Animal cell culture: macroscopic modeling, estimation and control

Animal cell cultures are targeted namely at the production of biopharmaceutical products such as vaccines, recombinant proteins and antibodies. The industrial production of these promising target-specific drugs is fairly recent. There are largely unexplored aspects that could be improved on so that a lower end-user price can be sustainable, namely, a more rational use of the culture medium.

In order to explore this, a model needs to be considered. In this thesis, we propose a practical approach of identifying model parameters with two experimental case studies (CHO-S and interferon-gamma producing CHO-320 cultures): a step-by-step parameter identification procedure that uses gradually more complex models and takes its inspiration from the analysis of the sensitivity functions.

The model is then useful for the control of a bioreactor. A focus is put on the continuous perfused production regime that allows a small volume to be cultured for a longer period and allows for a faster downstreaming. A controller can generate a manipulation of flowrates such that concentrations remain close to setpoints.

This research work then addresses the mathematical possibility of estimating some concentrations that cannot be measured in real life from the knowledge of other concentrations that can be measured online through probes currently available on the market. An illustration is provided via an implementation of the extended Kalman filter algorithm.

Finally, model-based automatic control is studied and the usefulness of nonlinear model predictive control highlighted. A model-plant mismatch case study serves to pinpoint the importance of how well biochemical kinetics are captured by the model in order to prevent persistent waste of culture medium.

Ines SARAIVA

Born in London, grown up in Portugal, Ines Saraiva is a chemical engineer from the University of Coimbra (5-year degree obtained in 2003, Master equivalence in 2008). She was a R&D engineer in the industry (process design, optimization and quality) and then started her research work in Biotechnology and Biosystems at the University of Mons, culminating with this thesis in 2015. Her main publications deal with modeling, estimation and control of animal cell cultures producing biopharmaceuticals. She is also a post graduate specialist in Drug Development Sciences (ULB, 2015).

Université de Mons
20, Place du Parc, B7000 Mons - Belgique
Tél: +32(0)65 373111
Courriel: info.mons@umons.ac.be
www.umons.ac.be

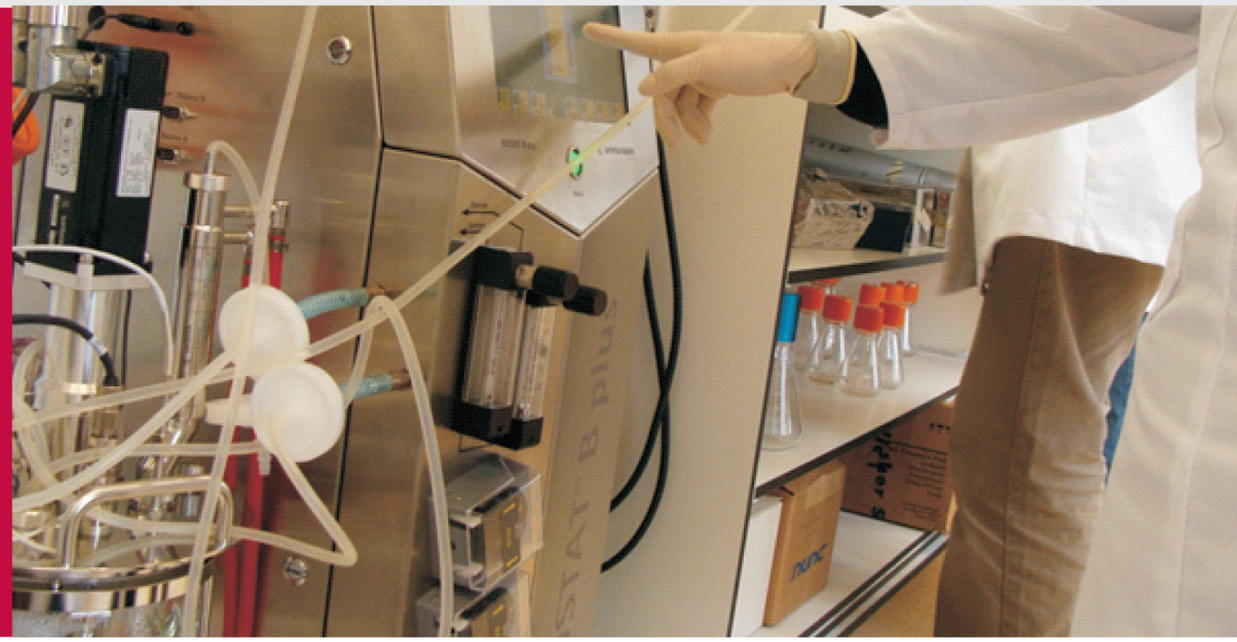
UMONS Université de Mons | POLYTECH MONS

Ines SARAIVA
Animal cell culture: macroscopic modeling, estimation and control
Thèse de Doctorat 2015

Thèse de Doctorat
2015

Animal cell culture: macroscopic modeling, estimation and control

Ines SARAIVA



UMONS Université de Mons | POLYTECH MONS

Animal cell culture: macroscopic modeling, estimation and control

Ines Saraiva
Research Assistant
M.Sc. in Chemical Engineering
(Universidade de Coimbra)

Ph.D. Thesis
submitted at the
Université de Mons
in fulfilment of the requirements for the degree of
Docteur en Sciences de l'Ingénieur

Jury members

Prof. Dr. Ir. M. Remy	(Université de Mons, Belgium) - *President*
Prof. Dr. Ir. L. Dewasme	(Université de Mons, Belgium) - *Secretary*
Prof. Dr. Ir. A. Vande Wouwer	(Université de Mons, Belgium) - *Advisor*
Prof. Dr. Ir. P. Bogaerts	(Université Libre de Bruxelles, Belgium)
Prof. Dr. Ir. A.-L. Hantson	(Université de Mons, Belgium)
Prof. Dr. Ir. O. Henry	(École Polytechnique de Montréal, Canada)
Prof. Dr. Ir. L. O. Santos	(Universidade de Coimbra, Portugal)

Acknowledgments

The years spent at the University of Mons have very honestly been some of the best of my life. Muito obrigada, Lino Oliveira Santos, da Universidade de Coimbra, pela janela de oportunidade que me abriste num momento em que vida se via cinzenta, pelo acompanhamento científico, a amizade e sorriso sempre presentes. Merci à mon promoteur, Alain Vande Wouwer, pour la confiance, audace scientifique, bonne humeur et pour toute son énergie pour faire en sorte que nous ayons une ambiance de travail au sein du Service d'Automatique de l'Université de Mons vraiment unique, scientifiquement et humainement. Merci infiniment à Monsieur Remy qui m'a accueilli au sein de ce service avec plein de gentillesse et qui a toujours su partager ses connaissances, expérience et esprit scientifique. Merci à Anne-Lise Hantson pour la coordination du vrai challenge qui a été le set-up du nouveau laboratoire de culture cellulaire. I would like to thank all members of the jury for all the scientific input and attention put in the follow-up of the research work culminating in this thesis and also the OCPAM project that made it possible.

I am closing a cycle of my life where I have learned a lot, developed a lot of know-how and gained immensely with all the people I've met. So, to all of you (including those I'm about to forget), thank you: Véronique, Vincent, William, Johan, Laurent, Christine, Renato, Carlos, Francisca, Cristina, Guillaume, Anne-Catherine, Mihaela, Edmundo, Lamia, Daniel, Thomas, Razvan, Radouane, Christian, Christophe(s), Alex, Romina, Andrés, Laurent(s), Julián, Gerardo, Bob, Alejandro, Jaime, Jonathan, Zakaria, Magali, Lionel, David, Joelle, Liza, Louise, Marie-Eve, Amaury, Aldo, Henri, Michel, Christelle, Charlotte and not least... Pedro, Sofia, Daniel, Paul, Giannina, Patrick, Micaela, Jimmy, Bruna, Guilherme, Marita, Fer, Gilles, Chris (our favorite thesis proof-reader!), Arnaud, Amy, Ciler, Per, Mauricio, Maria, Benoit, Onur, Danielle, René, Stéphanie, Ibeliz, Roxanne. Thank you Marc, Pierre(s) and Kris, biking buddies, Sylvie et tous les copains de la voile, Nounou, Nico and Titi, voor de lekkere pintjes rond Brussel, Catarina, Pedro, Carmen,

Susana, Artur, Hélder, Ricardo, Milita pela alma portuguesa com certeza, Georges, Élodie, Olivier, Natacha, Térence, Sylvie, Jacques, Marco, Thomas, Stéphane de YouFM, la radio de l'UMons, ceux de Radio Campus, Maxime et ceux de la scène musicale que j'ai pu retrouver ici, Isabelle et les chouettes volontaires de la Croix-Rouge.

Obrigada à minha família, dispersa pelos cantos deste planeta e na eternidade. Não sou de nenhum lugar, o meu país é o vento e os caminhos do mar, disse um dia Teolinda Gersão. À la fin, merci aussi à ce charmant pays d'accueil.

Summary

Context Animal cell cultures are targeted at the production of biopharmaceutical products such as vaccines (eg. rotarix, polio, smallpox), recombinant proteins and antibodies (eg., monoclonal antibodies, interferon-γ).

The complexity of these biomolecules is such that production through common chemistry is difficult, if not impossible. These substances can, however, sometimes be synthesised by cells programmed (eg. by transfection) to produce them, hence the common use of the name biologicals. Sometimes this can be done using cells with a genetic information sufficiently close to that of humans (eg. mammals). Animal cells are cells extracted typically from tissues of organs of animals. For example, CHO (Chinese Hamster Ovary) are the most commonly used mammalian cells in this field, being known for their capacity to correctly fold and post-translationally modify recombinant proteins compatible with humans (Kildegaard et al, 2013).

As these biopharmaceutical products have a growing demand, the quest is on to seek better production processes in terms of quality, quantity and end user price.

For process optimization and control, these cultures can be described by mathematical models that estimate the evolution of the concentrations of biomass (cells), the substrates they are fed with (eg. glucose and glutamine), the product of interest and other metabolites they produce in the course of the culture and which may affect (eg. inhibit) their own growth (eg. lactate, ammonia). These models allow to predict culture behaviour and to study, monitor and control different production scenarios.

Motivation This thesis focuses on a field where much is still to discover, namely the study of cultures of animal cells, such as CHO, HEK or hybridoma, in suspension, common in industrial production (Zhang, 2010). Thus, cultures where the cells are suspended in a culture medium capable of providing them with the substrates that they need in order to grow, multiply and even-

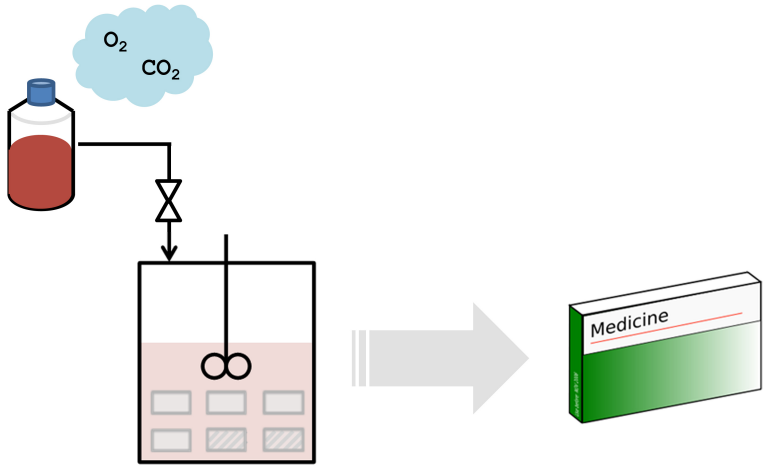


Figure 1: Illustration of the general idea behind an animal cell culture.

tually synthesise bioproducts of interest, typically around body temperature (37°C) inside a bioreactor.

More specifically, this thesis focuses on studying an interesting and still largely unexplored production regime called continuous perfused since a perfusion filter is placed in an outstream. This filter is meant to retain cells inside a bioreactor while guaranteeing that some culture medium containing the bioproduct can be taken out and sent to a purifying downstreaming unit. This operation regime is thought for the operation of small volumes during a long period of time (eg. one to six months) and is already used in industrial practice (Boedeker, 2013; Chu and Robinson, 2001), particularly for cases when one or more of the following factors occur (Drugmand, 2011):

- cell growth is somewhat slower than average;
- bioproduct production is not completely growth-associated;
- the bioproduct stability is somewhat limited and, therefore, it is best to send it as fast as possible to the downstreaming purifying unit[1];
- another component may degrade fast and, therefore, a production regime with a slower residence time is preferable (ie. perfusion is a better choice than fedbatch/batch);
- inhibition phenomena are important, and, for the same reason, a smaller residency time is preferable;

- there is limited space in the sterile production zone;
- initial investment is bugdet-limited[2].

This thesis seeks also to address the question of a management of culture medium that is more intelligent and less expensive, that could thus allow to perform a more efficient culture where, for example, not a lot of unused substrates would be detected in the outstream (a medium able to provide nutrients in a dynamical manner in function of cell's evolving needs, for example). If the industrial production of these biologicals is improved, then the end-user price of precision medicine such as this can, potentially, become more sustainable and allow market access to innovative medicines to be broader.

Thesis organisation and contribution This thesis is organised as follows.

Animal cell cultures are introduced in Chapter 1, with a presentation of what a cell culture looks like in terms of the timecourse of concentrations of components and how its behaviour can be captured in a snapshot, along with a motivation for modeling.

Chapter 2 reviews approaches for modeling. The choice of unsegregated macroscopic models, in the scope of this thesis, is stated and reasons for doing so are presented. Briefly, this type of models presents a good compromise between descriptive insight and simplicity for practical purposes.

Once a model structure is chosen, the values of the parameters of these models still need, in real practice, to be identified. One of the major contributions of this thesis is to be found in Chapter 3, where, first of all, the identification problem is presented. Starting the procedure with a good initial guess for a large set of parameters represents a challenge. In order to tackle this challenge, a procedure is proposed. It is a step-by-step identification approach which gradually considers more detailed models. To exemplify it, a very simple model is first identified for a limited subset of experimental data. This allows to roughly determine the value of one parameter. The next step considers more states and parameters and a still limited data subset. Hence, it is more complex but, on the other hand, a good initial guess of

[1]This is the case eg. of recombinant plasma-based factor VIII (a clotting factor missing from the blood of people with hemophilia A) produced in perfusion by Baxter Belgium and Bayer Healthcare.

[2]A smaller bioreactor is needed, in comparison to the high initial costs of a large fedbatch/batch bioreactor. On the other hand, the perfusion operation requires a bigger investment on workers expertise.

one of the parameters is already roughly known. Other steps follow, with increasing model complexity, where the identification algorithm has gradually more and more knowledge about the initial guess for the parameter set. In the end, the identification procedure is terminated whenever the user decides that he accepts a particular model structure and is happy with both its descriptive quality (fit to whole dataset) and its qualitative power (phenomena such as limitation or inhibition that the model is robust enough to account for). Chapter 3 also presents the rationale supporting the proposal of this identification procedure. The inspiration came from an analysis of how sensitive model states are when a value of a model parameter is changed. A simulation case study using a well known animal cell culture model from the literature is used to illustrate this in the beginning of the chapter. It is then that real data from experimental campaigns performed at the University of Mons is used to illustrate the step-by-step identification proposal in two experimental practical case studies.

Once a model is identified and available for use in a real scenario, the question of observability is then looked into. Chapter 4 illustrates how such type of models can be used by sofware sensors (observers) that allow to estimate the timecourse of concentrations not being measured in reality. More precisely, in a practical scenario, the equipment available may not allow to measure all the concentrations needed to solve the model. An observer will, however, allow to estimate online how unmeasured concentrations evolve. This is done given both the knowledge of the model and the information about the variables being measured. The property dealing with this being mathematically possible is called observability. The most common approaches to study this property may sometimes not be very helpful with this type of animal cell culture models. In this chapter, a contribution on a manner of studying the observability of animal cell culture models is also presented. It is based on recent works of Moreno et al (2014) that address the indistinguishability properties of the system.

Chapter 5 focuses on control, starting off with a study of a control scheme already presented in the literature for a very simple biomass-substrate model describing animal cell cultures in continuous perfused operation. Its adaptive backstepping control strategy is compared to one that can be achieved by a nonlinear model predictive controller (NMPC). NMPC is found to be a promising control strategy. Next, an illustration of how NMPC can be applied to a model of higher complexity (5 to 7 states, such as the ones chosen in the scope of this thesis) is shown[3]. Finally, the question of model

choice is tackled: how complex should a model be in a continuous perfused production scenario. Model-plant mismatch is studied in a case study with different models identified for the same databank of interferon-γ producing CHO-320 cell cultures performed at the University of Mons (and previously listed in Chapter 3). The importance of having kinetics correctly captured is highlighted.

Finally, a global round-up of the thesis contributions and some future prospects are presented in Chapter 6.

[3]This study is further developed in Sbarciog et al (2013) with a study that addresses controller tuning and robustness analysis.

List of publications

Journal articles

I. Saraiva, A. Vande Wouwer, A.-L. Hantson, Parameter identification of a dynamic model of CHO cultures. Bioprocess and Biosystems Engineering (2015), submitted.

M. Sbarciog, I. Saraiva, A. Vande Wouwer, Accelerating animal cell growth in perfusion mode by multivariable control: Simulation studies. Bioprocess and Biosystems Engineering (2013), 36(5):517-530.

Conferences with proceedings

I. Saraiva, A. Vande Wouwer, A.-L. Hantson, J. Moreno, Observability analysis and software sensor design for an animal cell culture in perfusion mode. In proceedings of the 12th IFAC Symposium on Computer Applications in Biotechnology (2013).

Conferences without proceedings

I. Saraiva, A.-L. Hantson, A. Vande Wouwer, Identification of Animal Cell Culture Models: a sequenced approach. Matinée des Chercheurs (2013), Mons (Belgium), poster.

I. Saraiva, M. Sbarciog, A. Vande Wouwer, Practical Design of an Extended Kalman Filter for an Animal Cell Culture. 31st Benelux Meeting on Systems and Control (2012), Heijen (Netherlands).

I. Saraiva, M. Sbarciog, A. Vande Wouwer, Multivariable Control of Animal Cell Cultures in Perfusion Mode. 30th Benelux Meeting on Systems and Control (2011), Lommel (Belgium).

I. Saraiva, M. Sbarciog, D. Coutinho, A. Vande Wouwer, Performance and robustness assessment of multivariable control strategies for a high-cell-density perfusion bioreactor. IAP DYSCO Study Day (2011), Leuven (Belgium), poster.

I. Saraiva, M. Sbarciog, A. Vande Wouwer, Multivariable Control of Animal Cell Cultures in Perfusion. Matinée des Chercheurs (2011), Mons (Belgium), poster.

I. Saraiva, L. O. Santos, A. Vande Wouwer, Nonlinear model predictive control of animal cell cultures in perfusion mode. 29th Benelux Meeting on Systems and Control (2010), Heeze (Netherlands).

I. Saraiva, A.-L. Hantson, A. Vande Wouwer, L. Santos, Control of Animal Cell Cultures. 29th Benelux IAP DYSCO Study Day (2009), Leuven (Belgium), poster.

Contents

List of Figures

List of Tables

Nomenclature

Acronyms and abbreviations used (components)

AA	Amino acids
Ala	Alanine
Amm	Ammonia
Arg	Arginine
Asp	Aspartic acid
Cys	Cysteine
Glc	Glucose
Gln	Glutamine
Glu	Glutamic acid
Gly	Glycine
His	Histidine
$IFN-\gamma$	Interferon-gamma
Ile	Isoleucine
Lac	Lactate
Leu	Leucine
Lys	Lysine
MAb	Monoclonal antibodies
Met	Methionine
Phe	Phenylalanine
Pro	Proline
Prot	Generic protein (product)
Ser	Serine
Thr	Threonine
Tyr	Tyrosine
Val	Valine
Xd	Dead biomass
Xv	Viable living biomass

Acronyms and abbreviations used (others)

CHO	Chinese hamster ovary (cells)
DMEM	Dulbecco's Modified Eagle Medium
EKF	Extended Kalman filter
FBS	Fetal bovine serum
FMI	Fisher information matrix
GMP	Good manufacturing practices
IgG	Immunoglobulin G
IgM	Immunoglobulin M
LB	Lower bound
NMPC	Nonlinear model predictive control
UB	Upper bound
UMons	University of Mons

Variables

$a_{i,j}$	generic element of a matrix
Amm	ammonium concentration
$b_{i,j}$	generic element of a matrix
c_i	logistic equation parameter
D	dilution ratio
$f(t)$	generic function
F_i	flowrate
Glc	glucose concentration
Gln	glutamine concentration
J	cost function
k_{S_i}	limitation constant of substrate S_i
k_{d,P_i}	limitation constant of product P_i
k_{P_i}	inhibition constant of product P_i
Lac	lactate concentration
MAb	monoclonal antibodies concentration
m_{Glc}	specific cell maintenance rate for glucose
O	observability matrix
p	prediction horizon
P_i	product i concentration
q	observability map (in the observability chapter)
q	prediction horizon (in the control chapter)
Q_{ij}	measurement covariance vector
S_i	substrate i concentration[1]
S_{x_i,θ_j}	sensitivity of state x_i to parameter θ_i
t	time
$t_{95\%}$	t-student value for a level of 95%
t_{peak}	switch time between cell growth and cell death in batch
u_j	manipulated variable
V	volume
V_i	Lyapunov function
w_i	weighting factors of the cost function
x	state variable
x_{meas}	measured state variable
Xd	dead cell concentration
Xv	viable cell concentration
$y(x)$	generic function
$Y_{a/b}$	b-to-a yield coefficient
z_i	homologue variables

[1] on subsection 2.1.2.6 on p.61 it is used for "Serum" instead of "Substrate"

Variables

γ_i	adaptation gain
ε_i	deviation variables
θ_i	generic parameter i
θ^0	initial guess of the set of parameter values
θ^*	optimal set of parameter values
μ	specific rate
μ_d	specific death rate
μ_g	specific growth rate
μ_m	specific maintenance rate
μ_{max}	maximum specific growth rate
ν	stoichiometric coefficient
σ	standard deviation
τ	age
φ	reaction rate
Ω_i	generic model i
$\mathcal{H}_i$	hypothesis i

Chapter 1

Introduction

Animal cell cultures are introduced in this chapter, with a presentation of what a cell culture looks like in terms of the timecourse of concentrations of components and how its behaviour can be captured in a snapshot, along with a motivation for modeling.

Models are of great importance in the study of better manners of performing animal cell cultures. In order to establish a model some steps are necessary:

- Formulation of kinetic relations;
- Establishment of balances;
- Parameter identification (fitting);
- Model validation;

Through observation of dominant kinetic phenomena taking place, the conversion of substrates into products and how they affect the evolution of living cell concentration throughout a culture can be better understood. Cell growth can typically be limited by some substrates, when they become scarce, and by some products, when they build up excessively. Some substrate

consumption can also be sometimes attributed not to cell growth but to cell maintenance purposes.

This knowledge, coming partially from theoretical background and from observation, is summed up into balance equations (mass balances: component-wise and total). The values of the parameters present in these equations still need to be identified, such that the model structure chosen fits well enough with experimental data and provides a validated prediction performance in different culture scenarios.

The purpose of suspending cells in a culture medium is in some way to replicate the environment that they would naturally have before having been extracted from the animal, such that they will stay alive for the duration of the culture. Typical candidates for system states are, thus, the concentration of the bioproduct excreted if one is produced (usually measured offline after filtration), cell concentration (biomass), substrates consumed and metabolites produced.

With regard to kinetics, typical states are, thus, the concentration of living and dead biomass, major substrates such as glucose and glutamine, and metabolites such as lactate and ammonia, aminoacids and the bioproduct.

According to the production regime chosen, the model also describes fluid dynamics phenomena by considering variables such as the flowrates of streams being fed or taken out of the bioreactor.

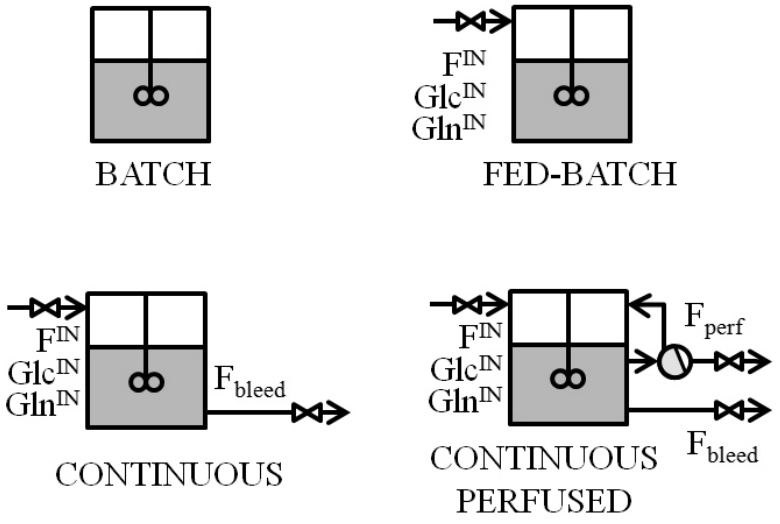


Figure 1.1: Layout according to operation regime.

Figure 1.1 shows different production regimes. When operating in batch mode, besides the culture medium initially placed in the bioreactor where

cells are inoculated, no stream is fed or taken out. Cell concentration grows and, after some time, it decreases mostly because substrates become scarce and/or some inhibiting metabolites have built up.

In fedbatch mode, cell growth is accompanied by an input of fresh medium, such that the duration of the culture can be extended.

When an outstream is added, a continuous regime can be operated, generally with a smaller volume. In order to keep the culture volume constant, the incoming flow equals the outgoing flow, with the constraint that cell concentration must at least be maintained (otherwise the biomass is washed out of the bioreactor simply because medium renewal is too fast for the speed with which biomass is actually growing).

Longer cultures are generally operated in continuous perfused regime. In this case, the main outstream lets out all components apart from living biomass that is kept inside by means of a perfusion filter. A small bleed outstream allows a better control of the desired cell concentration.

Figure 1.2 provides an example of concentration profiles obtained for a batch culture of CHO-S performed at the UMons. Biomass concentration initially rises in an approximately exponential manner while the major substrates glucose and glutamine are consumed. At a given moment around $t = 90h$, glutamine is extinguished and biomass begins to decrease. While cells were consuming glucose, they were also producing lactate. In parallel, consumption of glutamine is associated with ammonia production.

Figure 1.3 shows how during the same culture several other species are either consumed or produced, albeit in a smaller scale. Metabolic studies often focus on finding out more insight about the role that these aminoacids may have in a culture.

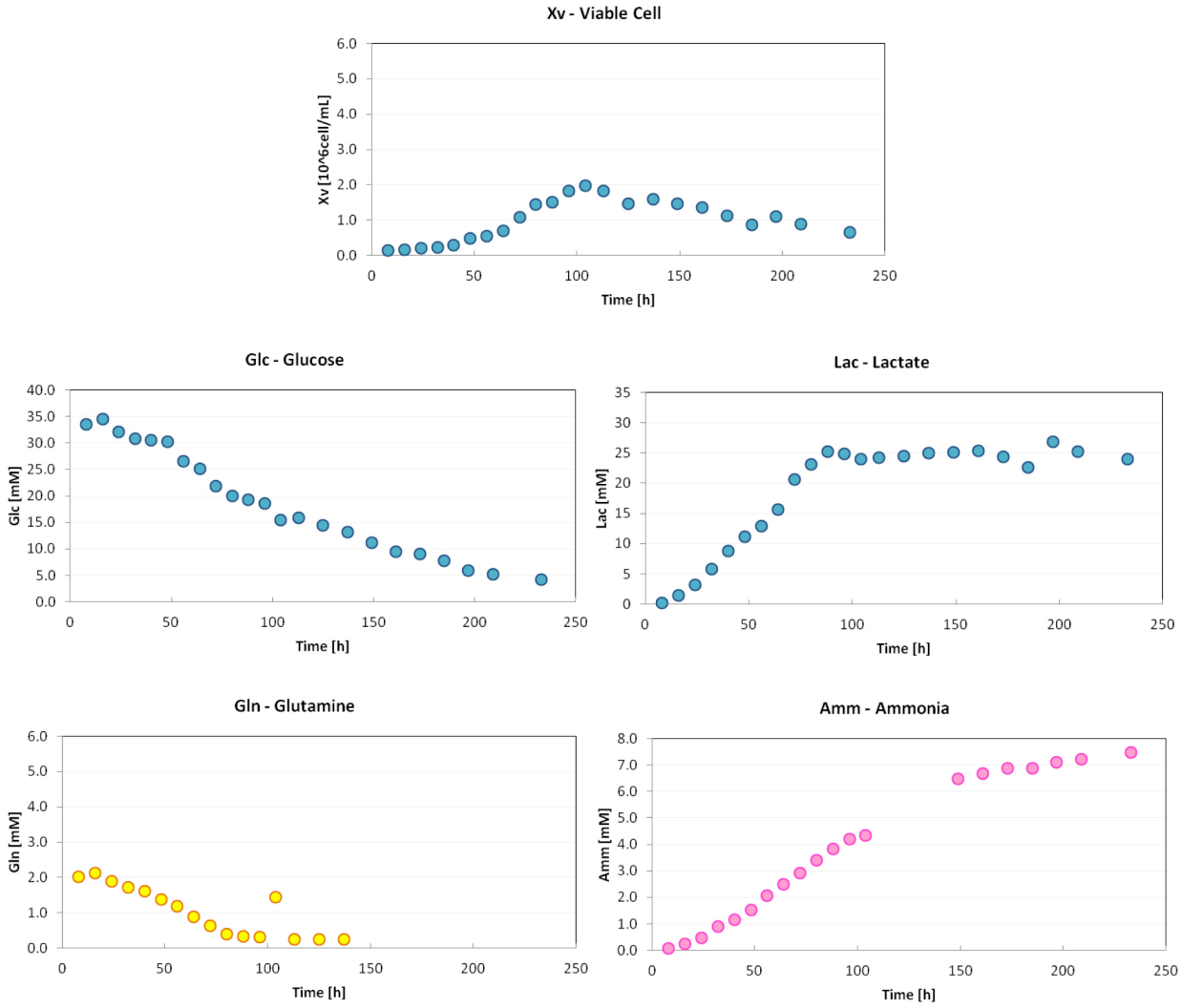


Figure 1.2: CHO-S batch culture performed at the UMons cell laboratory: concentrations of major components.

Figure 1.4 illustrates different ranges of variation regarding the concentrations of substrates and products in a culture. Those denominated major substrates and metabolites have concentration ranges of several dozen mM (here about 0-40mM for the pair glucose/lactate and 0-10mM for the pair glutamine/ammonia). Minor scale components, such as the aminoacids depicted in Figure 1.3, are typically detected below 3mM. For living cell concentration, usual operation involves some millions per mL, depending mostly on if boosters such as serum are present or not. Time-wise, a common batch duration may be one or two weeks, a fedbatch may last a week longer, and continuous or continuous perfused cultures may last some months (in theory, longer).

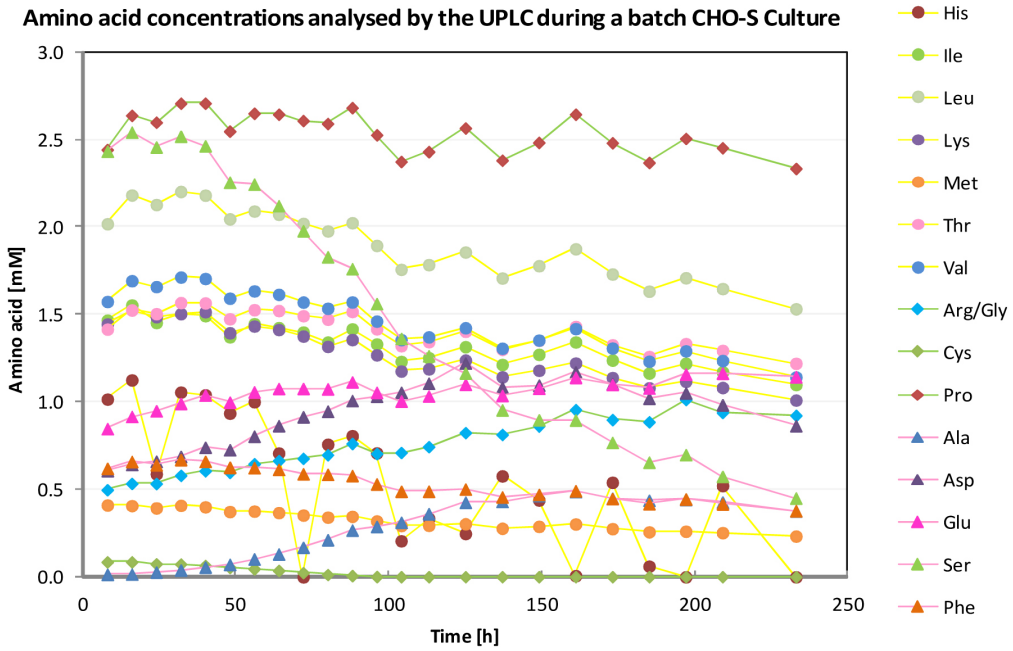


Figure 1.3: CHO-S batch culture performed at the UMons cell culture laboratory: concentration of smaller-scale components.

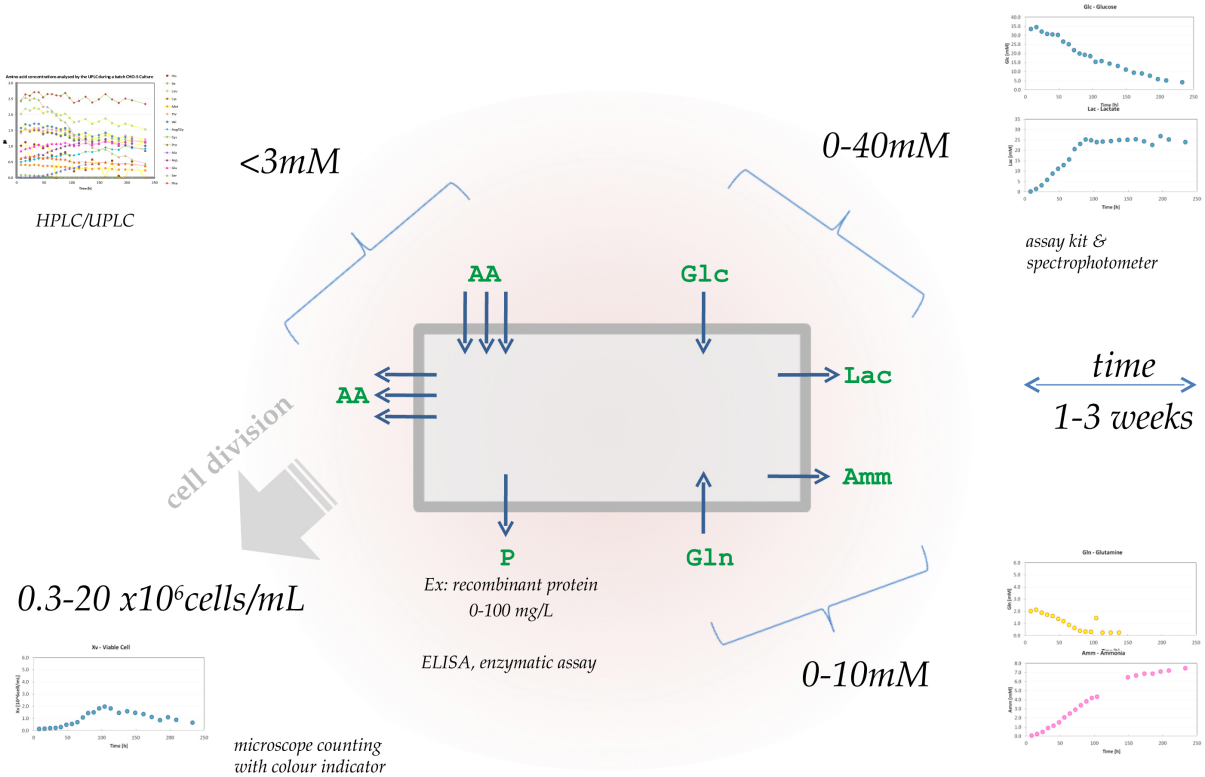


Figure 1.4: CHO-S culture: different ranges of concentrations of components.

Typically, cell concentration is estimated offline by sample counting under the microscope using a colour indicator (trypan blue exclusion) in order to distinguish between those living from dead. Major substrates and metabolites can be measured offline by assay kits using a spectrophotometer. The product of interest, the biopharmaceutical (eg. with concentrations between 0-100mg/L), can usually be measured offline with an ELISA kit.

1.1 Purpose of modelling

Models come both from biological prior knowledge and from observation of reality. Biology proposes several possible reaction pathways taking place inside a cell. This list can be quite extensive, with several hundred reactions proposed, such as those described in Zamorano et al (2010) and Zamorano (2012). Thus a choice to be made is the degree of complexity one wishes to include while modeling. This leads to the consideration of the various purposes of the model. One possible purpose is the description of a simple data set planned for tasks such as prediction. Optimization, control and monitoring can be others. The following subsections will provide illustrations of these purposes.

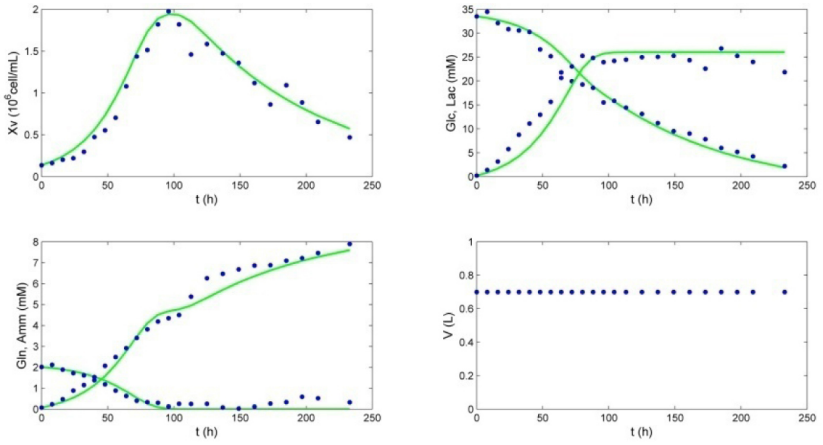
1.1.1 Description and prediction

A model describing a data set such as in Figure 1.4 can be meant to simulate hypothetical scenarios. For example, what would happen if the culture were to be initiated with half the amount of glutamine present on the medium (in this case, overall growth would be lower) or half as much (higher).

1.1.2 Optimization

Optimization is another motivation for modeling. For example, estimating which particular operating conditions would correspond to the best value for a certain criterion, such as the maximum biomass/product produced or concentration attained.

description



prediction

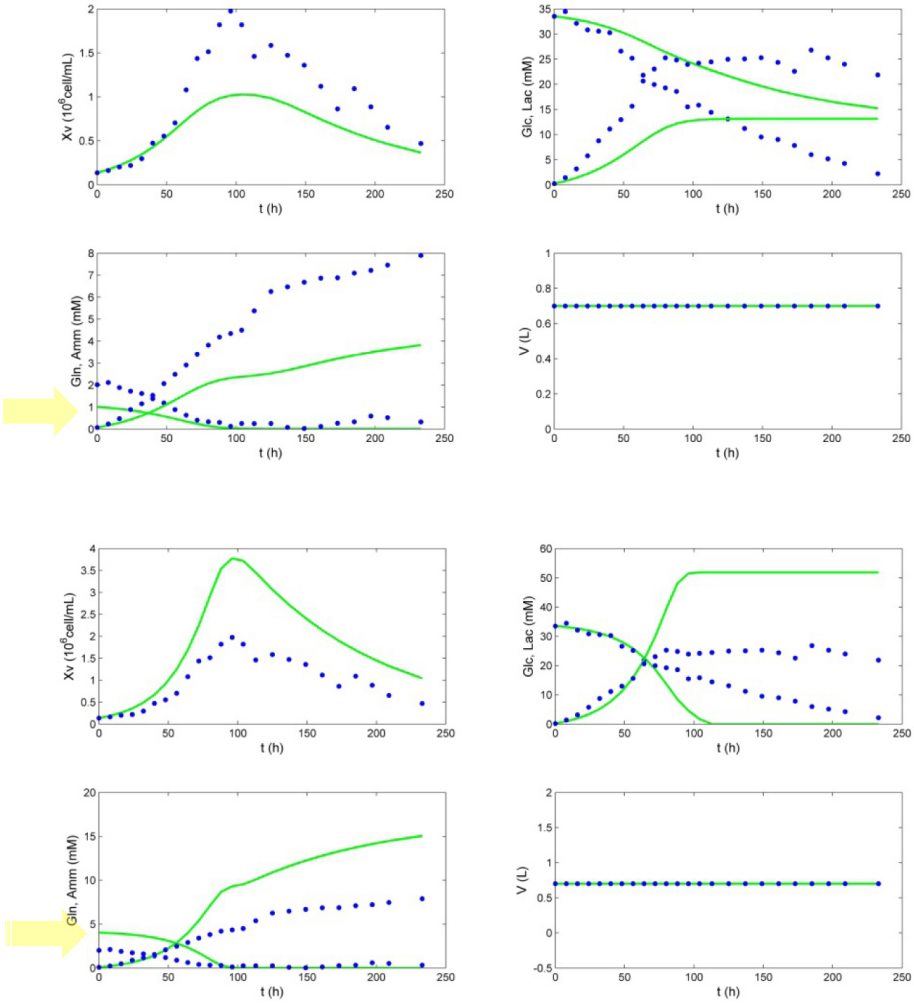


Figure 1.5: Model predictions and data for a CHO-S batch culture performed at the UMons cell laboratory: concentrations of living cells (Xv), glucose (Glc), lactate (Lac), glutamine (Gln), ammonia (Amm) and volume (V). On top: use of the model for descriptive purposes. Bottom: use of the model for prediction purposes (what would happen with half or twice that glutamine medium concentration.)

Figure 1.6 illustrates one such example: given about 2 weeks and around 2L of medium, examining which operation regime could seem more appealing.[1]

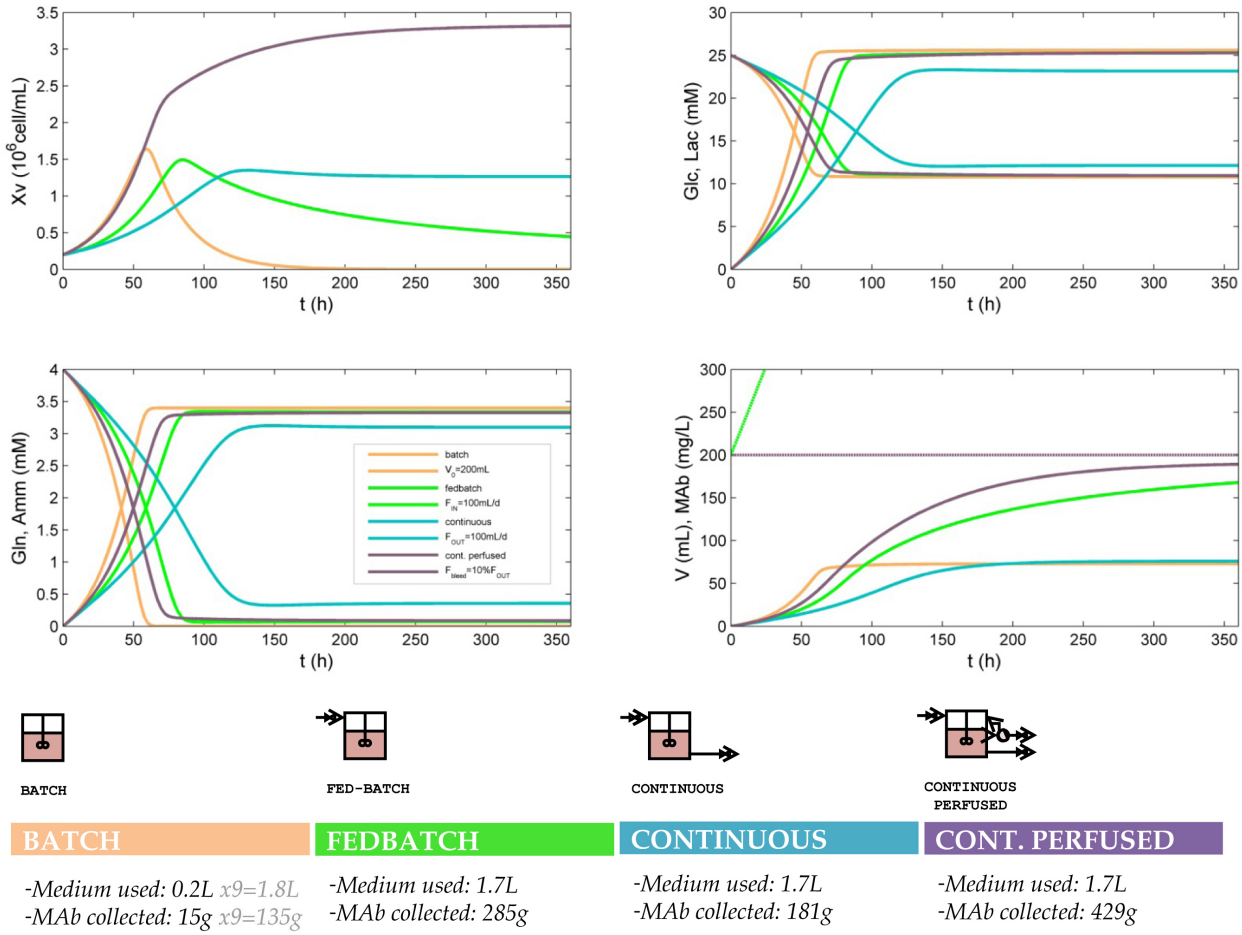


Figure 1.6: Model predictions: concentrations of living cells (Xv), glucose (Glc), lactate (Lac), glutamine (Gln), ammonia (Amm), monoclonal antibodies (MAb) and volume (V). Model used: de Tremblay et al (1992), 15-day time window comparison.

One could place the medium inside a bioreactor, inoculate it with cells and perform a batch, obtaining the orange profiles and, after 2 weeks, having collected 135g of antibodies for 1.8L of medium used. One could also consider using the available medium for a fedbatch where 200mL are gradu-

[1]Values used in the simulation: $V_0 = 200mL$ for batch, $[V_0 \ F^{IN}]^T = [200mL \ 100mL/d]^T$ for fedbatch, $[V_0 \ F^{IN} \ F_{bleed}]^T = [200mL \ 100mL/d \ 100mL/d]^T$ for continuous, $[V_0 \ F^{IN} \ F_{bleed} \ F_{perf}]^T = [200mL \ 100mL/d \ 10mL/d \ 90mL/d]^T$ for continuous perfused considering the nomenclature used in Fig. 1.1 on p. 31.

ally supplemented with additional medium as cells start growing. The total amount of antibodies collected would be 285g, for 1.7L of medium used. A pure continuous regime would lead to 181g collected, and if a perfusion filter were to be placed, a bigger quantity would be obtained: 429g.

In this example, the model was used for insight for the early choice of the operation regime to implement in the future.

1.1.3 Monitoring and control

Thirdly, another example of a motivation for modeling relates to its use for control and monitoring of the system. Let us consider that at a given moment the desired setpoint for biomass concentration in a continuous perfused culture needs to be changed while keeping the substrate concentration unchanged. In Saraiva et al (2010) the model presented in Deschenes et al (2006b) was used with this goal. A model predictive controller was implemented as in Santos et al (2010, 2012) and proposed a set of manipulations in the inflow and outflow rates such that the setpoint concentrations of biomass and substrate would be attained, as can be seen in Figure 1.7.

1.2 Aspects of animal cell behaviour

As mentioned previously, biology lists an incredible number of possible reaction pathways taking place in a cell or a bacteria. Figure 1.8 is an illustration of how intricate this network can be. In the case of CHO cell metabolism, for example, Zamorano (2012) considers around 100 intracellular reactions. Nevertheless, global metabolism can be significantly simplified for modeling purposes. Figure 1.9 presents an abridged overview of the metabolism of animal cells.

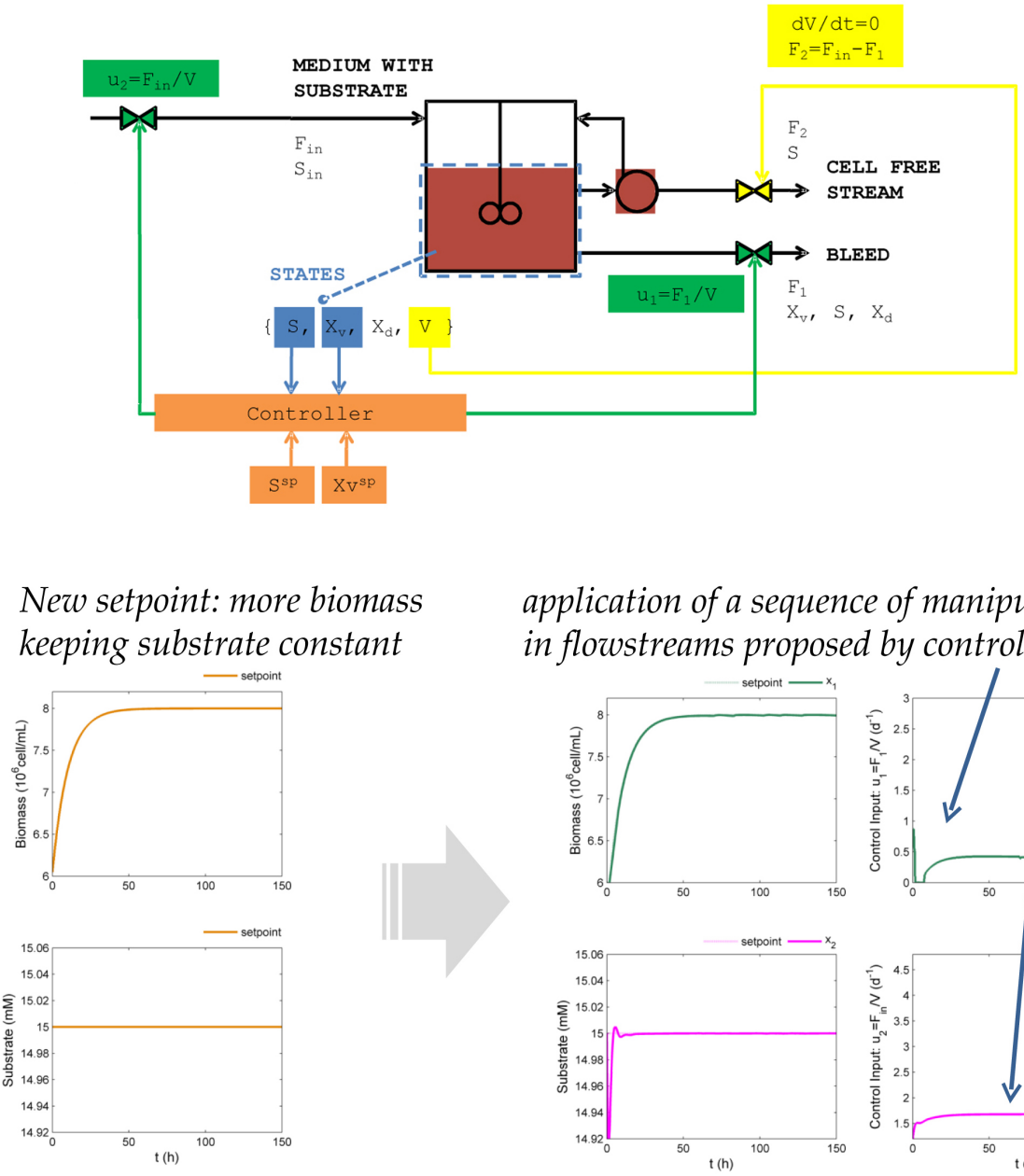


Figure 1.7: Model predictive controller set on a constant volume continuous perfused culture: setpoint and measured concentrations for biomass and substrate, manipulations of inflow rate F_{in} and outflow rate F_1. Source: Saraiva et al (2010). Model used: Deschenes et al (2006b).

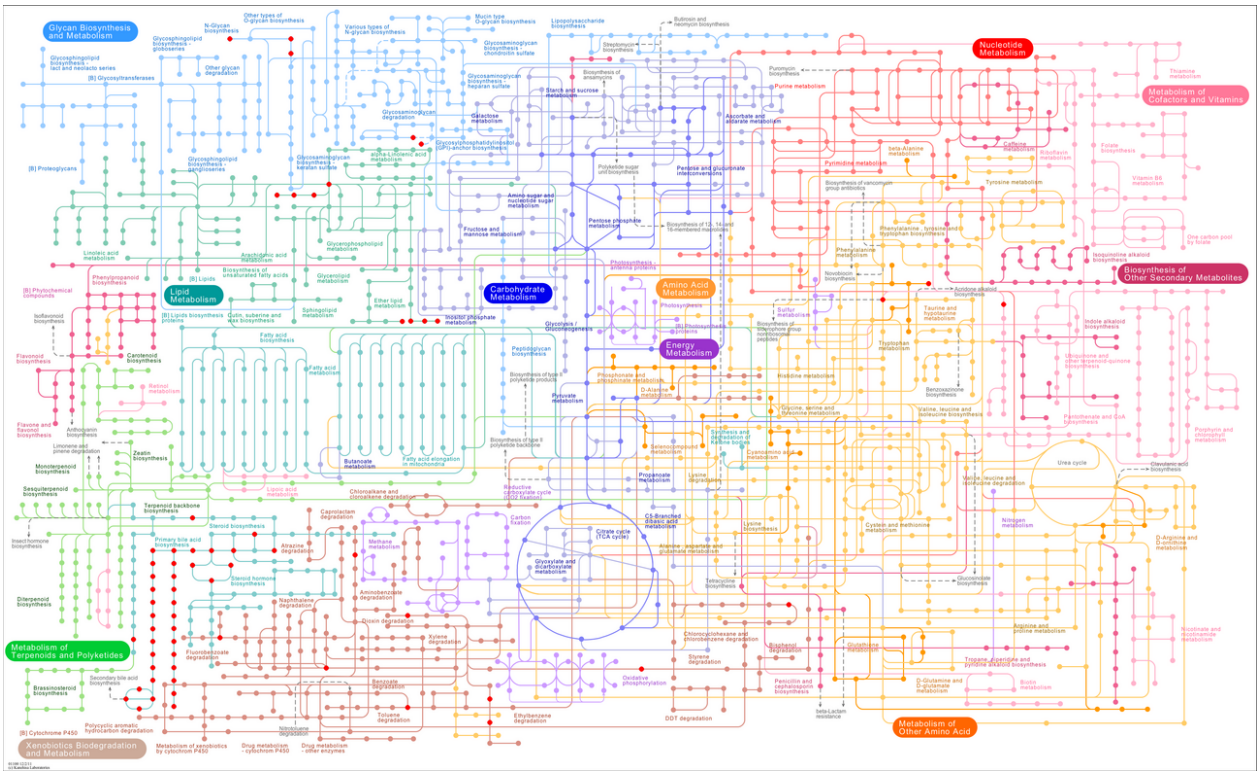


Figure 1.8: Metabolic pathways for Bacillus anthracis (strain A0248). Source: Kyoto Encyclopedia of Genes and Genomes (KEGG) Database.

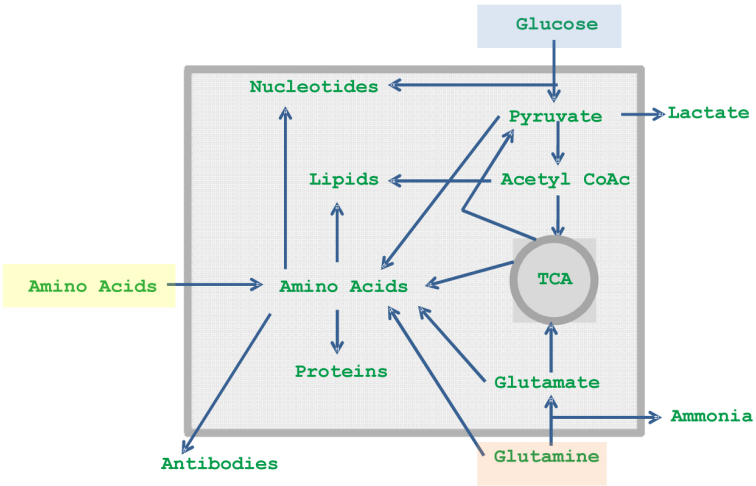


Figure 1.9: General simplified metabolism of animal cells. Adapted from Batt and Kompala (1989).

Concerning substrates, at least two are vital. One is usually glucose, a carbon source, entering the cell and following five types of pathways: pentose

pathway, lipid formation, lactate, amino acids and Krebs cycle. Another vital substrate is a source of nitrogen, typically glutamine, which contributes to the formation of amino acids nucleotides, proteins and lipids.

A particular feature of animal cells in comparison to microorganisms such as bacteria and yeast is the fact that, because, from an evolutionary perspective, they are not autonomous living beings, they are not able to produce all the amino acids that they need. These must then necessarily be fed through the culture medium.

1.2.1 Glucose

Regarding glucose consumption, it is known that the rate of glycolysis is usually much faster than the rate of utilisation of glycolytic intermediates. Therefore, most glucose is metabolised to lactate, a product that may inhibit cell growth (Batt and Kompala, 1989; Glacken et al, 1986; Kovacevic et al, 1991; Ljunggren and Lena, 1992; Miller et al, 1988a; Reitzer et al, 1979; Zielke et al, 1976).

This seems to indicate that a state of overflow is happening frequently during cell cultures, ie, the incomplete oxidation under aerobic conditions of an abundant energy source resulting in the excretion of often inhibitory metabolic byproducts. Since glucose is a cheap substrate, one would then question the risks of the associated lactate production being inhibitory to cell growth. Miller et al (1988a) state that lactate is less inhibitory than ammonia to cells at constant pH. They add that human hybridoma in batch culture have been found not to be affected by the addition of 4.9mM lactate, and that no lactate inhibition was detected at 40mM. A mouse hybridoma line has been found to suffer no inhibition from added lactate up to 22mM but to suffer from it above 28mM. Critical levels for myeloma cell growth inhibition have been indicated to be 40mM for lactate by Simpson et al (1998) and Zhou et al (2006). Researchers like Provost et al (2006) and Ljunggren and Lena (1994) observed that produced lactate can be reconsumed by cells when glucose is extinguished.

The scale of phenomena such as lactate inhibition varies from cell line to cell line. But globally, it seems that levels above 40mM lactate can be undesired. In reality this seldom happens in cultures. Additionally, it seems also that lactate can provide an *in extremis* carbon source if needed.

1.2.2 Glutamine

Glutamine, the more common nitrogen source in the medium, is also subject to overflow. In this case, leading to the formation of ammonia (and eventually also some extra lactate and amino acids such as alanine, proline, aspartic acid, glutamic acid, serine and glycine) as described by several authors (Amribt, 2014; Batt and Kompala, 1989; Glacken et al, 1986; Kovacevic et al, 1991; Lee et al, 2003; Ljunggren and Lena, 1992; Miller et al, 1988a; Reitzer et al, 1979; Zielke et al, 1976).

In fact, glutamine is usually added to the medium just before starting the culture since it is an unstable molecule that degrades spontaneously, particularly at culture temperature ($37°C$). Some authors have estimated this spontaneous decomposition to be as high as 11% per day at room temperature.

In most batch cultures, glutamine is the first major substrate extinguishing and triggering cell death. Fedbatchs extend culture time by sporadic supplement of glutamine concentrate.

Since glutamine is a more expensive substrate and more delicate to store than glucose, one can then ask how risky this inhibition effect of cell growth can be. Miller et al (1988b) state that the inhibitory ammonia concentration varies substantially among types of animal cell lines and it depends also on whether serum is used. They report ammonia inhibition at 5mM for human hybridoma in batch. Other authors indicate the same concentration (Simpson et al, 1998; Zhou et al, 2006). Indeed, ammonia can easily reach a 5mM concentration in common batch and fedbatch cultures. It is thus an important metabolic aspect to consider while modeling.

1.2.3 Amino acids

Amino acids, present on a smaller scale, can be divided into 3 categories. Non-essential are those that cells can synthesise, while conditionally essential are those that may be produced in certain circumstances (for example, CHO used in Provost (2006) have been programmed to produce a protein, and during the transfection, the cell line became auxotrophic with respect to proline, ie. lost the proline production path and became dependant on its external supply (Zamorano et al, 2009). Finally, essential amino acids are those that cells cannot produce and must thus forcibly be present in the culture medium.

1.2.4 Snapshot

On the whole, essential phenomena for modeling can be captured by the snapshot provided in Figure 1.10.

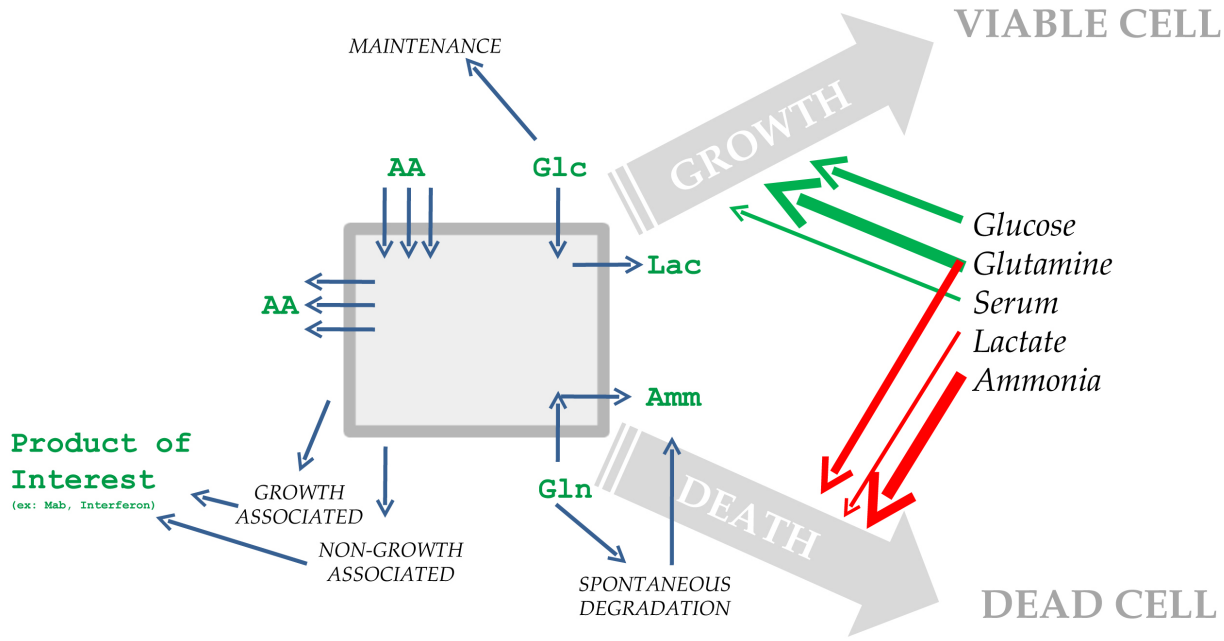


Figure 1.10: Snapshot of global cell metabolism. Adapted from Ghoul et al (1991)

It can be said that cell growth is associated with the presence of substrates glucose, glutamine and serum or other growth factors[1]. Cell death is mostly connected to the presence of ammonia, lactate and lack of glutamine. It is known that glucose can partially be used, not for growth purposes, but for cell maintenance purposes. It is also known that glutamine decomposes spontaneously. The production of the biopharmaceutical product can be, in part, growth-associated and, in part, non-growth-associated (dos Reis Castilho, 2008).

[1]More information about the composition of serum and other growth factors and the rationale for avoiding their use in the production of biologicals is given in Brunner et al (2010), for example. A more recent scientific discussion is to be seen in EMA (2013). The European GMP guidelines are provided in Eudralex (2003).

Chapter 2

Modeling

This chapter reviews approaches for modeling. The choice of unsegregated macroscopic models, in the scope of this thesis, is stated and reasons for doing so are presented. Briefly, this type of models presents a good compromise between descriptive insight and simplicity for practical purposes.

A large variety of modeling approaches can be categorised into four types as shown in Figure 2.1. Non-segregated unstructured models are the most simple since they suppose that all cells are equal to an average cell which processes extracellular components measured in the surrounding medium. Another class, non-segregated structured models, hypothesizes a network of reaction pathways inside the cell relaying products and substrates measured outside. Furthermore, if instead of an average cell, a population of different cells is considered, then an additional layer of complexity is gained. This is the case of models considering cells in different phases of their life cycle.

In the following subsections, some examples will be given.

2.1 Non-segregated unstructured models

These are macroscopic models since they account for more abundant substrates and products measured outside the cell. Typically, measurements are

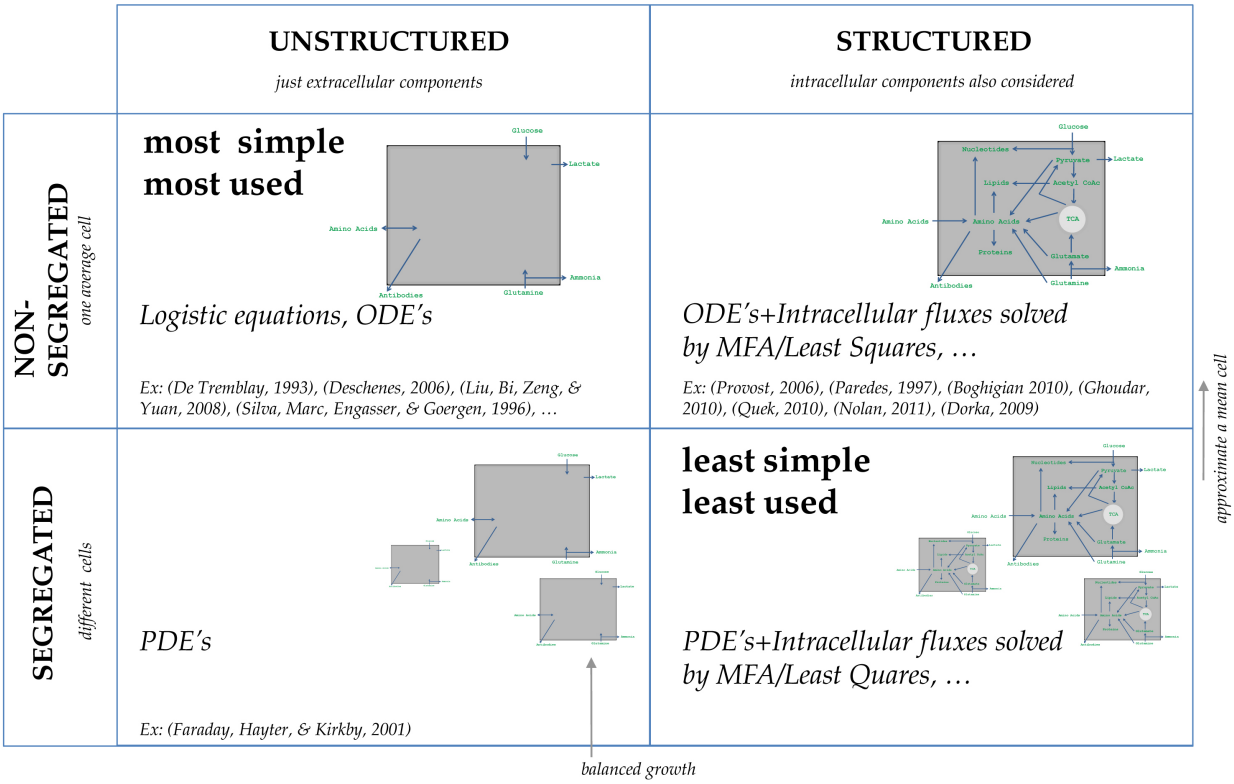


Figure 2.1: Classification of modeling approaches. Adapted from Fredrickson et al (1970) and Bailey (1998).

made outline and samples analysed with enzymatic kits or an HPLC/UPLC with a PDA/ELSD detector.

2.1.1 Logistic equations

In this model class, probably one of the most simple modeling approaches are logistic equations, whose inspiration comes from differential equations developed for population models. In what concerns animal cell cultures, the literature commonly proposes, not equations for the derivatives, but rather equations that directly and explicitly express concentrations over time.

It is very simple to fit these equations into a data set. For example, Figure 2.2 shows a quick fit of ammonia from culture data obtained at the UMons, and the values found for logistic equation (2.1).

$$\begin{cases} f(t) = c_1 + \dfrac{c_2}{c_3 + c_4 e^{-c_5 t}} \\ \{c_1; c_2; c_3; c_4; c_5\} = \{-2.22; 10; 1; 3.5; 0.02\} \end{cases} \tag{2.1}$$

Goudar et al (2005a) propose a four-parameter generalized logistic equation that can be used to describe three types of concentration variables: cell concentration, substrate consumption and product formation. We have illustrated it with generic profiles in Figure 2.2 generated with values in eq. (2.2).

$$\begin{cases} f(t) = \dfrac{1}{c_1 e^{t/c_2} + c_3 e^{-t/c_4}} \\ \{c_1; c_2; c_3; c_4\} = \{0.01; 1; 0.01; 1000\} \quad \text{for biomass} \\ \text{or } = \{0.1; 1; 0.1; 1\} \quad \text{for substrate} \\ \text{or } = \{0.1; 1000; 10; 1\} \quad \text{for product} \end{cases} \tag{2.2}$$

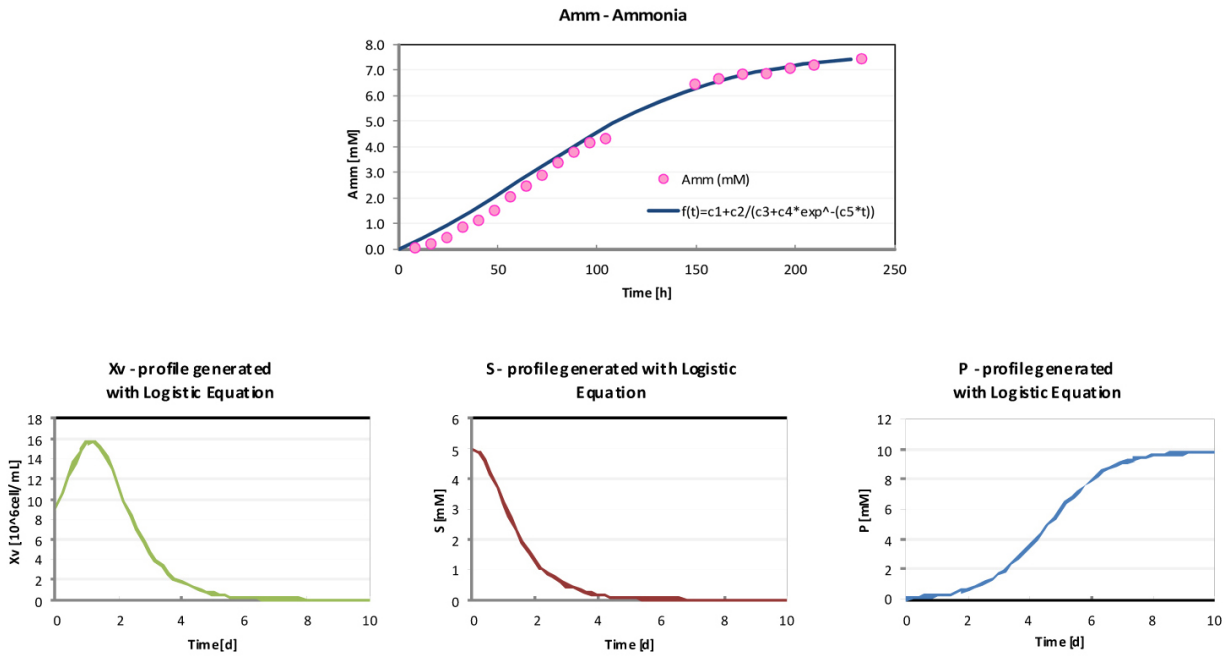


Figure 2.2: Modeling culture data with logistic equations. Top: fit to ammonia data for a culture performed at UMons. Bottom: qualitative examples of profiles generated for biomass, substrate and product.

These very simple equations for batches can be easily analytically differentiated and integrated even if the available multitude of possible shapes with exponentials may lack some biological meaning.

More recently, asymmetric logistic equations (ALE) have been proposed in order to correctly explain the asymmetric evolution of cell concentration and substrate consumption for a batch (Goudar et al, 2005a,b, 2007). For example:

$$Xv(t) = c_1 + c_2 e^{-\frac{t+c_3 \ln c_4 - c_5}{c_3}} \left(1 + e^{-\frac{t+c_3 \ln c_4 - c_5}{c_3}}\right)^{-c_4-1} c_4^{-c_4} + (c_4+1)^{c_4+1} \tag{2.3}$$

$$S(t) = c_6 + \frac{c_7}{\left(1 + e^{\left(-\frac{t-c_8 \ln(2^{1/c_9}-1)-c_5}{c_8}\right)}\right)^{c_9}} \tag{2.4}$$

The equations are highly nonlinear, but since the states $x(t)$ are explicitly time-dependant, differentiation can be easily done analytically. This allows for the estimation of some parameters for that batch. For example, for a particular batch, the specific cell growth μ can either be estimated from experimental biomass concentration data, $Xv(t)$, via eq. (2.5) or by differentiation via eq. (2.6).

$$\frac{dXv(t)}{dt} = \mu(t) Xv(t) \Longrightarrow \mu(t) = \frac{dXv(t)/dt}{Xv(t)} \tag{2.5}$$

$$\begin{aligned}\frac{dXv(t)}{dt} = \frac{-c_2}{c_3} e^{\left(-\frac{t+c_3 \ln c_4 - c_5}{c_3}\right)} c_4^{-c_4} (c_4+1)^{(c_4+1)} \left(1 + e^{\left(-\frac{t+c_3 \ln c_4 - c_5}{c_3}\right)}\right)^{(-c_4-1)} \times \\ \times \left(1 + e^{\left(-\frac{t+c_3 \ln c_4 - c_5}{c_3}\right)} (-c_4-1) \left(1 + e^{\left(-\frac{t+c_3 \ln c_4 - c_5}{c_3}\right)}\right)^{-1}\right)\end{aligned} \tag{2.6}$$

A limitation of this type of equations deals with its application: they are valid for one specific batch (descriptive purpose), not expressing conveniently the link between substrate presence and biomass (predictive purpose). Some models possess common parameters. For example, equations (2.3) and (2.4), for biomass and substrate respectively, have a common parameter c_5 providing a link. However, the equations cannot be used to predict a scenario where the evolution of substrate concentration would be different (eg. for control purposes). Another weaker point is that they can have many parameters and take many shapes. It is thus difficult to compare parameter values coming from different logistic equations.

However, these models may be useful if the focus is not on the predictive power but rather on topics such as batch-to-batch repeatability (eg. a repeated industrial batch operation always performed under the same conditions, where it is interesting for regulatory reasons to show that the value of some parameters has not varied much and therefore potentially neither has the quality of the pharmaceutical product).

2.1.2 ODE-based models

The model class that seems to be of more common use is based on ordinary differential equations (ODE) describing mass balances by means of terms for kinetic phenomena (reactions) and terms for fluid dynamics (flowrates entering or leaving the bioreactor).

The kinetic terms comprise stoichiometric coefficients and reaction rates. Due to the high nonlinearity of these terms, the analytical integration of the equations is difficult. It it thus common to use numerical integration in order to obtain the time profiles for biomass, substrates and products.

The general equations can be written in a canonical form well described in Bastin and Dochain (1990). In equations (2.7) and (2.8)[1], ξ_i are the concentrations of the i components considered, v_{ij} the pseudo-stoichiometric coefficient of component i in reaction j, φ_j the reaction rate of reaction j, F^{IN} the instream flowrate with substrate concentration ξ_i^{IN}, V the volume, and F_{perf} and F_{bleed} the outstream flowrates for perfused output and bleed output, respectively.

$$\frac{d\xi_i}{dt} = \sum_{j=1}^{M} v_{ij}\varphi_j + \frac{F^{IN}}{V}\xi_i^{IN} - \frac{F^{IN}}{V}\xi_i \left(+ \frac{F_{perf}}{V}\xi_i \text{ for } \xi_i \equiv \text{biomass} \right) \tag{2.7}$$

$$\frac{dV}{dt} = F^{IN} - F_{bleed} - F_{perf} \tag{2.8}$$

Notice that biological knowledge is incorporated through the set of reactions considered. Inspiration often comes from enzyme kinetics. Equations (2.9) to (2.10) show an example of 3 reactions to consider in a simple ODE-based model for a cell culture. The first reaction indicates that growth, ie, the formation of living cells (Xv), involves the uptake of substrates glucose (Glc) and glutamine (Gln) while metabolites lactate (Lac) and ammonia (Amm) are formed at the same time as an interesting biopharmaceutical: monoclonal

[1]The corresponding layout is in Fig. 1.1, p. 31.

antibodies (*MAb*). The second reaction indicates that part of the glucose consumption is due to cell maintenance activities. Finally, the third reaction states that part of the living cells in the system will become dead cells (X_d). This is an important feature of an animal cell model. Whilst for microorganisms models often do not incorporate death, for animal cells the phenomenon is considered, in vue of their fragility to multiple outside conditions (dos Reis Castilho, 2008) and because the model can be intended for use outside a pure-exponential growth initial phase of the culture.

An example of an ODE-based model for animal cell cultures is presented below:

$$(-\nu_{11})\text{Glc} + (-\nu_{21})\text{Gln} \xrightarrow{\varphi_1} \circlearrowleft\text{Xv} + (\nu_{41})\text{Lac} + (\nu_{51})\text{Amm} + (\nu_{61})\text{MAb} \tag{2.9}$$

$$(-1)\text{Xv} \xrightarrow{\varphi_2} (1)\text{Xd} \tag{2.10}$$

$$(-1)\text{Glc} + (-\nu_{32})\text{Xv} \xrightarrow{\varphi_3} (\nu_{32})\text{Xv} \tag{2.11}$$

For batch operation:

$$\frac{d}{dt}\begin{bmatrix} Xv \\ Glc \\ Lac \\ Gln \\ Amm \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ -\nu_{11} & 0 & -1 \\ \nu_{41} & 0 & 0 \\ -\nu_{21} & 0 & 0 \\ \nu_{51} & 0 & 0 \end{bmatrix} \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \varphi_3 \end{bmatrix} \tag{2.12}$$

$$\varphi_1 = \mu_g \times Xv \quad \text{(cell growth)} \tag{2.13}$$

$$\varphi_2 = \mu_d \times Xv \quad \text{(cell death)} \tag{2.14}$$

$$\varphi_3 = \mu_S \times Xv \quad \text{(cell maintenance)} \tag{2.15}$$

$$\nu_{11} = \frac{1}{Y_{Xv/Glc}}; \nu_{21} = \frac{1}{Y_{Xv/Gln}}; \nu_{41} = \frac{Y_{Lac/Glc}}{Y_{Xv/Glc}}; \nu_{51} = \frac{Y_{Amm/Gln}}{Y_{Xv/Glc}} \tag{2.16}$$

The arrow in $\circlearrowleft$Xv denotes that the cell culture behaves like an autocatalytic reaction, since it takes the division of one cell to obtain more and thus the rate of growth is proportional to the biomass that is present. Notice also

that one variable, biomass concentration, Xv, affects all dynamic equations (2.13)-(2.15). The quality of its modeling is, thus, of utmost importance.

While observing animal cell batch cultures, it can be noticed that a plot of the logarithms of concentrations indicates regions with roughly straight lines, implying that some approximately exponential-type phases can be identified. This is shown in Figure 2.3 as portrayed in the textbook Dunn et al (2003).

1. A short (sometimes negligible) lag phase where cells are thought to be adapting to the culture medium where they are suspended. Cell concentration remains constant;

2. A period of exponential growth for cells ($\ln(Xv)$ is thus a line with positive slope) and exponentially proportional substrate consumption meanwhile ($\ln(S)$ is constant);

3. A short moment when (for substrate depletion reasons or others) cell growth becomes limited ($\ln(Xv)$ is more or less constant);

4. Eventually, a later period where cell growth is outweighed by cell death ($\ln(Xv)$ becomes a line with negative slope);

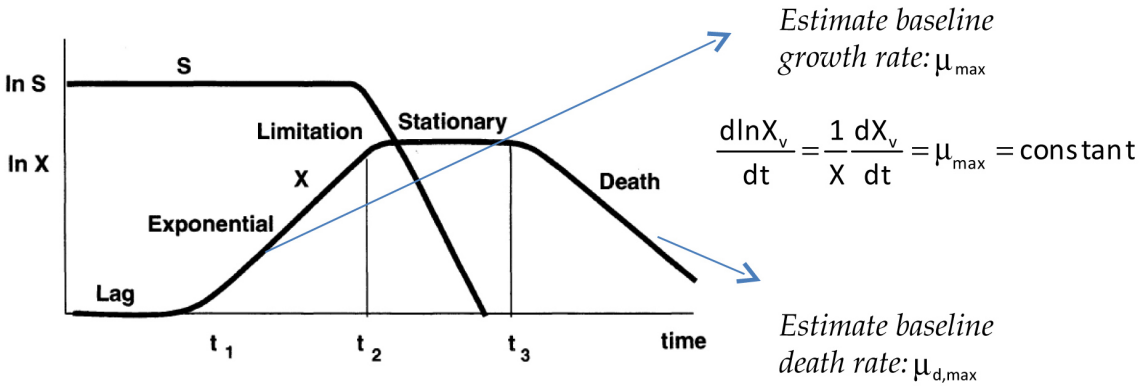


Figure 2.3: Estimation of some ODE-based model parameters. Log profiles as shown in (Dunn et al, 2003).

A practical illustration where some of these aspects are present is provided in Fig. 2.4, where the model of de Tremblay et al (1992) was used to generate biomass and substrate profiles in batch mode.

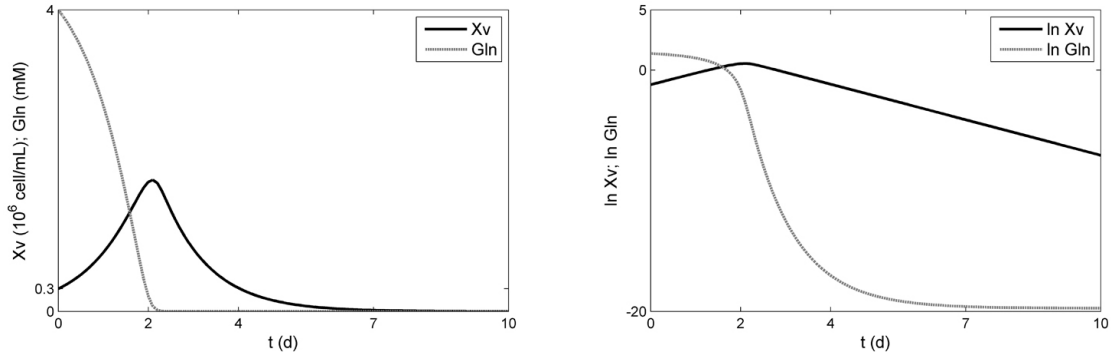


Figure 2.4: Left: numerical simulation of biomass and substrate (glutamline) profiles using the model of de Tremblay et al (1992) in batch mode. Right: corresponding log profiles.

2.1.2.1 Cell growth and death rates

To sum up, the slope of the exponential phase allows to estimate a baseline maximum growth rate μ_{max} based on cell concentration and the slope of the death phase to estimate a baseline death rate $\mu_{d,max}$. Since it is well known that limitation and inhibition phenomena take place during a culture, the maximum growth rate can be multiplied by terms taking values from 0 and 1. For example, factor (2.21) becomes zero after substrate depletion implying zero cell growth when multiplied to μ_{max}. For cell death, similarly. Some common examples are provided below an an illustration of the form taken provided in Figure 3.47.

$$\varphi_{growth} = \mu Xv \tag{2.17}$$

where $\mu = \mu_{max} \times$ phenomena lowering maximal growth

$$\varphi_{death} = \mu_d Xv \tag{2.18}$$

where $\mu_d = \mu_{d,base} \times$ phenomena enhancing baseline cell death

$$\varphi_{maintenance} = m_{Glc} Xv \tag{2.19}$$

$$\text{product inhibition: } \frac{k_{P_i}}{P_i + k_{P_i}} \tag{2.20}$$

$$\text{substrate limitation: } \frac{S_i}{S_i + k_{S_i}} \tag{2.21}$$

$$\text{product limitation: } \frac{P_i}{P_i + k_{d,P_i}} \tag{2.22}$$

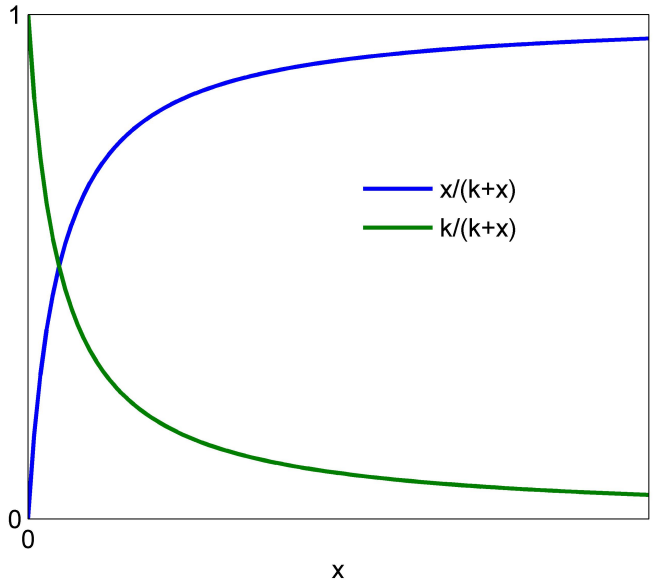


Figure 2.5: Mathematical forms commonly used do describe inhibition and limitation phenomena.

Literally hundreds of forms have been proposed for different cell lines cultured in different media. Some of the most referenced are listed in dos Reis Castilho (2008) and sampled in Tables 2.1 and 2.2.

Globally, most hypotheses used in practice are that glucose and glutamine limit cell growth (μ_g) and that death (μ_d) is accentuated with ammonia and lactate build-up. The combination of the terms chosen for the description of μ_g and μ_d depend strongly on the cell line and conditions for which the model is intended. It is thus necessary to evaluate whether our culture medium possesses alternative carbon and nitrogen sources other than common glucose and glutamine that cells may use (eg. Batt and Kompala (1989) and Dunn et al (2003)). In that case, it should be observed that cell growth is not null after simultaneous glucose and glutamine depletion because other substitutes are still being uptaken. Another important hypothesis is whether substrate limitation is considered additive or multiplicative, ie., if the presence of one substrate is sufficient or if cell growth needs the the simultaneous presence of glucose and glutamine.

Table 2.1: Growth terms (dos Reis Castilho, 2008).

Terms	Source
$\mu_g = \mu_{g,max}\frac{Gln}{k_{Gln}+Gln}$	Portner et al (1996)
$\mu_g = \mu_{g,max}\frac{Glc}{k_{Glc}+Glc}\frac{Gln}{k_{Gln}+Gln}$	de Tremblay et al (1992)
$\mu_g = \mu_{g,max}\frac{Glc}{k_{Glc}+Glc}\frac{k_{i,Lac}}{k_{i,Lac}+Lac}$	Kurokawa et al (1994)
$\mu_g = \mu_{g,max}\frac{Gln}{k_{Gln}+Gln}\frac{k_{i,Amm}}{k_{i,Amm}+Amm}\frac{k_{i,Lac}}{k_{i,Lac}+Lac}$	Bree et al (1988)
$\mu_g = \mu_{g,max}\frac{Glc}{k_{Glc}+Glc}\frac{Gln}{k_{Gln}+Gln}\frac{k_{i,Amm}}{k_{i,Amm}+Amm}\frac{k_{i,Lac}}{k_{i,Lac}+Lac}$	Miller et al (1988a)
$\mu_g = \mu_{g,max} + (\mu_{g,max} - \mu_{g,min})\frac{Glc-Glc_{thres}}{k_{Glc}+(Glc-Glc_{thres})}$	Frame and Hu (1991a)
$\mu_g = \mu_{g,max}Serum\frac{Glc}{k_{Glc}+Glc}$	Dalili et al (1990)
$\mu_g = \mu_{g,max}\frac{Serum}{Serum+k_{Serum,0}Xv^{-\beta*}}\frac{Gln}{k_{Gln}+Gln}\frac{k_{i,Amm}}{k_{i,Amm}+Amm}$	Glacken et al (1989)
$\mu_g = a_1\frac{B-a_2}{B}$	Gaertner and Dhurjati (1993)
$\mu_g = D + d_0e^{d_1/\mu_g}$	Linardos et al (1991)
$\mu_g = \mu_{g,max}(1 - a_1\frac{Xv}{D})\frac{Glc}{k_{Glc}+Glc}\frac{Gln}{k_{Gln}+Gln}$	Zeng et al (1998)

Table 2.2: Death terms (dos Reis Castilho, 2008).

Terms	Source
$\mu_d = \mu_{d,max}\frac{Amm}{k_{d,Amm}+Amm}\frac{Lac}{k_{d,Lac}+Lac}$	Batt and Kompala (1989)
$\mu_d = \mu_{d,max}\frac{Amm}{k_{d,Amm}+Amm}\frac{Lac}{k_{d,Lac}+Lac}\frac{k_{d,i,Gln}}{k_{d,i,Gln}+Gln}$	Bree et al (1988)
$\mu_d = \mu_{d,max}\frac{1}{(\mu_{g,max}-k_{d,Lac}Lac)(\mu_{g,max}-k_{d,Amm}Amm)}\frac{k_{d,i,Gln}}{k_{d,i,Gln}+Gln}$	de Tremblay et al (1992)
$\mu_d = \mu_{d,min} + (k_{d,max} - k_{d,min})\frac{k_{d,i,Gln}}{k_{d,i,Gln}+Gln}$	Dalili et al (1990)
$\mu_d = (\mu_{g,min} - D_{min}) - \mu_{d,max}\frac{Glc-Glc_{thres}}{k_{d,Glc}+(Glc-Glc_{thres})}$	Frame and Hu (1991a)
$\mu_d = b_1 + \frac{b_2}{Gln+b_3}$	Portner et al (1996)
$\mu_d = c_1e^{c_2\mu_g}$	Glacken et al (1989)
$\mu_d = d_0e^{(d_1/\mu_g)}$	Linardos et al (1991)
$\mu_d = (\beta_0 + \beta_1\mu_g)\frac{Xv+Xd}{D}$	Zeng et al (1998)

For example, in a medium where two alternative substrates such as glucose and another hexose are present a multiplicative form would not be valid.

However, an additive form such as double-Monod or a diauxic-Monod could be employed (Batt and Kompala, 1989; Dunn et al, 2003):

- Double Monod
  Each substrate allows a different maximal growth.;

$$\mu_g = \mu_{g,max}\left(\frac{k_1 S_1}{k_1 + S_1} + \frac{k_2 S_2}{k_2 + S_2}\right)\left(\frac{1}{k_1 + k_2}\right) \tag{2.23}$$

- Diauxic Monod
  The consumption of substrate S_2 is inhibited until S_1 is exhausted (eg, for bacteria *E.Coli* the uptake of lactose is repressed while glucose is present.;

$$\mu_g = \mu_{g,max,1}\frac{S_1}{k_1 + S_1} + \mu_{g,max,2}\frac{S_2}{k_2 + S_2 + S_1^2/k_1} \tag{2.24}$$

Again, model application must be thought of before complicating it. If it is meant to be applied in situations where certain phenomena will not occur, then it is unnecessary to include terms for these phenomena. For example, the culture presented in Figure 1.5 (p. 36) has abundant glucose. It may, thus, not be necessary to model glucose limitation at all, just glutamine limitation, since the term $Glc/(k_{Glc}+Glc)$ would always be approximately 1 in those conditions.

2.1.2.2 Substrate consumption

In ODE-based models, the evolution of substrate concentrations is usually either related to its use for cell growth or to cell maintenance purposes. The later is a phenomenon sometimes observed when, during the death phase, the substrate concentration is still diminishing. Table 2.3 presents the more common simple terms and some less used and more complicated expressions, as listed in dos Reis Castilho (2008). μ_g is present in growth-related terms and maintenance related ones are typically noted by a m_S parameter.

Table 2.3: Substrate consumption (dos Reis Castilho, 2008).

Terms	Source
$\mu_S = \frac{1}{Y^{max}_{Xv/S}}\mu_g$	de Tremblay et al (1992); Hiller et al (1991)
$\mu_S = \frac{1}{Y^{max}_{Xv/S}}\mu_g + m_S$	Harigae et al (1994); Hiller et al (1991); Miller et al (1988a); Kurokawa et al (1994)
$\mu_S = \frac{1}{Y^{max}_{Xv/S}}\mu_g + m_S - e^{\mu_d}$	Linardos et al (1991)
$\mu_S = \frac{1}{Y^{max}_{Xv/S}}\left(\mu_g - \mu_{min}\right)$	Frame and Hu (1991a)
$\mu_S = \frac{1}{\frac{1}{Y_{Xv/S}} - \frac{\lambda}{Y_{P/S}}}\mu_g + \frac{\alpha_2}{Y_{P/S}} - \frac{1}{Y_{Xv/S}}\mu_{g,min}$	Frame and Hu (1991b)
$\mu_S = \frac{1}{Y^{max}_{Xv/S}}\mu_g + m_S\frac{S}{k^S_S+S}$	de Tremblay et al (1992)
$\mu_S = \frac{h_1 S}{h_2+S}$	Portner et al (1996); Gaertner and Dhurjati (1993)
$\mu_S = \left(\frac{1}{Y^{max}_{Xv/S}}\mu_g + m_S\right) + \Delta\mu^m_S\frac{S}{S+Xv\times k_S}$	Zeng (1996b)

2.1.2.3 Product production

Byproduct formation is related to substrate consumption. For instance, the formation of byproduct lactate is related to glucose variation (consumption, if in batch mode) and can be modeled by equation (2.25).

$$\frac{dLac}{dt} = -Y_{Lac/Glc}\frac{dGlc}{dt} \tag{2.25}$$

For the more special case of the biopharmaceutical product of interest, it is generally assumed that its synthesis is partially related to cell growth μ_g and partially independent, as stated in the general Luedeking-Piret equation (Dunn et al, 2003):

$$\frac{dP}{dt} = \left(\frac{1}{Y_{Xv/P}}\mu_g + b\right)Xv \tag{2.26}$$

Tables 2.4 and 2.5 present some other possible forms.

Table 2.4: Byproduct formation (dos Reis Castilho, 2008).

Terms	Source
$\mu_{Lac} = \mu_{Lac,max}\frac{Glc}{k_{Glc}^{Lac}+Glc}$	Gaertner and Dhurjati (1993)
$\mu_{Lac} = Y_{Lac/Xv}\mu_g + m_{Lac} + \Delta\mu_{Lac,Glc}^{max}\frac{Glc-Glc^*}{Glc-Glc^*+k_{Glc}^{Lac}}$	Zeng (1995)
$\mu_{Lac} = Y_{Lac/Xv}\mu_g + m_{Lac} + \Delta\mu_{Lac,Glc}^{max}\frac{Glc}{Glc+k_{Glc}^{Lac\#}Xv} + \ldots +$ $+\Delta\mu_{Lac,Gln}^{max}\frac{Gln}{Gln+k_{Gln}^{Lac\#}Xv}$	Zeng (1996b)
$\mu_{Amm} = \frac{E_1+E_2Glc}{E_3+Glc}$	Gaertner and Dhurjati (1993)
$\mu_{Lac} = \frac{G_1+G_2Lac}{G_3+Lac}$	Gaertner and Dhurjati (1993)
$\mu_{Amm} = Y_{Amm/Xv}\mu_g + m_{Amm} + \Delta\mu_{Amm,Gln}^{max}\frac{Gln}{Gln+k_{Gln}^{Amm}} +$ $+\ldots+\Delta\mu_{Amm,Glc}^{max}\frac{Glc}{Glc+k_{Glc}^{Amm}}$	Zeng (1995)
$\mu_{Amm} = Y_{Amm/Xv}\mu_g + m_{Amm} + \Delta\mu_{Amm,Gln}^{max}\frac{Gln}{Gln+k_{Gln}^{Amm\#}Xv}$	Zeng (1996b)

Table 2.5: Product formation (monoclonal antibodies) (dos Reis Castilho, 2008).

Terms	Source
$\mu_{MAb} = \beta$	Portner et al (1996)
	Hiller et al (1991)
$\mu_{MAb} = \alpha\mu_g + \beta$	Frame and Hu (1991b)
$\mu_{MAb} = \frac{\alpha_0}{k_\mu+\mu_g}\mu_g + \beta$	de Tremblay et al (1992)
$\mu_{MAb} = \alpha_1\mu_d + \beta_1$	Linardos et al (1991)
$\mu_{MAb} = \beta_2 Serum\frac{Gln}{k_{Gln}^{MAb}+Gln}$	Dalili et al (1990)
$\mu_{MAb} = (\alpha_1\mu_d + \beta_1)\frac{Gln}{k_{Gln}^{MAb}+Gln}\frac{k_{i,Glc}^{MAb}}{Glc-Glc^*+k_{i,Gln}^{MAb}}(F_1 + e^{-F_2\Delta t})$	Zeng (1996a)
$\mu_{MAb} = (\alpha_1\mu_d + \beta_1)\frac{Gln}{k_{Gln}^{MAb\#}Xv+Gln}\frac{k_{i,Glc}^{MAb\#}Nv}{Glc+k_{i,Glc}^{MAb\#}Xv}$	Zeng (1996b)
$\mu_{MAb} = \delta\frac{D_{Per}}{D_{Per}+k_{D_{Per}}^{MAb}}\frac{k_{i,Glc}^{MAb}}{Glc+k_{i,Glc}^{MAb}}$	Zeng (1996b)

2.1.2.4 Example of a simple model

A simple example of an ODE-based model is the one used in Deschenes (2007) to describe mammalian cell line HEK 293-SF cultivated in NSFM13 medium in continuous perfused regime. The kinetics assumed by the author to describe 3 states (glucose, living biomass and dead biomass) can be transposed into the following equations:

$$(-\nu_{11})\text{Glc} \xrightarrow{\varphi_1} \circlearrowleft\text{Xv} \quad \text{(cell growth reaction)} \tag{2.27}$$

$$(-1)\text{Xv} \xrightarrow{\varphi_2} (1)\text{Xd} \quad \text{(cell death reaction)} \tag{2.28}$$

$$\frac{d\xi_i}{dt} = \nu_{ij}\varphi_j + \cdots \rightarrow \frac{d}{dt}\begin{bmatrix} Glc \\ Xv \\ Xd \end{bmatrix} = \begin{bmatrix} -\nu_{11} & 0 \\ 1 & -1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} \varphi_1 \\ \varphi_2 \end{bmatrix} + \cdots \tag{2.29}$$

$$\varphi_1 = \mu_1 Xv \text{ with } \mu_1 = \mu_{max}\frac{Glc}{k_C Xv + Glc} \tag{2.30}$$

$$\varphi_2 = \mu_2 Xv \text{ with } \mu_2 = Xv + Xd \tag{2.31}$$

The choice of a Contois (over a Monod) form limiting cell growth relates to the use of the model for high cell concentrations typical of perfusion regimes: real substrate availability may become limited when many cells surround one cell. The data presented do not include important states such as limiting glutamine or inhibiting ammonia, which are thus not comprised in the model. Regarding cell death rate, is it simply assumed to be proportional to the total cell concentration. This means that Xd, a variable which in reality is difficult to measure (dead cells eventually break down), was used. The author bases his choice on the fact that, among several models tried, this was the sole model he could find to reproduce an overshoot visible on the biomass profile data presented to him. Actually, this phenomenon can also be reproduced in some conditions for perfused cultures using more descriptive models. In Figure 2.6 this is illustrated with a comparison of Deschenes (2007)'s results (in black) with a simulation of de Tremblay et al (1992)'s model (in green).

The model used in Deschenes (2007) is meant to simulate well an available set of data so that some control strategies for perfusion may be studied. It is, thus, very simple and may lack some predictive power for different culture conditions since it was identified with limited data. The risk of using it would

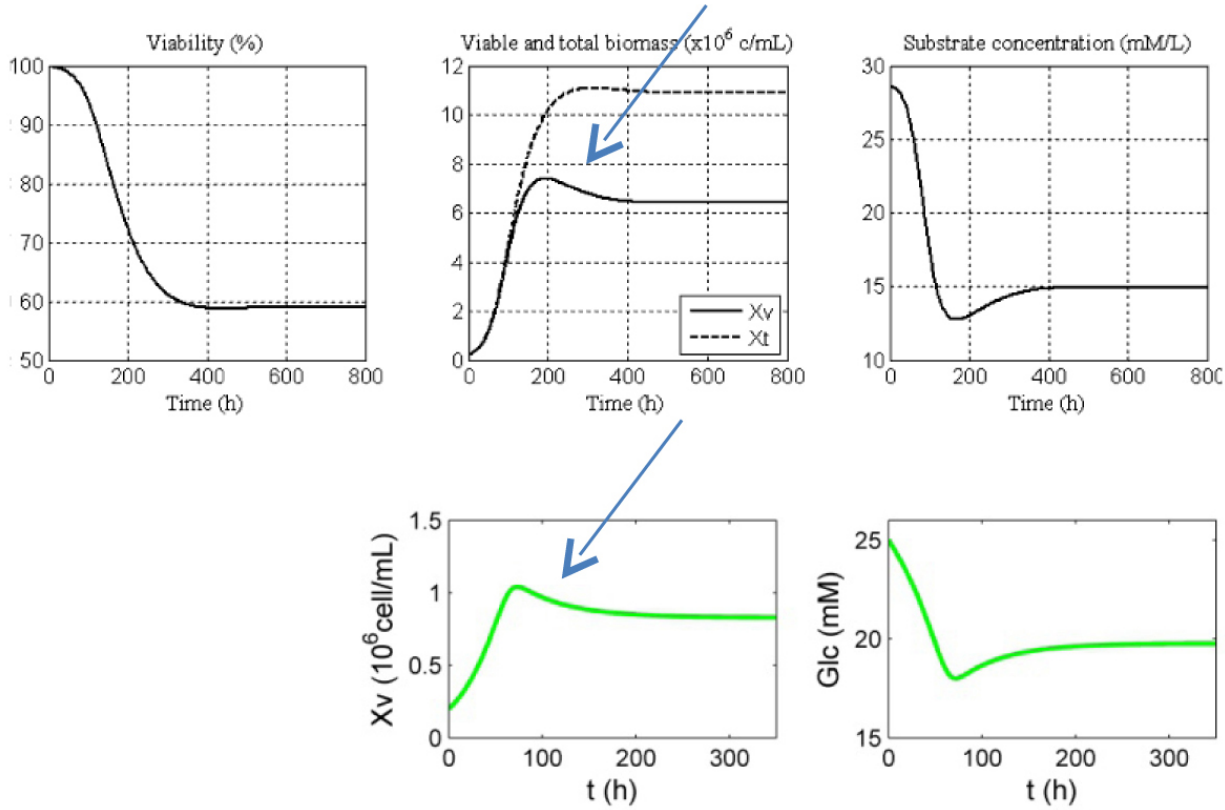


Figure 2.6: Continuous perfused cultures presenting an overshoot in the biomass profile: (Deschenes et al, 2006b)'s results in black and simulation of de Tremblay et al (1992)'s model in green.

be that phenomena not considered in the model (limitation, inhibition) would occur and the controller would not be able to overcome them (eg. not properly adjust flowrates in order to maintain a good setpoint compliance).

2.1.2.5 Example of a reasonably comprehensive model: De Tremblay

An example of a more comprehensive, yet still reasonably simple, model is the one developed and identified by de Tremblay (1991) for mouse-mouse CBM-P2C hybridoma cell line producing IgM monoclonal antibodies in a customised DMEM base medium with 1% FBS.

It comprises 6 states variables (7 if volume is considered) which are the major component concentrations: living biomass, glucose/lactate, glutamine/ammonia and the product of interest, monoclonal antibodies. The model structure comprises the possible occurrence of typical cell culture phenomena and may thus be considered of fairly general applicability. The

reasonable biological assumptions considered are that cell growth is limited by glucose and glutamine availability, cell death is promoted by accumulation of lactate and ammonia and glutamine exhaustion, a part of glucose is spent in cell maintenance, and the production of the product of interest, monoclonal antibodies, is partially growth-related and partially not.

De Tremblay considered several batch and fedbatch runs in order to develop and identify a model structure. She used it to study the optimal static[1] media feeding trajectories in fedbatch mode and, at this point, significantly simplified it.[2]

The model can be translated into the following set of reactions and canonical-form equations:

$$(-\nu_{11})\mathrm{Glc} + (-\nu_{21})\mathrm{Gln} \xrightarrow[\text{(cell growth)}]{\varphi_1} \circlearrowleft\mathrm{Xv} + \nu_{41}\mathrm{Lac} + \nu_{51}\mathrm{Amm} \tag{2.32}$$

$$(-1)\mathrm{Glc} + (-\nu_{32})\mathrm{Xv} \xrightarrow[\text{(maintenance)}]{\varphi_2} (\nu_{32})\mathrm{Xv} + (\nu_{42})\mathrm{Lac} \tag{2.33}$$

$$(-1)\mathrm{Xv} \xrightarrow[\text{(cell death)}]{\varphi_3} (1)\mathrm{Xd} \tag{2.34}$$

$$(-\nu_{34})\mathrm{Xv} \xrightarrow[\text{(product formation)}]{\varphi_4} (1)\mathrm{Xv} + (1)\mathrm{MAb} \tag{2.35}$$

$$\frac{d\xi_i}{dt} = \nu_{ij}\varphi_j + \cdots \rightarrow \frac{d}{dt}\begin{bmatrix} Glc \\ Gln \\ Xv \\ Lac \\ Amm \\ MAb \end{bmatrix} = \begin{bmatrix} -\nu_{11} & -1 & 0 & 0 \\ -\nu_{21} & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ \nu_{41} & \nu_{42} & 0 & 0 \\ \nu_{51} & 0 & 0 & 0 \\ 0 & 0 & 0 & \nu_{64} \end{bmatrix} \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \varphi_3 \\ \varphi_4 \end{bmatrix} + \cdots \tag{2.36}$$

$$\varphi_1 = \mu_1 Xv \text{ with } \mu_1 = \mu_{max}\frac{Glc}{k_{Glc} + Glc}\frac{Gln}{k_{Gln} + Gln} \tag{2.37}$$

[1] its composition remains constant throughout the culture, unlike a dynamic composition medium

[2] see (de Tremblay et al, 1993)

$$\varphi_2 = \mu_2 Xv \text{ with } \mu_2 = \underbrace{m_{Glc}\left(\frac{Glc}{k_{m,Glc} + Glc}\right)}_{(1)} \tag{2.38}$$

$$\varphi_3 = \mu_3 Xv \text{ with}$$
$$\mu_3 = \mu_{d,max} \underbrace{\left(\frac{1}{\mu_{max} - k_{d,Lac}Lac}\right)\left(\frac{1}{\mu_{max} - k_{d,Amm}Amm}\right)}_{(1)}\left(\frac{k_{d,Gln}}{k_{d,Gln} + Gln}\right) \tag{2.39}$$

$$\varphi_4 = \mu_4 Xv \text{ with } \mu_4 = \frac{\alpha}{k_\mu + \mu_1}\mu_1 + \underbrace{\beta}_{(1)} \tag{2.40}$$

Where $\nu_{11} = 1/Y_{Xv/Glc}$; $\nu_{21} = 1/Y_{Xv/Gln}$; $\nu_{41} = Y_{Lac/Glc}/Y_{Xv/Glc}$; $\nu_{42} = Y_{Lac/Glc}$;$\nu_{51} = Y_{Amm/Gln}/Y_{Xv/Gln}$; $\nu_{64} = 1$ and the parameter values are those in Table 3.2.

Simulations of what the model predicts for different operating modes are provided in Figure 1.6 on page 37.

It should be noticed that the phenomena considered are generic and may not happen with all combinations of animal cell lines, medium and operating conditions. But it is precisely its generic character that makes it an interesting model to work with. For example, in Saraiva et al (2012) the model was used to study a situation that may often occur: a limited capability of measuring all components at the laboratory. An extended Kalman filter (EKF) served as a software sensor to reconstruct all system states from limited measurements, as illustrated in Figure 2.7.

In Sbarciog et al (2013) the model was integrated in a nonlinear model predictive controller used in a continuous perfused culture.

Several other authors have also considered this model, such as Aehle et al (2011); Chen et al (2002); Franco-Lara and Weuster-Botz (2005); Nguang et al (2001); Portner and Schafer (1996); Roubos et al (1997, 1999); Sarkar and Modak (2004).

[1]Terms later dropped out in de Tremblay et al (1993) where the authors, for the purpose of another study, did not consider glucose maintenance nor lactate or ammonia inhibition.

Table 2.6: De Tremblay's model parameter values (de Tremblay, 1991; de Tremblay et al, 1992, 1993).

Parameter	Value	Units
μ_{max}	1.09	d^{-1}
k_{Glc}	1.0	mM
k_{Gln}	0.3	mM
$\mu_{d,max}$	0.09	d^{-1}
$k_{d,Lac}$	0.01	$mM^{-1}d^{-1}$
$k_{d,Amm}$	0.06	$mM^{-1}d^{-1}$
$k_{d,Gln}$	0.02	mM
$Y_{Xv/Glc}$	1.09×10^{-1}	$10^9 cell\, mmol^{-1}$
$Y_{Lac/Glc}$	1.8	$mmol\, mmol^{-1}$
m_{Glc}	$0.17 \times 10^{+1}$	$mmol\, (10^9 cell)^{-1}\, d^{-1}$
$k_{m,Glc}$	19.0	mM
$Y_{Xv/Gln}$	3.80×10^{-1}	$10^9 cell\, mmol^{-1}$
$Y_{Amm/Gln}$	0.85	$mmol\, mmol^{-1}$
β	$0.35 \times 10^{+1}$	$mg\, (10^9 cell)^{-1}\, d^{-1}$
α	$2.57 \times 10^{+1}$	$mg\, (10^9 cell)^{-1}\, d^{-1}$
k_μ	0.02	d^{-1}

2.1.2.6 Example of a more complex model

An example of a model that is very complex regarding practical applications such as control and monitoring is that of Silva et al (1996), developed for a 6H2 murine cell line cultured in DMEM/HamF12 custom medium and producing a IgG2a monoclonal antibody directed against a melanoma-associated antigen. It can be transposed into the following reactions and equations:

$$\begin{array}{c}(-\nu_{11})\mathrm{Glc} + (-\nu_{21})\mathrm{Gln} + (-\nu_{31})\mathrm{S} + (-\nu_{41})\mathrm{Met} + (-\nu_{51})0_2 \xrightarrow{\varphi_1} \\ \xrightarrow{\varphi_1} \quad \curvearrowleft (\nu_{61})\mathrm{Xv} + (\nu_{71})\mathrm{Lac} + (\nu_{81})\mathrm{Amm} + (\nu_{91})\mathrm{MAb} \\ \text{(cell growth)}\end{array} \tag{2.41}$$

$$\begin{array}{c}(-1)\mathrm{Xv} \xrightarrow{\varphi_2} (1)\mathrm{Xd} \\ \text{(cell death)}\end{array} \tag{2.42}$$

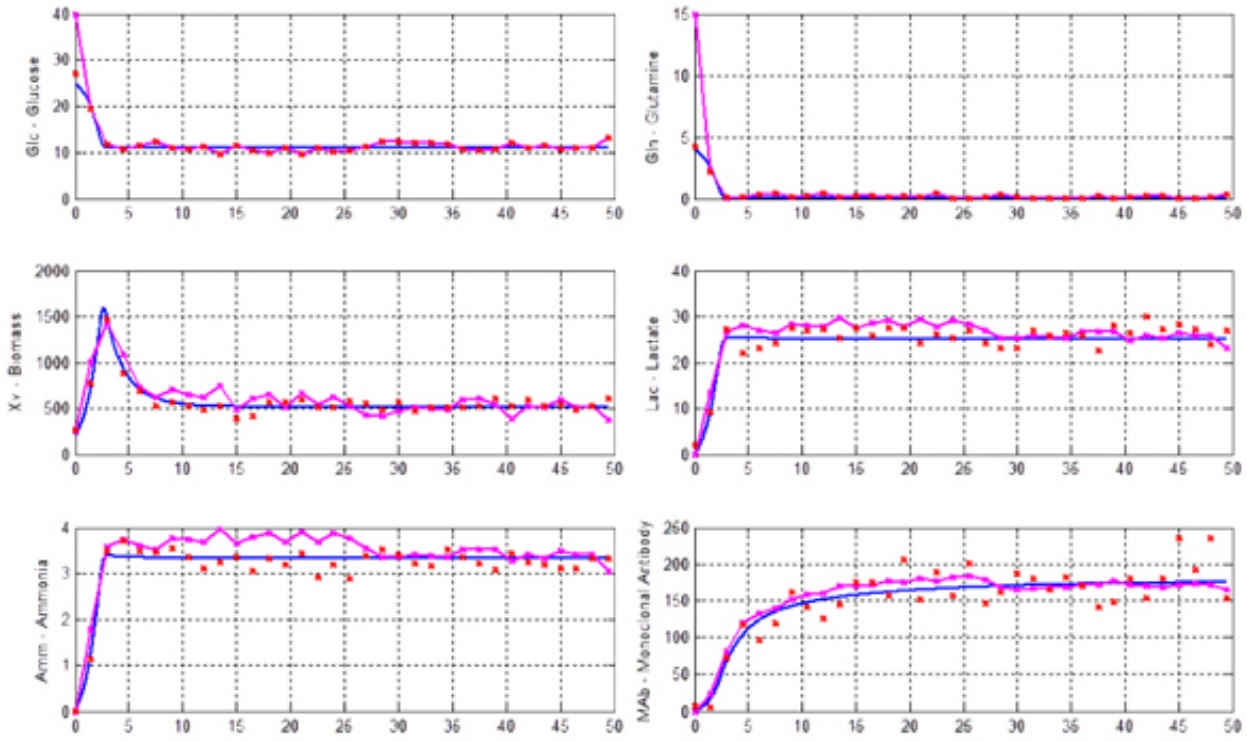


Figure 2.7: Extended Kalman filter estimating concentrations of biomass, lactate, ammonia and monoclonal antibodies from knowledge of initial conditions and measurements of glucose and glutamine. Model (blue), noisy model-generated measurements (red), filter predictions (magenta). Model used: de Tremblay et al (1992). Source: Saraiva et al (2012).

$$\frac{d\xi_i}{dt} = \nu_{ij}\varphi_j + \cdots \rightarrow \frac{d}{dt}\begin{bmatrix} Glc \\ Gln \\ S \\ Met \\ O_2 \\ Xv \\ Lac \\ Amm \\ MAb \end{bmatrix} = \begin{bmatrix} -\nu_{11} & 0 \\ -\nu_{21} & 0 \\ -\nu_{31} & 0 \\ -\nu_{41} & 0 \\ -\nu_{51} & 0 \\ 1 & -1 \\ \nu_{71} & 0 \\ \nu_{81} & 0 \\ \nu_{91} & 0 \end{bmatrix} \begin{bmatrix} \varphi_1 \\ \varphi_2 \end{bmatrix} + \cdots \tag{2.43}$$

The model considers several factors varying between 0 and 1 that can affect growth rate: glucose *Glc*, glutamine *Gln*, methionine *Met*, serum *S*, an amino acids pool *AA*, oxygen O_2, lactate *Lac* and ammonia *Amm*. Since terms are multiplied, a null term is sufficient to cease cell growth:

$$\varphi_1 = \mu_1 X v \text{ with}$$
$$\mu_1 = \mu_{max}\left(\frac{Glc}{Glc+k_{Glc}}\right)\left(\frac{Gln}{Gln+k_{Gln}}\right)\left(\frac{S}{S+k_S}\right)\left(\frac{AA}{AA+k_{AA}}\right)\left(\frac{O_2}{O_2+k_{O_2}}\right)\left(\frac{k_{Lac}}{Lac+k_{Lac}}\right)\left(\frac{k_{Amm}}{Amm+k_{Amm}}\right) \quad (2.44)$$

Cell death, presenting additive terms, cannot be null:

$$\varphi_2 = \mu_2 X v \text{ with}$$
$$\mu_2 = \mu_{d,max}\left[\left(\frac{1}{1+\beta Glc}\right) + \left(\frac{1}{1+\beta Gln}\right) + \left(\frac{1}{1+\alpha S^2}\right) + (k_1 Lac) + (k_2 Amm) + \left(\frac{1}{1+\gamma O_2}\right)\right] \quad (2.45)$$

It is a model with many states and parameters which makes the task of identifying the parameter values difficult. In fact, the authors do not provide them fully. Furthermore, the use of serum is not practical either, since it is by definition a mixture of composition not entirely known and whose concentration in the culture is only known at the initial condition, t_0, once the medium has been prepared.

2.2 Non-segregated unstructured models

Another class of models considers hypotheses regarding the existence of intracellular components and the network of reaction paths connecting extracellular components (substrates and products) measured outside the cell, in the surrounding medium. In these models, a pseudo-stationary state is presumed, meaning that internal metabolites hardly accumulate inside the cell, since reaction rates inside the cell are by far higher than those outside. Thus, for a culture phase where a pseudo-stationary state can be assumed valid, metabolism can be represented by a reaction network with some constant values for fluxes. This is typical when considering the growth phase of a culture, for example, when reaction rates are at their highest since plenty of substrate is still available. There is a limited ability to predict dynamic cell responses to changes on the whole though, since the model only holds during certain limited time intervals of the culture.

2.2.1 Metabolic flux analysis (MFA)

On the whole, two scenarios are possible, according to the number of reactions and the number of fluxes (in Figure 2.8 measured fluxes are green and fluxes

to be determined are red). The system is either determined or overdetermined and can be solved through least squares methods, or it is underdetermined. For this case[1], the use of a toolbox such as Metatool is helpful (Pfeiffer et al, 1999; Schuster et al, 1999).

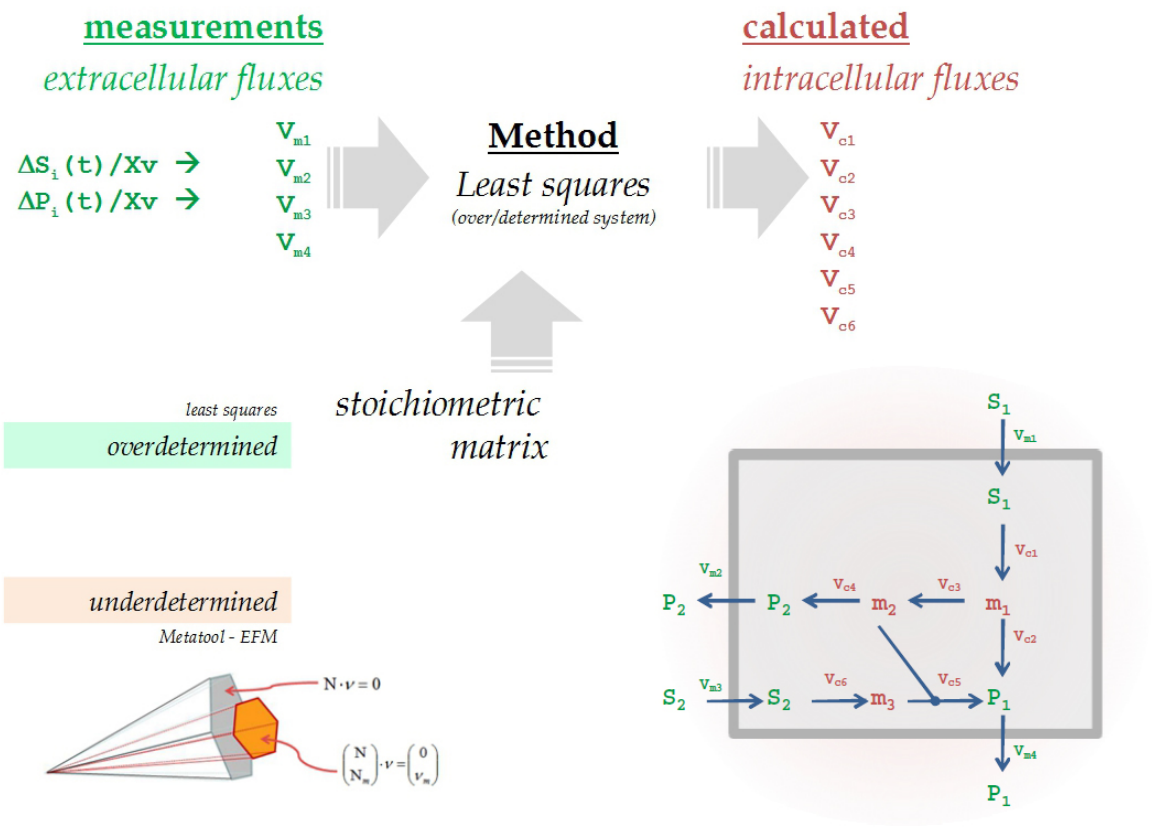


Figure 2.8: Left: metabolic flux analysis: the overdetermined and underdetermined case. Right: example of a metabolic network adapted from Goudar et al (2007).

Two notations exist (eq. (2.46) and eq. (2.49)) and are equivalent. The first is that used in Stephanopoulos et al (1998):

$$G^T v = 0 \iff \left[\begin{array}{cc} G_m^T & G_c^T \end{array} \right] \times \left[\begin{array}{c} v_m \\ v_c \end{array} \right] = 0 \tag{2.46}$$

For example, for a metabolic network of a size similar to the one in Figure 2.8, the system to solve could be something like Goudar et al (2007):

[1]Metatool does not solve exclusively underdetermined systems.

$$
\underbrace{\begin{bmatrix}
\overbrace{\begin{matrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{matrix}}^{a = 4 \text{ measured rates}} & \overbrace{\begin{matrix} -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & -0.5 & -0.5 & 0 \end{matrix}}^{n = 6 \text{ internal fluxes to calculate}} \\
\begin{matrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{matrix} & \begin{matrix} 2 & -1 & -1 & 0 & 0 & 0 \\ 2 & -1 & 4 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 \\ 2 & 0 & 1 & 2.5 & 1.5 & -1 \end{matrix}
\end{bmatrix}}_{8 \times 10}
\begin{bmatrix} -1.48 \\ 1.73 \\ 5.83 \\ -5.14 \\ v_{c1} \\ v_{c2} \\ v_{c3} \\ v_{c4} \\ v_{c5} \\ v_{c6} \end{bmatrix}_{10 \times 1}
=
\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}_{8 \times 1}
$$

($m = 8$ components; $a = 4$ measured components; $b = 4$ unmeasured intracellular components)

(2.47)

Four species with extracellular exchanges are measured and another 4 species are considered intracellular since they undergo no exchange with the outside medium and are, therefore, not measured in the medium. The metabolic network inside the cell consists of 6 reactions whose fluxes v_{ci} we wish to compute. This system is overdetermined by 2 degrees of freedom. In order to have a perfectly determined system with zero degrees of freedom, 2 equations could be taken out.

The rank and condition number of matrix G^T can be computed with software tools such as Matlab. In this case, the rank is 8 and the condition number is 7,6. Thus, matrix G^T is full rank: its rank is equal to $min(rows, columns)$, ie., all eight metabolites have independent mass balances. The matrix also has a condition number of 7.6, not very far from its rank 8 which suggests low sensitivity of the calculated fluxes v_{ci} to the measured rates v_{mi}.

The result can be computed with Matlab by least squares resolution of the overdetermined system, yielding the following result for the unknown fluxes within the cell:

$$\begin{bmatrix} 1.65 \\ 1.64 \\ 1.86 \\ 8.98 \\ 1.75 \\ 30.2 \end{bmatrix} = \begin{bmatrix} v_{c1} \\ v_{c2} \\ v_{c3} \\ v_{c4} \\ v_{c5} \\ v_{c6} \end{bmatrix} \tag{2.48}$$

A common tool used in metabolic flux analysis is the Metatool algorithm, which deals with the case of underdetermined networks by computing the extreme rays of the polyhedral cone of solutions (Pfeiffer et al, 1999; Schuster et al, 1999). This software uses, however, another arrangement of the mass balance equations which represents a second nomenclature, used by Bastin, for example in Bastin (2008); Fernandes et al (2015); Provost and Bastin (2004); Provost et al (2006); Zamorano et al (2010, 2013). For the same system it would be:

$$\overbrace{\begin{bmatrix} N & 0 \\ N_m & -v_m \end{bmatrix}}^{\text{Metatool matrix M}} \times \begin{bmatrix} v \\ 1 \end{bmatrix} = 0 \iff \tag{2.49}$$

$$\begin{bmatrix} \overbrace{\qquad\qquad}^{n=6 \text{ internal fluxes to be measured}} & \overbrace{\quad}^{\text{measured rates}} \\ \left.\begin{matrix} 2 & -1 & -1 & 0 & 0 & 0 \\ 2 & -1 & 4 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 \\ 2 & 0 & 1 & 2.5 & 1.5 & -1 \end{matrix}\right\} b=4 \text{ internal metabolites} & \begin{matrix} 0 \\ 0 \\ 0 \\ 0 \end{matrix} \\ \left.\begin{matrix} -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & -0.5 & -0.5 & 0 \end{matrix}\right\} a=4 \text{ external metabolites} & -\begin{pmatrix} -1.48 \\ 1.74 \\ 5.83 \\ -5.14 \end{pmatrix} \end{bmatrix}_{8\times 7} \begin{bmatrix} v_{c1} \\ v_{c2} \\ v_{c3} \\ v_{c4} \\ v_{c5} \\ v_{c6} \\ \\ 1 \end{bmatrix}_{7\times 1} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}_{8\times 1} \tag{2.50}$$

$m = 8$ components

2.2.2 Reduction of a bigger network

An example of this type of modeling approach is the thesis of Agnès Provost (Provost, 2006). She divided a batch data set into 3 different phases (exponential growth, transition, and death, as illustrated in Figure 2.3 on page 50).

She attributed a reduced metabolic network to each phase, hoping to capture a general picture of what happens at sequential moments of the culture. Finally, in order to have an expression valid for any time t, the 3 models were united by means of continuous functions ϕ_i that vary from 0 to 1 at precise moments t_i chosen by her to represent the switches between culture phases. This is illustrated in Figure 2.9 .

Provost's data represents a batch culture (with 2 replicas) responding to one set of initial conditions. In the course of the culture, two phenomena, glucose and glutamine extinction, happen more or less simultaneously, which makes the relative importance of each hard to understand. It would be interesting if the model would allow to predict the switch time t_{peak} where biomass reaches its maximum concentration, but here it is imposed. As for lactate re-consumption, once glucose is exhausted, this is predicted by the model since the metabolic network considers the reaction to be direct for phases 1 and 2, and reverse for phase 3. It is, therefore, the user's choice of the switch time that determines the beginning of lactate re-consumption. Since it is based on one data set, it is hard to insure its validity for other conditions.

A very interesting point of Provost's thesis is the proposition of a method to reduce bigger pathways to simpler sets of reactions with the intention of facilitating model purposes such as control and optimization.

In fact, Provost's data set could also be quickly modeled with simple macroscopic reactions and an ODE-based model. We have qualitatively illustrated this in Figure 2.10, where Provost's data and model predictions (left) stand side by side with profiles predicted with the following illustrative ODE-based model and presumptions:

1. Glutamine limits growth;

2. ammonia enhances death;

3. Lactate reconsumption is triggered by low growth rate.

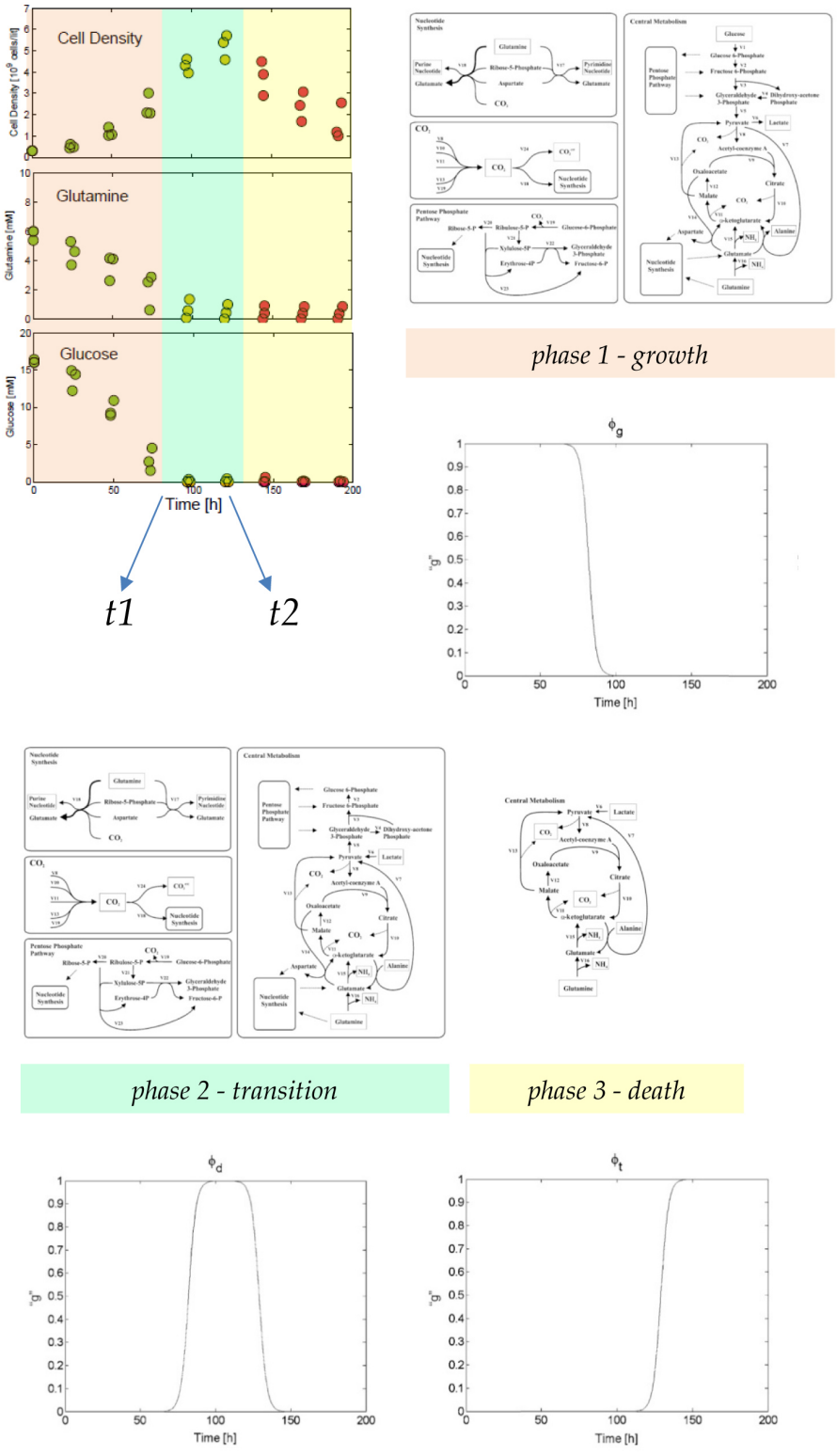


Figure 2.9: The idea behind Provost's model. Adapted from Provost (2006)

$$\Phi_{growth} = \mu_{growth} X_v \tag{2.51}$$

$$\text{where } \mu_{growth} = \mu_{max} \frac{Gln}{k_{Gln} + Gln}$$

$$\Phi_{death} = \mu_{death} X_v \tag{2.52}$$

$$\text{where } \mu_{death} = \mu_{d,max} \frac{Amm}{k_{Amm} + Amm}$$

$$\Phi_{reconsumption} = \mu_{reconsumption} X_v \tag{2.53}$$

$$\text{where } \mu_{reconsumptionth} = \alpha \frac{k_\mu}{k_\mu + \mu_{growth}}$$

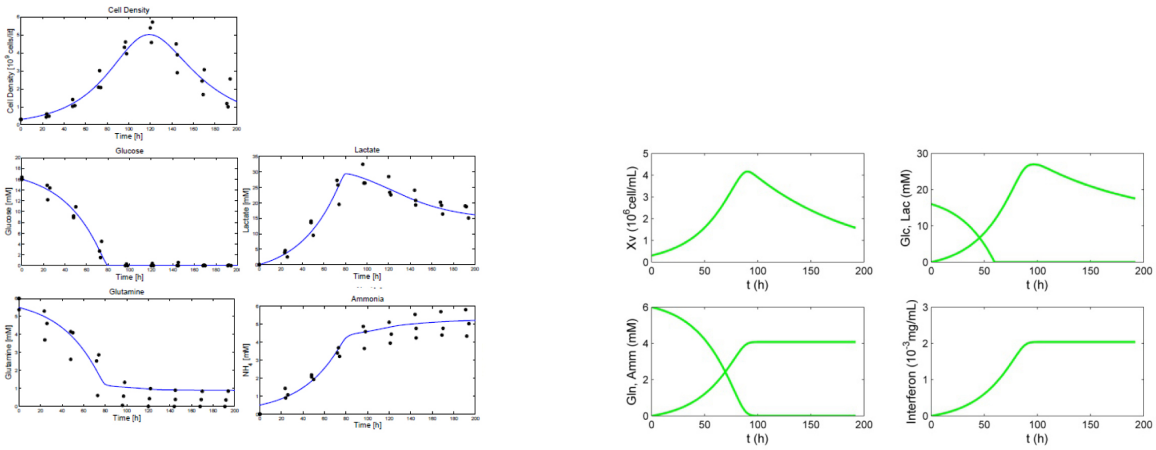


Figure 2.10: Left: Provost's model (Provost, 2006). Right: Illustrative model predictions for the same batch initial conditions.

Globally, models that consider intracellular components, such as Provost's, often give insight into cell metabolic states that may be useful mostly to biologists. Hopefully, when developing one, a reduced-order form may be found and used for a period of the culture when the balanced-growth condition holds. Hopefully also, this type of model will still be simple enough to use in a real scenario or be useful and informative in the development of regular macroscopic ODE-based models.

2.3 Segregated structured models

One classical example of not supposing, while modeling, that the culture is composed of an average cell is to consider a population of cells at different stages of their life cycle. The section below describes one such model.

2.3.1 Cell population distributed in different phases

The fact that cells in a culture may be at different phases of their life cycle is considered in Faraday et al (2001) as illustrated in Figure 2.11.

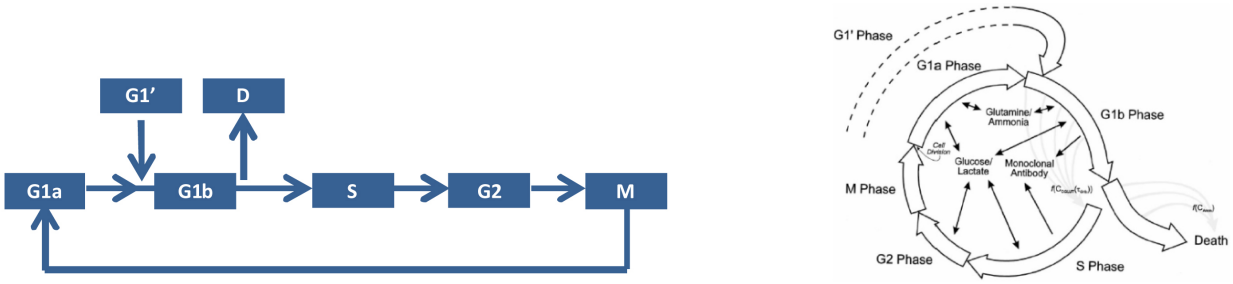


Figure 2.11: Cell life cycles considered in Faraday et al (2001)

A consequence is that the system becomes mathematically more complex. The authors suggest a procedure to approximately solve the set of first order differential equations and first order quasi linear hyperbolic partial differential equations.

The change in population density at any point of a generic phase X is given by:

$$\frac{\partial n_X(t,\tau_X)}{\partial t} = \frac{F(t)}{V(t)} n_X(t,\tau_X) - \sum_{J=1}^{G_X} r_{JX}\left(\tau_X, n_X(t,\tau_X), C_C(t,\tau_X), C(t,\tau_X)\right) - \frac{\partial n_X(t,\tau_X)}{\partial \tau} \tag{2.54}$$

where τ is the biological age (h), C is the medium state vector containing the concentrations of all the medium components of interest ($kg.m^{-3}$), C_C is the cytological state vector containing the concentrations of intra-cellular components of interest ($kg.cell^{-3}$), F is the flow rate (m^3h^{-1}), G_X is the number of transition rules in a phase, J is a transition rule, n_X is the population density function, ie. the number of cells per volume per biological age in an arbitrary

phase X ($cell.m^{-3}$), r_{JX} is the rate of transition ($cell.m^3h^{-1}$), t is time (h) and V the volume (m^3).

Some biological hypotheses are used to express the rates at which cells undergo transitions between phases. Below are the change in population density at any point of a generic phase X and the boundary conditions:

$$r_{1X}(\tau_X, n_X(t,\tau_X), C_C(t,\tau_X), C(t)) = \begin{cases} 0 & \text{if } \tau_X \in [0; T_X[\\ n_X(t,T_X) & \text{if } \tau_X = T_X \end{cases} \tag{2.55}$$

$$r_{2G1b}(\tau_{G1b}, n_{G1b}(t,\tau_X), C_C(t,\tau_X), C(t)) = \frac{2n_{G1b}}{C_{C,Gln} - S_{max}} \frac{\partial C_{Gln}}{\partial t} \tag{2.56}$$

$$r_{2D}(\tau_D, n_D(t,\tau_X), C_C(t,\tau_X), C(t)) = k_{Amm} C_{Amm}^{1.5} n_D \tag{2.57}$$

Boundary Conditions:

$$n_{G1a}(t,0) = \begin{cases} 2n_M(t,T_M) & \text{when } C_{Gln}(t) > 0 \\ 0 & \text{when } C_{Gln}(t) = 0 \end{cases} \tag{2.58}$$

$$n_{G1b}(t,0) = n_{G1a}(t,T_{G1a}) + n_{G1'}(t,T_{G1'}) \tag{2.59}$$

$$n_S(t,0) = \int_0^{T_{G1b}} \frac{2n_{G1b}(t,\tau_{G1b})}{C_{C,Gln}(t,\tau_{G1b}) - S_{max}} \frac{\partial C_{C,Gln}(t,\tau_{G1b})}{\partial t} d\tau_{G1b} \tag{2.60}$$

$$n_{G2}(t,0) = n_S(t,T_S) \tag{2.61}$$

$$n_M(t,0) = n_{G2}(t,T_{G2}) \tag{2.62}$$

$$n_D(t,0) = \begin{cases} n_{G1b}(t,T_{G1b}) & \text{for } C_{Gln} > 0 \\ n_{G1b}(t,T_{G1b}) + 2n_M(t,T_M) & \text{for } C_{Gln}(t) = 0 \end{cases} \tag{2.63}$$

More hypotheses are made regarding other state variables. For instance, substrates are consumed during certains phases (5 phases for glucose, 2 phases for glutamine). Consumption and production are expressed by:

- Glucose
  consumed during G1a, G1b, S, G2, M with first order kinetics;

$$\frac{dGlc}{dt} = \left(C^{IN}_{Glc}(t) - C_{Glc}(t)\right)\frac{F(t)}{V(t)} - R_{Glc}\left[\int_0^{T_{G1a}} n_{G1a}(t,\tau_{G1a})d\tau + \right.$$
$$+ \int_0^{T_{G1b}} n_{G1b}(t,\tau_{G1b})d\tau + \int_0^{T_S} n_S(t,\tau_S)d\tau + \int_0^{T_{G2}} n_{G2}(t,\tau_{G2})d\tau +$$
$$\left. + \int_0^{T_M} n_M(t,\tau_{G1b})d\tau\right] \tag{2.64}$$

$$R_{Glc} = k_{Glc}C_{Glc} \tag{2.65}$$

- Lactate
  produced proportionally to glucose consumed;

$$\frac{dLac}{dt} = \left(C^{IN}_{Lac}(t) - C_{Lac}(t)\right)\frac{F(t)}{V(t)} + Y_{Lac/Glc}R_{Glc}\left[\int_0^{T_{G1a}} n_{G1a}(t,\tau_{G1a})d\tau + \right.$$
$$+ \int_0^{T_{G1b}} n_{G1b}(t,\tau_{G1b})d\tau + \int_0^{T_S} n_S(t,\tau_S)d\tau + \int_0^{T_{G2}} n_{G2}(t,\tau_{G2})d\tau +$$
$$\left. + \int_0^{T_M} n_M(t,\tau_{G1b})d\tau\right] \tag{2.66}$$

- Glutamine
  consumed during G1a, G1b with zero order kinetics;

$$\frac{dGln}{dt} = \left(C^{IN}_{Gln}(t) - C_{Gln}(t)\right)\frac{F(t)}{V(t)} - R_{Gln}\left[\int_0^{T_{G1a}} n_{G1a}(t,\tau_{G1a})d\tau + \right.$$
$$\left. + \int_0^{T_{G1b}} n_{G1b}(t,\tau_{G1b})d\tau\right] \tag{2.67}$$

$$R_{Glc} = k_{Glc}C_{Glc} \tag{2.68}$$

- ammonia
  produced proportionally to glutamine consumed;

$$\frac{dAmm}{dt} = \left(C^{IN}_{Amm}(t) - C_{Amm}(t)\right)\frac{F(t)}{V(t)} - Y_{Amm/Gln}R_{Gln}\left[\int_0^{T_{G1a}} n_{G1a}(t,\tau_{G1a})d\tau + \right.$$
$$\left. + \int_0^{T_{G1b}} n_{G1b}(t,\tau_{G1b})d\tau\right] \tag{2.69}$$

- Monoclonal Antibodies
  produced during G1b, S with fixed rate.

$$\frac{dMAb}{dt} = (0 - C_{MAb}(t))\frac{F(t)}{V(t)} + R_{MAb}\left[\int_0^{T_{G1b}} n_{G1b}(t,\tau_{G1b})d\tau + \right.$$
$$\left. + \int_0^{T_S} n_S(t,\tau_S)d\tau\right] \tag{2.70}$$

The model takes the form of a system with a large list of parameters that are also difficult to estimate. The authors propose some values for a few of them:

Table 2.7: Faraday's model parameter values (Faraday et al, 2001).

Parameter	Value	Units
k_{Amm}	6.2×10^{-5}	$mL^{1.5}mg^{-1.5}h^{-1}$
k_{Gln}	3.6×10^{-8}	$mL\,cell^{-1}h^{-1}$
$Y_{Amm/Gln}$	0.1	$mg\,Amm/mg\,Gln$
$Y_{Lac/Glc}$	0.79	$mg\,Lac/mg\,Glc$
R_{Gln}	1.5×10^{-8}	$mg\,cell^{-1}h^{-1}$
S_{max}	2.6×10^{-7}	$mg\,cell^{-1}$
R_{Anti}	1.1×10^{-8}	$mg\,cell^{-1}h^{-1}$

Globally the vector of model states is rather big:

$$\frac{d\xi_i}{dt} = \begin{bmatrix} n_{G1a} \\ n_{G1'} \\ n_{G1b} \\ n_S \\ n_{G2} \\ n_M \\ Gln \\ Glc \\ Lac \\ Amm \\ MAb \end{bmatrix} = \begin{bmatrix} \text{cells in phase } G1a \\ \text{cells in phase } G1' \\ \text{cells in phase } G1b \\ \text{cells in phase } S \\ \text{cells in phase } G2 \\ \text{cells in phase } M \\ \text{glutamine} \\ \text{glucose} \\ \text{lactate} \\ \text{ammonia} \\ \text{monoclonal antibodies} \end{bmatrix} \tag{2.71}$$

The model comprises 11 states, several of which are not measurable, and more than 13 parameters. The system is of complex resolution, let alone the difficulty of identifying all its parameters. For example, in order to assess the cycle phase where a cell currently is, flow cytometry was employed - however, the method technically only differentiates phases $G1$, S and $G2+M$, meaning that the cell phase distribution and the phase length are hardly identifiable. It is also an expensive technology that is difficult to transpose to online implementation. On the whole, it is an interesting model for biological exploratory research (it could provide insight, eg. into finding out an interesting antibody production pattern, and then one could try and arrest more cells in that phase). This is unlikely to be useful in a context of real control application and, for this purpose, the data could be better modeled with a much simpler macroscopic model.

2.4 Round-up

There are many models in the literature. Some review articles such as Boghigian et al (2010); Portner and Schafer (1996); Sidoli et al (2004); Tziampazis and Sambanis (1994) shortlist some of them.

The book dos Reis Castilho (2008) provides a comprehensive outlook. Others, such as Dunn et al (2003); Torres and Voit (2002), do so as well.

More recently, further models have been proposed. For example, Nolan and Lee (2011) consider that some outer intracellular reactions have kinetic rate expressions based on extracellular metabolites and applies this hypothesis to the growth and transition phase. Another example is Amribt (2014)

who bases his modeling approach on the overflow phenomenon notorious in the growth-phase modeling of bacterial cultures.

In the scope of this thesis, the choice has fallen upon simple phenomenological macroscopic ODE-based models as the best compromise between descriptive quality, predictive power and practicality with regard to real-scenario control, monitoring and optimisation applications in view of an on-line continuous implementation of perfused high-cell-density cultures. We believe that they can provide enough information regarding phenomena that we consider important (eg. limitation, inhibition) and are yet simple enough to use for the intended purposes.

This ODE-based macroscopic model with the best joint performance and simplicity for practical control implementation will most likely consider 6 states besides volume: living biomass, glucose/lactate, glutamine/ammonia and the product of interest, if one is produced by cells.

The model will thus most likely take the following general form:

$$\frac{d\xi_i}{dt} = \nu_{ij}\varphi_j + \cdots \rightarrow \frac{d}{dt}\begin{bmatrix} Glc \\ Gln \\ Xv \\ Lac \\ Amm \\ P \end{bmatrix} = \begin{bmatrix} -\nu_{11} & 0 & 0 \\ -\nu_{21} & 0 & 0 \\ 1 & -1 & 0 \\ \nu_{41} & 0 & 0 \\ \nu_{51} & 0 & 0 \\ 0 & 0 & \nu_{63} \end{bmatrix} \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \varphi_3 \end{bmatrix} + \cdots \tag{2.72}$$

$$(-\nu_{11})\mathrm{Glc} + (-\nu_{21})\mathrm{Gln} \xrightarrow[\text{(cell growth)}]{\varphi_1} \circlearrowleft\mathrm{Xv} + \nu_{41}\mathrm{Lac} + \nu_{51}\mathrm{Amm} \tag{2.73}$$

$$(-1)\mathrm{Xv} \xrightarrow[\text{(cell death)}]{\varphi_2} (1)\mathrm{Xd} \tag{2.74}$$

$$(-1)\mathrm{Xv} \xrightarrow[\text{(product formation)}]{\varphi_3} (1)\mathrm{Xv} + (\nu_{63})\mathrm{P} \tag{2.75}$$

A data bank was built from cultures performed at the new cell laboratory of Chemistry and Applied Chemistry department of the Biosystems Pole at the University of Mons. Initially, the cultures were performed with hybridoma cells, but these proved delicate to cultivate. A more robust type of animal cells, CHO (chinese hamster ovary cells), was then used to accomplish

enough cultures to build the data bank. In the following chapter, the issue of identifying a model for these cultures will be addressed and a model suitable for control will be presented.

Chapter 3

Animal cell culture model identification

Once a model structure is chosen, the values of the parameters still need to be identified. This chapter presents one of the major contributions of this thesis. First, the identification problem is presented. In order to tackle the challenge of starting the procedure with a good initial guess for a large set of parameters, a procedure is proposed: a step-by-step identification approach that gradually considers more detailed models. This chapter presents the rationale supporting it. It is then that real data from experimental campaigns serves to illustrate it in two experimental practical case studies.

3.1 Introduction

The ultimate goal of the identification procedure is to estimate the values of parameters θ_j in the mathematical model describing the evolution of states x_i (the concentrations of biomass (living cells), the substrates that the cells are fed with, the product of interest and other metabolites that the cells produce in the course of the culture and which may affect their own growth). The

model, let's recall, is a set of equations that allows to predict culture behavior and can be used to study and control different production scenarios.

Briefly, identification considers the following steps:

- Propose a model structure;
- Propose a set of possible values for parameters θ_j;
- Simulate model predictions x based on that initial guess;
- Estimate how far this simulation is from real experimental data x_{meas} by computing a cost J;
- Implement an optimization algorithm that searches for other values for the parameters leading to a lower cost;
- The final optimal set of values θ^* for model parameters will be the one having led to the lowest value of J.

3.2 Step-by-step identification

To correctly identify parameter values from experimental data, one should firstly plan to perform experiments that are informative enough to investigate the phenomena that the model intends to describe.

Due to financial and resource constraints[1], experimental data is often limited. Many parameters are thus to be drawn from this limited data bank, which renders the procedure quite complex. One of the first hurdles is, in fact, the initialization of the algorithm aiming at minimizing the distance of model predictions to real data.

In order to overcome this, a step-by-step reduced order model identification procedure is proposed in this chapter. It is based on the analysis of the sensitivities of model outputs to changes in parameter values. Simplifying the identification problem is thus possible thanks to helpful insight brought by the analysis of these sensitivity functions $S_{xi,\theta j}$ (for example, $S_{x1,\theta 2}$ describes the evolution of how sensitive model state x_1 is to changes in the value of parameter θ_2).

On the whole, the procedure aims at being more efficient with less data by laying its foundations on one of the keys to a good performance of an

[1]Besides the culture medium, sample analysis of the concentrations can be expensive and time-consuming.

identification algorithm: to have an initial guess that is close enough to the optimal values. In order to do so, box constraints (based on some a priori biological knowledge) can sometimes be imposed in order to limit the space from where the random initial guess used to initialize the algorithm will be drawn. However, the search space remains very vast: in the case of De Tremblay's model, it is at least a 16-parameter combination set.

Initialization of the identification procedure has already been addressed in several ways. For instance Mairet et al (2011) proposed a semi-analytic procedure to estimate the parameters of Droop model (a classical model describing substrate limitation in micro-algal cultures), that can be used as a starting point for a numerical procedure. Besides analytic or semi-analytic approaches (which are restricted to relatively modest-sized models), another strategy is to use models that are linear in the parameters, or that can be linearized, such as for instance the model proposed by Grosfils et al (2007).

In this chapter, a systematic identification procedure will be proposed, based on the examination of the parametric sensitivities $S_{xi,\theta j}$. Indeed, these functions allow to distinguish specific conditions or periods of the culture during which the model states are mostly influenced by a subset of parameters. Therefore, simpler models are likely to fit the experimental data in these specific periods, and a divide and conquer approach to the identification problem can be imagined on this basis. The procedure consists in several identification steps. The parameters estimated in a previous step can be used as initial guess in the next. Step-by-step, the model is refined and the parameter set becomes more consistent. Finally, the full parameter set can be re-estimated from the estimated parameters (which hopefully are now in close distance from the optimum).

For example, given several models Ω_i:

$$\Omega_i : \dot{x}_{\Omega_i} = f_{\Omega_i}(\theta_{\Omega_i}, x_{\Omega_i}) \tag{3.1}$$

the sequence of identification steps to take, culminating in the identification of the original 6-states 16-parameters full model Ω_F, could be, for example, the one given in Table 3.1.

Globally, the chapter is organised as follows. The next section states and explains the identification problem. Section 3.4 introduces the sensitivity concept. In Section 3.5 the rationale behind the proposal of the step-by-step approach is presented with a sensitivity study using De Tremblay's model. This model is then further used in a practical simulation example that illus-

Table 3.1: Illustrative sequence of identification steps.

Step	Model	States considered	Parameters to identify	Based on initial guess	Culture phase
1	Ω_1	$x_{\Omega_1} = \begin{bmatrix} x_1 \end{bmatrix}$	$\theta_{\Omega_1} = \begin{bmatrix} \theta_1 \end{bmatrix}$	$\theta^0_{\Omega_1} = \begin{bmatrix} \theta_1, \text{random} \end{bmatrix}$	A
2	Ω_2	$x_{\Omega_2} = \begin{bmatrix} x_1 \\ x_2 \\ x_4 \end{bmatrix}$	$\theta_{\Omega_2} = \begin{bmatrix} \theta_1 \\ \theta_8 \\ \theta_{12} \end{bmatrix}$	$\theta^0_{\Omega_2} = \begin{bmatrix} \theta_1, \text{found in step 1} \\ \theta_8, \text{random} \\ \theta_{12}, \text{random} \end{bmatrix}$	A
3	Ω_3	$x_{\Omega_3} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}$	$\theta_{\Omega_3} = \begin{bmatrix} \theta_1 \\ \theta_8 \\ \theta_9 \\ \theta_{12} \\ \theta_{13} \end{bmatrix}$	$\theta^0_{\Omega_3} = \begin{bmatrix} \theta_1, \text{found in step 2} \\ \theta_8, \text{found in step 2} \\ \theta_9, \text{found in step 2} \\ \theta_{12}, \text{random} \\ \theta_{13}, \text{random} \end{bmatrix}$	A
⋮	⋮	⋮	⋮	⋮	⋮
F	Ω_F	$x_{\Omega_F} = \begin{bmatrix} x_1 \\ \vdots \\ x_6 \end{bmatrix}$	$\theta_{\Omega_F} = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_{16} \end{bmatrix}$	$\theta^0_{\Omega_F} = \begin{bmatrix} \theta_1, \text{found previously} \\ \vdots \\ \theta_{13}, \text{found previously} \\ \theta_{14}, \text{random} \\ \theta_{15}, \text{random} \\ \theta_{16}, \text{random} \end{bmatrix}$	all

trates the potential of the step-by-step identification procedure. Finally, data gathered from real animal cell cultures performed at the UMons is used in two experimental practical case studies where the proposed procedure is applied and some conclusions are drawn. The first experimental case study deals with CHO-S cells (Section 3.6) and the second experimental case study with CHO-320 cells producing a well known therapeutic biological: interferon-γ recombinant protein (Section 3.7). The particularity of the second experimental campaign (with CHO-320) is that this cell line is transfected to produce a bioproduct of pharmaceutical interest (unlike the CHO-S campaign). Therefore, the final model has in this case an additional state variable, the bioproduct concentration, and this may be interesting in the study of observability and control strategies which are subjects dealt with in the following chapters.

3.3 The identification problem

As already mentioned, given a certain mathematical model structure with the power of qualitatively describing cell culture behaviour, the identification problem consists of finding the values of parameters θ such that the model predictions x are close enough to experimental data x_{meas}. This is typically done by solving an optimization problem where the optimal set θ^* found is the one leading to the lowest value of J, a cost that depicts how far predictions are from data. A common approach is to minimize the least-squares criterion (sum of squared differences between model predictions and measurements). The algorithm can be implemented with Matlab®'s function fminsearch and ode15s solver. The states, having different physical units, can be normalised in order to vary between 0 and 1.

$$J(\theta) = \sum_{i=1}^{nm \times ne} \sum_{j=1}^{ns} \left(x_{ij}(\theta) - x_{meas,ij}(\theta)\right)^T \cdot Q_{ij}^{-1} \left(x_{ij}(\theta) - x_{meas,ij}(\theta)\right), \tag{3.2}$$

where θ is the vector of parameters to be identified, x_{ij} is the value of state j for timepoint i (nm measurements along ne experiments), $x_{meas,ij}$ are the measurements of these states, Q_{ij}^{-1} is the measurement covariance error, a symmetric positive-definite weighting matrix. This matrix can either defined by equation (3.4) if, for example, various measuring accuracies are considered, or by equation (3.3) if not (since minimizing J or J/σ would in this case lead to the same results).

$$Q_{ij} = diag\left([1 \ldots 1]\right), \tag{3.3}$$

$$Q_{ij} = diag\left([\sigma(x_1)^2 \ldots \sigma(x_j)^2]\right). \tag{3.4}$$

If the curvature of the hyper area at the minimum is small in the direction of a certain parameter, then that parameter is not being very well estimated since a change in its value does not greatly affect the cost. This curvature can be computed with the eigenvalues of the Fisher Information matrix (FIM) (Lindner and Hitzmann, 2006). The bigger they are, the bigger the curvature and, thus, the more accurate the parameter estimation. The inverse of the FIM is the Cramer-Rao lower bound that indicates the minimal variance of the parameter values which can be obtained from the selected measurements.

3.4 Sensitivity analysis

3.4.1 Introduction

As mentioned previously, one must firstly be sure that the chosen model equations are structurally able to reproduce the shape of the concentration plots observed in real data and, on a more general view, some aspects of culture behavior (eg. biomass growth decrease when the substrate is limited). The next consideration focuses on the importance of exploring how sensitive model response is to a variation of the values of the parameters. Parametric sensitivities are particularly important to assess the influence of the parameters on the model states, depending on the operating conditions. Sensitivity analysis allows either to propose more informative experiments (i.e. experiments in which parametric sensitivities take larger values and are linearly independent), or on the contrary, when the experimental conditions are imposed, to simplify the model by eliminating (or fixing at specific values) parameters that have little influence. As a byproduct, sensitivities can also be exploited in gradient-based optimization algorithms that can be used to minimize the cost function measuring the deviation between the model outputs and the experimental data.

3.4.2 Definitions

Let $y(t, \theta)$ be the response of a certain variable of the model. If parameter θ were to vary slightly, then the new response would be (Murray-Smith, 2013):

$$y(t, \theta + \Delta\theta) = y(t, \theta) + \frac{\partial y}{\partial \theta}\Delta\theta + \frac{1}{2!}\frac{\partial^2 y}{\partial \theta^2}(\Delta\theta)^2 + \dots \tag{3.5}$$

For a $\Delta\theta$ small enough, the contribution of higher order terms is ignored and equation (3.5) becomes:

$$y(t, \theta + \Delta\theta) \simeq y(t, \theta) + \frac{\partial y}{\partial \theta}\Delta\theta, \tag{3.6}$$

where $\partial y/\partial\theta$ is the first order sensitivity of $y(t, \theta)$. This linearisation allows the use of the superposition principle to find the effect of simultaneous changes of parameters (Murray-Smith, 2013).

For a state space vector, the (first order) sensitivity matrix $\boldsymbol{S}$ for $i = 1, \ldots, n$ states and $j = 1, \ldots, p$ parameters is thus defined by the elements (Keesman, 2011):

$$S_{x_i,\theta_j} = \frac{\partial x_i(t)}{\partial \theta_j} \tag{3.7}$$

For example, $S_{x_1\theta_2}(t)$ describes the evolution of how sensitive model state x_1 is to changes in the value of parameter θ_2.

Possible practical approaches to evaluate sensitivities include finite difference approximations (with some rounding errors), internal differentiation (using the chain rule and Clairaut's theorem), and use of the Taylor series for higher order sensitivities (Murray-Smith, 2013; Zivari, 2009). In this thesis internal differentiation will be used.

For common cell culture models, however, explicit expressions $x(t)$ are not available. The state dynamics are available, thus a joint numerical integration of state and sensitivity dynamics is possible:

$$\text{Model:} \quad \frac{dx_i}{dt} = f_i(x, \theta), \tag{3.8}$$

$$\text{Sensitivities:} \quad \frac{dS_{x_i,\theta_j}}{dt} = \partial f_i(x,\theta) \sum_{i=1}^{n} \frac{1}{\partial x_i} S_{x_i,\theta_j} + \frac{\partial f_i(x,\theta)}{\partial \theta_j} =$$

$$= \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix} \begin{bmatrix} S_{x_1,\theta_1} & \cdots & S_{x_1,\theta_p} \\ \vdots & \ddots & \vdots \\ S_{x_n,\theta_1} & \cdots & S_{x_n,\theta_p} \end{bmatrix} + \begin{bmatrix} \frac{\partial f_1}{\partial \theta_1} & \cdots & \frac{\partial f_1}{\partial \theta_p} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial \theta_1} & \cdots & \frac{\partial f_n}{\partial \theta_p} \end{bmatrix}. \tag{3.9}$$

Given the physical nature of the system in study, we have chosen to work with dimensionless sensitivities. These are, thus, normalised because parameter values assume different numerical amplitudes and states have different ranges of variation during their open-loop response. The procedure consists of dividing parameters θ_j by nominal values $\theta_{j,nom}$ and states x_i by the maximum values assumed during the set of experiments considered (the minimum value is always zero, since we are dealing with concentrations and volume). Thus sensitivities will have a range of variation around $[-1; 1]$:

$$S_{i,j}^{natural} = \frac{\partial x_i}{\partial \theta_j} \quad \xrightarrow{normalization} \quad S_{i,j} = \frac{\partial (x_i/x_{i,max})}{\partial (\theta_j/\theta_{j,nom})} = S_{i,j}^{natural} \frac{\theta_{j,nom}}{x_{i,max}}. \tag{3.10}$$

3.5 Case study: De Tremblay's model

In order to lay out the fundamentals of the step-by-step model identification procedure, let us firstly consider a real dynamic model of hybridoma cells producing monoclonal antibodies (de Tremblay et al, 1992). It comprises 7 states (concentrations of biomass, glucose, lactate, glutamine, ammonia, monoclonal antibodies and volume), 16 parameters and describes typical animal cell culture phenomena.

Four biological macroreactions are considered[1]:

$$(-\nu_{21})\text{Glc} + (-\nu_{41})\text{Gln} \xrightarrow{\varphi_1} \circlearrowleft\text{Xv} + (\nu_{31})\text{Lac} + (\nu_{51})\text{Amm}, \tag{3.11}$$

$$(-1)\text{Xv} \xrightarrow{\varphi_2} \text{Xd}, \tag{3.12}$$

$$(-1)\text{Glc} + (-\nu_{13})\text{Xv} \xrightarrow{\varphi_3} \nu_{13}\text{Xv} + \nu_{33}\text{Lac}, \tag{3.13}$$

$$(-\nu_{14})\text{Xv} \xrightarrow{\varphi_4} \nu_{14}\text{Xv} + \text{MAb}, \tag{3.14}$$

Mass balance equations are given by:

$$\frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1/\theta_8 & 0 & -1 & 0 \\ \theta_9/\theta_8 & 0 & \theta_9 & 0 \\ -1/\theta_{12} & 0 & 0 & 0 \\ \theta_{13}/\theta_{12} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \varphi_3 \\ \varphi_4 \end{bmatrix} + \begin{bmatrix} -x_1 & x_1 \\ (-x_2 + k_1) & 0 \\ -x_3 & 0 \\ (-x_4 + k_2) & 0 \\ -x_5 & 0 \\ -x_6 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}. \tag{3.15}$$

For the sake of simplicity, the following symbols represent states, kinetic parameters, constants and inputs: $x_1 = \text{Xv}$; $x_2 = \text{Glc}$; $x_3 = \text{Lac}$; $x_4 = \text{Gln}$; $x_5 = \text{Amm}$; $x_6 = \text{MAb}$; $\theta_1 = \mu_{max}$; $\theta_2 = k_{Glc}$; $\theta_3 = k_{Gln}$; $\theta_4 = \mu_{d,max}$; $\theta_5 = k_{d,Lac}$; $\theta_6 = k_{d,Amm}$; $\theta_7 = k_{d,Gln}$; $\theta_8 = Y_{Xv/Glc} = 1/\nu_{21}$; $\theta_9 = Y_{Lac/Glc} = \nu_{31}\theta_8$; $\theta_{10} = m_{Glc}$; $\theta_{11} = k_{m,Glc}$; $\theta_{12} = Y_{Xv/Gln} = 1/\nu_{41}$; $\theta_{13} = Y_{Amm/Gln} = \nu_{51}\theta_{12}$; $\theta_{14} = \beta$; $\theta_{15} = \alpha$; $\theta_{16} = k_\mu$; $k_1 = Glc^{IN}$; $k_2 = Gln^{IN}$; $u_{10} = F^{IN}$; $u_{20} = F_{perf}$; $u_{30} = F_{bleed}$; $u_1 = D = F^{IN}/V$; $u_2 = D_{perf} = F_{perf}/V$.

[1]These describe cell growth, cell death, cell maintenance and bioproduct production, respectively.

The reaction rates are given by:

$$\varphi_i = \mu_i x_1, \tag{3.16}$$

$$\mu_1 = \theta_1 \cdot \frac{x_2}{(\theta_2 + x_2)} \cdot \frac{x_4}{(\theta_3 + x_4)}, \tag{3.17}$$

$$\mu_2 = \theta_4 \cdot \frac{1}{(\theta_1 - \theta_5 x_3)} \cdot \frac{1}{(\theta_1 - \theta_6 x_5)} \cdot \frac{\theta_7}{(\theta_7 + x_4)}, \tag{3.18}$$

$$\mu_3 = \theta_{10} \cdot \frac{x_2}{(\theta_{11} + x_2)}, \tag{3.19}$$

$$\mu_4 = \theta_{14} + \theta_{15} \cdot \frac{\mu_1}{\theta_{16} + \mu_1}. \tag{3.20}$$

If volume is not constant then it needs to be included in the model ($V = x_7$). In this case, besides the reaction terms $\nu\varphi$, the flow dynamics terms are to be considered:

$$\frac{dx_7}{dt} = \begin{cases} 0 & ,batch \\ u_{10} & ,fedbatch \\ u_{10} - u_{30} & ,continuous \\ u_{10} - u_{30} - u_{20} & ,continous\ perfused \end{cases} \tag{3.21}$$

A schematic illustration is provided in Fig. 1.1 on p. 31 for different operation regimes. Parameter values are listed below.

De Tremblay's model proposes the interdependencies of system dynamics described in Table 3.3. Biomass dynamics are influenced by almost all states (itself, substrates consumed and metabolites produced). However, the production of the product of interest (x_6, monoclonal antibodies) does not dynamically influence any of the other states. Therefore, parameters related to the dynamics of x_1 to x_5 could be identified firstly without any prior knowledge of the antibodies being required.

As for the dynamics of the sensitivities, they are obtained by differentiation. The terms $\partial f_i/\partial x_i$ and $\partial f_i/\partial \theta_i$ defined in equation (3.9)) were computed and are listed in Appendix A, equations (A.1) to (A.99).

Table 3.2: De Tremblay's model parameter values (de Tremblay, 1991; de Tremblay et al, 1992, 1993).

Parameter	Value	Units
μ_{max}	1.09	d^{-1}
k_{Glc}	1.0	mM
k_{Gln}	0.3	mM
$\mu_{d,max}$	0.09	d^{-1}
$k_{d,Lac}$	0.01	$mM^{-1}d^{-1}$
$k_{d,Amm}$	0.06	$mM^{-1}d^{-1}$
$k_{d,Gln}$	0.02	mM
$Y_{Xv/Glc}$	1.09×10^{-1}	$10^9 cell\, mmol^{-1}$
$Y_{Lac/Glc}$	1.8	$mmol\, mmol^{-1}$
m_{Glc}	$0.17 \times 10^{+1}$	$mmol\, (10^9 cell)^{-1}\, d^{-1}$
$k_{m,Glc}$	19.0	mM
$Y_{Xv/Gln}$	3.80×10^{-1}	$10^9 cell\, mmol^{-1}$
$Y_{Amm/Gln}$	0.85	$mmol\, mmol^{-1}$
β	$0.35 \times 10^{+1}$	$mg\, (10^9 cell)^{-1}\, d^{-1}$
α	$2.57 \times 10^{+1}$	$mg\, (10^9 cell)^{-1}\, d^{-1}$
k_μ	0.02	d^{-1}

Table 3.3: Dependency of state dynamics on system states for De Tremblay's model.

dynamics	x_1	x_2	x_3	x_4	x_5	x_6	state	type
$\dot{x_1}$	●	●	●	●	●		biomass	◯
$\dot{x_2}$	●	●	●	●	●		glucose	□
$\dot{x_3}$	●	●	●	●	●		lactate	△
$\dot{x_4}$	●	●	●	●	●		glutamine	□
$\dot{x_5}$	●	●	●	●	●		ammonia	△
$\dot{x_6}$	●	●	●	●	●	●	antibodies	△

Legend: ◯=biomass; □=substrate; △=metabolite/product.

3.5.1 The experimental data bank

As already mentioned, one of the purposes of modeling is to predict behaviour that is observed in reality. In the case of animal cell cultures, it is usually expected from the model that it predicts real phenomena such as substrate limiting cell growth when it becomes scarce, metabolites building up and negatively affecting cell growth, etc. Therefore, when setting out to identify a model it is of utmost importance that these phenomena are present in the experimental data bank that will be considered.

Figure 3.1 exemplifies the predictions of De Tremblay's model for 2 different initial culture conditions. In the pink experiment, there is plenty of substrate glucose, but at $t = 2d$ the other substrate, glutamine, runs out ($\mu_2\,(t > 2) = 0$, no growth). On the other hand, there is plenty of glucose and its associated metabolite, lactate, builds up throughout the experiment during a second phase where the net cell growth is negative ($\mu_1 - \mu_2 < 0$, thus, Xv decreases on this second phase). The blue experiment illustrates the opposite: the substrate limiting growth is glucose and the excess glutamine contributes to more ammonia.

The two phases of the culture are easily identifiable on Figure 3.1. There is an initial cell growth phase (depicted as "A") where substrates are initially abundant and become gradually scarce, and a subsequent cell death phase ("B") when $Xv(t)$ begins to decrease because a vital substrate is no longer available and/or inhibiting metabolites have built up. In essence, $Xv(t)$ starts to decrease after t_{peak} (time when $Xv(t)$ reaches its maximum) because overall net cell growth $\mu_1 - \mu_2$ has suddenly become negative.

Net cell growth is composed of several terms f_i. The biomass balance equation (for batch operation) is given by:

$$\frac{dXv}{dt} = (\mu_1 - \mu_2)Xv = \left(\mu_{max} \overbrace{\frac{Glc}{k_{Glc} + Glc}}^{f1} \overbrace{\frac{Gln}{k_{Gln} + Gln}}^{f2} \right) Xv -$$

$$- \left(\mu_{d,max} \overbrace{\frac{1}{\mu_{max} - k_{d,Lac} Lac}}^{f3} \overbrace{\frac{1}{\mu_{max} - k_{d,Amm} Amm}}^{f4} \overbrace{\frac{k_{d,Gln}}{k_{d,Gln} + Gln}}^{f5} \right) Xv \tag{3.22}$$

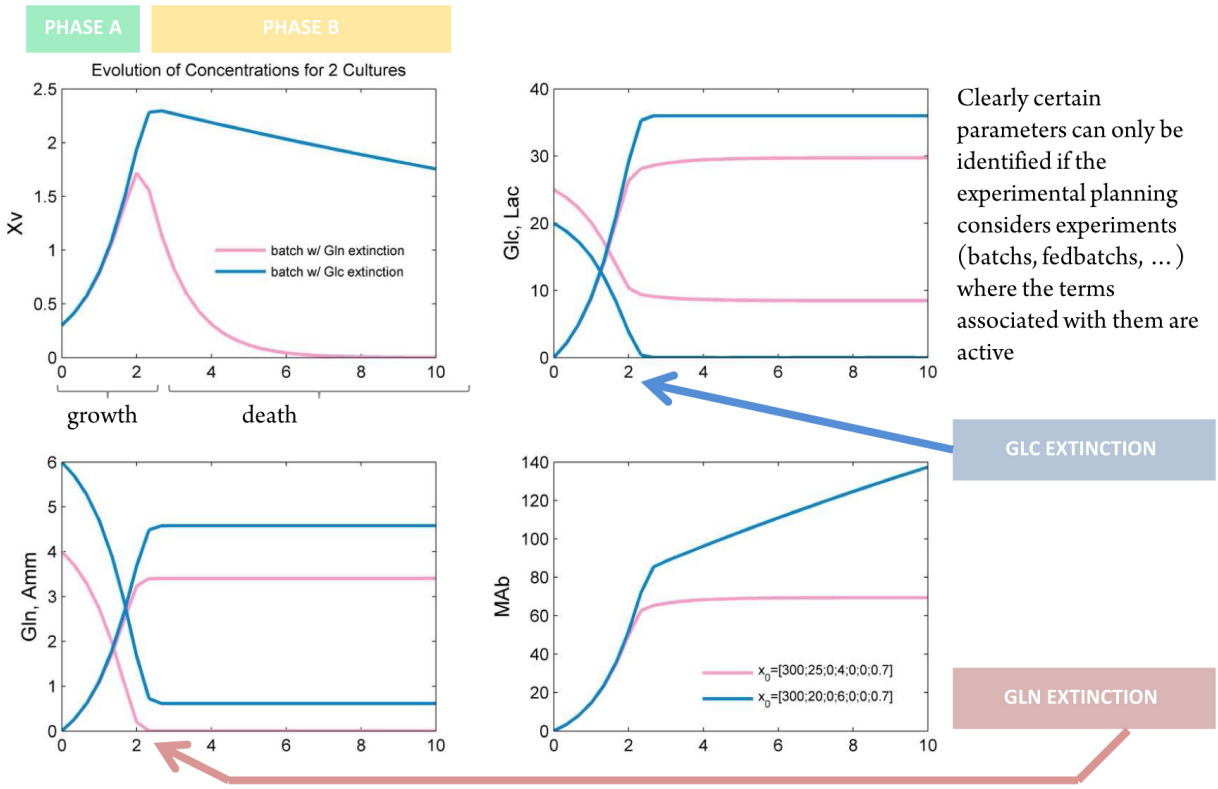


Figure 3.1: Open-loop response of the system for 2 batch cultures: in the pink culture the substrate that firstly runs out is glucose; in the blue culture, the substrate that runs out first is glutamine.

If a parameter is not very active throughout an experiment, then model states should not be greatly influenced by small changes in its value. This will be dealt further on with the sensitivity analysis.

Regarding animal cell cultures, given the phenomena involved and a pair of major substrates considered, it is recommended, whenever possible, the strategic inclusion of at least the following 4 experiments in the data bank:

- #1: both substrates never run out, death is triggered by inhibiting metabolites;
- #2: substrate S_1 runs out, metabolite M_2 accumulates;
- #3: substrate S_2 runs out, metabolite M_1 accumulates;
- #4: experiment at the expected operating conditions, used for cross-validation.

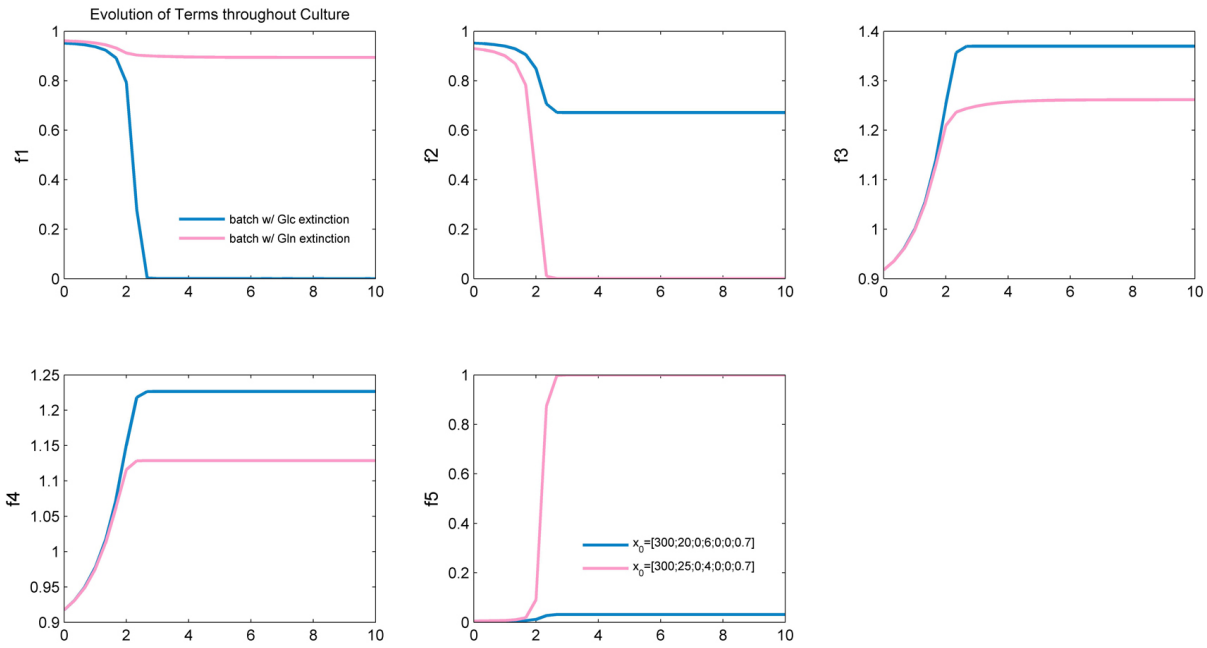


Figure 3.2: Two cultures: evolution of terms f_i for glucose-limited experiment (blue) and glutamine-limited experiment (pink).

Figure 3.2 shows that these terms are not always active. For example, in the pink experiment there is always enough glucose throughout the culture, thus the term dealing with glucose limitation of growth, f_1, is naturally not active and remains close to 1 for all t.

This minimum experimental set will allow, in principle, to capture the limitation of both substrates and the inhibition by associated metabolites. The parameters found with data sets 1-3 could then ideally be cross-validated against data set 4, close to the desired daily operating conditions. Given common culture media composition, typically S_1/M_1 will be the pair glucose/lactate and S_2/M_2 glutamine/ammonia.

3.5.2 Sensitivity analysis

The sensitivities for De Tremblay's 16-parameter model have been computed (equation terms in Appendix A.1, p. 220-227) for several experimental conditions. An example of their evolution throughout culture time is plotted in Fig. 3.3 up to Fig. 3.9, for batch initial condition $x(t_0) = [0.3\ 25\ 4\ 0\ 0\ 0\ 0.7]^T$. The grey vertical line separates phase A from phase B when $t = t_{peak}$ (maximum

of biomass concentration).

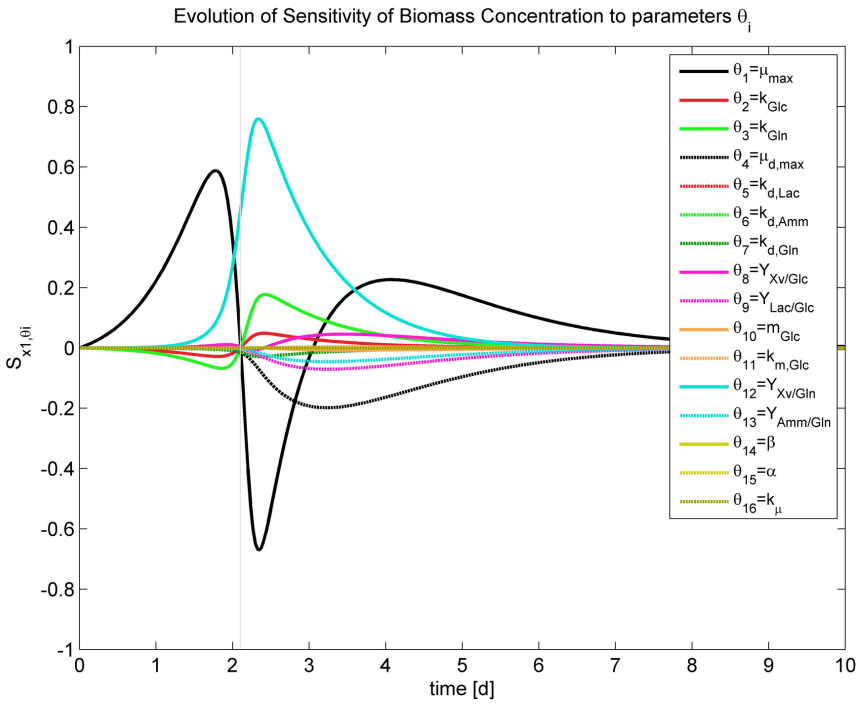


Figure 3.3: $S_{x_1,\theta_i}(t)$ for batch initial condition $x(t_0) = [0.3\ 25\ 4\ 0\ 0\ 0\ 0.7]^T$.

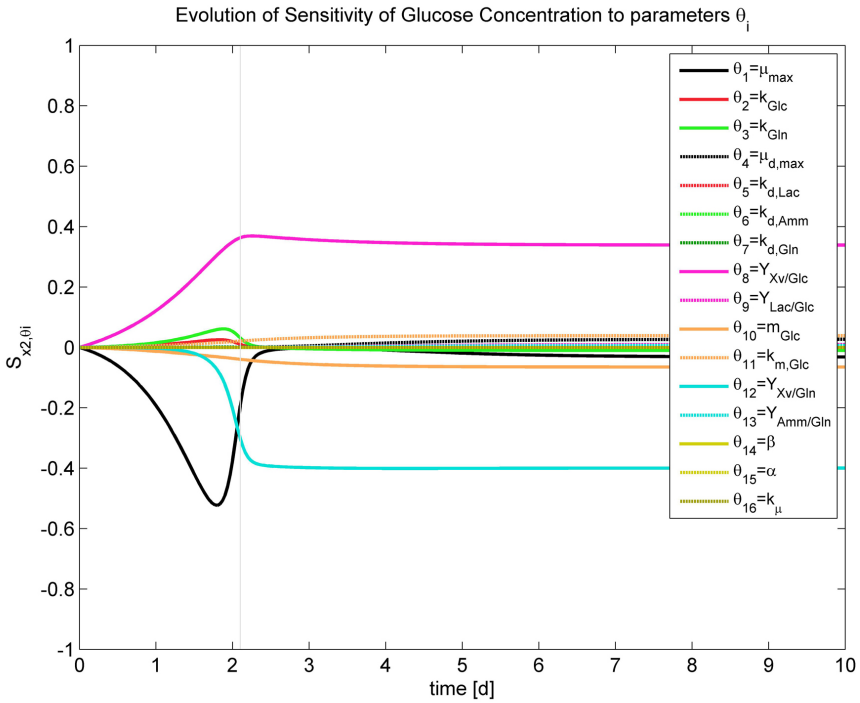


Figure 3.4: $S_{x_2,\theta_i}(t)$ for batch initial condition $x(t_0) = [0.3\ 25\ 4\ 0\ 0\ 0\ 0.7]^T$.

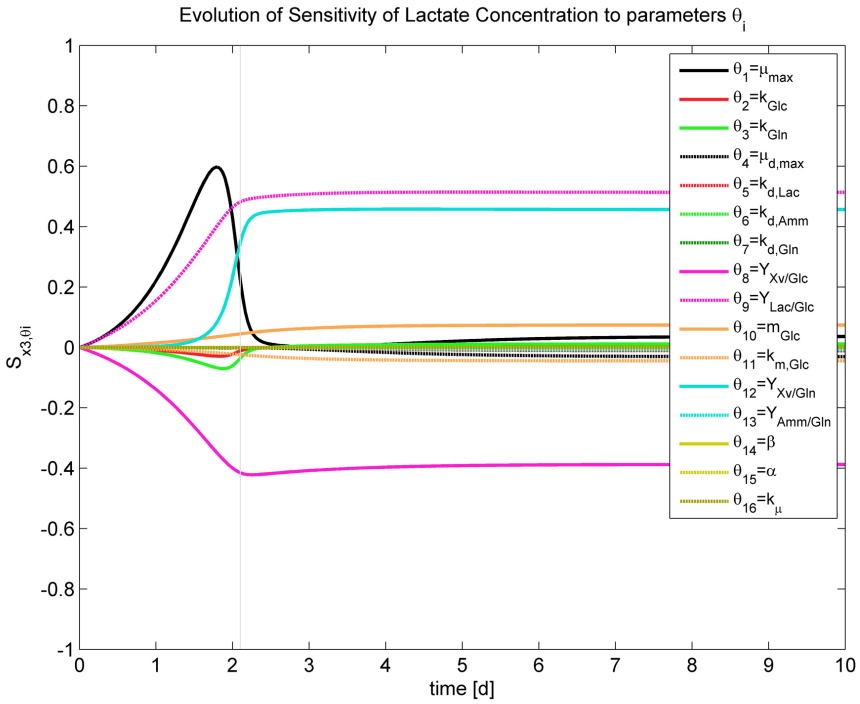


Figure 3.5: $S_{x_3,\theta_i}(t)$ for batch initial condition $x(t_0) = [0.3\ 25\ 4\ 0\ 0\ 0\ 0.7]^T$.

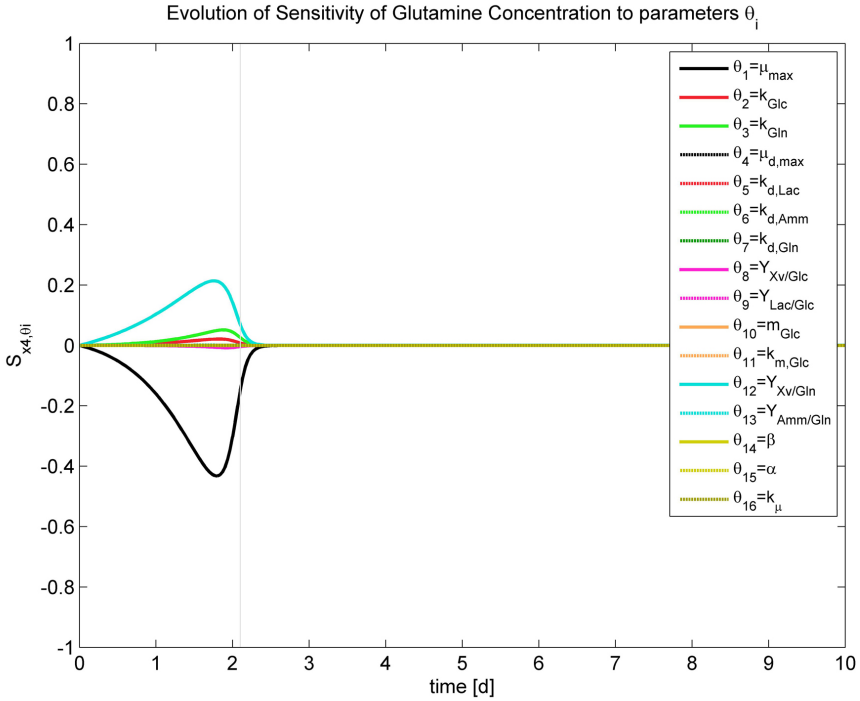


Figure 3.6: $S_{x_4,\theta_i}(t)$ for batch initial condition $x(t_0) = [0.3\ 25\ 4\ 0\ 0\ 0\ 0.7]^T$.

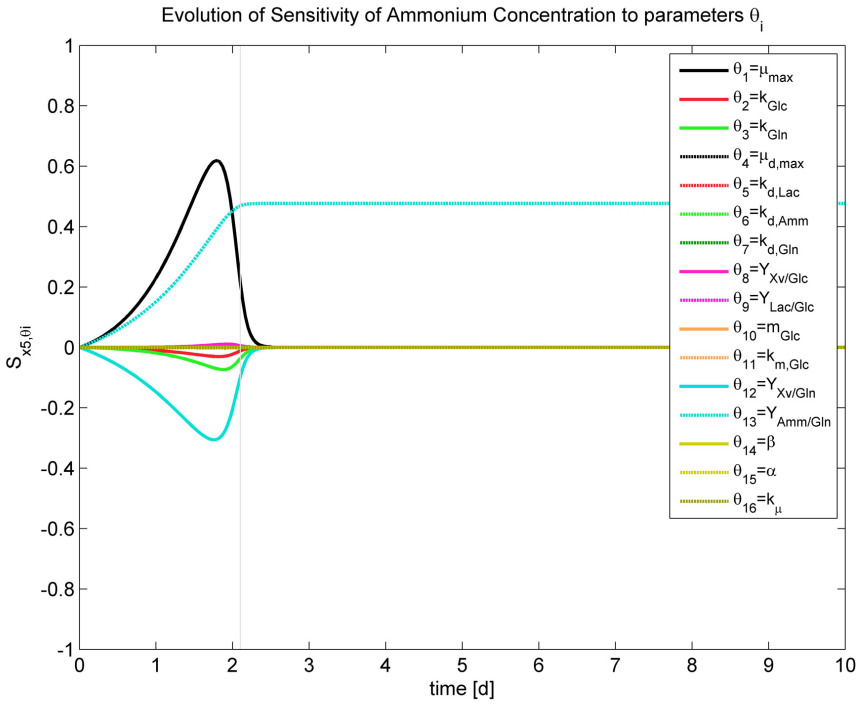


Figure 3.7: $S_{x_5,\theta_i}(t)$ for batch initial condition $x(t_0) = [0.3\ 25\ 4\ 0\ 0\ 0\ 0.7]^T$.

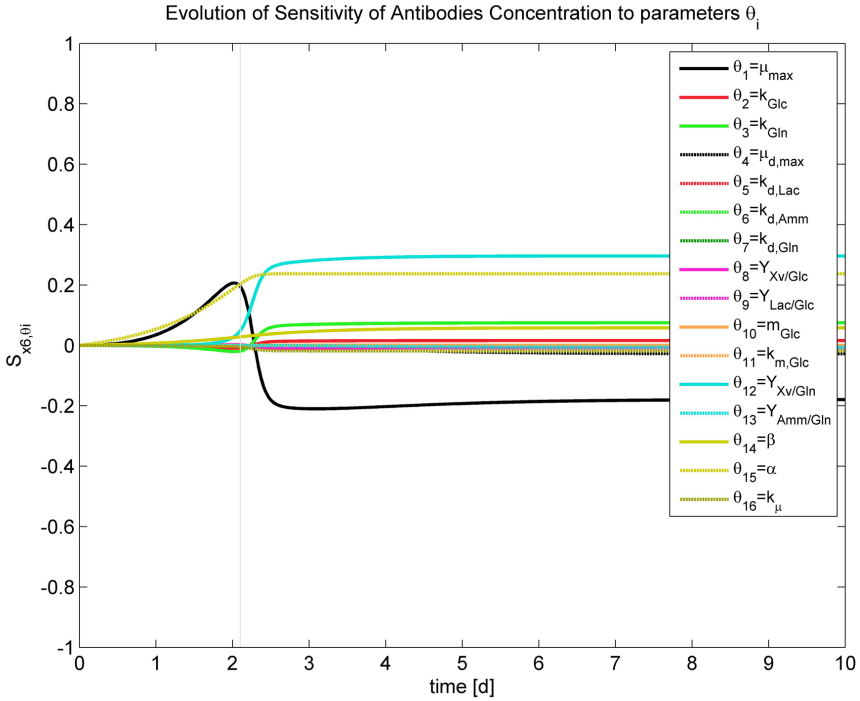


Figure 3.8: $S_{x_6,\theta_i}(t)$ for batch initial condition $x(t_0) = [0.3\ 25\ 4\ 0\ 0\ 0\ 0.7]^T$.

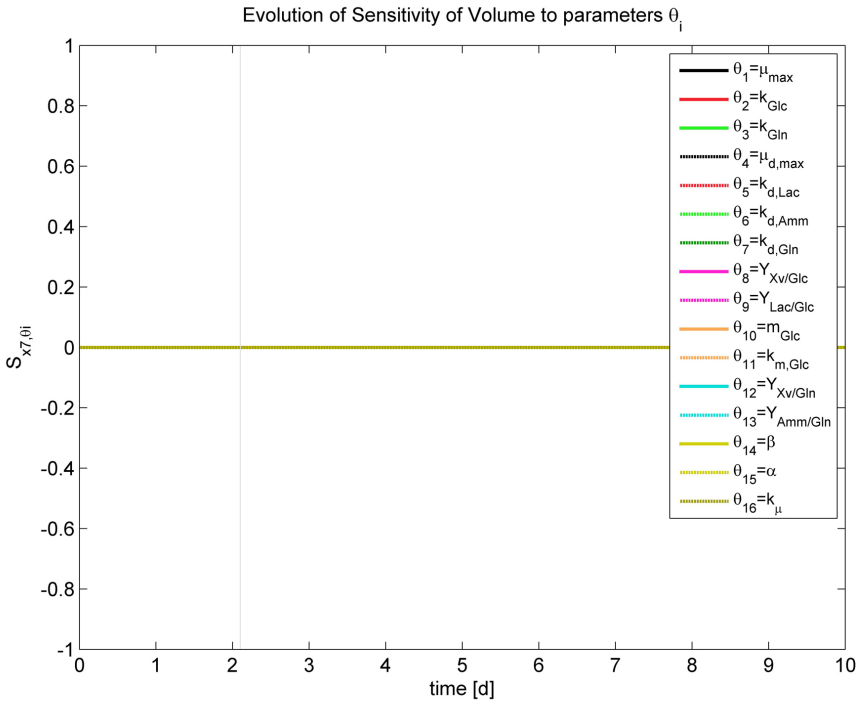


Figure 3.9: $S_{x_7,\theta_i}(t)$ for batch initial condition $x(t_0) = [0.3\ 25\ 4\ 0\ 0\ 0\ 0.7]^T$.

Several observations can be drawn from these figures. Firstly, a trivial observation: volume is obviously independent of any parameter value (its dynamics do not depend on kinetics). This can be seen in Figure 3.9.

Secondly, generally speaking, there is clearly a difference of sensitivities during phase A and phase B. In fact, if one analyses model equations, it is evident that during the initial growth phase, system dynamics is much more simple. During this phase A, stretching from t_0 till t_{peak}, biomass growth is approximately exponential and substrate consumption and product production follow proportionally. The simplest way of modeling phase A is thus:

$$\begin{cases} \frac{dx_1}{dt} = \theta_1 x_1, \\ \frac{dx_2}{dt} = \frac{-1}{\theta_8}\theta_1 x_1, \quad \frac{dx_3}{dt} = \frac{\theta_9}{\theta_8}\theta_1 x_1, \\ \frac{dx_4}{dt} = \frac{-1}{\theta_{12}}\theta_1 x_1, \quad \frac{dx_5}{dt} = \frac{\theta_{13}}{\theta_{12}}\theta_1 x_1, \end{cases} \tag{3.23}$$

This reduced model can be decomposed into smaller ones if one analyses the dependencies of system dynamics to state variables:

$$\begin{cases} \frac{dx_1}{dt} = f(x_1); \\ \frac{dx_2}{dt} = constant \times \frac{dx_1}{dt}; \\ \frac{dx_3}{dt} = constant \times \frac{dx_2}{dt}; \end{cases} \tag{3.24}$$

$$\begin{cases} \frac{dx_1}{dt} = f(x_1) \\ \frac{dx_4}{dt} = constant \times \frac{dx_1}{dt}; \\ \frac{dx_5}{dt} = constant \times \frac{dx_4}{dt}; \end{cases} \tag{3.25}$$

Namely, apart from the 5-variable model (equations (3.23)), a smaller 3-variable (biomass and substrates) model could be considered or a just 1-variable (biomass) one:

$$\begin{cases} \frac{dx_1}{dt} = \theta_1 x_1 \\ \frac{dx_2}{dt} = \frac{-1}{\theta_8}\theta_1 x_1 \\ \frac{dx_4}{dt} = \frac{-1}{\theta_{12}}\theta_1 x_1 \end{cases} \tag{3.26}$$

$$\begin{cases} \frac{dx_1}{dt} = \theta_1 x_1 \end{cases} \tag{3.27}$$

Sensitivity analysis specifically supports this model reduction since it is clear from Fig. 3.3 up to Fig. 3.9 that state variables are only significantly influenced by a few parameters during phase A: noticeably the maximum growth rate $\theta_1 = \mu_{max}$ and substrate/product coefficients $\theta_8 = Y_{Xv/Glc}$, $\theta_9 = Y_{Lac/Glc}$, $\theta_{12} = Y_{Xv/Gln}$, $\theta_{13} = Y_{Amm/Gln}$.

However, growth is not infinite. There is a moment in the culture, t_{peak}, where biomass Xv(t) reaches it peak and begins to decrease. This is what makes animal cell models more complicated than, for example, bacterial culture models: they are usually meant to predict also the death phase, meaning that the term $\mu_{growth} - \mu_{death}$ is allowed to become negative so that Xv(t) can decrease:

$$\frac{dXv}{dt} = \left(\mu_{growth} - \mu_{death}\right) Xv \quad \text{(batch regime)} \tag{3.28}$$

The term can become null or negative when phenomena such as substrate limitation or metabolite inhibition take place.

As for limitation, its constants, $\theta_2 = k_{Glc}$ and $\theta_3 = k_{Gln}$, are of more impact around t_{peak}. In order to identify them, at least two experiments should be considered: one where cell death is triggered by glucose running out, and the other by glutamine doing so. Although sensitivities to these parameters are much smaller than the sensitivities to $\theta_1 = \mu_{max}$ (the parameter with the biggest impact on states), the sensitivity of $\theta_2 = k_{Glc}$ is higher when glucose limitation occurs and the sensitivity for $\theta_3 = k_{Gln}$ is higher when it is glutamine.

In fact, sensitivity analysis could be used to optimise the experiments to perform such that the impact of a specific parameter will be maximal. Table 3.4 exemplifies briefly some interesting initial concentrations of glucose and glutamine in the medium of a 10-day batch such that model sensitivity to $\theta_2 = k_{Glc}, \theta_3 = k_{Gln}, \ldots$ would be maximised.

Table 3.4: De Tremblay's model: initial medium conditions favouring high sensitivities to specific parameters.

Parameter	$Glc_0(mM)$	$Gln_0(mM)$	Cost J obtained
$\theta_1 = \mu_{max}$	23	6.2	18
$\theta_2 = k_{Glc}$	8.8	2.4	14
$\theta_3 = k_{Gln}$	24	6.5	18
$\theta_4 = \mu_{d,max}$	23	6.3	18
$\theta_5 = k_{d,Lac}$	21	5.5	17

Note: optimization initiated with randomly generated initial medium concentration within plausible bounds. The cost maximized is the positive area (trapeze rule) of all the sensitivities for a given parameter i: $\sum_{x_1,\ldots,x_5} \int \|S_{x_i}\| dt$.

The optimization algorithm needs, however, to use parameter values in order to make model predictions and compute the cost. Yet, at the beginning of the identification task, there may be no prior certain knowledge of these values: the goal of the identification lays precisely in finding them out. But once some estimates are found, then performing this optimization could in reverse validate the experimental planning that was carried out.

Figure 3.3 up to Figure 3.9 also reveal that most other model parameters have a significantly lower impact on system states, but their action spreads out throughout both phases, A and B. Overall, these parameters may be helpful in fine-tuning the data to the form provided by the mathematical model.

If the model is over-parameterized, then it will be hard to identify some of these lesser-impact parameters. But hopefully all major impact parameters will be identified following a carefully planned sequential identification procedure.

This sequential strategy was tried out to identify some of De Tremblay's model parameters with promising results. The submodels Ω_i considered and the fits obtained are illustrated below.

- Step #1 (phase A):

$$\Omega_1 : \begin{cases} \dfrac{dx_1}{dt} = \theta_1 x_1 \end{cases} \tag{3.29}$$

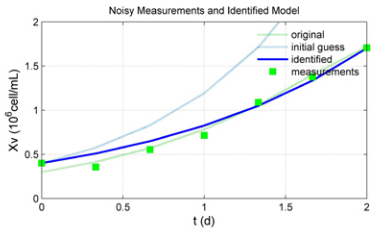


Figure 3.10: Step 1: data (green) and model predictions (blue) for submodel Ω_1 during initial phase A.

- Step #2 (phase A):

$$\Omega_2 : \begin{cases} \dfrac{dx_1}{dt} = \theta_1 x_1, & \\ \dfrac{dx_2}{dt} = \dfrac{-1}{\theta_8}\theta_1 x_1, & \dfrac{dx_3}{dt} = \dfrac{\theta_9}{\theta_8}\theta_1 x_1, \\ \dfrac{dx_4}{dt} = \dfrac{-1}{\theta_{12}}\theta_1 x_1, & \dfrac{dx_5}{dt} = \dfrac{\theta_{13}}{\theta_{12}}\theta_1 x_1, \end{cases} \tag{3.30}$$

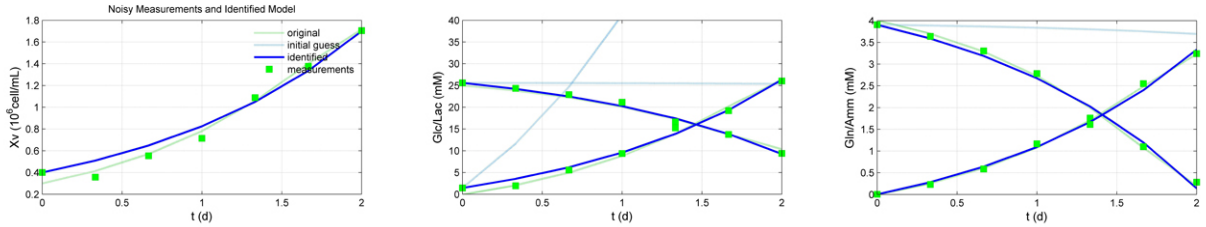


Figure 3.11: Step 2: data (green) and model predictions (blue) for submodel Ω_2 during initial phase A.

- Step #3 (phases A and B):

$$\Omega_3 : \begin{cases} \dfrac{dx_1}{dt} = \left(\theta_1 \cdot \dfrac{x_2}{(\theta_2 + x_2)} \cdot \dfrac{x_4}{(\theta_3 + x_4)}\right) x_1 - \theta_4 x_1, & \\ \dfrac{dx_2}{dt} = \dfrac{-1}{\theta_8}\theta_1 x_1, & \dfrac{dx_3}{dt} = \dfrac{\theta_9}{\theta_8}\theta_1 x_1, \\ \dfrac{dx_4}{dt} = \dfrac{-1}{\theta_{12}}\theta_1 x_1, & \dfrac{dx_5}{dt} = \dfrac{\theta_{13}}{\theta_{12}}\theta_1 x_1, \end{cases} \tag{3.31}$$

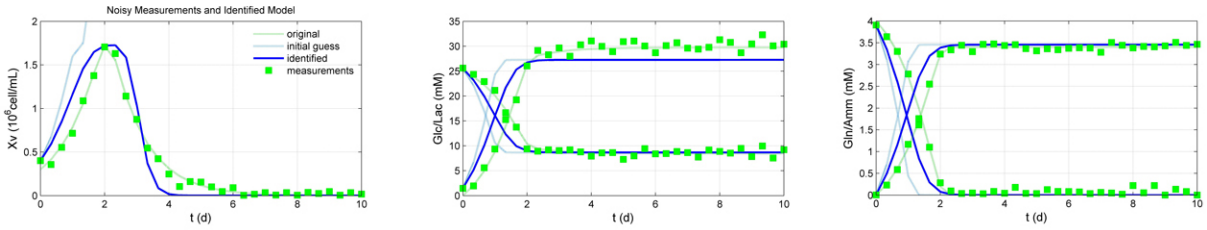


Figure 3.12: Step 3: data (green) and model predictions (blue) for submodel Ω_3 throughout the culture (phases A and B).

By the end of step 3, the 8 most significant parameters in terms of impact on system states have already been roughly identified. The next step would then be, from this knowledge, to identify the remaining ones ($\theta_5 = k_{d,Lac}, \theta_6 = k_{d,Amm}, \theta_7 = k_{d,Gln}, \theta_{11} = m_{Glc}, \theta_{12} = k_{m,Glc}$) if the available data bank is sufficiently informative. As for the parameters concerning antibody production, $\theta_{14} = \beta, \theta_{15} = \alpha, \theta_{16} = k_\mu$, they can be identified *a posteriori*, since the dynamics are decoupled as previously shown on Table 3.3 from page 86.

3.6 Step-by-step identification: CHO-S cells

The step-by-step approach will now be applied to the identification of models for animal cell cultures performed at the University of Mons in a newly set-up cell culture laboratory. A brief presentation of the site can be found in Appendix D on p. 253-254.

3.6.1 Materials and methods

Two CHO cell lines were kindly provided by Dr Emmanuelle Adam (Institute of Molecular Biology and Medicine (IBMM), Université Libre de Bruxelles, Belgium)[1]: CHO-S and CHO-S clone 4922-69 transfected to produce hypoallergenic ProDer p 1 FC (a precursor of Der p 1, a major dust mite allergen). CHO-S cells derive from CHO-K1 and have been adapted to grow in suspension. The first successful cultures performed at the new laboratory with this cell line in shake flasks and bioreactor are described in (Zamorano, 2012), but do not provide enough informative richness for the model identification intended for this chapter.

[1]Further details about this laboratory, the OCPAM project and the collaboration with the ULB are given in Appendix D on page 253.

Therefore, a new set of experiments was designed in which initial conditions varied such that different phenomena could take place and be captured by a more comprehensive dynamical model. The new experiments were carried out with CHO-S cultivated with serum-free glucose-free chemically defined PowerCHO medium (Lonza, cat. nr. BE02-042Q) supplemented with glucose (AppliChem, cat. nr. 3666) and glutamine (Sigma, cat. nr. G7513) so as to obtain a data bank with different initial concentrations:

Table 3.5: Experimental planning for 4 batch CHO-S cultures.

Code	B1	B2	B3	B4
Glc_0 (mM)	33	25	8	16
Gln_0 (mM)	7	4	8	4

B1 was meant as an experiment where glucose and glutamine are overabundant, thus, metabolite inhibition phenomena may be particularly important in explaining how the $\mu_g - \mu_d$ term becomes negative during the cell concentration decrease phase; B2 was meant as an experiment at presumably good operational conditions (those used also in de Tremblay et al (1993)'s examples); B3 as an experiment where there may be a shortage of glucose (limitation); and B4 a shortage of glutamine (limitation).

Each experiment was performed in triplicate (flasks A, B, C) in order to discard possible contamination effects or unusual behaviour and to account for the inherent biological variability: in biomedical statistical sciences the true value of a variable is added to measurement errors, in addition to intrinsic biological variability which is generally bigger (De Maertelaer, 2014)[1]. The sampling took place once per day with a schedule that allowed the gathering of the necessary information whilst not losing more than 15% of the initial volume through sampling:

The twelve $120mL$ culture flasks, each with $60mL$ of medium, were inoculated with a cell density around 0.2 to $0.3 \times 10^9 cell/L$ and kept in an incubator at $37°C$ under 5% of CO_2. To prevent contamination, $2mL$ of antibiotics (Sigma, cat. nr. P4333) were added to each $200mL$ custom medium bottle.

[1]Not all cells behave like an "average cell", there is biological variability. When measuring biomass concentration, for example, it is sensible to estimate this biological variability as a more important cause of data dispersion than errors related to the measurement probes.

Table 3.6: Sampling planning for the triplicate flasks of each experiment.

Day	1	2	3	4	5	6	7	8	9	10	11
flask A	•	•		•	•		•	•		•	•
flask B	•	•	•		•	•		•	•		•
flask C	•		•	•		•	•		•	•	

Note: first sample is taken one hour after inoculating.

Each sample was measured offline for the concentrations of biomass, glucose, lactate, glutamine and ammonia. From each $1.5mL$ sample, $20\mu L$ was used for viability and cell density counting via Trypan blue dye exclusion method (the dye penetrates dead cell membranes, discriminating them from live viable ones, and making microscopic counting possible with a Bürker hemocytometer). The remaining sample was filtrated and then assayed with Megazyme enzymatic kits K-GLNAM, K-GLUC and K-LATE and a spectrophotometer (Shimadzu UVmini-1240) at $340nm$. Glucose was also assayed by DNS (Dinitrosalicylic Colorimetric) method and absorbances read at $540nm$. The DNS data was chosen since the method proved to be better in terms of precision and much better in terms of trueness.

3.6.2 Data bank

The data bank built from experiments performed with CHO-S cells is plotted in Figure 3.13. If the intrinsic variability of the triplicate flasks in each experiment is considered, the average values can be plotted with the confidence intervals of Figure 3.14.

In terms of cell growth, it can be seen that both the first and the second experiments (B1 and B2) present similar achievements (ie, B2's initial glutamine concentration of 5mM seems enough to achieve that biomass profile). It can also be seen that experiment B3 presents an early switch to the death phase associated with a very quick glucose disappearance. A glucose limitation phenomena can, therefore, be assumed. No noticeable lactate reconsumption is observed. As for experiment B4, it was meant to be the glutamine-limited one, but, in fact, it seems that glucose is the substrate disappearing faster from the culture.

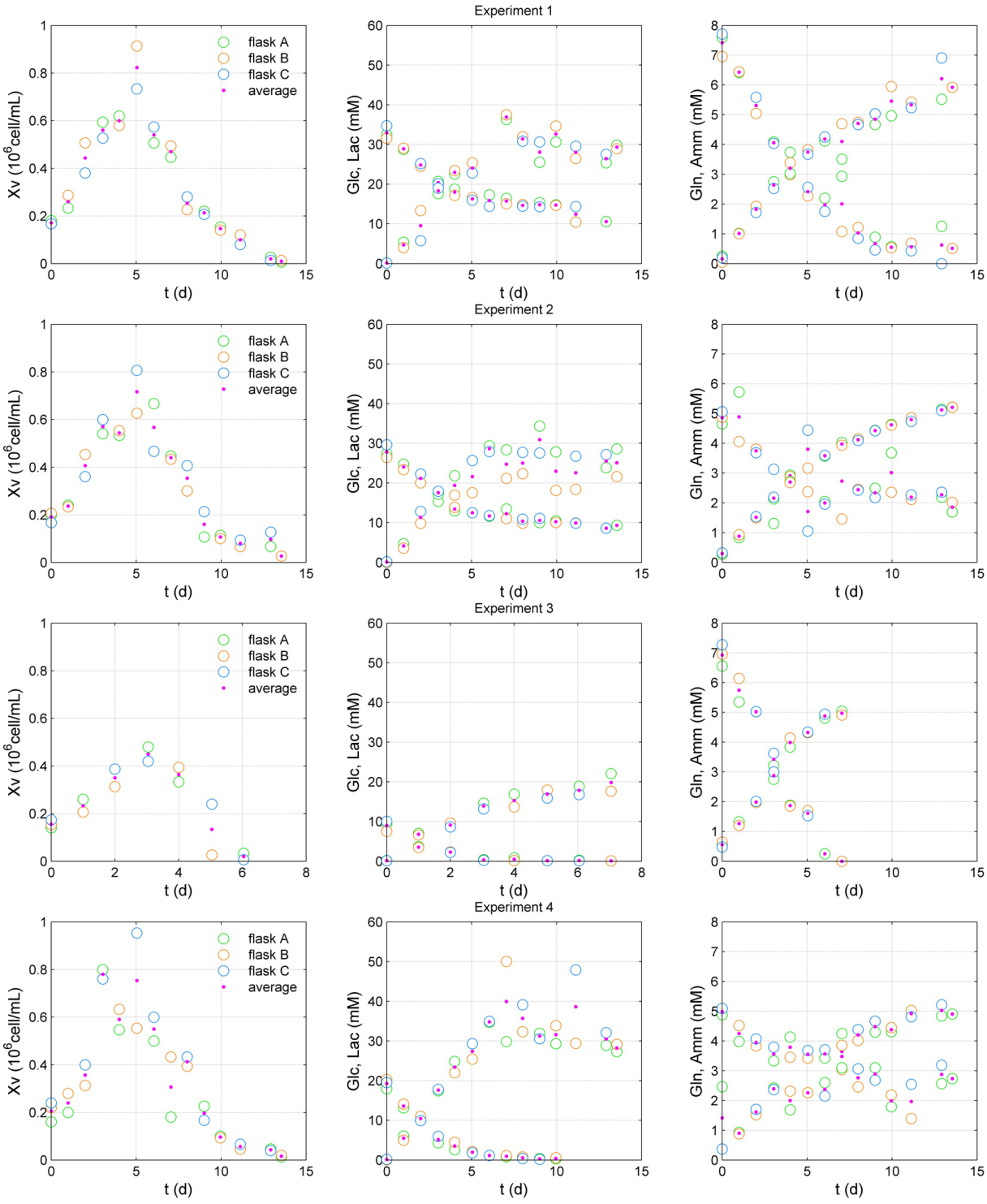


Figure 3.13: CHO-S cultures: 4 experimental conditions, each with triplicate flasks A,B,C (green, orange, blue). Averages are plotted in pink.

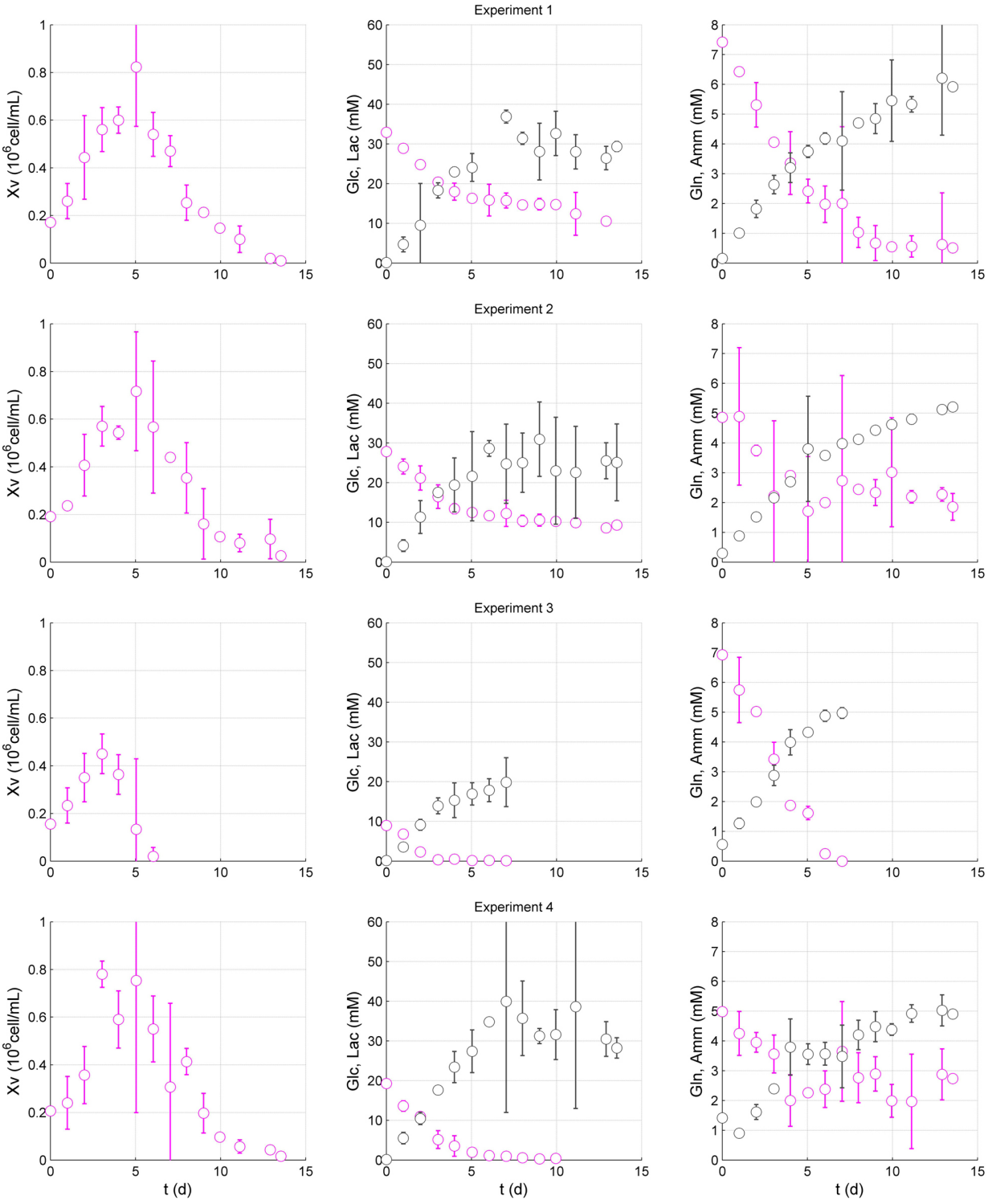


Figure 3.14: CHO-S cultures: average values for triplicate flasks with confidence intervals (2σ). Metabolites lactate and ammonia are plotted in grey.

There is obviously no prior knowing of which precise initial conditions

will lead to the phenomena we intend to check. On the whole, there is presumably potentially enough information to try and check several model structures assuming different phenomena.

3.6.3 Model identification

3.6.3.1 Confidence intervals on parameters

The values for the model parameters that one wishes to identify will have some uncertainty for many reasons: some experiments may not be informative enough, there is noise associated with the measurements, the algorithm implementation has limitations, or the model isn't obviously perfect, among others.

One way of assessing some uncertainty in parameter values is to consider confidence intervals and to compute them under certain hypotheses.

In order to compute confidence intervals, the elements of the Fisher information matrix $\boldsymbol{F}$ are computed:

$$F_{ij} = \sum_{k=1}^{n\ states} \frac{1}{\sigma_{kk}^2} \frac{\partial x_k}{\partial \theta_i} \frac{\partial x_k}{\partial \theta_j} = \sum_{k=1}^{n\ states} \frac{1}{\sigma_{kk}^2} S_{x_k,\theta_i} S_{x_k,\theta_j} \tag{3.32}$$

This matrix predicts how well the experiments are able to constrain the parameters and needs a prior knowledge of the measurement uncertainties via σ_{kk}. Under certain assumptions (identifiability), the Fisher matrix can be inverted to compute the covariance matrix that informs on the uncertainties on model parameters. The Fisher matrix thus assesses how informative the experimental set is (when multiple experiments are available, the Fisher matrix is the sum of the matrix for each experiment). We will use it to compute a lower and upper bound on parameter values.

3.6.3.2 Step #1: model Ω_1 (phase A)

Let us firstly consider data from the initial cell growth phase and reduced order model Ω_1 (1-state, 1-parameter) to identify the parameter maximum growth rate $\theta_1 = \mu_{max}$. The sensitivity terms (for batch) were computed and are listed in Appendix A, p. 227.

$$\Omega_1 : \begin{cases} \dfrac{dx_1}{dt} = \theta_1 x_1, \end{cases} \tag{3.33}$$

where $x_1 = Xv$ and $\theta_1 = \mu_{max}$.

A random initial estimate of the parameter value is generated between the following bounds (these are related to biological significance: we suppose that animal cells take at least 6h to divide):

$$\left[\theta_1 \right]_0 = random(0 - 2.8\ d^{-1}). \tag{3.34}$$

The algorithm then goes on to compute an estimate of θ_1 according to the procedure described on page 81. Although the generation of a θ_1^0 is bounded, the algorithm is free to investigate values of θ_1 outside this interval. Even if it is not necessary with this simple model, a best practice is followed: in order to avoid being trapped in a local minimum of the cost function, a multistart procedure is used (10 random initial values are tested). The following value was proposed as the best:

$$\left[\theta_1 \right]^* = 0.49 d^{-1}. \tag{3.35}$$

Figures 3.15 up to 3.18 show the results: a simulation of the identified model (in blue) with upper and lower bounds relating to a 95%-confidence interval (in green) of the simulated concentrations. Measurements (circles) are also plotted with error bars relating to the intrinsic biological variability (this information is available for time points where the number of measurements was enough).

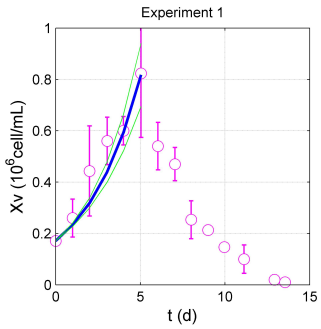


Figure 3.15: Model Ω_1, experiment 1, CHO-S: model simulation (blue), 95%-confidence interval for states (green), measurements (circles) and their variability (error bars, 2σ) .

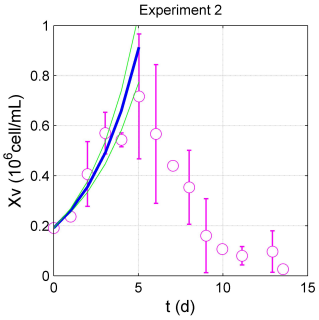


Figure 3.16: Model Ω_1, experiment 2, CHO-S: model simulation (blue), 95%-confidence interval for states (green), measurements (circles) and their variability (error bars, 2σ) .

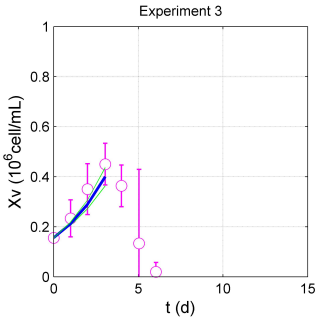


Figure 3.17: Model Ω_1, experiment 3, CHO-S: model simulation (blue), 95%-confidence interval for states (green), measurements (circles) and their variability (error bars, 2σ) .

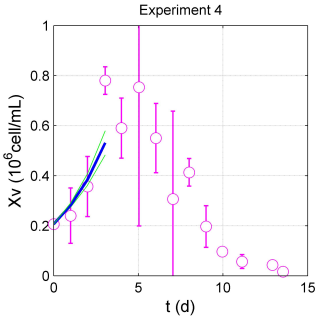


Figure 3.18: Model Ω_1, experiment 4, CHO-S: model simulation (blue), 95%-confidence interval for states (green), measurements (circles) and their variability (error bars, 2σ) .

For the fourth experiment it can be noticed that the real t_{peak} (highest value of Xv attained) is maybe a bit higher. Overall the results are good, even

if this is not for the moment very important. In fact, the goal of this first identification step is to find out a good estimate of the value of θ_1 that could later on be used with a more complex model. It is, thus, not very relevant if the fit is excellent or not since there is an obvious model-plant mismatch (ie. this reduced model is too simple to capture the behaviour of the biological system being considered).

All the initial estimates tested in the multistart led to basically the same final value for μ_{max}. In fact, there can only be one value describing this exponential early phase of biomass concentration and the algorithm converges quite quickly given that the model is mathematically very simple so far.

3.6.3.3 Step #2: model Ω_2 (phase A)

Let us now further expand the model by considering also the substrates. The initial growth phase can be described by reduced order model Ω_2 (3-states, 3-parameters) to identify the maximum growth rate (again) and parameters θ_8 and θ_{12} related to the stoichiometry of the consumption of the substrates:

$$\Omega_2 : \begin{cases} \dfrac{dx_1}{dt} = \theta_1 x_1, \\ \dfrac{dx_2}{dt} = \dfrac{-1}{\theta_8}\theta_1 x_1, \\ \dfrac{dx_4}{dt} = \dfrac{-1}{\theta_{12}}\theta_1 x_1, \end{cases} \tag{3.36}$$

where $x_1 = Xv$; $x_2 = Glc$; $x_4 = Gln$; $\theta_1 = \mu_{max}$; $\theta_8 = Y_{Xv/Glc}$; $\theta_{12} = Y_{Xv/Gln}$;

The sensitivity terms (for batch) were computed and are listed in Appendix A on p. 227.

As for the initial estimate, one value is already available from the previous step. For the other parameters, a random initial guess θ_0 is generated between the following plausible bounds (where a large safety margin was put on typical literature values[1] for θ_8 and θ_{12}):

$$\begin{bmatrix} \theta_1 \\ \theta_8 \\ \theta_{12} \end{bmatrix}_0 = \begin{bmatrix} 0.49 & d^{-1} \\ random(0-10) & 10^9 cell\, mmol^{-1} \\ random(0-10) & 10^9 cell\, mmol^{-1} \end{bmatrix}. \tag{3.37}$$

Again, despite the fact that the generation of the initial guess θ_0 is constrained, the identification algorithm is still free to search elsewhere. After a

[1] See, for example, (dos Reis Castilho, 2008), (Xing et al, 2010), (de Tremblay et al, 1992), (Dunn et al, 2003).

multistart with 10 initial guesses, the optimal vector θ^* found was:

$$\begin{bmatrix} \theta_1 \\ \theta_8 \\ \theta_{12} \end{bmatrix}^* = \begin{bmatrix} 0.30 & d^{-1} \\ 0.03 & 10^9 cell\, mmol^{-1} \\ 0.12 & 10^9 cell\, mmol^{-1} \end{bmatrix}. \tag{3.38}$$

This corresponds to a model that is able to describe the data as plotted in Figure 3.20 to 3.22.

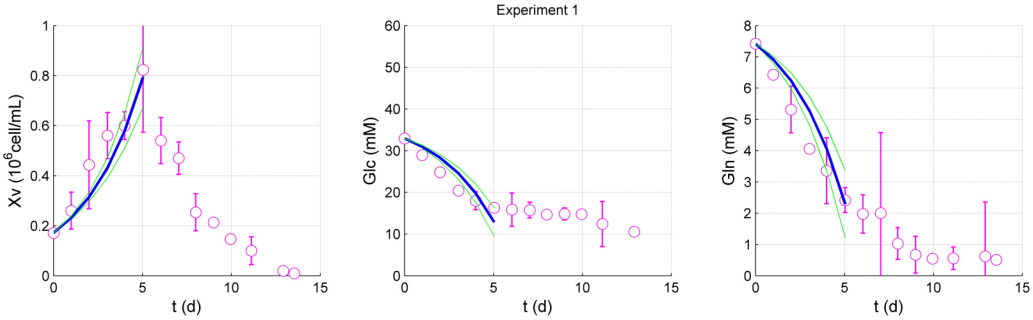


Figure 3.19: Model Ω_2, experiment 1, CHO-S: model simulation (blue), 95%-confidence interval for states (green), measurements (circles) and their variability (error bars, 2σ) .

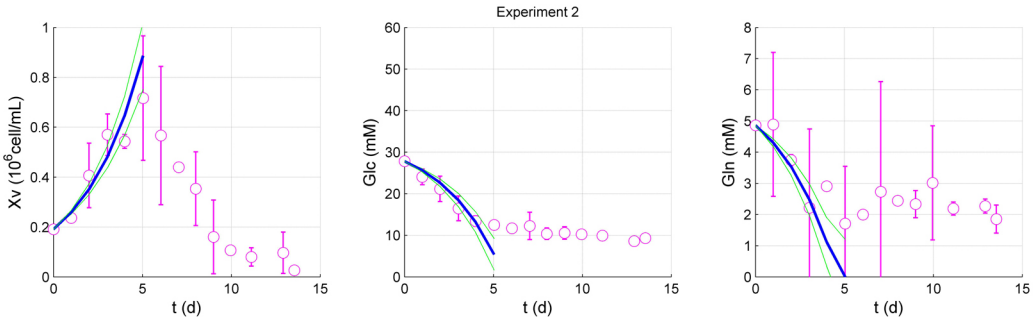


Figure 3.20: Model Ω_2, experiment 2, CHO-S: model simulation (blue), 95%-confidence interval for states (green), measurements (circles) and their variability (error bars, 2σ) .

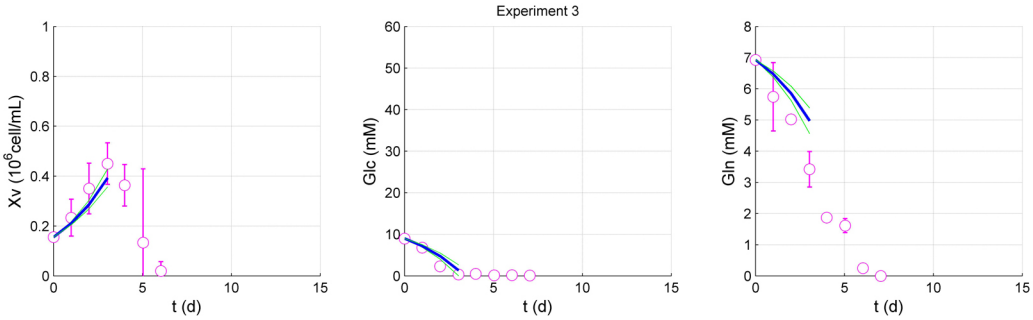


Figure 3.21: Model Ω_2, experiment 3, CHO-S: model simulation (blue), 95%-confidence interval for states (green), measurements (circles) and their variability (error bars, 2σ) .

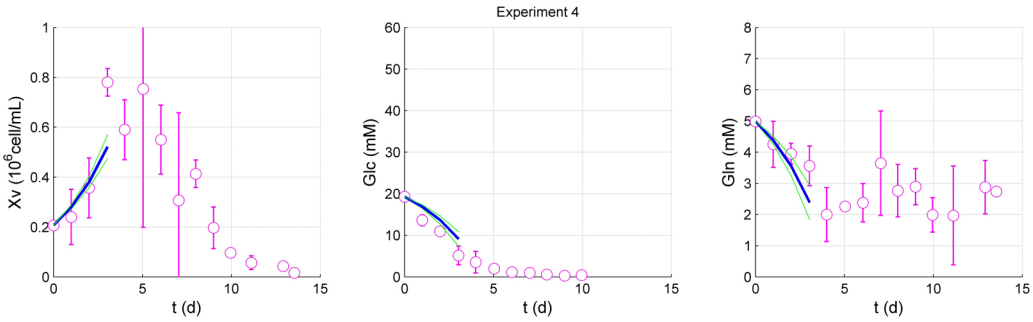


Figure 3.22: Model Ω_2, experiment 4, CHO-S: model simulation (blue), 95%-confidence interval for states (green), measurements (circles) and their variability (error bars, 2σ) .

It can be seen that the estimates found for θ_1, θ_8 and θ_{12} are overall acceptable.

3.6.3.4 Step #3: model Ω_3 (phase A)

Let us now include also the production of metabolites associated with the two major substrates. The initial growth phase can be described by reduced order model Ω_3 (5-states, 5-parameters) to identify not only θ_1, θ_8 and θ_{12} (again) but also new parameters θ_9 and θ_{13} related to the stoichiometry of the production of metabolites:

$$\Omega_3 : \begin{cases} \dfrac{dx_1}{dt} = \theta_1 x_1, & \\ \dfrac{dx_2}{dt} = \dfrac{-1}{\theta_8}\theta_1 x_1, & \dfrac{dx_3}{dt} = \dfrac{\theta_9}{\theta_8}\theta_1 x_1, \\ \dfrac{dx_4}{dt} = \dfrac{-1}{\theta_{12}}\theta_1 x_1, & \dfrac{dx_5}{dt} = \dfrac{\theta_{13}}{\theta_{12}}\theta_1 x_1, \end{cases} \tag{3.39}$$

where $x_1 = Xv$; $x_2 = Glc$; $x_3 = Gln$; $x_4 = Gln$; $x_5 = Gln$; $\theta_1 = \mu_{max}$; $\theta_8 = Y_{Xv/Glc}$; $\theta_9 = Y_{Xv/Glc}$; $\theta_{12} = Y_{Xv/Gln}$; $\theta_{13} = Y_{Xv/Gln}$;

The sensitivity terms (for batch) were computed and are listed in Appendix A on p. 228.

Three parameter estimates are available from the previous step. We have made use of this information by including them in the initial estimate vector θ_0 whilst the remaining values were randomly generated within bounds (again, a large safety margin was put on typical literature values for θ_9 and θ_{13}):

$$\begin{bmatrix} \theta_1 \\ \theta_8 \\ \theta_9 \\ \theta_{12} \\ \theta_{13} \end{bmatrix}_0 = \begin{bmatrix} 0.30 & d^{-1} \\ 0.03 & 10^9 cell\, mmol^{-1} \\ random(0-10\,000) & mmol\, mmol^{-1} \\ 0.12 & 10^9 cell\, mmol^{-1} \\ random(0-10\,000) & mmol\, mmol^{-1} \end{bmatrix}. \tag{3.40}$$

The model to be identified has now more states and is more complex. Furthermore, time vectors are various, not all timepoints were actually measured and sometimes variance information is lacking for a given state, timepoint and experiment. It is thus more important to look in further detail at the form of the cost function J_{norm}. Up to now, an intuitive form of the cost function in terms of normalized variables, J_{norm}, has been used:

$$J_{norm} = \sum_{j=1}^{n_E} \sum_{i=1}^{n_S} \sum_{k=1}^{n_{t_{ij}}} \frac{k_{t,i,j}\left(\bar{x}^j_{i,k,sim} - \bar{x}^j_{i,k,meas}\right)^2}{n^j_{i,k}} \tag{3.41}$$

where n_E is the number of experiments, n_S is the number of states, $n_{t_{i,j}}$ is the number of measured timepoints for state i in experiment j, $k_{t,i,j}$ is either 1 if state i in experiment j in timepoint $t_{k_{i,j}}$ was measured or 0 if not, $\bar{x}^j_{i,k,sim}$ is the simulated normalised value of state i in experiment j and timepoint $t_{k_{i,j}}$, $\bar{x}^j_{i,k,measured}$ is the measured normalised value of state i in experiment j

and timepoint $t_{k_{i,j}}$ and $n_{i,k}^{j}$ is the number of times a quadratic deviation was computed (ie, the number of timepoints where measurements were available, all experiments of the data bank and all state variables comprised).

Firstly, for each experiment, states x were normalised with respect to the maximum measured value attained by the state. The reason for doing so is the fact that scales and units are quite different[1]. By doing this, all normalised states $\bar{x}$ will now vary between 0 and 1 (if measured) or around it (if simulated).

Secondly, for some timepoints, some concentrations were not measured and it is, thus, not possible to evaluate the quadratic deviation between measured and simulated values. Furthermore, some experiments last more than others. Since we intend to have a cost that treats all experiments with the same importance irrespective of the amount of timepoints available, we have, therefore, chosen to compute the average quadratic distance per timepoint.

Cost J_{norm} illustrates thus how distant model simulations are to measured values, on an average timepoint for all normalised states and all experiments.

However, more information can be used in the construction of the cost function J. Namely, if a measurement is more certain than another, then ideally it should weigh less on the cost that we intend to minimise, while more uncertain values should weigh more. This weighting can be done by introducing the variance matrix, the general formula being:

$$J = Q \times V^{-1} \times Q \tag{3.42}$$

where Q is the matrix of quadratic deviations and V a diagonal matrix with the variances of the states, σ^2.

A consequence of dividing by the variance is that (while not exactly normalising with respect to the highest measured value) we attenuate the different scales of values assumed by different physical states of the system since each state deviation is divided by the standard deviation, yielding also, obviously, in this case, a dimensionless value:

$$dim\left(\frac{(x_{sim} - x_{meas})^2}{\sigma^2}\right) = dim\left(\frac{(x_{sim} - x_{meas})}{\sigma}\right) \times dim\left(\frac{(x_{sim} - x_{meas})}{\sigma}\right) = 1 \tag{3.43}$$

Typically, when variance is accounted for in the cost function, a hypothesis is made that measuring probes will have a certain ϵ% error associated and this

[1]For example, while typical open loop values for biomass can vary between 0 and $10^6 cell/mL$, glutamine concentrations can fluctuate between 0 and $8mM$, a much smaller span.

will be the error behind the dispersion observed throughout one experiment. This is usually an optimistic estimate of the real dispersion of data since it is frequently based on equipment supplier technical sheets (if at all available[2]) or a common sense free guess of its absolute/relative numerical magnitude. In this case, one could, for example, for one curve of biomass over time, add confidence intervals corresponding to a constant 10% error.

In our case, since each experiment was done in triplicate[3], sample variance σ^2 can be computed in timepoints t_k where a concentration was measured more than once (n times, as shown in Figure 3.14) with its definition:

$$\sigma^2_{t_k} = \sum_i \frac{(x_i - x_{average})^2}{n-1} \tag{3.44}$$

We have thus a better measure of the real dispersion of data. Not only do we have more information on dispersion over time, over different experiments and over different states, we also have real information that embodies the natural biological variability which really has the bigger weight in animal cell cultures, ie. the different behaviour of cells in the culture has more impact in data dispersion than probe related errors.

One can thus consider another form for the cost function J:

$$J_{var} = \sum_{j=1}^{n_E} \sum_{i=1}^{n_S} \sum_{k=1}^{n_{t_{ij}}} \frac{k_{t,i,j} \left(x^j_{i,k,sim} - x^j_{i,k,meas} \right)^2 / \sigma^2_{ij}}{n^j_{i,k}} \tag{3.45}$$

where $x^j_{i,k,sim}$ is the simulated value of state i in experiment j and timepoint $t_{k_{i,j}}$, $x^j_{i,k,measured}$ is the measured value of state i in experiment j and timepoint $t_{k_{i,j}}$, $\sigma^2_{i,j}$ is the time-average[4] variance for state i in experiment j, and the other variables are defined as previously in equation (3.41).

[2]For instance, cell counting under the microscope (for the determination of biomass concentration) has not only errors associated with pipettes and general equipment but also a user error that can be bigger or smaller, depending on the user counting, his/her way of counting, attention, etc. It is common practice to suppose a value between 5% and 20%.

[3]Ideally, a good practice is to consider a sample size around 10 for a good estimate of dispersion or 30 for a very good one, if it is to be used in hypothesis tests leading to critical consequences (eg. patients undergoing a treatment). However, with respect to animal cell cultures, performing the same measurement/experiment 10 times is really too much given financial, time and resource constraints.

[4]For example, $\sigma^2_{2,4}$ refers to state 2 (*Glc*) in experiment 4. It is the average of variances in the timepoints of experiment 4 where variance could be computed since state x_2 was measured more than once.

Both cost functions were trialed for 10 multistarts departing from initial estimate as defined in equation (3.40). With the variance-based cost function J_{var} the simulation resulted in:

$$\begin{bmatrix} \theta_1 \\ \theta_8 \\ \theta_9 \\ \theta_{12} \\ \theta_{13} \end{bmatrix}^* = \begin{bmatrix} 0.24 & d^{-1} \\ 0.73 & 10^9 cell\, mmol^{-1} \\ 1.39 & mmol\, mmol^{-1} \\ 0.07 & 10^9 cell\, mmol^{-1} \\ 0.02 & mmol\, mmol^{-1} \end{bmatrix}. \tag{3.46}$$

Using J_{norm} the results were:

$$\begin{bmatrix} \theta_1 \\ \theta_8 \\ \theta_9 \\ \theta_{12} \\ \theta_{13} \end{bmatrix}^* = \begin{bmatrix} 0.29 & d^{-1} \\ 0.84 & 10^9 cell\, mmol^{-1} \\ 1.34 & mmol\, mmol^{-1} \\ 0.04 & 10^9 cell\, mmol^{-1} \\ 0.007 & mmol\, mmol^{-1} \end{bmatrix}. \tag{3.47}$$

Both results are acceptable upon a visual inspection (Figures 3.23 to 3.26, where results obtained with J_{norm} are plotted in blue and those obtained using J_{var} are in green).

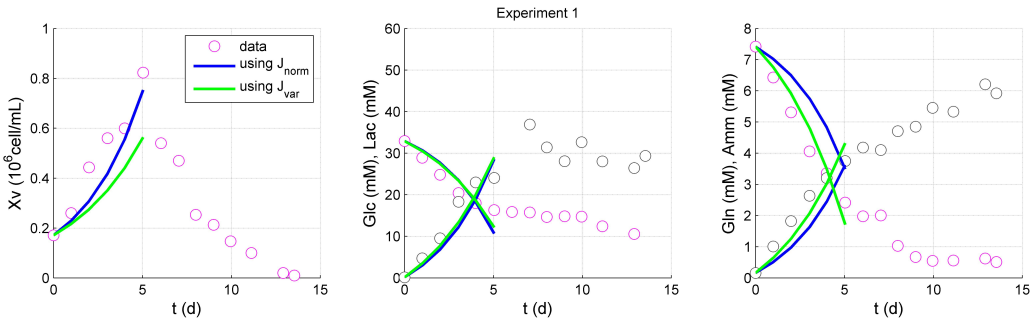


Figure 3.23: Model Ω_3, experiment 1, CHO-S: model simulation using J_{norm} (blue), using J_{var} (green), measurements (circles).

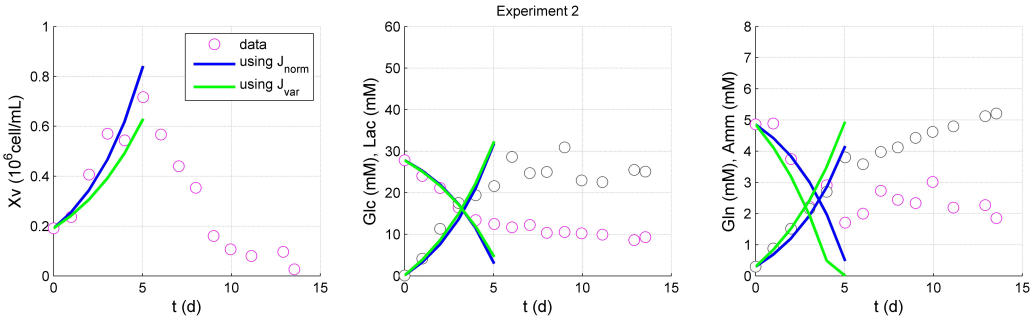


Figure 3.24: Model Ω_3, experiment 2, CHO-S: model simulation using J_{norm} (blue), using J_{var} (green), measurements (circles).

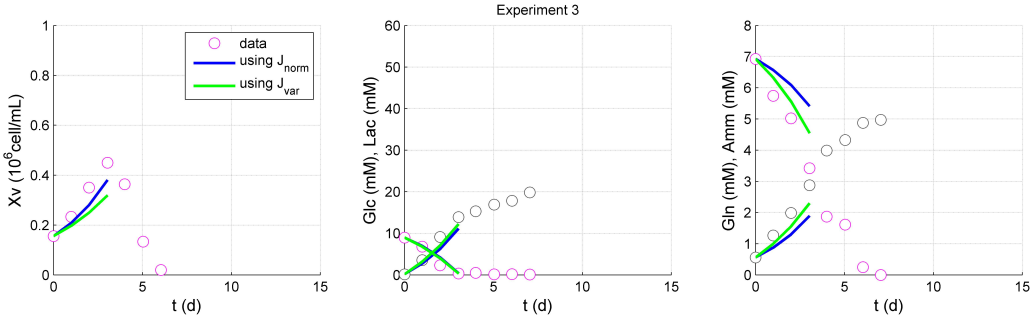


Figure 3.25: Model Ω_3, experiment 3, CHO-S: model simulation using J_{norm} (blue), using J_{var} (green), measurements (circles).

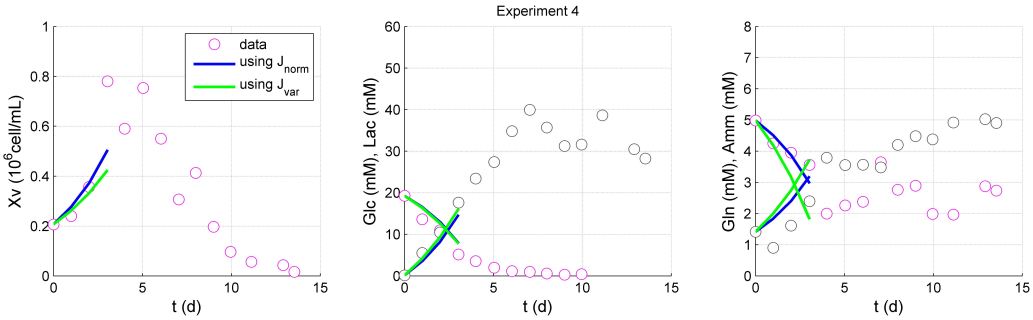


Figure 3.26: Model Ω_3, experiment 4, CHO-S: model simulation using J_{norm} (blue), using J_{var} (green), measurements (circles).

It seems that, compared to J_{norm}, if J_{var} is used, more emphasis is put on glutamine model-data compliance and less on biomass compliance. The final decision of which is more important remains with the final user, of course, but both results are acceptable. We will proceed using J_{norm} in the following steps since it seems to be slightly faster.

3.6.3.5 Step #4: model Ω_4 (phases A and B)

The following step is a critical one since all of the culture data will now be included, namely growth phase A (period when cell concentration rises) and death phase B (when cell concentration starts to decrease). It is critical to correctly chose a mathematical form that allows the net growth rate μ_{net} to become negative whenever biomass concentration Xv decreases.

$$\frac{dXv}{dt} = \mu_{net} Xv \tag{3.48}$$

The simplest approach is to firstly try out very simple sub-models for the whole culture (growth phase and death phase) based on some hypothesis on phenomena that can put an end to the rise of biomass concentration.

Some basic hypotheses are:

- $\mathcal{H}_{4a}$: growth is limited by glucose disappearance

$$\Omega_{4a} : \begin{cases} \dfrac{dx_1}{dt} = \theta_1 \dfrac{x_2}{\theta_2 + x_2} x_1 \equiv \dfrac{dXv}{dt} = \mu_{max} \dfrac{Glc}{k_{Glc} + Glc} Xv \end{cases} \tag{3.49}$$

- $\mathcal{H}_{4b}$: growth is limited by glutamine disappearance

$$\Omega_{4b} : \begin{cases} \dfrac{dx_1}{dt} = \theta_1 \dfrac{x_4}{\theta_3 + x_4} x_1 \equiv \dfrac{dXv}{dt} = \mu_{max} \dfrac{Gln}{k_{Gln} + Gln} Xv \end{cases} \tag{3.50}$$

- $\mathcal{H}_{4c}$: growth is limited both by glucose and glutamine disappearance

$$\Omega_{4c} : \begin{cases} \dfrac{dx_1}{dt} = \theta_1 \dfrac{x_2}{\theta_2 + x_2} \dfrac{x_4}{\theta_3 + x_4} x_1 \equiv \dfrac{dXv}{dt} = \mu_{max} \dfrac{Glc}{k_{Glc} + Glc} \dfrac{Gln}{k_{Gln} + Gln} Xv \end{cases} \tag{3.51}$$

Sub-models Ω_{4a} to Ω_{4c} provide typical forms of expressing substrate limitation. They are, however, not eligible since they only allow μ_{net} to be either positive (in that case Xv will increase) or null (Xv will remain constant).

In order to allow μ_{net} to become negative, a baseline death rate θ_4 will be introduced:[1]

[1] Notice also that a sub-model where simply $\mu_{net} = \mu_{max} - \mu_{d,max} = constant$ is not eligible either since it would not allow both a growth and a death phase. It would allow only a growth phase ($\mu_{net} > 0$), only null growth ($\mu_{net} = 0$) or only a death phase ($\mu_{net} < 0$).

- $\mathcal{H}_{4d}$: growth is limited by glucose disappearance and there is a baseline death rate

$$\Omega_{4d} : \left\{ \frac{dx_1}{dt} = \left(\theta_1 \frac{x_2}{\theta_2 + x_2} - \theta_4 \right) x_1 \right. \tag{3.52}$$

- $\mathcal{H}_{4e}$: growth is limited by glutamine disappearance and there is a baseline death rate

$$\Omega_{4e} : \left\{ \frac{dx_1}{dt} = \left(\theta_1 \frac{x_4}{\theta_3 + x_4} - \theta_4 \right) x_1 \right. \tag{3.53}$$

- $\mathcal{H}_{4f}$: growth is limited both by glucose and glutamine disappearance and there is a baseline death rate

$$\Omega_{4f} : \left\{ \frac{dx_1}{dt} = \left(\theta_1 \frac{x_2}{\theta_2 + x_2} \frac{x_4}{\theta_3 + x_4} - \theta_4 \right) x_1 \right. \tag{3.54}$$

All of these will now be tested. Note that these sub-models cannot be solved autonomously. The system is multivariable and the evolution of states x is done by simultaneous integration of all the various differential equations. More precisely, how other components are modeled will also have an impact on $Xv(t)$ since dXv/dt depends on other states (the substrates). In order to test all submodels Ω_{4d} to Ω_{4f}, these will be coupled to the other following differential equations where μ_{growth} is defined in each submodel:

$$\Omega_{4,rest} : \left\{ \begin{array}{ll} \dfrac{dx_2}{dt} = -\dfrac{1}{\theta_8} \mu_{growth} x_1 & , \\ \dfrac{dx_4}{dt} = -\dfrac{1}{\theta_{12}} \mu_{growth} x_1 & , \end{array} \right. \tag{3.55}$$

The following initial guesses were considered:

$$[\theta_{4d}]_0 : \left[\begin{array}{c} \theta_1 \\ \theta_2 \\ \theta_4 \\ \theta_8 \\ \theta_{12} \end{array} \right]_0 = \left[\begin{array}{cc} 0.24 & d^{-1} \\ random(0-11) & mM \\ random(0-24) & d^{-1} \\ 0.73 & 10^9 cell\, mmol^{-1} \\ 0.07 & 10^9 cell\, mmol^{-1} \end{array} \right]. \tag{3.56}$$

$$[\theta_{4e}]_0 : \begin{bmatrix} \theta_1 \\ \theta_3 \\ \theta_4 \\ \theta_8 \\ \theta_{12} \end{bmatrix}_0 = \begin{bmatrix} 0.24 & d^{-1} \\ random(0-11) & mM \\ random(0-2.8) & d^{-1} \\ 0.73 & 10^9 cell\, mmol^{-1} \\ 0.07 & 10^9 cell\, mmol^{-1} \end{bmatrix}. \tag{3.57}$$

$$[\theta_{4f}]_0 : \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \theta_8 \\ \theta_{12} \end{bmatrix}_0 = \begin{bmatrix} 0.24 & d^{-1} \\ random(0-11) & mM \\ random(0-11) & mM \\ random(0-2.8) & d^{-1} \\ 0.73 & 10^9 cell\, mmol^{-1} \\ 0.07 & 10^9 cell\, mmol^{-1} \end{bmatrix}. \tag{3.58}$$

The upper bounds for the random generation of a guess value of new parameters $\theta_2 = k_{Glc}$, $\theta_3 = k_{Gln}$, $\theta_4 = \mu_{d,max}$ were set by choosing values much bigger than those provided in the literature. As for $\theta_4 = \mu_{d,max}$, it needs to be smaller than $\theta_1 = \mu_{max}$ so that there will be an initial growth phase and a later death phase (otherwise μ_{net} would always be negative and biomass would always decrease throughout the culture). The value found in the previous step was, thus, used as an upper bound for this random generation.

The algorithm was run using J_{var} and J_{norm} as cost functions (with similar results). The identified values were added to Table 3.7 that summarises all results gathered so far.

Note that at this stage some extra information from system dynamics could be useful in this identification step: we expect from the sensitivity analysis that parameters such as k_{Glc} will be more influential throughout a glucose-poor experiment (in this CHO-S case study, experiments 3 and 4), and parameters such as k_{Gln} will be more influential in glutamine-poor experiments (experiments 2 and 4).[1] Since the cost function is a sum of the costs for the different experiments, some weighting factors w_i could be used to put special emphasis on more informative experiments:

$$J = w_1 J_{Exp1} + w_2 J_{Exp2} + w_3 J_{Exp3} + w_4 J_{Exp4} \tag{3.59}$$

One can thus expect the algorithm's convergence to be faster if w_3 and w_4 are bigger while identifying option $\{\Omega_{4d} + \Omega_{4rest}\}$.[2], and, likewise, if greater

[1] This is illustrated in Figure 3.2 on page 89, for example.

w_2 and w_4 are used while identifying $\{\Omega_{4e} + \Omega_{4rest}\}$. However, it is important to consider the informative richness of all experiments. In fact, as an extreme example, if only one experiment is accounted for by the algorithm as the "training dataset", then most likely the identified model will offer a reasonably good fit for that particular experiment and yet be inadequate for the remaining experiments (which are the independent "cross-validation datasets"). This is illustrated in Figures 3.27 to 3.30: model $\{\Omega_{4d} + \Omega_{4rest}\}$ was firstly identified considering only data from experiment 1: the identified model unsurprisingly describes experiment 1 (red solid lines in Figure 3.30) well, but does not predict the outcome of the culture for the conditions of the other experiments (red dashed lines in Figures 3.28-3.30) well. The same happens when the algorithm considers only data from experiment 2 (blue), 3 (green) and 4 (black): they can only describe well the experiment from which the parameters were identified.

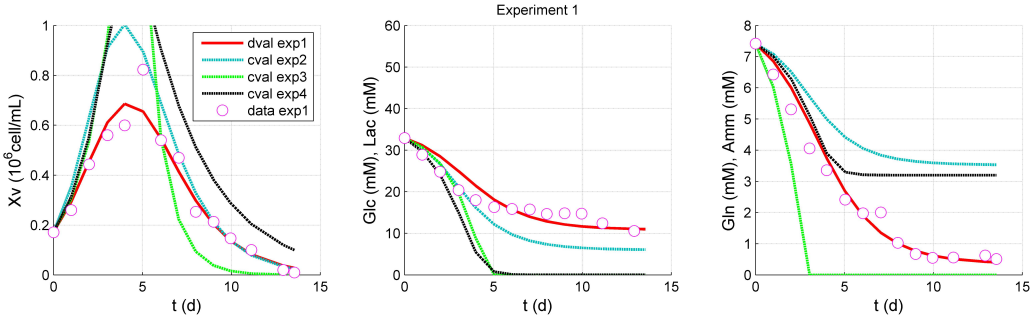


Figure 3.27: Model $\{\Omega_{4d} + \Omega_{4rest}\}$, experiment 1, CHO-S: model simulation using parameters identified (direct validation) with experiment 1 (red) or (cross-validation) using only experiment 2 (blue) or 3 (green) or 4 (black). Circles represent experiment 1's dataset.

[2]Since this model considers the existence of a k_{Glc}, ie. a limitation of growth by progressive glucose depletion.

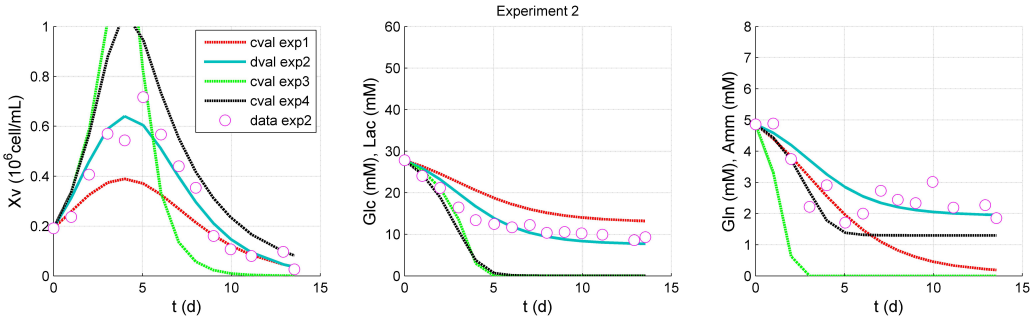


Figure 3.28: Model $\{\Omega_{4d} + \Omega_{4rest}\}$, experiment 2, CHO-S: model simulation using parameters identified (direct validation) with experiment 2 (blue) or (cross-validation) using only experiment 1 (red) or 3 (green) or 4 (black). Circles represent experiment 2's dataset.

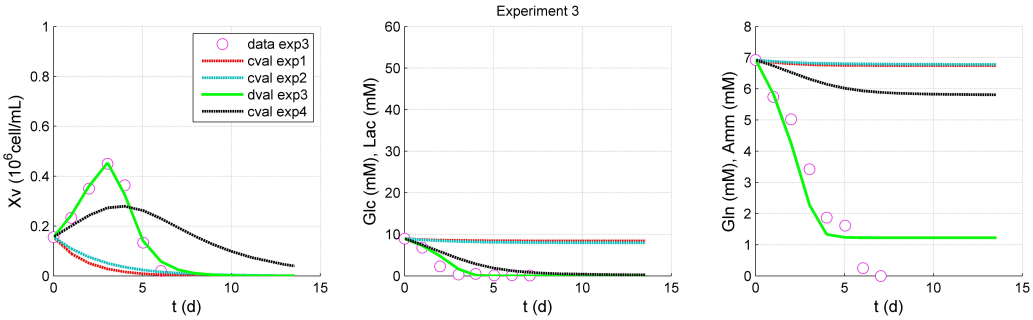


Figure 3.29: Model $\{\Omega_{4d} + \Omega_{4rest}\}$, experiment 3, CHO-S: model simulation using parameters identified (direct validation) with experiment 3 (green) or (cross-validation) using only experiment 1 (red), 2 (blue) or 4 (black). Circles represent experiment 3's dataset.

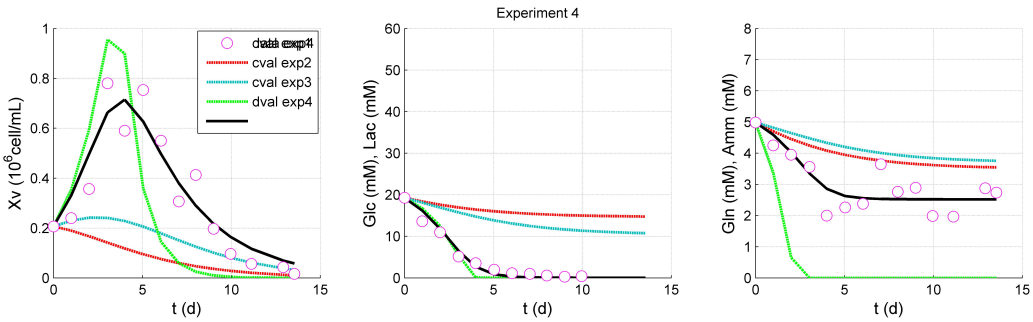


Figure 3.30: Model $\{\Omega_{4d} + \Omega_{4rest}\}$, experiment 4, CHO-S: model simulation using parameters identified (direct validation) with experiment 4 (black) or (cross-validation) using only experiment 1 (red), 2 (blue) or 3 (green). Circles represent experiment 4's dataset.

This model is simple enough to be able to adjust itself and describe (only) one experimental set of conditions but not powerful enough to capture culture behaviour and to predict how it would react to other different conditions.

It is, thus, important at this stage to consider all datasets since our purpose is to have a rough idea of good starting values for some parameters that will be used later in more complex models. All experimental datasets considered, the best fits are presented in Figures 3.31 to 3.34 for the three simple submodels (Ω_{4d}, Ω_{4e}, Ω_{4f}) proposed to estimate the substrate limitation phenomena:

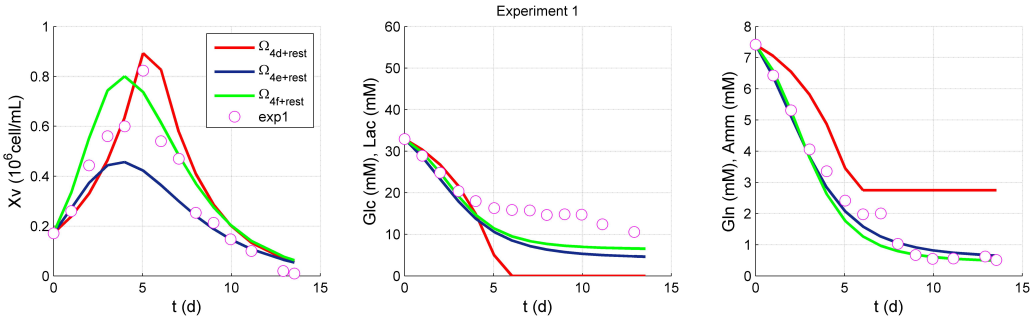


Figure 3.31: Models $\{\Omega_{4d} + \Omega_{4rest}\}$, $\{\Omega_{4e} + \Omega_{4rest}\}$, $\{\Omega_{4f} + \Omega_{4rest}\}$, experiment 1, CHO-S: simulations with models identified using the complete databank. Circles represent experiment 1's dataset.

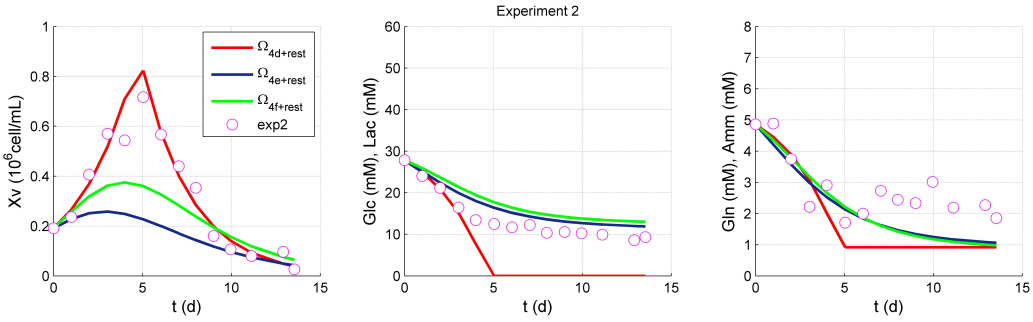


Figure 3.32: Models $\{\Omega_{4d} + \Omega_{4rest}\}$, $\{\Omega_{4e} + \Omega_{4rest}\}$, $\{\Omega_{4f} + \Omega_{4rest}\}$, experiment 2, CHO-S: simulations with models identified using the complete databank. Circles represent experiment 2's dataset.

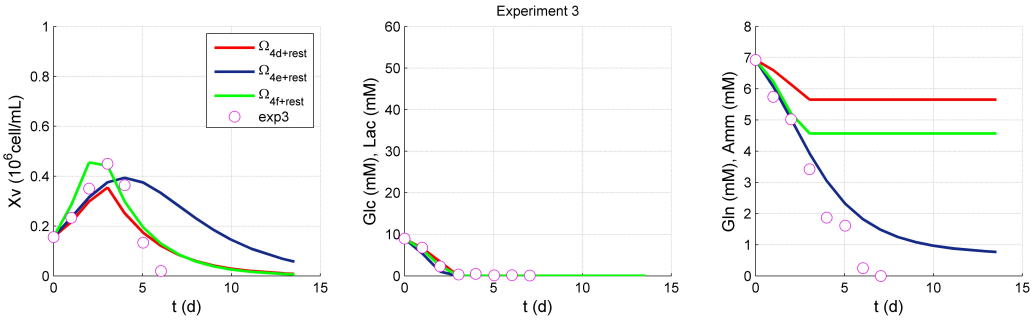


Figure 3.33: Models $\{\Omega_{4d} + \Omega_{4rest}\}$, $\{\Omega_{4e} + \Omega_{4rest}\}$, $\{\Omega_{4f} + \Omega_{4rest}\}$, experiment 3, CHO-S: simulations with models identified using the complete databank. Circles represent experiment 3's dataset.

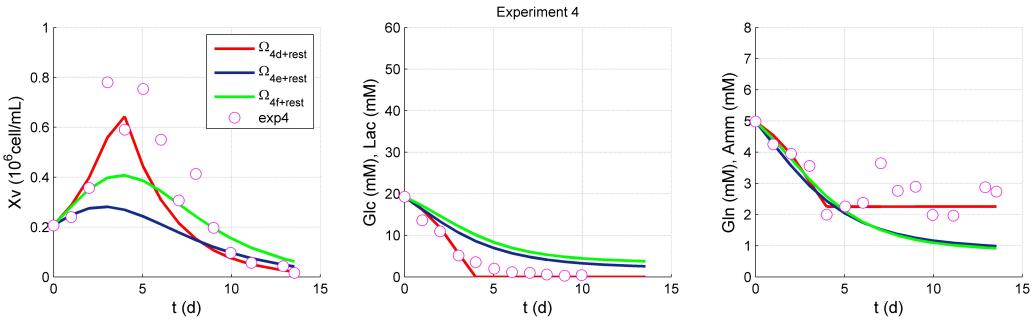


Figure 3.34: Models $\{\Omega_{4d} + \Omega_{4rest}\}$, $\{\Omega_{4e} + \Omega_{4rest}\}$, $\{\Omega_{4f} + \Omega_{4rest}\}$, experiment 4, CHO-S: simulations with models identified using the complete databank. Circles represent experiment 4's dataset.

In Table 3.7, it can be seen that although the best cost value was obtained with model $\Omega_{4f+rest}$ (both glucose and glutamine limitation are assumed), different runs of the algorithm converge, in this case, to different values for some of the parameters. In fact, it seems that the available data may not be informative enough to estimate glutamine limitation and, therefore, different combinations of values for $\{k_{Gln}; \mu_{max}; k_{Glc}\}$ lead to similar curves.

On the other hand, for glucose, another substrate, limitation is a phenomenon well present in the data. Therefore, k_{Glc} can be easily estimated using model $\Omega_{4d+rest}$: multiple runs of the algorithm all lead to a similar final value.

Table 3.7: Identification results for steps #1 to #4f.

Parameter / Step	Ω_1	Ω_2	Ω_3	Ω_{4d}	Ω_{4e}	Ω_{4f}
$\theta_1 \equiv \mu_{max}$	0.49	0.30 (0.49)	0.24 (0.30)	0.69 (0.24)	(*) (0.24)	(*) (0.24)
$\theta_2 \equiv k_{Glc}$				0.14		(*)
$\theta_3 \equiv k_{Gln}$					(*)	(*)
$\theta_4 \equiv \mu_{d,max}$				0.053	0.36	0.41
$\theta_8 \equiv Y_{Xv/Glc}$		0.03	0.02 (0.03)	0.053 (0.02)	0.042 (0.73)	0.082 (0.73)
$\theta_9 \equiv Y_{Lac/Glc}$			1.39			
$\theta_{12} \equiv Y_{Xv/Gln}$		0.12	0.08 (0.12)	0.38 (0.07)	0.17 (0.07)	0.31 (0.07)
$\theta_{13} \equiv Y_{Amm/Gln}$			0.74			
Cost J				0.677	0.535	0.472

Note: for each column (step/model), the initial parameter guess, whenever not randomly generated, is shown in parentheses. Cells shaded in grey indicate parameters considered in that model. Remarks: (*) means results are inconclusive because, despite convergence obtained in each run, the identified values are very different, they have different orders of magnitude, eg. they range from 10 to 10^7.

In any case, the main purpose of this step is to check whether an idea of the amplitude of the substrate limitation constants can be found. In this reduced simplistic model, these are the parameters responsible for the moment when cell growth switches to a cell death phase. But it is clear from Figures 3.31 to 3.34 that the phenomenon alone is not able to fully describe the death phase. We will consider some other hypotheses.

3.6.3.6 Step #5: model Ω_5 (phases A and B)

Let us presume that another phenomenon can occur also in the culture: metabolite inhibition, namely, the fact that ammonia and lactate may inhibit cell growth and contribute to the negativeness of term μ_{net} during the death phase.

The terms proposed below in models $\Omega_{5a} - \Omega_{5d}$ are inspired by (de Tremblay, 1991)'s model. Taking into consideration the parameter values of this model (for hybridoma cells) and the range of concentrations of our data bank

(similar but with CHO cells), these terms assume the curves described in Figure 3.35.

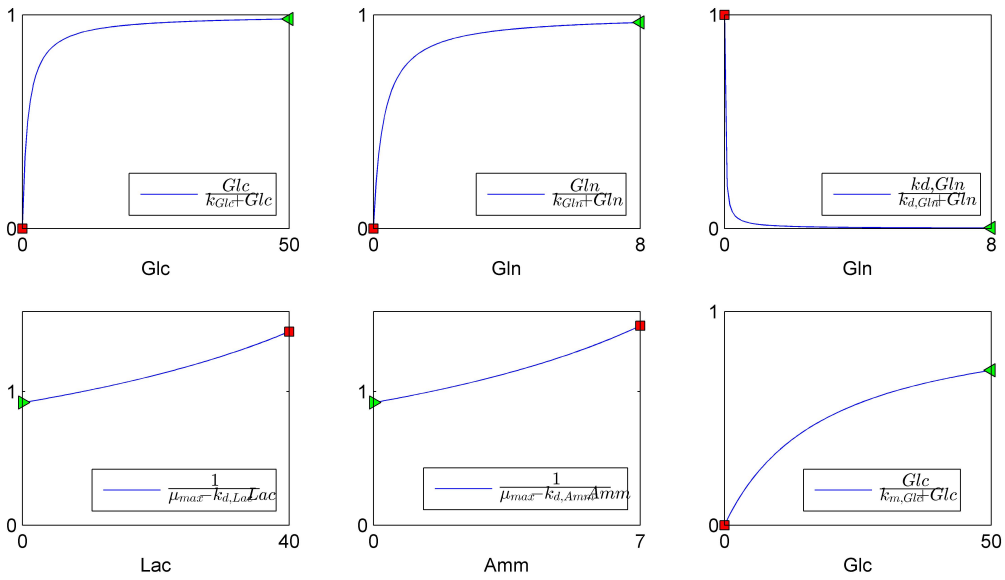


Figure 3.35: How terms used in models $\{\Omega_{5a}, \cdots, \Omega_{5d}\}$ vary for our range of concentrations (mM) when given De Tremblay's parameter values. Green triangles indicate initial value at t_0 and red squares indicate final value at t_f.

To start with, very simple hypotheses will be considered:

- $\mathcal{H}_{5a}$: growth is limited by glucose disappearance, death is enhanced by lactate accumulation (ie, the pair substrate/metabolite Glc/Lac plays a major role)

$$\Omega_{5a} : \begin{cases} \dfrac{dx_1}{dt} = \left(\theta_1 \dfrac{x_2}{\theta_2 + x_2} - \theta_4 \dfrac{1}{\theta_1 - \theta_5 x_3}\right) x_1 \equiv \\ \dfrac{dXv}{dt} = \left(\mu_{max} \dfrac{Glc}{k_{Glc} + Glc} - \mu_{d,max} \dfrac{1}{\mu_{max} - k_{d,Lac} Lac}\right) Xv \end{cases} \tag{3.60}$$

- $\mathcal{H}_{5b}$: growth is limited by glutamine disappearance, death is enhanced by ammonia accumulation (ie, the pair Gln/Amm plays a major role)

$$\Omega_{5b} : \begin{cases} \dfrac{dx_1}{dt} = \left(\theta_1 \dfrac{x_4}{\theta_3 + x_4} - \theta_4 \dfrac{1}{\theta_1 - \theta_6 x_5}\right) x_1 \equiv \end{cases}$$

$$\frac{dXv}{dt} = \left(\mu_{max}\frac{Gln}{k_{Gln}+Gln} - \mu_{d,max}\frac{1}{\mu_{max}-k_{d,Amm}Amm}\right)Xv \tag{3.61}$$

- $\mathcal{H}_{5c}$: growth is limited by glucose and glutamine disappearance, death is enhanced by lactate and ammonia accumulation (ie, both pairs Glc/Lac and Gln/Amm play a major role)

$$\Omega_{5c}: \left\{\frac{dx_1}{dt} = \left(\theta_1\frac{x_2}{\theta_2+x_2}\frac{x_4}{\theta_3+x_4} - \theta_4\frac{1}{\theta_1-\theta_5x_3}\frac{1}{\theta_1-\theta_6x_5}\right)x_1 \equiv \right.$$
$$\frac{dXv}{dt} = \left(\mu_{max}\frac{Glc}{k_{Glc}+Glc}\frac{Gln}{k_{Gln}+Gln} - \right.$$
$$\left.-\mu_{d,max}\frac{1}{\mu_{max}-k_{d,Lac}Lac}\frac{1}{\mu_{max}-k_{d,Amm}Amm}\right)Xv \tag{3.62}$$

- $\mathcal{H}_{5d}$: (de Tremblay, 1991)'s model structure, ie. growth limited by glucose and glutamine disappearance, death enhanced by lactate and ammonia accumulation, insignificant death as long as glutamine abundant, glucose consumption partially explained by cell maintenance activities

$$\Omega_{5d}: \left\{\frac{dx_1}{dt} = \left(\theta_1\frac{x_2}{\theta_2+x_2}\frac{x_4}{\theta_3+x_4} - \theta_4\frac{1}{\theta_1-\theta_5x_3}\frac{1}{\theta_1-\theta_6x_5}\frac{\theta_7}{\theta_7+x_4}\right)x_1 \equiv \right.$$
$$\frac{dXv}{dt} = \left(\mu_{max}\frac{Glc}{k_{Glc}+Glc}\frac{Gln}{k_{Gln}+Gln} - \right.$$
$$\left.-\mu_{d,max}\frac{1}{\mu_{max}-k_{d,Lac}Lac}\frac{1}{\mu_{max}-k_{d,Amm}Amm}\frac{k_{d,Gln}}{k_{d,Gln}+Gln}\right)Xv \tag{3.63}$$

Again, notice that Ω_{5a}, ... are submodels and need to be coupled with the remaining differential equations that further define relations with the other states so that the model is complete:

$$\Omega_{5,rest}: \left\{\begin{array}{ll} \frac{dx_2}{dt} = \frac{-1}{\theta_8}\mu_{growth}x_1, & \frac{dx_3}{dt} = \frac{\theta_9}{\theta_8}\mu_{growth}x_1, \\ \frac{dx_4}{dt} = \frac{-1}{\theta_{12}}\mu_{growth}x_1, & \frac{dx_5}{dt} = \frac{\theta_{13}}{\theta_{12}}\mu_{growth}x_1, \end{array}\right. \tag{3.64}$$

except for submodel Ω_{5d} where glucose, while abundant, is partially used for maintenance purposes and thus:

$$\Omega_{5d,rest}: \begin{cases} \frac{dx_2}{dt} = -\left(\frac{1}{\theta_8}\mu_{growth} + \theta_{10}\frac{x_2}{\theta_{11}+x_2}\right)x_1, & \frac{dx_3}{dt} = \frac{\theta_9}{\theta_8}\mu_{growth}x_1, \\ \frac{dx_4}{dt} = \frac{-1}{\theta_{12}}\mu_{growth}x_1, & \frac{dx_5}{dt} = \frac{\theta_{13}}{\theta_{12}}\mu_{growth}x_1, \end{cases} \tag{3.65}$$

$$where: \theta_{10}\frac{x_2}{\theta_{11}+x_2} \equiv m_{Glc}\frac{Glc}{k_{m,Glc}+Glc}$$

The following initial values were considered for the parameters present overall in the four models:

$$[\theta_{M5abcd}]_0: \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \theta_5 \\ \theta_6 \\ \theta_7 \\ \theta_8 \\ \theta_9 \\ \theta_{10} \\ \theta_{11} \\ \theta_{12} \\ \theta_{13} \end{bmatrix}_0 = \begin{bmatrix} 0.69 & d^{-1} \\ 0.14 & mM \\ random(0-11) & mM \\ 0.35 & d^{-1} \\ random(0-14) & mM^{-1}\,d^{-1} \\ random(0-28) & mM^{-1}\,d^{-1} \\ random(0-900) & mM \\ 0.02 & 10^9 cell\,mmol^{-1} \\ 1.39 & mmol\,mmol^{-1} \\ random(0-2.5\times10^3) & mmol\,10^{-9}cell\,d^{-1} \\ random(0-11) & mM \\ 0.08 & 10^9 cell\,mmol^{-1} \\ 0.74 & mmol\,mmol^{-1} \end{bmatrix}. \tag{3.66}$$

Since the complexity of the models is now greater, starting from a bad initial guess has a bigger impact on the duration of the identification procedure. Therefore, an additional step was added where many random θ_0 are initially generated (eg. 20,000) and trialled (their cost is computed) so as to scan the area of possibilities, and then the algorithm goes on to be launched for the k most promising ones (the best 10 sets of values for θ_0, for example).

In order to improve the speed of the identification procedure, an extra constraint was added, as well: in the initial guess, μ_{max} needs to be bigger than $\mu_{d,max}$ (otherwise, no initial cell growth would be observed, which would be a waste of time).

There are now also more states (5, instead of 3 in step 4), thus the weight of biomass compliance is smaller in the total cost account, J. In order to balance this out and keep a similar weighting to that used in the previous step, the following weighting of the cost function was introduced:

$$J = (1/3)\, J_{x_1} + \left[\frac{(1/3)}{2} J_{x_2} + \frac{(1/3)}{2} J_{x_3}\right] + \left[\frac{(1/3)}{2} J_{x_4} + \frac{(1/3)}{2} J_{x_5}\right] \tag{3.67}$$

This means that, for any experiment, 1/3 of the total cost comes from biomass compliance, another 1/3 from Glc/Lac compliance and the remaining 1/3 from Gln/Amm compliance. In fact, for the final user of the model, the compliance of model predictions to real data is often more important when it comes to biomass for several reasons: one of the culture's purposes may be to culture cells, it may also be to produce a protein (and their submodels are usually rather a function of biomass concentration and not substrate/metabolite concentrations), or simply because mathematically the most important state is biomass as the dynamics of other states are usually based on its dynamics (see equation (3.64) on page 123, for instance).

Table 3.8 presents the results obtained for models defined in equations (3.60) to (3.64).

Figures 3.36 to 3.39 present the corresponding plots for the 4 experiments. It can be seen that with this type of models, towards the end of the culture, when cells are dying out, final lactate and ammonia concentrations present constant profiles. Real data, however, is more "round". This can be attributed possibly to measurement noise and possibly to other phenomena, such as some glucose being partially used for maintenance purposes and glutamine spontaneously decomposing into ammonia. The switch time t_{peak} from the growth phase to the death phase is sometimes not correctly predicted by the models (eg. for experiment 1, models Ω_{5a} and Ω_{5b}). If the biomass curve is not correctly predicted, then the substrate and metabolite curves are also arguable.

Table 3.8: Identification results for step #5.

Parameter / Step	Ω_{5a}	Ω_{5b}	Ω_{5c}	Ω_{5d}
$\theta_1 \equiv \mu_{max}$	1.37 (0.69)	0.80 (0.69)	1.04 (0.69)	1.37 (0.69)
$\theta_2 \equiv k_{Glc}$	0.0014 (0.14)		0.0009 (0.14)	0.0246 (0.14)
$\theta_3 \equiv k_{Gln}$		1.19	2.11	1.21
$\theta_4 \equiv \mu_{d,max}$	1.33 (0.35)	0.22 (0.35)	0.39 (0.35)	1.37 (0.35)
$\theta_5 \equiv k_{d,Lac}$	0.022		0.019	0.019
$\theta_6 \equiv k_{d,Amm}$		0.133	0.037	0.000
$\theta_7 \equiv k_{d,Gln}$				155
$\theta_8 \equiv Y_{Xv/Glc}$	0.212 (0.021)	0.059 (0.021)	0.097 (0.021)	0.165 (0.021)
$\theta_9 \equiv Y_{Lac/Glc}$	1.54 (1.39)	1.23 (1.39)	1.54 (1.39)	1.67 (1.39)
$\theta_{10} \equiv m_{Glc}$				1.31
$\theta_{11} \equiv k_{m,Glc}$				189
$\theta_{12} \equiv Y_{Xv/Gln}$	0.43 (0.08)	0.26 (0.08)	0.42 (0.08)	0.27 (0.08)
$\theta_{13} \equiv Y_{Amm/Gln}$	0.40 (0.74)	0.88 (0.74)	0.83 (0.74)	0.80 (0.74)
Cost J	48.3	63.9	39.1	41.6

Note: for each column (step/model), the initial parameter guess, whenever not randomly generated, is shown in parentheses. Cells shaded in grey indicate parameters considered in that model. Remarks: $^{(*)}$ inconclusive results since, despite convergence obtained in several runs, identified values range from 10 to 10^7 in amplitude.

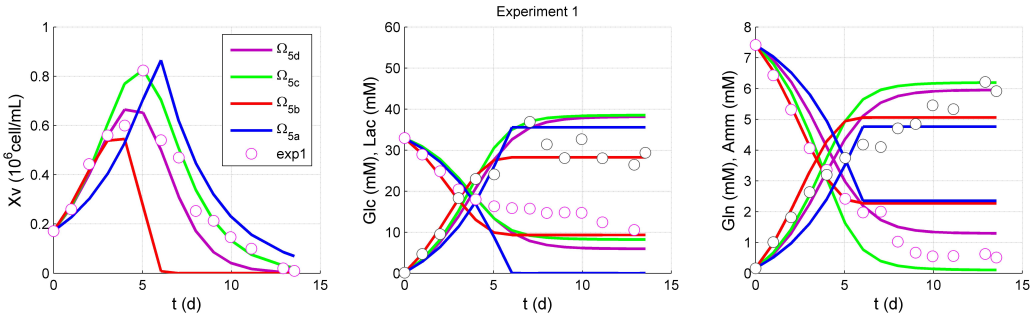


Figure 3.36: Models $\{\Omega_{5a} + \Omega_{5,rest}\}$ in blue, $\{\Omega_{5b} + \Omega_{5,rest}\}$ in red, $\{\Omega_{5c} + \Omega_{5,rest}\}$ in green, $\{\Omega_{5d} + \Omega_{5d,rest}\}$ in purple, experiment 1, CHO-S: simulations with models identified using the complete databank. Circles represent experiment 1's dataset.

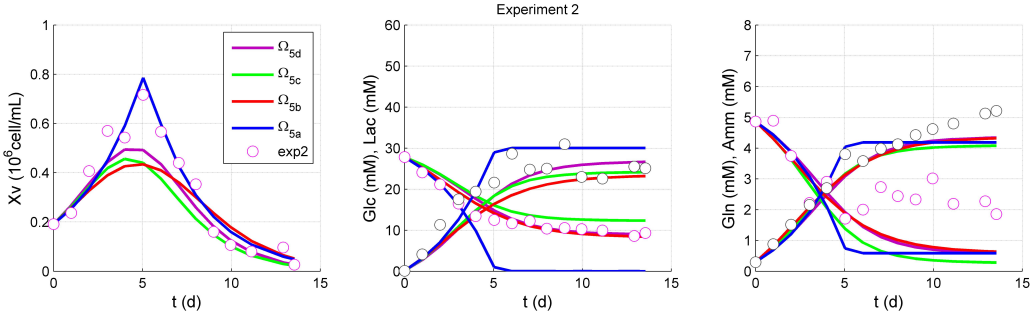


Figure 3.37: Models $\{\Omega_{5a} + \Omega_{5,rest}\}$ in blue, $\{\Omega_{5b} + \Omega_{5,rest}\}$ in red, $\{\Omega_{5c} + \Omega_{5,rest}\}$ in green, $\{\Omega_{5d} + \Omega_{5d,rest}\}$ in purple, experiment 2, CHO-S: simulations with models identified using the complete databank. Circles represent experiment 2's dataset.

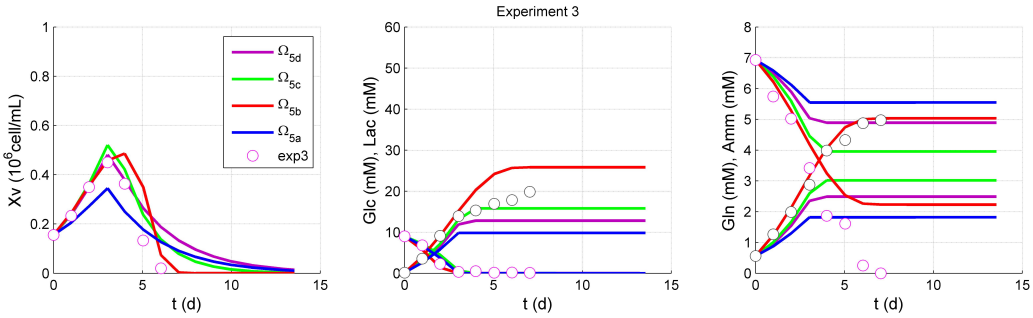


Figure 3.38: Models $\{\Omega_{5a} + \Omega_{5,rest}\}$ in blue, $\{\Omega_{5b} + \Omega_{5,rest}\}$ in red, $\{\Omega_{5c} + \Omega_{5,rest}\}$ in green, $\{\Omega_{5d} + \Omega_{5d,rest}\}$ in purple, experiment 3, CHO-S: simulations with models identified using the complete databank. Circles represent experiment 3's dataset.

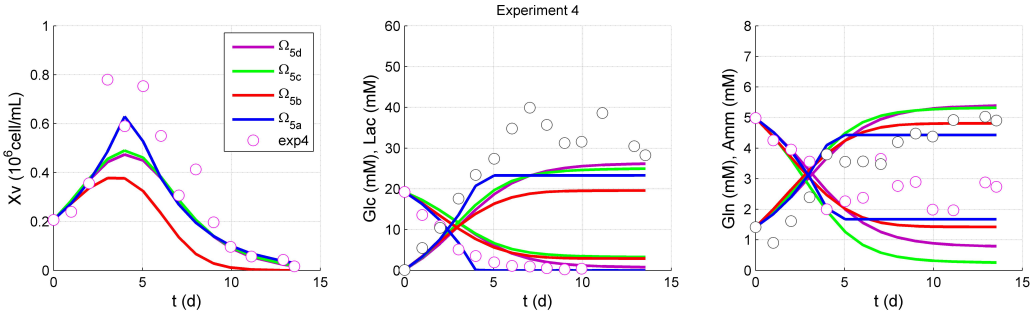


Figure 3.39: Models $\{\Omega_{5a} + \Omega_{5,rest}\}$ in blue, $\{\Omega_{5b} + \Omega_{5,rest}\}$ in red, $\{\Omega_{5c} + \Omega_{5,rest}\}$ in green, $\{\Omega_{5d} + \Omega_{5d,rest}\}$ in purple, experiment 4, CHO-S: simulations with models identified using the complete databank. Circles represent experiment 4's dataset.

Models are imperfect, but it is important that, whatever the model-plant mismatch may be, at least the switch time between the growth and the death phase is more or less correctly identified. In fact, this will help us analyse fitted curves and decide if other phenomena should be considered in the model form. A better estimate of the switch time translates into a bigger compliance around t_{peak}. This wish has been expressed by introducing a timewise weighting w_i in the cost function that accounts for all i points of a component's concentration during any particular experiment:

$$J = \sum_i w_i J_i \tag{3.68}$$

It was also considered that the compliance of points around t_{peak} should be greatly weighted. The values chosen are $w(t_0) = 10$, $w(t_{peak}) = 100$, $w(t_f) = 1$ as shown in Figure 3.40 which can in some way be interpreted as an oversimplification of a typical $Xv(t)$ curve.

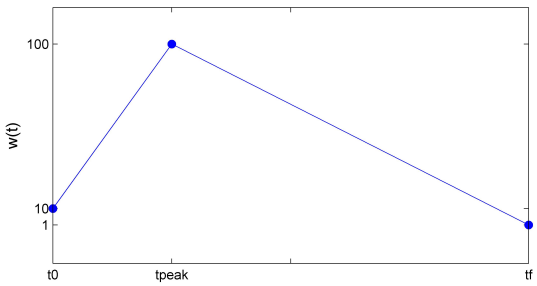


Figure 3.40: Timewise weighting.

The algorithm was run again for these models using timewise weighting and achieved better results. Figure 3.41 exemplifies how these can be improved when the switch time t_{peak} is correctly estimated.[1]

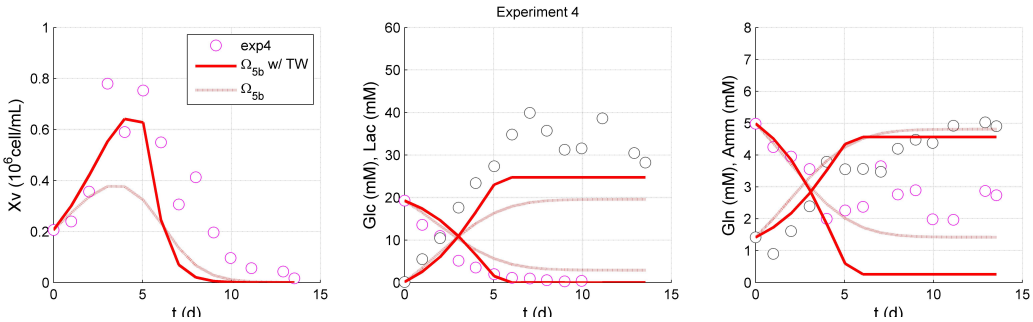


Figure 3.41: Model $\{\Omega_{5b} + \Omega_{5,rest}\}$ without timeweighting the cost function (light red) and timeweighting it (dark red). Circles represent experiment 4's dataset.

Whilst without timeweighting, the biomass concentration is, in this case, underestimated (and thus the identified values for the stoichiometric coefficients should be further from the real values),[2] when timeweighting is introduced, compliance for points around t_{peak} gains more importance. The curve in bold presents a more interesting fit for biomass concentration. It is thus expected that the stoichiometric coefficients will be better estimated with this cost function. With regard to better or worse compliance around the last points of each experiment, this will allow us to investigate whether other phenomena (eg. spontaneous glutamine degradation) are still unexplained and need to be included in the model.

The new identified values (now with timeweighting) are comparable and fits better, as shown on Table 3.9 and Figures 3.42 to 3.45.

[1]Notice that the cost function value is not comparable anymore to the previous values, since the function is built differently. A visual check was thus used to inspect the identification results.

[2]Since substrate consumption and metabolite production per cell are, in this case, based on a bad fit for $Xt(t)$.

Table 3.9: Identification results for step #5 using timeweighting.

Parameter / Step	Ω_{5a}	Ω_{5b}	Ω_{5c}	Ω_{5d}
$\theta_1 \equiv \mu_{max}$	1.41 (0.69)	0.76 (0.69)	1.26 (0.69)	0.97 (0.69)
$\theta_2 \equiv k_{Glc}$	0.0970 (0.14)		0.0425 (0.14)	0.0001 (0.14)
$\theta_3 \equiv k_{Gln}$		1.00	2.35	2.82
$\theta_4 \equiv \mu_{d,max}$	1.41 (0.35)	0.21 (0.35)	0.74 (0.35)	0.23 (0.35)
$\theta_5 \equiv k_{d,Lac}$	0.020		0.015	0.016
$\theta_6 \equiv k_{d,Amm}$		0.125	0.000	0.037
$\theta_7 \equiv k_{d,Gln}$				7.64×10^8
$\theta_8 \equiv Y_{Xv/Glc}$	0.225 (0.021)	0.059 (0.021)	0.111 (0.021)	0.078 (0.021)
$\theta_9 \equiv Y_{Lac/Glc}$	1.68 (1.39)	1.22 (1.39)	1.66 (1.39)	1.57 (1.39)
$\theta_{10} \equiv m_{Glc}$				0.747
$\theta_{11} \equiv k_{m,Glc}$				0.287
$\theta_{12} \equiv Y_{Xv/Gln}$	0.43 (0.08)	0.26 (0.08)	0.42 (0.08)	0.27 (0.08)
$\theta_{13} \equiv Y_{Amm/Gln}$	0.40 (0.74)	0.88 (0.74)	0.83 (0.74)	0.80 (0.74)
Cost J	48.3	63.2	39.2	41.6

Note: for each column (step/model), the initial parameter guess, whenever not randomly generated, is shown in parentheses. Cells shaded in grey indicate parameters considered in that model.

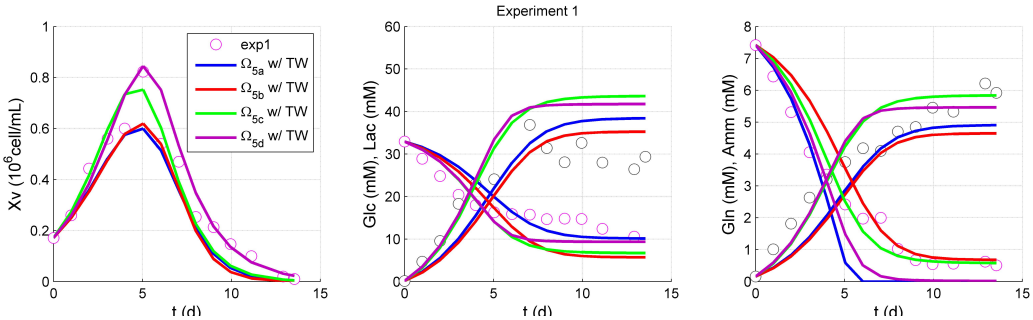


Figure 3.42: Models $\{\Omega_{5a} + \Omega_{5,rest}\}$ in blue, $\{\Omega_{5b} + \Omega_{5,rest}\}$ in red, $\{\Omega_{5c} + \Omega_{5,rest}\}$ in green, $\{\Omega_{5d} + \Omega_{5d,rest}\}$ in purple, experiment 1 (circles), CHO-S: simulations with models identified using the complete databank. Timeweighting of the cost function was used as described in eq. (3.68), p. 128.

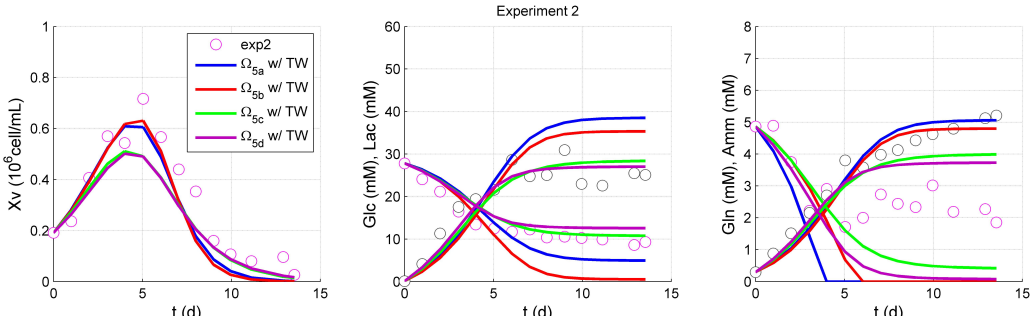


Figure 3.43: Models $\{\Omega_{5a} + \Omega_{5,rest}\}$ in blue, $\{\Omega_{5b} + \Omega_{5,rest}\}$ in red, $\{\Omega_{5c} + \Omega_{5,rest}\}$ in green, $\{\Omega_{5d} + \Omega_{5d,rest}\}$ in purple, experiment 2 (circles), CHO-S: simulations with models identified using the complete databank. Timeweighting of the cost function was used as described in eq. (3.68), p. 128.

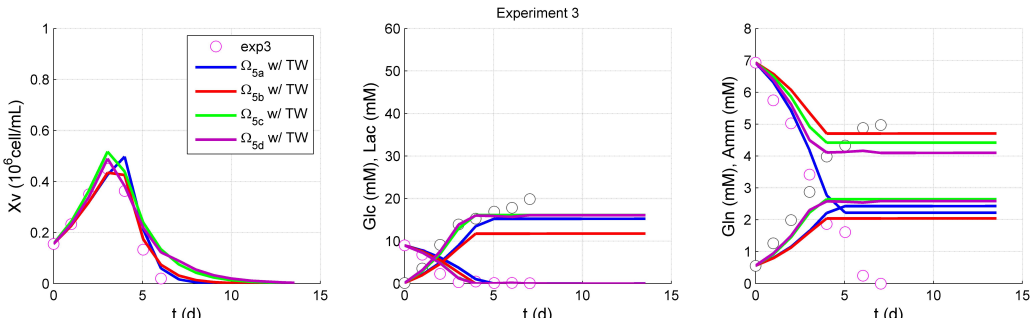


Figure 3.44: Models $\{\Omega_{5a} + \Omega_{5,rest}\}$ in blue, $\{\Omega_{5b} + \Omega_{5,rest}\}$ in red, $\{\Omega_{5c} + \Omega_{5,rest}\}$ in green, $\{\Omega_{5d} + \Omega_{5d,rest}\}$ in purple, experiment 3 (circles), CHO-S: simulations with models identified using the complete databank. Timeweighting of the cost function was used as described in eq. (3.68), p. 128.

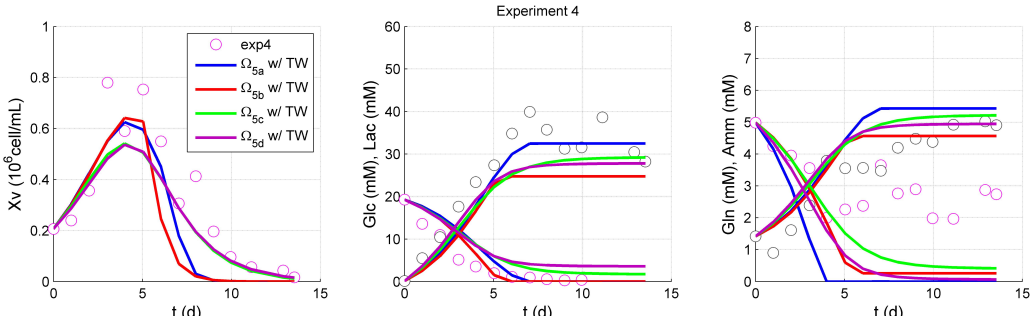


Figure 3.45: Models $\{\Omega_{5a} + \Omega_{5,rest}\}$ in blue, $\{\Omega_{5b} + \Omega_{5,rest}\}$ in red, $\{\Omega_{5c} + \Omega_{5,rest}\}$ in green, $\{\Omega_{5d} + \Omega_{5d,rest}\}$ in purple, experiment 4 (circles), CHO-S: simulations with models identified using the complete databank. Timeweighting of the cost function was used as described in eq. (3.68), p. 128.

Now that some parameter values were identified for our CHO cell culture, we can re-plot Figure 3.35 (terms in De Tremblay's model) for our models Ω_{5a} to Ω_{5d} in order to check if and how they vary within the ranges of concentrations of our cultures. This is done in Figure 3.46.

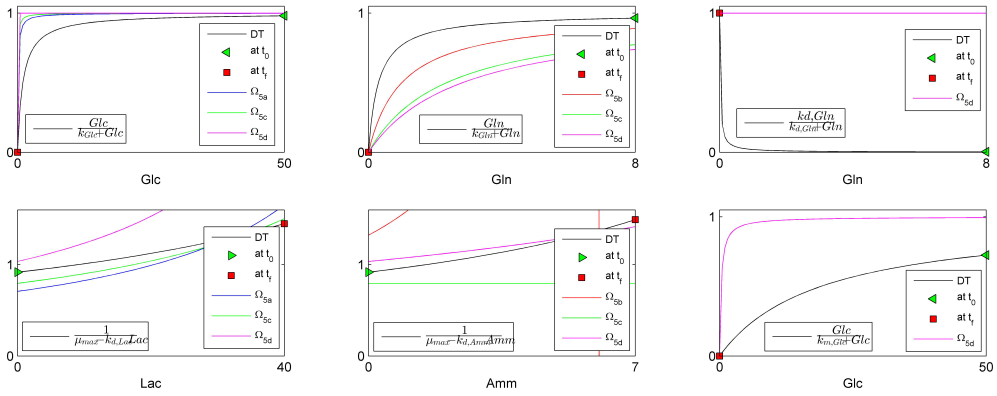


Figure 3.46: How terms used in models Ω_{5a} (blue), Ω_{5b} (red), Ω_{5c} (green), Ω_{5d} (magenta) vary for our range of concentrations. In black, using De Tremblay's values for another type of animal cells. Green triangles indicate initial value at t_0 and red squares indicate final value at t_f.

It can be seen globally that terms are varying within the ranges of concentration of the culture. One exception is $k_{d,Gln}/(k_{d,Gln} + Gln)$ which is always

roughly equal to 1 in model Ω_{5d} (eq. (3.63) on p. 123). This means that this model could actually be reduced (the term in question is being multiplied and could be discarded).

The term $1/(\mu_{max} - k_{d,Amm}Amm)$ presents a problem though (eq. (3.61), p. 123): the denominator has a zero that accounts for unusual behavior. Using (de Tremblay, 1991)'s model values, this behaviour happens well above the range of ammonia concentrations (black line in Figure 3.46). However, with the values identified for Ω_{5b}, this happens within the range. A different expression to describe this phenomenon (ammonia inhibition) could therefore be used.

On the whole, the best model so far seems to be Ω_{5c} since it reached the lowest value for the cost function. Model Ω_{5d} (De Tremblay) is not very far, with the remarks that the term $k_{d,Gln}/(k_{d,Gln} + Gln)$ could in this case be discarded (as discussed above) and that it might not be necessary to introduce term $Glc/(k_{m,Glc} + Glc)$ to express limitation of the partial use of glucose in cell maintenance purposes (eq. (3.65), p. 124).

Let us now try to improve the very simple model forms considered previously by considering either new structures or changes to their structures. One of the improvements will be the choice of other forms for representing ammonia and lactate inhibition phenomena.

- $\mathcal{H}_{5e}$: growth is limited by glucose and glutamine disappearance and inhibited by lactate and ammonia accumulation)

$$\Omega_{5e} : \left\{ \frac{dx_1}{dt} = \left(\overbrace{\theta_1 \frac{x_2}{\theta_2 + x_2} \frac{x_4}{\theta_3 + x_4} \frac{\theta_5}{\theta_5 + x_3} \frac{\theta_6}{\theta_6 + x_5}}^{\mu_{growth}} - \overbrace{\theta_4}^{\mu_{death}} \right) x_1 \equiv \right.$$

$$\frac{dXv}{dt} = \left(\mu_{max} \frac{Glc}{k_{Glc} + Glc} \frac{Gln}{k_{Gln} + Gln} \frac{k_{i,Lac}}{k_{i,Lac} + Lac} \frac{k_{i,Amm}}{k_{i,Amm} + Amm} - \mu_{d,max} \right) Xv \tag{3.69}$$

In comparison to model Ω_{5c} which presumes the same set of phenomena, we have now chosen to test an alternative classic form of expressing it (from the list of forms available in Tables 2.1 and 2.2). Before, it was presumed that baseline death was enhanced by accumulation of

metabolites, here the baseline death is constant and it is growth that decreases with it. Let us recall the shape of the mathematical forms used:

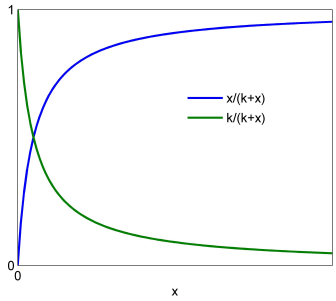


Figure 3.47: Typical model forms used to express inhibition or limitation in cell cultures.

As a side note, let us mention that, in order to be sure that the search space is being well covered (when x random guesses are generated so that the best ones can be chosen to initialize the algorithm), we have increased them from 20.000 to 100.000 and checked the results. An example is provided in Figure 3.48. No great improvement was found and, therefore, it was considered that 20.000 is a number big enough to scrutinize the space of possible θ_0 for the amount of parameters with which we are dealing[1].

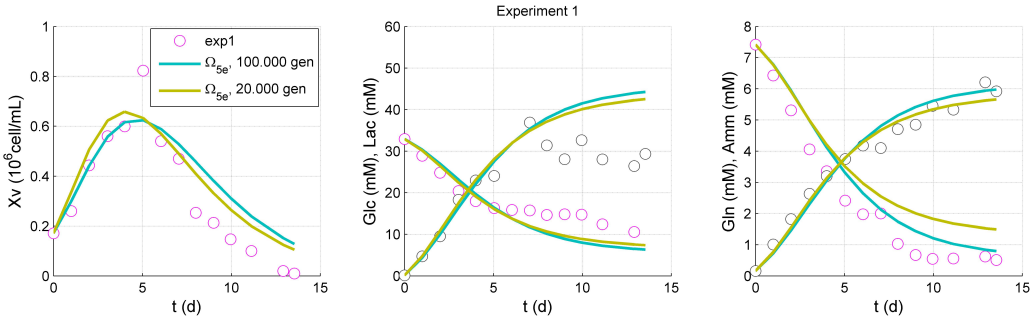


Figure 3.48: Model $\{\Omega_{5e} + \Omega_{5e,rest}\}$ identified using the best θ_0 values out of 20.000 random ones (in yellow) and out of 100.000 random ones (in blue). Experiment 1, CHO-S.

- $\mathcal{H}_{5f}$: equal to Ω_{5e} but the possibility of extra phenomena happening is added, namely that some glutamine may disappear due to spontaneous

[1]Throughout this case study, on an Intel(R) Core(TM)2 Duo CPU P8400 @2.26GHz Memory (RAM) 4GB 32-bit Windows operating system, this procedure lasted a duration of a couple of minutes to a couple of hours, depending on the complexity of the model structure and number of parameters considered in the step.

degradation (rate θ_7) and that some glucose may be used also for cell maintenance purposes (rate θ_{10}).

We would also like to test another way of expressing the limitation/inhibition of net growth:

- $\mathcal{H}_{5g}$: growth is limited by glucose and glutamine disappearance, death is null if lactate or ammonia are absent but enhanced if they are present.

$$\Omega_{5g}: \left\{ \frac{dx_1}{dt} = \left(\theta_1 \frac{x_2}{\theta_2 + x_2} \frac{x_4}{\theta_3 + x_4} - \theta_4 \frac{x_3}{\theta_5 + x_3} \frac{x_5}{\theta_6 + x_5} \right) x_1 \equiv \right.$$
$$\left. \frac{dXv}{dt} = \left(\mu_{max} \frac{Glc}{k_{Glc} + Glc} \frac{Gln}{k_{Gln} + Gln} - \mu_{d,max} \frac{Lac}{k_{i,Lac} + Lac} \frac{Amm}{k_{i,Amm} + Amm} \right) Xv \right. \tag{3.70}$$

  For submodel Ω_{5e}, as soon as one substrate concentration is null, then there is null growth and a constant death rate (the later time period of the culture being then largely defined by parameter $\mu_{d,max}$). Here in Ω_{5g}, the assumption is that a (maximum) death rate will be low as long as lactate and ammonia do not accumulate too much. The later period of the culture can be largely shaped by 3 parameters $\mu_{d,max}$, $k_{d,Lac}$ and $k_{d,Amm}$.

- $\mathcal{H}_{5h}$: equal to Ω_{5g} but, similarly to Ω_{5f}, extra phenomena are added to Ω_{5g}: glutamine partially decomposes spontaneously (rate θ_7) and glucose is partially used for maintenance (rate θ_{10}).

Once more, notice that Ω_{5e}, ... are submodels and need to be coupled with the remaining differential equations: eq. (3.64) on p. 123 for Ω_{5e} and Ω_{5g}, and, for the specific case of Ω_{5f} and Ω_{5h}:

$$\Omega_{5fh,rest}: \left\{ \begin{array}{ll} \frac{dx_2}{dt} = -\left(\frac{1}{\theta_8} \mu_{growth} + \theta_{10} \right) x_1, & \frac{dx_3}{dt} = \theta_9 \left(\frac{1}{\theta_8} \mu_{growth} + \theta_{10} \right) x_1, \\ \frac{dx_4}{dt} = -\left(\frac{1}{\theta_{12}} \mu_{growth} + \theta_7 \frac{x_4}{x_1} \right) x_1, & \frac{dx_5}{dt} = \theta_{13} \left(\frac{1}{\theta_{12}} \mu_{growth} + \theta_7 \frac{x_4}{x_1} \right) x_1, \end{array} \right. \tag{3.71}$$

As for initial guesses, the standard procedure was followed, except that models Ω_{5f} and Ω_{5h}, being more complex forms of models Ω_{5e} and Ω_{5f}, were tested using the input from the identified values of their simpler forms.

Overall, it seems that the inclusion of spontaneous glutamine degradation and maintenance glucose is positive: the cost of model Ω_{5e} (35.6) gets as low as 19.8 in Ω_{5f} and the cost of Ω_{5g} (39.8) is lowered to 31.2 in Ω_{5h}. This can be seen in Table 3.10 and Figures 3.49 to 3.52.

Table 3.10: More identification results for step #5 using timeweighting.

Parameter / Step	Ω_{5e}	Ω_{5f}	Ω_{5g}	Ω_{5h}
$\theta_1 \equiv \mu_{max}$	1.55 (0.69)	1.51 (1.55)	1.08 (0.69)	1.22 (1.08)
$\theta_2 \equiv k_{Glc}$	0.022 (0.14)	0.003 (0.022)	0.093 (0.14)	1.005 (0.093)
$\theta_3 \equiv k_{Gln}$	1.22	0.92 (1.22)	2.61	1.22 (2.61)
$\theta_4 \equiv \mu_{d,max}$	0.501 (0.354)	0.522 (0.501)	0.630 (0.354)	0.770 (0.630)
$\theta_5 \equiv k_{d,Lac}$	21.2	17.2 (21.2)	3.75	1.74 (3.75)
$\theta_6 \equiv k_{d,Amm}$	23.7 (23.7)	633	0.000 (0.000)	0.004
$\theta_7 \equiv k_{\alpha,Gln}$		0.184		0.195
$\theta_8 \equiv Y_{Xv/Glc}$	0.104 (0.021)	0.107 (0.104)	0.095 (0.021)	0.207 (0.095)
$\theta_9 \equiv Y_{Lac/Glc}$	1.66 (1.39)	1.58 (1.66)	1.69 (1.39)	1.53 (1.69)
$\theta_{10} \equiv m_{Glc}$		0.16		2.44
$\theta_{12} \equiv Y_{Xv/Gln}$	0.45 (0.08)	6.42 (0.45)	0.35 (0.08)	1.55 (0.35)
$\theta_{13} \equiv Y_{Amm/Gln}$	0.93 (0.74)	0.80 (0.93)	0.81 (0.74)	0.66 (0.81)
Cost J	35.6	19.8	39.8	31.2

Note: for each column (step/model), the initial parameter guess, whenever not randomly generated, is shown in parentheses. Cells shaded in grey indicate parameters considered in that model.

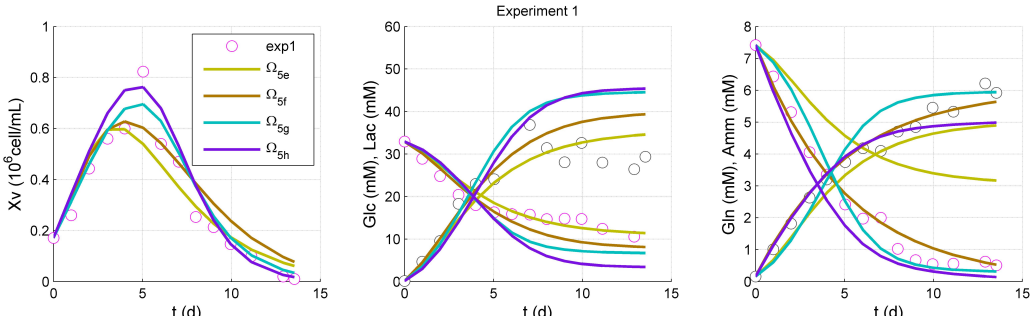


Figure 3.49: Models $\{\Omega_{5e} + \Omega_{5e,rest}\}$ in yellow, $\{\Omega_{5f} + \Omega_{5f,rest}\}$ in brown, $\{\Omega_{5g} + \Omega_{5g,rest}\}$ in cyan, $\{\Omega_{5h} + \Omega_{5h,rest}\}$ in purple, experiment 1 (circles), CHO-S: simulations with models identified using the complete databank. Timeweighting of the cost function was used as described in eq. (3.68), p. 128.

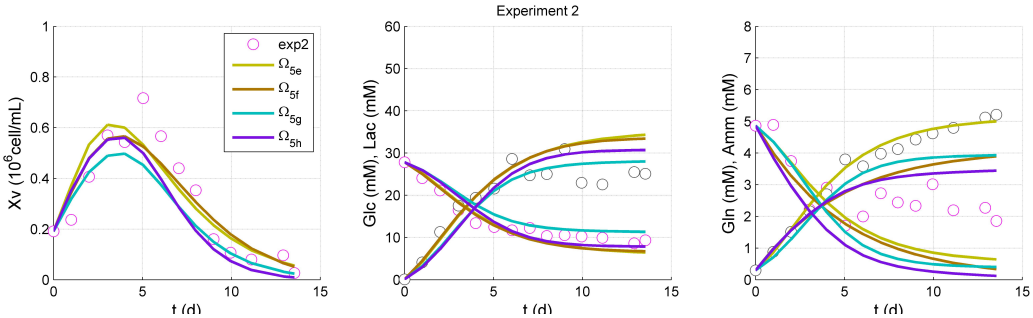


Figure 3.50: Models $\{\Omega_{5e} + \Omega_{5e,rest}\}$ in yellow, $\{\Omega_{5f} + \Omega_{5f,rest}\}$ in brown, $\{\Omega_{5g} + \Omega_{5g,rest}\}$ in cyan, $\{\Omega_{5h} + \Omega_{5h,rest}\}$ in purple, experiment 2 (circles), CHO-S: simulations with models identified using the complete databank. Timeweighting of the cost function was used as described in eq. (3.68), p. 128.

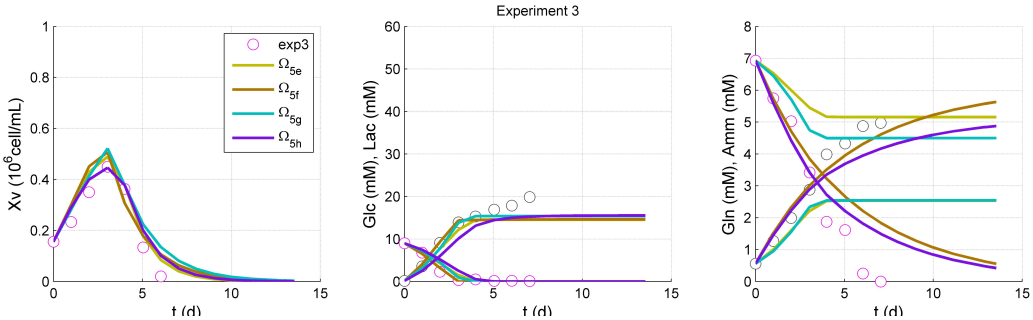


Figure 3.51: Models $\{\Omega_{5e} + \Omega_{5e,rest}\}$ in yellow, $\{\Omega_{5f} + \Omega_{5f,rest}\}$ in brown, $\{\Omega_{5g} + \Omega_{5g,rest}\}$ in cyan, $\{\Omega_{5h} + \Omega_{5h,rest}\}$ in purple, experiment 3 (circles), CHO-S: simulations with models identified using the complete databank. Timeweighting of the cost function was used as described in eq. (3.68), p. 128.

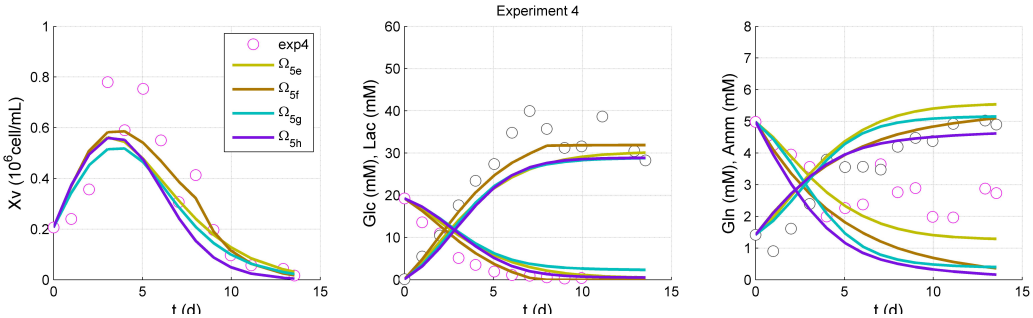


Figure 3.52: Models $\{\Omega_{5e} + \Omega_{5e,rest}\}$ in yellow, $\{\Omega_{5f} + \Omega_{5f,rest}\}$ in brown, $\{\Omega_{5g} + \Omega_{5g,rest}\}$ in cyan, $\{\Omega_{5h} + \Omega_{5h,rest}\}$ in purple, experiment 4 (circles), CHO-S: simulations with models identified using the complete databank. Timeweighting of the cost function was used as described in eq. (3.68), p. 128.

Model Ω_{5f} (eq. (3.83)) is apparently the best, presenting the lowest cost. Let us now describe the procedure for computing confidence intervals on the parameter values identified.

- For the identified set of parameters $\hat{\theta} = [1.51\ 0.003\ \cdots]_{1\times 12}$ retrieve simulated values for all experiments:

$$(t_{sim}, x_{sim}) = f(x, \hat{\theta}, u) \tag{3.72}$$

- Compute difference, throughout time, between simulated and measured value (if measured) for all states.

$$exp.\ 1 : \begin{cases} error_{x_1} = x_{1,meas} - x_{1,sim} \\ \vdots \end{cases} \\ \vdots \tag{3.73}$$

- Compute cost per state considering available datapoints all experiments comprised.

$$\begin{aligned} J_{x_1} &= \sum_{\substack{when \\ available}} error_{x_1}^2 \\ &\vdots \\ J_{x_5} &= \dots \end{aligned} \tag{3.74}$$

- Compute variance of the errors, ie. a measure, for each model state, of the average data-simulation error variability per point. This average step is important since not all timepoints are available (eg: sometimes, in a given experiment at a given timepoint, one of the concentrations was not measured). Therefore, for example, var_{x_1} expresses how distant, on average, simulated biomass concentrations are to measured values, globally throughout all experiments and all culture phases.

$$\begin{aligned} var_{x_1} &= \frac{J_{x_1}}{\text{nr avaible datapoints} - \text{nr parameters}} \\ &\vdots \\ var_{x_5} &= \dots \end{aligned} \tag{3.75}$$

- Compute time courses of sensitivities, for each state and each experiment, using $\hat{\theta}$ obtained and $S_{x_i,\theta_j}(t_0) = 0$ as initial condition. The terms needed for the simultaneous integration of equations (3.8)-(3.9) on p. 83 were derived and are listed in Appendix A.2.4 throughout pages 230-237.

$$exp.\ 1 : \begin{cases} [S_{x_1,\theta_1} \dots S_{x_1,\theta_{12}}](t) = \dots \\ \vdots \\ [S_{x_5,\theta_1} \dots S_{x_1,\theta_{12}}](t) = \dots \end{cases} \\ \vdots \tag{3.76}$$

- Compute Fisher Information Matrix (FMI).

$$exp.\,1: \begin{cases} \text{for each } t_k: \\ \frac{dy}{d\theta} = \begin{bmatrix} S_{x_1,\theta_1} \dots S_{x_1,\theta_{12}} \\ \vdots \\ S_{x_5,\theta_1} \dots S_{x_1,\theta_{12}} \end{bmatrix}; \quad \sigma = \begin{bmatrix} var_{x_1} & 0 & \dots & 0 \\ 0 & var_{x_2} & \dots & 0 \\ \vdots & & \ddots & \\ 0 & & & var_{x_5} \end{bmatrix} \\ F = \frac{dy}{d\theta}^T \sigma^{-1} \frac{dy}{d\theta} \\ \\ \text{all } t_k \text{ considered}: F = F_{t_0} + F_{t_1} + \dots + F_{t_f} \end{cases}$$
$$\vdots \tag{3.77}$$

- Compute covariance.

$$cov_{(np \times np)} = F^{-1} \tag{3.78}$$

- Compute intervals on parameters (upper and lower bounds for values identified, considering a certain confidence level for a t-student distribution with $t_{95\%} = 1.96$).

$$\theta_i^{UB} = \hat{\theta}_i + t_{95\%} \sqrt{cov(i,i)} \tag{3.79}$$

$$\theta_i^{LB} = \hat{\theta}_i - t_{95\%} \sqrt{cov(i,i)} \tag{3.80}$$

In order to compute bounds on the prediction of the evolution of concentrations (ie, the system states), the procedure below was followed:

- Compute covariance, for each t_k in each experiment.

$$exp.\,1: \begin{cases} G_{exp1}(t_k) = \begin{bmatrix} S_{x_1,\theta_1}(t_k) & \dots & S_{x_1,\theta_{12}}(t_k) \\ \vdots & & \vdots \\ S_{x_6,\theta_1}(t_k) & \dots & S_{x_6,\theta_{12}}(t_k) \end{bmatrix} \\ \\ var_{exp.1}(t_k) = G_{exp1}(t_k) \times cov \times G_{exp1}^T(t_k) \\ \\ \text{ie, for a particular } t_k,\ var_{exp1}(t_k) = [b_{i,j}]_{(ns \times ns)} \text{ such that} \\ b_{i,j} = \sum_{f=1}^{np=12} \left[S_{x_i,\theta_f}(t_k) \left(\sum_{l=1}^{np=12} cov_{f,l} S_{x_j,\theta_l}(t_k) \right) \right] \end{cases}$$
$$\vdots \tag{3.81}$$

- Compute bounds on state estimates for each timepoint t_k of each experiment.

$$\begin{array}{l} exp.\,1: \left\{ \begin{array}{l} x_1^{UB}(t_k) = \hat{x_1}(t_k) + t_{95\%}\sqrt{var_{x_1,x_1}(t_k)} \\ x_1^{LB}(t_k) = \hat{x_1}(t_k) - t_{95\%}\sqrt{var_{x_1,x_1}(t_k)} \\ \vdots \end{array} \right. \\ \vdots \end{array} \tag{3.82}$$

The identified values $\hat{\theta}$ and bounds are listed in Table 3.11. Model simulation along with their bounds can be seen in Figure 3.53. For the four experiments, most data points (variability considered in pink) include the range of the predicted values (the interval between lower and upper green lines), which is good. The confidence intervals are on the whole quite large, essentially meaning that the collected data may not be as informative as it should ideally be, and also that the data itself may contain considerable variability (both biological and probe-related). The core parameters, however, (maximum growth and death rate and stoichiometric coefficients) seem to be sound for this model, Ω_{5f}. On the other hand, two parameters, θ_2 and θ_6, were not well extracted from the databank (the experiments were not informative enough) and will be, thus, randomly made equal to the values 0.003 and 633, respectively.

$$\text{Model } \Omega_{5f}: \frac{dx_1}{dt} = \left(\overbrace{\theta_1 \frac{x_2}{\theta_2 + x_2} \frac{x_4}{\theta_3 + x_4} \frac{\theta_5}{\theta_5 + x_3} \frac{\theta_6}{\theta_6 + x_5}}^{\mu_{growth}} - \overbrace{\theta_4}^{\mu_{death}} \right) x_1$$

$$\begin{array}{ll} \dfrac{dx_2}{dt} = -\left(\dfrac{1}{\theta_8}\mu_{growth} + \theta_{10}\right)x_1, & \dfrac{dx_3}{dt} = \theta_9\left(\dfrac{1}{\theta_8}\mu_{growth} + \theta_{10}\right)x_1, \\ \dfrac{dx_4}{dt} = -\left(\dfrac{1}{\theta_{12}}\mu_{growth} + \theta_7\dfrac{x_4}{x_1}\right)x_1, & \dfrac{dx_5}{dt} = \theta_{13}\left(\dfrac{1}{\theta_{12}}\mu_{growth} + \theta_7\dfrac{x_4}{x_1}\right)x_1, \end{array} \tag{3.83}$$

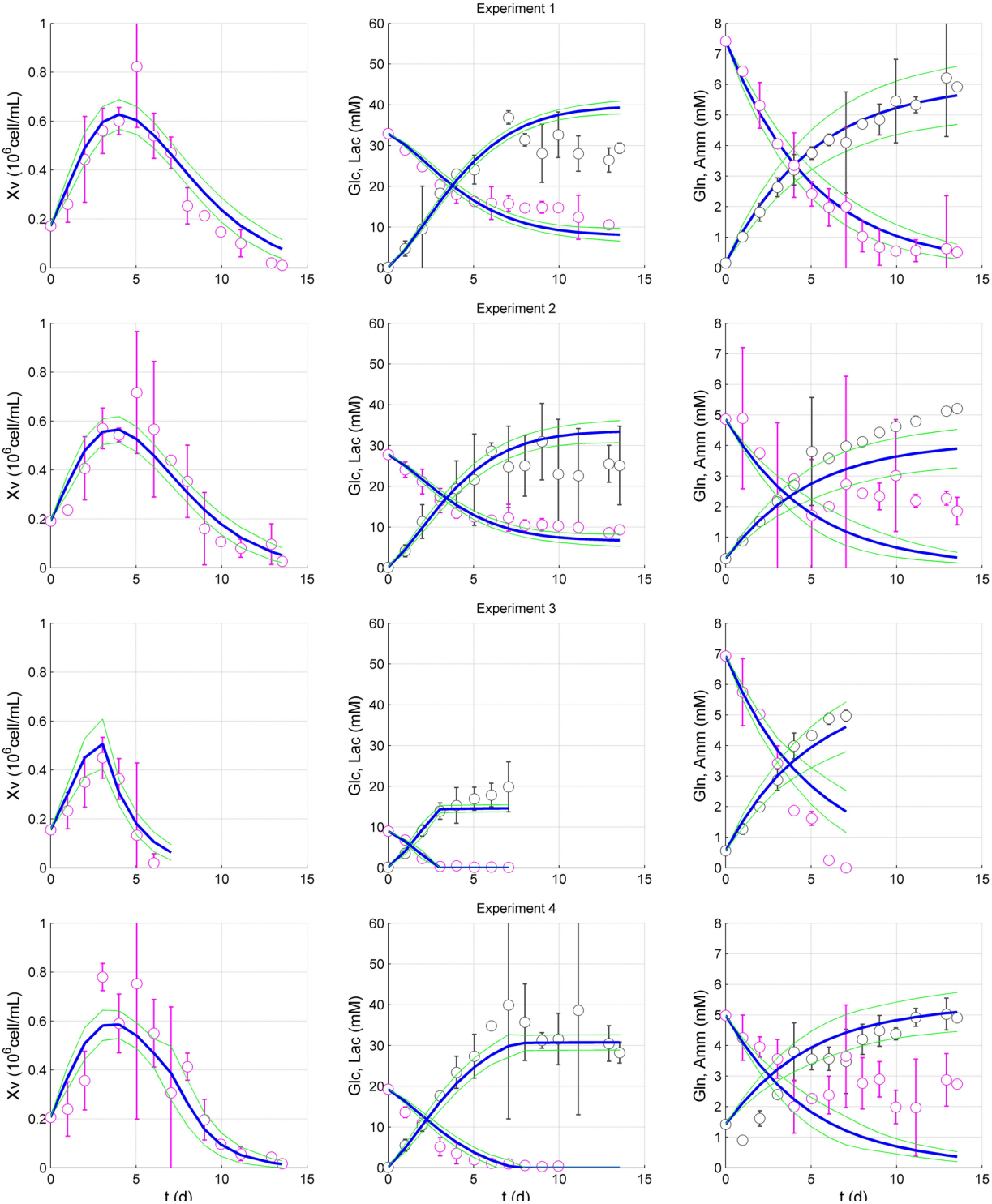


Figure 3.53: Model Ω_{5f}, CHO-S: model simulation (blue), 95%-confidence interval for states (green), measurements (circles) and their variability (error bars, 2σ) .

Table 3.11: Confidence intervals for model Ω_{5f}.

Parameter	LB	Identified Value	UB	Max. Rel Error
$\theta_1 \equiv \mu_{max}$	1.23	1.51	1.79	0.19
$\theta_2 \equiv k_{Glc}$	0	0.003	0.874	291
$\theta_3 \equiv k_{Gln}$	0	0.92	2.17	1.35
$\theta_4 \equiv \mu_{d,max}$	0.167	0.522	0.877	0.680
$\theta_5 \equiv k_{d,Lac}$	0	17.2	38.7	1.24
$\theta_6 \equiv k_{d,Amm}$	0	633	35200	54.5
$\theta_7 \equiv k_{\alpha,Gln}$	0.086	0.184	0.282	0.533
$\theta_8 \equiv Y_{Xv/Glc}$	0.067	0.107	0.146	0.370
$\theta_9 \equiv Y_{Lac/Glc}$	1.45	1.58	1.71	0.08
$\theta_{10} \equiv m_{Glc}$	0	0.16	4.1	25
$\theta_{12} \equiv Y_{Xv/Gln}$	0	6.42	60.3	8.40
$\theta_{13} \equiv Y_{Amm/Gln}$	0.60	0.80	0.99	0.24

Note: values for a 95% confidence level supposing a t-student distribution: $\theta_{LB/UB} = \hat{\theta} \pm t_{95\%}\sqrt{\sigma^2}$ where σ^2 is computed from the inverse of the Fisher matrix, all experiments considered, as defined in eq. (3.32) on p. 103, or more explicitly in eq. (3.79)-(3.80), p. 140.

It is, thus, reasonable to say that, given the available data, model Ω_{5f} is globally acceptable. The focus was put on compliance around t_{peak} rather than towards the end of the culture in order to correctly capture the switch time. Furthermore, concerning cross-validation, if, for example, only experiments A, B and C are accounted for in the identification of parameters, then it is also possible to reasonably predict experiment D, as exemplified in Figure 3.54.

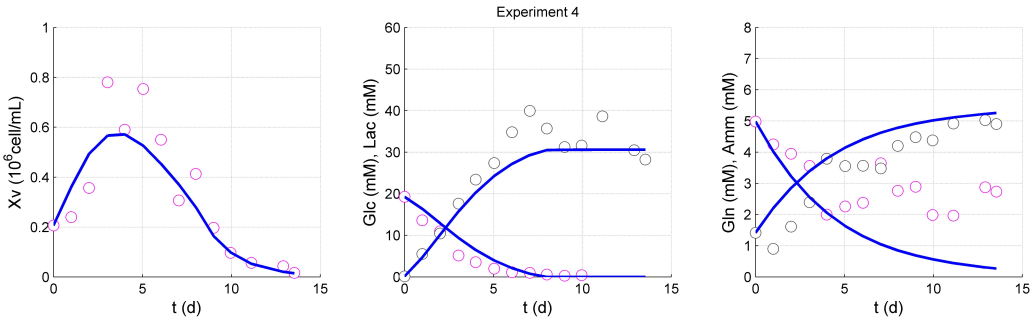


Figure 3.54: Model Ω_{5f}, experiment 4, CHO-S: cross-validation of model simulations (blue) on experiment 4, when only data from experiments 1 to 3 is considered.

However, it should be stressed that the databank is only made out of 4 experiments, each designed to potentially capture a particular phenomenon. Therefore, there isn't enough redundancy to allow for other more extensive cross-validation tests.

3.6.4 Final model

Considering any of the possible operation regimes of Fig. 1.1 (p. 31), the model statement for this CHO-S animal cell line comprising 6 states[1] and 12 parameters can be done as follows:[2]

Reactions:

$$(-\nu_{21})\mathrm{Glc} + (-\nu_{41})\mathrm{Gln} \xrightarrow{\varphi_1} \circlearrowleft\mathrm{Xv} + \nu_{31}\mathrm{Lac} + \nu_{51}\mathrm{Amm} \tag{3.84}$$

$$(-1)\mathrm{Xv} \xrightarrow{\varphi_2} \mathrm{Xd} \tag{3.85}$$

$$(-1)\mathrm{Glc} + (-\nu_{13})\mathrm{Xv} \xrightarrow{\varphi_3} \nu_{13}\mathrm{Xv} + \nu_{33}\mathrm{Lac} \tag{3.86}$$

$$(-1)\mathrm{Gln} \xrightarrow{\varphi_4} \nu_{54}\mathrm{Amm} \tag{3.87}$$

Mass balance equations are given by:

$$\frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ -\nu_{21} & 0 & -1 & 0 \\ \nu_{31} & 0 & \nu_{33} & 0 \\ -\nu_{41} & 0 & 0 & -1 \\ \nu_{51} & 0 & 0 & \nu_{54} \end{bmatrix} \times \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \varphi_3 \\ \varphi_4 \end{bmatrix} + \begin{bmatrix} -x_1 & x_1 \\ (-x_2 + k_1) & 0 \\ -x_3 & 0 \\ (-x_4 + k_2) & 0 \\ -x_5 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \tag{3.88}$$

$$\frac{dx_6}{dt} = \begin{cases} 0 & ,batch \\ u_{10} & ,fedbatch \\ u_{10} - u_{30} & ,continuous \\ u_{10} - u_{30} - u_{20} & ,perfused \end{cases} \tag{3.89}$$

[1] For example, eliminating the only experiment with low glutamine from the training set would very likely prevent the identification algorithm to capture this phenomenon at all.

[1] Only 5 if volume is a process constant. For example: batch mode or constant volume continuous perfused mode.

[2] Parameter nomenclature has been reordered for clarity's sake.

with the following notation:
$\nu_{21} = 1/\theta_8$; $\nu_{31} = \theta_9/\theta_8$; $\nu_{33} = \theta_9$; $\nu_{41} = 1/\theta_{11}$; $\nu_{51} = \theta_{12}/\theta_{11}$;
$k_1 = Glc^{IN}$; $k_2 = Gln^{IN}$; $u_1 = D = F^{IN}/V$; $u_2 = D_{perf} = F_{perf}/V$; $u_{10} = F^{IN}$; $u_{20} = F_{perf}$; $u_{30} = F_{bleed}$.

Table 3.12: Model states.

State	Variable	Units
$x_1 = Xv$	viable biomass concentration	$10^9 cells/L$
$x_2 = Glc$	glucose concentration	mM
$x_3 = Lac$	lactate concentration	mM
$x_4 = Gln$	glutamine concentration	mM
$x_5 = Amm$	ammonia concentration	mM
$x_6 = V$	volume	L

The reaction rates are given by:

$$\begin{cases} \varphi_1 = \mu_1 x_1 \\ \varphi_2 = \mu_2 x_1 \\ \varphi_3 = \mu_3 x_1 \\ \varphi_4 = \mu_4 x_4 \end{cases} \tag{3.90}$$

$$\text{cell growth: } \mu_1 = \theta_1 \frac{x_2}{\theta_2 + x_2} \frac{x_4}{\theta_3 + x_4} \frac{\theta_5}{\theta_5 + x_3} \frac{\theta_6}{\theta_6 + x_5} \tag{3.91}$$

$$\text{cell death: } \mu_2 = \theta_4 \tag{3.92}$$

$$\text{cell maintenance: } \mu_3 = \theta_{10} \tag{3.93}$$

$$\text{spontaneous glutamine degradation: } \mu_4 = \theta_7 \tag{3.94}$$

Parameter values are as follows:

Table 3.13: Model parameter values.

Parameter	Value	Units
$\theta_1 = \mu_{max}$	1.51	d^{-1}
$\theta_2 = k_{Glc}$	0.003	mM
$\theta_3 = k_{Gln}$	0.92	mM
$\theta_4 = \mu_{d,max}$	0.522	d^{-1}
$\theta_5 = k_{d,Lac}$	17.2	mM
$\theta_6 = k_{d,Amm}$	633	mM
$\theta_7 = k_{\alpha,Gln}$	0.184	d^{-1}
$\theta_8 = Y_{Xv/Glc}$	0.107	$10^9 cell\ mmol^{-1}$
$\theta_9 = Y_{Lac/Glc}$	1.58	$mmol\ mmol^{-1}$
$\theta_{10} = m_{Glc}$	0.16	$mmol\ (10^9 cell)^{-1}\ d^{-1}$
$\theta_{11} = Y_{Xv/Gln}$	6.42	$10^9 cell\ mmol^{-1}$
$\theta_{12} = Y_{Amm/Gln}$	0.80	$mmol\ mmol^{-1}$

3.6.5 Summary

To summarise, this case study started out with experimental campaigns where data was collected: concentrations of biomass, substrates and products during cultures of a CHO cell line. The initial experimental conditions were planned envisioning that relevant phenomena would take place (substrate limitation, metabolite inhibition). The truth is, it is only once the experiment is complete and samples have been measured that it can be verified whether these phenomena actually happened and to what extent (for instance, glutamine limitation phenomena happened below expectations for this CHO cell line at the provided initial conditions). This model is further found to be globally identiafiable using the DAISY (Differential Algebra for Identifiability of SYstems) software tool (Bellu et al, 2007; Saccomani et al, 2003, 2010).

A step-by-step identification procedure was proposed. It allows a better alternative to the standard strategy of picking up a complex model and trying to identify all parameters at once, not knowing very well with which initial values to start. Alternatively, the proposed procedure starts off with a very simple model, valid for a limited amount of data, but for which one global parameter can be estimated rather well. Subsequent steps use other data subsets and increasingly complex models, taking advantage of what was already previously identified by using this input in the initial condition. At

this moment, a few parameters have already been estimated. In a final step, several more comprehensive models (complex enough to describe all states and all culture phases) are tried.

The parameters identified in the early phases are the maximum growth and death rate and the stroichiometric coefficients. These are high-impact parameters (ie, if we change their value slightly, a substantial change will be observed in the evolution of concentrations). At the moment when a full comprehensive model is trialled, the new parameters that are added are now of lesser impact and this results in bigger confidence intervals.

A form of improving their identification is (now that the model has been identified) to compute another experimental planning (ie. the initial culture conditions) so that the sensitivity of these lesser-impact parameters will be bigger. This can be done, for instance, by solving an optimisation problem to find out the best initial concentrations of glucose and glutamine in the feeding medium (within admissible bounds), the length t_k of the experiment, and eventually a rate of feeding (fed-batch mode). The new experiments have then to be performed in the laboratory and new samples measured so that the full model can be re-identified (using as initial guess the values already obtained) with this new data that will hopefully[1] be more informative regarding the lesser-impact parameters.

3.7 Step-by-step identification: CHO-320 cells

3.7.1 Materials and methods

Some of the many possible criteria in optimising mammalian cell culture operation (and possibly controlling it) deal with the production of an eventual bioproduct of pharmaceutical interest. Therefore, in order to develop a model where such a product is present, an additional data bank was set up using CHO-320, a clone of CHO-K1 cells genetically modified to secrete a recombinant protein: human interferon-gamma (IFN-γ), used therapeutically for its non specific antiviral activity. This cell line has been studied recently by (Zamorano, 2012) and (Provost, 2006), namely.

[1] That is, hoping that these lesser-impact parameters have been sufficiently well estimated so that, while optimising the new experimental planning, we will find out better means of seeing them have a bigger impact on model states. If however, they have not been properly estimated, then there is no guarantee that the new experimental planning is so optimal after all.

The CHO-320 cells used to build this data bank were kindly provided by Professor Annie Marc (École Nationale Supérieure d'Agronomie et des Industries Alimentaires, Institut National Polytechnique de Lorraine (ENSAIA-INPL), Université de Lorraine, Nancy, France) and its protocol by Doctor Frédérique Balandras (ENSAIA-INPL).

The experiments were again carried out at the University of Mons in serum-free medium: chemically defined glucose free PowerCHO supplemented with glucose and glutamine so as to obtain a data bank with different initial concentrations and expected phenomena (similarly to the CHO-S cultures described on page 148):

Table 3.14: Experimental planning for CHO-320.

Code	B1	B2	B3	B4
Glc_0 (mM)	33	25	8	16
Gln_0 (mM)	9	4	9	3

Again, each experiment was performed in triplicate and sampling took place as described in Table 3.6 on page 100.

The twelve $100mL$ culture flasks each with $80mL$ of medium were inoculated with a cell density around 0.2 to $0.3 \times 10^9 cell/L$ and kept in an incubator at $37°C$ under 5% of CO_2.

Each sample was measured for biomass, glucose, lactate, glutamine and ammonia concentrations following the same procedure as in the CHO-S case. From each $1mL$ sample, 20 μL was used for viability and cell density counting. The remaining sample was filtrated and then assayed: 800 μL with enzymatic kits and 500 μL in the estimation of interferon presence by ELISA testing (IFN-γ Human Direct ELISA Kit, Life Technologies, ref. KAC1231) at $492nm$.

3.7.2 Data bank

The data bank built from cultures performed with CHO-320 cells is plotted in Figures 3.55 and 3.56.

If the intrinsic variability of the triplicate flasks in each experiment is considered, the average values can be plotted with the following confidence intervals (Figures 3.57 to 3.58).

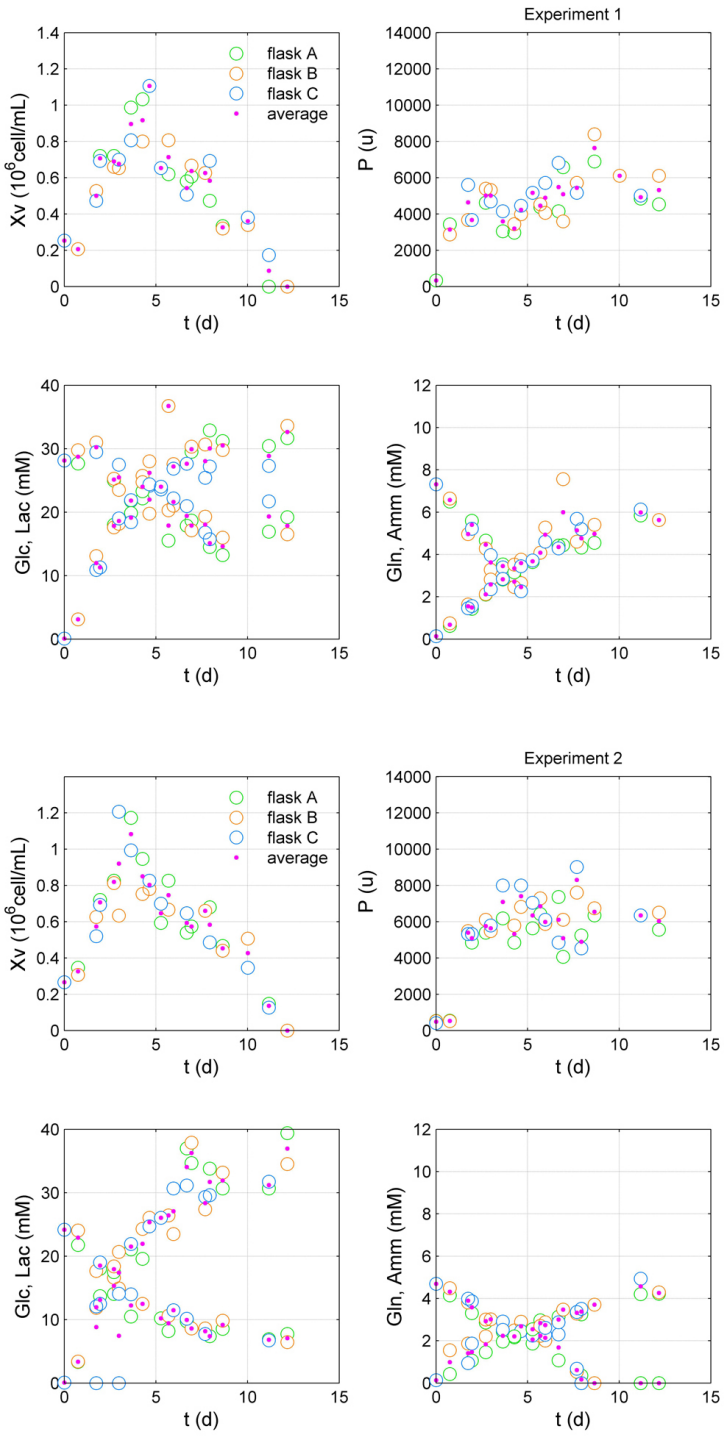


Figure 3.55: CHO-320 cultures: experiment 1 and experiment 2, each with triplicate flasks A,B,C (green, orange, blue). Averages are plotted in pink.

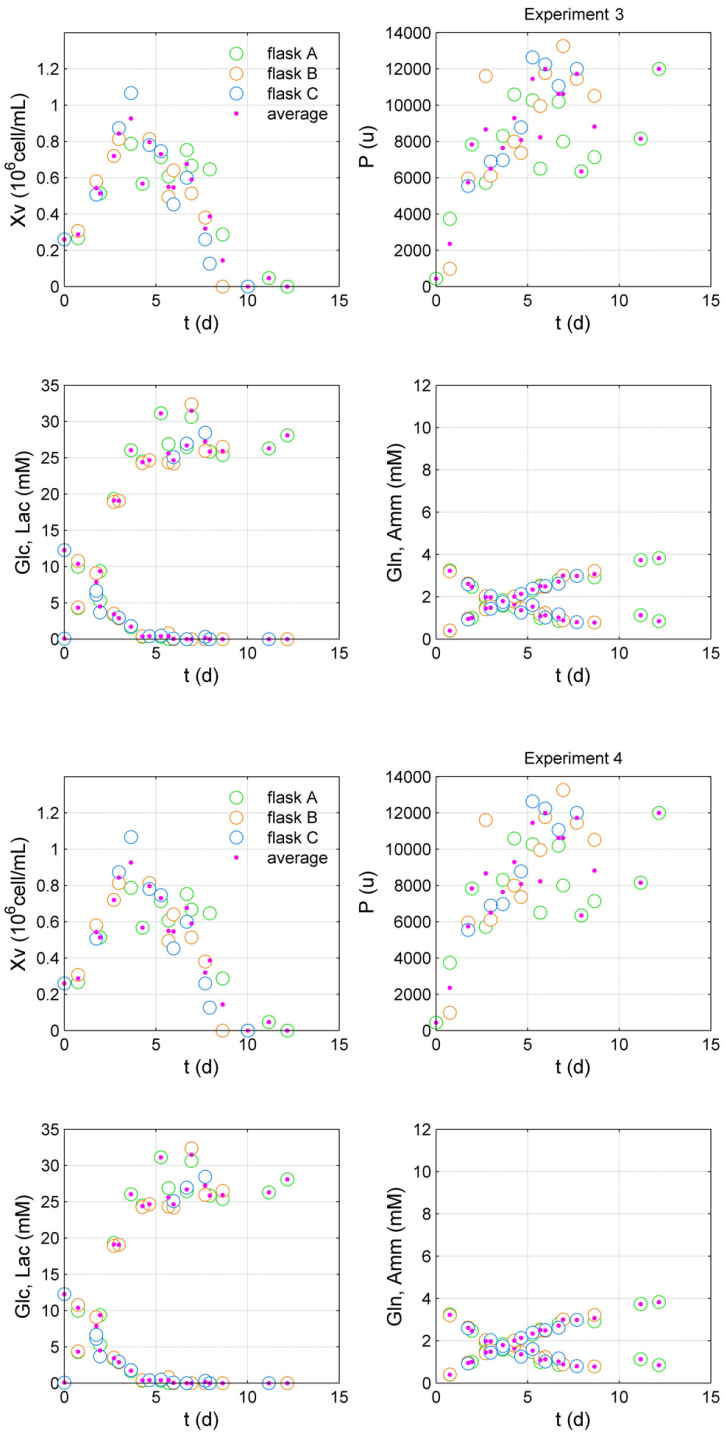


Figure 3.56: CHO-320 cultures: experiment 3 and experiment 4, each with triplicate flasks A,B,C (green, orange, blue). Averages are plotted in pink.

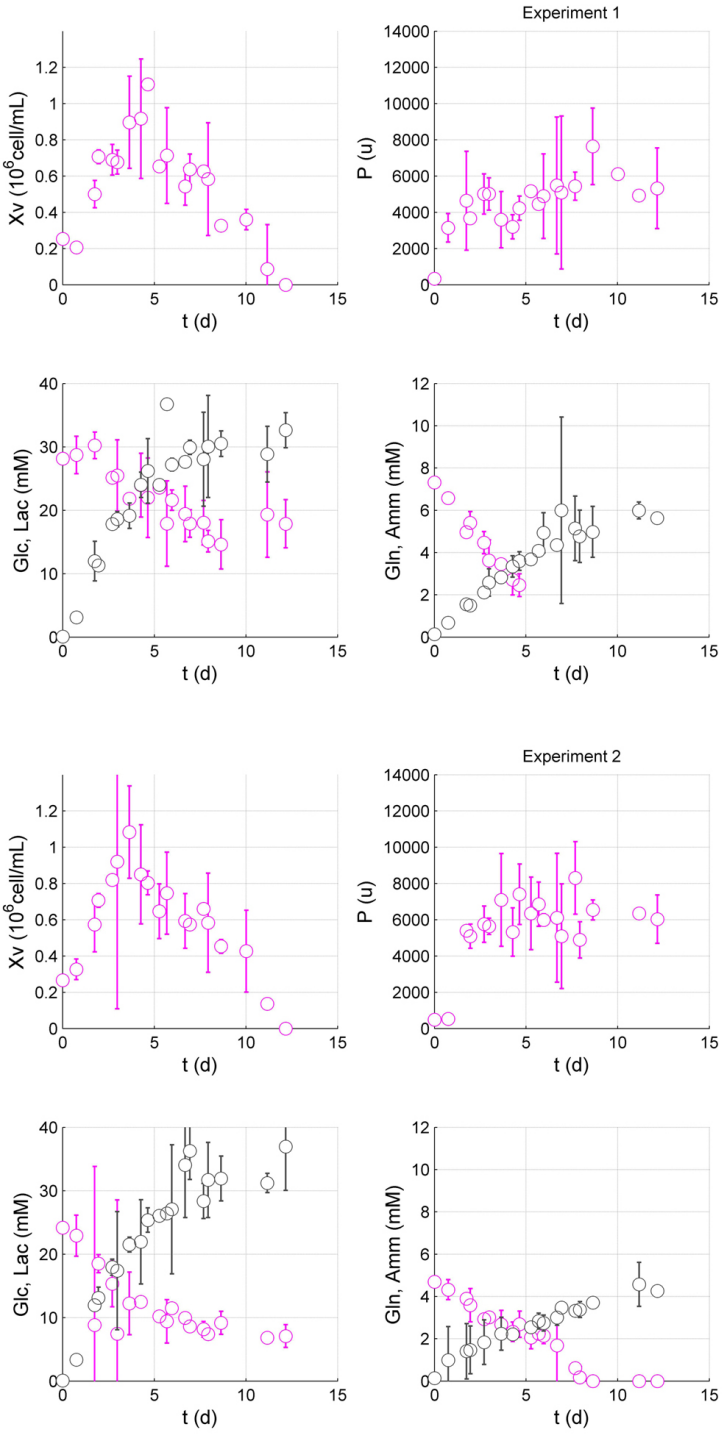


Figure 3.57: CHO-320 cultures: experiment 1 and experiment 2, average values for triplicate flasks with confidence intervals (2σ). Metabolites lactate and ammonia are plotted in grey.

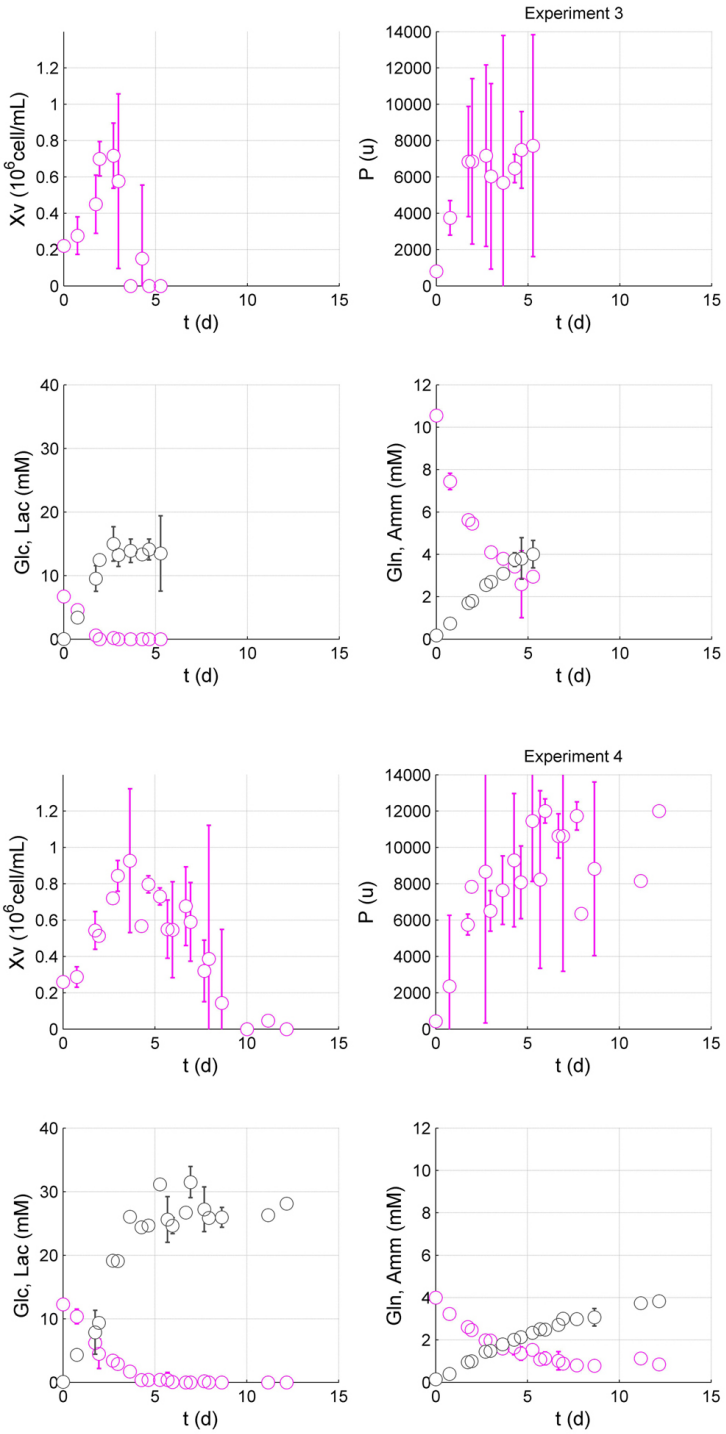


Figure 3.58: CHO-320 cultures: experiment 3 and experiment 4, average values for triplicate flasks with confidence intervals (2σ). Metabolites lactate and ammonia are plotted in grey.

It is flagrant that a lot more variability is present in the CHO-320 data when compared to the CHO-S case. This may be linked to the fact that, for logistic reasons, samples were frozen for a longer period before being assayed. Despite this fact, we will still try to identify a model for this cell line that expresses a therapeutic biological medicine: the IFN-γ protein.

3.7.3 Model identification

The goal is now to identify a model that captures the variety of the response behaviour of this cell line under different conditions (represented by the 4 different experiments). We have applied again the step-by-step identification procedure previously presented in Section 3.6, starting off with simple reduced models and gradually increasing the complexity.

Tables 3.15 and 3.16 summarise results for the sequence that led to a final model describing all states except for IFN-γ concentration (since it is presumed that it has no influence on other states). The equations considered are the same as in the previous CHO-S case study, namelly: Ω_1 (eq. (3.33), p. 103), Ω_2 (eqs. (3.36), p. 106), Ω_3 (eqs. (3.39), p. 109), Ω_{4d} (eqs. (3.52), p. 115), Ω_{5e} (eqs. (3.69), p. 133), Ω_{5f} (eqs. (3.83), p. 141), Ω_{5g} (eqs. (3.70), p. 135), Ω_{5h} (eqs. (3.70), p. 135).

This step-by-step step identification procedure has, thus, allowed us to choose the best promising model and identify its parameter values, which are shown in the following equations:

$$\text{Model } \Omega_{5f}: \frac{dx_1}{dt} = \left(\overbrace{\theta_1 \frac{x_2}{\theta_2 + x_2} \frac{x_4}{\theta_3 + x_4} \frac{\theta_5}{\theta_5 + x_3} \frac{\theta_6}{\theta_6 + x_5}}^{\mu_{growth}} - \overbrace{\theta_4}^{\mu_{death}} \right) x_1$$

$$\begin{aligned} \frac{dx_2}{dt} &= -\left(\frac{1}{\theta_8}\mu_{growth} + \theta_{10}\right)x_1, & \frac{dx_3}{dt} &= \theta_9\left(\frac{1}{\theta_8}\mu_{growth}\right)x_1, \\ \frac{dx_4}{dt} &= -\left(\frac{1}{\theta_{12}}\mu_{growth} + \theta_7\frac{x_4}{x_1}\right)x_1, & \frac{dx_5}{dt} &= \theta_{13}\left(\frac{1}{\theta_{12}}\mu_{growth} + \theta_7\frac{x_4}{x_1}\right)x_1, \end{aligned} \tag{3.95}$$

Table 3.15: Identification results for steps #1, #2, #3, #4.

Parameter / Step	Ω_1	Ω_2	Ω_3	Ω_{4d}
$\theta_1 \equiv \mu_{max}$	0.364	0.351 (0.364)	0.352 (0.351)	0.581 (0.352)
$\theta_2 \equiv k_{Glc}$				0.0004
$\theta_4 \equiv \mu_{d,max}$				0.352
$\theta_8 \equiv Y_{Xv/Glc}$		0.0474	0.0478 (0.0474)	0.0771 (0.0478)
$\theta_9 \equiv Y_{Lac/Glc}$			1.798	
$\theta_{12} \equiv Y_{Xv/Gln}$		0.219	0.221 (0.219)	0.297 (0.221)
$\theta_{13} \equiv Y_{Amm/Gln}$			0.784	
Cost J	0.06	0.37	0.44	0.83
States	x_1	x_1, x_2, x_4	x_1, x_2, x_3, x_4, x_5	x_1, x_2, x_4
Parameters	θ_1	$\theta_1, \theta_8, \theta_{12}$	$\theta_1, \theta_8, \theta_9, \theta_{12}, \theta_{13}$	$\theta_1, \theta_2, \theta_4, \theta_8, \theta_{12}$
Phases	A	A	A	$A + B$

These results suggest that the experiments performed were overall more informative with respect to model parameters than in the CHO-S experimental case study (if one compares Table 3.17 to Table 3.11 on p. 143). For this CHO-320 case study, parameter $\theta_{10} \equiv m_{Glc}$ appears to be less influential (and worse identified) and could, therefore, eventually be discarded.

One way of visually capturing the step-by-step procedure followed is illustrated in Figures 3.59-3.62: early identification steps with reduced order models are shown in light colours (grey, yellow and green, which superpose for phase A, and light blue for phase A+B) and the final model is shown in dark blue.

Table 3.16: Identification results for step #5.

Parameter / Step	Ω_{5e}	Ω_{5f}	Ω_{5g}	Ω_{5h}
$\theta_1 \equiv \mu_{max}$	1.69 (0.581)	1.60 (1.69)	1.176 (0.581)	1.121 (1.176)
$\theta_2 \equiv k_{Glc}$	0.0619 (0.0004)	0.0196 (0.0619)	0.0004 (0.0004)	0.0004 (0.0004)
$\theta_3 \equiv k_{Gln}$	0.5387	0.6536 (0.5387)	0.9337	0.9313 (0.9337)
$\theta_4 \equiv \mu_{d,max}$	0.501 (0.352)	0.531 (0.501)	1.052 (0.352)	1.114 (1.052)
$\theta_5 \equiv k_{d,Lac}$	10.6	16.3 (10.6)	0.6338	0.568 (0.6338)
$\theta_6 \equiv k_{d,Amm}$	13144	28.0 (13144)	0.333	0.516 (0.333)
$\theta_7 \equiv k_{\alpha,Gln}$		0.186		0.037
$\theta_8 \equiv Y_{Xv/Glc}$	0.1676 (0.0478)	0.1753 (0.1676)	0.2230 (0.0478)	0.215 (0.2230)
$\theta_9 \equiv Y_{Lac/Glc}$	2.149 (1.798)	2.067 (2.149)	2.060 (1.798)	1.966 (2.060)
$\theta_{10} \equiv m_{Glc}$		0.005		0.441
$\theta_{12} \equiv Y_{Xv/Gln}$	0.624 (0.221)	5.094 (0.624)	0.790 (0.221)	0.788 (0.790)
$\theta_{13} \equiv Y_{Amm/Gln}$	0.878 (0.784)	0.717 (0.878)	0.834 (0.784)	0.878 (0.834)
Cost J	39.9	22.8	50.1	29.7
States	x_1, x_2, x_3, x_4, x_5	x_1, x_2, x_3, x_4, x_5	x_1, x_2, x_3, x_4, x_5	x_1, x_2, x_3, x_4, x_5
Parameters	$\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_8, \theta_9, \theta_{12}, \theta_{13}$	$\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8, \theta_9, \theta_{10}, \theta_{12}, \theta_{13}$	$\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_8, \theta_9, \theta_{12}, \theta_{13}$	$\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8, \theta_9, \theta_{10}, \theta_{12}, \theta_{13}$
Phases	$A + B$	$A + B$	$A + B$	$A + B$

Table 3.17: Confidence intervals for model Ω_{5f}.

Parameter	LB	Identified Value	UB	Max. Rel Error
$\theta_1 \equiv \mu_{max}$	1.41	1.60	1.79	0.12
$\theta_2 \equiv k_{Glc}$	0	0.020	0.373	18
$\theta_3 \equiv k_{Gln}$	0	0.65	1.58	1.42
$\theta_4 \equiv \mu_{d,max}$	0.380	0.531	0.681	0.28
$\theta_5 \equiv k_{d,Lac}$	0	16.3	33.7	1.07
$\theta_6 \equiv k_{d,Amm}$	0	28.0	159	4.67
$\theta_7 \equiv k_{\alpha,Gln}$	0.163	0.186	0.208	0.12
$\theta_8 \equiv Y_{Xv/Glc}$	0.120	0.175	0.231	0.32
$\theta_9 \equiv Y_{Lac/Glc}$	1.42	2.07	2.71	0.31
$\theta_{10} \equiv m_{Glc}$	0	0.005	1.15	186
$\theta_{12} \equiv Y_{Xv/Gln}$	0	5.09	12.0	1.36
$\theta_{13} \equiv Y_{Amm/Gln}$	0.618	0.717	0.816	0.14

Note: values for a 95% confidence level supposing a t-student distribution: $\theta_{LB/UB} = \hat{\theta} \pm t_{95\%} \sqrt{\sigma^2}$ where σ^2 is computed from the inverse of the Fisher matrix, all experiments considered, as defined in equation (3.32) on p. 103, or more explicitly in eq. (3.79)-(3.80), p. 140.

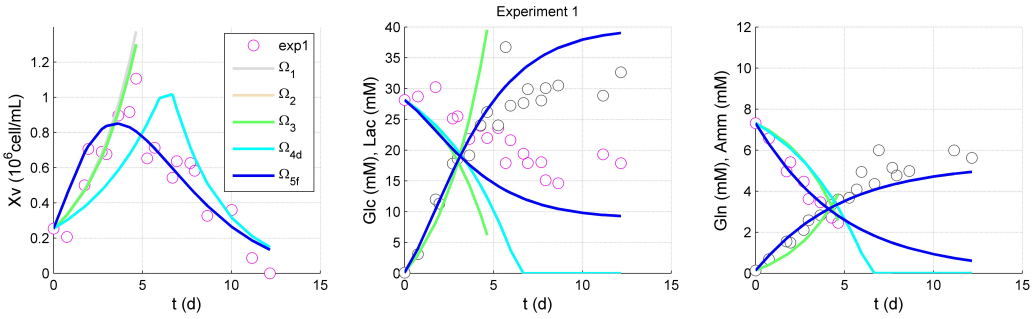


Figure 3.59: Experiment 1, CHO-320: sequence of identified models, from Ω_1 leading to Ω_{5f}.

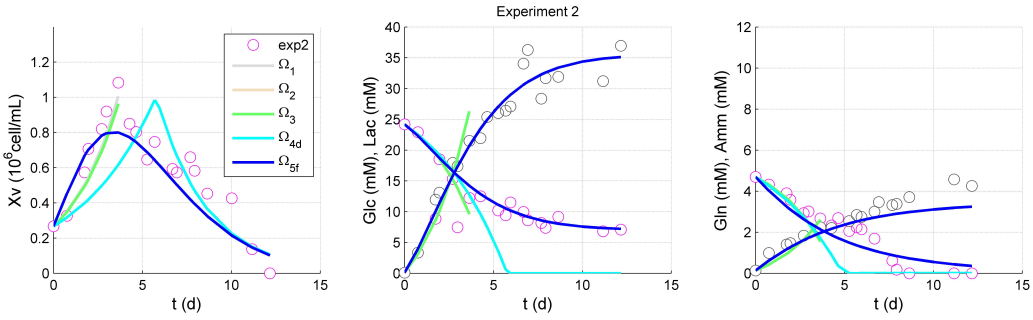


Figure 3.60: Experiment 2, CHO-320: sequence of identified models, from Ω_1 leading to Ω_{5f}.

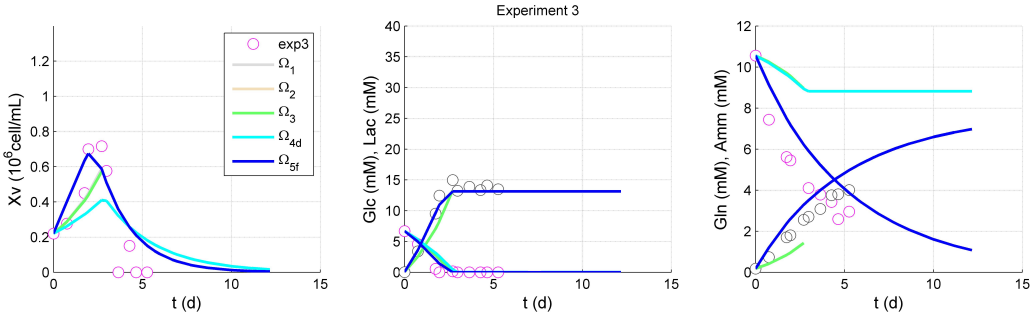


Figure 3.61: Experiment 3, CHO-320: sequence of identified models, from Ω_1 leading to Ω_{5f}.

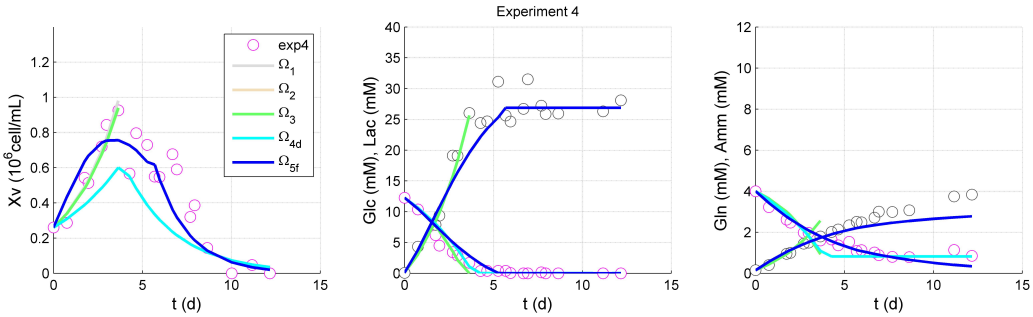


Figure 3.62: Experiment 4, CHO-320: sequence of identified models, from Ω_1 leading to Ω_{5f}.

The next challenge is to incorporate in the model also the concentration of the biologic, target protein IFN-γ, that these cells have been transfected to express. We have tried out several typical forms from the literature (see Table 2.5, p. 56) plus an extra one, all of them dynamic equations that were coupled to those of model Ω_{5f} in order to be solved:

- $\mathcal{H}_{6a}$: protein production is growth-associated (it is produced mostly during phase A, of cell growth) with rate θ_{15}.

$$\Omega_{6a} : \begin{cases} \dfrac{dx_6}{dt} = \left(\theta_{15}\mu_{growth}\right) x_1 \\ \dfrac{dProt}{dt} = \left(\alpha\mu_{growth}\right) Xv \end{cases} \tag{3.96}$$

Identified (cost $J = 0.3762$):

$$\frac{dx_6}{dt} = \left(2632\mu_{growth}\right) x_1 \tag{3.97}$$

- $\mathcal{H}_{6b}$: protein production is partially growth-associated (rate θ_{15}) and partially not (rate θ_{14}).

$$\Omega_{6b} : \begin{cases} \dfrac{dx_6}{dt} = \left(\theta_{15}\mu_{growth} + \theta_{14}\right) x_1 \\ \dfrac{dProt}{dt} = \left(\alpha\mu_{growth} + \beta\right) Xv \end{cases} \tag{3.98}$$

Identified (cost $J = 0.3762$):

$$\frac{dx_6}{dt} = \left(2632\mu_{growth} + \cancelto{\simeq 0}{10^{-13}}\right) x_1 \equiv \Omega_{6a} \tag{3.99}$$

- $\mathcal{H}_{6c}$: (de Tremblay et al, 1992)'s model form, where a factor θ_{16} is presumed to affect the growth-related production of the protein.

$$\Omega_{6c} : \begin{cases} \dfrac{dx_6}{dt} = \left(\dfrac{\theta_{15}}{\theta_{16} + \mu_{growth}}\mu_{growth} + \theta_{14}\right) x_1 \\ \dfrac{dProt}{dt} = \left(\dfrac{\alpha}{k_\mu + \mu_{growth}}\mu_{growth} + \beta\right) Xv \end{cases} \tag{3.100}$$

Identified (cost $J = 0.3762$):

$$\frac{dx_6}{dt} = \left(\frac{1.7977 \times 10^{308}}{6.8308 \times 10^{304} + \mu_{growth}}\mu_{growth} + \cancelto{\simeq 0}{10^{-2}}\right) x_1 \rightarrow$$

$$\xrightarrow{\mu_{growth}<<10^{304}} \simeq \left(2632\mu_{growth}\right) x_1 \equiv \Omega_{6a} \tag{3.101}$$

- $\mathcal{H}_{6d}$: protein production occurs mainly in the death phase.

$$\Omega_{6d} : \begin{cases} \dfrac{dx_6}{dt} = \left(\theta_{15}\mu_{death} + \theta_{14}\right) x_1 \\ \dfrac{dProt}{dt} = \left(\alpha\mu_{death} + \beta\right) Xv \end{cases} \tag{3.102}$$

Identified (cost $J = 0.4821$):

$$\frac{dx_6}{dt} = \left(2815\mu_{death} + 2.9392\right) x_1 \tag{3.103}$$

- $\mathcal{H}_{6e}$: protein production can occur anytime and be growth-related or death-related.

$$\Omega_{6e} : \begin{cases} \dfrac{dx_6}{dt} = \left(\theta_{14} + \theta_{15}\mu_{growth} + \theta_{16}\mu_{death}\right) x_1 \\ \dfrac{dProt}{dt} = \left(\beta + \alpha\mu_{growth} + k_\mu\mu_{death}\right) Xv \end{cases} \tag{3.104}$$

Identified (cost $J = 0.3762$):

$$\frac{dx_6}{dt} = \left(\cancel{10^{-19}}^{\simeq 0} + 2632\mu_{growth} + \cancel{10^{-13}}^{\simeq 0} \right) x_1 \equiv \Omega_{6a} \tag{3.105}$$

The quality of the protein concentration fit will obviously depend on the quality of the previously identified submodel Ω_{5f} (and particularly the quality of the biomass fit) but, in any case, it is clear for this data bank that protein production seems to be growth associated and that submodel Ω_{6a} describes it best among all submodel structures considered. The corresponding plots are presented in Figures 3.63 to 3.66.

Table 3.18: Confidence intervals for model Ω_{6a}.

Parameter	LB	Identified Value	UB	Max. Rel Error
$\theta_{15} \equiv \alpha$	577	2632	4686	0.78

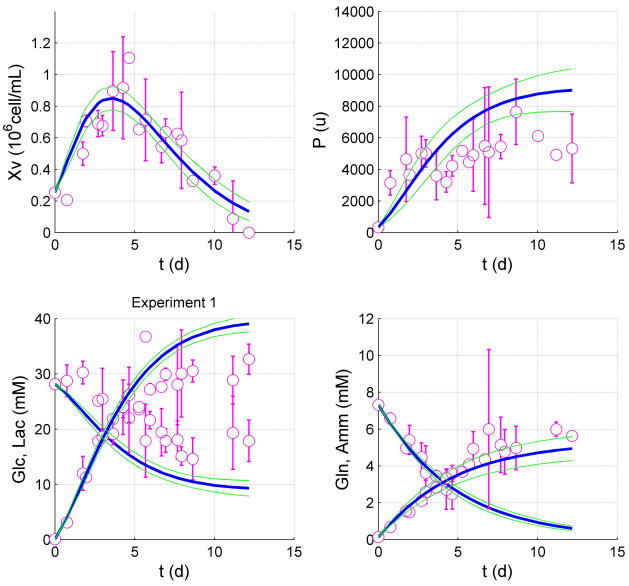


Figure 3.63: Final model: Ω_{5f}+Ω_{6a}, experiment 1, CHO-320: model simulation (blue), 95%-confidence interval for states (green), measurements (circles) and their variability (error bars, 2σ) .

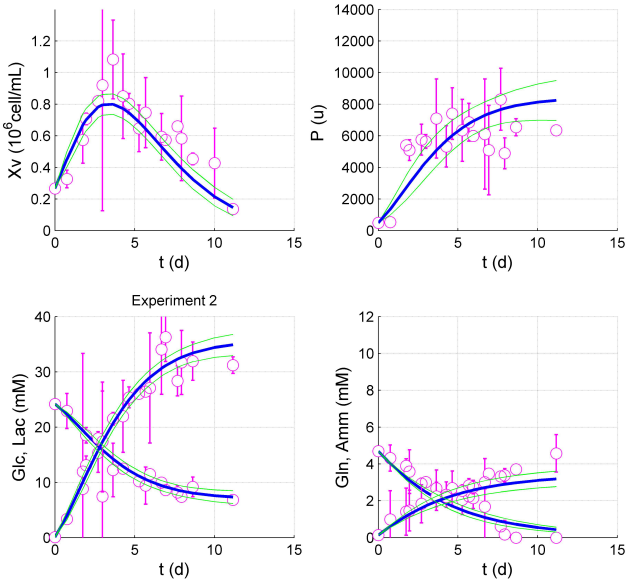


Figure 3.64: Final model: Ω_{5f}+Ω_{6a}, experiment 2, CHO-320: model simulation (blue), 95%-confidence interval for states (green), measurements (circles) and their variability (error bars, 2σ) .

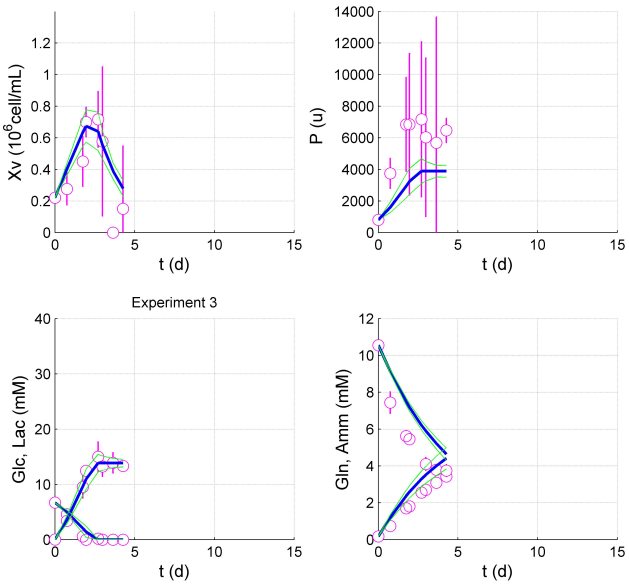


Figure 3.65: Final model: Ω_{5f}+Ω_{6a}, experiment 3, CHO-320: model simulation (blue), 95%-confidence interval for states (green), measurements (circles) and their variability (error bars, 2σ) .

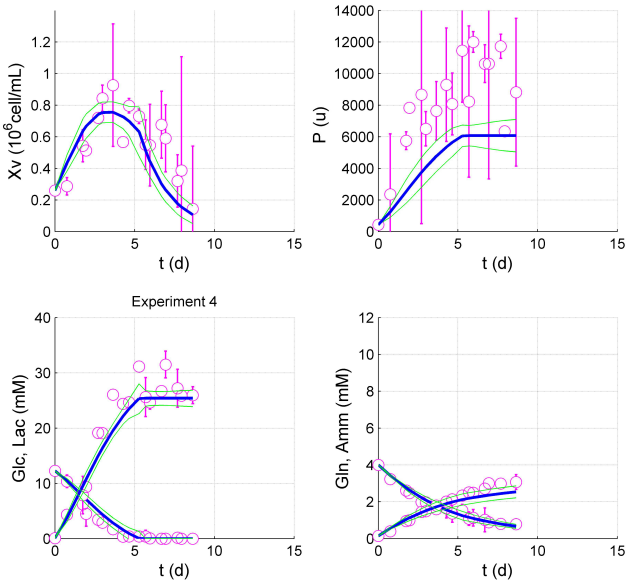


Figure 3.66: Final model: Ω_{5f}+Ω_{6a}, experiment 4, CHO-320: model simulation (blue), 95%-confidence interval for states (green), measurements (circles) and their variability (error bars, 2σ) .

3.7.4 Final model

Overall, considering any operation regime of Fig. 1.1 (p. 31), the model for this CHO-320 cell line producing the INF-γ protein comprises 7 states[1] and 13 parameters. It can be stated as below:[2]

Reactions:

$$(-\nu_{21})\text{Glc} + (-\nu_{41})\text{Gln} \xrightarrow{\varphi_1} \circlearrowleft\text{Xv} + \nu_{31}\text{Lac} + \nu_{51}\text{Amm} + \nu_{61}\text{Prot} \tag{3.106}$$

$$(-1)\text{Xv} \xrightarrow{\varphi_2} \text{Xd} \tag{3.107}$$

$$(-1)\text{Glc} + (-\nu_{13})\text{Xv} \xrightarrow{\varphi_3} \nu_{13}\text{Xv} \tag{3.108}$$

$$(-1)\text{Gln} \xrightarrow{\varphi_4} \nu_{54}\text{Amm} \tag{3.109}$$

Mass balance equations are given by

$$\frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ -\nu_{21} & 0 & -1 & 0 \\ \nu_{31} & 0 & 0 & 0 \\ -\nu_{41} & 0 & 0 & -1 \\ \nu_{51} & 0 & 0 & \nu_{54} \\ \nu_{61} & 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \varphi_3 \\ \varphi_4 \end{bmatrix} + \begin{bmatrix} -x_1 & x_1 \\ (-x_2 + k_1) & 0 \\ -x_3 & 0 \\ (-x_4 + k_2) & 0 \\ -x_5 & 0 \\ -x_6 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \tag{3.110}$$

$$\frac{dx_7}{dt} = \begin{cases} 0 & ,batch \\ u_{10} & ,fedbatch \\ u_{10} - u_{30} & ,continuous \\ u_{10} - u_{30} - u_{20} & ,perfused \end{cases} \tag{3.111}$$

with the following notation:

$\nu_{21} = 1/\theta_8$; $\nu_{31} = \theta_9/\theta_8$; $\nu_{41} = 1/\theta_{11}$; $\nu_{51} = \theta_{12}/\theta_{11}$; $\nu_{54} = \theta_{12}$; $\nu_{61} = \theta_{13}$;

$k_1 = Glc^{IN}$; $k_2 = Gln^{IN}$; $u_1 = D = F^{IN}/V$; $u_2 = D_{perf} = F_{perf}/V$; $u_{10} = F^{IN}$; $u_{20} = F_{perf}$; $u_{30} = F_{bleed}$.

[1] Only 6 if volume is a process constant. For example: batch mode or constant volume continuous perfused mode.

[2] Parameter nomenclature has been reordered for the sake of simplicity.

Table 3.19: Model states.

State	Variable	Units
$x_1 = Xv$	viable biomass concentration	$10^9 cells/L$
$x_2 = Glc$	glucose concentration	mM
$x_3 = Lac$	lactate concentration	mM
$x_4 = Gln$	glutamine concentration	mM
$x_5 = Amm$	ammonia concentration	mM
$x_6 = Prot$	IFN-γ protein concentration	u/L
$x_7 = V$	volume	L

The reaction rates are given by:

$$\varphi_i = \mu_i x_1 \tag{3.112}$$

$$\text{cell growth: } \mu_1 = \theta_1 \frac{x_2}{\theta_2 + x_2} \frac{x_4}{\theta_3 + x_4} \frac{\theta_5}{\theta_5 + x_3} \frac{\theta_6}{\theta_6 + x_5} \tag{3.113}$$

$$\text{cell death: } \mu_2 = \theta_4 \tag{3.114}$$

$$\text{cell maintenance: } \mu_3 = \theta_{10} \tag{3.115}$$

$$\text{spontaneous glutamine degradation: } \mu_4 = \theta_7 \frac{x_4}{x_1} \tag{3.116}$$

Parameter values are as follows:

Table 3.20: Model parameter values.

Parameter	Value	Units
$\theta_1 = \mu_{max}$	1.60	d^{-1}
$\theta_2 = k_{Glc}$	0.020	mM
$\theta_3 = k_{Gln}$	0.65	mM
$\theta_4 = \mu_{d,max}$	0.531	d^{-1}
$\theta_5 = k_{d,Lac}$	16.3	mM
$\theta_6 = k_{d,Amm}$	28.0	mM
$\theta_7 = k_{\alpha,Gln}$	0.186	d^{-1}
$\theta_8 = Y_{Xv/Glc}$	0.175	$10^9 cell\ mmol^{-1}$
$\theta_9 = Y_{Lac/Glc}$	2.07	$mmol\ mmol^{-1}$
$\theta_{10} = m_{Glc}$	0.005	$mmol\ (10^9 cell)^{-1}\ d^{-1}$
$\theta_{11} = Y_{Xv/Gln}$	5.09	$10^9 cell\ mmol^{-1}$
$\theta_{12} = Y_{Amm/Gln}$	0.717	$mmol\ mmol^{-1}$
$\theta_{13} = \alpha$	2632	$u\ (10^9 cell)^{-1}$

3.7.5 Summary

Again, as for the CHO-S case study, the proposed sequential identification procedure was applied, allowing for increasing model complexity. Instead of trying to identify the many parameters of a very complex model all at once (which can be very time-consuming and may not lead to good results), in this step-wise procedure, there is backup information on how to choose some initial guesses and initialize the identification algorithm. This is illustrated in Figure 3.67 where the model identified step-by-step is represented in blue, and the results using the usual approach of trying to identify all parameters at once are represented in grey (several runs of the algorithm)[1]. The step-by-step procedure seems to lead to better and more consistent results.

[1]Not all runs lead to useful results, sometimes the algorithm didn't converge.

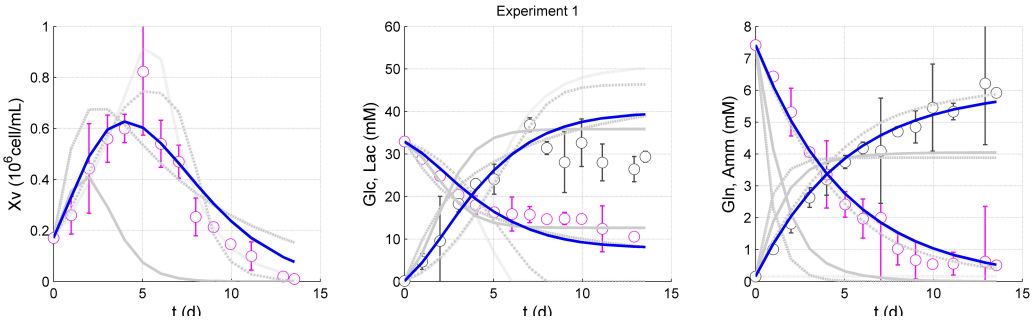


Figure 3.67: CHO-S, experiment 1 (circles). Model $\{\Omega_{5f} + \Omega_{5f,rest}\}$ in blue if identified by the step-by-step procedure and in grey if identified by the all-at-once approach.

The final model chosen for this CHO-320 cell line has a similar structure to that chosen for the CHO-S line. On the whole, it presumes that net cell growth is limited whenever substrate concentration (guclose/glutamine) becomes low and inhibited when some metabolites (lactate/ammonia) build up.

If this model was to be used in a control scheme, then it could allow, for example, a controller to account for these phenomena when computing which manipulations the culture should undergo (eg. more feeding) in order to maintain a certain setpoint (eg. biomass concentration). These aspects will be taken up in more detail in the following chapter.

Chapter 4

Animal cell culture model observability

Once a model is identified and available for use, the question of observability follows. This chapter deals with sofware sensors (observers) that allow to estimate the timecourse of concentrations that cannot be measured in real practice. Assessing if this is possible (studying observability) may not be an easy task. In this chapter, a contribution on a manner of studying the observability of animal cell culture models is presented, based on recent works of Moreno et al (2014) that address the indistinguishability properties of the system.

4.1 Introduction

Models such as those explored in this thesis can be used in the control of the operation of an animal cell culture taking place in a bioreactor. However, a model-based controller needs to have knowledge of the system states (eg. concentrations) in order to compute the best line of action. If eventually some system states are not being measured, they may be observed by mathematical means, using an observer. Generally speaking, observability is a property

dealing with this being possible.

Lets recall that common system states of animal cell culture models are the concentrations of biomass (Xv, viable living cells), substrate glucose and metabolite lactate (Glc/Lac), substrate glutamine and metabolite ammonia (Gln/Amm), and eventually a product of biomedical interest that these cells express such as, for example, monoclonal antibodies (MAb), as shown in Figure 4.1.

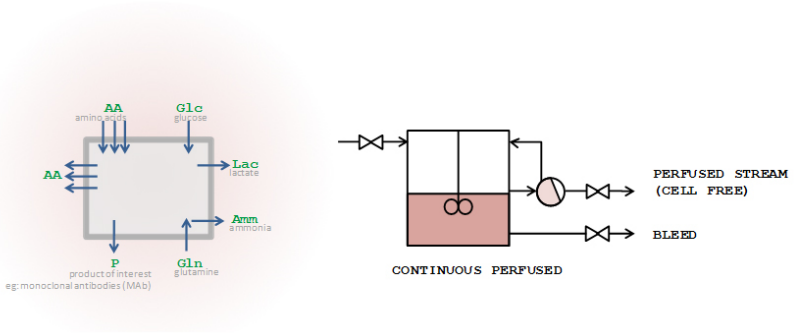


Figure 4.1: Acting components of the animal cell culture system (left). Bioreactor operating in a perfused regime (right).

The production regime of interest for this study is the continuous perfused. In this regime, cultures are intended to last longer since a medium with substrates is continuously fed to the bioreactor, while a workforce of cells is retained by means of a perfusion filter that lets all components out except for cells. This means that the biomolecule of interest can be extracted, purified and stocked on a continuous basis. On the other hand, culture medium inside the reactor is renewed to prevent the accumulation of cell growth inhibiting metabolites (Lac, Amm). Volume is kept constant and a small bleed outflow serves as an additional degree of freedom to adjust biomass concentration and keep the system at steady state.

While several analyses are possible by taking samples of the culture medium and assaying them offline in the laboratory, not many probes exist that could be coupled to the bioreactor and measure these concentrations in a continuous online manner so that this data could be transmitted on to a controller and the controller could then decide how to act on the system.

If the development of control strategies requires the availability of a number of on-line measurements that can be difficult to achieve in practice (availability of the probe, costs, processing time, etc), then the development of

software sensors (observers) is of paramount importance for the implementation of these control strategies. The challenge of this chapter is to study which concentrations can be followed online in order to study and perfect an intelligent production strategy. The big question being asked is: "Is it possible to build an observer that will estimate the concentrations not being measured in the system?".

4.1.1 Probes

Up to present bioanalysers have been used in the online follow-up of concentrations of bioreactor animal cell cultures. Models on the market analyse one or two substrate/metabolite pairs: either *Glc*/*Lac*, or *Gln*/*Amm* or both, transposing the concept of a standard enzymatic reaction kit into an online procedure (eg. YSI (2014)). Bioanalysers present high operation costs (reactants) and low sampling time. The samples removed on a frequent basis also pose sterility concerns as the system is open to a non-sterile exterior environment.[1] The error can also be considerable. In this scenario, components of the macroscopic reaction that are being measured are those highlighted in green below.[2]

$$\mathrm{Glc} + \mathrm{Gln} \xrightarrow{\varphi} \mathrm{Xv} + \mathrm{Lac} + \mathrm{Amm} + \mathrm{MAb} \tag{4.1}$$

More recently, capacitance probes have been put on the market, taking advantage of the dielectric properties of living cells (Fogale-nanotech, 2013; Logan et al, 2011). These probes measure viable living biomass concentration with very low operation costs, an intensive sampling time (1-5min) and little noise and error. They can be operated in a more sterility-friendly manner since they are submergible and can be autoclaved[3] along with the bioreactor without involving any removal of culture samples. The equipment allows thus biomass to be measured online while other components are not, as highlighted below.

$$\mathrm{Glc} + \mathrm{Gln} \xrightarrow{\varphi} \mathrm{Xv} + \mathrm{Lac} + \mathrm{Amm} + \mathrm{MAb} \tag{4.2}$$

[1] Sterility is of utmost importance in the production of biologicals. Regulatory guidelines are provided eg. in ICH Q5 and can be consulted in (EMA, 2014).

[2] Apart from bioanalysers, other alternatives on the market that could eventually be considered are online glucose probes and the NIR.

[3] Autoclaving is a sterilization procedure performed prior to starting the culture by inoculating the bioreactor with animal cells.

4.1.2 Motivation

In this chapter, the opportunity presented by this recent probe is studied, the final goal being to develop an observer that will provide an online follow-up of the culture for an intelligent production strategy. An interesting scenario would be to rely on biomass measurements alone or perhaps on the knowledge of biomass and one or two substrates. Some scenarios will be scrutinised:

Table 4.1: Online follow-up scenarios

Equipment	Measured	To Estimate	Scenario
Bioanalyser	Glc/Lac Gln/Amm	Xv MAb	current practice
Fogale	Xv	Glc/Lac Gln/Amm MAb	very interesting
Fogale + Bioanalyser	Xv Glc	Lac Gln/Amm MAb	interesting (less reactants needed)

For this purpose, we consider in this chapter a nonlinear dynamic model of hybridoma cell cultures producing monoclonal antibodies, initially proposed in de Tremblay et al (1992, 1993), and since then considered in several further works (Aehle et al, 2011; Chen et al, 2002; Franco-Lara and Weuster-Botz, 2005; Nguang et al, 2001; Portner and Schafer, 1996; Roubos et al, 1997, 1999; Sarkar and Modak, 2004).

We first analyse the observability/detectability properties of this model using a method based on a natural dynamical interpretation of the observability/detectability concepts (Moreno et al, 2012, 2014), leading to the description of the indistinguishable dynamics of the system. Following this analysis, a Kalman filter is designed to reconstruct on-line variables which are difficult or expensive to measure directly with a hardware sensor. Various sensor configurations are considered, showing promising results in simulation.

The chapter is organized as follows. The next section introduces observability analysis methods and illustrates them with an oversimplified fictitious

model of an animal cell culture, created here with the sole purpose of helping underpin the conceptual rationale for the method chosen and helping clear out why other more common methods were discarded. Section 4.3 deals with a real hybridoma cell cultivation model and the observability analysis of this more complex system. A Kalman filter is then designed and tested in simulation in Section 4.4. Section 4.5 is devoted to some conclusions.

4.2 Observability/detectability analysis

As it is well-known, the possibility of constructing an observer is tied to the observability/detectability properties of the system's model. When only the initial conditions are unknown, observability corresponds to the (theoretical) possibility of estimating the state in a finite time-horizon, whereas if the system is only detectable, the state estimation can only be attained asymptotically.

The observability/detectability analysis proposed in Moreno et al (2012) is based on a natural dynamical interpretation of the observability/detectability concepts, and is introduced in the following subsection using a simple example for the sake of illustration.

4.2.1 Definitions

In order to evaluate scenarios where some concentrations will be measured and others estimated by means of an observer, one should first check the observability properties of the system, which is nonlinear.

A system is observable if any state is distinguishable from any other state in the sense that states produce a recorded output y (the measurements) that is exclusive/unique. More precisely, a nonlinear system is observable if there exists an input u(t) which leads to different output trajectories for different initial conditions.

$$\text{Nonlinear system:} \quad \frac{dx(t)}{dt} = f\Big(x(t), u(t)\Big); \quad x(t_0) = x_0 \tag{4.3}$$

$$\text{Measurement equation:} \quad y(t) = h\Big(x(t)\Big) \tag{4.4}$$

Different measured trajectories for different initial conditions:

$$y(t, x_{0,A}, u) \neq y(t, x_{0,B}, u) \tag{4.5}$$

That is, a system is observable for an input $u(t)$ if different initial conditions $x_{0,A}$, $x_{0,B}$ will lead to distinct trajectories of the measured variable y.

A non-observable example is illustrated in Figure 4.2. Two different initial states produce exactly the same recorded output y: ie, the same set of measurements (green trajectory for biomass) can occur with the light pink trajectories or also in association with the dark pink trajectories.

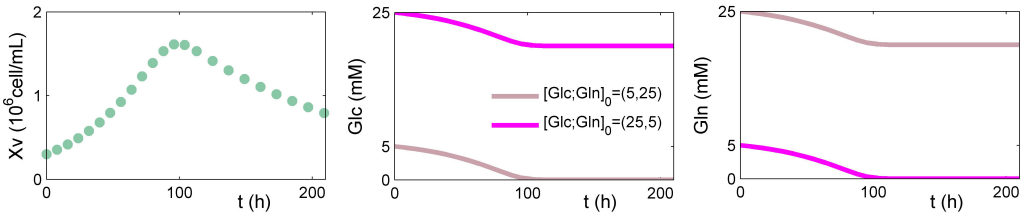


Figure 4.2: Given a certain initial biomass Xv_0, departing from initial condition $[Glc\ Gln]_0^T = [5\ 25]^T$ (light pink) leads to the green trajectory of biomass. Departing from $[Glc\ Gln]_0^T = [25\ 5]^T$ (dark pink) also leads to exactly the same biomass trajectory. Therefore, knowledge of the green trajectory alone is not enough to distinguish which of the two scenarios happened (light pink or dark pink).

4.2.2 Approaches for assessing observability

This subsection intends to overview available approaches of assessing observability that could be useful in the case of animal cell culture models such as those identified in this thesis. Firstly, a very simple fictitious toy model will be used for illustration purposes.

A continuously perfused bioreactor operated at constant volume is considered where a culture of suspended cells takes place. Besides the feeding of substrates, a perfusion filter allows cell retention, whilst a small bleed stream lets all components out.

Only two biological macroreactions comprising substrates glucose (Glc) and glutamine (Gln), living (Xv) and dead (Xd) biomass are considered:

$$(-\nu_{21})\mathrm{Glc} + (-\nu_{31})\mathrm{Gln} \xrightarrow{\varphi_{growth}} \circlearrowleft \mathrm{Xv} \tag{4.6}$$

$$(-1)\text{Xv} \xrightarrow{\varphi_{death}} (1)\text{Xd} \tag{4.7}$$

This leads to a system of mass balance equations where the following symbols represent states, kinetic parameters and inputs: $x_1 = Xv$; $x_2 = Glc$; $x_3 = Gln$; $u_1 = D = F/V$; $u_2 = D_{perf} = F_{perf}/V$; $k_1 = \nu_{21} = 1/Y_{Xv/Glc} = \nu_{31} = 1/Y_{Xv/Gln}$; $k_2 = Glc^{IN}$; $k_3 = Gln^{IN}$; $k_4 = \mu_{max}$; $k_5 = k_{Glc}$; $k_6 = k_{Gln}$; $k_7 = \mu_{d,max}$.

$$\frac{dx_1}{dt} = \left(k_4 \frac{x_2}{k_5 + x_2} \frac{x_3}{k_6 + x_3}\right) x_1 - k_7 x_1 - u_1 x_1 + u_2 x_1 \tag{4.8}$$

$$\frac{dx_2}{dt} = -k_1 \left(k_4 \frac{x_2}{k_5 + x_2} \frac{x_3}{k_6 + x_3}\right) x_1 - u_1 x_2 + k_2 u_1 \tag{4.9}$$

$$\frac{dx_3}{dt} = -k_1 \left(k_4 \frac{x_2}{k_5 + x_2} \frac{x_3}{k_6 + x_3}\right) x_1 - u_1 x_3 + k_3 u_1 \tag{4.10}$$

with values $k = 0.6mM$; $\mu_{max} = 0.0318h^{-1}$; $\nu = 400 \times 10^6 cell \cdot mmol^{-1} = \nu_{21} = \nu_{31}$; $Xv_0 = 0.3 \times 10^6 cell/mL$. This model has already been illustrated in Figure 4.2.

Lets now assess its observability considering the case where biomass is measured and glucose and glutamine have to be estimated. The following approaches are available:

- The rank condition of the nonlinear observability matrix;
- Linearization;
- Canonical forms;
- Indistinguishable dynamics;

The practicality of the use of these approaches in the scope of animal cell cultures will also be evaluated. In fact, the rank condition will prove to be of complex application; the linearization will only allow to conclude that the original nonlinear system is observable but not that it is unobservable (plus, for systems with more than 3 states, expressions become very complex); the canonical form will not be of use in this scenario; a method presented in Moreno et al (2012), on the other hand, will prove itself useful. This can be particularly helpful for animal cell culture models of a certain complexity such as the ones identified in the previous Chapter 3: models that can have 5 to 7

states and around 15 parameters describing cell growth and death. In fact, the observability of such systems is not frequently studied in the literature. Only very simple cases are, usually for systems up to 3 states, or with elementary models valid for the growth phase alone and generally meant for bacterial cultures, or in case studies where many states are measured and few need estimation[1].

4.2.2.1 Rank condition

This general method involves computing the observability map of the system and then the observability matrix O. For example, in batch mode:

$$\text{System:} \quad \frac{d\xi_i}{dt} = \nu_{ij}\varphi_j \Rightarrow \frac{d}{dt}\begin{bmatrix} Xv \\ Glc \\ Gln \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -\nu_{21} & 0 \\ -\nu_{31} & 0 \end{bmatrix} \times \begin{bmatrix} \varphi_{growth} \\ \varphi_{death} \end{bmatrix} =$$

$$= \begin{bmatrix} \left(\mu_{max}\frac{Glc}{k+Glc}\frac{Gln}{k+Gln}\right)Xv & -\mu_{d,max}Xv \\ -\nu\left(\mu_{max}\frac{Glc}{k+Glc}\frac{Gln}{k+Gln}\right)Xv & \\ -\nu\left(\mu_{max}\frac{Glc}{k+Glc}\frac{Gln}{k+Gln}\right)Xv & \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix} \tag{4.11}$$

$$\text{If only Xv is measured the output is:} \quad y = [h_1(x)] = x_1 \tag{4.12}$$

$$n = dim\left\{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right\} = 3 \tag{4.13}$$

$$\text{Observability map:} \quad q_1 = \begin{bmatrix} q_{1,1} \\ q_{1,2} \\ q_{1,3} \end{bmatrix} = \begin{bmatrix} L_f^0 h_1 \\ L_f^1 h_1 \\ L_f^2 h_1 \end{bmatrix}$$

$$\text{in this case } q_1 = \begin{bmatrix} x_1 \\ (1\times f_1 + 0\times f_2 + 0\times f_3) \\ \frac{\partial(f_1)}{\partial x_1}f_1 + \frac{\partial(f_1)}{\partial x_2}f_2 + \frac{\partial(f_1)}{\partial x_3}f_3 \end{bmatrix} \tag{4.14}$$

[1]Some examples are provided in Dewasme et al (2011); Dochain and Chen (1992); Doyle and Henson (1996); Goffaux (2010); Haag (2003); Hulhoven (2006); Veloso et al (2008).

$$\text{Observability matrix:} \quad O = \frac{\partial q}{\partial x} = \begin{bmatrix} 1 & 0 & 0 \\ \frac{\partial q_{1,2}}{\partial x_1} & \frac{\partial q_{1,2}}{\partial x_2} & \frac{\partial q_{1,2}}{\partial x_3} \\ \frac{\partial q_{1,3}}{\partial x_1} & \frac{\partial q_{1,3}}{\partial x_2} & \frac{\partial q_{1,3}}{\partial x_3} \end{bmatrix} \tag{4.15}$$

If O is full column rank, then the system is locally observable at x. The system is globally observable if a partition of q is injective for all x (however, this analytical solution is often difficult if not impossible to obtain).

$$rank(O) = rank \begin{bmatrix} 1 & 0 & 0 \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \stackrel{?}{=} dim(x) \tag{4.16}$$

For this simple toy model with 3 states only, the equations already tend to be quite difficult to manipulate analytically, as shown in Appendix B (p. 239-242) where they are developed. Global conclusions may be hard to prove, but a local analysis could be performed.

It is clear that for systems of dimension bigger than 2 (the case of animal cell culture models considered in this thesis), analytical manipulations become rapidly more complex.

4.2.2.2 Linearization

This method involves linearizing a system around a reference state ξ^*.

$$\text{Nonlinear system:} \quad \frac{d\xi(t)}{dt} = v\varphi - D\xi + F = f(\xi, D, F) \tag{4.17}$$

$$\text{Linearized system:} \quad \frac{dx}{dt} = A(\xi^*)x + B(\xi^*)u \quad \text{with:}$$

$$\begin{cases} A(\xi^*) = \left[\frac{\partial f}{\partial \xi}\right]_{\xi^*} = v\left[\frac{\partial \varphi}{\partial \xi}\right]_{\xi^*} - D^* I_n \\ B(\xi^*) = \left[\frac{\partial f}{\partial u}\right]_{\xi^*} \quad \text{and } x = \xi - \xi^* \end{cases} \tag{4.18}$$

The linear stationary system is observable if and only if $rank(O) = dim(x)$. This test is a sufficient local observability condition for the original nonlinear system: if the linearized model is observable at $\xi = \xi^*$, then the nonlinear

system is observable around this point (Dochain and Chen, 1992). This method cannot be used, however, to prove global non-observability.

$$y = Cx = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} Xv & Glc & Gln \end{bmatrix}^T = Xv \tag{4.19}$$

$$O = \begin{bmatrix} C \\ CA \\ CA^2 \end{bmatrix} \tag{4.20}$$

Once more, it is clear that the complexity of analytically analysing the rank for culture models with more than 3 states is considerable. This makes global conclusions hard to achieve, but a local analysis is a quite common use application.

4.2.2.3 Canonical forms

Some methods for the observability analysis rely on canonical forms. For instance, the invertibility of the map is guaranteed if, among others, the system can be put into a Lower Hessenberg triangular form (Gauthier and Kupka, 1994; Zeitz, 1984). This has, for example, been applied to bacterial cultures in Dewasme et al (2012). However, it is not always applicable and this is the case of the higher order models used in the scope of this thesis since it applies only to situations where the measurement partition is bigger or equal in size to the following partitions of states and this is not possible here for two reasons: we intend to measure few variables and estimate many, on one hand, and, on the other hand, the measurements (biomass) have dynamics that depend on too many states (two substrates and two metabolites on top of biomass).

4.2.2.4 Indistinguishable dynamics

As shown in Moreno et al (2014), a dynamical interpretation of the concepts of observability/detectability can be obtained considering the system and a copy of it. For this model, lets consider state variables x_i, $i = 1, 2, 3$ and copy variables z_i, $i = 1, 2, 3$. For a generic operation regime:

$$\text{System:} \quad \begin{cases} \frac{dx_1}{dt} = \left(k_4 \frac{x_2}{k_5+x_2} \frac{x_3}{k_6+x_3}\right) x_1 - k_7 x_1 - u_1 x_1 + u_2 x_1 \\ \frac{dx_2}{dt} = -k_1 \left(k_4 \frac{x_2}{k_5+x_2} \frac{x_3}{k_6+x_3}\right) x_1 - u_1 x_2 + u_1 k_2 \\ \frac{dx_3}{dt} = -k_1 \left(k_4 \frac{x_2}{k_5+x_2} \frac{x_3}{k_6+x_3}\right) x_1 - u_1 x_3 + u_1 k_3 \end{cases} \tag{4.21}$$

$$\text{Copy:} \quad \begin{cases} \frac{dz_1}{dt} = \left(k_4 \frac{z_2}{k_5+z_2} \frac{z_3}{k_6+z_3}\right) z_1 - k_7 z_1 - u_1 z_1 + u_2 z_1 \\ \frac{dz_2}{dt} = -k_1 \left(k_4 \frac{z_2}{k_5+z_2} \frac{z_3}{k_6+z_3}\right) z_1 - u_1 z_2 + u_1 k_2 \\ \frac{dz_3}{dt} = -k_1 \left(k_4 \frac{z_2}{k_5+z_2} \frac{z_3}{k_6+z_3}\right) z_1 - u_1 z_3 + u_1 k_3 \end{cases} \tag{4.22}$$

The question to ask is, given two versions of the same system, can it be proved that, by presuming some trajectories equal, others will also be equal or tend to be so (ie., does the system allow for the mathematical possibility of one trajectory of measurements co-existing with several trajectories of variables to estimate)? The first step is to define the deviations ε_i between the system states x_i and their homologs z_i:

$$\text{Error:} \quad \begin{cases} \varepsilon_1 = x_1 - z_1 \\ \varepsilon_2 = x_2 - z_2 \\ \varepsilon_3 = x_3 - z_3 \end{cases} \Leftrightarrow \begin{cases} z_1 = x_1 - \varepsilon_1 \\ z_2 = x_2 - \varepsilon_2 \\ z_3 = x_3 - \varepsilon_3 \end{cases} \tag{4.23}$$

The next step is to try to make indistinguishable dynamics $d\varepsilon_i/dt$ appear. This can be done by considering the original system (or alternatively, its copy) plus a subtraction of both. Substracting equations (4.22) from (4.21) leads thus to:

$$
\left\{
\begin{array}{rcl}
\frac{dx_1}{dt} - \frac{dz_1}{dt} = & \mu x_1 - \mu^z(x_1 - \varepsilon_1) + \\
& +[\cancel{(-k_7 x_1)} - \cancel{(-k_7(x_1} - \varepsilon_1))] + \\
& +[\cancel{(-u_1 x_1)} - \cancel{(-u_1(x_1} - \varepsilon_1))] + \\
& +[\cancel{u_2 x_1} - \cancel{u_2(x_1} - \varepsilon_1)] \\
\\
\frac{dx_2}{dt} - \frac{dz_2}{dt} = & (-k_1 \mu x_1) - (-k_1 \mu^z(x_1 - \varepsilon_1)) + \\
& +[\cancel{(-u_1 x_2)} - \cancel{(-u_1(x_2} - \varepsilon_2))] + \\
& +[\cancel{u_1 k_2} - \cancel{u_1 k_2}] \\
\\
\frac{dx_3}{dt} - \frac{dz_3}{dt} = & (-k_1 \mu x_1) - (-k_1 \mu^z(x_1 - \varepsilon_1)) + \\
& +[\cancel{(-u_1 x_3)} - \cancel{(-u_1(x_3} - \varepsilon_3))] + \\
& +[\cancel{u_1 k_3} - \cancel{u_1 k_3}]
\end{array}
\right. \quad \Leftrightarrow \quad (4.24)
$$

$$
\Leftrightarrow \quad \left\{
\begin{array}{rcl}
\frac{d\varepsilon_1}{dt} = & \mu x_1 - \mu^z(x_1 - \varepsilon_1) - k_7 \varepsilon_1 - u_1 \varepsilon_1 - u_2 \varepsilon_1 \\
\frac{d\varepsilon_2}{dt} = & -k_1 \mu x_1 + k_1 \mu^z(x_1 - \varepsilon_1) - u_1 \varepsilon_2 \\
\frac{d\varepsilon_3}{dt} = & -k_1 \mu x_1 + k_1 \mu^z(x_1 - \varepsilon_1) - u_1 \varepsilon_3
\end{array}
\right. \quad (4.25)
$$

$$
\text{with:} \quad \left\{
\begin{array}{l}
\mu = k_4 \frac{x_2}{k_5 + x_2} \frac{x_3}{k_6 + x_3} \\
\mu^z = k_4 \frac{(x_2 - \varepsilon_2)}{k_5 + (x_2 - \varepsilon_2)} \frac{(x_3 - \varepsilon_3)}{k_6 + (x_3 - \varepsilon_3)}
\end{array}
\right. \quad (4.26)
$$

The idea is, at this step, to verify the implications on ε_i of assuming that some variables will be measured. More precisely, we wish to verify the possibility of the measurements trajectories co-existing with two or more trajectories for each of the variables that we intend to estimate/observe. If two or more trajectories of a particular variable to estimate can co-exist with the measurements trajectories, then it is not possible to distinguish one from the other (since this particular variable will not be measured).

In this case, if biomass, $x_1 = Xv$, is being measured throughout time, then it can be presumed that $\varepsilon_1 = 0$ throughout time. Therefore, $d\varepsilon_1/dt = 0$, as well. Possible consequences may be that:

$$
\text{if, for } t \geq 0, \quad \varepsilon_1 = 0 \Rightarrow \left\{ \begin{array}{l} \varepsilon_2 = 0 \\ \varepsilon_3 = 0 \end{array} \right. \text{ then } \left\{ \begin{array}{l} \varepsilon_2 \text{ is observable} \\ \varepsilon_3 \text{ is observable} \end{array} \right. \quad (4.27)
$$

$$
\text{if, for } t \geq 0, \quad \varepsilon_1 = 0 \Rightarrow \left\{ \begin{array}{l} \varepsilon_2 \to 0 \\ \varepsilon_3 \to 0 \end{array} \right. \text{ then } \left\{ \begin{array}{l} \varepsilon_2 \text{ is detectable} \\ \varepsilon_3 \text{ is detectable} \end{array} \right. \quad (4.28)
$$

$$\text{if, for } t \geq 0, \quad \varepsilon_1 = 0 \Rightarrow \begin{cases} \varepsilon_2 \neq 0 \\ \varepsilon_3 \neq 0 \end{cases} \text{ then } \begin{cases} \varepsilon_2 \text{ is non observable} \\ \varepsilon_3 \text{ is non observable} \end{cases} \tag{4.29}$$

In order to check the consequence of measuring biomass on indistinguishable dynamics, lets consider that $\varepsilon_1 = 0$ for $t \geq 0$ in equations (4.25). It follows, in this case, that:

$$\begin{cases} 0 = & \mu x_1 - \mu^z x_1 \\ \frac{d\varepsilon_2}{dt} = & -k_1(\mu - \mu^z)x_1 - u_1\varepsilon_2 \\ \frac{d\varepsilon_3}{dt} = & -k_1(\mu - \mu^z)x_1 - u_1\varepsilon_3 \end{cases} \Leftrightarrow \tag{4.30}$$

$$\Leftrightarrow \quad \begin{cases} \mu = & \mu^z \quad (\text{for } x_1 \neq 0) \\ \frac{d\varepsilon_2}{dt} = & -u_1\varepsilon_2 \\ \frac{d\varepsilon_3}{dt} = & -u_1\varepsilon_3 \end{cases} \Leftrightarrow \tag{4.31}$$

$$\Leftrightarrow \quad \begin{cases} k_4 \frac{x_2}{k_5+x_2}\frac{x_3}{k_6+x_3} = & k_4 \frac{(x_2-\varepsilon_2)}{k_5+(x_2-\varepsilon_2)}\frac{(x_3-\varepsilon_3)}{k_6+(x_3-\varepsilon_3)} \wedge x_1 \neq 0 \\ \frac{d\varepsilon_2}{dt} = & -u_1\varepsilon_2 \\ \frac{d\varepsilon_3}{dt} = & -u_1\varepsilon_3 \end{cases} \tag{4.32}$$

The analytical solutions for the second and third equations converges asymptotically to zero when a, in the general form below, is a positive constant:

$$\frac{df}{dt} = -af \Leftrightarrow f(t) = f_0 e^{-at} \tag{4.33}$$

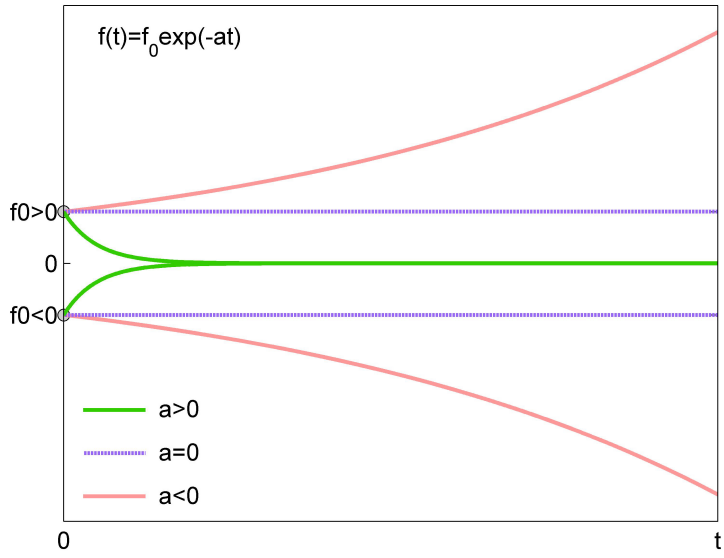


Figure 4.3: Analytical solution of an equation in the form of that of $d\varepsilon_2/dt$ and $d\varepsilon_3/dt$.

We are, in this case, under conditions (4.28). This method allows us to conclude that asymptotic convergence occurs for a dilution ratio $u_1 > 0$. The system is detectable (asymptotic convergence) because:

$$\lim_{t\to\infty} \varepsilon_2(t) = 0 \tag{4.34}$$

$$\lim_{t\to\infty} \varepsilon_3(t) = 0 \tag{4.35}$$

Thus, except for the batch mode, system states x_2 and x_3 are distinguishable if x_1 is measured[1].

To conclude, several methods are helpful in the study of the observability/detectability properties of this simple illustrative toy model and measurement configuration. In particular, the indistinguishable dynamics approach allows, in this specific case, to reach global conclusions in a rather straighforward manner. We will, therefore, apply it to a more complex animal cell culture model in the following section.

[1]In batch, the initial condition is not forgotten throughout the experiment and the error is always equal to the initial one: $d\varepsilon_i/dt = 0$ thus $\varepsilon_i(t) = \varepsilon_i(t_0)$ for $i = 2, 3$.

4.3 Case study: hybridoma cell culture

If different limitation/inhibition phenomena are to be captured by a model intended for use in a control strategy of an animal cell culture in continuous perfused regime, then this model will undoubtedly have a higher order than the model of the previous example: more states (concentrations) are needed to express these phenomena, and, therefore, the complexity of the observability analysis may rapidly increase. We will now study a higher order model by using the indistinguishable dynamics approach to reach conclusions (global, hopefully, if not, local) about the system. Different measurement scenarios will be examined.

4.3.1 Process Model and Analysis

A real dynamic model of hybridoma cells producing monoclonal antibody is considered (De Tremblay et al, 1992). It comprises 7 states (biomass, glucose, lactate, glutamine, ammonia, monoclonal antibodies, volume), 16 parameters and describes typical animal cell culture phenomena.

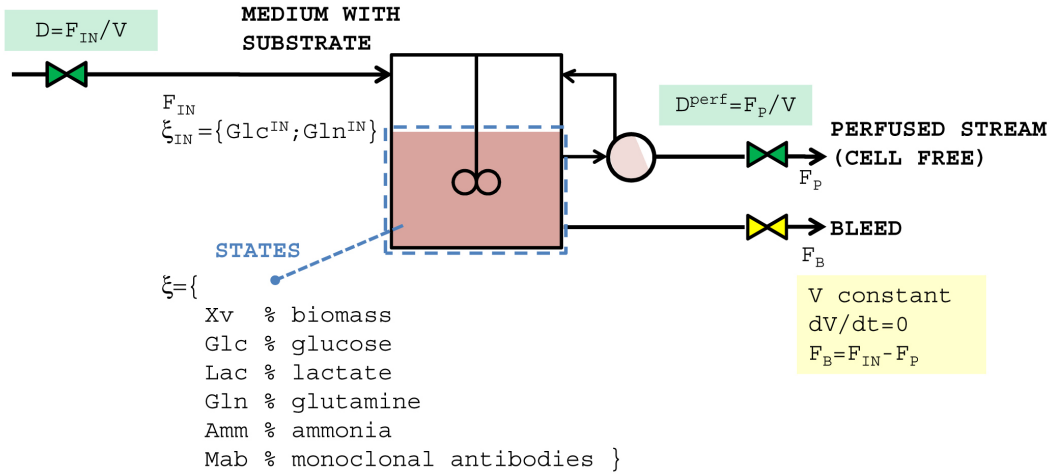


Figure 4.4: System layout for an animal cell culture performed in a bioreactor.

Four biological macroreactions are considered[1]:

$$(-\nu_{21})\text{Glc} + (-\nu_{41})\text{Gln} \xrightarrow{\varphi_1} \circlearrowleft\text{Xv} + \nu_{31}\text{Lac} + \nu_{51}\text{Amm} \tag{4.36}$$

[1]These describe cell growth, cell death, cell maintenance and bioproduct production, respectively.

$$(-1)\mathrm{Xv} \xrightarrow{\varphi_2} \mathrm{Xd} \tag{4.37}$$

$$(-1)\mathrm{Glc} + (-\nu_{13})\mathrm{Xv} \xrightarrow{\varphi_3} \nu_{13}\mathrm{Xv} + \nu_{33}\mathrm{Lac} \tag{4.38}$$

$$(-\nu_{14})\mathrm{Xv} \xrightarrow{\varphi_4} \nu_{14}\mathrm{Xv} + \mathrm{MAb} \tag{4.39}$$

Mass balance equations are given by

$$\frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ -k_{14} & 0 & -1 & 0 \\ k_{16} & 0 & k_{17} & 0 \\ -k_{15} & 0 & 0 & 0 \\ k_{13} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \varphi_3 \\ \varphi_4 \end{bmatrix} + \begin{bmatrix} -x_1 & x_1 \\ (-x_2 + k_{18}) & 0 \\ -x_3 & 0 \\ (-x_4 + k_{19}) & 0 \\ -x_5 & 0 \\ -x_6 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \tag{4.40}$$

with the following notation: $x_1 = Xv$; $x_2 = Glc$; $x_3 = Lac$; $x_4 = Gln$; $x_5 = Amm$; $x_6 = MAb$; $k_1 = \alpha$; $k_2 = \beta$; $k_3 = k_\mu$; $k_4 = k_{d,Amm}$; $k_5 = k_{d,Gln}$; $k_6 = k_{d,Lac}$; $k_7 = k_{Glc}$; $k_8 = k_{Gln}$; $k_9 = k_{m,Glc}$; $k_10 = m_{Glc}$; $k_{11} = \mu_{d,max}$; $k_{12} = \mu_{max}$; $k_{13} = \nu_{51} = Y_{Amm/Gln}/Y_{Xv/Gln}$; $k_{14} = \nu_{21} = 1/Y_{Xv/Glc}$; $k_{15} = \nu_{41} = 1/Y_{Xv/Gln}$; $k_{16} = \nu_{31} = Y_{Lac/Glc}/Y_{Xv/Glc}$; $k_{17} = \nu_{33} = Y_{Lac/Glc}$; $k_{18} = Glc^{IN}$; $k_{19} = Gln^{IN}$; $u_1 = D = F^{IN}/V$; $u_2 = D_{perf} = F_{perf}/V$; V constant. Parameter values are listed on p. 86.

The reaction rates are given by:

$$\varphi_i = \mu_i x_1 \tag{4.41}$$

$$\mu_1 = k_{12}\frac{x_2}{k_7 + x_2}\frac{x_4}{k_8 + x_4} \tag{4.42}$$

$$\mu_2 = k_{11}\frac{1}{(k_{12} - k_6 x_3)}\frac{1}{(k_{12} - k_{14}x_5)}\frac{k_5}{k_5 + x_4} \tag{4.43}$$

$$\mu_3 = k_{10}\frac{x_2}{k_9 + x_2} \tag{4.44}$$

$$\mu_4 = k_2 + k_1\frac{\mu_1}{k_3 + \mu_1} \tag{4.45}$$

It can be seen that biomass concentration x_1 depends on all states besides the antibodies concentration x_6. The antibodies, on the other hand, have no influence whatsoever on the dynamics of other states, meaning that they can be excluded from the observability analysis.

Table 4.2: Dependency of state dynamics on system states for De Tremblay's model.

dynamics	x_1	x_2	x_3	x_4	x_5	x_6	state	type
$\dot{x_1}$	●	●	●	●	●		biomass	◯
$\dot{x_2}$	●	●	●	●	●		glucose	□
$\dot{x_3}$	●	●	●	●	●		lactate	△
$\dot{x_4}$	●	●	●	●	●		glutamine	□
$\dot{x_5}$	●	●	●	●	●		ammonia	△
$\dot{x_6}$	●	●	●	●	●	●	antibodies	△

Legend: ◯=biomass; □=substrate; △=metabolite/product.

A copy of the original system is built.

$$\text{System} \quad \begin{cases} \frac{dx_1}{dt} = (k_{12}\frac{x_2}{k_7+x_2}\frac{x_4}{k_8+x_4})x_1 - (k_{11}\frac{1}{(k_{12}-k_6x_3)}\frac{1}{(k_{12}-k_{14}x_5)}\frac{k_5}{k_5+x_4})x_1 - \\ \qquad -x_1u_1 + x_1u_2 \\ \frac{dx_2}{dt} = -k_{14}(k_{12}\frac{x_2}{k_7+x_2}\frac{x_4}{k_8+x_4})x_1 - (k_{10}\frac{x_2}{k_9+x_2})x_1 \\ \qquad -x_2u_1 + k_{18}u_1 \\ \frac{dx_3}{dt} = k_{16}(k_{12}\frac{x_2}{k_7+x_2}\frac{x_4}{k_8+x_4})x_1 + k_{17}\left(k_{10}\frac{x_2}{k_9+x_2}\right)x_1 - x_3u_1 \\ \frac{dx_4}{dt} = -k_{15}(k_{12}\frac{x_2}{k_7+x_2}\frac{x_4}{k_8+x_4})x_1 - x_4u_1 + k_{19}u_1 \\ \frac{dx_5}{dt} = k_{13}(k_{12}\frac{x_2}{k_7+x_2}\frac{x_4}{k_8+x_4})x_1 - x_5u_1 \end{cases} \tag{4.46}$$

$$\text{Copy} \quad \begin{cases} \frac{dz_1}{dt} = (k_{12}\frac{z_2}{k_7+z_2}\frac{z_4}{k_8+z_4})z_1 - (k_{11}\frac{1}{(k_{12}-k_6z_3)}\frac{1}{(k_{12}-k_{14}z_5)}\frac{k_5}{k_5+z_4})z_1 - \\ \qquad -z_1u_1 + z_1u_2 \\ \frac{dz_2}{dt} = -k_{14}(k_{12}\frac{z_2}{k_7+z_2}\frac{z_4}{k_8+z_4})z_1 - (k_{10}\frac{z_2}{k_9+z_2})z_1 - z_2u_1 + k_{18}u_1 \\ \frac{dz_3}{dt} = k_{16}(k_{12}\frac{z_2}{k_7+z_2}\frac{z_4}{k_8+z_4})z_1 + k_{17}\left(k_{10}\frac{z_2}{k_9+z_2}\right)z_1 - z_3u_1 \\ \frac{dz_4}{dt} = -k_{15}(k_{12}\frac{z_2}{k_7+z_2}\frac{z_4}{k_8+z_4})z_1 - z_4u_1 + k_{19}u_1 \\ \frac{dz_5}{dt} = k_{13}(k_{12}\frac{z_2}{k_7+z_2}\frac{z_4}{k_8+z_4})z_1 - z_5u_1 \end{cases} \tag{4.47}$$

Following the procedure introduced in the previous section, the indistinguishable dynamics can be expressed by subtracting (4.47) to (4.46):

$$\begin{cases} \frac{d\varepsilon_1}{dt} = & \mu_1 x_1 - \mu_1^z(x_1-\varepsilon_1) - \mu_2 x_1 + \mu_2^z(x_1-\varepsilon_1) - \varepsilon_1 u_1 + \varepsilon_1 u_2 \\ \frac{d\varepsilon_2}{dt} = & -k_{14}\mu_1 x_1 + k_{14}\mu_1^z(x_1-\varepsilon_1) - \mu_3 x_1 + \mu_3^z(x_1-\varepsilon_1) - \varepsilon_2 u_1 \\ \frac{d\varepsilon_3}{dt} = & k_{16}\mu_1 x_1 - k_{16}\mu_1^z(x_1-\varepsilon_1) + k_{17}\mu_3 x_1 - k_{17}\mu_3^z(x_1-\varepsilon_1) - \varepsilon_3 u_1 \\ \frac{d\varepsilon_4}{dt} = & -k_{15}\mu_1 x_1 + k_{15}\mu_1^z(x_1-\varepsilon_1) - \varepsilon_4 u_1 \\ \frac{d\varepsilon_5}{dt} = & k_{13}\mu_1 x_1 - k_{13}\mu_1^z(x_1-\varepsilon_1) - \varepsilon_5 u_1 \end{cases} \tag{4.48}$$

where

$$\mu_1^z = k_{12}\frac{(x_2-\varepsilon_2)}{k_7+(x_2-\varepsilon_2)}\frac{(x_4-\varepsilon_4)}{k_8+(x_4-\varepsilon_4)} \tag{4.49}$$

$$\mu_2^z = k_{11}\frac{1}{k_{12}-k_6(x_3-\varepsilon_3)}\frac{1}{k_{12}-k_{14}(x_5-\varepsilon_5)}\frac{k_5}{k_5+(x_4-\varepsilon_4)} \tag{4.50}$$

$$\vdots$$

Two practical measurement configurations are considered:

- Case A: Living biomass and glucose are measured;
- Case B: Living biomass only is measured (eg. with a capacitance probe).

4.3.2 Case A - Biomass and glucose measurements

If we consider that biomass and glucose concentrations (x_1 and x_2) are being measured, then $\varepsilon_1 = 0$, $\varepsilon_2 = 0$ for $t \geq 0$ and it follows that $d\varepsilon_1/dt = 0$, $d\varepsilon_2/dt = 0$, as well.

The idea is now to check whether if, by assuming this, then indistinguishable dynamics equations (4.48) will allow us to make global conclusions about the observability/detectability of the system. Interesting conclusions could be, for example:

$$\text{if, for } t \geq 0, \quad \begin{cases} \varepsilon_1 = 0 \\ \varepsilon_2 = 0 \end{cases} \Rightarrow \begin{cases} \varepsilon_3 = 0 \\ \varepsilon_4 = 0 \\ \varepsilon_5 = 0 \end{cases} \text{ then } \begin{cases} \varepsilon_3 \text{ is observable} \\ \varepsilon_4 \text{ is observable} \\ \varepsilon_5 \text{ is observable} \end{cases} \tag{4.51}$$

$$\text{if, for } t \geq 0, \quad \begin{cases} \varepsilon_1 = 0 \\ \varepsilon_2 = 0 \end{cases} \Rightarrow \begin{cases} \varepsilon_3 \to 0 \\ \varepsilon_4 \to 0 \\ \varepsilon_5 \to 0 \end{cases} \text{ then } \begin{cases} \varepsilon_3 \text{ is detectable} \\ \varepsilon_4 \text{ is detectable} \\ \varepsilon_5 \text{ is detectable} \end{cases} \tag{4.52}$$

After some analytical manipulations (see proof in Appendix B, p. 242-245) we conclude that what effectively happens with this model under this measurements strategy, is that:

$$\text{if, for } t \geq 0, \quad \begin{cases} \varepsilon_1 = 0 \\ \varepsilon_2 = 0 \end{cases} \Rightarrow \begin{cases} \varepsilon_3 \to 0 \\ \varepsilon_4 = 0 \\ \varepsilon_5 \to 0 \end{cases} \text{ thus } \begin{cases} \varepsilon_3 \text{ is detectable}, D > 0 \\ \varepsilon_4 \text{ is observable} \\ \varepsilon_5 \text{ is detectable}, D > 0 \end{cases} \tag{4.53}$$

It is concluded that the measurements of x_1 and x_2 provide sufficient information to at least ensure detectability in any cell culture operation regime apart from batch.

Hence, this model could be used by an observer to provide an estimation of the concentrations of lactate, glutamine, ammonia (x_3, x_4, x_5) based on (non-negative) measurements sent by online probes (eg. a Fogale probe and a bioanalyser or a glucose probe or a NIR) for biomass and glucose concentrations (x_1, x_2).

Actually, it can also be proved similarly that the same is also possible by measuring the other substrate (ie. glutamine instead of glucose):

$$\text{if, for } t \geq 0, \quad \begin{cases} \varepsilon_1 = 0 \\ \varepsilon_4 = 0 \end{cases} \Rightarrow \begin{cases} \varepsilon_2 = 0 \\ \varepsilon_3 \to 0 \\ \varepsilon_5 \to 0 \end{cases} \text{ thus } \begin{cases} \varepsilon_2 \text{ is observable} \\ \varepsilon_3 \text{ is detectable}, D > 0 \\ \varepsilon_5 \text{ is detectable}, D > 0 \end{cases} \tag{4.54}$$

With respect to antibodies concentration x_6, up to now we have considered a reduced form of the system since x_6 has no influence on other states). Actually, x_6 is also detectable (for $D > 0$) if biomass and one of the substrates are measured since $\varepsilon_1 = \varepsilon_2 = \varepsilon_4 = 0$ and thus $\mu_1 = \mu_1^z$ and it follows that $d\varepsilon_6/dt = -\varepsilon_6 u_1$.

Therefore, for this model, if at least biomass and one of the substrates are measured, then an observer should have enough information to reconstruct the full system. But could we dare to measure only biomass in order to reconstruct the full system? We will investigate this in the following subsection.

4.3.3 Case B - Biomass measurements

Lets now consider that only the concentration of one component (biomass) is available and that other concentrations will need to be estimated with the help of this model. If x_1 is being measured, then $\varepsilon_1 = 0, d\varepsilon_1/dt = 0$. Equations (4.48), after the manipulations present in Appendix B, p. 245-247, lead to:

$$\mu_1' = \mu_2' \tag{4.55}$$

$$\frac{d\varepsilon_2}{dt} = \left[-k_{14}\mu_1' - \mu_3'\right] x_1 - \varepsilon_2 u_1 \tag{4.56}$$

$$\frac{d\varepsilon_3}{dt} = \left[k_{16}\mu_1' + k_{17}\mu_3'\right] x_1 - \varepsilon_3 u_1 \tag{4.57}$$

$$\frac{d\varepsilon_4}{dt} = \left[-k_{15}\mu_1'\right] x_1 - \varepsilon_4 u_1 \tag{4.58}$$

$$\frac{d\varepsilon_5}{dt} = \left[k_{13}\mu_1'\right] x_1 - \varepsilon_5 u_1 \tag{4.59}$$

where

$$\mu_i' = \mu_i - \mu_i^z \tag{4.60}$$

Unfortunately, for this particular very extreme situation of the measurement of one single variable, it is not easy to conclude in a straightforward manner about the observability/detectability conditions.

Computation of the observability map and linearization suggests that the system would be locally observable (Fig. 4.5) for illustrative normal operating conditions. The evolution of normalized values attained by the determinant is listed on the left column and plotted on the right (top) along with the states (middle and bottom).

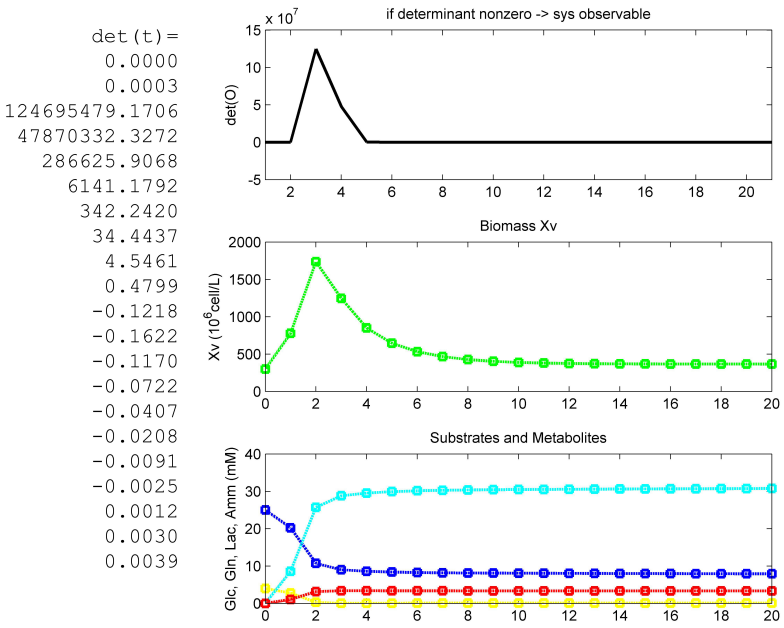


Figure 4.5: Evolution of the determinant of the observability matrix and concentrations during a normal culture.

4.3.4 Comments

We have seen several strategies for the study of observability/detectability properties of animal cell culture models. For a reasonably complex model describing a real hybridoma culture, we were able to achieve global conclusions with the help of indistinguishable dynamics: the system is detectable if at least biomass and one substrate are measured. In other situations, where a global conclusion was not possible, local observability was studied for a set of experimental conditions.

4.4 Kalman filter design

Following the observability/detectability analysis, let us now build the observer by designing a Kalman filter considering the two measurement configurations. The implemented algorithm extends the use of the filter to nonlinear systems by use of a linearization along the state trajectory.

System:

$$\begin{array}{l} \dot{x}(t) = f(\hat{x}(t), u(t)) + \eta(t); \quad x(t_0) = x_0 \\ y(t_k) = C(t_k)x(t_k) + \varepsilon(t_k) \end{array} \tag{4.61}$$

Gaussian white noises:

$$\begin{array}{l} \eta(t) \sim \mathbb{N}(0, R_\eta(t)) \\ \varepsilon(t) \sim \mathbb{N}(0, R_\varepsilon(t)) \end{array} \tag{4.62}$$

A continuous-discrete version of the extended Kalman filter (EKF) is considered.

- Initialization
  We consider a given initial condition and covariance P.

$$\begin{cases} m_{t_0|t_0} = x_0 \\ P_{t_0|t_0} = P_0 \end{cases} \tag{4.63}$$

- Continuous Prediction for $t_{k-1} < t < t_k$:

$$\begin{cases} \dot{\hat{x}}(t) = f(\hat{x}(t), u(t)) & \hat{x}(t_{k-1}) = m_{t_{k-1}|t_{k-1}} \\ \dot{P}(t) = AP(t) + P(t)A^T + R_\eta(t); & P(t_{k-1}) = P_{t_{k-1}|t_{k-1}} \\ \\ \text{where} \\ A = Jac(\hat{x}(t), u(t)) = \dfrac{\delta \hat{f}(\hat{x}(t), u(t))}{\delta x} \end{cases} \tag{4.64}$$

- Discrete-time correction at $t = t_k$

$$\begin{cases} m_{t_k|t_{k-1}} = \hat{x}(t_k) \\ P_{t_k|t_{k-1}} = P(t_k) \end{cases} \tag{4.65}$$

$$\begin{array}{l} K(t_k) = P_{t_k|t_{k-1}} C^T(t_k) \left(C(t_k) P_{t_k|t_{k-1}} C^T(t_k) + R_\varepsilon(t) \right)^{-1} \\ \begin{cases} m_{t_k|t_k} = m_{t_k|t_{k-1}} + K(t_k) \left(y(t_k) - C(t_k) m_{t_k|t_{k-1}} \right) \\ P_{t_k|t_{k-1}} = P(t_k) - K(t_k) C(t_k) P_{t_k|t_{k-1}} \end{cases} \end{array} \tag{4.66}$$

This algorithm is appropriate for bioprocesses, which can have low sampling frequencies of the available probes, whereas the process models are nonlinear mass balance equations. The state estimator can be coupled to a model predictive controller, such as the one described in Sbarciog et al (2013).

The convergence of the filter is illustrated in the case of biomass and glucose measurements in Figure 4.6.

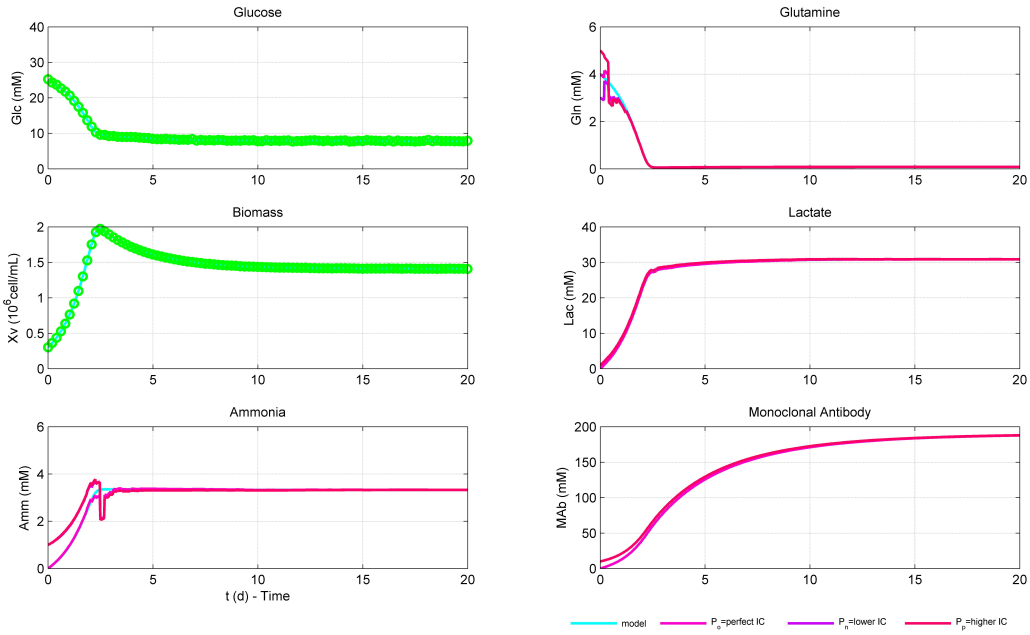


Figure 4.6: EKF based on biomass and glucose measurements. R_η=*diag*([0 0 0 0 0 0]); R_η=*diag*([0.1 0.01 30 0.1 0.01 1]); P_0=*diag*([10 10 10 10 10 10]). Estimation (magenta), real process variables (blue).

As for the case where only biomass measurements are available (Fig. 4.7), the filter appears to perform sufficiently well since estimates still converge to the real values even with an extremely poor knowledge of initial conditions.

In reality, an animal cell culture begins with initial conditions that are relatively well known: the medium is generally prepared beforehand and its concentrations of glucose and glutamine are fairly well known. On the other hand, before cells are inoculated, the concentration of metabolites lactate and ammonia are close to zero and the concentration of the synthesised bioproduct is zero.

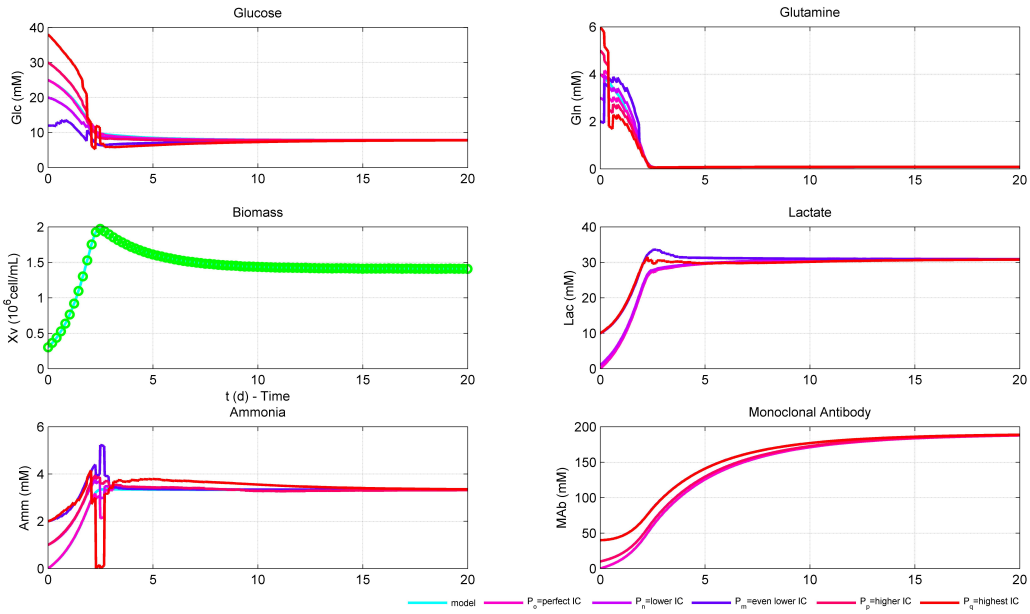


Figure 4.7: EKF based on biomass measurements. Lowest IC=[12 2 Xv_0 10 2 40]^T; lower IC=[20 3 Xv_0 1 1 10]^T; perfect IC=[25 4 Xv_0 0 0 0]^T; higher IC=[30 5 Xv_0 1 1 10]^T; highest IC=[38 6 Xv_0 10 2 40]^T.

Another issue to be addressed is the performance of the filter given measurement noise. The figures presented assume a small level of noise, which is, in fact, consistent with the performances of probes such as the Fogale, illustrated in Fig. 4.8.

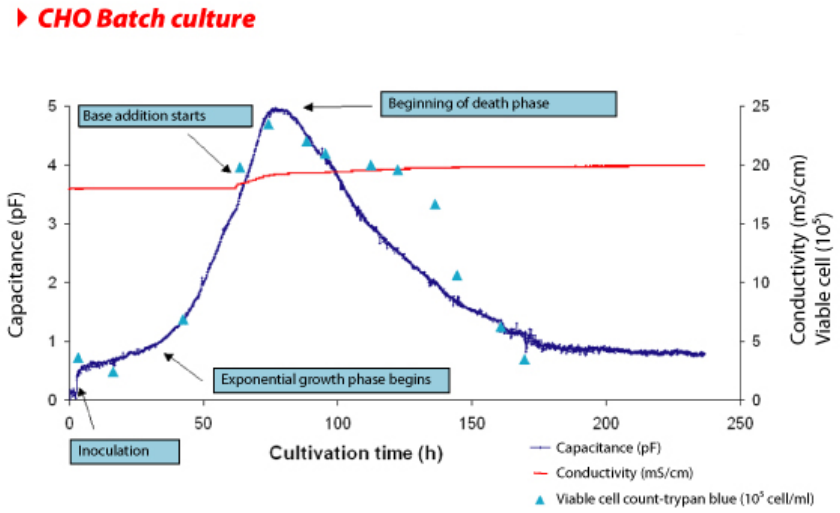


Figure 4.8: Biomass concentration signal given by a Fogale probe in a batch CHO culture (Fogale-nanotech, 2013).

4.5 Some conclusions

This application-oriented study addressed two questions. The first is the assessment of observability of nonlinear dynamic models of cell cultures, in the case of minimal measurement configurations. Given the complexity of these systems, most available methods can be helpful for a local analysis. Reaching global conclusions, however, involves long analytical manipulations that can prove frustrating. Canonical forms such as those used in bacterial cultures in Dewasme et al (2012) are also of difficult application. The indistinguishable dynamics method proposed in Moreno et al (2014) has shown an interesting approach, in this specific case, to come to global conclusions about observability/detectability in a straightforward manner. However, the solution of the differential algebraic system involved in that method may be delicate in other situations.

The second question is the design of observers for cell cultures using a few measurement probes, and in particular the relatively recent biomass probes which provide almost time-continuous evolution of the biomass with low levels of noise. This has to be contrasted with bioanalyzers that automate enzymatic kit analyses. Besides not being very frequent, these analyses have high operation costs and can present some errors. A more attractive configuration could be a biomass probe coupled to another probe (NIR, glucose sensor) in order to follow (at least) one substrate. In fact, the NIR could potentially be interesting for the development a simultaneous procedure of measuring online glucose, lactate, glutamine and ammonium (Didion et al, 2014).

Chapter 5

Animal cell culture control

This chapter focuses on control and begins with a study of a control scheme from the literature for a very simple biomass-substrate model describing animal cell cultures in continuous perfused operation. Its adaptive backstepping control strategy is compared to one that can be achieved by a nonlinear model predictive controller (NMPC). NMPC is found to be a promising control strategy. Next, an illustration of how NMPC can be applied to a model of higher complexity (5 to 7 states, such as the ones chosen in this thesis) is shown. Finally, the question of model choice is tackled: how complex should a model be in a continuous perfused production scenario. Model-plant mismatch is studied in a case study with different models identified for the same databank of interferon-γ producing CHO-320 cell cultures performed at the UMons (see Chapter 3). The importance of having kinetics correctly captured is highlighted.

5.1 Introduction

As mentioned already in Subsection 1.1 (p. 35), the purposes of modeling can be various. One of them may be to use the model in the real-time control of a system.

Let us recall that some biomedicines (such as recombinant proteins) are produced in cultures where animal cells, transfected in order to express these high value proteins, are suspended. In the industry, the large scale production of these biologicals appears to be commonly done in stirred tank reactors. The efforts for increasing the culture productivity in these systems focus on adjusting the media composition on the one hand, and the modes of operation on the other hand.

The most popular operating modes in animal cell cultivation are batch, fed-batch and perfusion modes (Fig. 1.1 on p. 31). Batch and fed-batch modes do not offer many alternatives for control, as in these cases the feed rate F^{IN} is either absent or limited and the growth is inhibited by the accumulation of toxic metabolites, which cannot be removed. In perfusion mode, fresh medium is fed to replenish the consumed nutrients, while an equal volume of spent medium is continuously withdrawn from it, allowing for the removal of toxic components. Cells are retained or recycled back to the reactor by some type of retention device, such as sonoperfusion filters. Higher cell concentrations and higher productivity are achieved in perfusion cultures than in conventional batch cultures. This is very clear in the example of Figure 1.6 (p. 37), for instance.

Some goals of the control strategy may be to:

- maximise the time of operation - hence the preference for a continuous regime in perfusion mode (see Fig. 1.5 on p.36);
- maximise the bioproduct of interest - in quantity, in easiness of purification, ...;
- minimise substrate waste - that is, to provide each substrate in a quantity adequate to cells' needs instead of by default overcharging the medium with it;
- maximise the automation of the process - to have a controller that is robust;
- set an optimum mode for animal cells - to define a useful cost function.

An appealing concept is that of the dynamic growth medium (Fig. 5.1), in which a multivariable controller would allow to adjust substrate availability to cells needs, since it is well known that these vary throughout a culture.

For example, it could act on several component concentrations simultaneously, according to estimated cell metabolism. Currently most cultures are performed with static composition media. This would present a possibility of optimizing operation by making available to a multivariable controller pools of different substrates (eg. carbon sources, such as glucose, or carbon and nitrogen sources, such as glutamine). The extra degrees of freedom gained by the possibility of varying the availability of different substrates could lead to better performances.

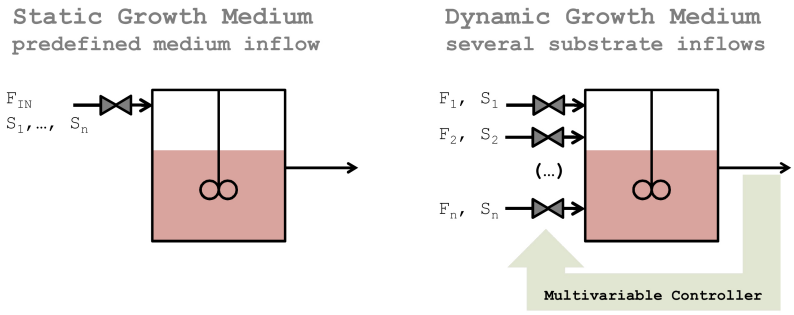


Figure 5.1: Static culture medium vs dynamic culture medium.

The successful operation of cultures in perfusion mode requires tight control, and some recent developments consider the use of multivariable control to manipulate the feed and bleed rates, F^{IN} and F_{bleed} (Deschenes et al, 2006a,b), as well as the composition of the feed flow (Saraiva et al, 2010, 2011; Sbarciog et al, 2013).

5.2 Case study: backstepping vs. NMPC

In Saraiva et al (2010), a nonlinear model predictive control (NMPC) of an animal cell culture in perfusion mode was studied and compared to another control strategy proposed in Deschenes et al (2006a); Deschenes (2007); Deschenes et al (2006b) with a simple model intended for a perfused animal cell culture (backstepping control). The same model was used in the two control strategies. The goal was to control biomass and substrate concentrations (Xv and S, respectively). This is a simple example of how biomass and substrate concentrations can be controlled in a continuous perfused culture and the nonlinearities of the system dealt with.

5.2.1 Model statement

The model considered has been presented in more detail on p. 57.

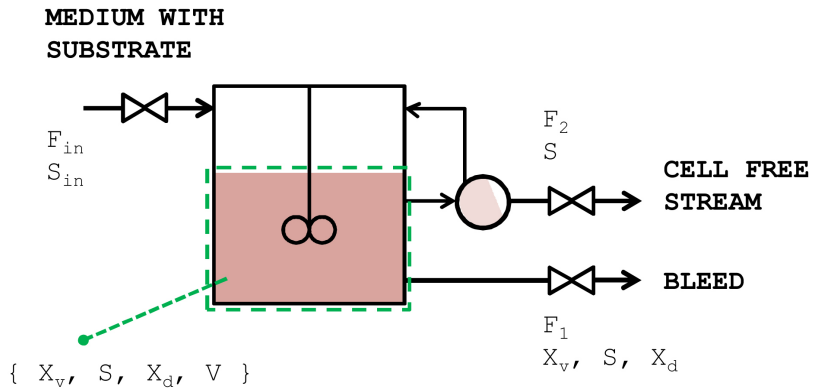


Figure 5.2: Bioreactor cell culture: layout for a continuous perfused regime operated at constant volume.

It is supposed that viable cells Xv grow on a substrate S. Some of the viable cells die (Xd). Cell net growth presumes a dependency on substrate and both viable and dead cells concentration.

$$S \xrightarrow{\varphi_{growth}} \circlearrowleft \mathrm{Xv} \tag{5.1}$$

$$\mathrm{Xv} \xrightarrow{\varphi_{death}} \mathrm{Xd} \tag{5.2}$$

For a continuous perfused regime operating at constant volume, as illustrated in Fig. 5.2:

$$\text{Balances:} \quad \begin{cases} \frac{dXv}{dt} = & \mu Xv & +\mu_d Xv & -\frac{F_{IN}}{V} Xv & +\frac{F_2}{V} Xv \\ \frac{dS}{dt} = & -k_1 \mu Xv & & -\frac{F_{IN}}{V} S & +\frac{F_{IN}}{V} S_{IN} \\ \frac{dXd}{dt} = & & \mu_d Xv & -\frac{F_{IN}}{V} Xd & \\ F_{IN} = & & & F_1 & +F_2 \end{cases} \tag{5.3}$$

$$\text{Cell growth (Contois):} \quad \mu = \mu_{max} \frac{S}{k_c Xv + S} \tag{5.4}$$

$$\text{Cell death:} \quad \mu_d = k_d(Xv + Xd) \tag{5.5}$$

$$\text{Constants:} \quad \begin{cases} \mu_{max} = 0.0504h^{-1} \\ k_c = 1.87mM/(10^6 cell/mL) \\ k_d = 0.000758h^{-1}/(10^6 cell/mL) \end{cases}$$

Units: $[Xv]$=$[Xd]$=$10^6 cell/mL$=$10^9 cell/L$, $[S_i]$=mM=$mmol/L$, $[\mu]$=$[\mu_d]$=h^{-1}, $[F_i]$=L/h, $[V]$=L.

5.2.2 Backstepping control

Backstepping is a technique dating from the 1990's that relies on a recursive method to stabilize nonlinear dynamical systems of lower triangular forms (Zhou and Wen, 2008):

$$\begin{cases} \dot{x_1} = f_1(x_1, x_2) \\ \dot{x_2} = f_2(x_1, x_2, x_3) \\ \vdots \\ \dot{x_n} = f_1(x_1, x_2, \ldots, x_n, u) \end{cases} \tag{5.6}$$

The recursive algorithm serves to design a control law in as many steps as the relative degree of freedom of the system. Adaptive features are possible, presuming that some parameter values may vary.

The control strategy proposed in Deschenes et al (2006b) considers a strategy with two controllers where the dilution ratios are manipulated in order to control state variables Xv and S. This is illustrated in Fig. 5.3.

The following notation is used:

$$\begin{cases} dXv/dt = & (\mu - \mu_d) & Xv & -\frac{F_{IN}}{V}Xv + \frac{F_2}{V}Xv \\ dS/dt = & -k_1 \quad \mu & Xv + \frac{F_{IN}}{V}S_{IN} & -\frac{F_{IN}}{V}Xv \end{cases} \rightsquigarrow \tag{5.7}$$

$$\rightsquigarrow \begin{cases} dx_1/dt = & \theta_1 & x_1 & -u_1 x_1 \\ dx_2/dt = & \theta_2 & x_1 + u_2 x_2^{IN} & -u_2 x_2 \end{cases} \tag{5.8}$$

The hypothesis that $y_1 = x_1$, $y_2 = x_2$ (no model-plant mismatch) is made. θ_1 and θ_2 are parameters, u_1 and u_2 are the manipulated variables, and x_1 and x_2 are the controlled ones[1].

[1]Notice that at constant volume $-F_{IN} + F_2 = -F_1$.

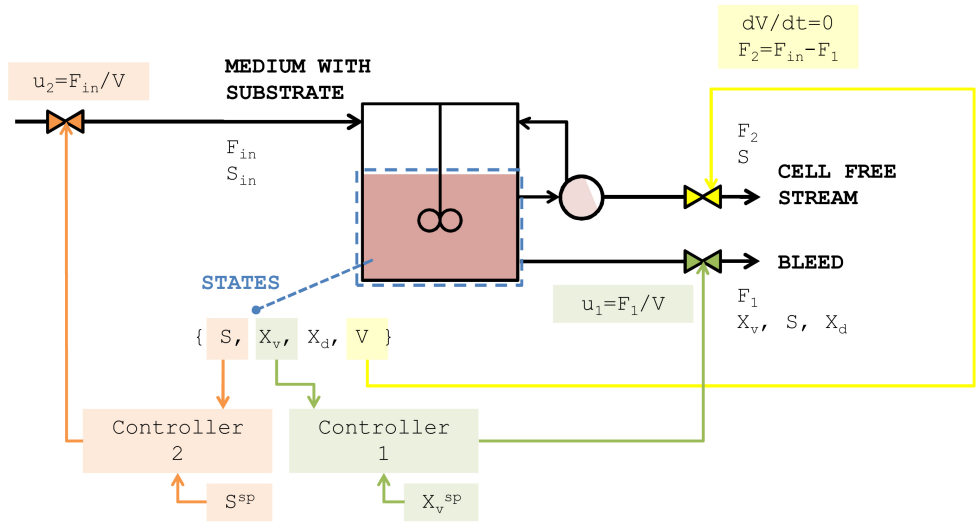


Figure 5.3: Backstepping control applied to a bioreactor cell culture: layout for a continuous perfused regime operated at constant volume.

Since dx_1/dt is independent of x_2 and u_2, then the multivariable controller can be designed sequentially:

1. Design a law to control y_1 (=Xv, biomass) by manipulation of u_1 (=F_1/V);
2. Design a law to control y_2 (=S, substrate) by manipulation of u_2 (=F_{IN}/V).

The idea is to use Lyapunov's theory by choosing a positive Lyapunov function V that is in some way a measure of the error (between the state and its setpoint y_{Ri}) and by forcing this function to have a negative derivative. A controller is then built where this measure of the error V is bound to decrease and converge to zero (therefore, the state will converge to its setpoint). This function may also include an adaptation term if parameters θ_i (eg. kinetic) are presumed uncertain or time-varying. In this way, the existence of $V > 0$ and $dV/dt < 0$ will ensure the stability of the adaptive control scheme.

This case considers the following Lyapunov function:

$$V_1 = \frac{1}{2}z_1^2 + \frac{1}{2\gamma_1}\tilde{\theta}_1^2 \quad \text{where} \tag{5.9}$$

$$\begin{cases} \text{estimation error:} & \tilde{\theta}_1 = \theta_1 - \hat{\theta}_1 \\ \text{adaptation gain:} & \gamma_1 \text{ ,positive} \\ & \overbrace{\phantom{k_{P1}\epsilon_1}}^{\text{proportional}} \quad \overbrace{\phantom{+k_{I1}\int \epsilon_1 dt}}^{\text{integral}} \\ \text{augmented error:} & z_1 = k_{P1}\epsilon_1 + k_{I1}\int \epsilon_1 dt \\ \text{closed loop error:} & \epsilon_1 = y_{R1} - y_1 \end{cases} \tag{5.10}$$

Computing the derivative of V_1, after some manipulations, and presuming a slow parameter time variation occurring in reality ($\dot{\theta}_1 \simeq 0$):

$$\dot{V}_1 = z_1[k_{P1}(\dot{y}_{R1} - \hat{\theta}_1 x_1 + x_1 u_1) + k_{I1}\epsilon_1] + \tilde{\theta}_1[\frac{-\dot{\hat{\theta}}_1}{\gamma_1} - k_{P1} z_1 x_1] \tag{5.11}$$

In order to force this derivative to be negative, one considers also that $\dot{V}_1 = -c_1 z_1^2$ where from eq. (5.11) it follows that:

$$\begin{cases} -c_1 z_1^2 = & z_1[k_{P1}(\dot{y}_{R1} - \hat{\theta}_1 x_1 + x_1 u_1) + k_{I1}\epsilon_1] \\ 0 = & [\frac{-\dot{\hat{\theta}}_1}{\gamma_1} - k_{P1} z_1 x_1] \rightarrow \dot{\hat{\theta}} = -\gamma_1 k_{P1} z_1 x_1 \end{cases} \tag{5.12}$$

Explicitating u_1 and $\dot{\hat{\theta}}_1$, for controller #1:

$$\begin{array}{rl} \text{control law} & \left\{ \quad u_1 = \frac{1}{x_1}\left(-\frac{c_1 z_1}{k_{P1}} - \frac{k_{I1}\epsilon_1}{k_{P1}} - \dot{y}_{R1}\right) + \hat{\theta}_1 \right. \\ \text{adaptation law} & \left\{ \quad \dot{\hat{\theta}}_1 = -\gamma_1 k_{P1} z_1 x_1 \text{with } c_1, \gamma_1 > 0 \right. \end{array} \tag{5.13}$$

The same procedure may be followed to obtain the laws for controller #2:

$$\begin{array}{rl} \text{control law} & \left\{ \quad u_2 = \frac{1}{x_2^{IN} - x_2}\left(\frac{c_2 z_2}{k_{P2}} + \frac{k_{I2}\epsilon_2}{k_{P2}} + \dot{y}_{R2} - \hat{\theta}_2 x_1\right) \right. \\ \text{adaptation law} & \left\{ \quad \dot{\hat{\theta}}_2 = -\gamma_2 k_{P2} z_2 x_1 \text{with } c_2, \gamma_2 > 0 \right. \end{array} \tag{5.14}$$

The controllers are tuned by choosing values for a number of control parameters: the proportional gains k_{Pi}, integral gains k_{Ii}, adaptation gains γ_i and Lyapunov function constants c_i. There is no systematic procedure for backstepping control tuning, but the following steps may be followed (Astrom and Wittenmark, 1997):

- **Step 1** Choose a small enough sampling time Δt for the control algorithm. Heuristics:

$$0.01 \leq \Delta t / \tau_{\text{dominant}} \leq 0.05 \tag{5.15}$$

- **Step 2** Turn down adaptation (assume that parameters θ_i are perfect) and roughly adjust the gains of the PI controller, making it very gentle (non-aggressive). Check controller's performance.
- **Step 3** Turn on adaptation and try small values (magnitudes $10^{-4}/10^{-1}$).
- **Step 4** Increase set point changes and check response. Increase sampling time and do the same.

Let it now be considered that the system is at steady state operating in continuous perfused mode. Starting from this stable point a gentle setpoint change (a 1^{st} order trajectory) in the biomass concentration is introduced while the substrate concentration setpoint remains unchanged:

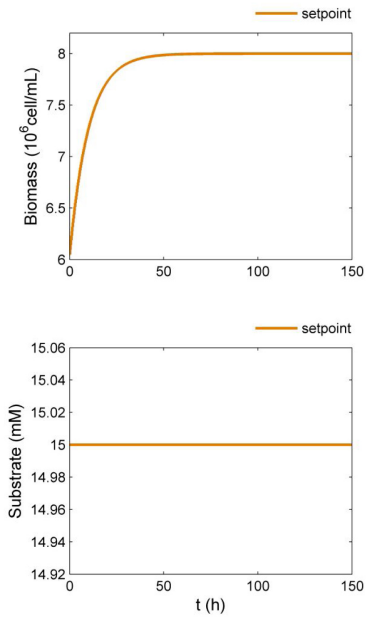


Figure 5.4: Perfused culture: setpoint changes introduced.

If no adaptation is considered (perfect model case), the system reacts as illustrated in Figure 5.5 for the following set of chosen values:

$$\begin{cases} k_{P1} = 0.100; & k_{I1} = 0.10 \\ k_{P2} = 0.100; & k_{I2} = 0.10 \end{cases} \tag{5.16}$$

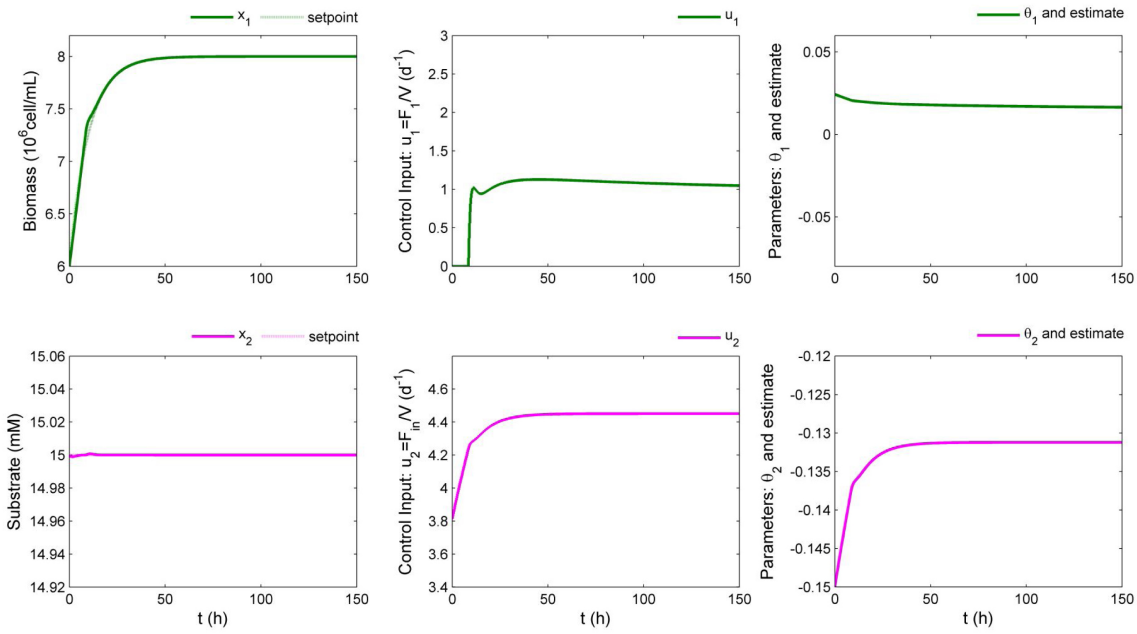


Figure 5.5: Backstepping control: closed-loop response without adaptation and tuning values listed in eq. (5.16).

It can be seen that the closed-loop response is rather good even if controller #1 is initially saturated: u_1 (the bleed dilution ratio) is initially meant to be negative in order for the new higher biomass concentration setpoint to be attained, but is, for physical reasons, forced to be null ($u_1 = 0$, valve closed).

Let us now introduce adaptation by assuming a 90% initial parameter estimation error ($\hat{\theta} = 90\%\theta$). Results using the set of tuning values[1] in eq. (5.17) are represented in Figure 5.6.

$$\begin{cases} k_{P1} = 0.015; \;\; k_{I1} = 0.01; \;\; \gamma_1 = 0.10; \;\; c_1 = 0.02 \\ k_{P2} = 0.015; \;\; k_{I2} = 0.01; \;\; \gamma_2 = 0.10; \;\; c_2 = 0.15 \end{cases} \tag{5.17}$$

Again, there is some saturation.

This adaptation in the control algorithm allows to deal with model-plant mismatch. Again, despite an initial saturation of u_1 (causing some oscillation), overall response is good.

Let us consider now a more drastic scenario (for the same tuning values) with a -60% initial parameter estimate error. The response (Figure 5.7) is now not so good, with some oscillation introduced in the biomass concentration

[1]These were chosen by trial and eror.

and in the substrate concentration (where setpoint compliance is weaker and slower). Again, u_1 is initially saturated.

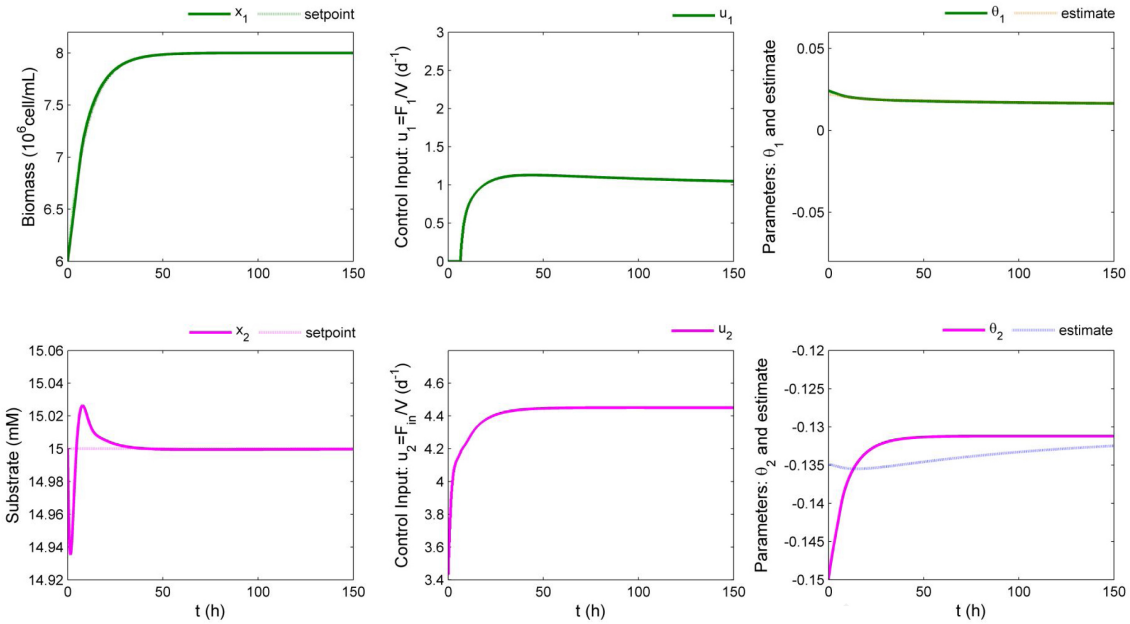


Figure 5.6: Backstepping control: closed-loop response with adaptation ($\hat{\theta} = 90\%\theta$) and tuning values listed in eq. (5.17).

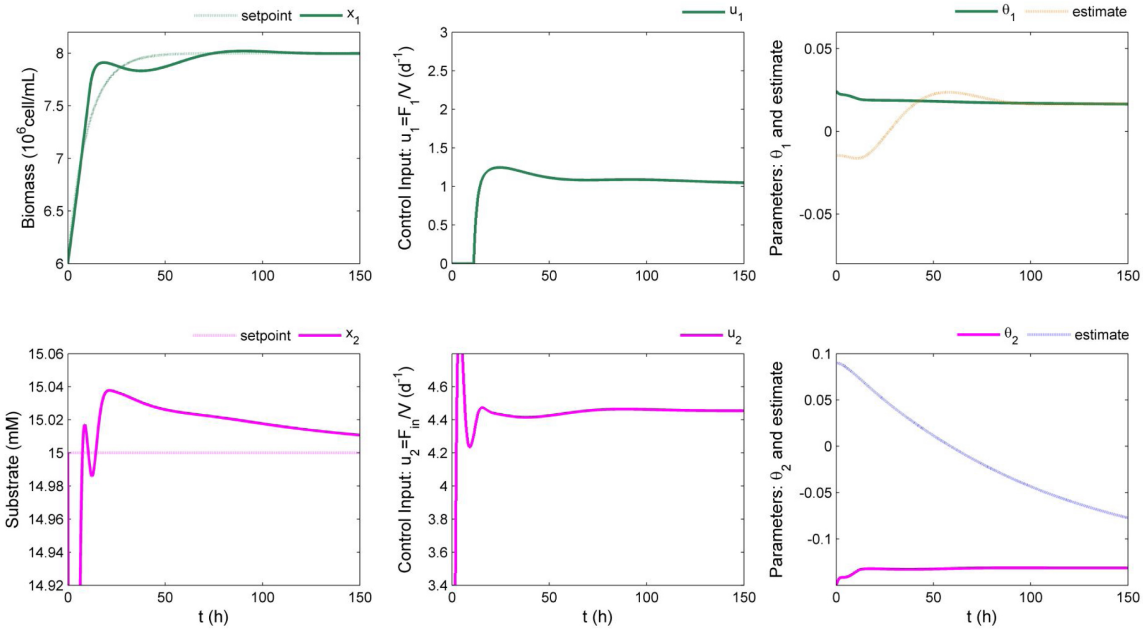


Figure 5.7: Backstepping control: closed-loop response with adaptation ($\hat{\theta} = -60\%\theta$) and tuning values listed in eq. (5.17).

5.2.3 Nonlinear model predictive control (NMPC)

Upon a setpoint change, a multivariable nonlinear model predictive controller will use simultaneously all model equations $dx_i/dt = f_i(x_i, u_j)$ in order to find the best trajectory of the manipulated variables u_j (Santos, 2001). It considers lower and upper bounds that represent natural process constraints.

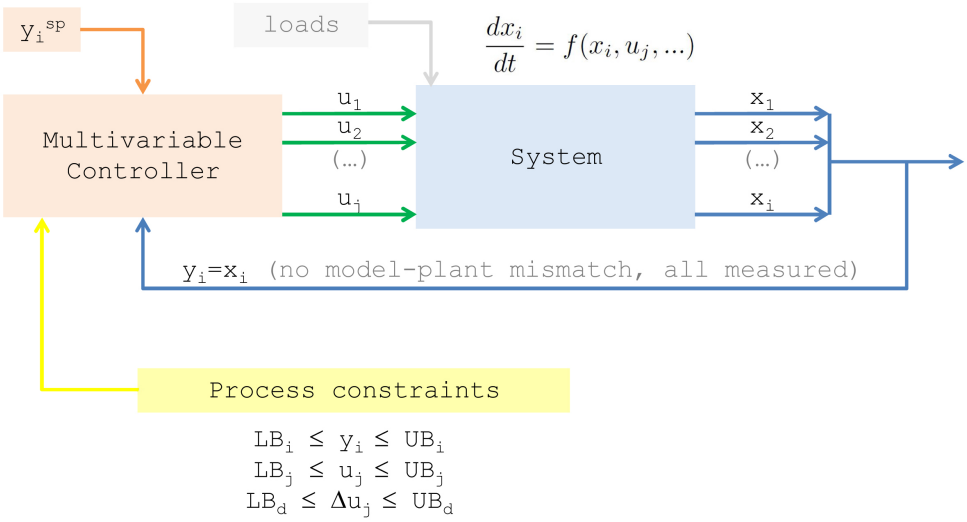


Figure 5.8: Model predictive controller.

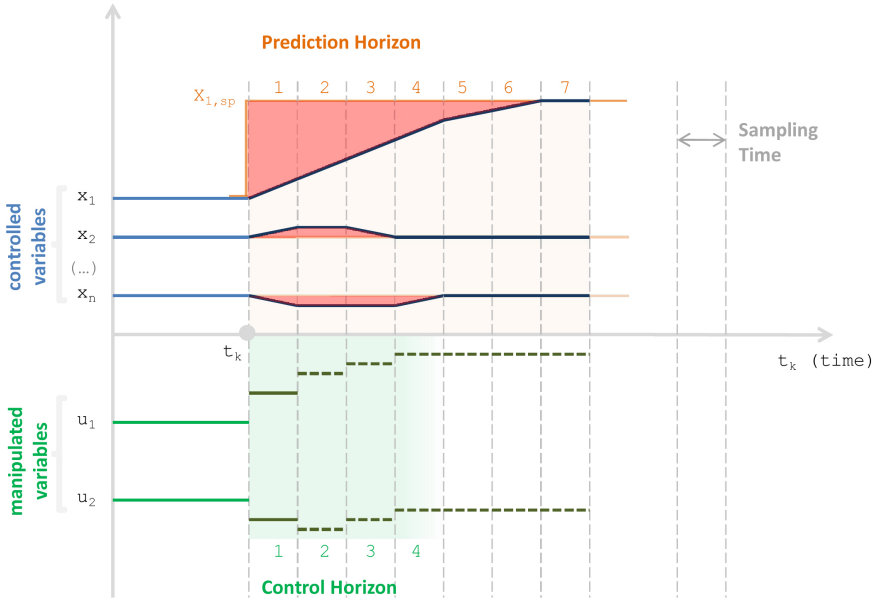


Figure 5.9: MPC: prediction and control horizons.

Figure 5.9 illustrates how, at each (discrete) timepoint t_k, the controller compares states to their setpoints and computes a sequence of control moves $u_{j,k}$ that will produce an evolution of system states within a prediction horizon leading to a better setpoint compliance.

The problem solved is open-loop (u_k is assumed constant during each interval). Therefore, the prediction horizon should be long enough to capture open-loop dynamics. $u_{i,ref}$ are the reference trajectories computed in t_k. Only the first control move (first interval) is applied. The remaining intervals are used as reference initial guesses for the algorithm in the following iteration t_{k+1}, where the open-loop problem will be solved again.

The optimal sequence of control moves can be computed through sequential quadratic programming by defining a cost function that is then minimised. The value of the cost function expresses a measure of how far predictions of systems states are to their setpoints and can also incorporate penalties regarding the amplitude of control moves, aggressiveness, and so on, through weighting factors w_i.

$$\text{minimize} \quad J = w_1\Phi_1 + w_2\Phi_2 + ... + w_t\Phi_t$$

$$\begin{aligned} \text{subject to} \quad & y_i^{min} \leq y_i \leq y_i^{max} \\ & u^{min} j \leq u_j \leq u_j^{max} \\ & \Delta u_j^{min} \leq \Delta u_j \leq \Delta u_j^{max} \end{aligned} \tag{5.18}$$

$$\begin{aligned} \text{setpoint compliance:} \quad & \Phi_1 \propto \sum (y - y_{ref})^2 \\ \text{control trajectory compliance:} \quad & \Phi_2 \propto \sum (u - u_{ref})^2 \\ \text{control move rate penalty:} \quad & \Phi_3 \propto \sum (\Delta u)^2 = \sum (u_k - u_{k-1})^2 \end{aligned} \tag{5.19}$$

Some trial and error is needed to avoid computationally intensive algorithms. Henson and Seborg (1997) and Agachi et al (2006) provide some rules of thumb:

- Sampling time Δt: small enough to capture system's dynamics (digital control theory can be applied) and complete open-loop computation for the next iteration;
- Prediction Horizon p: long enough to capture the system's natural open-loop dynamics (eg: after a step setpoint change, the time to reach a new steady state). Short horizons may lead to aggressive control and overshoot.

- Control Horizon m: smaller than prediction horizon. The smaller it is, the fewer degrees of freedom it will have (less chances of being aggressive). Example: $m \approx p/4$

- Ratio Prediction/Control Horizon (p,m). Example:

  $(p, m) = (8, 1)$ gentle controller, slower response, less sensitivity to disturbances

  $(p, m) = (8, 8)$ aggressive controller

- Cost function weight matrix: much of the above considerations for the prediction/control horizons can also be achieved through proper weighting.

In the case study, while the backstepping control scheme had two controllers, each acting individually on one state variable, in the NMPC approach, one controller will compute the best way of simultaneously manipulating the dilution ratios to control the two variables Xv and S, as illustrated in Figure 5.10. This is done by simultaneously integrating all model equations.

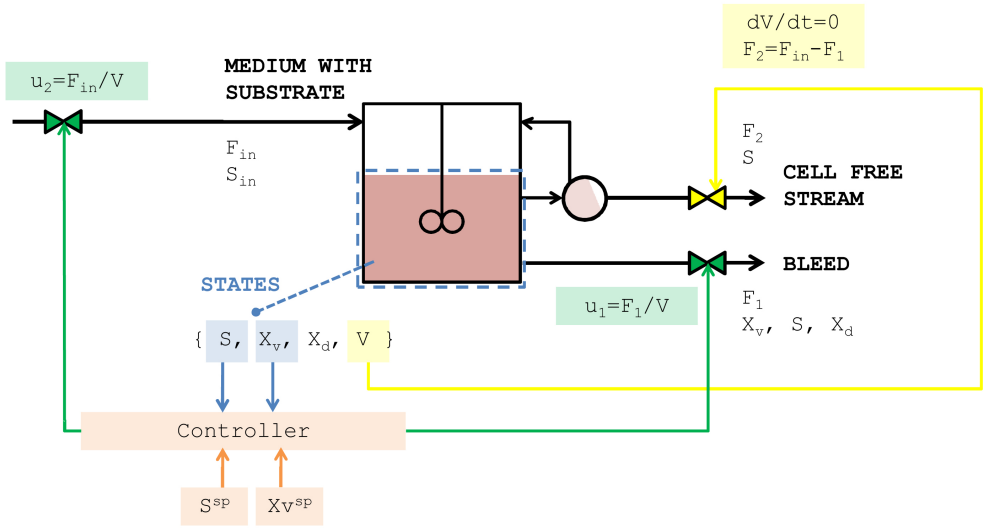


Figure 5.10: NMPC control applied to a bioreactor cell culture: layout for a continuous perfused regime operated at constant volume.

We have implemented the NMPC algorithm in Matlab as in (Santos et al, 2010). For the same setpoint change described previously in Figure 5.4 (p.

198), the response obtained is presented in Figure 5.11 for the following conditions:

$$\begin{aligned}
\text{sampling time:} \quad & \Delta t = 0.25h \ (\approx 15min) \\
\text{prediction horizon:} \quad & p = 8 \\
\text{control horizon:} \quad & m = 3 \\
\text{constraints:} \quad & u \geq 0 \\
\text{cost function} \quad & w_{y_1 - y_{1,ref}} = 1; w_{u_1 - u_{1,ref}} = 100; w_{u_{1,k} - u_{1,k-1}} = 10; \\
\text{weighting constants:} \quad & w_{y_2 - y_{2,ref}} = 10; w_{u_2 - u_{2,ref}} = 100; w_{u_{2,k} - u_{2,k-1}} = 5e4;
\end{aligned} \tag{5.20}$$

The response is good. When compared to the (best) backstepping control response for the same setpoint change (Fig. 5.6, p. 200), the NMPC controller uses slightly lower values of u_1 and u_2 (less dramatic changes of flows are involved in the corrective action).

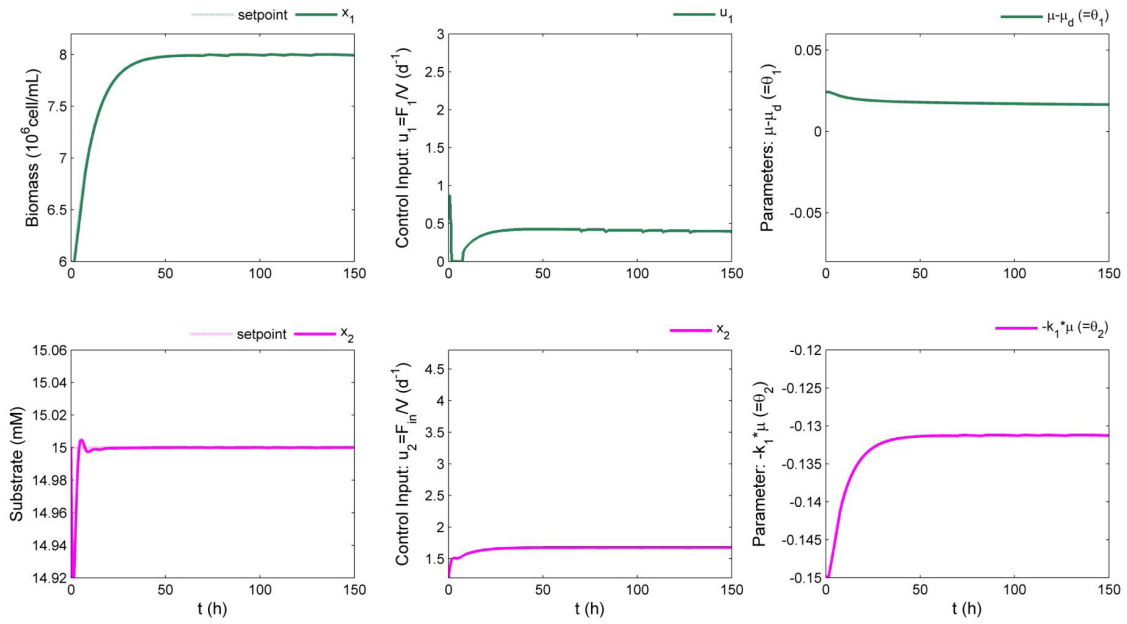


Figure 5.11: NMPC control: closed-loop response under conditions listed in eq. (5.20).

5.2.4 Comparison

In this case study a very simple model for continuous perfused animal cell cultures was used in two control approaches, backstepping and NMPC. A few comments can be made about the first:

- For given stable setpoints, the controllers can work well if the right set of parameters is found, although a systematic tuning procedure is not available.
- Lyapunov functions may be hard to find for more complex systems (models comprising 2 substrates and 2 metabolites, for instance).
- The coupling of manipulated and controlled variables in larger systems has to be chosen and imposed.
- Saturation in the manipulated variable u might cause loss of convergence/stability. Inclusion of process constraints can cause disturbances in controller performance.

The other approach tested, nonlinear model predictive control, seems to be a strategy with a better capacity of assimilating the multivariable nature of the process and incorporating its constraints in a very straightforward manner.

The NMPC also incorporates, very naturally in this case, some restrictions such as the non-negativeness of states (concentrations) and control inputs (dilution ratios/flowrates) and the eventuality of certain bounds on concentrations wanted (eg. if there is a prior knowledge of an optimal region of operation in terms of cell behavior).

It is the purpose of this thesis also to investigate the concept of the dynamic medium, by which the substrates that are fed to the cells would vary dynamically in order to be adjusted to their needs. This framework requires models where the impact of different substrates (and their metabolites) is evaluated. And this in turn means that the model will have a higher order and complexity. Also, we would like to incorporate several process constraints. Therefore, the NMPC strategy will be the one preferred in the scope of this thesis.

5.3 Case study: NMPC with a higher order model

Let it now be considered a real dynamic model of hybridoma cells producing monoclonal antibodies (de Tremblay et al, 1992). It comprises 7 states (biomass, glucose, lactate, glutamine, ammonia, monoclonal antibodies, volume), 16 parameters and describes typical animal cell culture phenomena. Its full description is provided in Section 3.5 on page 84.

The control scheme proposed is illustrated in Figure 5.12. The manipulation of two flowrates (feed and bleed) is proposed to the controller, who, with two degrees of freedom, controls two variables: biomass and glucose concentrations.

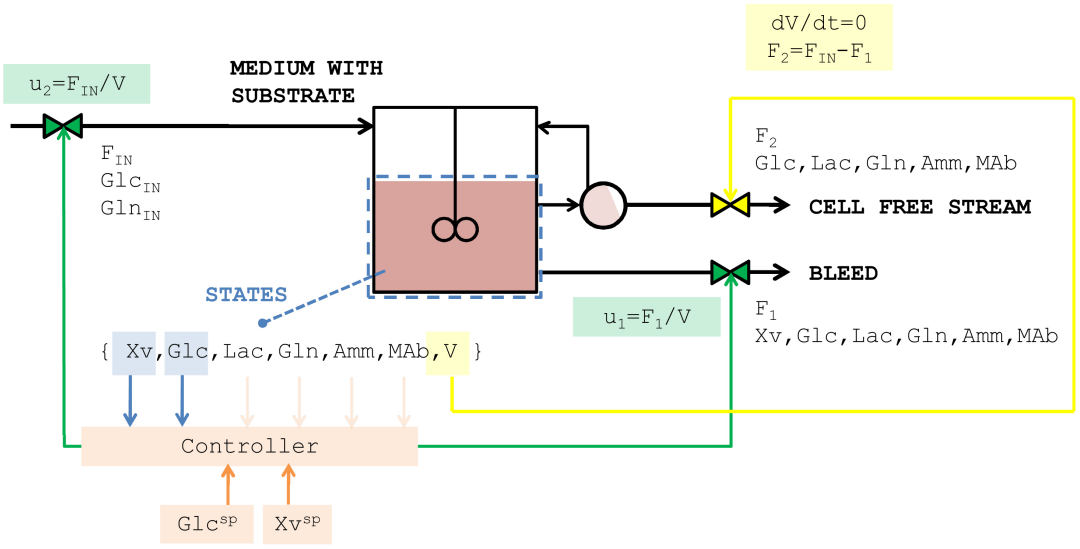


Figure 5.12: NMPC control applied to a bioreactor cell culture: layout for a continuous perfused regime operated at constant volume.

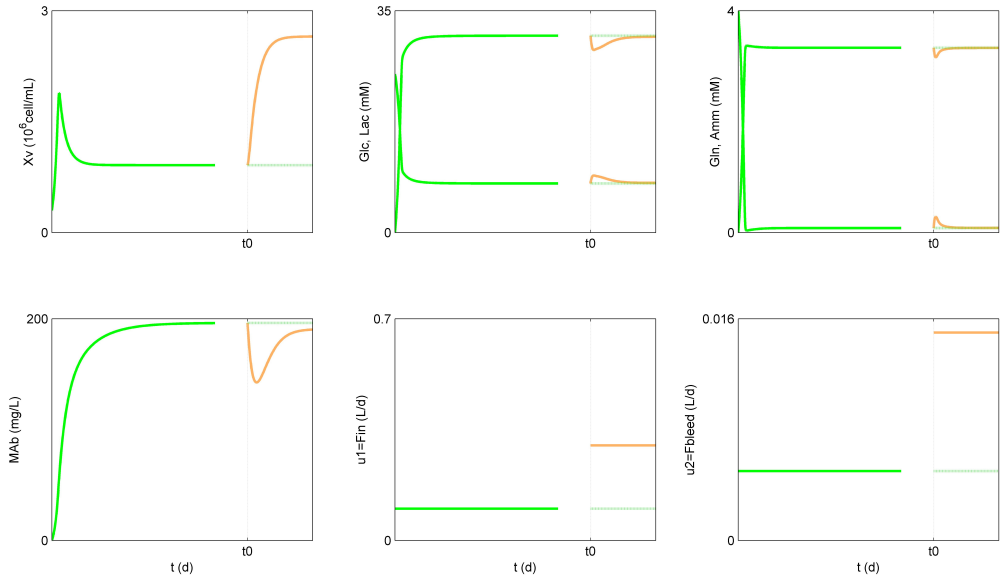


Figure 5.13: Continuous perfused bioreactor at constant volume: open-loop response to step changes in the feed and bleed flowrates (orange).

Let us suppose that the bioreactor is incubated and, after some time of operation in continuous perfused regime at constant volume, a stable point is attained. This is illustrated in green in Figure 5.13. If no further changes are made to F_{IN} and F_{bleed}, then all the concentrations will remain the same. However, if at $t = t_0$ step changes are introduced to the values of F_{IN} and F_{bleed}, then the system will evolve towards another equilibrium point. This open-loop response is plotted in orange.

A closed-loop response can be obtained by turning on a NMPC controller at t_0. Every Δt_{shift} this controller is called to compute the best line of action to drive the system to the new setpoint biomass and glucose concentration. This action is then implemented by adjusting the flowrates $u_1 = F_{IN}$ and $u_2 = F_{bleed}$. Gradually, after some time, compliance to the references is attained, as shown in Figure 5.14 for the conditions listed in eq. (5.21). Therefore, even if it is possible to compute the values of F_{IN} and F_{bleed} leading to specific values of biomass and glucose concentrations, the NMPC controller should be able to drive the system to these values faster (and automatically).

$$\begin{aligned}
\text{controller called every :} \quad & \Delta t_{shift} = 0.25d \quad \text{(4times/day)} \\
\text{sampling time:} \quad & \Delta t = 2d \\
\text{prediction horizon:} \quad & p = 6 \quad \text{(12d moving window)} \\
\text{control horizon:} \quad & m = 2 \\
\text{constraints:} \quad & u \geq 0.001L/d \\
& u \leq 2L/d \\
\text{cost function} \quad & w_{y_1 - y_{1,ref}} = 100 \\
\text{weighting constants:} \quad & w_{y_2 - y_{2,ref}} = 1
\end{aligned} \tag{5.21}$$

The NMPC control approach to the continuous perfused operation has been further studied with another implementation in Sbarciog et al (2013) and Saraiva et al (2011). Simulation studies showed better controller performances whenever nutrient concentrations ($x_2, x_4 \equiv Glc, Gln$) evolved in closed-loop nearer to the equilibrium line that relates them (equations (5.22) state these equilibrium relationships).

More precisely, in cases where nutrients were evolving closer to this *Glc*/*Gln* equilibrium line, the controller ended up needing less time to drive biomass to comply with its setpoint.

Through analysis of the equilibria (see p. 248), the exact relationship implicitly relating nutrients in equilibrium (x_2^*, x_4^*) can be found as stated in eq. (5.23). The equation can be approximated by a quadratic explicit description of x_4^*, which is easier to incorporate in the controller.

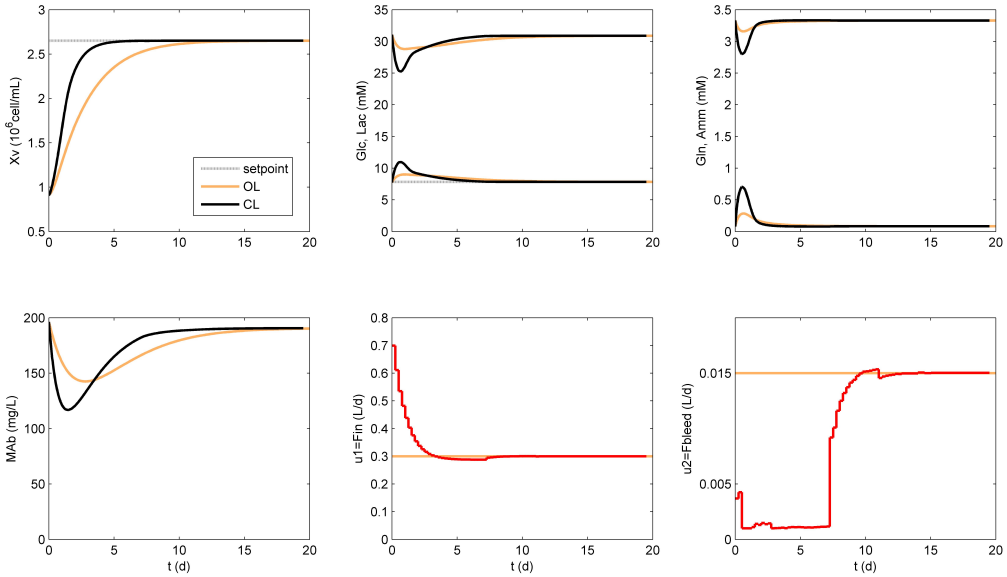


Figure 5.14: NMPC control: closed-loop response under conditions listed in eq. (5.21) (black and red lines) vs. open-loop (orange).

$$\text{Steady state:} \quad \begin{cases} 0 = & (k_{12}\frac{x_2}{k_7+x_2}\frac{x_4}{k_8+x_4})x_1 - (k_{11}\frac{1}{(k_{12}-k_6x_3)}\frac{1}{(k_{12}-k_{14}x_5)}\frac{k_5}{k_5+x_4})x_1 - \\ & -x_1u_1 + x_1u_2 \\ 0 = & -k_{14}(k_{12}\frac{x_2}{k_7+x_2}\frac{x_4}{k_8+x_4})x_1 - (k_{10}\frac{x_2}{k_9+x_2})x_1 \\ & -x_2u_1 + k_{18}u_1 \\ 0 = & k_{16}(k_{12}\frac{x_2}{k_7+x_2}\frac{x_4}{k_8+x_4})x_1 + k_{17}\left(k_{10}\frac{x_2}{k_9+x_2}\right)x_1 - x_3u_1 \\ 0 = & -k_{15}(k_{12}\frac{x_2}{k_7+x_2}\frac{x_4}{k_8+x_4})x_1 - x_4u_1 + k_{19}u_1 \\ 0 = & k_{13}(k_{12}\frac{x_2}{k_7+x_2}\frac{x_4}{k_8+x_4})x_1 - x_5u_1 \end{cases} \tag{5.22}$$

$$\frac{x_2^* - k_{18}}{x_4^* - k_{19}} = \frac{k_{14}}{k_{15}} + \frac{k_{10}\dfrac{x_2^*}{k_9 + x_2^*}}{k_{15}\left(k_{12}\dfrac{x_2^*}{k_7 + x_2^*}\dfrac{x_4^*}{k_8 + x_4^*}\right)} \tag{5.23}$$

Therefore, this knowledge was incorporated in the control law by adding a penalty term. This term is an assessment of how far actual glutamine concentration x_4 is to the value x_4^* that it should have, if it was in equilibrium for that particular value of glucose x_2. Results are further discussed in Sbarciog et al (2013) and Saraiva et al (2011).

5.4 CHO-320 case study: which model for control?

A very interesting question is which model to use for control purposes: how descriptive, detailed, complex or simple should it be? In order to address this, we will consider the many models identified for the CHO-320 cultures databank. We will presume that this cell line cultivated in this medium is perfectly described by model Ω_{5f+6a}. Other (less effective[1]) forms of describing it are models Ω_{5g}, Ω_{5e}, Ω_{4d} or Ω_3 (parameter values listed on p. 155-154). Table 5.1 lists the phenomena included in the model, its variables[2], and the phases for which they apply.

Table 5.1: Phenomena and variables included in different models identified for the CHO-320 cell line databank.

model	Glc lim.	Gln lim.	Lac inh.	Amm inh.	Glc mai.	Gln dec.	states	valid for phases
Ω_{5f+6a}	•	•	•	•	•	•	Xv,Glc,Lac,Gln,Amm,Prot	$A+B$
Ω_3							Xv,Glc,Lac,Gln,Amm	A
Ω_{4d}	•						Xv,Glc,Gln	$A+B$
Ω_{5e}	•	•	•	•			Xv,Glc,Lac,Gln,Amm	$A+B$
Ω_{5g}	•	•	•	•			Xv,Glc,Lac,Gln,Amm	$A+B$

Legend: lim. ≡ limitation, inh. ≡ inhibition, mai. ≡ maintenance, dec. ≡ spontaneous decomposition. In a batch, phase A is the early cell growth phase, phase B is the latter when cell concentration begins to decrease.

Figure 5.15 illustrates the timecourse of concentrations predicted by these models during a batch experiment[3].

Let us return now to the continuous perfused production regime where a NMPC controller is set out to keep biomass and glucose concentrations close to reference setpoints by manipulating the inflow and bleed outstream. Let us suppose that there will be no model-plant mismatch if the controller resorts to model Ω_{5f+6a} to compute the best line of action. On the other hand, if it considers one of the other possible models, there will be some mismatch since the real plant is described by Ω_{5f+6a}.

[1] In the sense that they obtained a lower score (value of the cost function used by the identification algorithm).

[2] Volume and bioproduct are left out of the analysis.

[3] Example chosen: under the exact conditions of experiment A of the databank.

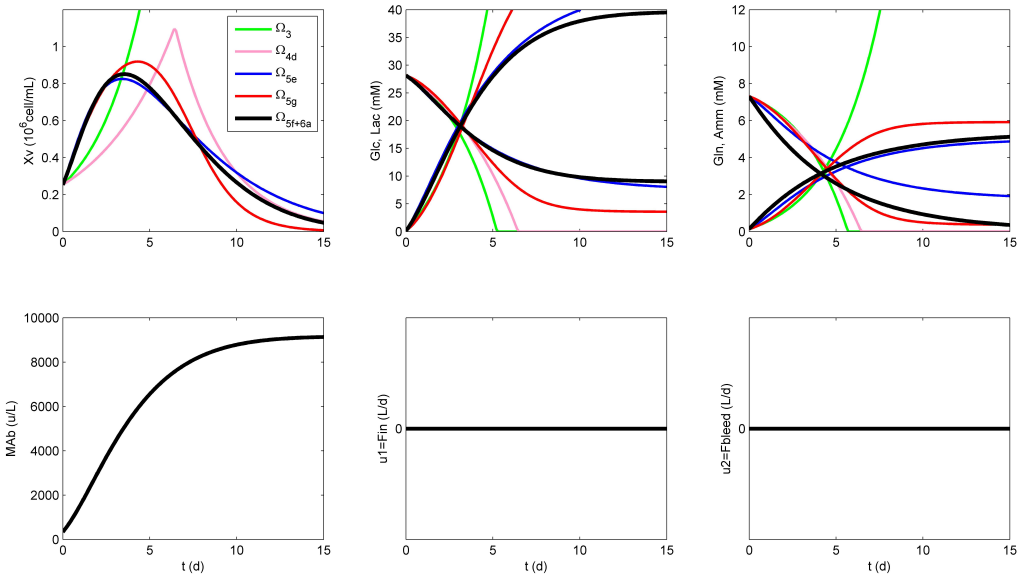


Figure 5.15: CHO-320: batch culture under conditions of experiment A of the databank. Description through several models. Ω_{5f+6a} (in grey) is considered the perfect description of the plant.

After some trial and error of tuning parameters (general procedure on p. 202), the following were chosen:

$$\begin{aligned}
\text{controller called every :} \quad & \Delta t_{shift} = 0.25d \quad \text{(4times/day)} \\
\text{sampling time:} \quad & \Delta t = 2d \\
\text{prediction horizon:} \quad & p = 6 \quad \text{(12d moving window)} \\
\text{control horizon:} \quad & m = 2 \\
\text{constraints:} \quad & u \geq 0.001L/d \\
& u \leq 2L/d \\
\text{cost function} \quad & w_{y_1 - y_{1,ref}} = 250 \\
\text{weighting constants:} \quad & w_{y_2 - y_{2,ref}} = 1
\end{aligned} \tag{5.24}$$

The controller's performance was tested in different circumstances (starting points and setpoint changes). Figures 5.16 and 5.17 illustrate two such examples. In all cases, the controller was able to drive states to their setpoints in a faster time than the open-loop response.

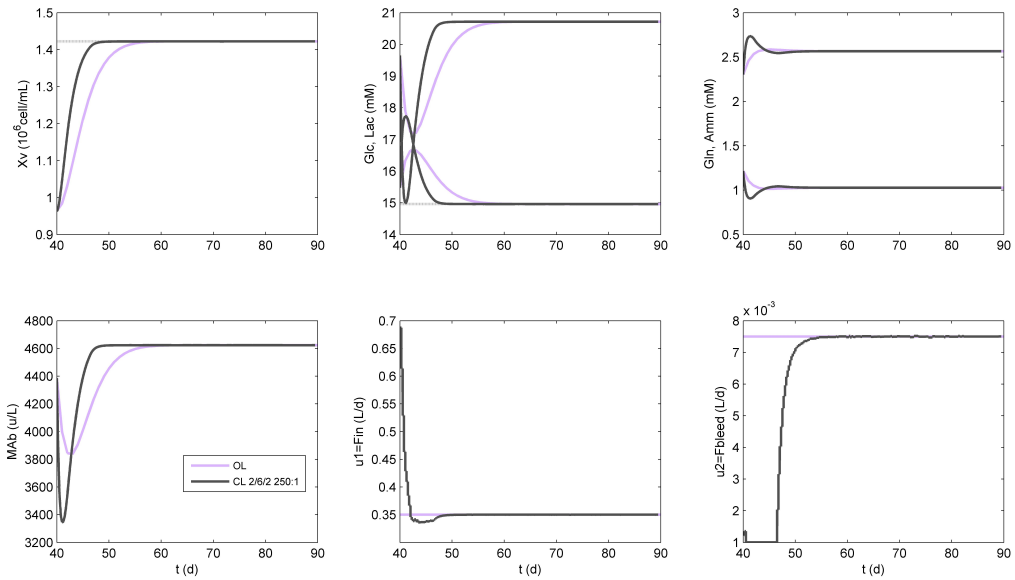


Figure 5.16: CHO-320 continuous perfused culture, NMPC control: closed-loop response under conditions listed in eq. (5.24) (black lines) vs. open-loop (lines in lilac). Setpoint changes: higher biomass with lower substrate.

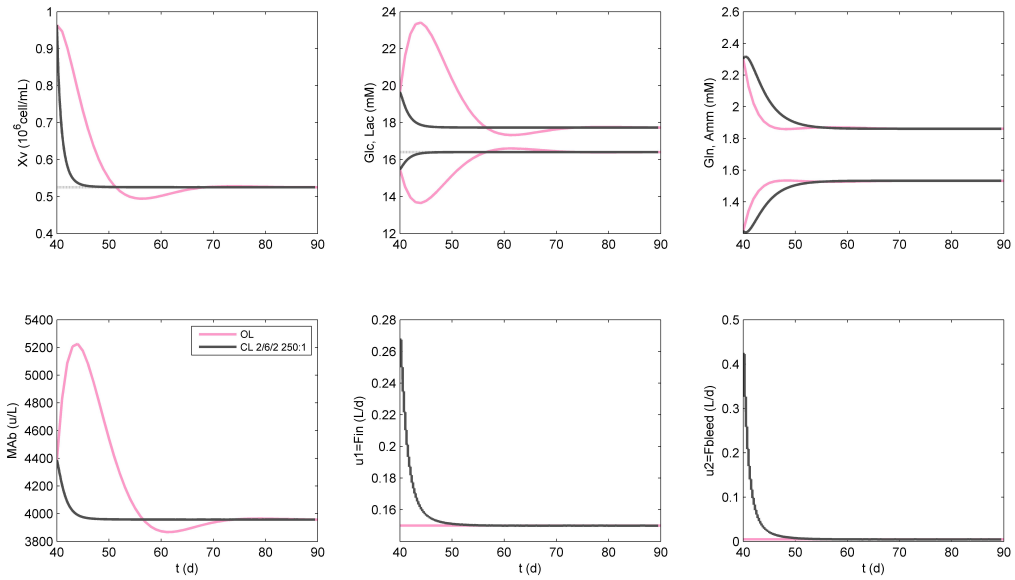


Figure 5.17: CHO-320 continuous perfused culture, NMPC control: closed-loop response under conditions listed in eq. (5.24) (black lines) vs. open-loop (lines in pink). Setpoint changes: lower biomass with higher substrate.

These figures express different types of setpoint changes. In fact, whenever a lower biomass concentration is wanted, the controller can act[1] essentially on the flow dynamics part of the model by adjusting the flowrates, which is rather fast (manipulation of valves). If a higher biomass concentration becomes the new setpoint, then, besides flow dynamics, the kinetic part of the model will also be particularly important. This part has slower dynamics (cells may need up to a day to divide, while valves can be manipulated in a matter of seconds). We will see two practical examples where the controller is asked to move from high to low biomass and the reverse (from points B1 to B2 and then from B2 to B1, as detailed on p. 251).

5.4.1 Lower biomass point

In closed-loop, the controller was asked to drive the system from an operation point to another operation point compliant with some biomass and glucose setpoints. In this case the cell concentration was to become lower. The real plant is described by model Ω_{5f+6}. The controller, in its optimization algorithm, considered either this model (no model-plant mismatch) or some other models also identified for the same experimental databank. In Figure 5.18 the no-mismatch case is represented in black and cases where model-plant mismatch exists are represented in green, pink, blue and red.

It can be seen that a biomass compliance is more or less attained even with the simpler models (Ω_3 and Ω_{4d} in green and pink). This is done at the expense of the use of more feeding medium (F_{IN} around 1.7L/d after 3 days) than what would in reality be needed (around 1L/d), leading to significant costs since operation in continuous perfused regime is meant to last very long (in the industry sometimes up to 3 months). It can also be seen that after 3 days, the system gets to a new steady state in all situations.

In what regards glucose, there is, for this setpoint change, a significant steady-state error when the plant's model is not correctly known by the controller (the exception being Ω_{5e} which is very close to the real one, Ω_{5f}, but doesn't consider phenomena such as glutamine decomposition and glucose for maintenance).

[1] Within operational constraints such as pump maximum allowed flowrate ($u_i = F_i \leq UB$), minimum flowrate intended for downstreaming purification (eg. $F_{perf} \geq LB$), minimum flowrate to prevent clogging in pipes (eg. $F_{bleed} \geq LB$), etc.

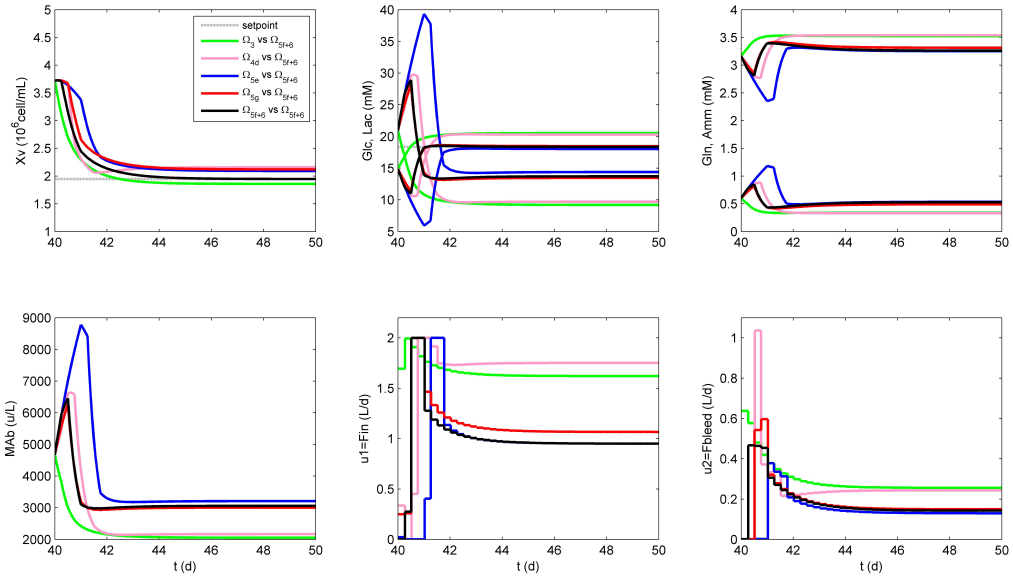


Figure 5.18: CHO-320 continuous perfused culture, NMPC control: closed-loop response when the controller knows the model perfectly (black) or when it uses different models (green, pink, blue and red lines). Setpoint change to lower biomass.

5.4.2 Higher biomass point

The reverse situation (that is, going now from low to high biomass) takes more time (around 6 days) since on top of adjusting flow dynamics, the kinetic part of the model (namely, cell division) plays an important role in trying to attain setpoint compliance. This is illustrated in Fig. 5.19.

Once more, the simpler models (green and pink) involve the use of more medium than necessary (2L/d instead of 1L/d in stationary phase). After some time, there is even saturation (eg. Ω_3, in green, predicts that a bigger F_{IN} would still be needed, whilst the pump is already at its maximum flowrate capacity).

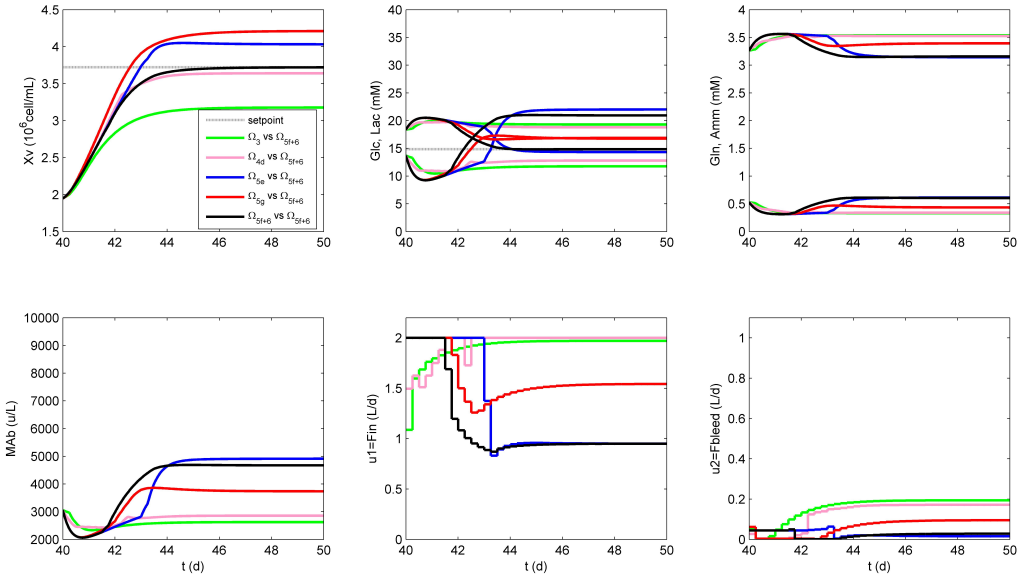


Figure 5.19: CHO-320 continuous perfused culture, NMPC control: closed-loop response when the controller knows the model perfectly (black) or when it uses different models (green, pink, blue and red lines). Setpoint change to higher biomass.

5.4.3 Some conclusions

This study has served to draw attention to the fact that knowing the kinetics of the model well is of particular importance for the control of continuous perfused cultures since it may lead to persistent excessive use of medium during a long operating time.

Besides this fact, a more obvious note is that if kinetics are not properly identified, then it is difficult to determine a priori the steady state conditions of the plant that would be optimal. These optimal conditions would allow to achieve, as best as possible, a certain goal, for example:

- high biomass concentration if the culture's purpose is to produce cells;
- high bioproduct concentration if bioproduct is the focus;
- the ideal bioproduct concentration that could be processed by the downstreaming purification unit.

Chapter 6

Conclusions and future prospects

The purposes of the work presented in this thesis are manifold and will be overviewed in the following sections.

6.1 Modeling and Identification

Firstly, it is to contribute to the mathematical modeling of animal cell cultures, in particular to its parametric identification. To this purpose, a step-by-step methodology has been proposed and illustrated in two experimental case studies that explore data from cultures that we have performed at the UMons, with two CHO cell lines using medium which is serum-free, a present industrial concern (Brunner et al, 2010).

Experimental planning We have proposed a guideline for an early-stage minimalist experimental planning where major cell behavior phenomena is present and could be captured:

- an experiment where low levels of substrate glucose limit cell growth while normal levels of glutamine may induce a build-up of corresponding metabolite ammonia,
- an experiment where low levels of substrate glutamine limit cell growth while normal levels of glucose may induce a build-up of corresponding metabolite lactate,

- an experiment where both substrates glucose and glutamine are over-abundant leading to significant metabolite inhibition phenomena,
- an experiment near the desired operation conditions.

This planning serves as a first collection of data. In reality, it is only after the experiments are performed that a better idea of the inhibition/limitation regions is possible.

Modeling Among the many choices of modeling approaches, we have opted for a macroscopic unstructured model capturing relevant cell behaviour so that the best compromise between complexity and usefulness can be achieved in what regards its implementation in a control scheme meant for continuous perfused cultures.

Identification: step-by-step procedure Instead of identifying the values of the many parameters of a complex animal cell culture model all at once, we have proposed a step-by-step methodology that allows for a gradual complexification of the model to identify. It is based on sensitivity analysis (namely the fact that some parameters are more influential at specific moments) and on model dynamics. Its main advantage is that, in each step, some knowledge is available about good initial values for some sets of parameters. For example:

- During an initial culture phase (phase A, while cell concentration is rising), parameters such as the stoichiometric coefficients and maximum growth rate are of major influence. Thus, several reduced order models can be easily identified.
- The knowledge of these values can serve as input for more complex models, now relating to the whole culture (both phase A, while cell concentration rises, and the latter phase B, while cell concentration decreases).
- In each of the steps, several reduced order models are possible. Looking carefully at the dynamics of complex macroscopic models, it is possible to reduce the number of states and/or parameters.

- While considering the entire culture time-series (phases A and B), the following parameters in terms of impact are those relating to the limitation of substrate, active around the switch time from phase A to phase B. It is important in this stage that the data bank illustrates this phenomena. Parameters relating to inhibition become active when phase B begins.

- Even if, globally speaking, some parameters will always tend to be more dominant than the remaining, one way of highlighting less influential parameters is introducing a timeweighting of the cost function. For example, by putting more weigh around t_{peak}, the impact of parameters such as the limitation constant (particularly active around t_{peak}) is amplified.

- Once this set of parameters is already roughly known, others can be added to the model form, but they will have less sensitivity (system states will be less influenced by changes in their values). So, in order to be more confident of their values, a bigger data bank is then convenient.

This step-by-step procedure has shown promising results. An illustration of better and more consistent results obtained with it, when compared with the traditional approach of trying to identify all parameters at once, is provided on page 165.

6.2 Observability

In real practice, it is rare to have the opportunity of measuring with online probes all the concentrations that represent states in the model. Therefore, in order to gather full information about the system, an observer can be built if observability/detectability properties can be guaranteed.

Observability/detectability study We have evaluated available approaches of doing so and found a method that, in these cases, can be a useful tool to reach global conclusions with a few analytical manipulations involved. This method lays its foundation on the analysis of the indistinguishable dynamics considering the system and a copy of it, and allows for the search of possible indistinguishable trajectories. We then applied this indistinguishable dynamics analysis to evaluate the observability/detectability property of a real animal cell model.

Case study A case study was presented, focusing on a recently available probe on the market that provides a quasi-continuous measurement of biomass of quality standard in what concerns cultures of animal cells (which are quite small in size) and presents a low-noise signal. An extended Kalman filter was then developed to observe all the other concentrations on the model not being measured. This full system knowledge (some variables are measured, others estimated) could be used in a control scheme where, for example, a model predictive controller uses this information to decide on how to act on the system.

6.3 Control

NMPC We have evidenced the usefulness of model predictive control in the scope of animal cell cultures. In particular, through its incorporation of real process constraints and acceptance of the multivariable nature of the process.

Higher order models The models stressed in this thesis intend to describe a full cell culture (all phases: growth and death). NMPC is able to handle this degree of information and we have illustrated it with case studies where the model incorporates many states (biomass, substrates, metabolites). The inclusion of these states is important because they help to express different possible phenomena taking place in a cell culture. This comprehensive model is also useful in determining interesting steady-state operation conditions of the continuous perfused culture.

Which model for control Model-plant mismatch occurs in real practice. We have highlighted the importance of having kinetics correctly captured by the model used in control in a continuous perfused production scenario. Among others, it may lead to a suboptimal use of substrate (implying waste of resources and extra costs). To pinpoint this, a case study was shown using different models identified for the same databank of interferon-γ producing CHO-320 cell cultures performed at the University of Mons.

6.4 Future prospects

It would be interesting to further continue this study with real data of continuous perfused cultures. The controller would then need to be coupled to an observer (eg. a Kalman filter algorithm such as that described previously) for a full system knowledge. Eventually, if in real practice process disturbances and noise prove to be limited, it may be interesting to provide the controller with a reduced-order model. For example, if, in a sufficiently big region around the steady state operating point, no ammonia inhibition is consistently verified, then this term could be eliminated from the controller's model and, therefore, the knowledge of this concentration would no longer be necessary in the loop. Ultimately, overlooking a phenomenon (thus, having the benefits of a simpler model) must be balanced with the risk of losing out in robustness.

Appendix A

Sensitivity equations terms

A.1 De Tremblay's model

The dynamics of the sensitivities for De Tremblay's model are obtained through equation (3.9), with the following terms:

$$\frac{\partial f_1}{\partial x_1} = z_{23}x_1 - \theta_4 z_{13}z_{14}z_{15} + \frac{u_{20}}{x_7} - \frac{u_{10}}{x_7}, \tag{A.1}$$

$$\frac{\partial f_1}{\partial x_2} = \theta_1 z_{12}x_1 z_{16}, \tag{A.2}$$

$$\frac{\partial f_1}{\partial x_3} = -\theta_4 z_{14}z_{15}x_1 z_{17}, \tag{A.3}$$

$$\frac{\partial f_1}{\partial x_4} = \theta_1 z_{11}x_1 z_{18} + \theta_4 z_{13}z_{14}x_1 z_{19}, \tag{A.4}$$

$$\frac{\partial f_1}{\partial x_5} = -\theta_4 z_{13}z_{15}x_1 z_{20}, \tag{A.5}$$

$$\frac{\partial f_1}{\partial x_6} = 0, \tag{A.6}$$

$$\frac{\partial f_1}{\partial x_7} = z_1 u_{10}\frac{x_1}{x_7^2} - z_3 u_{20}\frac{x_1}{x_7^2}, \tag{A.7}$$

$$\frac{\partial f_2}{\partial x_1} = -\left(\frac{z_{23}}{\theta_8} + \theta_{10}z_{21}\right), \tag{A.8}$$

$$\frac{\partial f_2}{\partial x_2} = -\left(\frac{\theta_1}{\theta_8} z_{12} x_1 z_{16} + \theta_{10} x_1 z_{22}\right) - \frac{u_{10}}{x_7}, \tag{A.9}$$

$$\frac{\partial f_2}{\partial x_3} = \frac{\partial f_2}{\partial x_5} = \frac{\partial f_2}{\partial x_6} = 0, \tag{A.10}$$

$$\frac{\partial f_2}{\partial x_4} = -\left(\frac{\theta_1}{\theta_8} z_{11} x_1 z_{18}\right), \tag{A.11}$$

$$\frac{\partial f_2}{\partial x_7} = z_1 u_{10} k_1 \left(\frac{-1}{x_7^2}\right) - z_1 u_{10} x_2 \left(\frac{-1}{x_7^2}\right), \tag{A.12}$$

$$\frac{\partial f_3}{\partial x_1} = \frac{\theta_9 \theta_1}{\theta_8} z_{11} z_{12}, \tag{A.13}$$

$$\frac{\partial f_3}{\partial x_2} = \frac{\theta_9 \theta_1}{\theta_8} z_{12} x_1 z_{16}, \tag{A.14}$$

$$\frac{\partial f_3}{\partial x_3} = -\frac{u_{10}}{x_7}, \tag{A.15}$$

$$\frac{\partial f_3}{\partial x_4} = \frac{\theta_9 \theta_1}{\theta_8} z_{11} x_1 z_{18}, \tag{A.16}$$

$$\frac{\partial f_3}{\partial x_5} = \frac{\partial f_3}{\partial x_6} = 0, \tag{A.17}$$

$$\frac{\partial f_3}{\partial x_7} = -z_1 u_{10} x_3 \left(\frac{-1}{x_7^2}\right), \tag{A.18}$$

$$\frac{\partial f_4}{\partial x_1} = -\frac{\theta_1}{\theta_{12}} z_{11} z_{12}, \tag{A.19}$$

$$\frac{\partial f_4}{\partial x_2} = -\frac{\theta_1}{\theta_{12}} z_{12} x_1 z_{16}, \tag{A.20}$$

$$\frac{\partial f_4}{\partial x_3} = \frac{\partial f_4}{\partial x_5} = \frac{\partial f_4}{\partial x_6} = 0, \tag{A.21}$$

$$\frac{\partial f_4}{\partial x_4} = -\frac{\theta_1}{\theta_{12}} z_{11} x_1 z_{18} - \frac{u_{10}}{x_7}, \tag{A.22}$$

$$\frac{\partial f_4}{\partial x_7} = z_1 u_{10} k_2 \left(\frac{-1}{x_7^2}\right) - z_1 u_{10} x_4 \left(\frac{-1}{x_7^2}\right), \tag{A.23}$$

$$\frac{\partial f_5}{\partial x_1} = \frac{\theta_{13}\theta_1}{\theta_{12}} z_{11} z_{12}, \tag{A.24}$$

$$\frac{\partial f_5}{\partial x_2} = \frac{\theta_{13}\theta_1}{\theta_{12}} z_{12} x_1 z_{16}, \tag{A.25}$$

$$\frac{\partial f_5}{\partial x_3} = \frac{\partial f_5}{\partial x_6} = 0, \tag{A.26}$$

$$\frac{\partial f_5}{\partial x_4} = \frac{\theta_{13}\theta_1}{\theta_{12}} z_{11} x_1 z_{18}, \tag{A.27}$$

$$\frac{\partial f_5}{\partial x_5} = -\frac{u_1 0}{x_7}, \tag{A.28}$$

$$\frac{\partial f_5}{\partial x_7} = -z_1 u_{10} x_5 \frac{-1}{x_7^2}, \tag{A.29}$$

$$\frac{\partial f_5}{\partial x_1} = \frac{\theta_{13}\theta_1}{\theta_{12}} z_{11} z_{12}, \tag{A.30}$$

$$\frac{\partial f_6}{\partial x_1} = \theta_{14} + \theta_{15} \frac{z_{23}}{\theta_{16} + z_{23}}, \tag{A.31}$$

$$\frac{\partial f_6}{\partial x_2} = \theta_{15} x_1 \frac{z_{24}(\theta_{16} + z_{23}) - z_{23} z_{24}}{(\theta_{16} + z_{23})^2}, \tag{A.32}$$

$$\frac{\partial f_6}{\partial x_3} = \frac{\partial f_6}{\partial x_5} = 0, \tag{A.33}$$

$$\frac{\partial f_6}{\partial x_4} = \theta_{15} x_1 \frac{z_{25}(\theta_{16} + z_{23}) - z_{23} z_{25}}{(\theta_{16} + z_{23})^2}, \tag{A.34}$$

$$\frac{\partial f_6}{\partial x_6} = -\frac{u_{10}}{x_7}, \tag{A.35}$$

$$\frac{\partial f_6}{\partial x_7} = -z_1 u_{10} x_6 \left(\frac{-1}{x_7^2}\right), \tag{A.36}$$

$$\frac{\partial f_7}{\partial x_1} = \frac{\partial f_7}{\partial x_2} = \frac{\partial f_7}{\partial x_3} = \frac{\partial f_7}{\partial x_4} = \frac{\partial f_7}{\partial x_5} = \frac{\partial f_7}{\partial x_6} = \frac{\partial f_7}{\partial x_7} = 0, \tag{A.37}$$

$$\frac{\partial f_1}{\partial \theta_1} = z_{26}, \tag{A.38}$$

$$\frac{\partial f_1}{\partial \theta_2} = -\theta_1 z_{12} x_1 z_{27}, \tag{A.39}$$

$$\frac{\partial f_1}{\partial \theta_3} = -\theta_1 z_{11} x_1 z_{28}, \tag{A.40}$$

$$\frac{\partial f_1}{\partial \theta_4} = -z_{13} z_{14} z_{15} x_1, \tag{A.41}$$

$$\frac{\partial f_1}{\partial \theta_5} = -\theta_4 z_{13} z_{15} x_1 \frac{(\theta_5 x_3 - \theta_1)}{(\theta_1 - \theta_5 x_3)^2}, \tag{A.42}$$

$$\frac{\partial f_1}{\partial \theta_6} = -\theta_4 z_{14} z_{15} x_1 \frac{(\theta_6 x_5 - \theta_1)}{(\theta_1 - \theta_6 x_5)^2}, \tag{A.43}$$

$$\frac{\partial f_1}{\partial \theta_7} = -\theta_4 z_{13} z_{14} x_1 z_{29}, \tag{A.44}$$

$$\frac{\partial f_1}{\partial \theta_c} = 0 \quad for\ c = 8, \ldots, 16, \tag{A.45}$$

$$\frac{\partial f_2}{\partial \theta_1} = \frac{-1}{\theta_8} z_{26}, \tag{A.46}$$

$$\frac{\partial f_2}{\partial \theta_2} = \frac{\theta_1}{\theta_8} z_{12} x_1 z_{27}, \tag{A.47}$$

$$\frac{\partial f_2}{\partial \theta_3} = \frac{\theta_1}{\theta_8} z_{28} x_1 z_{11}, \tag{A.48}$$

$$\frac{\partial f_2}{\partial \theta_c} = 0 \quad for\ c = 4, \ldots, 7, 9, 12, \ldots 16, \tag{A.49}$$

$$\frac{\partial f_2}{\partial \theta_8} = \frac{\theta_1}{\theta_8^2} z_{11} z_{12} x_1, \tag{A.50}$$

$$\frac{\partial f_2}{\partial \theta_{10}} = -x_2 \frac{x_1}{\theta_{11} + x_2}, \tag{A.51}$$

$$\frac{\partial f_2}{\partial \theta_{11}} = \theta_{10} x_1 \frac{x_2}{(\theta_{11} + x_2)^2}, \tag{A.52}$$

$$\frac{\partial f_3}{\partial \theta_1} = \frac{\theta_9}{\theta_8} z_{26}, \tag{A.53}$$

$$\frac{\partial f_3}{\partial \theta_2} = \frac{\theta_9 \theta_1}{\theta_8} z_{12} x_1 (-z_{27}), \tag{A.54}$$

$$\frac{\partial f_3}{\partial \theta_3} = \frac{\theta_9 \theta_1}{\theta_8} z_{11} x_1 (-z_{28}), \tag{A.55}$$

$$\frac{\partial f_3}{\partial \theta_c} = 0 \quad for\ c = 4, \ldots, 7, 10, \ldots 16, \tag{A.56}$$

$$\frac{\partial f_3}{\partial \theta_8} = \theta_9 \theta_1 z_{26} \frac{-1}{\theta_8^2}, \tag{A.57}$$

$$\frac{\partial f_3}{\partial \theta_9} = \frac{\theta_1}{\theta_8} z_{26}, \tag{A.58}$$

$$\frac{\partial f_4}{\partial \theta_1} = \frac{-1}{\theta_{12}} z_{26}, \tag{A.59}$$

$$\frac{\partial f_4}{\partial \theta_2} = \frac{-\theta_1}{\theta_1 2} z_{12} x_1 (-z_{27}), \tag{A.60}$$

$$\frac{\partial f_4}{\partial \theta_3} = \frac{-\theta_1}{\theta_{12}} z_{11} x_1 (-z_{28}), \tag{A.61}$$

$$\frac{\partial f_4}{\partial \theta_c} = 0 \quad for\ c = 4, \ldots, 11, 13, \ldots 16, \tag{A.62}$$

$$\frac{\partial f_4}{\partial \theta_{12}} = \frac{\theta_1}{\theta_{12}^2} z_{26}, \tag{A.63}$$

$$\frac{\partial f_5}{\partial \theta_1} = \theta_{13} \frac{\partial f_4}{\partial \theta_1}, \tag{A.64}$$

$$\frac{\partial f_5}{\partial \theta_2} = \theta_{13}\frac{\partial f_4}{\partial \theta_2}, \tag{A.65}$$

$$\frac{\partial f_5}{\partial \theta_3} = \theta_{13}\frac{\partial f_4}{\partial \theta_3}, \tag{A.66}$$

$$\frac{\partial f_5}{\partial \theta_c} = 0 \quad for\ c = 4, \ldots, 11, 14, \ldots 16, \tag{A.67}$$

$$\frac{\partial f_5}{\partial \theta_{12}} = -\theta_{13}\frac{\partial f_4}{\partial \theta_{12}},, \tag{A.68}$$

$$\frac{\partial f_5}{\partial \theta_{13}} = x_1\frac{z_{23}}{\theta_{12}}, \tag{A.69}$$

$$\frac{\partial f_6}{\partial \theta_1} = \frac{z_{26}\theta_{15}\theta_{16}}{(\theta_{16} + z_{23})^2}, \tag{A.70}$$

$$\frac{\partial f_6}{\partial \theta_2} = \frac{\theta_{15}x_1\theta_1 z_{12}(-z_{27})\theta_{16}}{(\theta_{16} + z_{23})^2}, \tag{A.71}$$

$$\frac{\partial f_6}{\partial \theta_3} = \frac{\theta_{15}x_1\theta_1 z_{11}(-z_{28})\theta_{16}}{(\theta_{16} + z_{23})^2}, \tag{A.72}$$

$$\frac{\partial f_6}{\partial \theta_c} = 0 \quad for\ c = 4, \ldots, 13, \tag{A.73}$$

$$\frac{\partial f_6}{\partial \theta_{14}} = x_1, \tag{A.74}$$

$$\frac{\partial f_6}{\partial \theta_{15}} = \frac{x_1 z_{23}}{(\theta_{16} + z_{23})}, \tag{A.75}$$

$$\frac{\partial f_6}{\partial \theta_{16}} = \frac{-x_1 z_{23}\theta_{15}}{(\theta_{16} + z_{23})^2}, \tag{A.76}$$

$$\frac{\partial f_7}{\partial \theta_c} = 0 \quad for\ c = 1, \ldots, 16, \tag{A.77}$$

with auxiliary equations:

$$z_1 = \begin{cases} 0 & \text{for batch,} \\ 1 & \text{for fedbatch/continuous/perfused,} \end{cases} \tag{A.78}$$

$$z_2 = \begin{cases} 0 & \text{for batch/fedbatch,} \\ 1 & \text{for continuous/perfused,} \end{cases} \tag{A.79}$$

$$z_3 = \begin{cases} 0 & \text{for batch/fedbatch/continuous,} \\ 1 & \text{for perfused,} \end{cases} \tag{A.80}$$

$$z_{11} = \frac{x_2}{x_2 + \theta_2}, \tag{A.81}$$

$$z_{12} = \frac{x_4}{x_4 + \theta_3}, \tag{A.82}$$

$$z_{13} = \frac{1}{\theta_1 - \theta_5 x_3}, \tag{A.83}$$

$$z_{14} = \frac{1}{\theta_1 - \theta_6 x_5}, \tag{A.84}$$

$$z_{15} = \frac{\theta_7}{x_4 + \theta_7}, \tag{A.85}$$

$$z_{16} = \frac{\theta_2}{(\theta_2 + x_2)^2}, \tag{A.86}$$

$$z_{17} = \theta_5 z_{13}^2, \tag{A.87}$$

$$z_{18} = \frac{\theta_3}{(\theta_3 + x_4)^2}, \tag{A.88}$$

$$z_{19} = \frac{\theta_7}{(\theta_7 + x_4)^2}, \tag{A.89}$$

$$z_{20} = \theta_6 z_{14}^2, \tag{A.90}$$

$$z_{21} = \frac{x_2}{x_2 + \theta_{11}}, \tag{A.91}$$

$$z_{22} = \frac{x_2}{x_2 + \theta_2}, \tag{A.92}$$

$$z_{23} = \theta_1 z_{11} z_{12}, \tag{A.93}$$

$$z_{24} = \theta_1 z_{12} z_{16}, \tag{A.94}$$

$$z_{25} = \theta_1 z_{11} z_{18}, \tag{A.95}$$

$$z_{26} = \theta_1 z_{11} z_{12}, \tag{A.96}$$

$$z_{27} = \frac{x_2}{(\theta_2 + x_2)^2}, \tag{A.97}$$

$$z_{28} = \frac{x_4}{(\theta_3 + x_4)^2}, \tag{A.98}$$

$$z_{29} = \frac{x_4}{(\theta_7 + x_4)^2}. \tag{A.99}$$

A.2 CHO-S cultures

A.2.1 Model Ω_1

The dynamics of the sensitivities for model Ω_1 are obtained through equation (3.9), with the following terms (for batch mode):

$$\frac{\partial f_1}{\partial x_1} = \theta_1, \tag{A.100}$$

$$\frac{\partial f_1}{\partial \theta_1} = x_1, \tag{A.101}$$

A.2.2 Model Ω_2

The dynamics of the sensitivities for model Ω_2 are obtained through equation (3.9), with the following terms (for batch mode):

$$\frac{\partial f_1}{\partial x_1} = \theta_1, \tag{A.102}$$

$$\frac{\partial f_1}{\partial x_2} = \frac{\partial f_1}{\partial x_4} = \frac{\partial f_2}{\partial x_2} = \frac{\partial f_2}{\partial x_4} = \frac{\partial f_4}{\partial x_2} = \frac{\partial f_4}{\partial x_4} = 0, \tag{A.103}$$

$$\frac{\partial f_2}{\partial x_1} = -\frac{1}{\theta_8}\theta_1, \tag{A.104}$$

$$\frac{\partial f_4}{\partial x_1} = -\frac{1}{\theta_{12}}\theta_1, \tag{A.105}$$

$$\frac{\partial f_1}{\partial \theta_1} = x_1, \tag{A.106}$$

$$\frac{\partial f_1}{\partial \theta_8} = \frac{\partial f_1}{\partial \theta_{12}} = \frac{\partial f_2}{\partial \theta_{12}} = \frac{\partial f_4}{\partial \theta_8} = 0, \tag{A.107}$$

$$\frac{\partial f_2}{\partial \theta_1} = -\frac{x_1}{\theta_8}, \tag{A.108}$$

$$\frac{\partial f_2}{\partial \theta_8} = \frac{\theta_1}{\theta_8}x_1, \tag{A.109}$$

$$\frac{\partial f_4}{\partial \theta_1} = \frac{-x_1}{\theta_{12}}, \tag{A.110}$$

$$\frac{\partial f_4}{\partial \theta_{12}} = \frac{\theta_1}{\theta_{12}}x_1, \tag{A.111}$$

A.2.3 Model Ω_3

The dynamics of the sensitivities for model Ω_3 are obtained through equation (3.9), with the following terms (for batch mode):

$$\frac{\partial f_1}{\partial x_1} = \theta_1, \tag{A.112}$$

$$\frac{\partial f_1}{\partial x_2} = \frac{\partial f_1}{\partial x_3} = \frac{\partial f_1}{\partial x_4} = \frac{\partial f_1}{\partial x_5} = 0, \tag{A.113}$$

$$\frac{\partial f_2}{\partial x_1} = -\frac{\theta_1}{\theta_8}, \tag{A.114}$$

$$\frac{\partial f_2}{\partial x_2} = \frac{\partial f_2}{\partial x_3} = \frac{\partial f_2}{\partial x_4} = \frac{\partial f_2}{\partial x_5} = 0, \tag{A.115}$$

$$\frac{\partial f_3}{\partial x_1} = \frac{\theta_1 \theta_9}{\theta_8}, \tag{A.116}$$

$$\frac{\partial f_3}{\partial x_2} = \frac{\partial f_3}{\partial x_3} = \frac{\partial f_3}{\partial x_4} = \frac{\partial f_3}{\partial x_5} = 0, \tag{A.117}$$

$$\frac{\partial f_4}{\partial x_1} = -\frac{\theta_1}{\theta_{12}}, \tag{A.118}$$

$$\frac{\partial f_4}{\partial x_2} = \frac{\partial f_4}{\partial x_3} = \frac{\partial f_4}{\partial x_4} = \frac{\partial f_4}{\partial x_5} = 0, \tag{A.119}$$

$$\frac{\partial f_5}{\partial x_1} = \frac{\theta_1 \theta_{13}}{\theta_{12}}, \tag{A.120}$$

$$\frac{\partial f_5}{\partial x_2} = \frac{\partial f_5}{\partial x_3} = \frac{\partial f_5}{\partial x_4} = \frac{\partial f_5}{\partial x_5} = 0, \tag{A.121}$$

$$\frac{\partial f_1}{\partial \theta_1} = x_1, \tag{A.122}$$

$$\frac{\partial f_1}{\partial \theta_8} = \frac{\partial f_1}{\partial \theta_9} = \frac{\partial f_1}{\partial \theta_{12}} = \frac{\partial f_1}{\partial \theta_{13}} = 0, \tag{A.123}$$

$$\frac{\partial f_2}{\partial \theta_1} = -\frac{x_1}{\theta_8}, \tag{A.124}$$

$$\frac{\partial f_2}{\partial \theta_1} = \frac{\theta_1 x_1}{\theta_8^2}, \tag{A.125}$$

$$\frac{\partial f_2}{\partial \theta_9} = \frac{\partial f_2}{\partial \theta_{12}} = \frac{\partial f_2}{\partial \theta_{13}} = 0, \tag{A.126}$$

$$\frac{\partial f_3}{\partial \theta_1} = \frac{\theta_9 x_1}{\theta_8}, \tag{A.127}$$

$$\frac{\partial f_3}{\partial \theta_8} = -\frac{\theta_1 \theta_9 x_1}{\theta_8^2}, \tag{A.128}$$

$$\frac{\partial f_3}{\partial \theta_9} = \frac{\theta_1 x_1}{\theta_8}, \tag{A.129}$$

$$\frac{\partial f_3}{\partial \theta_{12}} = \frac{\partial f_3}{\partial \theta_{13}} = 0, \tag{A.130}$$

$$\frac{\partial f_4}{\partial \theta_1} = -\frac{x_1}{\theta_{12}}, \tag{A.131}$$

$$\frac{\partial f_4}{\partial \theta_8} = \frac{\partial f_4}{\partial \theta_9} = \frac{\partial f_4}{\partial \theta_{13}} = 0, \tag{A.132}$$

$$\frac{\partial f_4}{\partial \theta_{12}} = \frac{\theta_1 x_1}{\theta_{12}^2}, \tag{A.133}$$

$$\frac{\partial f_5}{\partial \theta_1} = \frac{\theta_{13} x_1}{\theta_{12}}, \tag{A.134}$$

$$\frac{\partial f_5}{\partial \theta_8} = \frac{\partial f_5}{\partial \theta_9} = 0, \tag{A.135}$$

$$\frac{\partial f_5}{\partial \theta_{12}} = -\frac{\theta_1 \theta_{13} x_1}{\theta_{12}^2}, \tag{A.136}$$

$$\frac{\partial f_5}{\partial \theta_{13}} = \frac{\theta_1 x_1}{\theta_{12}}, \tag{A.137}$$

A.2.4 Model Ω_{5f}

The dynamics of the sensitivities for model Ω_{5f} are obtained through equation (3.9), with the following terms (for batch mode):

$$\frac{\partial f_1}{\partial x_1} = \frac{\theta_1 \theta_5 \theta_6 x_2 x_4}{(\theta_2 + x_2)(\theta_3 + x_4)(\theta_5 + x_3)(\theta_6 + x_5)} - \theta_4, \tag{A.138}$$

$$\frac{\partial f_1}{\partial x_2} = x_1 \frac{\theta_1 \theta_5 \theta_6 x_4}{(\theta_2 + x_2)(\theta_3 + x_4)(\theta_5 + x_3)(\theta_6 + x_5)} - x_1 \frac{\theta_1 \theta_5 \theta_6 x_2 x_4}{(\theta_2 + x_2)^2(\theta_3 + x_4)(\theta_5 + x_3)(\theta_6 + x_5)}, \tag{A.139}$$

$$\frac{\partial f_1}{\partial x_3} = -\frac{\theta_1\theta_5\theta_6 x_1 x_2 x_4}{(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)^2(\theta_6+x_5)}, \tag{A.140}$$

$$\frac{\partial f_1}{\partial x_4} = x_1\frac{\theta_1\theta_5\theta_6 x_2}{(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)} - \\ -x_1\frac{\theta_1\theta_5\theta_6 x_2 x_4}{(\theta_2+x_2)(\theta_3+x_4)^2(\theta_5+x_3)(\theta_6+x_5)}, \tag{A.141}$$

$$\frac{\partial f_1}{\partial x_5} = -\frac{\theta_1\theta_5\theta_6 x_1 x_2 x_4}{(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)^2};, \tag{A.142}$$

$$\frac{\partial f_1}{\partial x_6} = \frac{\partial f_2}{\partial x_6} = \frac{\partial f_3}{\partial x_6} = \frac{\partial f_4}{\partial x_6} = \frac{\partial f_5}{\partial x_6} = 0, \tag{A.143}$$

$$\frac{\partial f_2}{\partial x_1} = -\theta_{10} - \frac{\theta_1\theta_5\theta_6 x_2 x_4}{\theta_8(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)}, \tag{A.144}$$

$$\frac{\partial f_2}{\partial x_2} = -x_1\left(\frac{\theta_1\theta_5\theta_6 x_4}{\theta_8(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)} - \right. \\ \left. -\frac{\theta_1\theta_5\theta_6 x_2 x_4}{\theta_8(\theta_2+x_2)^2(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)}\right), \tag{A.145}$$

$$\frac{\partial f_2}{\partial x_3} = \frac{\theta_1\theta_5\theta_6 x_1 x_2 x_4}{\theta_8(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)^2(\theta_6+x_5)}, \tag{A.146}$$

$$\frac{\partial f_2}{\partial x_4} = -x_1\left(\frac{\theta_1\theta_5\theta_6 x_2}{\theta_8(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)} - \right. \\ \left. -\frac{\theta_1\theta_5\theta_6 x_2 x_4}{\theta_8(\theta_2+x_2)(\theta_3+x_4)^2(\theta_5+x_3)(\theta_6+x_5)}\right), \tag{A.147}$$

$$\frac{\partial f_2}{\partial x_5} = \frac{\theta_1\theta_5\theta_6 x_1 x_2 x_4}{\theta_8(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)^2}, \tag{A.148}$$

$$\frac{\partial f_3}{\partial x_1} = \theta_9\left[\theta_{10} + \frac{\theta_1\theta_5\theta_6\theta_9 x_2 x_4}{\theta_8(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)}\right], \tag{A.149}$$

$$\frac{\partial f_3}{\partial x_2} = \theta_9 x_1\left(\frac{\theta_1\theta_5\theta_6 x_4}{\theta_8(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)} - \right.$$

$$-\frac{\theta_1\theta_5\theta_6 x_2 x_4}{\theta_8(\theta_2+x_2)^2(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)}\Bigg), \tag{A.150}$$

$$\frac{\partial f_3}{\partial x_3} = -\frac{\theta_1\theta_5\theta_6\theta_9 x_1 x_2 x_4}{\theta_8(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)^2(\theta_6+x_5)}, \tag{A.151}$$

$$\frac{\partial f_3}{\partial x_4} = \theta_9 x_1\Bigg(\frac{\theta_1\theta_5\theta_6 x_2}{\theta_8(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)} - -\frac{\theta_1\theta_5\theta_6 x_2 x_4}{\theta_8(\theta_2+x_2)(\theta_3+x_4)^2(\theta_5+x_3)(\theta_6+x_5)}\Bigg), \tag{A.152}$$

$$\frac{\partial f_3}{\partial x_5} = -\frac{\theta_1\theta_5\theta_6\theta_9 x_1 x_2 x_4}{\theta_8(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)^2}, \tag{A.153}$$

$$\frac{\partial f_4}{\partial x_1} = -\frac{\theta_1\theta_5\theta_6 x_2 x_4}{\theta_{12}(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)}, \tag{A.154}$$

$$\frac{\partial f_4}{\partial x_2} = -x_1\Bigg(\frac{\theta_1\theta_5\theta_6 x_4}{\theta_{12}(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)} - -\frac{\theta_1\theta_5\theta_6 x_2 x_4}{\theta_{12}(\theta_2+x_2)^2(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)}\Bigg), \tag{A.155}$$

$$\frac{\partial f_4}{\partial x_3} = \frac{\theta_1\theta_5\theta_6 x_1 x_2 x_4}{\theta_{12}(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)^2(\theta_6+x_5)}, \tag{A.156}$$

$$\frac{\partial f_4}{\partial x_4} = -x_1\Bigg(\frac{\theta_7}{x_1} + \frac{\theta_1\theta_5\theta_6 x_2}{\theta_{12}(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)} - -\frac{\theta_1\theta_5\theta_6 x_2 x_4}{\theta_{12}(\theta_2+x_2)(\theta_3+x_4)^2(\theta_5+x_3)(\theta_6+x_5)}\Bigg), \tag{A.157}$$

$$\frac{\partial f_4}{\partial x_5} = \frac{\theta_1\theta_5\theta_6 x_1 x_2 x_4}{\theta_{12}(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)^2}, \tag{A.158}$$

$$\frac{\partial f_5}{\partial x_1} = \theta_{13}\Bigg(\frac{\theta_7 x_4}{x_1} + \frac{\theta_1\theta_5\theta_6 x_2 x_4}{\theta_{12}(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)}\Bigg) - \frac{\theta_{13}\theta_7 x_4}{x_1}, \tag{A.159}$$

$$\frac{\partial f_5}{\partial x_2} = \theta_{13} x_1 \left(\frac{\theta_1 \theta_5 \theta_6 x_4}{\theta_{12}(\theta_2 + x_2)(\theta_3 + x_4)(\theta_5 + x_3)(\theta_6 + x_5)} - \right.$$
$$\left. - \frac{\theta_1 \theta_5 \theta_6 x_2 x_4}{\theta_{12}(\theta_2 + x_2)^2(\theta_3 + x_4)(\theta_5 + x_3)(\theta_6 + x_5)} \right), \tag{A.160}$$

$$\frac{\partial f_5}{\partial x_3} = - \frac{\theta_1 \theta_5 \theta_6 \theta_{13} x_1 x_2 x_4}{\theta_{12}(\theta_2 + x_2)(\theta_3 + x_4)(\theta_5 + x_3)^2(\theta_6 + x_5)}, \tag{A.161}$$

$$\frac{\partial f_5}{\partial x_4} = \theta_{13} x_1 \left(\frac{\theta_7}{x_1} + \frac{\theta_1 \theta_5 \theta_6 x_2}{\theta_{12}(\theta_2 + x_2)(\theta_3 + x_4)(\theta_5 + x_3)(\theta_6 + x_5)} - \right.$$
$$\left. - \frac{\theta_1 \theta_5 \theta_6 x_2 x_4}{\theta_{12}(\theta_2 + x_2)(\theta_3 + x_4)^2(\theta_5 + x_3)(\theta_6 + x_5)} \right), \tag{A.162}$$

$$\frac{\partial f_5}{\partial x_5} = - \frac{\theta_1 \theta_5 \theta_6 \theta_{13} x_1 x_2 x_4}{\theta_{12}(\theta_2 + x_2)(\theta_3 + x_4)(\theta_5 + x_3)(\theta_6 + x_5)^2}, \tag{A.163}$$

$$\frac{\partial f_6}{\partial x_1} = \frac{\partial f_6}{\partial x_2} = \frac{\partial f_6}{\partial x_3} = \frac{\partial f_6}{\partial x_4} = \frac{\partial f_6}{\partial x_5} = \frac{\partial f_6}{\partial x_6} = 0, \tag{A.164}$$

$$\frac{\partial f_1}{\partial \theta_1} = \frac{\theta_5 \theta_6 x_1 x_2 x_4}{(\theta_2 + x_2)(\theta_3 + x_4)(\theta_5 + x_3)(\theta_6 + x_5)}, \tag{A.165}$$

$$\frac{\partial f_1}{\partial \theta_2} = - \frac{\theta_1 \theta_5 \theta_6 x_1 x_2 x_4}{(\theta_2 + x_2)^2(\theta_3 + x_4)(\theta_5 + x_3)(\theta_6 + x_5)}, \tag{A.166}$$

$$\frac{\partial f_1}{\partial \theta_3} = - \frac{\theta_1 \theta_5 \theta_6 x_1 x_2 x_4}{(\theta_2 + x_2)(\theta_3 + x_4)^2(\theta_5 + x_3)(\theta_6 + x_5)}, \tag{A.167}$$

$$\frac{\partial f_1}{\partial \theta_4} = -x_1, \tag{A.168}$$

$$\frac{\partial f_1}{\partial \theta_5} = x_1 \left(\frac{\theta_1 \theta_6 x_2 x_4}{(\theta_2 + x_2)(\theta_3 + x_4)(\theta_5 + x_3)(\theta_6 + x_5)} - \right.$$
$$\left. - \frac{\theta_1 \theta_5 \theta_6 x_2 x_4}{(\theta_2 + x_2)(\theta_3 + x_4)(\theta_5 + x_3)^2(\theta_6 + x_5)} \right), \tag{A.169}$$

$$\frac{\partial f_1}{\partial \theta_6} = x_1 \left(\frac{\theta_1 \theta_5 x_2 x_4}{(\theta_2 + x_2)(\theta_3 + x_4)(\theta_5 + x_3)(\theta_6 + x_5)} - \right.$$

$$-\frac{\theta_1\theta_5\theta_6 x_2 x_4}{(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)^2}\Bigg), \tag{A.170}$$

$$\frac{\partial f_1}{\partial \theta_7} = \frac{\partial f_1}{\partial \theta_8} = \frac{\partial f_1}{\partial \theta_9} = \frac{\partial f_1}{\partial \theta_{10}} = \frac{\partial f_1}{\partial \theta_{12}} = \frac{\partial f_1}{\partial \theta_{13}} = 0, \tag{A.171}$$

$$\frac{\partial f_2}{\partial \theta_1} = -\frac{\theta_5\theta_6 x_1 x_2 x_4}{\theta_8(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)}, \tag{A.172}$$

$$\frac{\partial f_2}{\partial \theta_2} = \frac{\theta_1\theta_5\theta_6 x_1 x_2 x_4}{\theta_8(\theta_2+x_2)^2(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)}, \tag{A.173}$$

$$\frac{\partial f_2}{\partial \theta_3} = \frac{\theta_1\theta_5\theta_6 x_1 x_2 x_4}{\theta_8(\theta_2+x_2)(\theta_3+x_4)^2(\theta_5+x_3)(\theta_6+x_5)}, \tag{A.174}$$

$$\frac{\partial f_2}{\partial \theta_4} = \frac{\partial f_2}{\partial \theta_9} = \frac{\partial f_2}{\partial \theta_{12}} = \frac{\partial f_2}{\partial \theta_{13}} = \frac{\partial f_2}{\partial \theta_7} = 0, \tag{A.175}$$

$$\frac{\partial f_2}{\partial \theta_5} = -x_1\Bigg(\frac{\theta_1\theta_6 x_2 x_4}{\theta_8(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)} - \frac{\theta_1\theta_5\theta_6 x_2 x_4}{\theta_8(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)^2(\theta_6+x_5)}\Bigg), \tag{A.176}$$

$$\frac{\partial f_2}{\partial \theta_6} = -x_1\Bigg(\frac{\theta_1\theta_5 x_2 x_4}{\theta_8(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)} - \frac{\theta_1\theta_5\theta_6 x_2 x_4}{\theta_8(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)^2}\Bigg), \tag{A.177}$$

$$\frac{\partial f_2}{\partial \theta_8} = \frac{\theta_1\theta_5\theta_6 x_1 x_2 x_4}{\theta_8^2(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)}, \tag{A.178}$$

$$\frac{\partial f_2}{\partial \theta_{10}} = -x_1, \tag{A.179}$$

$$\frac{\partial f_3}{\partial \theta_1} = \frac{\theta_5\theta_6\theta_9 x_1 x_2 x_4}{\theta_8(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)}, \tag{A.180}$$

$$\frac{\partial f_3}{\partial \theta_2} = -\frac{\theta_1\theta_5\theta_6\theta_9 x_1 x_2 x_4}{\theta_8(\theta_2+x_2)^2(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)}, \tag{A.181}$$

$$\frac{\partial f_3}{\partial \theta_3} = -\frac{\theta_1\theta_5\theta_6\theta_9 x_1 x_2 x_4}{\theta_8(\theta_2+x_2)(\theta_3+x_4)^2(\theta_5+x_3)(\theta_6+x_5)}, \tag{A.182}$$

$$\frac{\partial f_3}{\partial \theta_4} = \frac{\partial f_3}{\partial \theta_{12}} = \frac{\partial f_3}{\partial \theta_{13}} = \frac{\partial f_3}{\partial \theta_7} = 0, \tag{A.183}$$

$$\frac{\partial f_3}{\partial \theta_5} = \theta_9 x_1 \left(\frac{\theta_1\theta_6 x_2 x_4}{\theta_8(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)} - \right.$$
$$\left. - \frac{\theta_1\theta_5\theta_6 x_2 x_4}{\theta_8(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)^2(\theta_6+x_5)} \right), \tag{A.184}$$

$$\frac{\partial f_3}{\partial \theta_6} = \theta_9 x_1 \left(\frac{\theta_1\theta_5 x_2 x_4}{\theta_8(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)} - \right.$$
$$\left. - \frac{\theta_1\theta_5\theta_6 x_2 x_4}{\theta_8(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)^2} \right), \tag{A.185}$$

$$\frac{\partial f_3}{\partial \theta_8} = -\frac{\theta_1\theta_5\theta_6\theta_9 x_1 x_2 x_4}{\theta_8^2(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)}, \tag{A.186}$$

$$\frac{\partial f_3}{\partial \theta_9} = x_1 \left[\theta_{10} + \frac{\theta_1\theta_5\theta_6 x_1 x_2 x_4}{\theta_8(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)} \right], \tag{A.187}$$

$$\frac{\partial f_3}{\partial \theta_{10}} = \theta_9 x_1, \tag{A.188}$$

$$\frac{\partial f_4}{\partial \theta_1} = -\frac{\theta_5\theta_6 x_1 x_2 x_4}{\theta_{12}(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)}, \tag{A.189}$$

$$\frac{\partial f_4}{\partial \theta_2} = \frac{\theta_1\theta_5\theta_6 x_1 x_2 x_4}{\theta_{12}(\theta_2+x_2)^2(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)}, \tag{A.190}$$

$$\frac{\partial f_4}{\partial \theta_3} = \frac{\theta_1\theta_5\theta_6 x_1 x_2 x_4}{\theta_{12}(\theta_2+x_2)(\theta_3+x_4)^2(\theta_5+x_3)(\theta_6+x_5)}, \tag{A.191}$$

$$\frac{\partial f_4}{\partial \theta_4} = \frac{\partial f_4}{\partial \theta_8} = \frac{\partial f_4}{\partial \theta_9} = \frac{\partial f_4}{\partial \theta_{10}} = \frac{\partial f_4}{\partial \theta_{13}} = 0, \tag{A.192}$$

$$\frac{\partial f_4}{\partial \theta_5} = -x_1 \left(\frac{\theta_1\theta_6 x_2 x_4}{\theta_{12}(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)} - \right.$$

$$-\frac{\theta_1\theta_5\theta_6 x_2 x_4}{\theta_{12}(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)^2(\theta_6+x_5)}\Bigg), \tag{A.193}$$

$$\frac{\partial f_4}{\partial \theta_6} = -x_1\Bigg(\frac{\theta_1\theta_5 x_2 x_4}{\theta_{12}(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)}- -\frac{\theta_1\theta_5\theta_6 x_2 x_4}{\theta_{12}(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)^2}\Bigg), \tag{A.194}$$

$$\frac{\partial f_4}{\partial \theta_{12}} = \frac{\theta_1\theta_5\theta_6 x_1 x_2 x_4}{\theta_{12}^2(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)}, \tag{A.195}$$

$$\frac{\partial f_4}{\partial \theta_7} = -x_4, \tag{A.196}$$

$$\frac{\partial f_5}{\partial \theta_1} = \frac{\theta_5\theta_6\theta_{13} x_1 x_2 x_4}{\theta_{12}(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)}, \tag{A.197}$$

$$\frac{\partial f_5}{\partial \theta_2} = -\frac{\theta_1\theta_5\theta_6\theta_{13} x_1 x_2 x_4}{\theta_{12}(\theta_2+x_2)^2(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)}, \tag{A.198}$$

$$\frac{\partial f_5}{\partial \theta_3} = -\frac{\theta_1\theta_5\theta_6\theta_{13} x_1 x_2 x_4}{\theta_{12}(\theta_2+x_2)(\theta_3+x_4)^2(\theta_5+x_3)(\theta_6+x_5)}, \tag{A.199}$$

$$\frac{\partial f_5}{\partial \theta_4} = \frac{\partial f_5}{\partial \theta_8} = \frac{\partial f_5}{\partial \theta_9} = \frac{\partial f_5}{\partial \theta_{10}} = 0, \tag{A.200}$$

$$\frac{\partial f_5}{\partial \theta_5} = \theta_{13} x_1\Bigg(\frac{\theta_1\theta_6 x_2 x_4}{\theta_{12}(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)}- -\frac{\theta_1\theta_5\theta_6 x_2 x_4}{\theta_{12}(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)^2(\theta_6+x_5)}\Bigg), \tag{A.201}$$

$$\frac{\partial f_5}{\partial \theta_6} = \theta_{13} x_1\Bigg(\frac{\theta_1\theta_5 x_2 x_4}{\theta_{12}(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)}- -\frac{\theta_1\theta_5\theta_6 x_2 x_4}{\theta_{12}(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)^2}\Bigg), \tag{A.202}$$

$$\frac{\partial f_5}{\partial \theta_{12}} = -\frac{\theta_1\theta_5\theta_6\theta_{13} x_1 x_2 x_4}{\theta_{12}^2(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)}, \tag{A.203}$$

$$\frac{\partial f_5}{\partial \theta_{13}} = x_1 \left(\frac{\theta_7 x_4}{x_1} + \frac{\theta_1 \theta_5 \theta_6 x_2 x_4}{\theta_{12}(\theta_2 + x_2)(\theta_3 + x_4)(\theta_5 + x_3)(\theta_6 + x_5)} \right), \tag{A.204}$$

$$\frac{\partial f_5}{\partial \theta_7} = \theta_{13} x_4, \tag{A.205}$$

$$\frac{\partial f_6}{\partial \theta_1} = \frac{\partial f_6}{\partial \theta_2} = \frac{\partial f_6}{\partial \theta_3} = \frac{\partial f_6}{\partial \theta_4} = \frac{\partial f_6}{\partial \theta_5} = \frac{\partial f_6}{\partial \theta_6} = 0, \tag{A.206}$$

$$\frac{\partial f_6}{\partial \theta_8} = \frac{\partial f_6}{\partial \theta_9} = \frac{\partial f_6}{\partial \theta_{10}} = \frac{\partial f_6}{\partial \theta_{12}} = \frac{\partial f_6}{\partial \theta_{13}} = \frac{\partial f_6}{\partial \theta_7} = 0, \tag{A.207}$$

A.3 CHO-320 cultures

A.3.1 Model Ω_{5f}

Equations A.138 to A.148, A.150 to A.179, A.181 to A.186, A.189 to A.205 plus:

$$\frac{\partial f_3}{\partial x_1} = \frac{\theta_1 \theta_5 \theta_6 \theta_9 x_2 x_4}{\theta_8(\theta_2 + x_2)(\theta_3 + x_4)(\theta_5 + x_3)(\theta_6 + x_5)}, \tag{A.208}$$

$$\frac{\partial f_3}{\partial \theta_1} = \frac{\theta_1 \theta_5 \theta_6 \theta_9 x_1 x_2 x_4}{\theta_8(\theta_2 + x_2)(\theta_3 + x_4)(\theta_5 + x_3)(\theta_6 + x_5)}, \tag{A.209}$$

$$\frac{\partial f_3}{\partial \theta_9} = \frac{\theta_1 \theta_5 \theta_6 x_1 x_2 x_4}{\theta_8(\theta_2 + x_2)(\theta_3 + x_4)(\theta_5 + x_3)(\theta_6 + x_5)}, \tag{A.210}$$

$$\frac{\partial f_3}{\partial \theta_{10}} = \theta_9 x_1, \tag{A.211}$$

$$\frac{\partial f_i}{\partial x_7} = 0 \quad for\ i = 1, \ldots, 7, \tag{A.212}$$

$$\frac{\partial f_7}{\partial \theta_i} = 0 \quad for\ i = 1, \ldots, 13, \tag{A.213}$$

A.3.2 Model Ω_{6a}

$$\frac{\partial f_6}{\partial x_1} = \frac{\theta_1\theta_5\theta_6\theta_{15}x_2x_4}{(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)}, \tag{A.214}$$

$$\frac{\partial f_6}{\partial x_2} = \frac{\theta_1\theta_5\theta_6\theta_{15}x_1x_4}{(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)} - \\ - \frac{\theta_1\theta_5\theta_6\theta_{15}x_1x_2x_4}{(\theta_2+x_2)^2(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)}, \tag{A.215}$$

$$\frac{\partial f_6}{\partial x_3} = -\frac{\theta_1\theta_5\theta_6\theta_{15}x_1x_2x_4}{(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)^2(\theta_6+x_5)}, \tag{A.216}$$

$$\frac{\partial f_6}{\partial x_4} = \frac{\theta_1\theta_5\theta_6\theta_{15}x_1x_2}{(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)} - \\ - \frac{\theta_1\theta_5\theta_6\theta_{15}x_1x_2x_4}{(\theta_2+x_2)(\theta_3+x_4)^2(\theta_5+x_3)(\theta_6+x_5)}, \tag{A.217}$$

$$\frac{\partial f_6}{\partial x_5} = -\frac{\theta_1\theta_5\theta_6\theta_{15}x_1x_2x_4}{(\theta_2+x_2)(\theta_3+x_4)(\theta_5+x_3)(\theta_6+x_5)^2}, \tag{A.218}$$

$$\frac{\partial f_6}{\partial x_6} = \frac{\partial f_6}{\partial x_7} = 0, \tag{A.219}$$

$$\frac{\partial f_i}{\partial \theta_{15}} = 0 \quad for\ i = 1,\ldots,5,7, \tag{A.220}$$

$$\frac{\partial f_6}{\partial \theta_{15}} = \theta_1 \frac{x_2}{\theta_2+x_2}\frac{x_4}{\theta_3+x_4}\frac{\theta_5}{\theta_5+x_3}\frac{\theta_6}{\theta_6+x_5}x_1, \tag{A.221}$$

$$\frac{\partial f_7}{\partial x_i} = 0, \tag{A.222}$$

$$\frac{\partial f_7}{\partial \theta_{15}} = 0, \tag{A.223}$$

Appendix B

Observability analysis equations

B.1 General method of proving observability (rank condition)

B.1.1 Development of equations for a simple toy model

For the model provided in Section 4.2.2 (p. 171):

$$\text{Observability map:} \quad q = \begin{bmatrix} q_{1,1} \\ q_{1,2} \\ q_{1,3} \end{bmatrix} = \begin{bmatrix} L_f^0 h_1 \\ L_f^1 h_1 \\ L_f^2 h_1 \end{bmatrix} \tag{B.1}$$

$$q_{1,1} = L_f^0 h_x = h(x) = x_1 \tag{B.2}$$

$$q_{1,2} = L_f h_x = \begin{bmatrix} \frac{\partial h}{\partial x_1} & \dots & \frac{\partial h}{\partial x_3} \end{bmatrix} \times \begin{bmatrix} f_1(x_1, \dots, x_3) \\ \vdots \\ f_3(x_1, \dots, x_3) \end{bmatrix} =$$

$$= \left[\cancel{\frac{\partial h}{\partial x_1}}^{=1} f_1 + \cancel{\frac{\partial h}{\partial x_2}}^{=0} f_2 + \cancel{\frac{\partial h}{\partial x_3}}^{=0} f_3 \right] =$$

$$= \cancel{\frac{\partial}{\partial x_1}[x_1]}^{=1} \times f_1 = \overbrace{1 \times \left[\mu_{max}\frac{x_2}{k+x_2}\frac{x_3}{k+x_3}x_1 - \mu_{d,max}x_1\right]}^{=\Psi \text{ (auxiliary variable)}} = \frac{dx_1}{dt} \tag{B.3}$$

$$q_{1,3} = L_f^2 h_x = L_f[L_f h(x)] = L_f[\Psi] =$$

$$= \begin{bmatrix} \frac{\partial \Psi}{\partial x_1} & \cdots & \frac{\partial \Psi}{\partial x_3} \end{bmatrix} \times \begin{bmatrix} f_1(x_1, \ldots, x_3) \\ \vdots \\ f_3(x_1, \ldots, x_3) \end{bmatrix}$$

Since:

$$\frac{\partial \Psi}{\partial x_1} = \mu_{max}\frac{x_2}{k+x_2}\frac{x_3}{k+x_3} \times 1 - \mu_{d,max} \times 1 \tag{B.4}$$

$$\frac{\partial \Psi}{\partial x_2} = \mu_{max}\frac{x_3}{k+x_3}x_1\frac{k}{(k+x_2)^2} - 0 \tag{B.5}$$

$$\frac{\partial \Psi}{\partial x_3} = \mu_{max}\frac{x_2}{k+x_2}x_1\frac{k}{(k+x_3)^2} - 0 \tag{B.6}$$

Then:

$$q_{1,3} = \left(\frac{\partial \Psi}{\partial x_1}f_1 + \frac{\partial \Psi}{\partial x_2}f_2 + \frac{\partial \Psi}{\partial x_3}f_3\right) =$$

$$= \left(\mu_{max}\frac{x_2}{k+x_2}\frac{x_3}{k+x_3} - \mu_{d,max}\right)\left(\mu_{max}\frac{x_2}{k+x_2}\frac{x_3}{k+x_3}x_1 - \mu_{d,max}x_1\right) +$$

$$+ \left(\mu_{max}\frac{x_3}{k+x_3}x_1\frac{k}{(k+x_2)^2}\right)\left(-\nu\mu_{max}\frac{x_2}{k+x_2}\frac{x_3}{k+x_3}x_1\right) +$$

$$+ \left(\mu_{max}\frac{x_2}{k+x_2}x_1\frac{k}{(k+x_3)^2}\right)\left(-\nu\mu_{max}\frac{x_2}{k+x_2}\frac{x_3}{k+x_3}x_1\right) \tag{B.7}$$

To build the observability matrix:

$$O(x) = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \tag{B.8}$$

ones computes:

$$O(x) = \frac{\partial}{\partial x} q = \begin{bmatrix} \cancel{\frac{\partial q_{1,1}}{\partial x_1}}^{=1} & \cancel{\frac{\partial q_{1,1}}{\partial x_2}}^{=0} & \cancel{\frac{\partial q_{1,1}}{\partial x_3}}^{=0} \\ \frac{\partial q_{1,2}}{\partial x_2} & \frac{\partial q_{1,2}}{\partial x_2} & \frac{\partial q_{1,2}}{\partial x_3} \\ \frac{\partial q_{1,3}}{\partial x_3} & \frac{\partial q_{1,3}}{\partial x_2} & \frac{\partial q_{1,3}}{\partial x_3} \end{bmatrix} \tag{B.9}$$

with:

$$a_{11} = \frac{\partial q_{1,1}}{\partial x_1} = \frac{\partial}{\partial x_1}(x_1) = 1 \tag{B.10}$$

$$a_{12} = \frac{\partial q_{1,1}}{\partial x_2} = \frac{\partial}{\partial x_2}(x_1) = 0 \tag{B.11}$$

$$a_{13} = \frac{\partial q_{1,1}}{\partial x_3} = \frac{\partial}{\partial x_3}(x_1) = 0 \tag{B.12}$$

$$\begin{aligned} a_{21} = \frac{\partial q_{1,2}}{\partial x_1} = \frac{\partial \Psi}{\partial x_1} = \frac{\partial}{\partial x_1}\left[\mu_{max}\frac{x_2}{k+x_2}\frac{x_3}{k+x_3}x_1 - \mu_{d,max}x_1\right] = \\ = \mu_{max}\frac{x_2}{k+x_2}\frac{x_3}{k+x_3} \times 1 - \mu_{d,max} \times 1 \end{aligned} \tag{B.13}$$

$$a_{22} = \frac{\partial q_{1,2}}{\partial x_2} = \frac{\partial \Psi}{\partial x_2} = \mu_{max}\frac{x_3}{k+x_3}x_1\frac{k}{(k+x_2)^2} \tag{B.14}$$

$$a_{23} = \frac{\partial q_{1,2}}{\partial x_3} = \frac{\partial \Psi}{\partial x_3} = \mu_{max}\frac{x_2}{k+x_2}x_1\frac{k}{(k+x_3)^2} \tag{B.15}$$

$$a_{31} = \frac{\partial q_{1,3}}{\partial x_1} = \frac{\partial}{\partial x_1}\left[\frac{\partial \Psi}{\partial x_1}f_1 + \frac{\partial \Psi}{\partial x_2}f_2 + \frac{\partial \Psi}{\partial x_3}f_3\right] = \dots \tag{B.16}$$

$$a_{32} = \frac{\partial q_{1,3}}{\partial x_2} = \frac{\partial}{\partial x_2}\left[\frac{\partial \Psi}{\partial x_1}f_1 + \frac{\partial \Psi}{\partial x_2}f_2 + \frac{\partial \Psi}{\partial x_3}f_3\right] = \dots \tag{B.17}$$

$$a_{33} = \frac{\partial q_{1,3}}{\partial x_3} = \frac{\partial}{\partial x_3}\left[\frac{\partial \Psi}{\partial x_1}f_1 + \frac{\partial \Psi}{\partial x_2}f_2 + \frac{\partial \Psi}{\partial x_3}f_3\right] = \dots \tag{B.18}$$

It is evident that the equations will become longer and ever more complex. It is very difficult then to enquire about the matrix rank since it involves

working with gigantic equations. Let us then simply generically represent the matrix by:

$$O = \begin{bmatrix} 1 & 0 & 0 \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \tag{B.19}$$

B.2 Assessment of observability/detectability

B.2.1 Development of equations for case A (real cell culture model)

From equations (4.48) on p. 183 it follows that:

- ε_1 dynamics

$$\begin{cases} \varepsilon_1 = 0 \\ d\varepsilon_1/dt = 0 \end{cases} \Rightarrow 0 = \mu_1 x_1 - \mu_1^z x_1 - \mu_2 x_1 + \mu_2^z x_1 - 0 + 0 \Leftrightarrow \tag{B.20}$$

$$\Leftrightarrow 0 = (\mu_1 - \mu_1^z - \mu_2 + \mu_2^z) x_1 \tag{B.21}$$

$$\therefore x_1 = 0 \vee (\mu_2 - \mu_2^z) = (\mu_1 - \mu_1^z) \quad (\text{for } x_1 \neq 0) \tag{B.22}$$

- ε_2 dynamics

$$\begin{cases} \varepsilon_2 = 0 \\ d\varepsilon_2/dt = 0 \end{cases} \Rightarrow 0 = -k_{14}\mu_1 x_1 + k_{14}\mu_1^z x_1 - \mu_3 x_1 + \mu_3^z x_1 - 0 \Leftrightarrow \tag{B.23}$$

$$\Leftrightarrow 0 = (-k_{14}\mu_1 + k_{14}\mu_1^z - \mu_3 + \mu_3^z) x_1 \Leftrightarrow \tag{B.24}$$

Because $\mu_3 \equiv \mu_3(x_2)$ and $\mu_3^z \equiv \mu_3^z(x_2, \varepsilon_2)$, given that $\varepsilon_2 = 0$, then $\mu_3 = \mu_3^z$, thus:

$$\Leftrightarrow 0 = -k_{14}(\mu_1 - \mu_1^z) x_1 \tag{B.25}$$

Since $k_{14} = \nu_{21}$ is a positive stoichiometric constant:

$$\therefore x_1 = 0 \vee \mu_1 = \mu_1^z \quad (\text{for } x_1 \neq 0) \tag{B.26}$$

Since:

$$\mu_1 = \mu_1^z \Leftrightarrow \cancel{k_{12}} \frac{x_2}{k_7 + x_2} \frac{x_4}{k_8 + x_4} = \cancel{k_{12}} \frac{x_2 - 0}{k_7 + x_2 - 0} \frac{x_4 - \varepsilon_4}{k_8 + x_4 - \varepsilon_4} \Leftrightarrow \tag{B.27}$$

$$\Leftrightarrow \frac{x_4}{k_8 + x_4} = \frac{x_4 - \varepsilon_4}{k_8 + x_4 - \varepsilon_4} \tag{B.28}$$

is an equation of the generic form below:

$$\frac{a}{b} = \frac{a - \varepsilon}{b - \varepsilon} \Rightarrow \begin{array}{c} a(b - \varepsilon) = (a - \varepsilon)b \\ \cancel{ab} - a\varepsilon = \cancel{ab} - \varepsilon b \\ 0 = (a - b)\varepsilon \\ (a - b) = 0 \vee \varepsilon = 0 \end{array} \tag{B.29}$$

In this case:

$$\begin{array}{c} [(\cancel{x_4}) - (k_8 + \cancel{x_4})] = 0 \\ -k_8 = 0 \\ (\text{impossible} \\ \text{since } k_8 = k_{Gln}) \end{array} \quad \vee \varepsilon_4 = 0 \quad \checkmark \tag{B.30}$$

In fact all model constants (as model states) have to be positive in order to present a physical meaning.

- ε_3 dynamics

$$\frac{d\varepsilon_3}{dt} = k_{16}(\mu_1 - \mu_1^z)x_1 + k_{17}(\mu_3 - \mu_3^z)x_1 - \varepsilon_3 u_1 \tag{B.31}$$

It is known from eq. (B.23) and (B.30) that $\varepsilon_2 = \varepsilon_4 = 0$ and therefore:

$$\begin{cases} \mu_1 \equiv \mu_1(x_2, x_4) & = k_{12} \frac{x_2}{k_7 + x_2} \frac{x_4}{k_8 + x_4} \\ \mu_1^z \equiv \mu_1^z(x_2, x_4, \varepsilon_2, \varepsilon_4) & = k_{12} \frac{(x_2 - \varepsilon_2)}{k_7 + (x_2 - \varepsilon_2)} \frac{(x_4 - \varepsilon_4)}{k_8 + (x_4 - \varepsilon_4)} \end{cases} \xrightarrow{\varepsilon_2 = \varepsilon_4 = 0} \mu_1 = \mu_1^z \tag{B.32}$$

$$\begin{cases} \mu_3 \equiv \mu_3(x_2) & = k_{10} \frac{x_2}{k_9 + x_2} \\ \mu_3^z \equiv \mu_3^z(x_2, \varepsilon_2) & = k_{10} \frac{(x_2 - \varepsilon_2)}{k_9 + (x_2 - \varepsilon_2)} \end{cases} \xrightarrow{\varepsilon_2 = 0} \mu_3 = \mu_3^z \tag{B.33}$$

Equation (B.31) becomes then:

$$\frac{d\varepsilon_3}{dt} = 0 + 0 - \varepsilon_3 u_1 \tag{B.34}$$

The analytical solution is, as seen in equation (4.33) on p. 178:

$$\varepsilon_3(t) = \varepsilon_3(t_0)e^{-u_1 t} \tag{B.35}$$

ε_3, the error on the estimation of lactate concentration (x_3), will converge assymptotically to zero because the dilution ratio $u_1 = D$ is positive.

$$\lim_{t\to\infty} \varepsilon_3(t) = 0 \tag{B.36}$$

This is valid for all regimes (fedbatch, continuous and continuous perfused) except for the particular case of batch when $D = 0$ and thus $\varepsilon_3(t) = \varepsilon_3(t_0)$.

- ε_4 dynamics

From eq. (B.30) and eq. (B.32) it follows that:

$$\frac{d\varepsilon_4}{dt} = \cancel{-k_{15}\mu_1 x_1} + \cancel{k_{15}\mu_1^z(x_1 - 0)} - \cancel{\varepsilon_4 u_1} = 0 = 0 \tag{B.37}$$

$$\varepsilon_4(t) = 0 \tag{B.38}$$

x_4 can, thus, be observed from the knowledge of x_1 and x_2.

- ε_5 dynamics

From eq. (B.32) it follows that:

$$\frac{d\varepsilon_5}{dt} = \cancel{k_{13}\mu_1 x_1} - \cancel{k_{13}\mu_1^z(x_1 - 0)} - \varepsilon_5 u_1 \tag{B.39}$$

Again, the analytical solution is, as seen in equation (4.33) on p. 178:

$$\varepsilon_5(t) = \varepsilon_5(t_0)e^{-u_1 t} \tag{B.40}$$

ε_5, the error on the estimation of ammonia concentration (x_5), will converge assymptotically to zero because the dilution ratio $u_1 = D$ is positive.

$$\lim_{t\to\infty} \varepsilon_5(t) = 0 \tag{B.41}$$

This is valid for all regimes (fedbatch, continuous and continuous perfused) except for the particular case of batch when $D = 0$ and, thus, $\varepsilon_5(t) = \varepsilon_5(t_0)$.

B.2.2 Development of equations for case B (real cell culture model)

From equations (4.48) on p. 183 it follows that:

- ε_1 dynamics

$$\begin{cases} \varepsilon_1 = 0 \\ d\varepsilon_1/dt = 0 \end{cases} \Rightarrow 0 = \mu_1 x_1 - \mu_1^z x_1 - \mu_2 x_1 + \mu_2^z x_1 - 0 + 0 \Leftrightarrow \tag{B.42}$$

$$\Leftrightarrow \therefore x_1 = 0 \vee (\mu_1 - \mu_1^z) = (\mu_2 - \mu_2^z) \quad (\text{for } x_1 \neq 0) \tag{B.43}$$

- ε_2 dynamics

$$\frac{d\varepsilon_2}{dt} = \left[-k_{14}(\mu_1 - \mu_1^z) - (\mu_3 - \mu_3^z)\right] x_1 - \varepsilon_2 u_1 \tag{B.44}$$

- ε_3 dynamics

$$\frac{d\varepsilon_3}{dt} = \left[k_{16}(\mu_1 - \mu_1^z) + k_{17}(\mu_3 - \mu_3^z)\right] x_1 - \varepsilon_3 u_1 \tag{B.45}$$

- ε_4 dynamics

$$\frac{d\varepsilon_4}{dt} = \left[-k_{15}(\mu_1 - \mu_1^z)\right] x_1 - \varepsilon_4 u_1 \tag{B.46}$$

- ε_5 dynamics

$$\frac{d\varepsilon_5}{dt} = \left[k_{13}(\mu_1 - \mu_1^z)\right] x_1 - \varepsilon_5 u_1 \tag{B.47}$$

If the system is detectable or observable then this means that the following is the only attractive equilibrium point:

$$\begin{bmatrix} \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{B.48}$$

Let us check then the equilibrium points of the system:

$$\begin{cases} d\varepsilon_2/dt = f_2(\varepsilon) \\ d\varepsilon_3/dt = f_3(\varepsilon) \\ d\varepsilon_4/dt = f_4(\varepsilon) \\ d\varepsilon_5/dt = f_5(\varepsilon) \end{cases} \tag{B.49}$$

When $d\varepsilon_i/dt = 0$, it follows that:

$$\begin{cases} \mu_1' = \mu_2' \\ 0 = \left(-k_{14}\mu_1' - \mu_3'\right) x_1 - \varepsilon_2 u_1 \\ 0 = \left(k_{16}\mu_1' + k_{17}\mu_3'\right) x_1 - \varepsilon_3 u_1 \quad \text{where, for simplicity, } \mu_i' = \mu_i - \mu_i^z. \\ 0 = \left(-k_{15}\mu_1'\right) x_1 - \varepsilon_4 u_1 \\ 0 = \left(k_{13}\mu_1'\right) x_1 - \varepsilon_5 u_1 \end{cases} \tag{B.50}$$

It is not straightforward anymore, in this case, to apply the procedure described previously.

One might argue that, if $\mu_1' \neq 0$, then:

$$\begin{cases} \left.\begin{array}{r} \frac{\varepsilon_2 u_1}{-k_{14}\mu_1' - \mu_3'} = x_1 \\ \\ \frac{\varepsilon_3 u_1}{k_{16}\mu_1' + k_{17}\mu_3'} = x_1 \end{array}\right\} \rightarrow \quad \dfrac{\varepsilon_2}{(-k_{14}\mu_1' - \mu_3')} = \dfrac{\varepsilon_3}{(k_{16}\mu_1' + k_{17}\mu_3')} \quad \text{(for } u_1 \neq 0\text{)} \\ \\ \left.\begin{array}{r} \frac{\varepsilon_4 u_1}{-k_{15}\mu_1'} = x_1 \\ \\ \frac{\varepsilon_5 u_1}{k_{13}\mu_1'} = x_1 \end{array}\right\} \rightarrow \quad \dfrac{\varepsilon_4}{-k_{15}\mu_1'} = \dfrac{\varepsilon_5}{k_{13}\mu_1'} \quad \text{(for } u_1 \neq 0\text{)} \end{cases} \tag{B.51}$$

This means that actually there might be certain specific possible combinations of the errors leading to $d\varepsilon_i/dt = 0$, meaning that there are indistinguishable trajectories possible for the variables to estimate that co-exist with the same trajectory of the measured variable $x_1 = Xv$:

$$\frac{\varepsilon_2}{\varepsilon_3} = \frac{-k_{14}\mu_1' - \mu_3'}{k_{16}\mu_1' + k_{17}\mu_3'} \tag{B.52}$$

$$\frac{\varepsilon_4}{\varepsilon_5} = \frac{-k_{15}}{k_{13}} \tag{B.53}$$

In real practice the question to consider now is how often a case might occur where the observer's predictions at a given moment t_k would be such that, compared to values measured, $\varepsilon_i(t_k)/\varepsilon_j(t_k)$ would be exactly equal to the values described by equations (B.52) and (B.53). Throughout a culture, these situations might not even happen (at least persistently) since real measurements will have random errors, sometimes positive, others negative, with different amplitudes throughout time.

Appendix C

Additional information for control

C.1 Equilibrium equations for De Tremblay's model

Equilibrium points[1] x^* for a continuous perfused regime operated at constant volume can be computed using the equations on p. 208:

$$\text{Steady state:} \quad \begin{cases} 0 = & (k_{12}\frac{x_2^*}{k_7+x_2^*}\frac{x_4^*}{k_8+x_4^*})x_1^* - (k_{11}\frac{1}{(k_{12}-k_6x_3^*)}\frac{1}{(k_{12}-k_{14}x_5^*)}\frac{k_5}{k_5+x_4^*})x_1^* - \\ & -x_1^*u_1 + x_1^*u_2 \\ 0 = & -k_{14}(k_{12}\frac{x_2^*}{k_7+x_2^*}\frac{x_4^*}{k_8+x_4^*})x_1^* - (k_{10}\frac{x_2^*}{k_9+x_2^*})x_1^* \\ & -x_2^*u_1 + k_{18}u_1 \\ 0 = & k_{16}(k_{12}\frac{x_2^*}{k_7+x_2^*}\frac{x_4^*}{k_8+x_4^*})x_1^* + k_{17}\left(k_{10}\frac{x_2^*}{k_9+x_2^*}\right)x_1^* - x_3^*u_1 \\ 0 = & -k_{15}(k_{12}\frac{x_2^*}{k_7+x_2^*}\frac{x_4^*}{k_8+x_4^*})x_1^* - x_4^*u_1 + k_{19}u_1 \\ 0 = & k_{13}(k_{12}\frac{x_2^*}{k_7+x_2^*}\frac{x_4^*}{k_8+x_4^*})x_1^* - x_5^*u_1 \end{cases} \tag{C.1}$$

One equilibrium point is the washout point. This solution corresponds to the absence of biomass in the bioreactor at steady state and is not interesting:

[1] Antibodies concentration was excluded from the analysis. It does not impact on the dynamics of other states.

$$\text{Washout:} \quad \begin{cases} x_1^* = 0 \\ x_2^* = k_{18} \\ x_3^* = 0 \\ x_4^* = k_{19} \\ x_5^* = 0 \\ \forall u_1, u_2 \geq 0 \end{cases} \tag{C.2}$$

If biomass is existant at steady state ($x_1 > 0$), then the system will have equilibrium points defined by eq. (C.3) where:

$$\begin{cases} x_1^* > 0 \\ 0 < x_2^* < k_{18} \\ x_3^* \geq 0 \\ 0 < x_4^* < k_{19} \\ x_5^* \geq 0 \end{cases} \tag{C.3}$$

The kinetics are quite complex and it will generally not be possible to derive explicit analytical expressions of steady states. Nevertheless, some relationships can be found which could be interesting (eg. for the design of a controller).

One such expression relates the concentrations of substrates glucose and glutamine in equilibrium. Explicitating x_1^* in the second and fourth equations of system (C.1)[2]:

$$\begin{cases} 0 = \left(-k_{14}\mu_1^* - \mu_3(x)\right) & x_1^* - x_2^* u_1 + k_{18} u_1 \\ 0 = \left(-k_{15}\mu_1^*\right) & x_1^* - x_4^* u_1 + k_{19} u_1 \end{cases} \Leftrightarrow \tag{C.4}$$

$$\Leftrightarrow \begin{cases} x_1^* = \dfrac{x_2^* u_1 - k_{18} u_1}{-k_{14}\mu_1^* - \mu_3^*} \\ x_1^* = \dfrac{x_4^* u_1 - k_{19} u_1}{-k_{15}\mu_1^*} \end{cases} \tag{C.5}$$

Therefore:

$$u_1 \frac{(x_2^* - k_{18})}{-k_{14}\mu_1^* - \mu_3^*} = u_1 \frac{(x_4^* - k_{19})}{-k_{15}\mu_1^*} \Rightarrow \tag{C.6}$$

[2]Note that $\mu_i(x)$ are defined in equations (3.17)-(3.20) on p. 85.

$$\frac{x_2^* - k_{18}}{x_4^* - k_{19}} = \frac{-k_{14}\mu_1^* - \mu_3^*}{-k_{15}\mu_1^*} \tag{C.7}$$

This is thus the exact relationship relating glucose and glutamine concentration at steady state:

$$\frac{x_2^* - k_{18}}{x_4^* - k_{19}} = \frac{k_{14}}{k_{15}} + \frac{k_{10}\dfrac{x_2^*}{k_9 + x_2^*}}{k_{15}\left(k_{12}\dfrac{x_2^*}{k_7 + x_2^*}\dfrac{x_4^*}{k_8 + x_4^*}\right)} \tag{C.8}$$

C.2 CHO-320 Ω_{5f+6a} model

Several stable operating points have been computed for a constant volume continuous perfused culture of a CHO-320 cell line described by model Ω_{5f+6a} (identification results on p. 162). Figure C.1 illustrates some. The initial condition chosen corresponds to that of experiment A of the databank.

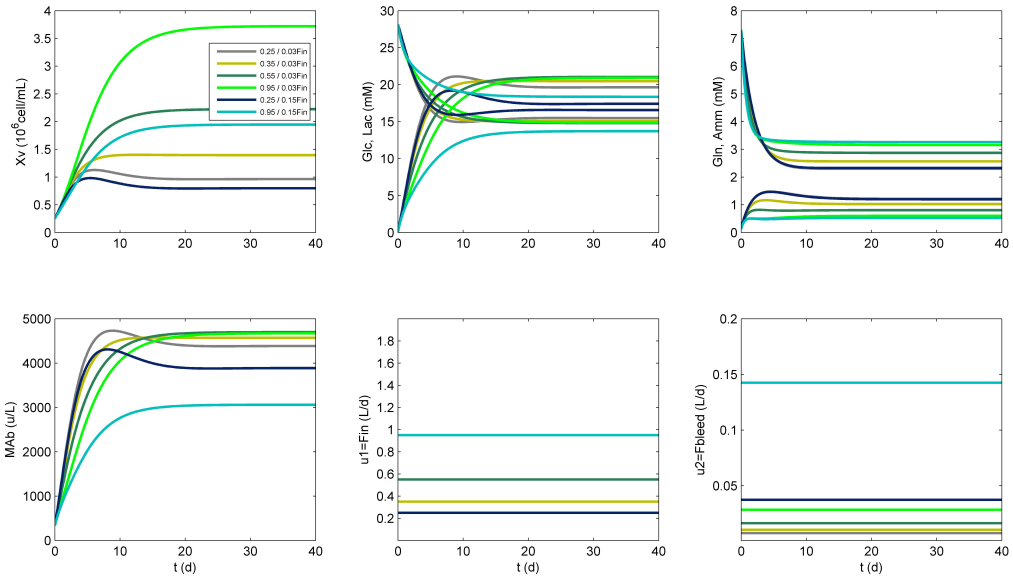


Figure C.1: CHO-320 culture described by model Ω_{5f+6a} in open-loop: several operation points departing from the initial condition of experiment A and using different values for the inflow rate F_{IN} and bleed outflow rate F_{bleed}. Regime: continuous perfused at constant volume.

For these simulations, the impact of various phenomena considered by the model is illustrated in Figure C.2. Different terms are listed on Table C.1. Lactate inhibition shows to be the most overwhelming phenomenon decreasing the potential maximum growth rate situation ($\mu_{growth} = \mu_{max} \times 1 \times 1 \times 1 \times 1$).

Table C.1: Terms in model Ω_{5f+6a}.

Term	Expression	Meaning
f_1	$\frac{Glc}{k_{Glc}+Glc}$	glucose limitation
f_2	$\frac{Gln}{k_{Gln}+Gln}$	glutamine limitation
f_3	$\frac{k_{Lac}}{k_{Lac}+Lac}$	lactate inhibition
f_4	$\frac{k_{Amm}}{k_{Amm}+Amm}$	ammonia inhibition
μ_{growth}	$\mu_{max} f_1 f_2 f_3 f_4$	growth rate
μ_{net}	$\mu_{growth} - \mu_{d,max}$	net growth rate

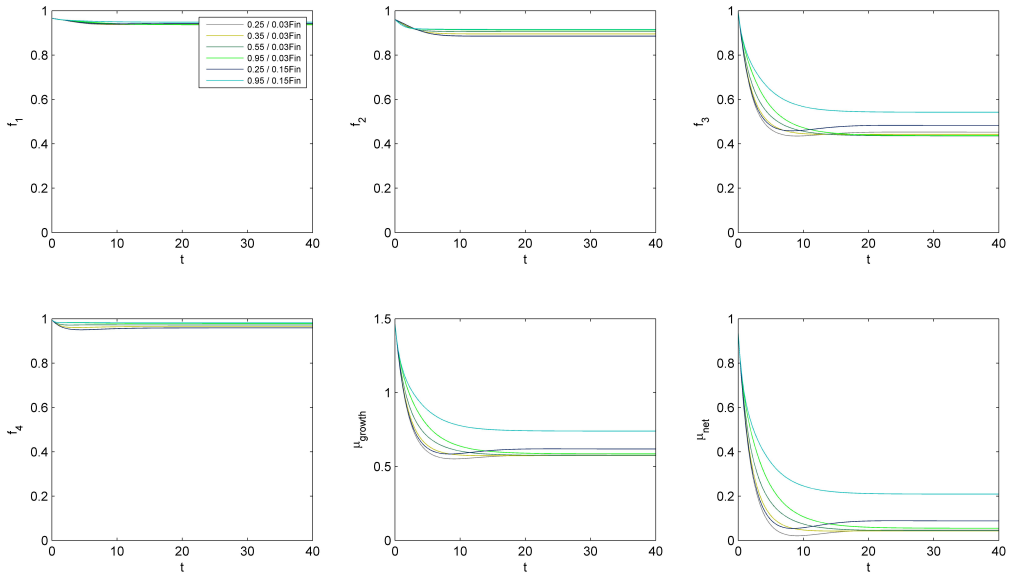


Figure C.2: CHO-320 culture described by model Ω_{5f+6a} in open-loop: impact of different phenomena considered by the model. Regime: continuous perfused at constant volume.

The two following points were chosen for the analysis of model-plant mismatch on page 212:

- Point B1 (lower biomass): obtained with $F_{IN} = 0.95$ and $F_{bleed} = 0.03F_{IN}$;
- Point B2 (higher biomass): obtained with $F_{IN} = 0.95$ and $F_{bleed} = 0.15F_{IN}$;

In Section 5.4.1 (p. 212) the system is driven in closed-loop from B1 to B2 (thus to a lower biomass concentration point) and in Section 5.4.2 (p. 213) from B2 to B1 (thus to a higher biomass concentration point).

Appendix D

Cell culture laboratory

The cell cultures used in the scope of this thesis were performed at the University of Mons in a newly set-up cell culture laboratory.

This laboratory is the joint effort of the Chemical and Biochemical Process Engineering Department and the Automatic Control Department of the Engineering Faculty of the University of Mons and was made possible through investment of the Engineering Faculty in overall laboratory equipment (Sartorius B-Plus Exclusive Flow 2-L bioreactor, ESCO reverse horizontal flow cabinet, Fedegari Autoklaven sterilizer, Beckman Coulter Allega X-15R centrifuge, Thermo Scientific Forma Series II incubator, Taylor-Wharton Liquid Nitrogen cryogenic freezer, furniture, refrigerators, water pre-treatment unit, gas and electrical emergency line installations, gas exhaust analyser, ...) and through support of the European Regional Development Fund (ERDF) for the 2007 to 2013 programming and the Belgian Walloon Region (for specific analysers such as the Waters UPLC-ELS/PDA, Flownamics Seg-flow automated online sterile sample collector, Bruker FT-NIR online spectrometer, Fogale Nanotech high resolution capacitive online probe, multifeed pump system, switch/link data management, daily laboratory material, etc). Other equipment such as spectrophotometers, HPLCs and RMN are available at other laboratories within the Chemical and Biochemical Process Engineering Department and at the General, Organic and Biomedical Chemistry Department of the University of Mons.

More detailed information on the set-up of this laboratory, its equipment and cultures performed can be found in Zamorano (2012).

Figure D.1: Animal cell culture laboratory at the University of Mons.

First row: UPLC, sample collector, bioreactor, Flownamics, pump system, Fogale probe, NIR, flow cabinet (with spinner flask culture); second row: centrifuge, fridge, incubator (with shake flask and t-flask cultures); third row: microscope, cell counter, cryogenic freezer; fourth row: autoclave sterilizer, bioreactor culture.

Bibliography

Aehle M, Kuprijanov A, Schaepe S, Simutis R, Lubbert A (2011) Increasing batch-to-batch reproducibility of cho cultures by robust open-loop control. Cytotechnology 63:41–47

Agachi P, Nagy Z, Cristea M, Imre-Lucaci A (2006) Model Based Control: Case Studies in Process Engineering. Wiley

Amribt Z (2014) Macoscopic modelling of hybridoma cell fed-batch cultures with overflow metabolism: model-based optimization and state estimation. PhD thesis, Universite Libre de Bruxelles

Astrom K, Wittenmark B (1997) Computer controlled systems: theory and design, 3rd edn. Prentice-Hall, USA

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

Bastin G (2008) Quantitative analysis of metabolic networks and design of minimal bioreaction models: a brief tutorial. Revue africaine de la recherche en informatique et mathématiques appliquées 9(1):41–55, URL http://perso.uclouvain.be/georges.bastin/paper89.pdf

Bastin G, Dochain D (1990) On-line Estimation and Adaptive Control of Bioreactors. Elsevier Science Publishers

Batt BC, Kompala DS (1989) A structured kinetic modeling framework for the dynamics of hybridoma growth and monoclonal antibody production in continuous suspension cultures. Biotechnology and Bioengineering 34(4):515–531

Bellu G, Saccomani M, Audoly S, Angio L (2007) Daisy: A new software tool to test global identifiability of biological and physiological systems. Computer Methods and Programs in Biomedicine 88:52–61

Boedeker BGD (2013) Recombinant Factor VIII (Kogenate®) for the Treatment of Hemophilia A: The First and Only World-Wide Licensed Recombinant Protein Produced in High-Throughput Perfusion Culture, Wiley-VCH Verlag GmbH & Co. KGaA, pp 429–443. DOI 10.1002/9783527669417.ch19, URL http://dx.doi.org/10.1002/9783527669417.ch19

Boghigian B, Seth G, Kiss R, Pfeifer B (2010) Metabolic flux analysis and pharmaceutical production. Metabolic Engineering 12(2):81–95

Bree M, Dhurjati P, Geoghegan R, Robnett B (1988) Kinetic modeling of hybridoma cell growth and immunoglobulin production in a large-scale suspensions culture. Biotechnology and Bioengineering 32:1067–1072

Brunner D, Frank J, Appl H, Schoffl H, Pfaller W, Gstraunthaler G (2010) Serum-free cell culture: The serum-free media interactive online database. Altex 27(1):53–62

Chen L, Nguang S, Chen X (2002) On-line identification and optimization of feed rate profiles for high productivity fed-batch culture of hybridoma cells using genetic algorithms. ISA Transactions 41(4):409–419

Chu L, Robinson D (2001) Industrial choices for protein production by large-scale cell culture. Current Opinion in Biotechnology 12(2):180–187

Dalili M, Sayles G, Ollis D (1990) Glutamine-limited batch hybridoma growth and antibody production: experiment and model. Biotechnol Bioen 36:74–82

De Maertelaer V (2014) Biostatistics - notes for pharmed (post-graduate programme in pharmaceutical medicine & medicines development sciences). Université Libre de Bruxelles, Brussels, Belgium

Deschenes J, Desbiens A, Perrier M, Kamen A (2006a) Multivariable nonlinear control of biomass and metabolite concentrations in a high-cell-density perfusion bioreactor. Industrial & Engineering Chemistry Research 45:8985–8997

Deschenes JS (2007) Commande non-lineaire des bioprocedes a des fins d'optimisation. PhD thesis, Universite Laval

Deschenes JS, Desbiens A, Perrier M, Kamen A (2006b) Use of cell bleed in a high cell density perfusion culture and multivariable control of biomass and

metabolite concentrations. Asia-Pacific Journal of Chemical Engineering 1(1-2):82–91

Dewasme L, Goffaux G, Hantson AL, Wouwer AV (2011) Extended kalman filter design for acetate estimation in e. coli cultures. Proceedings of the 18th IFAC World Congress 18

Dewasme L, Goffaux G, Hantson AL, Wouwer AV (2012) Experimental validation of an extended kalman filter estimating acetate concentrations in e. coli cultures. Journal of Process Control 23(2):148–157

Didion D, Norrant E, Hantson AL (2014) Comparison of nir and raman spectroscopies for in-line prediction of glucose, lactate and ammonia in a suspension culture of cho cells. In: EuroPACT

Dochain D, Chen L (1992) Local observability and controllability of stirred tank reactors. Journal of Process Control 2(3):139–144

Doyle F, Henson M (1996) Nonlinear process control. Chapter III Nonlinear Systems Theory, Prentice-Hall

Drugmand J (2011) Technologie des cellules animales. Cefochim Centre de formation aux métiers de production de l'Industrie Chimique et Pharmaceutique

Dunn I, Heinzle E, Ingham J, Prenosil J (2003) Biological Reaction Engineering: Dynamic Modelling Fundamentals with Simulation Examples. Wiley-VCH Verlag GmbH

EMA (2013) Raw materials for the production of cell-based and gene therapy products. Report of the Symposium organised by the European Directorate for the Quality of Medicines & HealthCare (EDQM), Council of Europe and the European Medicines Agency (EMA)

EMA (2014) Biological guidelines (home > human regulatory > scientific guidelines > biologicals). http://www.ema.europa.eu/, visited 17/12/2014

Eudralex (2003) Volume 4: Good manufacturing practice (gmp) guidelines. http://ec.europa.eu/health/documents/eudralex/vol-4/, visited 12/06/2015

Faraday D, Hayter P, Kirkby N (2001) A mathematical model of the cell cycle of a hybridoma cell line. Biochemical Engineering Journal 7(1):49–68

Fernandes S, Bastin G, Wouwer AV (2015) Metabolic flux analysis of hybridoma cells: underdetermined network and influence of batch and perfusion operating modes. Proceedings of the 8th Vienna International Conference on Mathematical Modelling

Fogale-nanotech (2013) Bioprocessing, animal cells, cho batch culture. http://www.fogale.fr/biotech/pages/appli11.php, visited 07/03/2013

Frame K, Hu W (1991a) Kinetic study of hybridoma cell growth in continuous culture. i. a model for non-producing cell. Biotechnology and Bioengineering 37:55–64

Frame K, Hu W (1991b) Kinetic study of hybridoma cell growth in continuous culture. ii. behaviour of producers and comparison to non-producers. Biotechnology and Bioengineering 38:1020–1028

Franco-Lara E, Weuster-Botz D (2005) Estimation of optimal feeding strategies for fed-batch bioprocesses. Bioprocess and Biosystems Engineering 27(4):255–262

Fredrickson A, McGee R, Tsuchiya H (1970) Mathematical models in fermentation processes. Advances in Applied Microbiology 23:419–426

Gaertner J, Dhurjati P (1993) Fractional factorial study of hybridoma behavior. 2. kinetics of nutrient uptake and waste production. Biotechnology Progress 9:309–316

Gauthier JP, Kupka I (1994) Observability and observers for nonlinear systems. SIAM Journal on Control and Optimization 32(4):975–994

Ghoul M, Dardenne M, Fonteix C, Marc A (1991) Extended kalman filtering technique for the on-line control of okt3 hybridoma cultures. Biotechnology Techniques 5(5):367–370

Glacken M, Huang C, Sinskey A (1989) Mathematical descriptions of hybridoma culture kinetics. iii simulation of fed-batch bioreactors. J Biotechnol 10:39–66

Glacken MW, Fleischaker RJ, Sinskey AJ (1986) Reduction of waste product excretion via nutrient control: Possible strategies for maximizing product and cell yields on serum in cultures of mammalian cells. Biotechnology and Bioengineering 28(9):1376–1389

Goffaux G (2010) Exploration of robust software sensor techniques with applications in vehicle positioning and bioprocess state estimation. PhD thesis, Universite de Mons

Goudar C, Heidemann R, Joeris K, Michaels J, Piret J, Konstantinov K (2005a) Generalized logistic equation modeling of mammalian cell batch cultures. In: Godia F, Fussenegger M (eds) Animal Cell Technology Meets Genomics, ESACT Proceedings, vol 2, Springer Netherlands, pp 601–604

Goudar C, Joeris K, Konstantinov K, Piret J (2005b) Logistic equations effectively model mammalian cell batch and fed-batch kinetics by logically constraining the fit. Biotechnology Progress 4:1109–18

Goudar C, Biener R, Piret J, Konstantinov K (2007) Animal Cell Biotechnology: Methods and Protocols, vol 14. Humana Press

Grosfils A, Wouwer AV, Bogaerts P (2007) On a general model structure for macroscopic biological reaction rates. Journal of Biotechnology 130:253–264

Haag J (2003) Dynamic modeling and state estimation of complex bioprocesses: theoretical issues and applications. PhD thesis, Universite de Mons

Harigae M, Matsumura M, Kataoka H (1994) Kinetic study on hbs-mab production in continuous cultivation. Journal of Biotechnology 34:227–235

Henson M, Seborg D (1997) Nonlinear Process Control. Prentice-Hall, USA

Hiller G, Aeschlimann A, Clark D, Blanc H (1991) A kinetic analysis of hybridoma growth and metabolism in continuous suspension culture on serum free medium. Biotechnology and Bioengineering 38:733–741

Hulhoven X (2006) Bioprocess software sensors development facing modelling and model uncertainties. PhD thesis, Universite Libre de Bruxelles

Keesman K (2011) Systems Identification: An Introduction. Springer-Verlag London

Kildegaard HF, Baycin-Hizal D, Lewis NE, Betenbaugh MJ (2013) The emerging cho systems biology era: harnessing the omics revolution for biotechnology. Current Opinion in Biotechnology 1(0):–

Kovacevic Z, Brkljac O, Bajin K (1991) Control and function of the transamination pathways of glutamine oxidation in tumour cells. Biochemical Journal 273:271 – 275

Kurokawa H, Park Y, Iijima S, Kobayashi T (1994) Growth characteristics in fedbatch culture of hybridoma cells with control of glucose and glutamine concentrations. Biotechnology and Bioengineering 44:95–103

Lee Y, Miranda G, Wei-Shou H, Kathy T (2003) Low-glutamine fed-batch cultures of 293-hek serum-free suspension cells for adenovirus production. Biotechnology Progress 19(2):501–509

Linardos T, Kalogerakis N, Behie L, Lamontagne L (1991) The effect of specific growth rate and death rate on monoclonal antibody production in hybridoma chemostat cultures. The Canadian Journal of Chemical Engineering 69:429–438

Lindner P, Hitzmann B (2006) Experimental design for optimal parameter estimation of an enzyme kinetic process based on the analysis of the fisher information matrix. Journal of Theoretical Biology 238:111–123

Ljunggren J, Lena H (1992) Glutamine limited fed-batch culture reduces the overflow metabolism of amino acids in myeloma cells. Cytotechnology 8(1):45–56

Ljunggren J, Lena H (1994) Catabolic control of hybridoma cells by glucose and glutamine limited fed batch cultures. Biotechnol Bioeng 44:808–818

Logan D, Carvell J, Lee M (2011) Creating new opportunities in process control through radio frequency impedance spectroscopy. BMC Proceedings 5(8):57

Mairet F, Bernard O, Masci P, Lacour T, Sciandra A (2011) Modelling neutral lipid production by the microalga isochrysis aff galbana under nitrogen limitation. Bioresource Technology 102:142–149

Miller W, Blanch H, Wilke C (1988a) A kinetic analysis of hybridoma growth and metabolism in batch and continuous suspension culture: effect of nutrient concentration, dilution rate, and ph. Biotechnology and Bioengineering 32:947–965

Miller W, Wilke C, Blanch H (1988b) Transient responses of hybridoma cells to lactate and ammonia pulse and step changes in continuous culture. Bioprocess Engineering 3(3):113–122

Moreno J, Wouwer AV, Rocha-Cozatl E (2012) Observability/detectability analysis for nonlinear systems with unknown inputs - application to biochemical processes. Proceedings of the 20th IEEE Mediterranean Conference on Control and Automation (MED) pp 151–156

Moreno J, Rocha-Cozatl E, Wouwer AV (2014) A dynamical interpretation of strong observability and detectability concepts for nonlinear systems with unknown inputs: application to biochemical processes. Bioprocess and Biosystems Engineering 37:37–49

Murray-Smith D (2013) The application of parameter sensitivity analysis methods to inverse simulation models. Mathematical and Computer Modelling of Dynamical Systems 19(1):67–90

Nguang S, Chen L, Chen X (2001) Optimisation of fed-batch culture of hybridoma cells using genetic algorithms. ISA Transactions 40:381–389

Nolan R, Lee K (2011) Dynamic model of cho cell metabolism. Metabolic Engineering 13:108–124

Pfeiffer T, Sanchez-Valdenebro I, Nuno J, Schuster S (1999) Metatool: for studying metabolic networks. Bioinformatics 15(3):251–257

Portner R, Schafer T (1996) Modelling hybridoma cell growth and metabolism - a comparison of selected models and data. Journal of Biotechnology 49:119–135

Portner R, Schilling A, Ludemann I, Markl H (1996) High density fed-batch cultures for hybridoma cells performed with the aid of a kinetic model. Bioprocess Engineering 15:117–124

Provost A (2006) Metabolic design of dynamic bioreaction models. PhD thesis, Université Catholique de Louvain, URL http://edoc.bib.ucl.ac.be:81/ETD-db/collection/available/BelnUcetd-11122006-203658/unrestricted/these.pdf

Provost A, Bastin G (2004) Dynamic metabolic modelling under the balanced growth condition. Journal of Process Control 14:717–728, URL http://www.inma.ucl.ac.be/publi/239019.pdf

Provost A, Bastin G, Agathos S, Schneider YJ (2006) Metabolic design of macroscopic bioreaction models: application to chinese hamster ovary cells. Bioprocess and Biosystems Engineering 29(5-6):349 – 366

dos Reis Castilho L (2008) Animal Cell Technology: From Biopharmaceuticals to Gene Therapy. Taylor and Francis Group

Reitzer LJ, Wice BM, Kennell D (1979) Evidence that glutamine, not sugar, is the major energy source for cultured hela cells. Journal of Biological Chemistry 254:2669–2676

Roubos J, de Gooijer C, van Straten G, van Boxtel A (1997) Comparison of optimization methods for fed-batch cultures of hybridoma cells. Bioprocess Engineering 17:99–102

Roubos J, van Straten G, van Boxtel A (1999) An evolutionary strategy for fed-batch bioreactor optimization: concepts and performance. Journal of Biotechnology 67:173–187

Saccomani M, Audoly S, Angio L (2003) Parameter identifiability of nonlinear systems: the role of initial conditions. Automatica 39:619–632

Saccomani M, Audoly S, Bellu G, Angio L (2010) Examples of testing global identifiability of biological and biomedical models. Computers in Biology and Medicine 40:402–407

Santos L (2001) Multivariable predictive control of chemical processes. PhD thesis, Universidade de Coimbra, Portugal

Santos L, Dewasme L, Hantson AL, Wouwer AV (2010) Nonlinear model predictive control of fed-batch cultures of micro-organisms exhibiting overflow metabolism. In: IEEE International Conference on Control Applications (CCA), pp 1608–1613

Santos L, Dewasme L, Coutinho D, Wouwer AV (2012) Nonlinear model predictive control of fed-batch cultures of micro-organisms exhibiting overflow metabolism: Assessment and robustness. Computers & Chemical Engineering 39:143–151

Saraiva I, Santos L, Wouwer AV (2010) Nonlinear model predictive control of animal cell cultures in perfusion mode. In: Proceedings of the 29th Benelux Meeting on Systems and Control, Heeze, The Netherlands, p 69, URL http://www.wfw.wtb.tue.nl/benelux2010/boaWWWversion1.pdf

Saraiva I, Sbarciog M, Wouwer AV (2011) Multivariable control of animal cell cultures in perfusion mode. In: Proceedings of the 30th Benelux Meeting on Systems and Control, Lommel, Belgium, p 183

Saraiva I, Sbarciog M, Wouwer AV (2012) Practical design of an extended kalman filter for an animal cell culture. In: Proceedings of the 31st Benelux Meeting on Systems and Control, Heijen, The Netherlands, p 147, URL http://www.wfw.wtb.tue.nl/benelux2012/boaWWWfinal.pdf

Sarkar D, Modak J (2004) Optimization of fed-batch bioreactors using genetic algorithm: multiple control variables. Computers & Chemical Engineering 28(5):789–798

Sbarciog M, Saraiva I, Vande Wouwer A (2013) Accelerating animal cell growth in perfusion mode by multivariable control: simulation studies. Bioprocess and Biosystems Engineering 36(5):517–530

Schuster S, Dandekar T, Fell D (1999) Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. TIBTECH 17:53–60

Sidoli F, Mantalaris A, Asprey S (2004) Modelling of mammalian cells and cell culture processes. Cytotechnology 44(1-2):27–46

Silva A, Marc A, Engasser J, Goergen J (1996) Kinetic model of hybridoma cultures for the identification of rate limiting factors and process optimisation. Mathematics and Computers in Simulation 42:197–205

Simpson N, Singh R, Perani A, Goldenzon C, Al-Rubeai M (1998) In hybridoma cultures, deprivation of any single amino acid leads to apoptotic death, which is suppressed by the expression of the bcl-2 gene. Biotechnol Bioeng 59(1):90 – 8

Stephanopoulos G, Aristidou A, Nielsen J (1998) Metabolic Engineering: Principles and Methodologies. Academic Press

Torres N, Voit E (2002) Pathway Analysis and Optimization in Metabolic Engineering. Cambridge University Press

de Tremblay M (1991) Modelisation, optimisation et commande d'un bioreacteur pour la culture d'hybridomes en mode "fed-batch". PhD thesis, Universite de Montreal

de Tremblay M, Perrier M, Chavarie C, Archambault J (1992) Optimization of fed-batch culture of hybridoma cells using dynamic programming: single and multi feed cases. Bioprocess Engineering 7(5):229–234

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

Tziampazis E, Sambanis A (1994) Modelling of cell culture processes. Cytotechnology 14:191–204

Veloso A, Rocha I, Ferreira E (2008) Monitoring of fed-batch e. coli fermentations with software sensors. Bioprocess and Biosystems Engineering 32(3):381–388

Xing Z, Bishop N, Leister K, Jian Z (2010) Modeling kinetics of a large-scale fed-batch cho cell culture by markov chain monte carlo method. Biotechnology Progress 26(1):208–219

YSI (2014) Ysi flownamics seg-flow biochemistry analyzer provides automated on-line analysis. http://www.ysilifesciences.com/index.php?page=ysi-2950-biochemistry-analyzer, visited 18/12/2014

Zamorano F (2012) Metabolic flux analysis of cho cell cultures. PhD thesis, Université de Mons

Zamorano F, Wouwer AV, Hantson AL, Bastin G (2009) Metabolic flux interval analysis of cho cells. In: Proceedings of the 6th Conference on Mathematical Modelling, URL http://www.inma.ucl.ac.be/publi/523033.pdf

Zamorano F, Wouwer AV, Bastin G (2010) A detailed metabolic flux analysis of an underdetermined network of cho cells. Journal of Biotechnology 150(4):497 – 508

Zamorano F, Wouwer AV, Jungers R, Bastin G (2013) Dynamic metabolic models of cho cell cultures through minimal sets of elementary flux modes. Journal of Biotechnology 164(3):409–422

Zeitz M (1984) Observability canonical (phase-variable) form for non-linear time-variable systems. International Journal of Systems Science 15(9):949–958

Zeng A (1995) A kinetic model for product formation of microbial and mammalian cells. Biotechnology and Bioengineering 46:314–324

Zeng A (1996a) Mathematical modeling and analysis of monoclonal antibody production by hybridoma cells. Biotechnology and Bioengineering 50:238–247

Zeng A (1996b) Quantitative assessment of cell density effect on the metabolism and antibody production rate of hybridoma cells at high cell density. Journal of Biotechnology 45:243–251

Zeng A, Deckwer W, Hu W (1998) Determinants and rate laws of growth and death of hybridoma cells in continuous culture. Biotechnology and Bioengineering 57:642–654

Zhang J (2010) Manual of industrial microbiology and biotechnology, vol Mammalian cell culture for biopharmaceutical production, 3rd edn. ASM Press

Zhou F, Bi JX, Zeng AP, Yuan JQ (2006) A macrokinetic and regulator model for myeloma cell culture based on metabolic balance of pathways. Process Biochemistry 41(10):2207 – 2217

Zhou J, Wen C (2008) Adaptive Backstepping Control of Uncertain Systems. Springer

Zielke H, Ozand P, Tildon J, Sevdalian D, Cornblath M (1976) Growth of human diploid fibroblasts in the absence of glucose utilization. Proceedings of the National Academy of Sciences of the United States of America 73(11):4110–4114

Zivari H (2009) Efficient simulation, accurate sensitivity analysis and reliable parameter estimation for delay differential equations. PhD thesis, University of Toronto