1 Model Development and analysis of mammalian cell culture systems

Alexandros Kiparissides, Michalis Koutinas, Efstratios N. Pistikopoulos and Athanasios Mantalaris

Biological Systems Engineering Laboratory,
Centre for Process Systems Engineering,
Imperial College London

1.1 Introduction

The advancements in molecular biology and analytical techniques over the last century have significantly elevated the biological industry in the economical scale. Monoclonal antibodies (MAb) alone have a projected market of 49bn$ by 2013, according to "Monoclonal antibodies Report, 2007". Considering that MAb industry is a mere fraction of the applications that utilise mammalian cell culture systems, one can appreciate the size of the biologics industry. However as initially Bailey (1998) and later Sidoli *et al.* (2004) argue the development of mathematically and computationally orientated research has failed to catch up with the recent developments in biology. Moreover, the little credit that mathematical modelling of biological systems receives from experimentalists may be the offspring of the lack of effective communication of the benefits of making a mathematical model (Bailey 1998).

In all aspects of science where modelling is involved, the first step, before making the model, is to determine its use and define *a priori* the problem the said model intends to address. Even the simplest possible bacterial strain, or the most exhaustively studied for that matter, is a complex network of a myriad of interconnected processes occurring on diverse time scales within a confined volume. To add to the complexity, the cell regulates its activities on multiple levels, deploying an elaborate control network, which to a large extent still remains grey territory. Moreover when cells grow in the neighbourhood of other cells, an intricate communication network of signals and interactions mediates the macroscopic behaviour of the culture. Therefore any attempt to elaborately model the function of even a single cell will undoubtedly fall short for numerous reasons. First of all, the amount of delicate intracellular measurements required to validate such a model is exhaustive both in terms of labour as well as

cost, not to mention parameter identifiability issues (Sidoli *et al.* 2004). Furthermore, not all aspects of biology are thoroughly understood, while many have been studied under very specific conditions. Therefore one will reasonably wonder what role mathematical modelling can play when studying biological systems.

Borrowing research principles from the Chemical and Process Engineering paradigms, mathematical modelling of biological systems can provide a systematic means to quantitatively study the characteristics of the complex and multilevel interactions that occur in cell bioprocessing. In a way, it can be viewed as an effective way to organise in a meaningful way the vast plethora of available biological information. Mathematical models have successfully been used to design optimal media (Xie and Wang 1993), identify previously ignored growth limiting factors (deZengotita *et al.* 2000), optimise culture growth and productivity (deTremblay *et al.* 1992, San and Stephanopoulos 1998, Dhir *et al.* 1999) and apply control principles to cell culture processes (Frahm *et al.* 2002). Thus the potential of modelling as a scientific and engineering tool, has proven its worth, however in order to maximise the gains from the ever increasing influx of information from biology, especially with the development of the -omics techniques, our view of modelling needs to be shifted towards a closed-loop framework from conception to optimisation.

The way biochemical engineers conceive of and mathematically describe biological processes is still defined by the framework presented by Tsuchiya *et al.* (1966) and Fredrickson *et al.* (1970). According to the framework, shown in Figure (1), a model can be *structured* (or unstructured) and/or *segregated* (or unsegregated). Most models fall in one of the four subcategories formed, depending on whether they possess one two or none of the above properties. Structured models enable the identification of a cell by assigning structure to it. The use of the term structure does not necessarily refer solely to the physical meaning of the term. Moreover besides physical structure which can be incorporated through the creation of intracellular compartments representing the various organelles, it can be incorporated by distinction amongst the various biochemical species hence giving rise to biochemical structure. Therefore a structured metabolic model would consist of the reactions of at least two intracellular species. On the other hand if cells are treated individually, so that they differ from each other in some distinct way we have a segregated model. If however the entire population is treated as a sum of averaged cellular behaviour, the model is unsegregated. In other words, segregated models can account for cells in different cycle phases and generally depict the inherent heterogeneity of a cell population, whereas unsegregated models describe a homogeneous culture composed by a number of “average” cells.

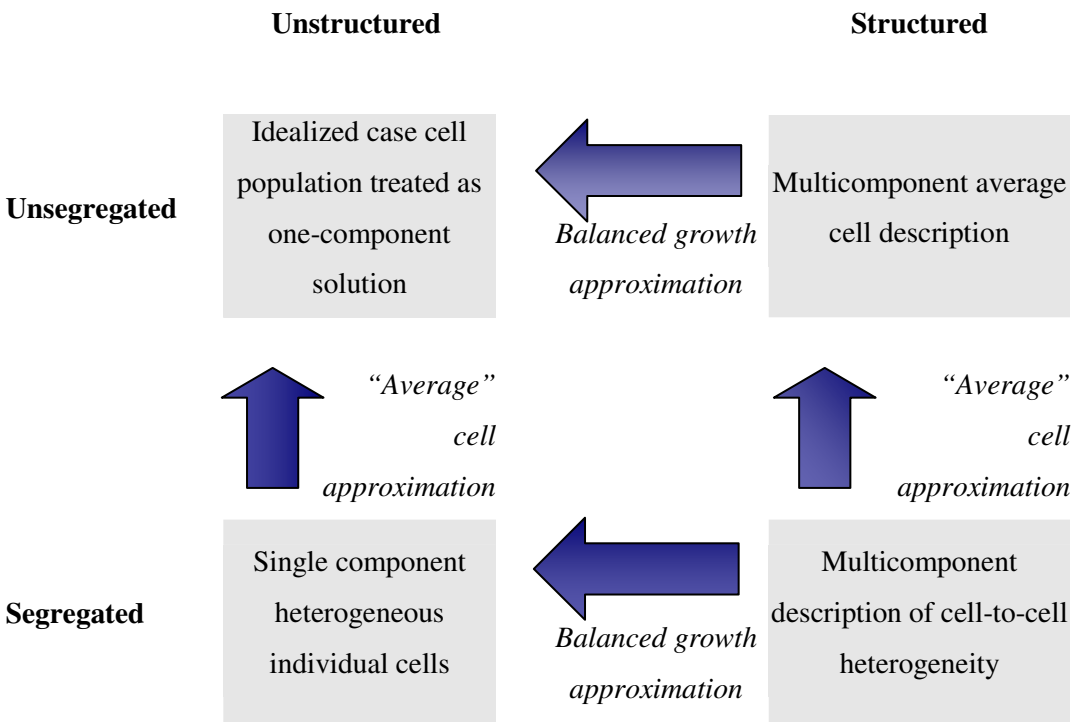


Figure 1 Classification of biological models according to A.G. Fredrickson (1970)

Discriminations that exist for mathematical models in general, such as stochastic or deterministic, static or dynamic, are also applicable to biological models. Stochastic models account for the uncertainty inherent in all systems and implement this through some probabilistic based variation of the input variables. Deterministic modeling is usually based on experimental observations, accounting in a straight forward manner for the most common behavior of the system under observation. The drawback of deterministic modelling is that it can not account for any possible set of inputs but only for the most probable ones. Dynamic models observe the evolution of the modelled system over a predetermined time horizon, whereas static models focus on a specific instance of the population. Dynamic models usually consist of systems of differential algebraic equations (DAE) and are computationally more demanding than static models which usually contain algebraic equations and can be used for more detailed modelling of a system while remaining tractable.

A typical example of the above can be found in metabolic models, which can be either stoichiometric or kinetic. A kinetic model is represented usually by a set of DAEs, which are integrated over a time domain of interest and result in well defined time trajectories of all the variables involved. A key drawback of kinetic modelling is that their additional predictive capability is associated with the incorporation of complex dynamic expressions which usually

result in non-linearities both in the parameters and the variables. On the other hand stoichiometric models are represented by a system of flux balance equations based on reaction stoichiometry of a metabolic network with accompanying constraints on flux values and solved as a constrained optimisation problem using some *a priori* assumed cellular objective. A key advantage of stoichiometric modelling is that it can take into account competing reactions, which enables the study of the relative activity of certain pathways under different culture conditions. However the main drawback of stoichiometric models is that they are not dynamic hence they can not provide information on the temporal evolution of the variables under study (Sidoli *et al.* 2004).

Two extremes exist on the dimension scale on which cell models are considered. The first one is the Single Cell Model (SCM) approach, first presented by Shuler *et al.* (1980), according to which a single cell is modelled exhaustively incorporating as much information as there is available. SCM models are detailed descriptions of the functions occurring within a single cell ignoring any interactions with other cells. The more holistic modelling counterpart to single cell modelling is Population Balance modelling (PB), where multiple populations with varying parameters can be studied. This type of modeling can account for cells being in different phases of the cell cycle and therefore displaying different behavior and different protein production rates. In the core of every population balance model, lies a simple model which describes cell metabolism, growth kinetics and when division occurs. The drawback of PB models is that they tend to be computationally demanding and include large numbers of parameters.

Notable studies that have shaped new sub-categories of biological systems modelling include, but are not limited to, cybernetic modelling presented by Ramkrishna (1982) and the introduction of structure as defined by Fredrickson (1970) at the genetic level by Lee and Bailey (1984a and b). In brief, the concept behind cybernetic modelling is the adaptation of a mathematically simple description of a complex organism which is compensated for the over-simplification by assigning an optimal control motive to its response (Kompala *et al.* 1984). For example, microbial cells growing in the presence of multiple substrates are assumed to be following an invariant strategy to optimise a certain goal by choosing which substrate to consume first. So assuming for example, a multisubstrate environment containing cells that follow different strategies of substrate consumption, cells that will, somehow, choose to grow first on the fastest substrate available will proliferate much faster than cells that respond differently. After some time all the cells that remain in the environment will be those that have responded in the optimal manner (Kompala *et al.* 1984). It is therefore reasonable to assume that

over the many years of evolution cells have acquired the ability to respond optimally to environmental conditions.

Lee and Bailey (1984a and b), extended the concept of structure as presented above to the level of nucleotide sequences. Lee introduced an explicit connection between a particular nucleotide sequence and the affinity of a particular protein for that sequence which in turn will influence the corresponding transcription event, thus deriving a quantitative mapping from nucleotide sequence to overall phenotype. Even though in his detailed review, Bailey (1998), predicted that this new "genetically structured model" would be widely embraced in the future, supported by the advancement of the omics techniques, little work has yet been done in that direction.

1.2 Review of mathematical models of mammalian cell culture systems

Mathematical biology, and biotechnology for that matter, can be subdivided in two broad categories depending on the type of cells studied. The oldest and more exhaustively studied category deals with microbial systems mostly and prokaryotic cells in general. On the other hand we have the more recently emerged field of eukaryotic (or mammalian) cell modelling. In light of recent advancements, both scientific and regulatory, it is worthwhile mentioning the prominent advancement of a third category, namely that of stem cell modelling. Even though all three types of cells share a lot of common elements with respect to their core metabolism, they have distinct differences in their behaviour in culture necessitating their study within these broad categories. Undoubtedly developments and research in prokaryotic cell modelling lead by a fair margin the respective developments in both mammalian and stem cell modelling. This can be attributed to a number of reasons, the main being that prokaryotic cells have to a certain degree simpler metabolic characteristics. 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 book chapter. Therefore we will attempt to review contributions that have either shaped or can successfully highlight a new way of approaching dynamical modelling of mammalian systems.

The earliest reference and possibly the most significant one is the mathematical formulation that describes enzyme kinetics, presented by Michaelis and Menten (1913). Although the hypothetical system studied was the simplest possible, the conversion of one molecule of a given substrate to a product via a single enzymatic reaction, in many ways it shaped the way we conceive of kinetic rates in biology. Since then the theory provided by Michaelis and Menten has

evolved, now being used as a starting point when attempting to describe much more complex enzyme kinetics. Around the same time that Michaelis and Menten presented their work, Archibald Vivian Hill in his effort to describe the sigmoidal binding curve of oxygen to Hemoglobin, derives what is now know as the Hill function (1910). In essence the Hill function describes the binding of a given ligand to a macromolecule when the latter is already saturated with ligands. Finally in 1948 Jacques Monod, presented a function identical to the Michaelis-Menten rate equation which successfully described microbial growth. The basic concept behind Monod's work was that the kinetics observed in every metabolic pathway are largely shaped by its rate limiting step, ultimately an enzyme catalysed reaction. All these kinetic equations are summarised in Table 1.1.

Table 1.1 Enzyme and Microbial growth kinetic expressions

Name	Expression	Function
Michaelis - Menten	$v_0 = \frac{V_{MAX}[S]}{K_M + [S]}$	Describes the kinetics of the simple enzyme catalysed reaction: $E + S \xleftrightarrow{k_1/k_{-1}} ES \xrightarrow{k_2} P$
Hill	$\theta = \frac{[L]^n}{(K_A)^n + [L]^n}$	Describes the fraction of the macromolecule saturated by ligand as a function of the ligand concentration.
Monod	$\mu = \mu_{MAX} \frac{[S]}{K_S + [S]}$	Describes microbial growth based on the consumption of one substrate.

Shifting our focus towards the area of mammalian cell metabolism, most mathematical models examine glucose and glutamine as the primary nutrients and lactate and ammonia as the main metabolites. A typical layout of an unstructured model for cell metabolism consists of mass balances on glucose, glutamine, ammonia and lactate around the bioreactor. These account for the uptake of glucose and glutamine from viable cells for cell growth, as well as glucose consumption by glucokinase, glucose maintenance energy and spontaneous degradation of glutamine in the medium. Lactate and ammonia production are described as functions of glucose and glutamine consumption, respectively (Jang and Barford, 2000). Monod-type kinetics are

used for most metabolic models (Jang and Barford, 2000; Tziampazis and Sambanis, 1994; Tatiraju et al., 1999).

One of the first attempts at developing a structured model for mammalian cells was that of Batt and Kompala (1988) who adapt ideas presented by Fredrickson (1970) and Shuler (1979) to mammalian cell culture systems. Cell mass is divided in four intracellular metabolic pools accounting for amino acids, nucleotides proteins and lipids. These are derived from the extracellular substrates, glucose glutamine and amino acids found in the culture media, while the secreted products include lactate, ammonia and monoclonal antibody. Borrowing experimental data from the extensive work of Miller *et al.* (1986), Batt and Kompala show that the model successfully describes experimental data but more importantly can be used in order to study the effects of various feeding strategies.

Bibila and Flickinger (1991) presented one of the most significant structured models describing MAb synthesis in hybridoma cells. Based on the mechanism proposed by Percy *et al.* (1975) for the covalent assembly of monoclonal antibodies, the authors present in detail the derivation of the structured model that successfully describes experimental data of monoclonal antibody synthesis and secretion. On subsequent studies (1991a and b) they move on to use the proposed model for both steady state and dynamic optimisation of the culture conditions, suggesting strategies that increase final antibody titre. Moreover they perform a parameter sensitivity analysis (1991a), through factorial design, in the steady state version of their model and draw conclusions on the parameters that affect antibody secretion positively. Finally they suggest the assembly step of antibodies within the Endoplasmic reticulum (ER) as the most probable candidate for a rate limiting step of the secretion process, based on perturbation studies conducted with their model (1991b).

Xie and Wang (1993) presented a detailed stoichiometric model for animal cell growth and utilise it to optimise culture media composition. Their stoichiometric analysis covers various aspects of cellular metabolism including energy requirements, lipid, carbohydrate, nucleotide and protein synthesis. Moreover they provide formulae for the derivation of stoichiometric coefficients both for nutrients and products by studying their roles in animal cell metabolism. Later work (1996a and b) by the same authors has provided valuable insight on mammalian cell metabolism. Utilising the devised model (1993), the authors reach a number of valuable conclusions including the necessity to control glucose feed at low concentrations in order to shift mammalian cell metabolism towards more energy efficient pathways. Finally (1996b) they were amongst the first to exhaustively study energy metabolism in mammalian cell culture systems by studying the stoichiometry of the simplified metabolic reaction network they devised.

De Tremblay and co-workers (1992) showcase the potential of dynamic programming for the optimisation of fed-batch hybridoma cultures. Having verified the applicability of dynamic programming, in 1993 they went one step forward and examined the benefits of using and optimal control approach versus a closed loop strategy on fed-batch hybridoma cultures, also presenting experimental data to support their results. Frahm and co-workers (2002a and b) presented a novel open-loop-feedback-optimal controller for the fed-batch cultivation of hybridomas. The utilised unstructured model accounts for monoclonal antibody production and culture growth based on the consumption of glucose and glutamine and the production of lactate and ammonia as basic by-products of metabolism.

DiMasi *et al.* (1995), present a mechanistic structured kinetic model of mammalian cell culture dynamics. The developed model specifically addresses the dynamics of substrate consumption and energy metabolism in mammalian cell culture. Borrowing experimental data from Miller *et al.* (1987) the authors compare their model to the unstructured model of Batt and Kompala (1989) and reach the conclusion that a structured model that successfully predicts specific growth rates and utilisation rates of the major substrates (glutamine, glucose, essential amino acids and oxygen) is a more suitable candidate for model based optimisation and control studies. Their work provides a solid framework for the development of structured dynamic models that capture the dynamics of mammalian energy metabolism, however parameters have been estimated from literature data, leading therefore to low confidence levels in the model output.

Even though cell growth is a well-studied area of animal cell cultures, there appear to be many differences between the mathematical models that describe it. These differences mainly involve its dependency on nutrients, metabolites and oxygen. Cell growth has been mathematically related to glucose concentration alone (Frame and Hu, 1991), glucose and glutamine (de Tremblay et al., 1992), glutamine, ammonia and lactate (Bree et al., 1988), glucose and lactate (Kurokawa et al., 1994) and to all four nutrients and metabolites (Miller et al., 2000; Jang and Barford, 2000). All of the above models assume Monod-type kinetics. Tatiraju et al. (1999) have also suggested an equation for oxygen consumption and its relation to cell growth, which is of little use as it is decoupled from the other nutrients with which it has been proven to be associated. Similar models have been obtained for cell death, which relate the rate of cell death to glutamine, lactate and ammonia concentrations (de Tremblay et al., 1992; Bree et al., 1988), or glucose (Frame and Hu, 1991), or glutamine (Dalili et al., 1990), or ammonia and lactate (Batt and Kompala, 1988), or ammonia (Jang and Barford, 2000).

Pörtner and Schäfer (1996) compared a selection of models and model parameters that existed in the literature at that time and carried out an analytic error and range of validity analysis. They found significant variations in the values of maximum growth rate, yields and nutrient Monod constants that were used by other 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 suggested that static batch cultures could be used, for example, for the determination of maximum specific growth rate, but not for establishing a relationship between the growth rate and substrate concentration, whereas continuous cultures could yield reliable data due to the steady state operation conditions. For very low substrate concentrations they suggest using fed-batch cultures. Finally, they recognised that for signicant improvements, parameter identication techniques and control strategies need to be applied to mammalian cell cultures as has previously been the case in other biotechnological processes.

Significant efforts, lead by Fredrickson and co-workers (Eakman *et al.* 1966, Tsuchiya *et al.* 1966), have been made to introduce population balance models (PBM) in biological systems modelling. Even though PBMs have the unique ability to account for the inherent heterogeneity in all cell cultures, unfortunately they are difficult to solve and usually lead to intractable models. Despite their promising characteristics, their limited usage in mathematical biology is mainly due to two major drawbacks (Srienc 1999, Villadsen 1999). They are complicated to handle and solve and accurate determination of model parameters is not possible due to the lack of distribution data.

One of the main contributors in the field of PBM has been Mantzaris and co-workers, who have presented a series of papers (2001a, b and c, 2002) covering a variety of different PBM cases, some of which were compared to models that could be analytically solved. The combination of SCMs and PBMs represents the next logical challenge. Despite the additional model fidelity, such a hybrid model is extremely computationally intensive; hence the solution of even the PBM component of the overall model becomes intractable. To overcome this problem, investigators have used finite-representation techniques to discretise populations avoiding the problems of continuous distributions and the integral differential features they bring. Sidoli et al. (2006) presented a coupled SCM-PBM model which uses a highly structured SCM to characterize single-cell growth and death rates in each stage of a multistage PBM. The model was validated against batch and fed-batch experimental data achieving a satisfactory agreement with some but not all of the modelled variables.

This section was intended as a brief overview of some, but not all, key contributions in the field of mammalian cell culture systems' modelling. The following section will pose the key questions that need to be answered in the near future and intends to motivate the reader to follow through the remainder of this chapter.

1.3 Motivation

A thorough overview of the previous sections reveals that the optimal point on the scale between tractability and fidelity does not lie near the boundaries. On the contrary, an approach that would attempt to exploit the advantages of structured models whilst maintaining tractability could result in a robust yet computationally flexible hybrid model. Such an approach has successfully been followed by Mantalaris and coworkers (Ho *et al.* 2006, Kontoravdi *et al.* 2007, Lam *et al.* 2008) for secreting mammalian cell culture systems. The idea was to maintain structure for the protein formation and secretion process, while using an unstructured model to describe growth and proliferation. All models successfully predict final antibody titres while some (Lam *et al.* 2008) have gone a step further being used for the derivation of an optimal feeding profile.

What ultimately discriminates a good model from a bad model is its ability to successfully describe the modelled process whilst minimising the uncertainty of its output variables. 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. One of the challenges in biochemical engineering 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 candidates for model-based control and optimisation. However, high fidelity models, inherently, contain a large number of parameters. Therefore use of a systematic framework that designs experiments in a way that minimises the necessary experimentation whilst simultaneously maximising information obtainable from the data, is the first step towards achieving a uniquely validated model (Sidoli *et al.* 2004). The increasingly available biological information, both theoretical and analytical, necessitates the use of such a framework from model conception to validation in order to avoid unnecessary experimentation and poorly informative experiments.

The work presented by Asprey and Machietto (2000), Asprey and Mantalaris (2001), Sidoli *et al.* (2004), Ho *et al.* (2006), Kontoravdi *et al.* (2007), Lam *et al.* (2008) and Kiparissides *et al.* (2009) defines a unique and systematic approach to modelling biological systems which is depicted in Figure (2).

Let us assume a first principle mathematical model, $g(x,\overline{P})$, formulated to describe a real life process, where x denotes the input vector, and $\overline{P_I}$ (where i = 1,…,ν) denotes the parameter vector. The first step of the model development algorithm is to determine, before actually designing or performing any experiments, whether model parameters can be uniquely identified from the mathematical structure of the model. Failure to pass the identifiability test implies mathematical singularity with respect to the model parameters, therefore there is no need to perform any further analysis or experiments for a model whose parameters are know *a priori* to be unidentifiable (Asprey and Mantalaris 2001, Sidoli *et al.* 2004). Models that fail the identifiability test should either be reformulated in a way that avoids singularity for the problematic parameters or discarded.

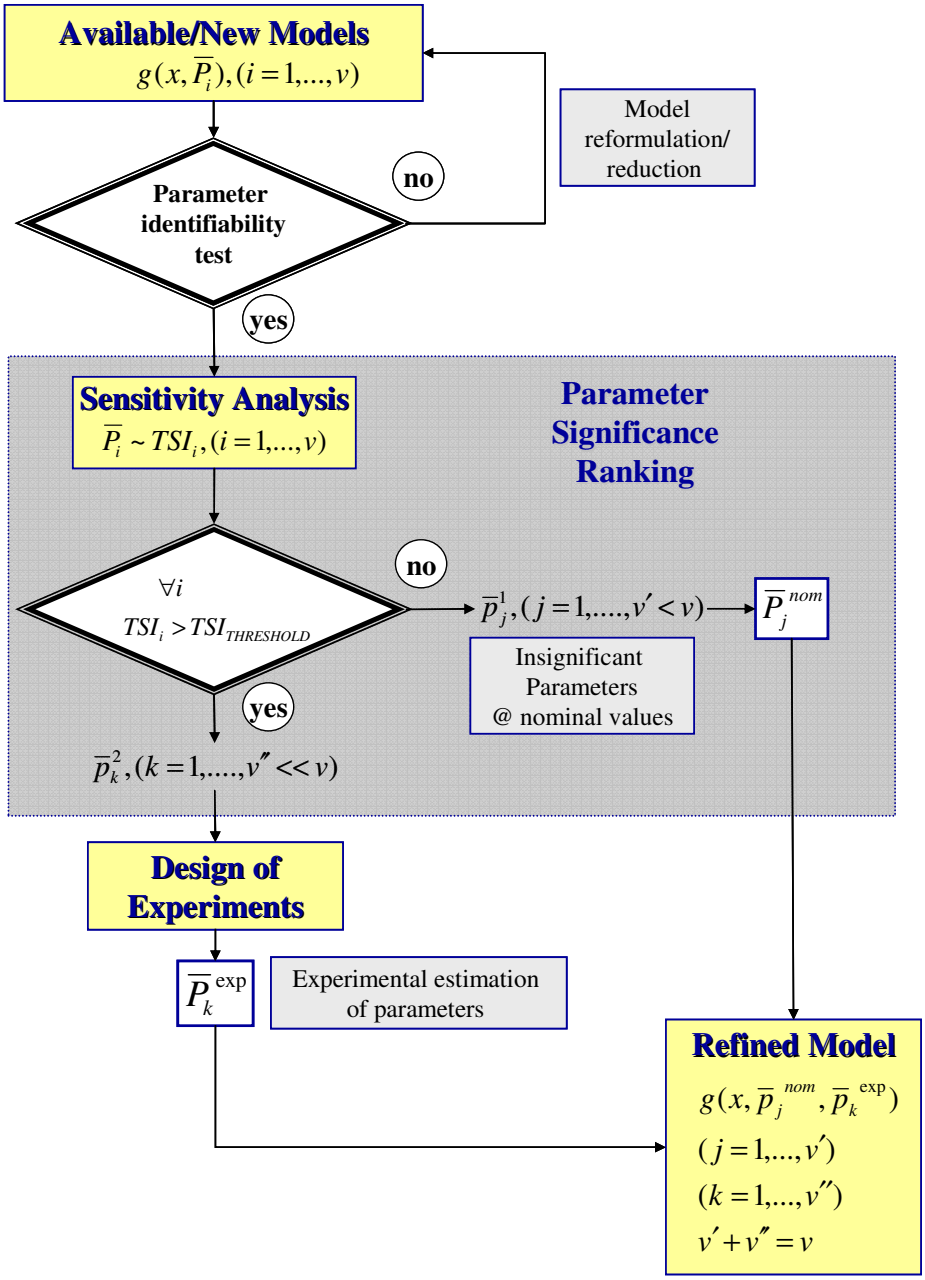
For models that satisfy the criteria of the identifiability test the next step is to apportion the uncertainty in the model output to the sources of variation. Model analysis techniques and sensitivity analysis (SA), in particular, can provide valuable insight regarding the dependence of the model output to its parameters. The output of SA will be a vector of size *v*, containing the sensitivity indices (SI) of the model parameters. As a rule of thumb if *v* larger than 20, the use of parameter grouping will become necessary (Kiparissides *et al.* 2009). This merely affects the notation of Figure (2) and in no way the algorithm itself. Therefore in the case of parameter grouping, $\overline{P_I}$ is the parameter vector and *v* is the number of partitions it contains, corresponding to the number of parameter groups formed.

Consequently an empirical criterion, determined by the modeller, is applied in order to discriminate the significant from the insignificant model parameters. The criterion is a threshold value for the SI, usually set between 0 and 0.2 (Saltelli, 2000). 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 $\overline{p}_j^1$ (j = 1,…,ν'). The remaining parameters whose SI is above the threshold value are allocated in a second partition of the parameter vector termed $\overline{p}_k^2$ (k = 1,…,ν''). The sum of *ν'* and *ν''* should of course equal the size of the parameter vector $\overline{P_I}$, at all times. The values of the parameters in partition $\overline{p}_j^1$ 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 $\overline{p}_j^{nom}$.

The values of the model affecting parameters in partition $\overline{p}_k^2$ need to be determined experimentally with accuracy in order to reduce the uncertainty in the model output. Therefore experiments are specifically designed (Asprey and Macchietto 2000 and 2002) for the

determination of the parameters in vector $\overline{p}_k^2$ and once the experimental data is available the values for parameters $\overline{p}_k^2$ are determined explicitly, yielding vector $\overline{p}_k^{\exp}$. Finally by substituting the initial parameter vector $\overline{P}_I$, with the newly derived $\overline{p}_j^{nom}$ and $\overline{p}_k^{\exp}$, we derive a refined version of the original model, $g(x, \overline{p}_j^{nom}, \overline{p}_k^{\exp})$.

Figure 2 Model development framework for biological systems



Following the above presented framework successfully minimises model uncertainty but more significantly minimises experimental costs and labour. Moreover it sets a scientific platform of communication between modeller and experimentalist, thus bridging the

communication gap between engineers and biologists. In the sections that follow the individual steps of the model development framework will be discussed in more detail, illustrated by the presentation of relative research examples.

1.4 Dynamic Modelling of Biological systems – an illustrative example.

The biological systems model building framework described in the previous section will be explained in detail through a "real life" illustrative example in this section. Let us utilise as an example an industrial process for the production of MAbs harvested from cultures of hybridoma cells. The reason we wish to model this process is the maximisation of final antibody titre in our culture through *in silico* experimentation. Batch and Fed-batch cultures are currently the cultivation methods of choice from the biologics industry for the large scale production of monoclonal antibodies, due to their operational simplicity, reliability, and flexibility for implementation in multipurpose facilities (Bibila and Robinson, 2008). Therefore for the purposes of our example a model capable of describing both batch and fed-batch cultures of antibody secreting hybridomas is required.

Bearing in mind that the model will ultimately be utilised for optimisation studies, which are inherently computationally expensive, renders structured models a less attractive idea. Moreover, as Sidoli *et al.* (2006) have proved, overparametrised models lead to parameter identifiability issues which in turn reduces the confidence in the model output. However the model should contain adequate level of information regarding the antibody formation process and how its various steps are affected by the growth characteristics of the culture and the availability of nutrients in order to yield meaningful results. Balancing the trade-off between tractability and fidelity is the first challenge we need to address. Hybridoma growth kinetics have been widely studied and unstructured models have time and again proven to be capable of capturing their dynamics. Therefore, structure can be avoided at a low cost and an unstructured model can be used to describe cell proliferation and nutrient uptake in batch and fed batch cultures. The model has been adapted by Kontoravdi *et al.* (2005) based on the work of Jang and Barford (2000) and models growth based on the consumption of two basic nutrients (glucose and glutamine) and the production of the two corresponding byproducts of the cell's metabolism (lactate and ammonia).

The formation and secretion of monoclonal antibodies is an inherently complex process (Figure 3). Monoclonal antibodies are Y shaped proteins formed from two identical heavy and two identical light polypeptide chains. The heavy and light chains are encoded from different genes and therefore a situation might easily arise where there is an abundance of one type of

chains yet a shortage in the other resulting in a small, and even zero, production rate. The formation of these proteins starts from the nucleus of the cell, where the chain specific DNA sequence is copied on an mRNA molecule in a process know as transcription. The mRNA molecule will migrate to the endoplasmic reticulum where it will bind to a ribosome and start the process of transcription. Once an antibody molecule has been formed by two heavy and two light chains it will be transferred to the Golgi apparatus where it will undergo posttranslational modifications in order to become a biologically active molecule prior to its secretion to the extracellular environment.

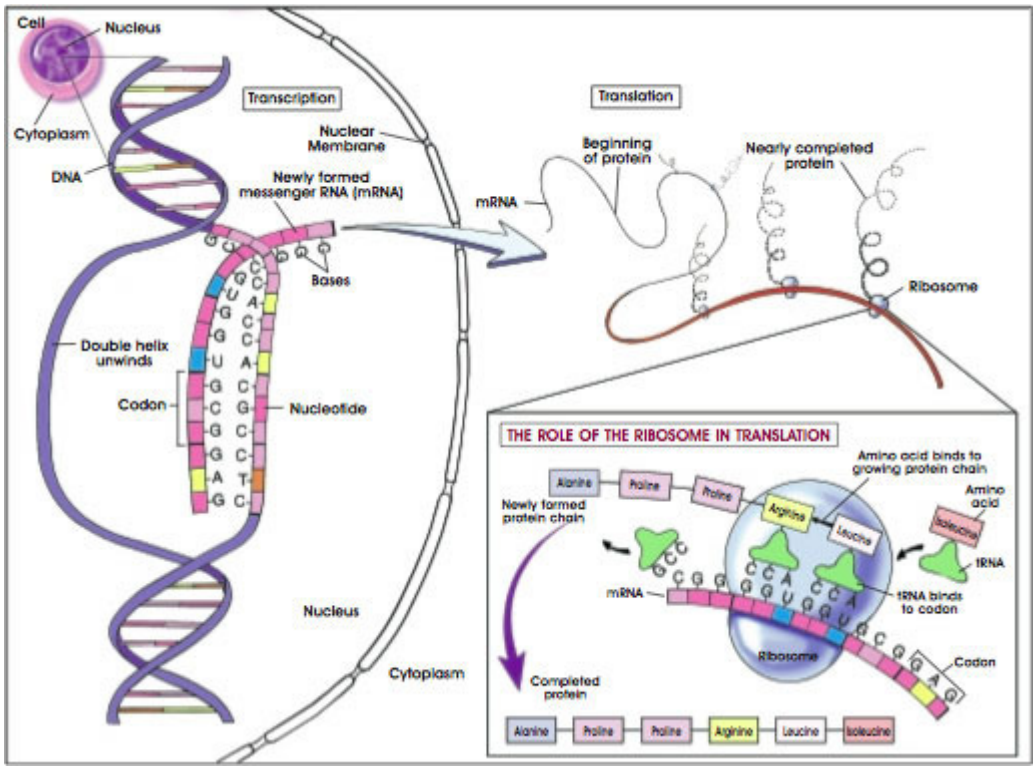


Figure 3 Protein Synthesis [adapted from: http://www.nih.gov/]

Since a significant number of processes, each occurring at a separate site, are involved in antibody formation, utilizing an unstructured model to describe the rate of antibody accumulation in the media would result in a significant loss of information. Therefore the structured model, presented by Bibila and Flickinger (1992), is an ideal candidate for the description of the antibody formation process. Kontoravdi *et al.* (2006) successfully managed to couple this structured model, to the unstructured model (mentioned above) describing cell growth and proliferation. Following this brief discussion around the conceptual formulation of the model, a first principles derivation of the hybrid model is presented below.

1.4.1 First principles model derivation

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

$$\frac{dVX_u}{dt} = \mu VX_u - \mu_d VX_u \,, \tag{1.4.1}$$

where X_u is the concentration of viable cells in the bioreactor measured in cell per liter and μ, 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:

$$\frac{dVX_t}{dt} = \mu VX_u \,, \tag{1.4.2}$$

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

The specific growth rate that appears in equations (1.4.1) and (1.4.2) is estimated through the following formula:

$$\mu = \mu_{\max} f_{\lim} f_{inh} \,, \tag{1.4.3}$$

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:

$$\begin{aligned} f_{\lim} &= \left(\frac{[GLC]}{K_{glc}+[GLC]}\right)\left(\frac{[GLN]}{K_{g\ln}+[GLN]}\right) \\ &\qquad\qquad , \\ f_{inh} &= \left(\frac{KI_{amm}}{KI_{amm}+[AMM]}\right)\left(\frac{KI_{lac}}{KI_{lac}+[LAC]}\right) \end{aligned} \tag{1.4.4}$$

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.

$$\mu_d = \frac{\mu_{d,\max}}{1+\left(\frac{K_{d,amm}}{[AMM]}\right)^n}, \qquad with \qquad n>1 \tag{1.4.5}$$

where, $\mu_{d,max}$ represents the maximum specific death rate (h^{-1}) and $K_{d,amm}$ 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 by-products of cellular metabolism within the bioreactor. Therefore, by performing material balances on each biological compound, 4 ordinary differential equations yielding the temporal evolution of the concentration of nutrients/metabolites are obtained. Specifically, the material balance for the concentration of glucose can be formulated as shown:

$$\frac{d(V[GLC])}{dt} = -Q_{glc} V X_u \,, \tag{1.4.6}$$

where Q_{glc} is the specific glucose consumption rate (mmol/cell/hr) and is defined as:

$$Q_{glc} = \frac{\mu}{Y_{x,glc}} + m_{glc} \,. \tag{1.4.7}$$

The parameters $Y_{x,glc}$ and m_{glc} which appear in equation (1.4.7) are the cell yield on glucose (cell/mmol) and maintenance energy of glucose (mmol/cell/hr), respectively. Equation (1.4.6) was originally presented (Jang and Barford, 2000) with an additional term for glucose consumption by glucokinase, which as Kontoravdi (2006) later argued, based on evidence by Tatiraju *et al.* (1998) has negligible effects. The material balance for glutamine similarly is described by the following equation:

$$\frac{d(V[GLN])}{dt} = -Q_{g\ln} V X_u - K_{d,g\ln} V[GLN] \quad . \tag{1.4.8}$$

The only difference is the term containing glutamine degradation. Glutamine is known to be spontaneously converted into pyrolidonecarboxylic acid at high temperatures and when in weakly acidic or alkaline solutions (Chibnall *et al.*1932). Bray *et al.* (1948) showed that even in medium temperatures, around 37 ^{o}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. The specific consumption rate for glutamine is calculated through a formulation containing the cell yield on glutamine, $Y_{x,gln}$, and the maintenance energy of glutamine, m_{gln}.

$$Q_{g\ln} = \frac{\mu}{Y_{x,g\ln}} + m_{g\ln} \ , \tag{1.4.9}$$

where

$$m_{g\ln} = \frac{a_1[GLN]}{a_2 + [GLN]} , \tag{1.4.10}$$

with a_1 and a_2 being the relevant kinetic constants. Equation (1.4.9) is presented in the updated version (Kontoravdi, 2006) and not as originally presented (Tatiraju, 1998).

Similarly, mass balances can be formulated to describe the temporal evolution of the concentrations of the primary by-products of cell metabolism. More specifically, the mass balance for ammonia is given by:

$$\frac{d(V[AMM])}{dt} = Q_{amm} V X_u + K_{d,g\ln} V[GLN] , \tag{1.4.11}$$

with,

$$Q_{amm} = Y_{amm,g\ln} Q_{g\ln} . \tag{1.4.12}$$

Similarly for lactate:

$$\frac{d(V[LAC])}{dt} = Q_{lac} V X_u \,, \tag{1.4.13}$$

with,

$$Q_{lac} = Y_{lac,glc} Q_{glc} \,. \tag{1.4.14}$$

Q_{lac} and Q_{amm}, represent the specific production rate (mmol/cell/hr) while $Y_{lac,glc}$ and $Y_{amm,gln}$ represent the yield 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 Kontoravdi (2006) consists of an intracellular heavy- and light- chain mRNA balance:

$$\frac{dm_H}{dt} = N_H S_H - K m_H \tag{1.4.15}$$

and

$$\frac{dm_L}{dt} = N_L S_L - K m_L \,, \tag{1.4.16}$$

where m_H and m_L are the intracellular heavy- and light-chain mRNA concentrations (mRNAs/cell), N_H and N are the heavy- and light-chain gene copy numbers (gene/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:

$$\frac{d[H]}{dt} = T_H m_H - R_H \tag{1.4.17}$$

and

$$\frac{d[L]}{dt} = T_L m_L - R_L \,, \tag{1.4.18}$$

where [H] and [L] are the free heavy and light chain concentrations in the ER (chain/cell), T_H and T_L are the heavy- and light-chain specific translation rates (chain/mRNA/h), and R_H and R_L are the rates of heavy- and light-chain consumption in assembly (chain/cell/h). MAbs consist of two heavy (H) and two light (L) amino acid chains. Each molecule is synthesised in the ER according to the following mechanism (Percy, 1975):

$$\begin{aligned} H + H &\longleftrightarrow H_2 \\ H_2 + L &\longleftrightarrow H_2L \\ H_2L + L &\longleftrightarrow H_2L_2 . \end{aligned} \tag{1.4.19}$$

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

$$\begin{aligned} R_H &= \frac{2}{3} K_A [H]^2 \\ R_L &= 2K_A [H_2][L] + K_A [H_2L][L] \end{aligned} , \tag{1.4.20}$$

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

$$\frac{d[H_2]}{dt} = \frac{1}{3} K_A [H]^2 - 2K_A [H_2][L] \tag{1.4.21}$$

$$\frac{d[H_2L]}{dt} = 2K_A [H_2][L] - K_A [H_2L][L], \tag{1.4.22}$$

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

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

$$\frac{d[H_2L_2]_{ER}}{dt} = K_A [H_2L][L] - K_{ER} [H_2L_2]_{ER}, \tag{1.4.23}$$

where $[H_2L_2]_{ER}$ is the MAb concentration in the ER (molecule/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 intraGolgi Mab balance yields:

$$\frac{d[H_2L_2]_G}{dt} = \varepsilon_1 K_{ER} [H_2L_2]_{ER} - K_G [H_2L_2]_G, \tag{1.4.24}$$

where $[H_2L_2]_G$ is the MAb concentration in the Golgi (molecule/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:

$$\frac{d(V[MAb])}{dt} = (\gamma_2 - \gamma_1\mu)Q_{MAb}VX_V\,, \tag{1.4.25}$$

where:

$$Q_{MAb} = \varepsilon_2\lambda K_G[H_2L_2]_G\,, \tag{1.4.26}$$

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 equation (1.4.25), [MAb] is the MAb concentration in the culture, and γ_1, γ_2 are constants.

Equations 1.4.1 to 1.4.26 form a first principles model consisting of a total of 16 differential equations and 30 model parameters. In order to save time and effort from performing tedious computations manually, models are usually implemented in a CAD (Computer Aided Design) tool of choice. Common choices amongst biochemical engineers include, but are not limited to, Fortran, C++, the Mathworks Matlab® suite, Mathematica®, Mathcad® and gPROMS®. For the purposes of our example we choose gPROMS® as the software of choice, due to its superior solvers and seamless integration of experimental data (PSE 2009).

The next step in the biological model development algorithm is the derivation of initial estimates for the model parameters from relevant experimental data. In the case that the utilised model already exists, this step can utilise parameter values obtained from relevant literature. For the derivation of estimates for the presented model's parameters we will borrow experimental data of batch hybridoma cultures from the work of Kontoravdi (2006). All parameter estimation experiments and model simulations where carried out on an Intel® Core™2 Duo (E4600 – 2.4, 2.39) personal computer with 3.24 GB of RAM memory and all model simulations and parameter estimation experiments where implemented in the advanced process modelling environment gPROMS® (Process Systems Enterprise, 2009).

gPROMS is an equation-oriented modelling system used for building, validating and executing first-principles models within a flow sheeting framework. Parameter Estimation in gPROMS is based on the Maximum Likelihood formulation which provides simultaneous estimation of parameters in both the physical model of the process as well as the variance model of the measuring instruments. gPROMS attempts to determine values for the uncertain physical and variance model parameters, θ, that maximise the probability that the mathematical model will predict the measurement values obtained from the experiments. Assuming independent, normally distributed measurement errors, ε_{ijk}, with zero means and standard deviations, σ_{ijk}, this maximum likelihood goal can be captured through the following objective function:

$$\Phi = \frac{N}{2}\ln(2\pi) + \frac{1}{2}\min_{\theta}\left\{\sum_{i=1}^{NE}\sum_{j=1}^{NVi}\sum_{k=1}^{NMij}\left[\ln\left(\sigma_{ihk}^2\right) + \frac{\left(\bar{z}_{ijk} - z_{ijk}\right)^2}{\sigma_{ihk}^2}\right]\right\} \qquad (1.4.27)$$

where N stands for total number of measurements taken during all the experiments, θ is the set of model parameters to be estimated, NE is the number of experiments performed, NV_i is the number of variables measured in the i th experiment and NM_{ij} is the number of measurements of the jth variable in the ith experiment. The variance of the kth measurement of variable j in experiment i is denoted as σ^2_{ijk}, while $\bar{z}_{ijk}$ is the kth measured value of variable j in experiment i and z_{ijk} is the kth (model-) predicted value of variable j in experiment i. The above formulation can be reduced to a recursive least squares parameter estimation if no variance model for the sensor is selected.

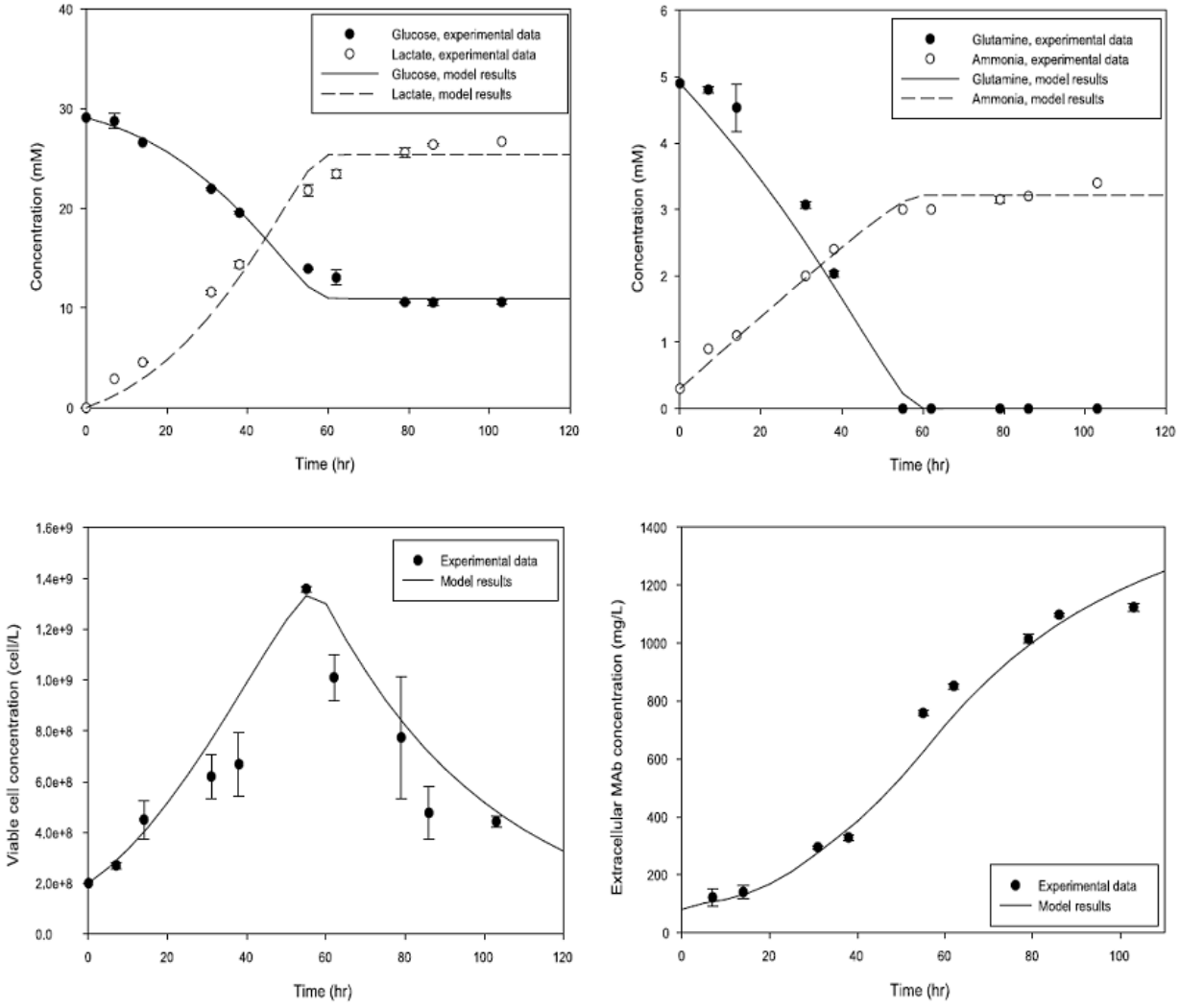


Figure 4 Experimental Data from batch hybridoma cultures and model predictions

Table (1.2) summarises the list of model parameter estimates obtained from the parameter estimation algorithm while Figure (4) 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 well posed model, capable of describing the process under study even with initial parameter estimates. Having obtained initial estimates for the values of the model parameters we can proceed to the next step of the algorithm, namely model analysis.

Table 1.2 Model parameter estimates derived from batch hybdridoma culture data

Symbol	*Units*	*Nominal Value*
μ_{max}	hr^{-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
$m_{d,max}$	hr^{-1}	0.03
$K_{d,Amm}$	mM	0.1386
n	Real integer	0.995
$Y_{Lac,Glc}$	Dimensionless	1.399
m_{Glc}	Mmol/cell/hr	$4.853*10^{-14}$
$Y_{x,Glc}$	Cell/mmol	$1.061*10^{8}$
$Y_{x,Gln}$	Cell/mmol	$5.565*10^{8}$
$K_{d,gln}$	hr^{-1}	$9.6*10^{-3}$
a_1	mM*L/cell/hr	$3.4*10^{-13}$
a_2	mM	4
$Y_{Amm,Gln}$	Dimensionless	0.4269
K	h^{-1}	0.1
N_H	gene/cell	139.8
S_H	mRNAs/gene/h	300
N_L	gene/cell	117.5
S_L	mRNAs/gene/h	4500
T_H	chain/mRNA/h	17
T_L	chain/mRNA/h	11.5
K_A	(molecule/cell) h^{-1}	10^{-6}
K_{ER}	h^{-1}	0.693
K_G	h^{-1}	0.1386
ε_1	Dimensionless	0.995

γ_1	Dimensionless	0.1
γ_2	Dimensionless	2
ε_2	Dimensionless	1
λ	g/mol	146000

1.4.2 Model Analysis

Model analysis techniques and sensitivity analysis, in particular, can provide valuable insight regarding the dependence of the model output to its parameters. Allocating model uncertainty to the various sources of uncertainty (i.e. model parameters) facilitates the targeted reduction of output uncertainty by accurately estimating model parameters 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 output, can be fixed at their literature values (if available) or approximated hence reducing unnecessary experimentation. There is a vast plethora of available model analysis techniques, enough to confuse even the experienced modeller. Below follows a rational discussion, leading to the proposal of the most suitable group of model analysis techniques, in the context of dynamical modelling of biological systems. A detailed study on the performance and applicability of sensitivity analysis techniques in the context of biological models can be found in the work of Kiparissides *et al.* (2009).

Dynamic models describing complex biological functions involve highly non-linear 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 analysis methods are commonly grouped in three main categories, namely screening, local and global methods.

Screening methods are randomised, one-at-a-time numerical experiments, which aim to indicate the most important factors amongst 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 non-linear models (Saltelli, 2000) Efforts to calculate higher order effects, through screening methods, have been recorded in the literature (Box *et al.* 1978b, Cotter *et al.* 1979), though these methods fall short either in terms of 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 (Saltelli, 2000) The main drawback of GSA methods is their extensive computational requirements for large models. GSA methods are commonly grouped in two categories, namely methods that utilise 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 can not 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.

$$f(X_1,...,X_n) = f_0 + \sum_{i=1}^{n} f_i(X_i) + \sum_{1\le i<j\le n} f_{ij}(X_i, X_j) + ... + f_{1,2,...,n}(X_1,...,X_n) \quad (4.2.1)$$

Under the assumption that each of the terms in (4.2.1) is orthogonal (Sobol' 2001, Homma and Saltelli 1996) 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 explore 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 can not be neglected for non-linear models. Chan *et al.* (1997) have illustrated the significance of higher order sensitivity

indices in understanding the behaviour 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) (Sobol' 1990, Homma and Saltelli 1996) The TSI for parameter *i* is estimated as the sum of all higher order terms in (4.2.1) which include parameter *i*. The vast majority of sensitivity analysis techniques do not include a decomposition similar to the one presented in equation (4.2.1) 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' (2001), Saltelli (2000), and Chan *et al.* (1997). The most commonly used variance based methods include the Sobol' global sensitivity indices and the Fourier Amplitude Sensitivity Test (FAST). As reported in Kiparissides *et al.* 2009, such methods can be computationally exhaustive, however robust and less cumbersome alternatives exist (Kucherenko *et al.,* 2008).

For the purposes of our example, we have chosen to use the global sensitivity analysis. The dimensionality of the sensitivity analysis problem is ultimately 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. A detailed discussion on parameter grouping and various methods for grouping can be found in the work of Kiparissides *et al.* 2009, and is summarised in table (1.3), adapted from the same work.

Table 1.3 Grouping Parameters in GSA (adapted from Kiparissides *et al.* 2009)

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

An overview of the above table, will help us decide upon the most beneficial for our analysis method of grouping. Having already tested our model's agreement with experimental data (Figure 4), we have concluded that our model is well posed and therefore there is no need for reduction or structural reformulation. Therefore, grouping according to biological significance seems to be the most suitable method of grouping our parameters prior to the application of sensitivity analysis.

As stated earlier, a batch operation mode was considered, the model was simulated for 120 hours of culture time, and the sensitivity analysis was performed at three characteristic time points (20, 50 and 120 hours). SI's have been known to change dynamically along the time trajectory of the model output. For example, as the culture progresses and nutrients start being depleted, the model output will become more sensitive towards the parameters affecting nutrient uptake and metabolism. This is a valuable property as it can provide information regarding the time point that would yield the most informative experiments. Sensitivity analysis was conducted at different phases of the cell culture in order to capture the dynamics of the various growth phases of a batch cell culture. Specifically time points from the lag phase, the exponential growth phase and the decline phase were evaluated. The output variables of interest, from a process point of view, are viable cell concentration and MAb concentration as these ultimately define the final amount of MAb titre available. The simulations involved scanning of all model parameters with respect to the output variables of interest. The uncertainty range associated with each of the 30 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 30 parameters were grouped into 4 unequal groups, which were formulated based on their biological function and can be seen table 1.4.

When the goal of sensitivity analysis is to indicate candidates for model based DOE, a "first layer" analysis of grouped parameters does not suffice. More specifically, considering the formulation of our problem, the first analysis will indicate one – or more – groups of parameters as significant instead of identifying individual parameters. A group of parameters with a high sensitivity index does not necessarily translate in all parameters within that group being significant with respect to the model output. Moreover, it is often the case, that different groups

will have high sensitivity indices at different points of the predetermined time horizon. Therefore, a second analysis, this time within the significant groups is required in order to identify individual parameters that affect the model output.

Table 1.4 Model Parameters: Biological Significance Grouping

Group 1 – Growth/Death related	**Group 2 - Metabolism related**
μ_{max}	$Y_{Lac,Glc}$
$K_{I,Amm}$	m_{Glc}
$K_{I,Lac}$	$Y_{x,Glc}$
K_{Glc}	$Y_{x,Gln}$
K_{Gln}	$K_{d,gln}$
$m_{d,max}$	a_1
$K_{d,Amm}$	a_2
N	$Y_{Amm,Gln}$
Group 3 – Mab Synthesis Related	**Group 4 - Mab secretion related**
K	γ_1
N_H	γ_2
S_H	ε_2
N_L	
S_L	
T_H	
T_L	
K_A	
K_{ER}	
K_G	
ε_1	

From a computational point of view, the "first layer" analysis is computationally more demanding from the "second layer" as the entirety of the parameter space is sampled. Furthermore, since the "first layer" ultimately is a stepping stone towards model based DOE, the acquisition of sensitivity indices of first order is somewhat of a luxury. Taking this into account we have chosen to use Derivative Based Global Sensitivity Measures (DGSM, Kucherenko *et al.*

2008) for the "first layer" of GSA. DGSM is a global screening method, proved (Kiparissides *et al.* 2009) to provide results similar to the variance based methods' total sensitivity index (TSI), and infact has been shown to have a direct correlation with the Sobol' TSI in most cases (Sobol' and Kucherenko 2008). The main benefit of using DGSM over a variance based method, is the significant gain in terms of computational time. DGSM provides only TSI information, and not first or higher order information, but as discussed previously this is not an issue for this stage of the analysis. Figure (5) presents the results of the analysis when studying parameters in the groups defined in table (1.4).

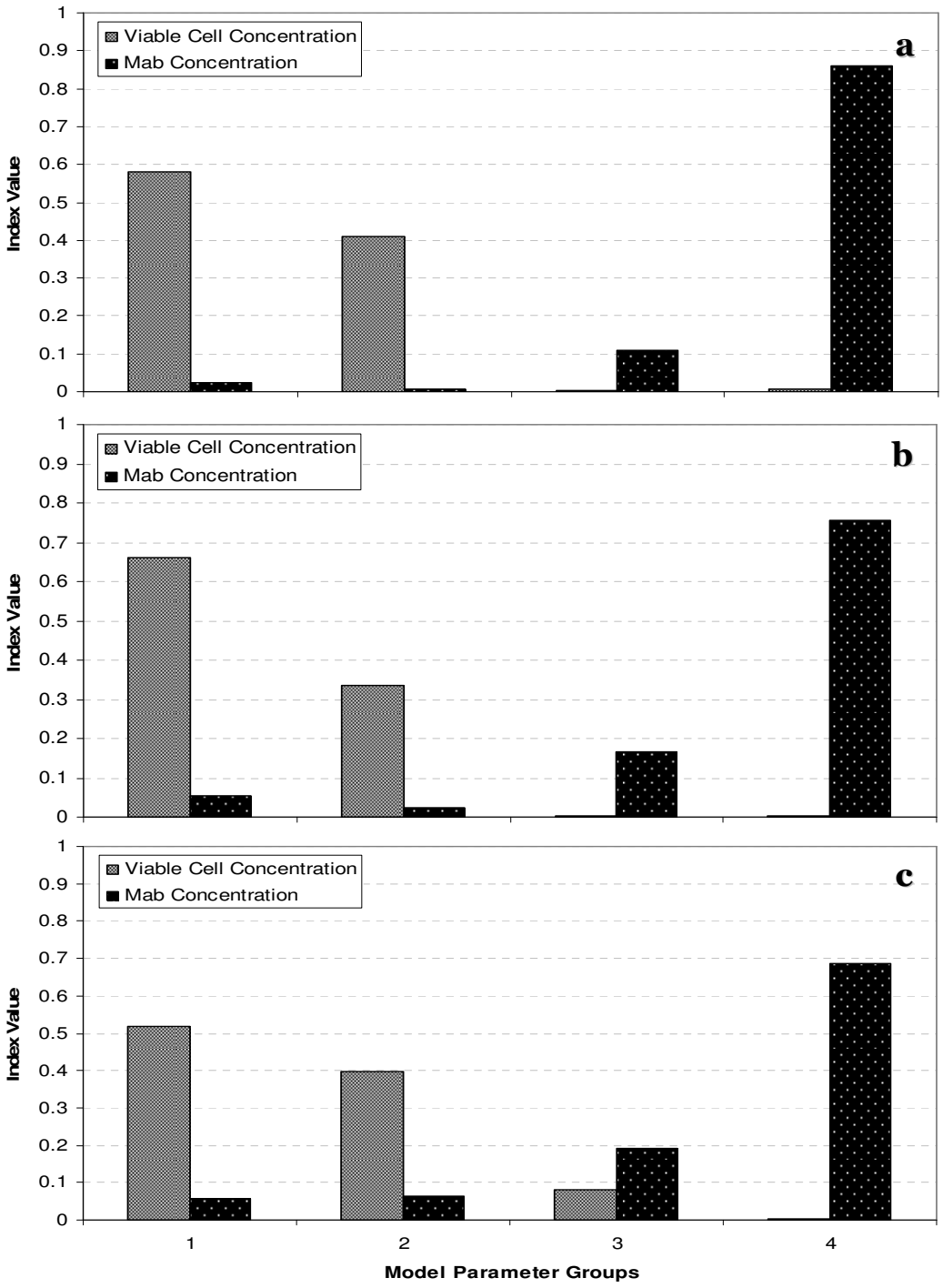


Figure 5 "First Layer" GSA: a. 20h, b. 50h and c.120h of culture time

Observing the results of Figure (5), one can conclude that different parameter groups affect different model outputs. This is both expected and logical. The unstructured model as seen through its output variable, namely viable cell concentration, is affected mainly by parameter groups 1 and 2, which contain the parameters of the unstructured model. In the early stages (20,50h) group 1 is the more significant by a fair margin while group 2 becomes increasingly significant as the culture progresses. Group 2 contains the parameters associated with nutrient

uptake and metabolism, and as discussed earlier is expected to become more significant as the depletion of nutrients starts being an increasingly crucial factor for the culture. On the other hand, the structured MAb formation model is affected mainly by parameter groups 3 and 4, which contain the respective model parameters, but as the culture progresses starts to be slightly affected by groups 1 and 2.

One could arguably discuss that this “first layer” of GSA has yielded no useful results as it appears that all parameter groups show some significance throughout the culture, indicating possibly a poor choice in parameter grouping. However a closer observation of the results of Figure (5) will yield valuable information. That is, group 4 for example is constantly significant throughout the culture with its sensitivity index varying slightly. Therefore a “second layer” analysis within group 4 at only one time point would be sufficient to indicate which parameters of that group are responsible for the high sensitivity index (SI) of that group. Due to the nature of the grouping we have chosen, the significant individual parameters of group 4 are expected to maintain their significance, in accordance to the behaviour of the SI of group 4, throughout the culture. Therefore we opt to perform a “second layer” analysis of the parameters within group 4 at the first time point (20h) for two reasons. First and foremost the earlier time points are less computationally demanding, even if marginally, than latter time points due to the required integration time. Moreover group 4 has the highest sensitivity index at 20h of culture time.

Group 1 has the highest sensitivity index at 50h of culture time, which makes sense from a biological point of view, since it’s the time point closest to the peak in viable cell concentration. Therefore the best time point to scan the parameters of group 1 is 50 hours of culture time. Using the same rationale, group 2 will be studied at 20 hours while group 3 will be studied at 120 hours. Therefore this “first layer” of GSA has indeed provided considerable amount of information, serving its purpose to guide us through the “second layer” which will indicate the actual model parameters and that affect the model output the most, and the point in time where they have the highest sensitivity index associated with them, leading to suggestions for model based DOE.

Since the “second layer” of GSA will involve smaller problems as only partitions of the parameter space will be sampled and individual parameters will be studied we have chosen to use the Sobol’ global sensitivity indices for this part of the analysis. The Sobol’ sensitivity indices, even though more computationally cumbersome will provide information on parameter interactions and can therefore be used to exclude parameters with a high level of non-linear interactions from DOE on the basis of singularity. Figures (6-9) present the results of the “second layer” of GSA, scanning individual parameters as discussed above.

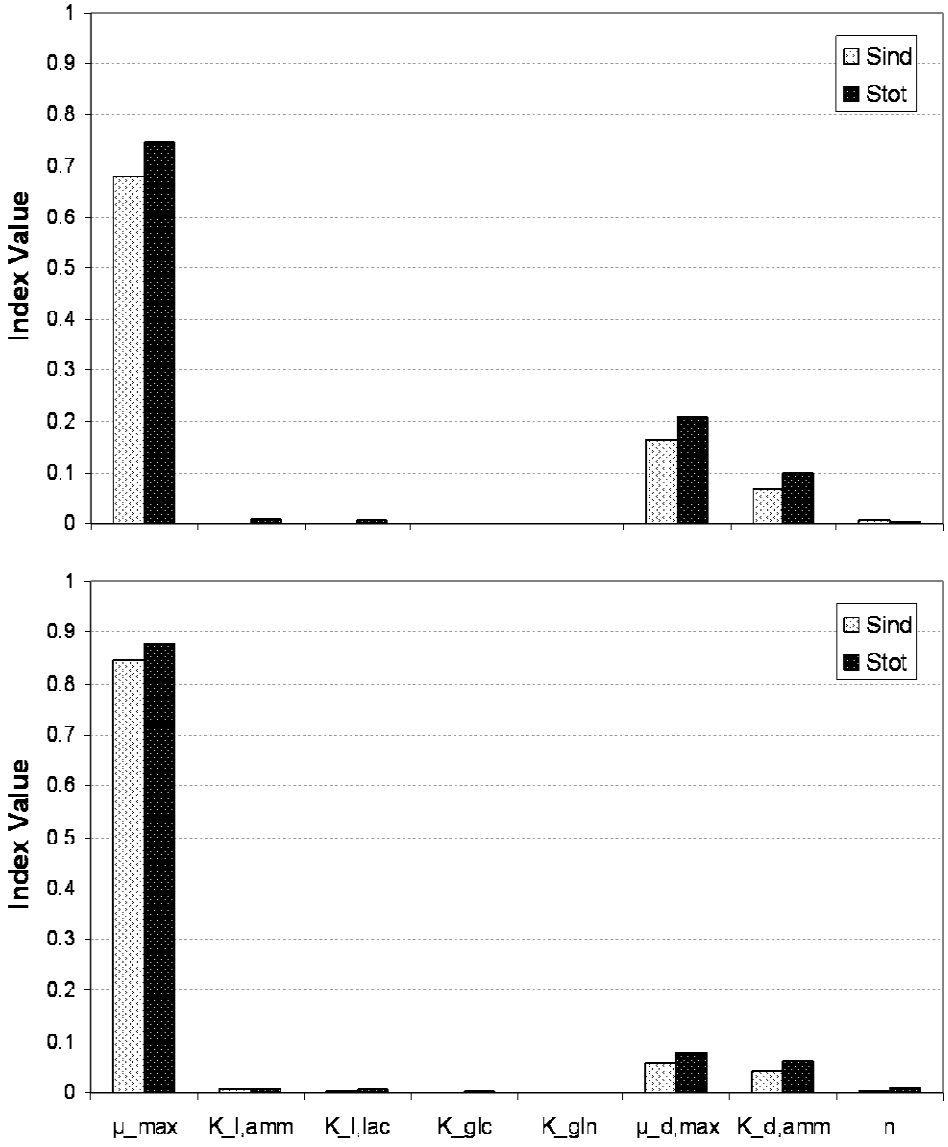


Figure 6 Sensitivity indices for parameters of group 1 at 50h of culture time; Top: Viable Cell concentration as the output, Bellow: MAb Concentration as the output

The largest contributor to the high sensitivity index of group 1 is the maximum growth rate (μ_{max}) as can be seen from Figure (6). From the remaining parameters only the maximum death rate ($\mu_{d,max}$) and to a lesser extent the death rate due to ammonia accumulation ($K_{d,amm}$) seem to affect the output of viable cell concentration. Another important conclusion drawn from figure (6) is that all parameters have a very small contribution of parameter interaction towards their overall sensitivity index. As discussed earlier the difference in value between the total index and the individual index is an indication of parameter interactions. Low level of interactions is a desirable property as it allows for more accurate parameter estimations.

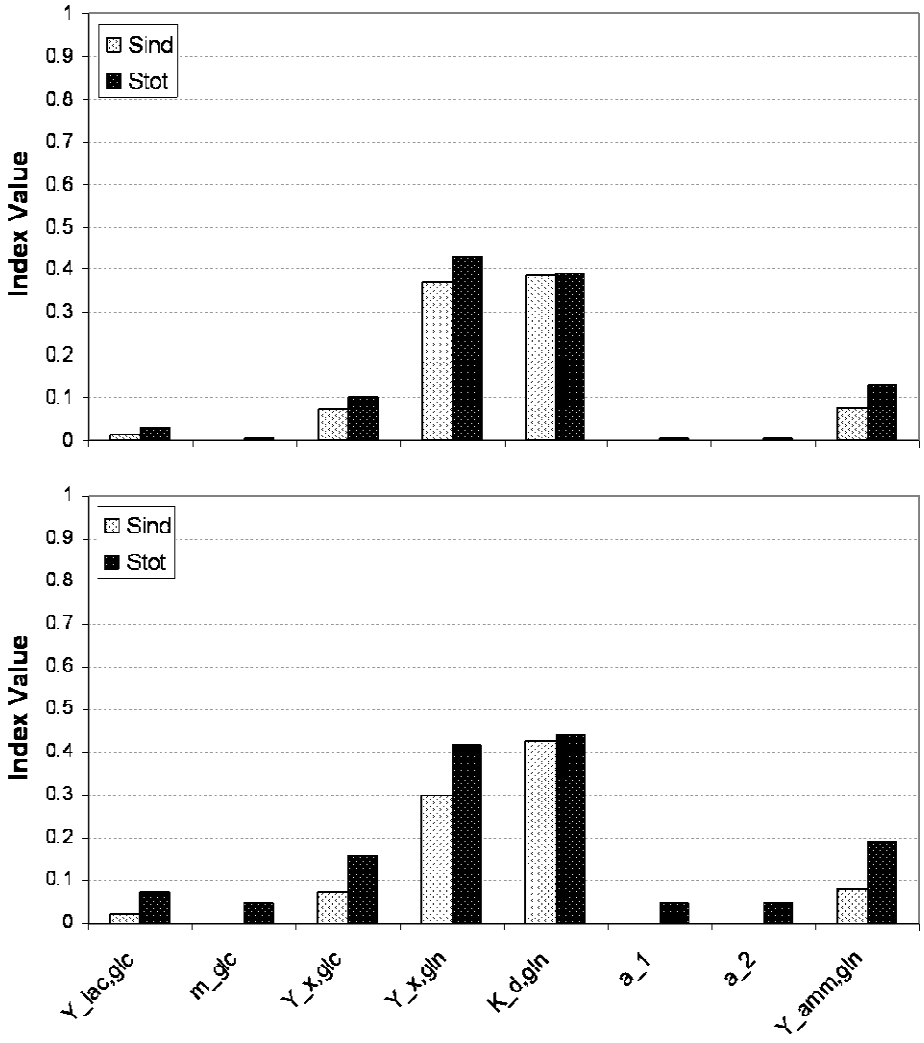


Figure 7 Sensitivity indices for parameters of group 2 at 20h of culture time; Top: Viable Cell concentration as the output, Bellow: MAb Concentration as the output

The sensitivity indices for the parameters of group 2 are quite similar between the two studied model outputs. Taking a closer look at the model formulation this result is both reasonable and expected. The parameters of group 2 are associated with the nutrient uptake rates and the metabolite accumulation rates. Therefore they ultimately define the overall growth rate which in turn affects the viable cell concentration profile. Moreover, the final antibody titre is a function of both the overall growth rate and the viable cell concentration. Figure (7) is a confirmation of the above, indicating the same parameters as significant for both model outputs. Furthermore a slight increase in the difference between the total and individual indices is noticeable when MAb concentration is the studied output, indicating the indirect effect these parameters have on the said output variable. The significant parameters are the yields of cell mass on both substrates ($Y_{x,glc}$, $Y_{x,gln}$) and the spontaneous degradation rate of glutamine ($K_{d,gln}$). The threshold level below which a parameter is considered insignificant is arbitrarily chosen by the modeller as stated earlier. In the present work parameters with sensitivity indices smaller

than 0.1 are considered insignificant, while others (Saltelli 2000) favour a higher cut-off point such as 0.2 or even 0.3.

According to figure (5) groups 3 and 4 have sensitivity indices equal to zero with respect to viable cell concentration as the model output. Referring back to the model equations this is readily justifiable since parameters from both groups are the parameters of the structured model describing MAb formation and secretion. The structured model is coupled to the unstructured model in a one-way manner. That is, the unstructured growth model affects the output of the structured model but the structured part of the model has no effect on the output of the growth model. Therefore for groups 3 and 4 only the sensitivity indices with respect to Mab concentration as the model output will be considered.

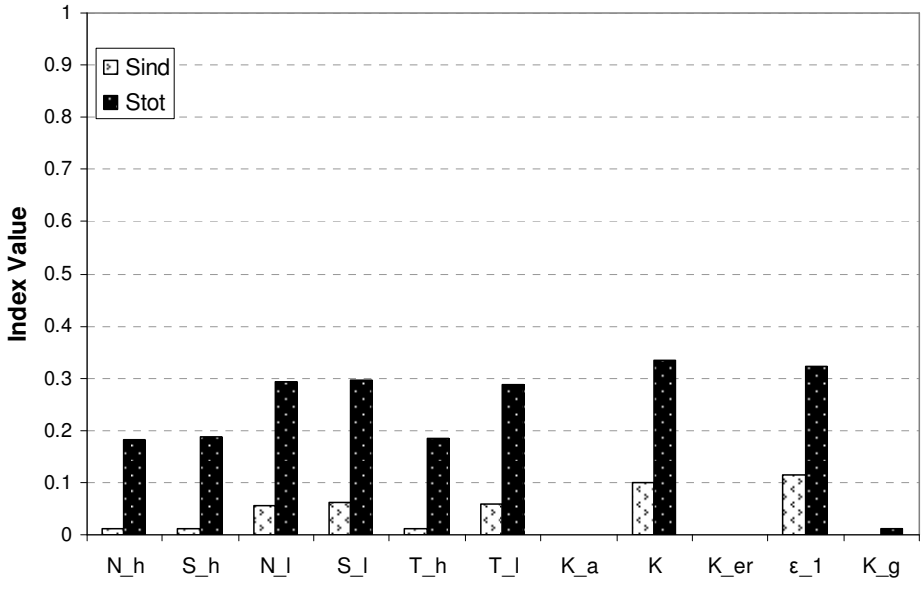


Figure 8 Sensitivity indices for parameters of group 3 at 120h of culture time; MAb Concentration as the output

Figure (8) seems difficult to interpret at a first glance. Almost all of the model parameters have a non-negligible total sensitivity index, yet only one (ε_1) has an individual index higher than 0.1. As mentioned earlier parameters whose total index in strongly influenced by non-linear interactions are poor candidates for parameter estimation as they can not be uniquely indentified (Sidoli *et al.* 2006). However this does not necessarily mean that none of the parameters of group 3 can be uniquely estimated. It is highly probable that the time point of the analysis was a poor choice and performing the analysis at a different time point might yield more informative results. Few other conclusions can be drawn from figure (7) alone. Therefore we have chosen to repeat the analysis of the parameters of group 3 at a different time point, namely at 20h of culture time.

Should the results of the new analysis resemble the results shown in figure (7) this would be an indication of an over-parameterised and ill-posed model.

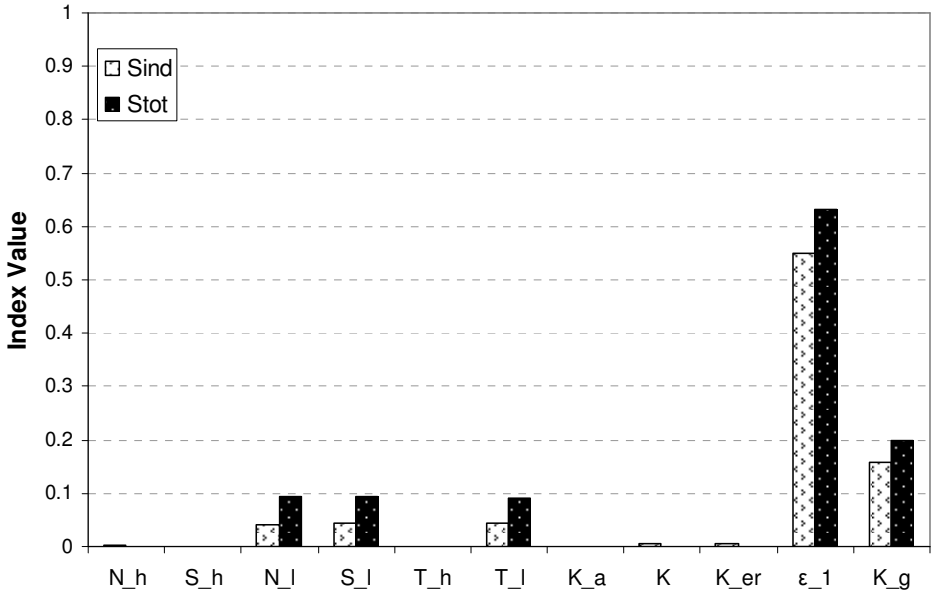


Figure 9 Sensitivity indices for parameters of group 3 at 20h of culture time; MAb Concentration as the output

Figure (9) contains the sensitivity indices for the parameters of group 3 as calculated after 20 hours of culture time. The results at this time point are indeed much more informative than the ambiguous results of figure (8). The glygosylation efficiency factor (ε_1) and the rate constant for Golgi to extracellular media transport (K_g) are the only parameters with a TSI greater than 0.1. Both parameters also have quite high individual indices which suggest that they could be uniquely identified from a suitably designed experiment. It is worthwhile identifying the reasons behind the differences between Figures (8) and (9). After 120 hours of culture time, glutamine has been completely depleted and the culture is well in its decline phase. A near zero concentration for glutamine would yield a near zero value for the specific growth rate as given from equations (1.4.3, 1.4.4). This in turn would affect the right hand side of equation (1.4.25) which in fact yields the output variable. Since GSA is a numerical tool it shares the same limitations as the numerical solvers it utilises for the integration of the DAE system. Therefore when the right hand side of equation (1.4.25) is known to lie at a near zero (if not exactly zero) value any variance from parameter value alterations is difficult to be quantified at the specific time point.

The analysis of the parameters of group 4 (Figure 10) highlighted both the glycosylation efficiency factor (ε_2) and one of the constants for MAb secretion (γ_2) as significant parameters. Both parameters display a high individual index value which, as mentioned earlier, is a desirable property from a parameter estimation point of view. However, even though a parameter may be

significant to the model output and 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 these parameters (ε_1, ε_2), 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 and therefore was excluded from the DOE algorithm. The difficulty of obtaining experimental measurements for certain parameters is often a "real life" problem. Some might argue that on the basis of experimental estimation, such parameters could be excluded from model analysis since they can not be estimated. However, the fact that a parameter can not 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.

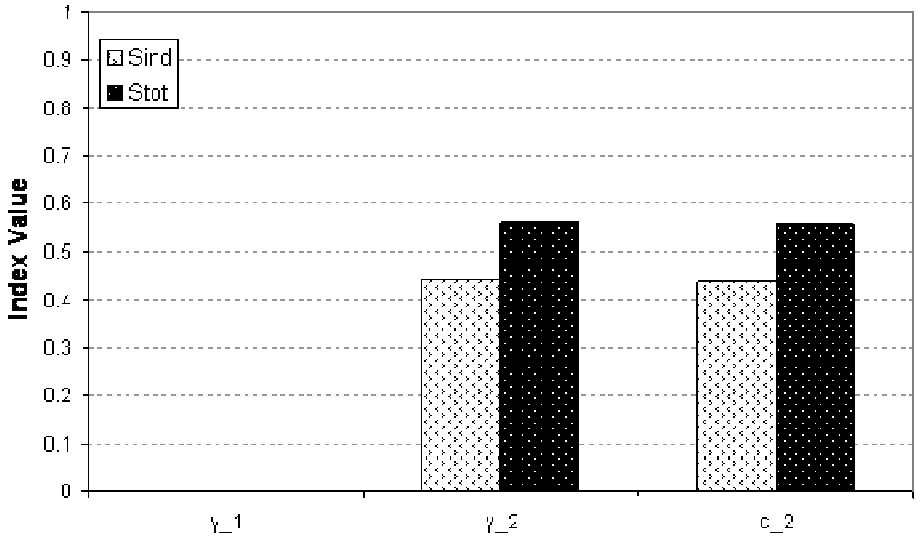


Figure 10 Sensitivity indices for parameters of group 4 at 20h of culture time; MAb Concentration as the output

The gain from the detailed analysis in this section is the reduction of the number of parameters that need to be experimentally validated in order to increase the fidelity of our model. Having started from a total of 30 parameters we have successfully narrowed down the parameters that need to be experimentally validated to a mere 7 summarised in table 1.5. Moreover we have gained valuable information regarding the time points that would yield the most informative experiments leading to a more accurate estimation of our model parameters. Therefore model analysis is a key step towards the development of a robust and well posed

dynamic model and should always be performed prior to experimentation in order to avoid unnecessary experimental costs and labour. The next step of the model development algorithm is DOE and finally model validation through an independent set of experiments.

Table 1.5 Summary of GSA results.

Parameters to be input to the DOE algorithm	
Parameter	**Time point of GSA (h)**
μ_{max}	50
$\mu_{d,max}$	50
$K_{d,amm}$	50
$Y_{x,glc}$	20
$Y_{x,gln}$	20
$K_{d,gln}$	20
K_g	20

1.4.3 Design of Experiments and Model Validation

Thus far we have presented the derivation of a model that can accurately describe batch cultures of MAb secreting hybridoma cultures as it flows through the model development algorithm of Figure (2). Using as an example the model and experimental results presented in the work of Kontoravdi (2006) we have successfully created the partitions of the parameter vector containing the significant parameters (Table 1.5) and the insignificant parameters that will be set at their nominal values. The next step of the model development algorithm is to design tailor made experiments for the significant parameters in order to facilitate their accurate estimation. In order to guide the reader through this step of the algorithm we will again use as an example the relevant work of Kontoravdi (2006) in an attempt to provide a "closed-loop" overview of the model development framework presented in Figure (2).

However, from a process engineering point of view, fed-batch operation is the most important for industrial applications as it can prolong culture longevity therefore increasing MAb productivity and final titre. Therefore the goal set out is to extend the model's predictive capabilities to fed-batch conditions, so that it can be used for the application of model based optimisation and control. Assuming that the model is valid under such conditions, the specific objective is to accurately estimate the significant model parameters from fed-batch experimental

data. As previous studies have discussed (Versyck et al., 1997; Nathanson and Saidel, 1985; Munack and Posten, 1989), optimal experimental design uses the model to design sufficiently informative experiments for this purpose. Borrowing experimental data from the work of Kontoravdi (2006)

Table 1.6 Optimal Experiment Schedule [Adapted from Kontoravdi 2006]

Time (h)	**Feed Volume (mL)**	**Time (h)**	**Feed Volume (mL)**
12		90	
F - 12.1	1.25	96	
18		108	
24		***F*** - 108.1	1.25
36		114	
F - 36.1	1.25	120	
42		132	
48		***F*** - 132.1	1.25
60		138	
F - 60.1	1.25	144	
66		156	
72		***F*** - 156.1	1.25
84		162	
F - 84.1	1.25	168	

Following the detailed analysis of the model and its parameters in the previous section we have already identified the most significant model parameters which can readily be input to the optimal experimental design algorithm. Reaping the benefits of using an advanced CAD tool like gPROMS, an optimal experimental design utility is already implemented trivialising the application of DOE. The concentrations of glucose and glutamine in the feed were set at 500mM and 100mM, respectively. The maximum total volume of feed was fixed at 8.75ml, which represents nearly 5% of the total culture volume (200ml), so as to avoid dilution effects.

Sampling times, at which measurements were conducted were determined (indicated from the work of Kontoravdi 2006) *a priori* and can be found in table 1.6. The output of the algorithm (table 1.7) provided us with 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 hours) (Kontoravdi 2006).

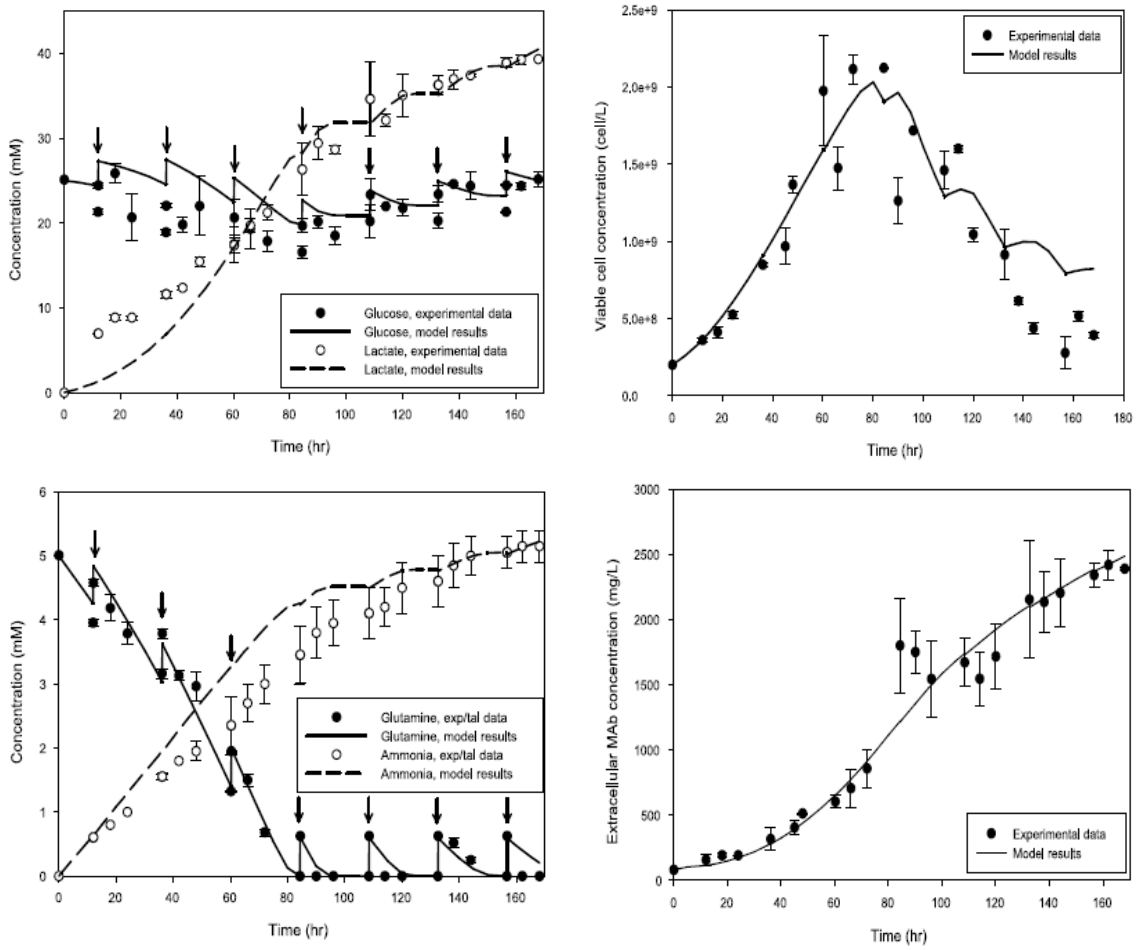


Figure 11 Fed-Batch Cultures of HFN 7.1 Hybridoma cells [Adapted from Kontoravdi 2006]

Borrowing fed-batch experimental data of HFN 7.1 hybridoma cultures from the work of Kontoravdi (2006) enables the re-estimation of the parameters identified as significant from model analysis. The "refined" version of the model is simulated for fed-batch operation and is plotted against relevant experimental data. 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 Figure (11). Table 1.7 contains the values of the "refined" model parameters.

Having verified the validity of the model structure for the simulation of fed-batch processes, its applicability should be examined against an independent set of experimental data. This will prove that model parameters were not just fitted to experimental data but were properly

estimated and the model is valid under various operating conditions. It is beyond the subject of this chapter however to provide such a validation and for a proof of concept the reader can refer to the work of Kontoravdi (2006,2008) where such data are available.

Table 1.7 Results of the parameter estimation algorithm

Parameter	New Estimated Value
μ_{max}	0.05439
$\mu_{d,max}$	0.02784
$K_{d,amm}$	1.932
$Y_{x,glc}$	$2.6*10^{8}$
$Y_{x,gln}$	$8*10^{8}$
$K_{d,gln}$	$8.75*10^{-3}$
K_g	0.14

1.5 Concluding remarks

The balancing point for the trade-off between fidelity and tractability is constantly shifting with the advancements both in numerical tools and raw computational power. Most of the models describing mammalian cell culture presented thus far in the literature are based on the consumption of up to 2 basic nutrients and the toxic effects from the accumulation of the corresponding end products of metabolism. However, in order to truly capture the dynamics and behaviour of a culture and achieve truly optimal feeding strategies, we need to start paying attention to a number of other components that have so far being ignored. The work of Xie and Wang (2001) has already taught us that excessive feed of glucose is not always the best means towards higher titres of product, as they proved it shifts metabolism towards energy inefficient pathways. Moreover, deZengotita *et al.* (2000) has shown that apart from the established growth limiting nutrients (namely glucose and glutamine), there are many other components that might be limiting the growth of a culture. The question that naturally arises is whether an optimal feeding profile should be derived on the basis of availability of nutrients while disregarding the energy requirements of the cell or on the provision of adequate yet not excessive amounts of

energy through the provision of controlled quantities of nutrients. Energy metabolism is a significant element of cell culture that has thus far been ignored from a modelling point of view.

Optimisation of cultures secreting a valuable end product has always been centred around the balance between prolonged culture life and increased specific productivity. Usually these two goals are reached through competing paths. Conditions that seem to prolong culture viability reduce specific productivity while conditions that increase specific productivity seem to affect culture longevity. Even though a lot of published studies have shown the potential of *in silico* experimentation, a lot of work still remains to be done. In order to achieve the global optimum between prolonged culture life and increased productivity, we must first understand and incorporate the significant elements of metabolism in our models. For example, there has been to the extent of our knowledge no model that takes into account the availability of amino-acids in the culture medium as a factor affecting either growth or productivity. However it is well known from biology that the building blocks for the synthesis of biological macromolecules are amino-acids and that not all of the amino-acids can be produced by the cell. Therefore seeking optimal feeding profiles based on the provision of glucose alone, and in the best case glucose and glutamine, might yield an increase in final titre on the one hand, however it is still quite far from the global optimum.

The visionary remarks of Bailey (1998) predicted the need to shift modelling focus upstream towards the gene level in order to truly understand the dynamics of cellular metabolism. Little – to no – work has been published since containing structure on the gene level where the kernel of the cell's control mechanism lies. 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. However from an engineering point of view, this information can only be utilised through a systematic and rigorous framework that will organise and prioritise necessary measurements and experiments.

Paving the way towards a "closed-loop" holistic framework for bio-process automation, this chapter covers the development of dynamical models of biological systems. The biological model development framework presented in figure (2) is explained in a step by step fashion, highlighting scientific concerns, challenges and "real life" problems associated with each step of the framework. Adapting a "real-life" example from the work of Kontoravdi (2006) we present the logical and systematic evolution of a model from conception to validation as it flows through the various steps of the model development framework. The key conclusion of this chapter is that by utilising a systematic way of organising available information, one can avoid conducting experiments for the sake of experimentation and develop models with an *a priori* set aim.

1.6 Cited Literature

[1] Bailey J.E., *Mathematical Modeling and Analysis in Biochemical Engineering: Past Accomplishments and Future Opportunities*, Biotechnol. Prog. 14, 8-20, 1998

[2] Batt B.C., Kompala D.S., *A Structured Kinetic Modelling Framework for the Dynamics of Hybridoma Growth and Monoclonal Antibody Production in Continuous Suspension Cultures*, Biotechnology and Bioengineering, Vol. 34, pp. 515-531, 1989

[3] Bibila T. A., Flickenger M. C., *A structured model for monoclonal antibody synthesis in exponentially growing and stationary phase hybridoma cells*, Biotechnol. Bioeng., 37, 210-226., 1991

[4] Bibila T. A., Flickenger M. C., *Use of a structured kinetic model of antibody synthesis and secretion for the optimisation of antibody product synthesis Part 1. Steady-state analysis*, Biotechnol. Bioeng., 39, 251-261, 1992

[5] Bibila T. A., Flickenger M. C., *Use of a structured kinetic model of antibody synthesis and secretion for the optimisation of antibody product synthesis Part 2. Transient analysis*, Biotechnol. Bioeng., 39, 262-272,1992

[6] Bray H.G., James S.P., Raffan I.M., Thorpe W.V., *The enzymic hydrolysis of Glutamine and its spontaneous decomposion in Buffer Solutions*, The Biochem, Journal, Vol. 44, 625-627, 1948

[7] Dalili M., Sayles G.D., Ollis D.F., *Glutamine-limited vatch hybridoma cell growth and antibody production: experiment and model*, Biotech. Bioeng., 36, 74-82, 1990

[8] De Tremblay M., Perrier M., Chavarie C., Archambault J., *Optimization of fed-batch culture of hybridoma cells using dynamic programming: single and multi feed cases*, Bioprocess Eng., 7, 229-234,1992

[9] 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, v. 9, 13-21, 1993

[10] deZengotita M. V., Miller, W. M., & Aunins, J. G., *Phosphate Feeding Improves High-Cell-Concentration NS0 Myeloma Culture Performance for Monoclonal Antibody Production*, Biotech. Bioeng., v. 69, 566-676, 2000

[11] Dhir, S., Morrow, K. J., Rhinehart, R. R., & Wiesner, T., *Dynamic optimisation of hybridoma growth in a fed-batch bioreactor.* Biotechnol. Bioeng, 67(2), 197-205, 1999

[12] DiMasi D., Swartz R.W., *An energetically structured model of mammalian cell metabolism. 1.Model development and application to Steady-State hybridoma cell growth in continuous culture*, Biotechnol. Prog., v.11(6), 664-676, 1995

[13] Eakman J.M., Fredrickson A.G., Tscuchiya H.M., *Statistics and Dynamics of microbial cell populations*, Chem. Eng. Prog., 62, 37-49, 1966

[14] 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, v. 26, 1-10, 2003

[15] Frame K.K., Hu W.S., *Kinetic Study of Hybridoma Cell Growth in Continuous Culture: I. A Model for Nonproducing Cells*, Biotechnol Bioeng, 37, 55-64, 1991

[16] Fredrickson, A. G., Mcgee, R. D., Iii, & Tsuchiya, H. M., *Mathematical Models in Fermentation Processes*, Adv. Appl. Microbiol, 23, 419, 1970

[17] 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, v. 22(6), 1560-9, 2006

[18] 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. Journal, 4, 153-168, 2000

[19] Kiparissides A., Kucherenko S., Mantalaris A., Pisitikopoulos E.N., *Global Sensitivity Analysis Challenges in biological systems modelling*, I&ECr, 2009

[20] 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., 21 (4), 1128 -1135, 2005

[21] 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

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

[23] Kompala D.S., Ramkrishna D., & Tsao G.T., *Cybernetic modeling of microbial growth on multiple substrates*, Biotechnol. Bioeng, 26, 1272-1281, 1984

[24] Kucherenko S., Rodriguez-Fernandez M., Pantelides C., Shah N., *Monte Carlo evaluation of Derivative based Global Sensitivity Indices*, accepted for publication, Reliab. Eng. Syst. Safety, 2007

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

[26] Lam M.C., Sriyudthsak K., Kontoravdi C., Kothari K., Park H.H., Pistikopoulos E.N., Mantalaris A., *Cell Cycle modelling for Off-line dynamic optimisation of mammalian culture*, Computer Aided Chemical Engineering, v. 25, 109-114, 2008

[27] Lee S.B., Bailey J.E., *A mathematical model for λdv plasmid replication: analysis of wild-type plasmid*, Plasmid, 11(2):151-65, 1984a

[28] Lee S.B., Bailey J.E., *A mathematical model for λdv plasmid replication: analysis of copy number mutants*, Plasmid, 11(2):166-77, 1984b

[29] Mantzaris N., Liou N.J., Daoutidis P., Srienc F., *Numerical solution of a mass structured cell population balance model in an environment of changing substrate concentration*, Journal of Biotechnology, 71, 157–174, 1999

[30] Mantzaris N., Daoutidis P., Srienc F., *Nonlinear productivity control using a multi-staged cell population balance model*, Chemical Engineering Science 57, 1–14, 2002

[31] Mantzaris N., Daoutidis P., Srienc F., *Numerical solution of multi-variable cell population balance models: I. finite difference method*s, Computers and Chemical Engineering, 25, 1411–1440, 2001a

[32] Mantzaris N., Daoutidis P., Srienc F., *Numerical solution of multi-variable cell population balance models: II. spectral methods*, Computers and Chemical Engineering, 25, 1441–1462, 2001b

[33] Mantzaris N., Daoutidis P., Srienc F., *Numerical solution of multi-variable cell population balance models: III. finite element methods*, Computers and Chemical Engineering 25, 1463–1481, 2001c

[34] Miller W.M., Blanch H., Wilke C., *Kinetic Analysis of Hybridoma Gorwth in Continuous Suspension Culture*, ACS Natl. Meet., Sept 11, 1986

[35] Miller W.M., Blanch H., Wilke C., *A kinetic analysis of hybridoma growth and metabolism in batch and continuous suspension culture: effect of nutrient concentration, dilution rate and ph*, Biotech. Bioeng., 67, 852-871, 2000

[36] Percy J.R., *A theoretical model for the covalent assembly of* immunoglobulins, J. Biol. Chem., 250(6), 2398-2400, 1975

[37] Ramkrishna D., *A cybernetic perspective of microbial growth,* In: Blanch HW, Papoutsakis ET, Stephanopoulos GN, editors. Foundations of biochemical engineering: kinetics and thermodynamics in biological systems. *Washington, DC: American Chemical Society.*, *p*, 161-178.

[38] Portner R., Schafer T., *Modelling hybridoma cell growth and metabolism – a comparison of selected models and data*, Journal of Biotechnology, 49, 119-135,1996

[39] Process Systems Enterprise, *gPROMS*, www.psenterprise.com/gproms, 1997-2009

[40] Saltelli A., Chan K., Scott E.M., *Sensitivity Analysis*, Wiley Press, 2000

[41] San, K. Y., & Stephanopoulos, G., *Optimization of fed-batch penicillin fermentation: A case of singular optimal control with state constraints*, Biotechnology and bioengineering, 34(1), 72-8, 1989

[42] Shuler M. L., Leung S., Dick C.C., *A MATHEMATICAL MODEL FOR THE GROWTH OF A SINGLE BACTERIAL CELL*, Annals of the New York Academy of Sciences, v. 326, 35-52, 1979

[43] Sidoli F.R., Mantalaris A., Asprey S.P., *Modelling of mammalian cells and cell culture processes*, Cytotechnology 44: 27-46, 2004

[44] 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., 44 (4), 868 -878, 2005

[45] Sobol' I.M., *On sensitivity estimation for nonlinear mathematical models*, Matem. Mod., v. 2, 112–118, 1990

[46] Sobol' I.M., *Global sensitivity idices for nonlinear mathematical models and their Monte Carlo estimates*, Math. and Comp. in Simulation 55, 271-280, 2001

[47] Sobol' I.M., Kucherenko S., *Derivative based Global Sensitivity Measures and their link with global sensitivity indices*, accepted for publication, Mathematics and Computer Simulations, 2008

[48] Srienc F., *Short communication: Cytometric data as the basis for rigorous models of cell population dynamics*, Journal of Biotechnology, 71, 233–238, 1999

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

[50] Tsuchiya, H. M., Fredrickson, A. G., & Aris, R. D., *Dynamics of microbial cell populations*, Adv. Chem. Eng, 6, 125-206, 1966

[51] Tziampazis E., Sambanis A., *Modelling of cell culture. Processes*, Cytotechnology, 14, 191-204, 1994

[52] Villadsen J., *Short communication: On the use of population balances*, Journal of Biotechnology, 71, 251–253, 1999

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

[54] Xie L., Wang D.I.C., *Material balance studies on animal cell metabolism using a stoichiometrically based reaction network*, Biotech. Bioeng., v.52(5), 578-590, 1996

[55] Xie L., Wang D.I.C., *Energy metabolism and ATP balance in animal cell cultivation using a stoichiometrically based reaction network*, Biotech. Bioeng., v.52(5), 591-601, 1996