*, *48*(15), 7168–7180.

Kontoravdi, C. (2006). 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.

Kontoravdi, C., Asprey, S. P., Pistikopoulos, E. N., & Mantalaris, A. (2007). Development of a dynamic model of monoclonal antibody production and glycosylation for product quality monitoring. *Computers & Chemical Engineering*, *31*, 392–400.

Kontoravdi, C., Pistikopoulos, E. N., & Mantalaris, A. (2010). Systematic development of predictive mathematical models for animal cell cultures. *Computers & Chemical Engineering*, *34*(8), 1192–1198.

Kurokawa, H., Park, Y. S., Iijima, S., & Kobayashi, T. (1994). Growth characteristics in fed-batch culture of hybridoma cells with control of glucose and glutamine concentration. *Biotechnology and Bioengineering*, *44*, 95–103.

Lam, M. C. (2009). In silico dynamic optimisation studies for batch/fed-batch mammalian cell suspension cultures producing biopharmaceuticals. *Ph.D. Thesis*. Department of Chemical Engineering and Chemical Technology Imperial College, London.

Lam, M. C., Sriyudthsak, K., Kontoravdi, C., Kothari, K., Park, H. H., Pistikopoulos, E. N., & Mantalaris, A. (2008). Cell cycle modelling for off-line dynamic optimisation of mammalian culture. *Computer Aided Chemical Engineering*, *25*, 109–114.

Ljung, L., & Glad, T. (1994). On global identifiability for arbritary model parameterizations. *Automatica*, *30*, 265–276.

Meadows, E. S., Crowley, T. J., Immanuel, C. D., & Doyle, F. J., III (2003). Nonisothermal modeling and sensitivity studies for batch and semibatch emulsion polymerization of styrene. *Industrial & Engineering Chemistry Research*, *42*(3), 555–567.

Munack, A., & Posten, C. (1989). Design of optimal dynamical experiments for parameter estimation. In *Proceedings of the American control conference ACC89*. Pittsburgh, PA, 2011-2016.

Percy, J. R. (1975). A theoretical model for the covalent assembly of immunoglobulins. *Journal of Biological Chemistry*, *250*(6), 2398–2400.

Pörtner, R., & Schäfer, T. (1996). Modelling hybridoma cell growth and metabolism — a comparison of selected models and data. *Journal of Biotechnology*, *49*, 119–135.

Process Systems Enterprise (1997–2010). *gPROMS*. www.psenterprise.com/gproms.

Rabitz, H., Krammer, M., & Dacol, D. (1983). Sensitivity analysis in chemical kinetics. *Annals of Review Physical Chemistry*, *34*, 419–461.

Saltelli, A., Ratto, M., Tarantola, S., & Campolongo, F. (2005). Sensitivity analysis for chemical models. *Chemical Review*, *105*, 2811–2827.

San, K. Y., & Stephanopoulos, G. (1989). Optimization of fed-batch penicillin fermentation: a case of singular optimal control with state constraints. *Biotechnology and Bioengineering*, *34*(1), 72–78.

Scaillet, O. (2003). Nonparametric estimation and sensitivity analysis of expected shortfall. *Mathematical Finance*, *14*, 115–129.

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

Sidoli, F. R., Mantalaris, A., & Asprey, S. P. (2005). Toward global parametric estimability of a large-scale kinetic single-cell model for mammalian cell culture. *Industrial & Engineering Chemistry Research*, *44*(4), 868–878.

Sobol', I. M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulations*, *55*, 271–280.

Tatiraju, S., Soroush, M., & Mutharassan, R. (1999). Multi-rate nonlinear state and parameter estimation in a bioreactor. *Biotechnology and Bioengineering*, *63*(1), 22.

Thilakavathi, M., Basak, T., & Panda, T. (2007). Modeling of enzyme production kinetics. *Applied Microbiology and Biotechnology*, *73*, 991–1007.

Thomas, I. M., & Kiparissides, C. (1984). Sensitivity analysis of a batch polymerization reactor. *Journal of Applied Polymer Science*, *29*, 2195–2204.

Versyck, K. J., Claes, J. E., & Impe, J. F. V. (1997). Practical identifcation of unstructured growth kinetics by application of optimal experimental design. *Biotechnology Progress*, *13*, 524–531.

Walter, E. (Ed.) (1987). *Identifiability of parametric models*. Pergamon Press.

Walter, E., & Pronzato, L. (1996). On the identifiability and distinguishability of non-linear parametric models. *Mathematics and Computers in Simulation*, *42*, 125–134.

Wood, S. N., & Thomas, M. B. (1999). Super-sensitivity to structure in biological models. *Proceedings of the Royal Society of London B*, *266*, 565–570.

Xie, L., & Wang, D. I. C. (1994). Applications of improved stoichiometric model in medium design and fed-batch cultivation of animal cells in bioreactor. *Growth Factors*, *1*, 17–29.

Yokobayashi, Y., Collins, C., Leadbetter, J. R., Weiss, R, & Arnold, F. H. (2003). Evolutionary design of genetic circuits and cell–cell communications. *Advances in Complex Systems*, *6*(1).

Zhou, W., Rehm, J., Europa, A., & Hu, W. S. (1997). Alteration of mammalian cell metabolism by dynamic nutrient feeding. *Cytotechnology*, *24*, 99–108.

Alexandros Kiparissides received his Diploma in Chemical Engineering from the Aristotle University of Thessaloniki, Greece in 2006. He is currently a Ph.D. candidate in the Biological Systems Engineering Laboratory in the department of Chemical Engineering at Imperial College, London. His main research interests lie in the coupling of mathematical and experimental techniques towards the optimisation of cell culture processes.

Michalis Koutinas graduated from the University of Patras, Greece in 2001, with a Diploma in Chemical Engineering. In 2002 he joined Imperial College London, UK where he obtained his Ph.D. in Chemical Engineering in 2006. He is currently a Post-Doctoral Research Associate at Imperial College London. His research interests are in the areas of biological systems modelling, fermentation technology, and the application of molecular biology methods in the field of biochemical engineering.

Cleo Kontoravdi received her M.Eng. and Ph.D. degrees in Chemical Engineering from Imperial College London. She was then employed as a Research and Development Scientist at Lonza Biologics where she was responsible for cell line development projects, before rejoining Imperial College London as a Lecturer in Biopharmaceuticals Processing. Her research interests include experimental and model-based analysis of mammalian cell metabolism and protein glycosylation.

Athanasios Mantalaris is a Professor of BioSystems Engineering heading the Biological Systems Engineering Laboratory (BSEL) in the Centre for Process Systems Engineering in the Department of Chemical Engineering at Imperial College London. He received his B.Sc. (Hons) in Biochemistry from the University of Western Ontario, Canada in 1991. He received his MSc and Ph.D. in Chemical Engineering from the University of Rochester, USA. In 2000 he was appointed as "Governor's Lecturer" at Imperial College, London before becoming a Professor in Chemical Enginerring (2010). His research interests are in bioprocess focusing on applied microbial, mammalian, and stem cell bioprocessing.

Efstratios N. Pistikopoulos, Director of the Centre for Process Systems Engineering (CPSE) from 2002 to September 2009, is a Professor of Chemical Engineering, at Imperial College London. He obtained a Diploma from the Aristotle University of Thessaloniki in 1984, a Ph.D. from Carnegie Mellon University, USA in 1988, and was with Shell Chemicals (Amsterdam, the Netherlands) before joining Imperial in 1991. His research interests include the development of theory, algorithms and computational tools for multi-parametric programming and explicit model predictive control, and their applications in biomedical, energy and process systems.