Novel Strategies for Process Control Based on Hybrid Semi-parametric Mathematical Systems

Moritz von Stosch

Porto, September 2011

Novel Strategies for Process Control Based on Hybrid Semi-parametric Mathematical Systems

Moritz von Stosch

Dissertação para a obtenção do grau de Doutor em Engenharia Química pela Faculdade de Engenharia da Universidade do Porto

Universidade do Porto

Faculdade de Engenharia

FEUP

Departamento de Engenharia Química

DEQ

tese realizada sob a orientação do

Professor Doutor Sebastião José Cabral Feyo de Azevedo

Professor Catedrático e Director da Faculdade de Engenharia da Universidade do Porto

e co-orientação do

Doutor Rui Manuel Freitas Oliveira

Professor Auxiliar da Faculdade de Ciência e Technologia da Universidade Nova de Lisboa

e co-orientação da

Doutora Maria Joana Monteiro de Carvalho Peres

Professora Auxiliar da Faculdade de Engenharia da Universidade do Porto

Acknowledgements

The question I heard the most during my PhD in Porto was: "Why would someone want to do his PhD in Porto? Why in Portugal?", which I would herewith like to answer:

When I was a student in Aachen, back in 2006, I was dreaming of what was happening to become reality in the following years. It all started with an appointment with Professor Wolfgang Marquardt, in which I confessed to him that I wanted to do my PhD (i) at a good university; (ii) in modeling; (iii) abroad; (iv) involving to learn a new language; and (v) close to the sea. Not asking me why nor trying to convince me otherwise, where at this point I want to deeply thank him for his support, he answered: "Well, I have a highly-qualified colleague in Porto, whom you might want to contact. You should tell him that I recommend you to him." At the very same day, I sent an email to this colleague, Professor Sebastião Feyo de Azevedo, and when he did not answer, I tried to phone him and I reached Joana Azeredo, who helped me out (as so many times throughout my PhD, which I herewith would like to gratefully acknowledge) giving me the correct number. I called him, introduced myself as "Maurice" and it turned out he had not received my email. We appointed for Whit Monday 2006 and I asked him three times whether he was sure about that date since it is a holiday in Germany. He said: "yes", and told me that even if it would be a holiday, he'd be most probably in the office.

So, I arrived for the first time in Porto on a Friday in 2006. The weather was perfect and what most people think of, when they hear about Portugal. I stayed on the campsite (Prelada) not too far from FEUP and on Monday I went there taking a taxi as the bus did not show up (a typical scenario I experienced more times during my stay). I arrived at FEUP, took the elevator to the second floor, and when I came out Professor Feyo de Azevedo was passing by the elevator door, looking at me and asking, in french since I had introduced myself as Maurice, whether this was me, then welcoming me and inviting me to his office. We went there, continuing to speak french until he looked at my resume and saw that I was german (whereupon he told me that he also spoke some german but we continued in english), and then he showed me around at FEUP. Back at his office, he told me that he did not have any research position at this time, but that I should come back tomorrow, he would arrange for some (as he always arranged/

provided opportunities, such as the courses PASI2008 and BEC2008 which I had the chance to visit, wherefore I want to thank him very much at this point.). Further, he gave me his business card, wrote his personal mobile number on it and told me to call him, at any time, if I would need help (and so it was throughout the entire PhD time. I always felt that I had his back-up, on which I could rely, and which in several bureaucratic situations made my life so much easier. Thank you very much therefore. Also I want to sincerely thank for all the time and effort put in the publications). Leaving his office, he introduced me to a current PhD student Helder (Dr. Helder Gomes da Silva) and Helder told me, that when I would move to Porto I should contact him, so that he could help me with all the stuff (and so he did, wherefore I am most thankful because without him my start to this new life in Porto would have been anything else but the dream it became). At the next day, back in the office of Professor Feyo de Azevedo, he and Doutora Joana Peres (my future PhD co-supervisor, whom I deeply want to thank for all the support along this PhD and the support when attempting to patent the developed software. Further, with her patience and organizational talent she helped me retaining my interests in all university matters) offered me a research position in a project on hybrid modeling, in which they were involved with Doktor Rui Oliveira (my other future PhD co-supervisor, whom I sincerely want to acknowledge for all the discussions, most of which at night on Skype, and for the housing and arrangements he provided whenever I came to Lisbon. His visions, different perspectives and the ability to keep (me) focused gave this work the necessary deepness) from the Universidade Nova de Lisboa. With this offer and the impression that Porto is a place where I could see myself living, I went back home and it did not take me long to sign in.

In October of the same year, I packed with the help of my brother, Christian von Stosch, my personal stuff into the car and drove (together with him, one of the many things he helped me with along the last five years wherefore I deeply want to thank him) from Aachen all the way to Porto. At this time, I was still concerned that something could go wrong, but since I had the support of my parents (where I deeply want to thank Dr. Rita Schaumann-von Stosch and Dr. Wolfgang Schleinzer for all the things they arranged for, always concerned with my well-being, and I also want to sincerely thank Thomas von Stosch and Christiane Klemm-von Stosch for their perspective and support along these years), and could always come back home, I was (and am) in a position where I do not need to fear the unknown and could (and can) reach out to take the adventure of five wonderful years in Porto which was about to start.

And then, of course, I came to Porto (even though I did not know this at that time) because of Cristiana Rodrigues de Azevedo a.k.a. Kita or my future wife, who along the last years supported me in all possible manners and made this PhD somewhat part of her life, by giving me the freedom I needed all those days and by always making me smile. Also, at this point I want to thank her brother, Daniel Rodrigues de

Azevedo, who so many times altruistically took care of my and Kita's concerns and I also want to thank her family (especially António and Ana) who made me feel a part of them.

Additionally I want to thank those people who contributed to that these five years:

The Swellhotel crew, namely Hannes Koerber and Reinhard Huber and his father also, especially for the unforgettable moments in Tamsweg and its surroundings.

The JUMB crew, namely Jan Strauch-Graf, Uwe Starr and Boris Peter, and in extension Judith who enabled Uwe and me to have some unforgettable moments in the alps.

My soulmates Christopher Przybylski and Boris Schinke.

So may other people like: Stefan Mauss, Maria Costa, Julio Paiva, Anthony Danko, Zita Soons, Helder (a.k.a. Tropa), Joana Perreira, Fernando Silva, Joaquim Fontes, Luis Paz, Andrea Topanka, Filipe Xavier, Carlos Miranda Afonso (a.k.a. Caca), Professor Urs von Stockar, Anibal Leite, Rita Araújo and Rita Alves.

Porto, 12.12.2011.

This thesis was realized with the financial support of the Fundação para a Ciência e a Tecnologia. The reference of the provided scholarship is: SFRH / BD / 36990 / 2007.

Abstract

The tight control of a process is important for various reasons, namely economical (e.g. to maximize profit), environmental (e.g. to reduce toxic components), or operational (e.g. to guarantee safety). The central question is what parameters or variables influence the desired outcome the most, and how can they be manipulated to achieve the best result. These issues are in this thesis addressed through the application of hybrid semi-parametric mathematical systems to process control.

At first, it is conceptually analyzed how different sources of process knowledge can be integrated into a hybrid modeling framework. The different options regarding the application of nonparametric and parametric models and the associated parameter identification are reviewed, and the manyfold practical applications of hybrid models are discussed.
At a second stage, hybrid modeling is investigated in those aspects related to structure and parameter identification. Two novel hybrid modeling methods are proposed. One of the methods addresses the incorporation of more mechanistic knowledge, namely the integration of different reaction time scales into hybrid structures. In this case it can be observed that the novel methodology leads to higher prediction accuracy than in cases when such knowledge is neglected. Further this methodology offers the opportunity to discover unknown process mechanisms. The other methodology, is based on the extension of a Nonlinear Partial Least Squares (NPLS) model to dynamic cases, namely a NPLS model is embedded into a hybrid material balance structure. With this hybrid model structure, the extraction of information from large amounts of (highly correlated) data is enabled. This hybrid model structure was extensively evaluated regarding its properties and compared to standard dynamic NPLS models, where it was found that, in the hybrid case, the model predictions are more accurate and the extrapolation properties are improved.
At a third stage, the Hybrid (N)PLS modeling method was applied for process monitoring in the context of Process Analytical Technology (PAT), namely to infer the concentrations of lactate, glutamate and biomass from at-time available NIR spectral data, temperature, dissolved oxygen concentration and pH data. It was observed that the hybrid (N)PLS not only keeps the classification properties of static (N)PLS, commonly employed in PAT for fault diagnosis, but also allows the development of improved regression models for process monitoring, requiring less calibration data.
At a fourth stage, different process controller structures are investigated that either are based on the hybrid process model or incorporate structural knowledge along with Artificial Neural Networks (ANNs) in a hybrid sense. Two complementary controller tuning methods are proposed. The incorporation of structural knowledge into a hybrid controller is observed to result into a better performance than pure nonparametric model based control.
At the end the conclusions are drawn anent the proposed methodologies in the context of process control and directions for future work are sketched.

All in all, this thesis contributes with hybrid model structures and controller structures which address eminent problems in bioprocesses.

Resumo

O controlo apertado de um processo industrial é importante por diversas razões, nomeadamente económicas (e.g. maximização de lucros), a nível ambiental (e.g. de modo a reduzir os componentes tóxicos), ou de um ponto de vista operacional (e.g. de maneira a garantir os critérios de segurança). A questão fulcral consiste em manter o processo controlado, i.e. definir quais os parâmetros ou variáveis que mais influenciam o resultado desejado, e de que maneira estas podem ser manipuladas por forma a alcançar o resultado óptimo.

Inicialmente, é explorada a forma como diferentes fontes de informação acerca dum processo podem ser integradas no contexto da modelação híbrida. São revistas as diversas formas de aplicar os modelos não-paramétricos e paramétricos e as várias formas de identificar os parâmetros respectivos. São ainda discutidas as várias aplicações práticas de diferentes estruturas híbridas.

Numa segunda fase, a modelização híbrida é investigada nos aspectos relacionados com a estrutura e a identificação de parâmetros. São propostas duas metodologias híbridas novas. Uma das metodologias visa a incorporação de mais conhecimento mecanístico, nomeadamente, a integração no modelo híbrido de diferentes escalas de tempo de reacção. Neste caso, constata-se que a metodologia híbrida proposta é mais exacta na previsão em relação àquelas em que esta informação não é considerada. Adicionalmente, esta metodologia permite a descoberta de mecanismos desconhecidos do processo. A segunda metodologia proposta é baseada na extensão, para uma versão dinâmica, de um modelo não-linear de Mínimos Quadrados Parciais (NPLS, Nonlinear Partial Least Squares), em que um modelo NPLS é embebido numa estrutura híbrida de balanços materiais. Este modelo híbrido torna a tarefa de extracção de informação de bases grandes de dados (altamente correlacionados) possível. Após uma extensa avaliação das propriedades deste modelo concluiu-se que comparado com modelos dinâmicos NPLS as previsões do modelo híbrido são mais exactas e as propriedades de extrapolação são melhoradas.

Seguidamente, o modelo híbrido NPLS foi aplicado a um processo de monotorização no contexto de PAT (Process Analytical Technology) com o objectivo de inferir as concentrações de lactato, glutamato e biomassa a partir de dados NIR espectrais "at-time" disponíveis, e ainda temperatura, concentração de oxigénio dissolvido e dados referentes ao pH. Observou-se que o modelo híbrido não só mantem as propriedades de classificação dos modelos estáticos (N)PLS, normalmente aplicados à detecção de falhas em PAT, como permite o desenvolvimento de modelos de regressão melhorados para monitorização de processos, requerendo menos dados para a sua calibração.

Posteriormente, são investigadas diferentes estruturas de controlo processual sendo estas baseadas em modelos híbridos do processo ou baseadas na incorporação, num sentido híbrido, de informação estrutural do processo juntamente com Redes Neuronais Artificiais (ANN, Artificial Neural Networks). Dois métodos complementares de sintonizão de controladores são propostos. Concluiu-se que um controlador híbrido incorporado com informação estrutural supera o desempenho de um controlador baseado apenas num modelo não paramétrico puro.

Nas conclusões finais, as metodologias propostas são enquadradas em controlo de processos e são apontadas as direcções futuras.

Em sûmula, esta tese contribui com novas estruturas híbridas e novas estruturas de controladores que respondem a problemas actuais da indústria de bioprocessos.

Index

List of Figures

List of Tables

Chapter 1

Introduction

A process consists of a set of operational units in which inputs are transformed to value added outputs. In this regard, process control seeks to manipulate the supply of inputs in such a manner that quality and/or quantity of the process outputs can be guaranteed and/or in relation to either time or supplied resources, maximized.
The concept of control is universal and can be found in all areas of activity (education, politics, religion, justice, economics, science and others). Control capitalizes the principle of cause and effect, that is a "controlled" manipulation of the cause should ideally result into the desired effect, see Fig. 1.1.

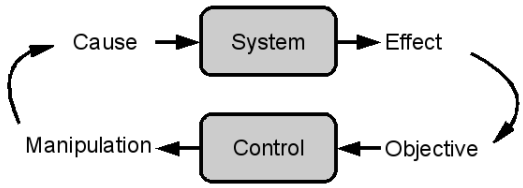

Figure 1.1 Control in the context of the principle of cause and effect

A process can be represented as a system in which input signals are transformed to output signals. The underlying system fundamentals are however *a priori* many times not known and can only be observed through cause and effect. Most times the system is not simple and instead of one cause and one effect there are several for each of which, i.e. multivariate. Additionally, the determination whether a quantity is a cause or an effect might be difficult, since the system might be self-triggering (positive or negative feedback). Moreover, it might be infeasible to either observe all causes or all effects, wherefore it becomes even harder to understand their interaction.

However, the development of a system effigy, referred to as model, might result into a "crude" but helpful understanding of the system. The unconsidered causes, effects and interactions are lumped and constitute stochastic uncertainties, which at least for the considered interactions should be negligible small. When this is the case then the derived model can be used to infer how the causes have to be manipulated in order to reach a certain objective. When applying these manipulations to the system, they ideally result into the desired effects. However, not all observable causes can be manipulated.

1.1 Biochemical Processes - Process monitoring and control

Control of biochemical processes, which is focused on in this thesis, is of major importance due to a number of reasons, e.g. safe and profitable operation, reduction of waste or minimization of energy requirements, to name just a few. The control of the production process usually breaks down to the independent control of several subunits (Rathore & al., 2010), such as the reactor. When referring to process control, it is mostly thought of the control of the reactor, since the reactor is the central unit of a bioprocess (short for biochemical processes), in which the educts are converted to products. The control strategy for the reactor depends on its operation mode, i.e. the reactor can be operated in continuous, batch or fed-batch mode. Batch and fed-batch modes are especially widespread in bioprocesses, due to several reasons Roubos (2002); Schuegerl (2001); Yamuna Rani and Ramachandra Rao (1999). The control of batch and fed-batch operations is more difficult than continuous operation, because the linearization of the process model around an operating point, for the design of a linear controller (e.g. a PID), is infeasible. Further, the control is challenging because the underlying system is usually dynamic and highly-nonlinear and awards with multi-variable aspects (Yamuna Rani and Ramachandra Rao, 1999). Additionally, the dynamics of the states can many times not be directly accessed, since most states cannot be directly at-time measured, nor measured frequently enough.

For bioprocesses, the development of measurement techniques has evolved significantly since the beginnings (Junker and Wang, 2006), because efficient process monitoring is needed for better process modeling and process control (Schuegerl, 2001), see Fig. 1.2. There are now many different techniques available for the monitoring of different state variables (Becker & al., 2007; Clementschitsch and Bayer, 2006; Gnoth & al., 2008a; Harms & al., 2002; Junker and Wang, 2006; Scheper & al., 1999; Vojinovic

& al., 2006, 2007). From a process control perspective it is interesting to divide those techniques into those which reveal measurement information at-time (meaning that they become available sufficiently fast to be used for feedback in control, which can comprise e.g. in-line, on-line, at-line techniques) and those which do not provide the information at-time, namely off-line. Those variables that cannot be measured at-time can sometimes be inferred using a model that has been beforehand calibrated from off-line and at-time measurements. Thereto it is referred as soft-sensors or software sensors (Becker & al., 2007; Gnoth & al., 2008a; Junker and Wang, 2006; Kadlec & al., 2009). However those soft-sensors can produce unreliable values when operating in regions which have not been accounted for during calibration. In general, there is not one optimal set of devices which suites all processes, but each process requires a particular set-up of measurement techniques, allowing to monitor specifically the key-components. For instance, the values of in-situ, at-time measurements might not be representative, since local concentration gradients can occur due to non-ideal mixing, which is frequently observed in industrial units. In such cases, so-called integral measurements might be better (Gnoth & al., 2008a). This also demonstrates the importance of the sensor placement or accordingly the importance of taking representative samples (van de Merbel & al., 1996).

In general, the collected data (also referred to as historical data) form a valuable source of information (Gnoth & al., 2008a; Luebbert and Bay Jorgensen, 2001). They can, for instance, be exploited to iteratively optimize the process operation policy (Bonvin, 1998). This means that through the utilization of a process model (derived from the data), e.g. the optimal feeding rate profile for the maximization of the product yield on substrate can be off-line identified. Such, the process can, to a certain degree, be optimized without the knowledge of at-time concentration measurements. However, occurring variations from the desired value cannot be corrected for, since there is no feed-back regarding the current values. It is commonly referred thereto as optimal open-loop control (Feyo de Azevedo & al., 2001). In fact, closed-loop control only makes sense when the current value of the controlled variable can be determined accurately and fast enough (Gnoth & al., 2008a). The ideal situation is, therefore, to determine at-time the concentrations and to choose the product concentration as controlled variable. However, since the product concentrations are typically very difficult to determine accurately enough at-time, indirect control strategies are sought. For instance, the biomass growth rate (which sometimes correlates with the product formation rate) is many times tended to be controlled. Approaches aiming thereat are for instance Dabros & al. (2010); Jenzsch & al. (2006a); Soons & al. (2006), where in all of which the identification of the biomass

growth rate is a key-point.

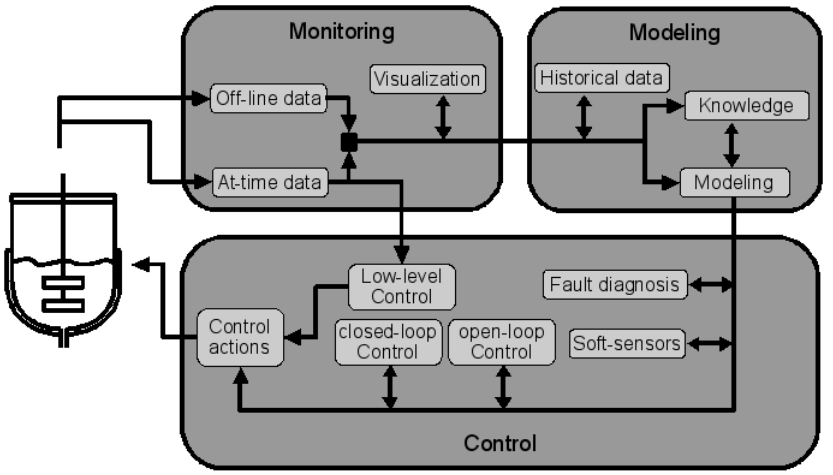

Figure 1.2 The different components for bioprocess control.

The application of advanced control schema to industrial settings remains quite occasional Alford (2006); Gnoth & al. (2008a); Jenzsch & al. (2006b); Rathore & al. (2010), partly due to the regulations that were imposed by government agencies, i.e. these processes were carried out manually, following an approved recipe which assumes that the product forms as defined in the regulatory issues (Clementschitsch and Bayer, 2006; Jenzsch & al., 2006b). Recent changes in the regulatory framework, e.g. the Process Analytical Technology (PAT) initiative by the Food and Drug Agency (FDA) in the United States PAT (2004), now offer industry the opportunity to have variance in their processes whenever the targeted product quality can be ensured, i.e. Quality by Design. This leaves room for frequent process re-optimization with respect to changing material, personal and energy costs. It will be interesting to see, how the at-time knowledge about the process state can be used for improved operation scheduling strategies, which aim at a lean process and ultimately at supply chain management (real-time release), since stock binds capital.
With this renewed interest in bioprocess control, the call is made for the development of efficient strategies which can be implemented right-away.

1.2 Modeling - a cornerstone for successful process control

Nonlinear dynamic systems are frequently encountered in biochemical (fed)-batch processes, but their control remains challenging, in the sense that quality and quantity of the process product can many times not be guaranteed in advance. Reasons are that:

1. even when the process are run in the same manner, the product properties will vary since variability enters the process through e.g. the raw materials, the environment or stochastic effects on the scale of the reactants (Rathore & al., 2010; Read & al., 2010a);

2. the equipment of the process plants (measurements) may only allow for low-level closed-loop control (e.g. thermostat, pHstat, DOstat), wherefore product properties cannot directly be subject to closed-control (Rathore & al., 2010; Read & al., 2010a); and

3. the underlying process system is predominantly poorly understood (Rathore & al., 2010).

The first listed point, namely product variability, can partially be accounted for by immediate control action, such as to minimize the effect of the disturbance on the process. This implies that closed-loop control is implemented on all process levels, wherefore, obviously, the adequate plant equipment is a prerequisite. The second reason, i.e. the plant equipment, is more and more ruled out since (i) government agencies recognized the advantages that closed-loop process control can bring, which already resulted in more flexible process guidelines, e.g. (PAT, 2004); and (ii) provided with these opportunities, industry recognized the potential of closed-loop control to minimize production costs while maximizing yield, wherefore substantial investment into equipment is made (Hinz, 2006). Thus the point that remains is the understanding of the process fundamentals, which can be achieved through process modeling.
The formulation of a process model can be viewed as an exercise of translation of the a priori knowledge into a condensed mathematical representation. In engineering science, the conservation laws, such as material or energy balances, provide an almost universally applicable modeling framework, which then "only" has to be complemented with the process specific details, e.g. the formulation of kinetic rate functions. However, the formulation of those specific details is rather laborious till the model description reaches the desired accuracy, because neither all details are observable nor are the underlying relationships *a-priori* known. Additionally, the inherent system dynamics and the multi-

variate context complicate the modeling exercise.

A different modeling approach that is rather quickly applicable and requires less mechanistic knowledge about the system is the one of data-driven models. As the name indicates this approach relies exclusively on experimental data. The concept bases on mathematics in order to describe the data underlying relations. As though, data-driven models can hardly differentiate between cause and effect. In comparison to mechanistic models (i) the physical meaning of data-driven models can barely be interpreted in terms of the underlying mechanisms; (ii) more data are necessary for the derivation of data-driven models; and (iii) their descriptive quality is good only in close vicinity to those regions for which they were derived. Ponton and Klemes (1993) even show that the most crude mechanistic model assumptions usually result into better prediction capabilities, than data-driven techniques.

An alternative to pure mechanistic/ phenomenological or pure data-driven modeling is the fusion of these methodologies in a hybrid model. Hybrid modeling provides a framework in which the integration of all kind of available process knowledge is intended. The advantages arising thereby are that (i) the model predictions (in the process regions the model was developed for) are more accurate than those obtained for each of the independent techniques alone; (ii) the model development is less cost intensive than the one for mechanistic modeling; (iii) the extrapolation properties are much better than those obtained for pure data-driven models; and (iv) the model can reveal mechanistic understanding. Therefore hybrid modeling provides an optimal platform in order to address nonlinear dynamic systems. This, of course, was also noted by other researchers and so the application of hybrid models is currently promoted for the PAT framework (Gernaey and Gani, 2010; Glassey *&* al., 2011; Teixeira *&* al., 2009). As a matter of fact, the use of hybrid models for bioprocess monitoring and control is promoted all along the way of their existence, e.g. (Alford, 2006; Bonvin, 1998; Gnoth *&* al., 2008a; Schuegerl, 2001; Yamuna Rani and Ramachandra Rao, 1999).

1.3 Objectives

So far, it was outlined that for efficient bioprocess control process understanding is a critical factor. In this regard, it was formulated that hybrid modeling can be expected to be an important vehicle to achieve this goal.
The application of hybrid models, comes along with the choice of the model structure

among many different possibilities, concerning the embedded nonparametric structure and underlying identification, which can have a large impact on the overall hybrid model performance, or which can even restrict the applicability of the hybrid model. Thus there is a need to systematically determine the influence that those decisions can have on the performance and moreover to investigate what other effects impact thereon.
This is especially important for the application of hybrid models to biochemical process, in order to be capable to distinguish between model mismatches that are due to the methodology (numerics) and those that are due to the process underlying cell-system.
The cell-based biological systems are highly complex and highly sensitive to changes in their environment (Read & al., 2010a). Their complexity stems from multi-variable interactions across all cell levels and from the different scales of intrinsic cell dynamics. It is not feasible (at this time) to consider all details about a cell system in an integrated framework (Stelling, 2004), but the consideration of all the details is also not mandatory for a lumped representation of the overall process system. Nevertheless, for good coherence of the model with the cell-based biological system, the inherent system dynamics have to be accounted for. Also changes in the cell environment, which are tented to be captured with at-time measurements, must be considered by the modeling approach. In this respect, methods are required that can evaluate and process huge amounts of highly-correlated data.
For process control there are, two main ways to profit from a process model. Either the model is integrated into the controller structure, or the model is used to tune a standalone controller. Both scenarios are interesting for investigation. Further, for the case of standalone controllers, the hybrid approach, namely the combination of different knowledge sources into a controller is interesting to be explored. Thus the list of objectives adds up to:

1. A systematic analysis of hybrid methodologies and their applications.

2. Development of a hybrid methodology that poses better coherence with the cell system through the integration of as much as possible information about the cell-environment, namely the development of a method that can process large amounts of correlated data.

3. Development of a hybrid methodology that constitutes better coherence with the cell-system, namely with the intrinsic dynamics that such systems do exhibit.

4. Investigation on the value of hybrid models and/or hybrid principles for process control.

1.4 Thesis Outline

This thesis is divided into four parts. In a first part (Chapter 2), a literature review on hybrid models is presented, where the focus is on the different hybrid methodologies and the varying fields of application. The work contained in this thesis is partly covered therein, in order to classify it in the overall context.
In a second stage, two complementary hybrid methodologies are developed. One methodology (Chapter 3) aims at the development of a serial hybrid model in which the nonparametric model can cope with correlated inputs and outputs, since this data cannot be adequately processed by the data-driven techniques that are traditionally used in serial hybrid modeling. In the other case (Chapter 4), a hybrid model is developed, which is more coherent with the true nature of cell-system dynamics.
In a third stage, hybrid models are applied for process monitoring in a PAT sense. The hybrid model is applied as a software sensor for processing highly dimensional input data for at-line monitoring of biomass, lactate and glutamate concentrations in a batch cultivation of *Bordetella pertussis*, Chapter 5.
In a fourth stage, a general hybrid control framework is developed in which both cases, i.e. the development of model based control structures and the utilization of the hybrid model for controller tuning, are considered. Additionally, the impact of knowledge integration into the controller formulation is studied.
In more detail the chapters deal with the following:

Chapter 2: In this chapter, the different existing structures of hybrid methodologies are explored. The focus is on (i) the different ways of process knowledge integration; (ii) the differences in nonparametric models and associated parameter identification; (iii) the purpose for which hybrid approaches are applied, such as modeling, control, optimization, etc.; (iv) . In such a way the hybrid methodologies state of the art is provided. (This chapter, in a more brief form, will be submitted to the Journal of Computers and Chemical Engineering.)

Chapter 3: (Nonlinear) Partial Least Squares (also referred to as (Nonlinear) Projection to Latent Structures, (N)PLS) is one of the most applied modeling approaches when large amounts of (highly correlated) data are required to be processed. In this chapter, a hybrid methodology is developed and evaluated which bases on the integration of (N)PLS into a material balance framework. This results into a inherent dynamic (N)PLS model which is shown to be better than other dynamic (N)PLS approaches.

(This chapter has been published, as von Stosch & al. (2011b).)

Chapter 4: A hybrid methodology is proposed in which the standard framework that bases on Ordinary Differential Equations (ODEs) is extended to Delay Differential Equations (DDEs). It is demonstrated that the proposed hybrid model can cope more accurately with the different orders of magnitude in reaction time scales, than the standard model. Additionally, it is shown that the "true" underlying system delay (in a mechanistic sense) might be unraveled through the novel approach. (This chapter has been published, as von Stosch & al. (2010).)

Chapter 5: The application of hybrid modeling for process monitoring is investigated in this chapter, with experimental data of *Bordetella pertussis* cultivations. In this regard the methodology, developed in chapter 3, is utilized to calibrate at-time available measurement data, i.e. pH, temperature, dissolved oxygen concentration and NIR spectral data, to concentration measurements that are only off-line available, and normally are sparse, infrequent and noisy. The hybrid methodology is benchmarked against static PLS models, which are usually applied in this context. (This chapter has been published, as von Stosch & al. (2011a).)

Chapter 6: A general hybrid control methodology is proposed. Different controller structures are presented that either base on the hybrid process model or incorporate structural knowledge along with Artificial Neural Networks (ANNs) in a hybrid sense. Two methodologies for off-line controller tuning are presented, that either use the process data or the process hybrid model. (This chapter will be submitted to the Journal of Process Control.)

Chapter 7: In this chapter the final conclusions are presented and possible directions for future work are developed.

Chapter 2

20 years of Hybrid gray-box modeling: A review

2.1 Abstract

Process or systems modeling is in its essence an exercise of translation of knowledge into a mathematical representation. In the classical view, different types of knowledge give rise to different model structures and identification schema. White-box modeling, founded on first principles, mechanisms or observed phenomena (phenomenological), gives rise to parametric structures. Black-box modeling is based exclusively on process data giving rise to nonparametric structures. In hybrid gray-box modeling, the distinctive feature is that different types of knowledge are viewed as complementary, thus their mathematical representation combines parametric and nonparametric into hybrid semi-parametric structures. The advantages of hybrid gray-box modeling, such as high prediction accuracy, good extrapolation properties, transparency of the model structure, small requirements on calibration data and cost-effective model development, have been widely recognized, not only in academia but also in the industry. However, in comparison to the parametric or the nonparametric techniques, hybrid modeling still lives in the shadows.

In this review, the most common hybrid modeling techniques and their identification techniques are revisited. Hybrid model applications in the areas of chemical, biochemical or mechanical engineering are mentioned. Further, the applications of hybrid modeling to Monitoring, Control, Optimization, Scale-Up and Model Reduction are reviewed. It is outlined that the application of hybrid techniques does not automatically lead into better results but that careful knowledge integration can result into significant enhancements.

2.2 Introduction

The beginnings - a sound foundation and source of inspiration.

Hybrid modeling evolved from the field of neural networks and was first reported in 1992 by Johansen and Foss (1992b); Kramer & al. (1992); Psichogios and Ungar (1992); Su & al. (1992). Johansen and Foss (1992b); Kramer & al. (1992); Su & al. (1992) proposed a so called parallel set-up, Psichogios and Ungar (1992) derived a serial approach. The central idea was to *a priori* structure the neural network model through the use of first-principle knowledge. The result was that when trained with the same amount of available process data, the hybrid model was capable to predict the process states better, was able to interpolate and extrapolate mostly more accurately and was easier to interpret than pure neural networks.

2.2.1 What is hybrid gray-box modeling?

Hybrid modeling is more than just the combination of different modeling techniques. It is a methodology in that all kind of knowledge about a system, which usually is distributed in several separated sources, can be fused and thus represented by one modeling approach. The advantage is that several synergistic effects come into play and therefore the overall representation of the system is, principally, enhanced.

There are various terms to refer to different types of knowledge. Typical terms for explicit formulated observation based knowledge, expressed mathematically in form of equations, are:

- first-principles (established laws of physics, no assumptions such as empirical model and fitting parameters are made);
- mechanistic (tending to explain phenomena only by reference to physical or biological causes);
- parametric (the knowledge formulated as equations rely on some modeling assumptions);
- phenomenological (knowledge that relates empirical observations of phenomena to each other);
- white-box (associated to be transparent, therefore representing approaches derived from observations); and

- empirical (information gained by means of observation).

Frequently used terms for methodologies that can capture data underlying patterns are:

- nonparametric (not based on any modeling assumptions, nor *a priori* structured);

- data-driven (data are the driving force behind those models);

- statistical (relating to the fact that those models capture statistical features of the data);

- black-box (the underlying mechanisms are unknown, thus nontransparent and therefore the opposite to white box models); and

- chemometric (extracting information from (bio)chemical systems by data-driven means of highly dimensional sets).

From these model representations the term hybrid modeling or gray-box modeling evolved, the latter standing for the combination of black-box and white-box modeling. However the distinction between gray box modeling and hybrid modeling is many times not clear and the terms therefore considered interchangeable. Gray-box modeling can, however, be understood as the broader term, in the sense that any kind of prior information incorporated into or along with a black-box model results into a gray-box model. It is proposed to understand hybrid modeling in this respect as a sub category, in that in a hybrid model, the black-box and the white-box can be separated from one another, see Sohlberg (2005).

2.2.2 Why hybrid-modeling? What is the gain?

White-box modeling and black-box modeling constitute two approaches which are different in their traits. While the development of a white-box model requires detailed knowledge about the process and is many times cumbersome or laborious, data-driven approaches are rather quickly applicable and require less knowledge. The concept of the latter bases on mathematics to describe the data underlying relations. In comparison to phenomenological models more data are necessary for the derivation of data-driven models; and its descriptive quality is good only in the vicinity to those regions for which it was derived.

An alternative to these model, that might be understood as a combination of both modeling techniques, is hybrid modeling. Hybrid modeling balances the advantages and disadvantages of pure mechanistic and nonparametric modeling. It awards with several attractive features, such as:

i) Higher estimation or prediction accuracy which stems from a number of factors, namely:

   a) A lower number of model parameters when compared to pure nonparametric techniques, which leads to a higher statistical confidence of the hybrid model;

   b) The opportunity to easily account for all measurable changes in variables that might impact on the predictions, i.e. all measurable quantities can be considered as inputs to the hybrid model structure where the correlation is achieved through the embedded nonparametric approach, while for phenomenological modeling the link must be established explicitly.

   c) Constraints and bounds can be incorporated into the hybrid model avoiding physically unfeasible solutions.

ii) Better calibration properties than with pure nonparametric models can be obtained, which is due the fact that the process operation space can be structured through the mechanistic or phenomenological knowledge decreasing the possible search space. Therefore the requirements on the number of data points and on the data quality are lower for hybrid modeling than for pure nonparametric techniques.

iii) Enhanced extrapolation properties, when compared to pure nonparametric approaches, can be achieved. The explanation is the same as for calibration, i.e. structuring of the process operation space. As a matter of fact calibration and extrapolation are closely related, meaning that the identification from the exact same data yields a hybrid model that has a lower uncertainty than the nonparametric model. When extrapolation, the higher uncertainty of the nonparametric models might be amplified through the lever, given between the region of identification and the region of application.

iv) Fewer phenomenological knowledge is required to construct a hybrid model than to develop a white-box model. However, all knowledge available can be integrated into the hybrid model.

v) Faster development of hybrid models than of white-box models, since less knowledge is required and so no additional effort is needed to unravel sufficient details for the model construction.

vi) The integration of the available phenomenological knowledge leads to a higher transparency of the hybrid model than of the nonparametric models.

Having outlined what hybrid models are, and why they should be considered, the questions to address in the following concerns the methodology of hybrid modeling and fields of application.

2.3 Hybrid-modeling - The framework

Hybrid-modeling is understood as the valuable combination of different sources of process knowledge, i.e. the integration of information which can be captured by miscellaneous models into one framework. Questions arising in this context are:

i) How to arrange the models?

ii) What kind of information can be integrated in what way?

iii) How can unknown parts be represented?

iv) How can the unknown parts be identified?

v) What model is performing best?

vi) What is the influence of the data?

These points will be addressed in detail below.

2.3.1 Hybrid model configurations: Parallel or Serial, One- or Multi-step

Two models can be arranged in three ways, see Fig. 2.1, where structure **A** is referred to as parallel and structures **B** and **C** are called serial, sequential, cascade or consecutive. These structures are theoretically addressed in Agarwal (1997) considering that the white box would represent phenomena founded information, and the black box consists of a nonparametric model. However, in the serial case, the series of the black and the white model might not always be interchangeable. This is for instance the case when the white box would represent material balances and the black box the kinetic rate model.

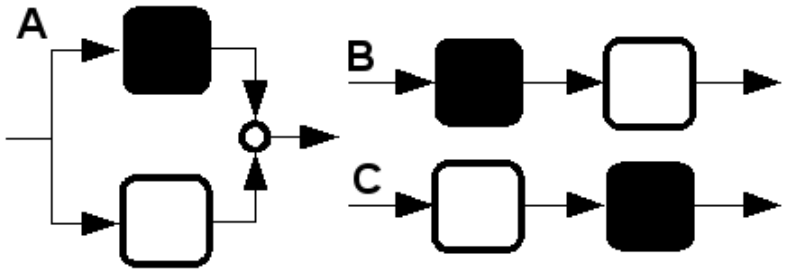

Figure 2.1 Schematic sketch of the three ways to combine two models (represented by a white and a black box). **A** shows a parallel configuration while **B** and **C** a serial structures.

The Parallel structure

The parallel structure **A** usually finds application if a process model (white box) is available, but its prediction power due to whatever reasons (e.g. unmodeled effects, nonlinearities, dynamic behavior) is limited. The parallel arrangement of a nonparametric model, in such a case, can lead to significantly improved predictions. Of course the prediction power of the nonparametric model remains poor in regions that it has not been trained on. The parallel approach is especially interesting if certain effects in the system can be uncoupled, e.g. a static non-linear and dynamic linear behavior, and thus by the separate model representations can accurately be captured (Abonyi *&* al., 1999; Chen *&* al., 2004; Klimasauskas, 1998; Masri, 1994; Narendra and Parthasarathy, 1990; Potocnik and Grabec, 1999; Su and McAvoy, 1993).

There exist several possible manners to combine the outputs of both models. Two fundamental ways are superposition and multiplication, see (Hu *&* al., 2009). The combination of both fundamental ways is also possible, as e.g. (Johansen and Foss, 1992a,b), wherein several models are combined in parallel and a weighting schema determines their contribution to the model's system representation. In Su and McAvoy (1993) the weighting schema accounting for the incorporation of the nonparametric model predictions, is part of the Hammerstein model. The approach proposed by Fellner *&* al. (2003) is somewhat the other way around, where the weighting of the mechanistic model outputs is accomplished by the nonparametric model. Klimasauskas (1998) proposes to use a confidence module for the weighting of the different model outputs. However, the most frequently applied is probably pure superposition, i.e. the summation of the outputs, in which case the nonparametric model predicts the residual between the first-principle model and the experimental data, (Su *&* al., 1992; Thompson and Kramer, 1994).

The Serial structure

The most popular serial combination is the one shown by structure **B**. In this formation the white box usually represents a model derived from first-principles that are the conservation laws, namely material, momentum, impulse, population or energy balances derived for the process at hand. The black box usually represents the underlying kinetics, for which a general valid model is not available, wherefore in the hybrid modeling context, nonparametric models are applied.

This serial structure **B** is especially suitable when no precise knowledge about the underlying mechanisms is available, but sufficient numbers of process data exist to infer the underlying patterns. Another scenario in which its application is worth, is addressed in (Teixeira *&* al., 2007b; von Stosch *&* al., 2011b), i.e. huge sets of data, rich in information about the process state but without direct physical interpretation, can be exploited by the nonparametric model resulting into more precise estimations of the kinetics.

The serial structure **C** can either be applied as an alternative to the parallel structure, i.e. the white-box model predictions are considered as inputs to the nonparametric model or to establish a link between the process state and certain process characterizing parameters (Aguiar and Filho, 2001; Hwang *&* al., 2009; Nascimento *&* al., 1999; Schenker and Agarwal, 2000; Zhang *&* al., 2006). An approach which is somewhat similar to a serial structure **C**, was proposed by Martinez and Wilson (1998); Tsen *&* al. (1996), i.e. to use a suitable first-principles to augment the number of experimental data, and then to use these data for the nonparametric model training. The prediction quality for this type of trained models was shown to be superior to models with structures **A** or **B** see (Tsen *&* al., 1996), which in this case is not astonishing since the model identification on the basis of much larger numbers of data will automatically result into more accurate models. However, neither this hybrid model nor hybrid structure **C** models did find much application so far. Further, other studies than the one by Tsen *&* al. (1996) did not compare the different structures.

Parallel or Serial

In case that the question is whether a serial or a parallel approach is more suitable for a given application, the main focus should be on the structural uncertainty of the mechanistic model. In cases that the structural uncertainty is high, the parallel schema

can be expected to perform better than the serial one, since the parallel nonparametric model can partially account for the structural miss-match (Examples would be (Bhutani & al., 2006; Lee & al., 2002)). Due to the fact that extrapolation properties are heavily determined by the underlying model structure (Fiedler and Schuppert, 2008; Mogk & al., 2002; Schuppert, 1999), the serial structure in such a case (high structural uncertainty) cannot be expected to perform well, i.e. any nonparametric model will probably perform better (Bhutani & al., 2006; Lee & al., 2002). When the structural uncertainty is low, then the serial hybrid model will have significantly better extrapolation properties than the parallel model (van Can & al., 1996) and also the prediction quality of the former can be expected to be increased (Conlin & al., 1997). In Corazza & al. (2005), the fact that the serial model will perform best when the provided structure is close to the "true" underlying structure is used to infer structural mechanistic knowledge.

One-step or Multi-step ahead prediction

Regardless whether the structure is serial or parallel, the hybrid model can be a one-step or a multi-step ahead predictor, see Fig. (2.2). Whether the model is one or the other depends on the kind of information that enters in the model. In the case that measured information is used as an input, the structure is a one-step ahead predictor, while in case that the only inputs are the model outputs the structure is a multi-step ahead predictor. It depends on the application and the availability of information which structure is to prefer or can at all be applied. In van Can & al. (1998) the hybrid models that are identified as one-step ahead predictors are applied as multi-step ahead predictors, as a rigorous model test. The different model properties that are associated with the structure being a one-step or multi-step ahead predictor are analyzed for a serial hybrid model in von Stosch & al. (2011b). Therein, it was observed that those models which feedback the state are, in general, better conditioned and provide enhanced predictions when compared to pure feed-forward models.

2.3.2 Integration of more knowledge into the basic serial or parallel structures

While the overall hybrid structure is usually assessed to categorize the hybrid approach into parallel or serial, the substructures can be versatile, because it is usually worthwhile to include additional knowledge into the hybrid model. This is due to the fact that additional knowledge, even though increasing the complexity of the hybrid from a spectator's view, can reduce and structure the space spanned by estimates and the

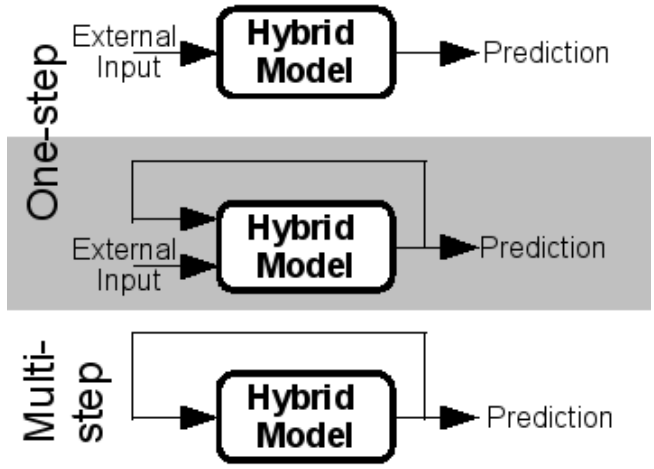

Figure 2.2 Schematic sketches of the model structure for one-step and multi-step ahead predictors.

parameters of the nonparametric model (Fiedler and Schuppert, 2008) (i.e. the curse of dimensionality). As a result, enhanced extrapolation properties, improved predictions and better calibration properties (i.e. less data are required for the calibration, the parameter identification converges faster and less variations in the optimal parameters are obtained) of the hybrid model are obtained.

Several approaches exist to integrate more knowledge into hybrid models than just the information provided by the conservation laws or the process data, which is especially true for the modeling of the kinetics in the serial hybrid structure. In the following the efforts undertaken to integrate more knowledge and schema for the fusion of knowledge are presented.

Additional mechanistic information

Many times additional phenomenological information about the studied system is available. In case that a complete white-box model is at hand, the certainty about the model structure and the certainty about the components should be used to decide which hybrid model structure might be the most suitable, i.e. parallel (**A**) or serial (**B** and **C**). In case that the structural certainty is relatively high it is one option to represent the most uncertain parts of the model through nonparametric techniques (Fu and Barford,

1995a; Georgieva & al., 2003), which results into a serial hybrid model (structure **B**). Similar thereto is the case in which only some phenomenological knowledge is available, in that only the completely unknown parts are necessarily represented through the non-parametric model.

Right from the beginning of hybrid modeling additional knowledge was incorporated, where it is considered that biomass is a catalyst to the reactions (i.e. the multiplication of the specific rates by the biomass) (Psichogios and Ungar, 1992; Schubert & al., 1994a,b; Tholudur and Ramirez, 1996; Zorzetto and Wilson, 1996). A general reaction schema in which two parameters were modeled by nonparametric models was proposed by Reuter & al. (1993) for batch and continuous mineral and metallurgical processes. The generated rate predictions could represent the reactor performance in various situations. Fu and Barford (1995a) used a complete cell model of hybridoma cell line cultivations and replaced the most uncertain parts by nonparametric models. In Chen & al. (2000) it was shown that knowledge about the matrix of stoichiometric/yield coefficients is of value in the formulation of macroscopic material balances. In Mogk & al. (2002) it is demonstrated that already the incorporation of the underlying mechanistic structure can improve the prediction and modeling accuracy. Oliveira (2004) formed a general framework for hybrid models that is based on the general framework of macroscopic material balances by Bastin and Dochain (1990). Therein (Oliveira, 2004) the reaction term is represented by a product of three terms, i.e. the stoichiometry, known and unknown rate terms. Bounded Input Bounded Output (BIBO) stability is discussed, for which a major point is that a reaction can only occur when all required educts are present. Thus the incorporation of more mechanistic knowledge can, beside the advantages mentioned above, also lead to better model properties concerning BIBO stability and/or adherence to physical constraints (Karama & al., 2010; Oliveira, 2004).

The impact that different levels of knowledge integration have on the model performance was studied by Vande Wouwer & al. (2004), applying three structures with different levels of sophistication, and similarly but with considerations regarding one-step and multi-step ahead predictor, studied by von Stosch & al. (2011b). Also Al-Yemni and Yang (2005) studied the difference between the integration of no knowledge at all and some knowledge. A state transformation technique, proposed originally by Bastin and Dochain (1990), was used by Vande Wouwer & al. (2004) to estimate the stoichiometric coefficients. The same state transformation technique (Bastin and Dochain, 1990) was chosen by Chen & al. (2000) to decouple the rates for the identification of the non-

parametric model and also was picked up by Georgieva and de Azevedo (2009) to infer unmeasured state variables and to on-line adapt uncertain stoichiometric coefficients. Brendel and Marquardt (2008) used the target factor analysis (proposed originally by (Bonvin and Rippin, 1990)) to obtain the stoichiometric coefficients. Teixeira & al. (2007a) applied Elementary Flux Modes to mammalian Baby Hamster Kidney (BHK) cultivations and integrated the gained information about the stoichiometry and the most important Elementary Modes (a reduced form of the metabolic network) into a hybrid model. This approach can be seen as a first step to integrate knowledge gained in systems biology into a hybrid framework.

In Al-Yemni (2003); Kasprow (2000); Mazutti & al. (2010) it is proposed to use "standard" formulations of the kinetic rates, and to represent the therein contained parameters by nonparametric model expressions. In Corazza & al. (2005) this idea was used to compare several expressions for inhibited kinetic rates, i.e. the inhibition parameter is given through the nonparametric model, such tenting to identify the "true" underlying structure and thus, ultimately, to infer mechanistic knowledge. The other way round, Costa & al. (2010) reported the use of several empirical expression along with Michaelis-Menten kinetics to determine which empirical expression has the biggest merits. A symbolic reformulation strategy for the underlying kinetic model is suggested by Lima and Saraiva (2007), aiming at a semi-empirical model representation that fits better to the data.

The integration of knowledge directly into the nonparametric model (embedded in a serial hybrid model) was studied by van Deventer & al. (2004), i.e. a semi-empirical regression network in which prior knowledge can directly be integrated was proposed. It was shown that the network structure was easier to identify and that the complete model has better extrapolation properties, than those networks without prior knowledge. Karama & al. (2010) investigates the integration of constraints into the nonparametric models, thus forcing the nonparametric estimates to adhere to physical restrictions. In the approach proposed by Fellner & al. (2003) the overall model structure is a highly connected network in which knowledge can be integrated in form of neurons. Also in this approach, the integration of knowledge leads to better model properties.

Fiedler and Schuppert (2008) addressed the integration of knowledge into a tree-structured scalar hybrid model, in which several parametric and nonparametric models can be integrated (Identifiability is also addressed). It is therein theoretically addressed

that such a structure can avoid the curse of dimensionality of nonparametric structures and thus also provides better extrapolation capabilities.

Combination of incorporated information

Two general ways to fuse information are superposition (as e.g. in most parallel structures) and multiplication (as e.g. proposed in Oliveira (2004)). If however the same quantity is predicted by two different techniques than other fusion approaches must be considered. Weighting schema can be used, as was already outlined for parallel structures in the respective section (Fellner & al., 2003; Johansen and Foss, 1992a,b; Klimasauskas, 1998; Su and McAvoy, 1993). Dors & al. (1995, 1996) applied a weighting function in a serial hybrid model in order to coordinate the predictions of the kinetic rates by heuristic rules (the Monod model) and the ones by a nonparametric model. The kinetic rate predictions and the nonparametric predictions were weighted by a clustering approach (for details see also (Galvanauskas & al., 2004)), where more weight is given to the nonparametric model in regions where process data are available, while restricting it when extrapolating. This combination schema was also picked up by Patnaik (2010), who however determined the weighting iteratively. As an extension thereto the Mixture of Experts framework proposed by Peres & al. (2000, 2001) can be understood. Therein several parallel submodels exist, whose contribution to the final prediction, is selected by a gating function. Note that the construct of the Mixture of Experts is similar to the structure of Fuzzy models, in that the gating function has its analogy in the rules (attendance part) and the submodel in the Fuzzy consequent part. However, the identification of the parameters in the mixture of experts approach is considerable more difficult than that of a Fuzzy model since the partitions (at which certain submodels are active) and the rules have to be learned from the data and are not given by the user, e.g. see Peres & al. (2001).

Another option for weighting different predictions of the same quantity is to use a nonparametric model, where all predictions are inputs to the nonparametric model and only the final prediction is the output (Bollas & al., 2003; Cao & al., 2004a,b; Fellner & al., 2003).

Operational knowledge-Rule based information

Knowledge about certain rule-alike procedures can be captured by Fuzzy models. Those Fuzzy systems make use of a logic structure to describe certain rule-based pro-

cedures, e.g. *if glucose concentration high, then increase biomass growth rate; elseif glucose concentration low, then decrease biomass growth rate*. The expressions such as low or high are associated to parameters that can either be determined manually, through the experience of an operator, or can be fitted to experimental data Roubos (2002); Roubos & al. (2000); Schubert & al. (1994a); van Lith & al. (2002, 2003).

There are several Fuzzy models available. The most popular is maybe the Takagi, Sugeno Kang (TSK) type (Takagi and Sugeno, 1985), in which the consequent part of each rule consists of a linear equation, (van Lith & al., 2002). Therefore the approach could be interpreted as several parallel linear models, where the contribution of each submodel is chosen according to some specified rule. This makes this type of Fuzzy model suitable for the modeling of nonlinear relations. As so, they can be used instead of nonparametric models. The biggest advantage of Fuzzy models, when compared to nonparametric techniques, is that they are interpretable and such they can offer transparency in situations where physical models are difficult to derive (van Lith & al., 2002, 2003). However, for their derivation considerable more knowledge is required than for nonparametric models.

The integration of Fuzzy models along with first-principles knowledge can, as before, be accomplished in parallel (Abonyi & al., 1999; Fu and Barford, 1995b) or in series (van Lith & al., 2002, 2003; Vieira & al., 2005). Moreover they can be complementary combined into an existing hybrid approach, e.g. in parallel to a nonparametric model (Dors & al., 1995, 1996; Peres & al., 2000, 2001) where a gating function decides about the degree of their involvement in the kinetic rate modeling; or in series as an input to the nonparametric model, providing a classification of the operational phase (Beluhan and Beluhan, 2000; Preusting & al., 1996; Schubert & al., 1994a; Simutis & al., 1995).

While the determination of the Fuzzy model parameters in the parallel hybrid case can be accomplished with standard techniques, not all of those techniques can be directly used in the serial approach, see (Preusting & al., 1996; Roubos, 2002; Roubos & al., 2000; Schubert & al., 1994a,b; van Lith & al., 2003) for examples.

2.3.3 Nonparametric Models

The structure of nonparametric models is not specified *a priori* but is instead determined from data. It is the nonparametric model that gives the hybrid model its

flexibility, e.g. to model systems with partially unknown underlying effects, and that also gives the hybrid approach its prediction power. The most frequently applied nonparametric models, are the MultiLayer Perceptron (MLP) (counted in are Feed Forward Neural Networks (FFNN), Artificial Neural Networks (ANN), Neural Networks (NN), Recurrent Neural Networks (RNN), and other approaches that base on the MLP concept) and the Radial Basis Function Network (RBFN), see Table 2.1. Both provide about equally good predictions (in favor of the former (James & al., 2002)), but the determination of the MLP takes considerably longer than the one of the RBFN. This is due to the fact that for the training of the RBFN approach direct use is made of the outputs. The advantage of the MLP is that the outputs (of the nonparametric model) do not need to be known explicitly for the model determination. This is especially important in the serial case, since the kinetic rates cannot be explicitly determined and their calculation from sparse, infrequent noisy measurements might be critical. The advantage of the RBFN is that those networks have certain, inherent stability characteristics, which make them suitable for control and monitoring (James & al., 2002).
However, there exist certain situations (e.g. in case that huge amounts or highly correlated data, such as when spectroscopic data, are integrated) where it is advisable to apply either different nonparametric methodologies, such as (Nonlinear) Partial Least Square models (Henneke & al., 2005; von Stosch & al., 2011b) or to pre-treat the inputs, for instance with Principal Component Analysis (PCA) (Cubillos and Lima, 1997, 1998) (Note that there are also other techniques available such as Independent Component Analysis (ICA), Singular Value Decomposition (SVD) etc., see (Clifford, 2005)).

Similar to the structure of NPLS models, but not capable to deal with huge amounts of highly correlated data, is the one of stacked neural networks. Tian & al. (2001) uses stacked neural networks in a parallel hybrid structure and found that they provide better predictions than if one single neural network is used for modeling the same task. In a similar manner Bollas & al. (2003) used a stack of ANNs whose outputs (various predictions for the same residual) were combined by an additional ANN to obtain the final residual prediction.

Table 2.1 List of nonparametric models and Fuzzy approaches that find application in serial (S) or parallel (P) hybrid models, and the respective publications.

Nonparametric Model	Hybrid Structure	Publications
MLP[1]	S	Abonyi & al. (2007); Acuna & al. (1999); Aguiar and Filho (2001); Al-Yemni (2003); Al-Yemni and Yang (2005); Anderson & al. (2000); Andrasik & al. (2004); Baldi and Chauvin (1996); Bazaei and Majd (2003); Bellos & al. (2005); Beluhan and Beluhan (2000); Bhutani & al. (2006); Boareto & al. (2007); Bollas & al. (2003); Braake & al. (1998); Brendel and Marquardt (2008); Cao & al. (2004a,b); Ccopa Rivera & al. (2006); Chabbi & al. (2008); Chorukova and Simeonov (2008); Conlin & al. (1997); Corazza & al. (2005); Cubillos & al. (1996); Cubillos and Lima (1998); Dadhe & al. (2001); Dors & al. (1995, 1996); Emmanuel & al. (2009); Eslamloueyan and Setoodeh (2011); Feil & al. (2004); Fellner & al. (2003); Feyo de Azevedo & al. (1997); Fu and Barford (1995a); Georgieva and de Azevedo (2009); Georgieva and Feyo de Azevedo (2007); Georgieva & al. (2003); Ghosh & al. (2000); Gnoth & al. (2008b); Goncalves & al. (2002); Guo & al. (1997); Gupta & al. (1999); Harada & al. (2002); Hinchliffe & al. (2003); Ibrehem & al. (2011); Ignova & al. (2002); James & al. (2002); Jenzsch & al. (2007); Kahrs and Marquardt (2007, 2008); Karama & al. (2001a,b, 2010); Kasprow (2000); Kim and Chang (2000); Lee & al. (2002); Leifsson & al. (2008); Madar & al. (2004, 2005); Mazutti & al. (2010); Mogk & al. (2002); Molga and Cherbanski (1999); Molga and Westerterp (1997); Molga (2003); Molga & al. (2000); Nascimento & al. (1999); Ng and Hussain (2004); Oliveira (1998, 2004); Patnaik (2001, 2010, 2003, 2004); Peres & al. (2008, 2000, 2001, 2003); Piron & al. (1997); Ploemen (1996); Porru & al. (2000); Preusting & al. (1996,?); Psichogios and Ungar (1992); Qi & al. (1999); Reuter & al. (1993); Roubos & al. (2000); Saraceno & al. (2010b,c); Saxen and Saxen (1996); Schenker and Agarwal (2000); Schubert & al. (1994a,b,b); Silva & al. (2000, 2001); Simon & al. (2006); Simutis & al. (1995); Simutis and Luebbert (1997); Simutis & al. (1997); Teissier & al. (1997); Teixeira & al. (2005b, 2007a, 2006); Thibault & al. (2000); Tholudur and Ramirez (1996); Tsen & al. (1996); van Can & al. (1996, 1999, 1998, 1997); von Stosch & al. (2010); Wei & al. (2007); Wilson and Zorzetto (1997); Zabot & al. (2011); Zahedi & al. (2005); Zander & al. (1999); Zbicinski & al. (1996); Zhang & al. (2006); Zorzetto & al. (2000); Zorzetto and Wilson (1996); Zuo & al. (2006); Zuo and Wu (2000)
RBFN[2]	S	Chen & al. (2000); Cubillos and Acuna (2007); Cubillos & al. (2001); Dadhe & al. (2001); Graefe & al. (1999); Hanomolo & al. (2000); James & al. (2002); Patnaik (2008); Peres & al. (2008); Thompson and Kramer (1994); Vande Wouwer & al. (2004); Zahedi & al. (2005, 2011)
PLS[3]	S	Carinhas & al. (2011); Henneke & al. (2005)
NPLS[4]	S	von Stosch & al. (2011a,b)
Fuzzy	S	Dors & al. (1995); Peres & al. (2000, 2001); Preusting & al. (1996); Roubos & al. (2000); Schubert & al. (1994a,b); van Lith & al. (2002, 2003); Vieira & al. (2005); Vieira and Mota (2005)
Wavenet	S	Safavi & al. (1999)
FLN[12]	S	Ccopa Rivera & al. (2006); Costa & al. (1999, 1998); Harada & al. (2002); Henriques & al. (1999); Mantovanelli & al. (2007)
TSE[6]	S	Sohlberg (2005)
SVM[8]	S	Kim and Kim (2005, 2006); Wang & al. (2010a); Yang & al. (2011)
NARX[9]	S	Vieira and Mota (2005)

ME[10]	S	Peres & al. (2008, 2001, 2003, 2004)
SERM[11]	S	van Deventer & al. (2004)
MLP[1]	P	Anderson & al. (2000); Bhutani & al. (2006); Bollas & al. (2003); Chen & al. (2004); Chungui & al. (2009); Conlin & al. (1997); Duarte & al. (2004); Genc (2006); Guclu and Dursun (2008); Hisbullah & al. (2002); Hussain and Ho (2004); Hussain & al. (2001, 2002); Johansen and Foss (1992a,b); Jones & al. (2007); Kamali and Mousavi (2008); Lee & al. (2002, 2005); Leifsson & al. (2008); Masri (1994); Sohn & al. (2008); Su & al. (1992); Su and McAvoy (1993); van Can & al. (1996); Vilim & al. (2001); Xiong and Jutan (2002)
RBFN[2]	P	Kramer & al. (1992); Lee & al. (2005); Linker and Seginer (2004); Potocnik and Grabec (1999); Potocnik & al. (2000); Vande Wouwer & al. (2004)
PLS[3]	P	Crowley & al. (2001); Doyle & al. (2003); Jia & al. (2011); Lee & al. (2005)
NPLS[4]	P	Klimasauskas (1998); Lee & al. (2005)
QPLS[7]	P	Lee & al. (2005)
Fuzzy	P	Abonyi & al. (1999)
MARS[5]	P	Duarte & al. (2004); Duarte and Saraiva (2003)
SVM[8]	P	Hu & al. (2011)
Stacked MLP	P	Tian & al. (2001)

MLP[1]: MultiLayer Perceptron
RBFN[2]: Radial Basis Function Network
PLS[3]: Partial Least Square or Projection to Latent Structures
NPLS[4]: Nonlinear PLS
MARS[5]: Multivariate Adaptive Regression Splines
TSE[6]: Taylor Series Expansion
QPLS[7]: Quadratic PLS
SVM[8]: Support Vector Machines
NARX[9]: Nonlinear AutoRegressive eXogenous
ME[10]: Mixture of Experts
SERM[11]: Semi-Empirical Regression Model
FLN[12]: Functional Link Network

Nonparametric models for specific problems

The concept to use more than one neural network was also explored by Cao & al. (2004a,b); Gnoth & al. (2008b); Gupta & al. (1999); Patnaik (2001, 2010, 2003); Piron & al. (1997); Preusting & al. (1996); Reuter & al. (1993); Silva & al. (2000, 2001) in serial hybrid models. Preusting & al. (1996) used two ANNs in parallel two model separate phenomena, i.e. one ANN to model the kinetics another to model the viscosity. Gupta & al. (1999) applies two parallel ANNs, each of which inferring a variable value, in series with another three parallel ANNs, each of which estimating a quantity that enters as an input to the mechanistic model. In Gnoth & al. (2008b); Silva & al. (2000, 2001) the prediction of one central kinetic rate (usually the specific biomass growth rate) by a first ANN, was used as an input (beside others) to another ANN, which in turn predicts another rate e.g. the product formation rate. It was shown that by doing so, lag phases which can occur when e.g. the main substrate in a fermentation is changed, can be modeled.

The modeling of each subtask in the hybrid model with one individual nonparametric model, as e.g. done by Patnaik (2001, 2010, 2003); Piron & al. (1997); Saraceno & al. (2010b) can help to make the model structure more transparent, and increase the accuracy of each predicted quantity. A difference to the individualization of the predictions is the structure by Cao & al. (2004a,b) who applied two individual nonparametric models to predict the same quantity but each model relying on different phenomena i.e. the inputs are different.

Other approaches incorporate nonparametric models that can, at least to some extent be analyzed and interpreted, such as a Taylor Series Extrapolation (Sohlberg, 2005), Multivariate Adaptive Regression Splines (MARS) (Duarte & al., 2004; Duarte and Saraiva, 2003), Semi-Empirical Regression Models (van Deventer & al., 2004) or Nonlinear AutoRegressive eXogenous (NARX) models (Vieira and Mota, 2005). The Fuzzy model applications, named above, could also be counted thereto.

Recently, Support Vector Machine (SVM) models, which are said to be fast and easy to train, find more and more application (Hu & al., 2011; Kim and Kim, 2005, 2006; Wang & al., 2010a; Yang & al., 2011). Fast training properties and fast on-line adaption is also reported for Functional Link Networks (Costa & al., 1999, 1998).

Comparison of nonparametric models

Comparisons between several in the hybrid model embedded nonparametric models have been carried out, but the findings are sometimes contradicting. This might be due to the fact that the performance of the nonparametric model is highly case dependent (what kind of function should be represented, how many data points are available, how many parameters does the nonparametric model have, what training algorithm is used, what are the properties of the in- and outputs, etc..) wherefore it is difficult to draw general conclusions. Nevertheless, the accomplished comparisons are listed in the following in order to provide some intuition for nonparametric model use.

MARS, regression analysis and ANN were compared by Duarte & al. (2004); Duarte and Saraiva (2003) in a parallel set-up, concluding that if sufficient training data are available the MARS is the best approach. Lee & al. (2005) made a comparison of several nonparametric models (FFNN, RBFN, PLS, QPLS and NPLS) embedded in a

parallel hybrid structure for the modeling of a full scale wastewater process, finding that the NPLS approach (through which fault detection is also enabled) provides the best predictions. Predictions of serial hybrid model in which RBFN and RB are embedded were found to be better than those with embedded ANN by Zahedi & al. (2005), which however might be due to the relative low number of experimental data. Kim and Kim (2005) compares an ANN to SVM model in a serial approach, and states that the SVM approach performs better. Wang & al. (2010a) compared the SVM to RBFN, incorporated into a serial structure, finding that the structure of the RBFN is, in general, greater. Also for a serial structure, Vieira and Mota (2005) compared a NARX to a Fuzzy approach, concluding that the performance of the NARX model depends stronger on the *a priori* knowledge than the one of the TSK Fuzzy model, but that it is considerably easier to adapt the NARX on-line. A comparison of an ANN to a more physical model, namely a polynomial approach, was already conducted by van Can & al. (1996) in a serial hybrid approach. Therein it was concluded that the development of the ANN requires less knowledge, while giving better extrapolation properties.

2.3.4 Identification Schema

The identification of the hybrid model unknown parts, most times comprises only the identification of the nonparametric model parameters (which is also referred to as training). This identification is accomplished by minimizing an objective function value through manipulation of the parameter values. The objective function usually consists of a part accounting for the fit of the model predictions to the experimental data. Additionally, the objective function can contain a regulation term which e.g. can enhance the generalization capabilities of the model (Hu & al., 2011; Kahrs and Marquardt, 2008; Vande Wouwer & al., 2004). While, in principle the same identification schema can be applied when also other hybrid model parts are unknown, e.g. yield/stoichiometric coefficients, it might, in this case, be advantageous to decompose the identification since e.g. usually the initial values of such parameters are known which might simplify the identification. Approaches explicitly dealing with this scenario are given in Kahrs and Marquardt (2008); Yang & al. (2011), the latter is shortly presented below.

In case of the serial structure **C**, Fig. 2.1, or the parallel structure **A**, where in the latter the nonparametric model predicts the residual between experimental data and mechanistic model predictions, the identification of the nonparametric models can, in

principle, be carried out with standard techniques (for instance for MLPs the well known back-propagation algorithm can find application (Werbos, 1974)).

In case of the serial hybrid structure **B**, Fig. 2.1, the determination is slightly more difficult since e.g. the kinetic rates cannot be measured and their reconstruction from sparse, infrequent and noisy experimental data is prone to error (Oliveira, 2004; Schubert & al., 1994a). Nevertheless, the direct approach, in which their reconstruction is required, is frequently considered. Two alternative approaches are the indirect approach, which is based on the sensitivities equations and the incremental approach. All three are described below.

The direct approach

For the direct approach at first the outputs e.g. the kinetic rates, are calculated from the experimentally measured state values. This can e.g. be applied through a Taylor-Series approximation (Tholudur and Ramirez, 1996) or through smoothing spline approximations (Schubert & al., 1994a). With these calculated outputs and the available inputs, readily available standard techniques can be used for the weights identification. However, a fact that has found few attention is the statistical optimality of the model state estimations with respect to the experimental data. This is interesting, since the identification is accomplished from kinetic data which were in turn calculated from the experimental data. Thereby the calculated kinetic data might be biased and so might be the model estimates.

The incremental approach

The incremental approach, proposed by Kahrs and Marquardt (2008), is ideal for relatively large systems, since the identification problem is at first decomposed into four smaller problems which are solved sequentially. During this phase standard training techniques for the identification of the nonparametric model can be used. Once the four sub-identifications are accomplished, an overall simultaneous parameter estimation is carried out in order to obtain predictions which are estimated in a statistically optimal sense. Theoretically, i.e. if the gradients with respect to the parameters can be analytically determined, the sensitivities approach can be utilized for the simultaneous identification step.

The indirect approach – the sensitivities equations

Right from the beginning of serial hybrid modeling, a schema for the identification of the neural network weights was required and such Psichogios and Ungar (1992) adapted the well known error back-propagation technique (Werbos, 1974) by using sensitivities equations (Caracotsios and Stewart, 1985). Acuna *&* al. (1999); Oliveira (2004); Schubert *&* al. (1994a) compared this so called sensitivities method to the direct identification approach. They noted that in the presence of few noisy measurement data the reliability of the calculated reaction rate suffers from the accurate determination of the time-derivative. The sensitivities approach can be used to train both one-step and multi-step ahead predictor models. Further, in case of a one-step ahead predictor structure, the number of input data that are used to establish the correlation between inputs and outputs can be significantly greater than with standard techniques wherefore better noise rejection properties are yielded, for details see von Stosch *&* al. (2011a).

Other alternative approaches

A state transformation technique was used by Chen *&* al. (2000) to decouple the kinetic rates, such decomposing the identification. Thereupon the application of an observer based estimator was proposed to estimate the values of the rate expressions. Subsequently standard techniques can find application for the training of the nonparametric model.

Another identification procedure, referred to as direct optimization (Madar *&* al., 2004), only requires the residual (between experimental data and hybrid model prediction), since a derivative free optimization routine is applied in which the prediction error is minimized through the adaptation of the weights (Madar *&* al., 2004; McKay *&* al., 1998; Roubos *&* al., 2000). While this approach of course has the advantage that it does not required the analytic determination of the derivatives, it might result in high computational costs (Roubos *&* al., 2000). However, with the ever increasing computation power and due to the fact that several random initiations of the parameters might not be required, this is an attractive solution for relatively small systems. Also in the case that the gradients are not continuous this might be a good choice (Istadi and Amin, 2006).

Another identification approach that addresses the generalization capabilities of parallel hybrid models is proposed by Potocnik and Grabec (1999). Therein those data that

are not rich in information are removed from the sample space on which the hybrid model is trained, in order to determine the "optimal parameters". However, while this might be a good procedure to develop hybrid models for certain phases of fermentations, the overall process representation might, most probably, suffer.

An identification approach for block-oriented hybrid models, that base e.g. on Hammerstein, Wiener or on feedback block-oriented models is proposed in Pearson and Pottmann (2000).

General Remarks about the identification

Two well known problems that are usually encountered during the identification are over-fitting and local minima. While the former is usually addressed with early-stopping/cross-validation or the above mentioned penalty term in the objective function, the latter is tackled by performing several identification runs for one structure starting from random parameters initializations.

Since convergence and success of the identification depend on the initialized parameter values (Kahrs and Marquardt, 2008), and since relatively small weight values are preferential due to the better generalization capabilities, the initialization values of the weights are, many times, constrained in size, e.g. smaller than one, greater than minus one. Additionally, in case that (i) only few experimental values exist and (ii) a simple model of the kinetic rates is available, the model can be used to provide kinetic rate data for a pre-identification (before the identification relying on the experimental data is carried out) of the nonparametric model parameters (Galvanauskas & al., 2004; Graefe & al., 1999; Henriques & al., 1999; Tsen & al., 1996).

Whenever the balance are posed in the form of Ordinary Differential Equations (ODEs), then, in principle, some boundary condition must be provided for the numerical integration, such as initial values. Since these initial values when taken from the experimental data most probably contain a certain amount of measurement noise, error propagation can occur (von Stosch & al., 2011b). It depends on the underlying set of ODEs whether the error is amplified or damped along time. In order to diminish the impact of such errors on the parameter identification, Vande Wouwer & al. (2004) proposed to include the initial values into the set of parameters (after those have been optimized to a certain threshold) and to, then, optimize all those values together.

2.3.5 Model structure and Extrapolation Capabilities

The model structure and the extrapolation capabilities are directly related. This not only concerns whether the structure is parallel or serial, but also concerns the structure of the nonparametric model, especially its size.

The nonparametric model and extrapolation

The determination of the nonparametric model size (e.g. in case of MLP the number of hidden layers and the therein covered numbers of nodes, or in case of PLS the number of latent variables) can be addressed with the Akaike Information Criterion (AIC) or Bayesian Information Criteria (BIC), the latter being more suitable for models with large numbers of parameters (Peres & al., 2008; von Stosch & al., 2010). Also other statistical criteria can be applied (Bollas & al., 2003; Kim and Chang, 2000) to evaluate the predictions obtained with different sized nonparametric models. In general, the estimation quality must be balanced against the number of involved parameters and against the number of data (the data content) that are available for the identification. A manual assessment, at least of the best candidate structures, is always advisable (Braake & al., 1998).

The number of parameters and the identified "optimal" parameter values determine the achievable prediction quality and the error when extrapolating. The larger the number of parameters and the smaller the number of data these parameters are identified on, the lower is the statistical confidence in the identified parameter values. This can, for instance, permit to locate economical process regions during an optimization (Mogk & al., 2002).

As a matter of fact, the integration of knowledge can significantly reduce the size of the nonparametric model, while enhancing the extrapolation properties, which is one of the reasons hybrid, semi-parametric, gray-box and similar techniques find application. A visual example for the impact that knowledge incorporation can have on the achievable extrapolation capability and on the size of the nonparametric model is demonstrated in Mogk & al. (2002); Schuppert (1999). A theoretical assessment of the relation between knowledge incorporation, the size of the nonparametric models, identifiability and extrapolation is made by Fiedler and Schuppert (2008) for scalar tree structured hybrid models.

Hybrid model structures and Extrapolation

A systematic investigation on the hybrid model extrapolation properties was conducted by van Can & al. (1996, 1999, 1998, 1997), distinguishing between dimensional extrapolation, range extrapolation, interpolation and frequency extrapolation, see Fig. 2.3.

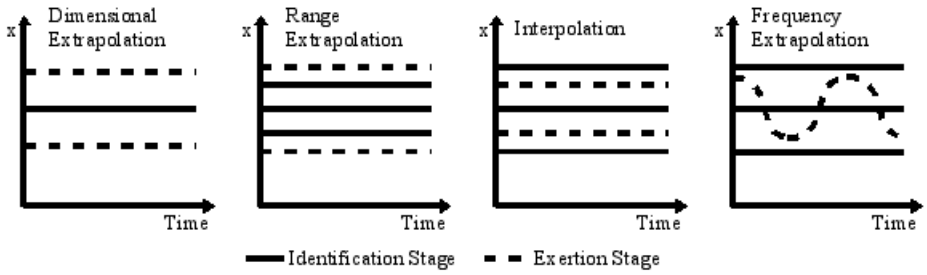

Figure 2.3 Schematic sketches for dimensional extrapolation, range extrapolation, interpolation and frequency extrapolation.

When testing serial and parallel hybrid models, through their incorporation into a model predictive control schema, experimentally for their dimensional extrapolation properties, van Can & al. (1996) observed that the serial hybrid model showed good dimensional extrapolation properties. These properties were found to be due to the accurately known terms in the balances. The parallel hybrid models, in contrast, did not show any advantage compared to pure nonparametric models. Similar observations were made by Anderson & al. (2000). Klimasauskas (1998) proposed to apply some measure, e.g. a confidence module, to restrict the influence of the parallel non-linear model on the prediction when extrapolating.

Range, dimensional and frequency extrapolation were studied in van Can & al. (1998) for different levels of incorporated mechanistic knowledge. It was found that due to the accurately known terms in the balances, good dimensional and reliable frequency extrapolation properties were obtained, and that the unknown terms could relatively easy be identified from the available data. Further, in comparison to more data-driven models, the serial gray-box models have better frequency and dimensional extrapolation properties. Thus with the same identification data, the model can be applied to a much wider range of conditions. This statement can also be translated into the fact that a smaller domain of identification data for the serial hybrid models is required, wherefore

the experimental effort is limited. The strong connection between the model properties and the identification data, which will be subject of experimental data section, is thus clearly outlined.

Measures for model extrapolation

The application of hybrid models to the off-line process optimization or to the off-line controller tuning can result in extrapolating situations, i.e. the nonparametric model is confronted with combinations of input values, which it has not been trained for. The risk of wrong predictions can be expected to rise, the larger the distance between the current combination and the trained combinations is. In such a case it is necessary to constraint the optimization by some measure to avoid false decisions.

As mentioned before, Klimasauskas (1998) proposes to use some kind of confidence module to restrain the impact of the nonlinear model when extrapolating, however no specifics are shown. In Simutis *&* al. (1995) a clustering procedure is applied to the ANNs inputs, in order to determine the contribution of different ANNs to the rate predictions. In Teixeira *&* al. (2005b) clustering of the nonparametric model inputs is carried out using the k-means algorithm. Then, the optimization is constrained by a user defined risk level (typically 80%) that takes the minimal distance between the inputs obtained during the optimization and the closest cluster mean into account. The risk level calculated along the trajectory is used to determine those instances of time in which the information content of the samples is the largest, i.e. where the risk is the highest. In Kahrs and Marquardt (2007) two complementary criteria, i.e. convex-hull criteria and a confidence interval criterion, are proposed to check the validity domain of hybrid models. It is checked with the convex-hull criteria whether each empirical model only interpolates the data encountered during model identification, while with the confidence interval criterion the confidence intervals for the hybrid model predictions are assessed. When comparing between the convex-hull and the clustering technique, the convex-hull criteria has the advantage that it can be implemented as a set of linear constraints, while the clustering technique is a non-linear constraint, but the convex-hull criteria might be too optimistic when the data distribution is strongly non-uniform, which is not the case for clustering. Thus a combination of both, even though increasing the computational cost, might in certain situations be advantageous.

A shortcoming of the clustering and the convex-hull criteria is the case of frequency extrapolation, mentioned above, since these criteria are somewhat focused on the distri-

bution of the measured points in the space and do not account for the dynamics. When optimizing for instance the feeding control policy in a fed-batch case, dynamics might however play a major role since the feeding rate might vary from one discretized instance to another. Investigations in this respect are especially interesting in cases in which the transient behavior is of importance such as for controller tuning.

2.3.6 Experimental data and pre-treatment

Data are necessary to identify the structure and the parameters of the hybrid model and basically all model properties (prediction quality, extrapolation capabilities) depend heavily not only on the quantity but also on the quality of the data.

Design of Experiments

While it is of course not feasible to manipulate experiments carried out in the past, and the attitude tellingly described in Sohlberg (2005), i.e. "you have to take what you can get" is dominant in industrial settings, it needs to be clear what is aimed at (Simutis & al., 1997). If no data at all, nor any knowledge about the system at hand, is available, then a systematic exploration of the process design space, through experimental design, can be a valuable choice. Examples would be factorial design (Ccopa Rivera & al., 2006; Gupta & al., 1999; Mantovanelli & al., 2007; Saraceno & al., 2010b) or sequential pseudo-uniform design (Chang & al., 2007). A screening design might, depending on the context, provide a very good starting point. In van Can & al. (1996) it is outlined that the design of an identification experiment should be such that the unknown part of the model is almost completely discovered, however it is rather difficult to know these in advance. Another option, if at least some knowledge or data are available, is to apply the coverage approach proposed by Brendel and Marquardt (2008), which proved to be better than the factorial design. Just recently, cell specific knowledge has been used along with data driven approaches to identify the optimal experimental design for culture media development (Ferreira & al., 2011).

A different idea is the one of iterative batch to batch optimization, where neither the exploration of the design space nor the model properties are in spotlight but the optimization of some objective (Doyle & al., 2003; Teixeira & al., 2006) e.g. the maximization of the total amount of product quantity (Teixeira & al., 2006). This means that the experiment is performed in such a way, i.e. the degrees of freedom are chosen in such a manner, as to meet the objective. It is of course intelligent to take samples

during the experiments at those instances of time at which the uncertainty about (the calculated risk of) the process trajectory is the highest (Teixeira & al., 2006).

The question arising for every case is thus, whether it is generally better to explore the design space first and to perform an optimization then or whether iterative batch-to-batch optimization can reach the optimum with less experiments. Of course the best strategy can also be a mixture of both.

Experimental Data Pretreatment

The pretreatment of the experimental data, especially of those that are inputs to the nonparametric model, was found to result in significant improvements of the nonparametric model performance (Bishop, 1995). In hybrid modeling, experimental data cannot only enter the model as inputs to the nonparametric submodel but also directly, e.g. as experimental data of the feeding rate or as concentration data considered in the semi-parametric model. It is for instance pointed out in Chabbi & al. (2008); Schubert & al. (1994a) that variances in the feeding concentration can cause big errors in the estimation of the respective substrate concentrations. Similar observations were made by von Stosch & al. (2011b). Studies on the impact of different levels of experimental noise on the identification results, performed by Yang & al. (2011), revealed that the variance of the identified model parameters increases with increasing level of noise. Thus pretreatment of the experimental data can be a valuable procedure to increase the model performance. There are many techniques available to filter the noise, remove off-sets, etc. It depends on the kind of measurement device used and on the context in which the measurement is performed, which pre-treatment technique is the most suitable. While the reader is referred to specialized literature, it can be said that AutoAssociative Neural Networks seem to enjoy great popularity (Galvanauskas & al., 2004; Patnaik, 2003, 2004; Simutis & al., 1995).

2.4 Application of hybrid modeling

In this section the application of hybrid approaches for (i) the modeling of systems in different areas are briefly presented; and (ii) monitoring, control, optimization, scale-up and model reduction are discussed.

2.4.1 Modeling

The number of hybrid models with application to chemical or biochemical engineering is about the same, see Fig. 2.4. There are also applications in water treatment processes, mechanical engineering and other areas, all of which are addressed in the following.

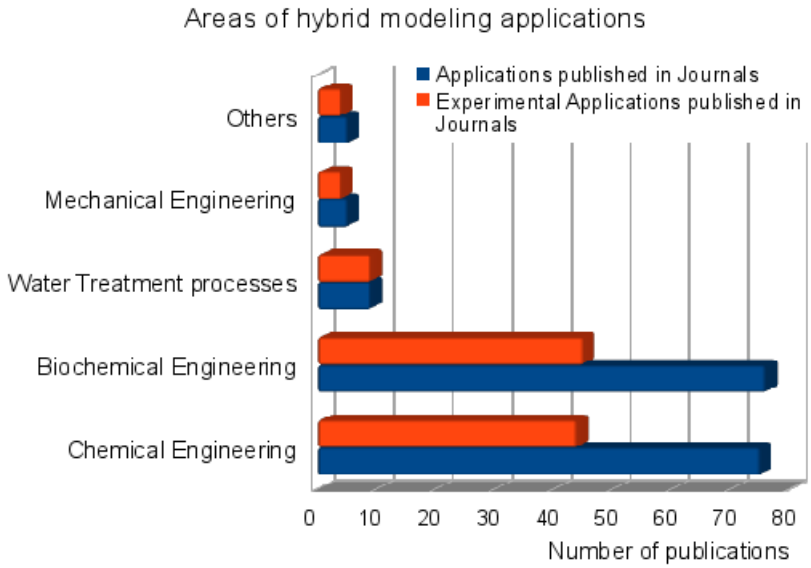

Figure 2.4 Number of publications on hybrid modeling over the area of applications and with respect to the type of data used.

Chemical Engineering

A list of hybrid model applications in the field of chemical engineering are compiled in Table 2.2. Since the number of applications is relatively large, only some of which are discussed in the following, namely those that award with solutions of more complex problems.

The particle size distribution, which is of major interest in many processes, can be modeled with population balances, for instance in crystallization (Georgieva & al., 2003; Lauret & al., 2000), milling (Kumar Akkisetty & al., 2010, 2009) or polymerization (Crowley & al., 2001; Doyle & al., 2003). The application of a complementary non-parametric model, e.g. in order to enhance the prediction quality, can also be beneficial

in this context.

While a parallel set-up (Crowley & al., 2001; Doyle & al., 2003) is relatively easy to apply and might be sufficient in many cases, a serial approach can help to understand the complex interactions. Further, those elements of the model that, due to variations in each batch, are uncertain can be linked to current process measurements, thus accounting for these variations (Kumar Akkisetty & al., 2010). For example Georgieva & al. (2003) applies a full set of mechanistic model equations, namely material, energy and population balances, wherein the most uncertain parts, namely the nucleation rate, growth rate and the agglomeration kernel, are modeled through nonparametric techniques. Similarly, Kumar Akkisetty & al. (2010, 2009) model the dynamics in the mass exchange between the different size intervals of milled ABI ribbons, applying a discrete one-dimensional population balance framework in which the selection and the breakage function are represented through ANNs that base on process measurements, such as the revolutions per minute or the milled mass.

In certain situations it might be necessary or desired to account for gradients in the temperature or concentration distribution along a spacial component. In Gupta & al. (1999) the material balances are formulated for the phosphate particles along the height of a flotation column, wherefore partial derivatives appear, thus resulting into Partial Differential Equations (PDEs). The reaction rate parameters in those balances, namely the flotation rate constants for phosphate and the flotation rate constants for gangue, are modeled through ANNs. Similarly, temperature and concentration gradients along the reactor length are represented in the component mass and energy balance of solid and fluid phases, by Zahedi & al. (2011). In Dadhe & al. (2001) the distillation column is divided into several theoretical stages, each of which assumed to be homogenous, wherefore the material and energy balances formulated for the liquid and vapor phases at each stage take the form of ordinary differential equations. The vapor-liquid equilibrium in this serial hybrid approach is described by a RBFN. Similar approaches are also proposed by Arahal & al. (2008); Hinchliffe & al. (2003), where Hinchliffe & al. (2003) divides the polymerization reactor into several stages whereas Arahal & al. (2008) uses discrete volume and wall segments. Also in Hinchliffe & al. (2003) the form of the molecular-weight distribution is specified, namely it is modeled via the weight fraction of the polymer chain length described by a Flory distribution.

The difficulty in the just named approaches is that for the training of the nonparametric model, sufficient data must be available, and that a nonparametric model trained with global data might not perform well locally. However, it is for instance shown in Molga and Cherbanski (1999) that a complex heterogeneous reaction system can be well presented by a serial hybrid model which is based on the overall material and energy balances. Similar observations were also made by Qi & al. (1999) who compared the hybrid models to two-dimensional models, finding that the hybrid is simpler in model structure, has lower computational costs and provides about the same prediction quality.

Table 2.2 Hybrid Model applications for the modeling of chemical processes

Reference	Application	Hybrid Structure	Nonparametric Model
Johansen and Foss (1992b)	Modeling of a pH neutralization process (exp.[1])	P	ANN
Johansen and Foss (1992a)	Modeling of a pH neutralization process (exp.[1])	P	ANN
Reuter & al. (1993)	Modeling of a 1. Zinc ferrite leaching; (exp.[1]) 2. a Jarosite precipitation; (exp.[1]) 3. a Tennessee copper rougher circuit; (exp.[1]) and 4. a Nchanga sulphide rougher circuit (exp.[1])	S	ANN
van Can & al. (1996)	Modeling and Model Predictive Control of a pressure vessel (exp.[1])	S/P	ANN / polynomial
Tsen & al. (1996)	Modeling and Model Predictive Control of a batch emulsion polymerization of vinyl acetate (exp.[1])	SC[3]	ANN
Cubillos & al. (1996)	Modeling of the dynamic behavior of two drying systems: 1: A direct flaw rotary dryer; (vir.[2]) and 2. A batch fluidized bed dryer (vir.[2])	S	ANN
Molga and Westerterp (1997)	Modeling of the kinetics of a catalytic hydrogenation reaction in a gas-liquid-solid system, namely the kinetics of the hydrogenation of 2,4-DNT over a palladium on alumina catalyst (exp.[1])	S	FNN
Guo & al. (1997)	Modeling of the gasification of two coals, carried out in a batch feed fluidized bed reactor at atmospheric pressure (exp.[1])	S	ANN
Qi & al. (1999)	Modeling of a benzene oxidization to maleic anhydride in a wall-cooled fixed-bed reactor (exp.[1])	S	NN
Gupta & al. (1999)	Modeling of a a phosphate flotation column, relating the effects of operating variables such as frother concentration to the column performance (exp.[1])	S	ANNs
Nascimento & al. (1999)	Modeling of the finishing stage of an industrial nylon-6,6 polycondensation in a twin-screw extruder reactor (exp.[1])	SC[3]	ANN
Zander & al. (1999)	Modeling of: 1. An ethane pyrolysis in a laboratory-scale plug-flow tubular fixed-bed reactor; (exp.[1]) 2. A base-catalyzed ethoxylation of dodecanol in a lab-scale continuous stirred tank reactor (exp.[1])	S	ANN
Safavi & al. (1999)	Optimization of a pilot scale distillation column which separates a binary mixture of water and ethanol in continuous operation (exp.[1])	S	Wavenet network

Molga and Cherbanski (1999)	Modeling the hydrolysis of propionic anhydrite catalysed with sulphuric acid in batch and semibatch stirred tank reactors (exp.[1])	S	MLP
Porru & al. (2000)	Modeling and Monitoring of a catalytic oxidation of carbon monoxide in a heterogeneous gas-solid reactor (exp.[1])	S	FNN
Molga & al. (2000)	Modeling the oxidation of 2-octanol with nitric acid in a reaction calorimeter (exp.[1])	S	NN
Ghosh & al. (2000)	Integrated product engineering of: 1. Fuel-additive performance; (exp.[1]) and 2. Rubber design (exp.[1])	S	ANN
Potocnik & al. (2000)	Modeling of the liquid phase methanol synthesis in a continuously stirred reactor (exp.[1])	P	RBFN
Lauret & al. (2000)	Modeling of the crystal growth rate in a batch-type evaporative crystallizer, called vacuum-pan in the sugar industry (exp.[1])	S	ANN
Aguiar and Filho (2001)	Prediction of pulping degree, i.e. prediction of the kappa number in a pulp mill (exp.[1])	SC[3]	ANN
Guo & al. (2001)	Modeling of the gasification of biomasses conducted in a fluidized bed gasifier at atmospheric pressure with steam as fluidizing medium (exp.[1])	S	ANN
Tian & al. (2001)	Modeling and optimal control of a free-radical solution polymerization of methyl methacrylate with a water solvent and benzoyl peroxide initiator (exp.[1])	P	Stacked neural networks
Vilim & al. (2001)	Incipient failure detection for a peristaltic pump (exp.[1])	P	ANN
Mogk & al. (2002)	1. Modeling of a continous polymerization; (exp.[1]) 2. Quality management for polymer compounding; (exp.[1]) 3. Metal hybride process development (exp.[1])	S	ANN
Xiong and Jutan (2002)	Control of the temperature in 1. An exothermic batch reactor (vir.[2]) and 2. A real-time CST process (exp.[1])	P	ANN
Molga (2003)	Modeling of a catalytic hydrogenation of 2,4-dinitrotoluene at non-steady state conditions in a multiphase stirred tank reactor (exp.[1])	S	ANN
Bollas & al. (2003)	Up-scaling of a pilot plant for fluid catalytic cracking process to an industrial unit (exp.[1])	S/P	ANNs
van Lith & al. (2003)	Modeling of a batch distillation column, including start-up (exp.[1])	S	Fuzzy TSK
Hinchliffe & al. (2003)	Modeling of a polyethylene production process, i.e. the prediction of important conditions, such as the reactor temperatures, conversions, and the molecular-weight distribution of the polymer (exp.[1])	S	FNN
Georgieva & al. (2003)	Modeling of an industrial scale fed-batch evaporative crystallization process in cane sugar refining (vacuum-pan) (exp.[1])	S	ANN
Chen & al. (2004)	Modeling and Internal Model Control of a continuous reactive industrial distillation column for producing epichlorhydrin (exp.[1])	P	ANN
Milanic & al. (2004)	Modeling of an industrial hydrolysis process for the production of titanium dioxide (exp.[1])	S	ANN
Feil & al. (2004)	Development of a product quality (melt index) soft sensor for the monitoring of industrial medium- and high-density polyethylene polymerization plant (exp.[1])	S	ANN
Bellos & al. (2005)	Modeling of an industrial hydrodesulfurization reactor (exp.[1])	S	ANN

Sohlberg (2005)	Modeling of an industrial pickling (steel) process (exp.[1])	S/(P)	TSE
Zahedi & al. (2005)	Modeling of a differential catalytic hydrogenation of carbon dioxide to methanol in a packed bed reactor (exp.[1])	S	MLP/RBFN
Zhang & al. (2006)	Control of fuel cell breathing in Proton Exchange Membrane Fuel Cells (exp.[1])	SC[3]	ANN
Simon & al. (2006)	Modeling of a three-phase industrial batch reactor (exp.[1])	S	FFNNs
Bhutani & al. (2006)	Optimization of an industrial hydrocracking unit (exp.[1])	S/P/SP	ANN
Cubillos and Acuna (2007)	Adaptive-predictive control of the combustion chamber temperature in a pilot-scale vibrating fluidized dryer (exp.[1])	S	RBFN
Jones & al. (2007)	Modeling of the final mechanical properties of rolled steel from the Port Talbot hot strip mill (exp.[1])	P	ANN
Fiedler and Schuppert (2008)	Modeling and optimization of an industrial continuous polymerization plant. (exp.[1])	S	ANN
Kamali and Mousavi (2008)	Thermodynamic modeling of the extraction of alpha-pinene using supercritical carbon dioxide (exp.[1])	P	ANN
Arahal & al. (2008)	Modeling of thermal storage tanks in the Plataforma Solar de Almeria (exp.[1])	S	-
Georgieva and de Azevedo (2009)	1. Modeling of the precipitation of calcium phosphate; (vir.[2]) and 2. Monitoring and Control of a sugar crystallization (vir.[2])	S	ANN
Hwang & al. (2009)	Modeling the performance of hollow fiber micro-filtration membranes for surface water treatment (exp.[1])	SC[3]	ANN
Kumar Akkisetty & al. (2010)	Modeling of the particle size distribution for the processing of a particulate material in a milling unit, i.e. compacted ABI ribbons milled in a lab scale Quadra conica (exp.[1])	S	ANN
Rusinowski and Stanek (2010)	Modeling of a steam boiler system to estimate the heat loss and the effects on heating (exp.[1])	S	ANN
Zahedi & al. (2011)	Modeling of ethylene to ethylene oxide heterogeneous fixed-bed reactor (exp.[1])	S	RBFN
Jia & al. (2011)	Monitoring of a copper extraction process in a cobalt hydrometallurgy pilot plant (exp.[1])	P	PLS
Hu & al. (2011)	Modeling of the leaching rate in hydrometallurgical process (exp.[1])	P	SVM

exp[1]: experimental
vir[2]: virtual
SC[3]: Serial Structure C model

Biochemical Engineering - Biotechnology

Hybrid modeling is frequently applied for the modeling of bioprocesses, as can be seen in Table 2.3. Most of the reported applications apply the basic framework formulated in Psichogios and Ungar (1992); Schubert & al. (1994a). There are some extensions to this framework, which were already mentioned above in the section on "Additional mechanistic information" or "Operational knowledge rule based information". Despite these, only a few others will be described in more detail below.

The modeling of a crossflow microfiltration process through the application of a serial hybrid model is considered in Piron & al. (1997). Therein, a physical model is derived for

the microfiltration process, wherein those parameters that are unknown, namely the cake resistance, the cake diffusion interface and the concentration gradient, are described by ANNs.

In Thibault & al. (2000) the spatial distribution of filamentous fungi is considered by the derivation of the material balance for the surface apex density. This results into a two-dimensional propagation model for the fungus, wherein the diffusion coefficient is represented by a FFNN.

The production process of bacterial cellulose with a pilot scale airlift reactor is, in Zuo & al. (2006), decomposed into two models, which are separately identified. The first is a standard serial hybrid model, consisting of material balances in which the specific kinetic rates are represented by ANNs, accounting for the biological part of the process. The second is a modified tanks-in-series model of the airlift reactor with wire-mesh draft tube, taking into consideration the hydrodynamic effects. Good results are obtained with both approaches and so the whole process, i.e. the cultivation systems in a modified airlift reactor with wire-mesh draft tubes, is appropriately represented.

Table 2.3 Hybrid Model applications for the modeling of biochemical processes

Reference	Application	Hybrid Structure	Nonparametric Model
Schubert & al. (1994a)	Monitoring, Control and Optimization of bioprocess, namely fed-batch *Saccharomyces cerevisiae* cultivations (exp.[1])	S	ANN/ Fuzzy
Schubert & al. (1994b)	Modeling of a fed-batch *Saccharomyces cerevisiae* cultivation performed in a standard pilot-scale fermenter (exp.[1])	S	ANN Fuzzy
Fu and Barford (1995a)	Modeling of *hybridoma* growth and metabolism for monoclonal antibody production (exp.[1])	S	ANN
Simutis & al. (1995)	Process supervision of an industrial beer production process (exp.[1])	S	Fuzzy ANN
Fu and Barford (1995b)	Modeling of *hybridoma* growth and metabolism for monoclonal antibody production (exp.[1])	S	Expert System
Dors & al. (1995)	Modeling, Monitoring and Optimization of an industrial recombinant protein production form mammalian cell cultures (exp.[1])	S	ANN
Preusting & al. (1996)	Optimization of an industrial penicillin production process (exp.[1])	S	ANN Fuzzy
Dors & al. (1996)	Monitoring, Control and Optimization of an industrial recombinant protein production form mammalian cell cultures (exp.[1])	S	ANN
Saxen and Saxen (1996)	Monitoring of a *Saccharomyces cerevisiae* batch fermentation process	S	FNN/ RNN
Feyo de Azevedo & al. (1997)	Comparison of three modeling approaches, applied to a baker's yeast production in a fed-batch fermenter at laboratory scale (exp.[1])	S	ANN

Piron & al. (1997)	Modeling of a crossflow microfiltration for a baker's yeast suspension (exp.[1])	S	ANN
Simutis and Luebbert (1997)	Considerations for hybrid model developments, demonstrated on a simulated *E.coli* fed-batch fermentation (vir.[2])	S	ANN
Simutis & al. (1997)	General consideration for model development and process optimizations with examples on 1. A fed-batch *Saccharomyces cerevisiae* cultivations; (vir.[2]) 2. An industrial beer brewing process; (exp.[1]) 3. Recombinant protein production by mammalian cell cultivations (exp.[1])	S	ANN
van Can & al. (1997)	Modeling of the enzymatic batch conversion of penicillin G (exp.[1])	S	ANN
Costa & al. (1998)	Optimal control for cell mass production and ethanol fermentation by *Saccharomyces cerevisiae*; (vir.[2]) and Modeling of a ethanol fermentation process by *Zymomonas mobilis* (exp.[1])	S	FLN
van Can & al. (1998)	Modeling of the pH effects on the enzymatic conversion of penicillin G (exp.[1])	S	NN
van Can & al. (1999)	Modeling of the enzymatic batch conversion of penicillin G (exp.[1])	S	NN
Potocnik and Grabec (1999)	Modeling of an industrial antibiotic production by a fed-batch fermentation process in which clavulanic acid is produced as a secondary metabolite by the microorganisms (exp.[1])	P	RBFN
Henriques & al. (1999)	Optimization a fed-batch alcoholic fermentation by *Zymomonas mobilis* using Pontryagin's maximum principle and the singular control theory (exp.[1])	S	FLN
Chen & al. (2000)	Modeling of a fungus cultivation in fed-batch for production of an antibiotic species (exp.[1])	S	RBFN
Peres & al. (2000, 2001)	Modeling of a fed-batch Baker's yeast fermentation process (exp.[1])	S	ME
Zuo and Wu (2000)	Optimization and control of fed-batch *Bacillus thuringiensis* cultivations for thuringiensin production (exp.[1])	S	ANN
Beluhan and Beluhan (2000)	Modeling of an industrial fed-batch yeast cultivation process (exp.[1])	S	ANN Fuzzy
Hanomolo & al. (2000)	Modeling of a batch animal cell culture (exp.[1])	S	RBFN
Thibault & al. (2000)	Modeling of 1. the fermentation of glucose to gluconic acid by the micro-organism *Pseudomonas ovalis*; (vir.[2]) 2. the growth of flamentous fungi in a solid state fermenter; (exp.[1]) and 3. the propagation of flamentous fungi growing on a two-dimensional solid substrate (exp.[1])	S	FNN
Roubos & al. (2000)	Comparison of modeling strategies for the production of clavulanic acid in batch fermentations with *Streptomyces clavuligerus* (exp.[1])	S	ANN Fuzzy
Silva & al. (2000)	Modeling and Monitoring of cephalosporin C fed-batch fermentation production with the aerobic fungus *Cephalosporium acremonium* (exp.[1])	S	FNNs
Ignova & al. (2002)	Online optimization of the feeding control policy for fed-batch penicillin production (exp.[1])	S	ANN
Silva & al. (2001)	Modeling of the duration lag that is cause by differences in inoculum of *cephalosporium acremonium* cultivated in fed-batch reactors (exp.[1])	S	FNNs

James & al. (2002)	Monitoring of biomass in a PHB production using *Alcaligenes eutrophus* fed-batch fermentations (exp.[1])	S	ANN/ RBFN
Goncalves & al. (2002)	Modeling of the synthesis of amoxicillin which is catalyzed by penicillin G in batch fermentations (exp.[1])	S	ANN
Fellner & al. (2003)	Modeling of diacetyl in brewery fermentations (exp.[1]) Functional Nodes dynamic neural network	-	-
Vande Wouwer & al. (2004)	Modeling of batch CHO animal cell cultures (exp.[1])	S/P	RBFN
Peres & al. (2004)	Modeling of a Polyhydroxyalkanoates production in a sequencing batch reactor with mixed cultures (exp.[1])	S	ME
Corazza & al. (2005)	A study on substrate and product inhibition observed in the enzymatic hydrolysis of cellobiose (exp.[1])	S	ANN
Teixeira & al. (2005a,b)	An optimization study of fed-batch Baby Hamster Kidney cultures expressing the human fusion glycoprotein IgG (exp.[1])	S	ANN
Henneke & al. (2005)	Modeling and monitoring of PHB concentrations in high cell density fermentations of *Ralstonia eutropha* based on spectrofluorometry measurements (exp.[1])	S	PLS
Zuo & al. (2006)	Modeling of the hydrodynamic and biological effects to describe the cultivation of *Acetobacter xylinum* for bacterial cellulose production in a modified airlift reactor with wire-mesh draft tubes (exp.[1])	S	ANNs
Jenzsch & al. (2007)	Control of the Carbon-dioxide Production Rate (CPR), hybrid model based monitoring of the specific biomass growth and inference of the CPR set-points using the hybrid model (exp.[1])	S	ANN
Teixeira & al. (2007a)	Optimization of recombinant Baby Hamster Kidney cultures producing a recombinant fusion glycoprotein (exp.[1])	S	ANN
Laursen & al. (2007)	Modeling of an industrial pharmaceutical process, namely production of a foreign protein product in a pilot scale fed-batch fermentation (exp.[1])	S	ANN
Boareto & al. (2007)	Modeling of the lipase production by *Candida rugosa* in batch and fed-batch operation (exp.[1])	S	ANN
Chorukova and Simeonov (2008)	Modeling of a fed-batch process for the enzymatic superoxide dismutase production (exp.[1])	S	ANNs
Gnoth & al. (2008b)	Monitoring of specific biomass growth in fermentation experiments on the laboratory scale with an *E.coli* strain producing a recombinant protein (exp.[1])	S	ANNs
Wang & al. (2010a)	Comparison of modeling strategies using a fed-batch penicillin fermentation (exp.[1])	S	LS SVM
Mazutti & al. (2010)	Modeling of an insulinase production in a batch bioreactor using agroindustrial residues as substrates (exp.[1])	S	ANN
Saraceno & al. (2010b,c)	Modeling of a fermentation of "ricotta cheese whey" for the production of ethanol (exp)	S	ANN
Zabot & al. (2011)	Modeling of a xanthan gum batch bioproduction process by *Xanthomonas campestris* pv. Mangiferaeindicae (exp.[1])	S	ANN
Carinhas & al. (2011)	A metabolic model of *Spodoptera frugiperda* cells (exp.[1])	S	PLS
Eslamloueyan and Setoodeh (2011)	Modeling and Optimization of batch and fed-batch fermentation of xylose-utilizing engineered *Saccharomyces cerevisiae* (exp.[1])	S	ANN

exp[1]: experimental

vir[2]: virtual

Mechanical Engineering

Even though most of the hybrid model applications are found for chemical or biochemical processes, hybrid modeling constitutes a valuable approach also for mechanical engineering, whenever different sources of knowledge can be fused. In Table 2.4 a number of mechanical applications are listed. It can be seen that most applications are oriented towards the modeling of one component in a complex overall system. The number of serial and parallel structure applications is about the same.

Table 2.4 Hybrid modeling in mechanical engineering

Reference	Application	Hybrid Structure	Nonparametric Model
Masri (1994)	Modeling of a physical nonlinear system incorporating the bearing friction phenomena (vir.[2])	P	ANN
Ploemen (1996)	Modeling of two degrees of freedom systems; 1. A flexible servo system; (exp.[1]) 2. An inverted pendulum (exp.[1])	S	ANN
Cao & al. (2004a)	Modeling of the dynamic friction component in powertrain systems (exp.[1])	S	ANNs
Cao & al. (2004b)	Modeling of the dynamic friction component in a vehicle automatic transmission system (exp.[1])	S	ANNs
Sohn & al. (2008)	Modeling of bushing in vehicle suspension systems (exp.[1])	P	ANN
Chungui & al. (2009)	Modeling the hysteretic restoring force of a wire cable vibration isolation system for electronic equipment (exp.[1])	P	ANN

exp[1]: experimental
vir[2]: virtual

Water treatment processes

Several applications of hybrid modeling to water treatment processes, mostly for wastewater treatment processes, can be found, e.g. see Table 2.5. It can be seen that parallel and serial approaches are about equally often utilized. The parallel model seems to provide better estimations than the serial model in case that the whole process is modeled and that the underlying mechanistic model is the so called ASM1 model (Lee & al., 2002).

Table 2.5 Hybrid Model of water treatment processes.

Reference	Application	Hybrid Structure	Nonparametric Model
Cote & al. (1995)	Modeling of an activated sludge process, comprising a biological reactor and a secondary settler (exp.[1])	P	FNN

Conlin & al. (1997)	Modeling of an industrial drinking water treatment plant (exp.[1])	S/P	ANN
Anderson & al. (2000)	1. Modeling of an activated sludge wastewater treatment; (exp.[1]) 2. Optimization and control of an activated sludge operation in an alternating aerobic-anoxic system (vir.[2])	S/P	NN
Karama & al. (2001b)	Modeling of anaerobic wastewater treatment processes (exp.[1])	S	Constrained ANN
Karama & al. (2001a)	Modeling of anaerobic wastewater treatment processes (exp.[1])	S	FNN
Lee & al. (2002)	Modeling of a coke-plant wastewater treatment process (conventional activated sludge unit) (exp.[1])	S/P	ANN
Lee & al. (2005)	Modeling of a coke-plant wastewater treatment process (conventional activated sludge unit) (exp.[1])	P	ANN/ RBFN/ PLS/ QPLS/ NPLS
Guclu and Dursun (2008)	Modeling of an activated sludge wastewater treatment plant (exp.[1])	P	ANN
Karama & al. (2010)	Modeling of anaerobic wastewater treatment processes (exp.[1])	S	Constrained ANN

exp[1]: experimental
vir[2]: virtual

Other areas

The food and beverages industry can profit from hybrid modeling, since the production processes are usually well equipped with instrumentation and at least some knowledge about the process streams can be incorporated e.g. in form of material balances. Other prior knowledge such as physical or chemical knowledge can of course also be incorporated. Wine and beer production are in fact bioprocesses and could be counted in above.

Table 2.6 Hybrid modeling in other areas.

Reference	Application	Hybrid Structure	Nonparametric Model
Teissier & al. (1997)	Monitoring and Control of a yeast fermentation in a wine base medium (exp.[1])	S	RNN
Zorzetto & al. (2000)	Modeling of a batch beer production process (exp.[1])	S	FNN
Saraceno & al. (2010a)	Modeling of Food Convective Drying (exp.[1])	S	ANN
Linker and Seginer (2004)	Modeling of the Greenhouse-climate (exp.[1])	P	RBFN
Leifsson & al. (2008)	Modeling the fuel consumption of a container vessel (exp.[1])	S/P	ANN

exp[1]: experimental

2.4.2 Monitoring

The opportunity to estimate unobserved process parameters and variables during the process through hybrid modeling was contemplated by Psichogios and Ungar (1992). A comparison between the hybrid methodology, a Nonlinear Programming (NLP) optimization and an Extended Kalman Filter (EKF), carried out on a virtual bioprocess, revealed that when no *a priori* model of the unobserved process parameters (specific kinetic rates) was available then the hybrid model estimates are better than those obtained by the other two approaches. Also, for this simulation case, a state reconstruction schema was investigated and similar observations regarding the comparison between the hybrid and the other methods were made.

The application of a serial hybrid methodology, consisting of material balances for substrate and biomass, an ANN for prediction of the specific kinetic rates from readily available on-line measurements and a Fuzzy Model for process phase estimation, to a pilot scale cultivation of *Saccharomyces cerevisiae* was conducted by Schubert *&* al. (1994a). They observed that biomass could be estimated with higher reliability than the substrate concentration, and that the prediction of the latter is very sensitive to the substrate concentration in the feeding. For a similar experimental set-up, Saxen and Saxen (1996) utilized a serial hybrid method based on material balances in which the rates, given by FFNN or RNN, are functions of the estimated biomass, substrate and product concentrations. When applied on-line, the rate predictions and the network parameters are adapted to the process conditions by adjusting a set of correction factors according to element balances and an electroneutrality condition, which can be build upon the online available measurements. By doing so the biphasic growth could be accurately described by the networks.

These three approaches illustrate two scenarios (schematically depicted in Fig. 2.5) to exploit hybrid modeling for monitoring, namely (i) the accurate prediction of certain quantities from the available on-line measurements and/or the model's own predictions is feasible; (Predictor/ Soft-sensor) and (ii) along with the hybrid model predictions a corrector schema can be applied to correct the state predictions and to adapt the model parameters (Corrector). In the following, these two classes will be used to classify other hybrid monitoring approaches.

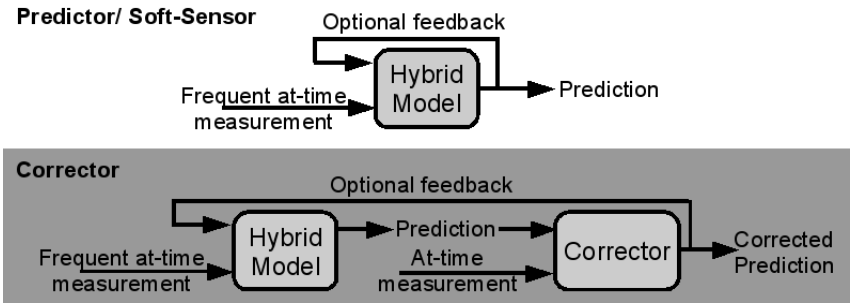

Figure 2.5 Diagram of two possibilities to use hybrid modeling for monitoring.

Soft-Sensor – Predictor Schema

The serial hybrid model is very attractive for the application as a soft-sensor since the kinetic rates, represented by a nonparametric model, can many times be estimated from at-time available measurements and/or the hybrid models' own prediction, see (James & al., 2002; Schubert & al., 1994a). Requirements for the application of these schema are that (i) the sampling rate of the at-time available measurements is more or less constant (a requirement that stems from the numerical integration of the material balances); (ii) that the sampling is carried out frequently enough (also due to the numerical integration); and (iii) that all inputs are available at the same time, eventually some kind of extrapolation schema is required.

When these requirements are met then the serial hybrid model will in principle provide better predictions than other models, since (a) either less parameters are required to achieve similar prediction qualities (when compared to pure nonparametric models) which reduces the statistical uncertainty or full advantage can be taken from the determined actual process conditions, reflected by a set of at-time available measurements (when compared to a pure mechanistic model); (b) the integration of the state variables leads to a smoothing effect which diminishes the influence of noisy measurements on the quality of the predictions (von Stosch & al., 2011a); (c) the hybrid model has better calibration properties; and (d) in the case that the sensitivities method is applied for nonparametric model training, more input data are used for the training (than e.g. for the direct approach), reducing the hybrid models sensitivity to noise, von Stosch & al. (2011a). Table 2.7 comprises a list of hybrid model soft-sensor applications.

The concentrations of cells, glucose and product are on-line inferred using a serial hybrid model by Silva & al. (2000, 2001) for fed-batch fermentations of the strictly aerobic fungus *Cephalosporium acremonium*. While the results do seem convincing, the presented material balances do not since the dilution terms are missing therein (Silva & al., 2000, 2001).

In order to infer the biomass concentration for Poly-β-HydroxyButyric acid (PHB) production in fed-batch fermentations of *Alcaligenes eutrophus* several possible schema (Linear model, FFNN, RBFN or serial hybrid models consisting of the material balance for the biomass concentration in which the specific growth rate is modeled by FFNN or RBFN) were evaluated by James & al. (2002). The results revealed that the hybrid model including the RBFN showed the best performance.

The direct prediction of PHB concentrations in high cell density fed-batch fermentations of *Ralstonia eutropha* through the application of a serial hybrid model (material balance of PHB wherein the reaction coefficient was modeled through a PLS model) was proposed by Henneke & al. (2005). Therein, it is stated that the predictions obtained with this hybrid modeling approach were, in contrast to previously used models for PHB concentration, for the first time sufficiently accurate. A similar approach is used by von Stosch & al. (2011b) for the prediction of concentrations from several on-line available measurements. It is shown that the hybrid approach in this case performs better than pure PLS, and that fault detection features of PLS are restored by the utilized serial approach.

The application of a serial hybrid model is also proposed by Gnoth & al. (2008b); Jenzsch & al. (2007) (same group, same approach, same fed-batch *Escherichia coli*) for the prediction of the specific biomass growth and additionally in Gnoth & al. (2008b) for the prediction of the specific product formation rate. The prediction quality in Gnoth & al. (2008b) is very good but stated to be due to a relative high number of available data. The hybrid model in Jenzsch & al. (2007) is not only used to monitor the process, but also to derive Carbon dioxide Production Rate (CPR) set-points for the control of CPR.

A somewhat similar approach is reported by Boareto & al. (2007), wherein also a serial hybrid model finds application and the inputs to the ANN, that estimates the product formation rate, are the Carbon dioxide Evolution Rate (CER), predicted biomass

concentration, substrate feeding rate. However, even so the results look promising the material balances as before in Silva & al. (2000, 2001) do not seem to be correct.

Table 2.7 The predictor (soft-sensor) approaches applying serial material balance based hybrid methodologies.

Reference	Application	Nonparametric model	On-line Inputs	Predicted Quantity
Schubert & al. (1994a)	Fed-batch *Saccharomyces cerevisiae*	Fuzzy-ANN	Carbon Production Rate (CPR), time, Oxygen Transfer Rate (OTR), Volume, Ethanol concentration and others	Biomass and Substrate concentration
Silva & al. (2000, 2001)	Fed-batch *Cephalosporium acremonium*	FNN	Exhaust gas concentrations of carbon dioxide and oxygen	Biomass, Glucose, Product concentration
James & al. (2002)	Fed-batch *Alcaligenes eutrophus*	FNN, RBFN	Exhaust gas concentrations of carbon dioxide and oxygen, off-gas stream, dissolved oxygen, pH and the models' own predictions	Biomass concentration
Henneke & al. (2005)	Fed-batch, high cell density *Ralstonia eutropha*	PLS	rate of change of fluorescence, the feed rate and previous PHB predictions	PHB concentration
Gnoth & al. (2008b); Jenzsch & al. (2007)	Fed-batch *Escherichia coli*	ANN	CPR, the induction time and the models' own predictions	Specific biomass growth and specific product formation rate
Boareto & al. (2007)	Fed-batch *Candida rugosa*	ANN	Carbon dioxide Evolution Rate (CER), predicted biomass concentration, substrate feeding rate	Lipase production
von Stosch & al. (2011a)	Batch *Bordetella pertussis*	NPLS	Near Infra Red, dissolved oxygen concentration, temperature and pH measurements	Biomass, glutamate and lactate concentrations

Corrector Schema

The application of the corrector schema is very interesting if the state variables (which are e.g. considered in the material or energy balances) can be measured at some instances during the process, because the predictions can be corrected and the model parameters can eventually be adapted. However, the corrector schema is subject to certain restrictions regarding the state observability (Dochain, 2003). The underlying hybrid model can either rely (as in the case of the soft-sensors) on other at-time available measurements or solely on its own predictions (Multi-step ahead predictor), such as in Saxen and Saxen (1996). In case that the hybrid model is serial and uses at-time available measurements the same requirements formulated above for the soft-sensor case

hold.

Wilson and Zorzetto (1997); Zorzetto and Wilson (1996) aimed at a general representation of the system for on-line state estimation, using an EKF along with hybrid models. The hybrid model's state estimate and the networks weights are therein corrected by the EKF, wherefore the system is locally linearized. Reliable state estimations were achieved in both simulation cases. An experimental application of a very similar EKF and hybrid framework to a catalytic reactor, in which the catalytic oxidation of carbon monoxide over Pt-aluminia supported by catalysts takes place, was reported by Porru & al. (2000). Therein it was observed that the reactor dynamics in both the ignition and extinction regions were well described by the hybrid model, even though the model was developed only with steady-state data.

The utilization of parallel hybrid model predictions as a basis for fault detection, demonstrated on a peristaltic pump to detect incipient failure, was reported by Vilim & al. (2001) to result into lower false alarm and missed detection rates than other methods. The nonparametric model in this approach was adapted using the online available state measurements.

A serial hybrid model based soft-senor design was applied by Feil & al. (2004) for the product quality (melt index) prediction in an industrial polymerization process. On the hybrid model basis a DD1 filter was used for the state correction and it was stated that excellent results were obtained.

For a class of serial hybrid models, namely models on which the state transformation technique (which was proposed in Bastin and Dochain (1990)) can be applied for state inferring of unmeasured states, an on-line state correction and an ANNs weight adaptation is proposed in Georgieva and de Azevedo (2009). Additionally the adaptation of uncertain stoichiometric coefficients is investigated in one of the two simulation case studies.

A soft-sensor based on a parallel hybrid model (consisting of a material balanced based model in parallel to a block-wise PLS schema) for copper extraction process in cobalt hydrometallurgy pilot plant is reported in Jia & al. (2011). Therein the hybrid model predictions are corrected by a rectification schema, which can utilize the off-line, time-lagged measured samples. It is stated that the hybrid model predictions are more

accurate and efficient than the other conventional models.

Other monitoring schema

An approach for the determination of the ethanol and sugar concentrations in wine fermentations was discussed in Teissier & al. (1997). This method, consisting of a dynamic RNN and a linear regression based on the assumption that the conversion yields of sugar into ethanol and into carbon dioxide are constant (called therein measurement model), has nothing of the beauty of the prior discussed approaches, but reliable predictions seem to be obtained.

Other approaches, which incorporate first-principle knowledge into the nonparametric model and are therefore classified as gray-box models, utilize e.g. on-line available spectroscopic data for the prediction of component concentrations (Gurden & al., 2001; Mouton & al., 2011; Ramaker & al., 2002; Ruckebusch & al., 2009). It was shown that due to the knowledge incorporation also in these cases better prediction qualities can in general be obtained than when compared to pure nonparametric models. In some of the cases otherwise ill-defined problem formulations could be avoided.

2.4.3 Control

Since hybrid models can accurately capture the process dynamics and nonlinearities their application for process control is logical. Various open- and closed-loop applications are reported, the former will be discussed in the section on optimization.

For the closed-loop case, there are two ways to maximally profit from a hybrid process model, namely (i) by employing a control structure that directly uses the process model equations for the calculation of the control action; or (ii) by application of the hybrid model for the controller tuning.

Hybrid model based controller structure

Schema that directly employ the hybrid process model and that frequently find application are for instance Model Predictive Control (MPC), Feedback Linearizing Control (FLC) and Generic Model Control (GMC).

Whenever the process model equations are invertible, i.e. an analytical explicit expression can be obtained through suitable manipulation, Direct FLC, GMC or Model

(Adaptive) Reference Control (MRAC) can be applied since these schema can account for nonlinearities and are, at the same time, computationally relative inexpensive.

In case that the process model equations are not invertible, FLC or MPC can be applied where the former is computational less expensive while the latter may provide better performance. A list of hybrid model based control applications can be found in Table 2.8.

Model Predictive Control (MPC) In van Can & al. (1996) various investigated hybrid structures were tested experimentally for their properties through their incorporation into the MPC schema. The MPC schema was, therefore, tuned without severely constraining the change of the manipulated variable, such that the closed-loop stability strongly depends on the models' accuracy. The serial hybrid model containing an ANN showed very good performance, when confronted with set-point changes and load disturbances, in this set-up.

The models used for the MPC of the dispersity and molecular weight distribution, in a batch reactor for emulsion polymerization by Tsen & al. (1996), are rather ANNs than hybrid models, see section (hybrid structures). However it is shown that the effect of unmeasured disturbances can be captured by using intermediate measurements which allows to correct for changing raw material properties and output specifications.

The hybrid parallel model discussed in Klimasauskas (1998) is also rather a collection of nonparametric modeling techniques than a hybrid model, defined here as the utilization of different sources of process knowledge. However, the advantages (such as robustness, adaptable modeling and the possibility for nonlinearity representation) which are discussed for a virtual simulation case, also hold for those hybrid models that are built upon different sources of knowledge.

A special form of MPC, namely a predictor corrector control schema, is used by Abonyi & al. (1999), such accounting for model uncertainties, which results into a good controller performance over a wide process range.

The adaptation of the hybrid's nonparametric model parameters, namely the output layer weights in a RBFN, during the MPC was studied by Cubillos & al. (2001). The

comparison to a pure FFNN-MPC schema revealed that due to the lower number of parameters of the hybrid model, those were quicker to adapt (maybe partly owed to the linear adaptation properties of the RBFN). Further, it was noted that the predictions are hard bound, which retains the model to be physically correct. When applying the hybrid model based MPC to an experimental case, namely the control of the combustion chamber temperature in a pilot-scale vibrating fluidized dryer (Cubillos and Acuna, 2007), a rapid adaption to new scenarios and the reduction of perturbation effects were observed.

In Ibrehem & al. (2011) the application of a serial hybrid model based MPC schema and an ANN based MPC approach are investigated for the control of a fluidized bed polymerization system. Both approaches are observed to have about equal performances, in case of output disturbances the hybrid model based controller performed a little better.

Table 2.8 Hybrid Model based Control Structures.

Reference	Application	Hybrid Structure	Nonparametric Model	Controller Design
van Can & al. (1996)	Control of a pressure vessel (exp.[1])	S or P	ANN or polynomial	MPC
Tsen & al. (1996)	Control of dispersity and molecular weight distribution in a batch reactor for emulsion polymerization of vinyl acetate (exp.[1])	SC[3]	ANN	MPC
Klimasauskas (1998)	Control of pH in two continuously stirred tank reactors (vir.[2])	P	NPLS / PLS	MPC
Cubillos and Lima (1998)	1. Control of the product concentration in a CSTR; (vir.[2]) 2. Control of a flotation process (vir.[2])	S	PCA-ANN	MPC / PC
Abonyi & al. (1999)	Control of pH neutralization process (vir.[2])	P	Fuzzy TSK	MPC
Cubillos & al. (2001)	Control of the quantity and quality of vinyl polymerization (vir.[2])	S	RBFN	MPC
Cubillos and Acuna (2007)	Control of the combustion chamber temperature in a pilot-scale vibrating fluidized dryer (exp.[1])	S	RBFN	MPC
Ibrehem & al. (2011)	Control of a fluidized bed polymerization system (vir.[2])	S	ANN	MPC
Cubillos and Lima (1997)	Control of a flotation process (vir.[2])	S	PCA-ANN	Optimal Control
Costa & al. (1998)	Control of the cell mass production and ethanol fermentation by *Sachromyces cerevisae* (vir.[2])	S	FLN	Optimal Control
Costa & al. (1999)	Control of the cell mass production and ethanol fermentation by *Sachromyces cerevisae* (vir.[2])	S	FLN	Optimal Control

Schenker and Agarwal (2000)	Profit Maximization (vir.[2])	SC[3]	ANN	Optimal Control
Anderson & al. (2000)	Minimization of energy costs (vir.[2])	P	ANN	Optimal Control
Vieira & al. (2005)	Control of the temperature in a gas water heater system (vir.[2])	S	Fuzzy TSK	Smith predictive controller
Oliveira (1998)	Control of the concentrations of the precursor and the nitrogen source (exp.[1])	S	ANN	Model Reference Control
Xiong and Jutan (2002)	1. Control of an exothermic batch reactor; (vir.[2]) 2. Control of real-time CST process (exp.[1])	P	ANN	GMC
Abonyi & al. (2007)	Temperature control of a CSTR (vir.[2])	S	ANN	GMC
Hussain and Ho (2004)	1. Control of liquid level in a non-interacting two-tank-in-series system; (vir.[2]) 2. Control of a CSTR with a first-order irreversible exothermic reaction (vir.[2])	P	ANN	Sliding Mode Control
Madar & al. (2005)	Temperature control of a continuous stirred tank (vir.[2])	S	ANN	Feedback Linearising Control
Hussain & al. (2001)	1. Control of an exothermic CSTR with first-order reaction; (vir.[2]) 2. Control of a fermentation process in a continuous biochemical reactor (vir.[2])	P	ANN	Adaptive Feedback Linearising Control
Bazaei and Majd (2003)	Control of a pressure vessel (vir.[2])	S	ANN	Feedback Linearisation

exp[1]: experimental
vir[2]: virtual
SC[3]: Serial Structure C model

Optimal Control & Predictive Control The utilization of optimal control strategies which base on a serial hybrid process model, were proposed by Cubillos and Lima (1997, 1998) for the optimizing control of simulated flotation process. The PCA-ANN model is adapted online through recursive least square and it is stated that: "The structure allows a satisfactory treatment of the main problems associated with flotation operations, such as: non measurable continuous perturbations, time delays, non-linearities, excessive number of degrees of freedom and multiple objective operation." Additionally, in Cubillos and Lima (1998) the serial hybrid model is used as a basis for a MPC of product concentration in a CSTR.

An adaptive optimal control schema based on a serial hybrid model was proposed by Costa & al. (1999, 1998) for the on-line determination of the optimal feeding control policy. The hybrid model predictions and the FLN weights were adapted during the simulation, and the optimal process trajectory, thereupon, recalculated.

An optimal control schema was also applied by Schenker and Agarwal (2000) for the maximization of the profit through manipulation of the switching time in a semi-batch chemical reactor. They observed that the utilization of on-line available measurements and the recalculation of the optimal switching time can increase the controller performance. The hybrid model performed significantly better for longtime predictions than all of the other investigated approaches, such as pure FFNN or EKF.

A similar control schema was applied to a virtual wastewater treatment process by Anderson & al. (2000). While a good fit of the parallel hybrid model estimations and the experimental data could be observed, the control performance utilizing the parallel hybrid model was inferior to the one using a linearized model. The reason for the inferior performance is that a control situation was considered which had an extrapolative character. That parallel hybrid models have poor extrapolation properties, if not restricted by some measure Klimasauskas (1998), was already reported in van Can & al. (1996), which is also resembled by the statement given in Anderson & al. (2000) that the utilization of a hybrid model is not a guaranty for better control performance.

A Smith predictive controller that is based on a serial hybrid model (consisting of energy balances and a Fuzzy TSK model) is used by Vieira & al. (2005) for the temperature control of a virtual gas water heater system. They state that the application of this control design is feasible only due to the better interpretability of their hybrid approach and they present good model predictions and controller performance.

Generic Model Control (GMC) GMC based on hybrid models is applied by Abonyi & al. (2007); Xiong and Jutan (2002). In Xiong and Jutan (2002) a parallel hybrid model is used to model the processes and it is stated that the hybrid model based GMC schema outperformed a self tuning PID and a generalized minimum variance controller in case of the experimental application. A serial hybrid model based GMC is applied by Abonyi & al. (2007) for the temperature control of an exothermic CSTR, where the kinetics are modeled by an ANN.

Feedback linearizing control Feedback linearizing control is used by Bazaei and Majd (2003); Hussain & al. (2001); Madar & al. (2005). In Hussain & al. (2001) a dynamic linear model is used in parallel with a neural network which accounts for all higher order terms. A dead-zone adaptive schema is used to online adapt the network parameters. The controller performance using the hybrid model is found to be better when compared

to that using a pure linear model or a pure neural network. Exact feedback linearizing and approximate feedback linearizing control schema are discussed in Bazaei and Majd (2003) for affine and non-affine serial hybrid model structures, which constitute the process model. It is shown that the knowledge about the structural invertibility can result in significant better controller performance. The best performance was obtained with an affine exact feedback linearizing controller. A feedback linearizing control schema based on serial hybrid models is also proposed by Madar & al. (2005), where Lie derivatives (which stem from the General linearizing control) are used to establish an invertible structure. It is demonstrated therein for a simulation case, that the hybrid model based control schema has certain advantages over the mechanistic model, in that less mechanistic knowledge is required for model derivation and the number of model variables can be reduced, namely to those which are on-line measurable.

Other model-based Control Schema In Hussain and Ho (2004) a parallel hybrid model is embedded into a sliding mode control schema with boundary layer approach. An on-line adaption schema for the ANN model weights is presented. Through the application of the hybrid model the modeling uncertainty is decreased (when compared to a nominal approach) which allows for a higher sensitivity of the controller to the model predictions, resulting into a good controller performance, being significantly better than for the nominal approach. However, for an experimental application the approach will need modifications since it is based on the canonical equations form.

A model adaptive reference control schema that is based on a serial hybrid model is applied in Oliveira (1998), for the control of an experimental penicillin production process. An on-line adaption procedure is applied to update the ANN weights during the process, whenever new data become available. The approach was found to work well during all fermentations.

Hybrid model based controller tuning

Instead of basing the controller upon the hybrid process model directly, the hybrid model can be exploited for the tuning of any controller. Frequently employed approaches comprise Internal Model Control (IMC) and Inverse Model Control (IVMC).

Almost at the beginning of hybrid modeling (Schubert & al., 1994a), an internal model controller that based on a neural net, was tuned with the hybrid model, (i) prior to application off-line and (ii) during the application while also updating the hybrid pro-

cess model weights. The performance was compared to PID control, concluding that the internal model controller can cope with the difficult nonlinear kinetics, in contrast to the PID controller. In Chen & al. (2004) the application of an internal model controller, which bases on the static gain of the parallel hybrid process model, to an industrial reactive distillation column is presented. The IMC is used along with the hybrid model and along with a detuning filter for the calculation of the control action. Significantly better controller performance is obtained when the plant is under the control of the proposed schema. An IMC schema is also used by Zhang & al. (2006), wherein the the controller is based on a neural network and the hybrid model has a serial structure **C**. The neural controller is tuned during the process using the serial hybrid model.

In Ng and Hussain (2004) a hybrid Inverse Model Control (IVMC) schema is proposed that linearizes the model equations and applies a nonparametric model, namely an ANN, in parallel in order to account for the error introduced through the linearization. The serial hybrid model functions as a soft sensor and further is applied for the tuning of the controller ANN in the adaptive schema. The controller was compared to other approaches such as the PID and the performance was observed to be better. This study was extended in Wei & al. (2007), where the controller was additionally implemented in an IMC fashion, which resulted in better performances than the IVMC. When compared to the same control structures (IMC and IVMC), where the underlying process model was a pure ANN, worse results were obtained in the ANN case.

A control structure that bases on a FFNN structure is proposed in Patnaik (2003) for the determination of three dilution rates. The online tuning of the FFNN controller weights is accomplished in such a way as to maximize the product concentration in the next interval. It seems that the hybrid serial process model finds application for this tuning. A comparison between this tuning schema and an optimizing control schema based on the same serial hybrid process model would in any case be interesting regarding (i) the maximization of the product concentration and (ii) the associated computational costs, since in case of the optimizing control schema only three dilution term values would have to be optimized instead of several weights. Almost the same methodology is also used in Patnaik (2010, 2004, 2008).

A serial hybrid model is used by Andrasik & al. (2004) for the tuning of a parallel hybrid PID – ANN controller. This neural-control structure is frequently applied and was e.g. studied in comparison to pure PI and pure ANN by Hisbullah & al. (2002) or similar

involving RBFN by Li & al. (2006).

In Georgieva and de Azevedo (2009); Georgieva and Feyo de Azevedo (2007) a Nonlinear MPC schema finds application, namely the MPC is based on an ANN. The serial hybrid model is in these cases applied to tune the ANN integrated in the MPC structure. In Georgieva and Feyo de Azevedo (2007) the ANN-MPC performance is compared to a Feedback Linearizing Control schema, that bases on Nonlinear AutoRegressive-Moving Average ANN models trained, and it is concluded that the MPC provides better set-point tracking than the FLC, but is computationally also much more involved.

Table 2.9 Hybrid Model tuned closed-loop controller

Reference	Application	Hybrid Structure	Nonparametric Model	Controller Design
Schubert & al. (1994a)	Control of Glucose in a fed-batch cultivation of *Saccharomyces cerevisiae* (exp.[1])	S	ANN	ANN -Internal Model Controller
Patnaik (2003)	Control of recombinant β- galactosidase production by *E. coli* in a fed-batch fermentation (vir.[2])	S	Elman NN	Optimal-Control FNN
Chen & al. (2004)	Control of an industrial reactive distillation column (exp.[1])	P	ANN	Internal Model Control
Andrasik & al. (2004)	Control of biomass concentration in a continuous *Saccharomyces cerevisiae* fermentation (vir.[2])	S	ANN	Parallel PID - ANN
Ng and Hussain (2004)	Temperature control of a nonlinear semi-batch polymerization reactor (vir.[2])	S	ANN	Inverse Hybrid Control
Patnaik (2004)	Control of the streptokinase activity in fed-batch fermentations of *Streptococcus equisimilis* (vir.[2])	S	Elman NNs	Optimal-Control FNN
Zhang & al. (2006)	Control for a fuel cell stack breathing control system (exp.[1])	SC[3]	ANN	Internal Model Control
Wei & al. (2007)	Temperature control of a nonlinear semi-batch polymerization reactor (vir.[2])	S	ANN	Inverse Hybrid Control / Internal Model Control
Georgieva and de Azevedo (2009); Georgieva and Feyo de Azevedo (2007)	Control for the supersaturation trajectory in a fed batch crystallization process (vir.[2])	S	ANN	ANN-MPC

Georgieva and Feyo de Azevedo (2007)	Control for the supersaturation trajectory in a fed batch crystallization process (vir.[2])	S	ANN	Feedback Linearising Control -NARM-ANN
Patnaik (2008)	Control of the poly-β-hydroxybutyrate (PHB) production through Ralstonia eutropha (vir.[2])	S	RBFNs	Optimal-Control FNN
Patnaik (2010)	Control of the poly-β-hydroxybutyrate (PHB) production through Ralstonia eutropha (vir.[2])	S	RBFNs	Optimal-Control FNN

exp[1]: experimental
vir[2]: virtual
SC[3]: Serial Structure C model

2.4.4 Optimization

The kind of optimization that is usually carried out with hybrid models addresses process control policies. In this respect, the optimization can be carried out off-line or on-line. On-line optimization (which essentially devolves to closed-loop (sub)optimal control) can be expected to achieve better performances than off-line optimization (implemented as open-loop control) (Oliveira, 1998), since e.g. process variations can be taken into account. However, on-line optimization may become infeasible due to the lack of reliable online measurements or high computational costs of the optimization. In industry open-loop control frequently finds application, e.g. in pharmaceutical industry "approved recipes" is tightly followed.

Model-based optimization is the "core business" for hybrid models, since the hybrid methodology bears major advantages over other models such as: (i) variables that influence the product value, e.g. Temperature or pH, can easily be incorporated into the hybrid model; (ii) hybrid models award with better extrapolation properties than pure nonparametric models (at least in the serial case); and (iii) the quality of the predictions is usually significantly increased. In general, it can be stated that the more knowledge has been used before to optimize the process, the more detailed, needs to be a model with which further improvements are sought to be obtained (Galvanauskas & al., 2004).

The possibility to apply hybrid models for process optimization was already disclosed by Psichogios and Ungar (1992), who showed that the optimal feed policy determined through hybrid model based optimal control, was very similar to the one obtained by the "true" model. Similarly, Schubert & al. (1994a) applied hybrid modeling for the optimization of a substrate feeding rate for maximization of the biomass yield on substrate. They also numerically studied different optimized sets and discussed the implications of

biomass estimation errors on the real process.

The maximization of the biomass growth rate, while limiting the formation of by-products is aimed at in Dors & al. (1995). Two feeding control strategies with different levels of complexity are investigated. The hybrid model predictions are adapted all 4-10 hours and the optimization procedure harvesting and optimal feeding are calculated. A systematic approach for the development of a model to be used in process optimization is discussed in Simutis & al. (1997).

Tholudur and Ramirez (1996) studied the optimization of the control policies, namely feeding rate control, for two fed-batch bioreactor case studies using serial hybrid models. The optimized control profiles obtained through the hybrid models were compared to those of the actual model and it was concluded that the hybrid model based approach offers a reliable method for optimizing the system performance. Preusting & al. (1996) developed a detailed hybrid model for an industrial penicillin process, a process that was well-studied before, wherefore the effort for further optimization is enhanced. Even so the optimization constitutes only a few percent in concentration, the gain is significant due to the process scale.

The derivation of the optimal set-points with respect to the nominal net profit for a steady-state distillation column through a serial hybrid model is studied by Safavi & al. (1999). Since, only the separation factor is modeled through a nonparametric model in the hybrid approach, the therethrough obtained optimized set-points are very similar to the ones derived for the mechanistic model.

The optimal control schema that was investigated in closed-loop by Costa & al. (1999, 1998) was applied for open-loop control of an experimental fermentation by Henriques & al. (1999). The experimental results, obtained for the optimal profile, outline the importance of using optimization techniques, especially in process in which high substrate concentrations inhibits fermentation.

A re-optimization strategy is applied by Zuo and Wu (2000), meaning that the optimal feeding policy is re-calculated whenever the new measured state values become available, i.e. the measured values are used as initial values in the hybrid model based optimization schema (this leading towards a closed-loop control schema).

In order to account for the effects of the dissolved oxygen concentration on the penicillin production, Ignova & al. (2002) applied a serial hybrid model. The derived hybrid model did however not meet the desired accuracy, wherefore an on-line adaptation schema for the biomass growth rate was developed. The optimization was at first carried out off-line and the profiles were re-optimized (similar to Zuo and Wu (2000)).

In Tian & al. (2001) two parallel hybrid model, comprising either a stacked ANNs or a single ANN, are applied to calculate the optimal temperature trajectories for a batch polymerization reactor. The predictions are assessed with the calculated 95% confidence interval and compared to the results of the experiments that were carried out using the optimized trajectories. It was observed that for both hybrid modeling approaches the process could be optimized, and that, when comparing the predictions to the experiment, the stacked ANNs are more reliable than the single ANN.

A batch-to-batch optimization schema is proposed in Crowley & al. (2001); Doyle & al. (2003), in order to meet target specifications in a polymerization process. Therein, a parallel hybrid model, consisting of a dynamic population balance based framework in parallel with a PLS model, is embedded into a MPC framework in order to derive the optimal process trajectory for a given target particle size distribution.

An iterative batch-to-batch approach for the risk-constrained optimization of bioprocesses was proposed in Teixeira & al. (2006) and studied on three simulation cases. This approach was then utilized for the maximization of the final amount of active product expressed by Baby Hamster Kidney (BHK) cultures in Teixeira & al. (2005a,b). Therein it was observed that the incorporation of more mechanistic information resulted into better model predictions, which in turn resulted into more reliable optimization results. The augmentation of the amount of mechanistic information through the integration of Elementary Modes, a methodology that allows to decompose metabolic models and opens the way to the quantification of the relative importance of certain pathways at a given process stage, was thereupon studied in Teixeira & al. (2007a) for further optimization.

For the modeling and optimization of a hydro cracking unit, Bhutani & al. (2006) compared the performances of first-principles, parallel, serial, parallel-serial and non-parametric models. They found that the pure nonparametric model, namely an ANN, had the best performance, followed closely by the parallel hybrid model. Therefore they

based the process optimization on the pure nonparametric model. However, the findings regarding the model performances are in contrast to those frequently reported. Since the number of available data is relatively large, the good performance of a pure nonparametric model can be expected. Striking is that the time required to derive a pure ANN process model is significantly lower than the time required for the derivation of those ANN embedded in the hybrid models. This is in contrary to what would be expected and points at a complication of the identification problem through the incorporation of the mechanistic knowledge (which in turn could indicate that the mechanistic structure is not appropriate). Additionally, in the serial hybrid case, the combination of measured model inputs and partial first-principle knowledge, might lead to a situation which abets error propagation, e.g. (von Stosch & al., 2011b).

Similar to a batch-to-batch optimization an incremental optimization is proposed in Kahrs and Marquardt (2007), i.e. (i) the optimization is constrained by the validity region of the hybrid process model; (ii) in those regions were the optimization is constrained, measurements are performed to enlarge the validity region; (iii) re-identification of the hybrid model is then performed, whereupon another optimization is carried out, and so on. The optimization of an industrial continuous polymerization plant, in form of set-point optimization applying hybrid modeling techniques is also mentioned in Fiedler and Schuppert (2008), i.e. "The resulting high correlation in the process parameters had caused the training of pure black-box models to fail. Therefore, optimization and model-based control of the melt index could only be performed using a structured hybrid model for the reactor."

The optimal operating policies for a two-step (aerobic followed by anaerobic step) fed-batch process of genetically modified *Saccharomyces cerevisiae* were studied by Eslamloueyan and Setoodeh (2011) for the production of ethanol. Therein a detailed intracellular model is constructed from experimental data using Flux Balance Analysis (FBA). The FBA model can be used for fed-batch simulation along with a reactor model based on material balances. However for the maximization of the final amount of ethanol, this modeling approach would result in a bi-level optimization problem since the FBA model is driven by the maximization of the specific growth rate at each time instance. In order to avoid the bi-level optimization the FBA is replaced by an ANN model, which results into a hybrid model. The optimal profiles, derived from the hybrid model, are then simulated along with suboptimal profiles utilizing the dynamic FBA model and it is observed that a small deviation from the optimal values can considerably decrease the

ethanol productivity.

Table 2.10 Hybrid model based optimization.

Reference	Application	Hybrid structure	Nonparametric model model	Objective
Psichogios and Ungar (1992)	Optimization of Process Operation Scheduling, i.e. maximization of product yield through determination of feeding policy in a fed-batch reactor (vir.[2])	S	ANN	Maximisation of the final amount of biomass
Schubert & al. (1994a)	Optimization of the feeding rate for yield maximisation with respect to the final amount of biomass in fed-batch process (exp.[1])	S	ANN/ Fuzzy	Yield maximisation
Tholudur and Ramirez (1996)	Optimization of the feeding control policies in a fed-batch protein production (vir.[2])	S	ANN	Maximisation of the profitability
Dors & al. (1995)	Optimization of the feeing control policy in an industrial recombinant protein production using mammalin cell cultures (exp.[1])	S	ANN	Biomass growth rate maximization, restricting by-products
Preusting & al. (1996)	Optimization of an industrial penicillin production (exp.[1])	S	ANN/Fuzzy Models	Maximisation of the final penicillin concentration
Henriques & al. (1999)	Optimization of the feeding policy in fed-batch alcoholic fermentations of *Zymomonas mobilis*	S	FLN	Maximisation of the final ethanol concentration
Safavi & al. (1999)	Maximisation of the nominal net profit for a stead-state distillation column process (exp.[1])	S	Wavelet-based NN	Profit Maximization
Zuo and Wu (2000)	Optimization of fed-batch *Bacillus thuringiensis* cultivations for thuringiensin production (exp)	S	ANN	Maximization of the final amount of product
Ignova & al. (2002)	Optimization of the feeding control policy with respect to dissolved oxygen concentration limitation (exp.[1])	S	ANN	Maximisation of the final amount of penicillin
Tian & al. (2001)	Optimal reactor temperature control policies for a batch polymerization reactor (exp.[1])	P	Stacked NN	Meeting target specifications
Crowley & al. (2001); Doyle & al. (2003)	Optimization of the process trajectories targeting specific size distributions using MPC in a semi-batch emulsion polymerization (vir.[2])	P	PLS	Meeting target specifications

Teixeira & al. (2005b)	Optimization of fed-batch Baby Hamster Kidney cultures expressing the human fusion glycoprotein IgG (exp.[1])	S	ANN	Maximisation of the final amount of active product
Teixeira & al. (2006)	A general risk constraint framework for the optimization of bioprocesses (vir.[2])	S	ANN	-
Bhutani & al. (2006)	Optimization of an industrial hydrocracking unit (exp.1)	-	ANN	Maximisation of the product
Teixeira & al. (2007a)	Optimization of recombinant Baby Hamster Kidney cultures producing a recombinant fusion glycoprotein (exp.[1])	S	ANN	Maximisation of the final amount of active product
Kahrs and Marquardt (2007)	Maximization of the amount of ethylene glycol in the product stream of a stead-state process through optimization of the educt flowrate, the flash temperature and the liquid hold- up of the reactor (vir.[2])	S	ANNs	Yield maximization
Fiedler and Schuppert (2008)	Optimization of the set-points in an industrial continuous polymerization plant. (exp.[1])	S	ANN	-
Eslamloueyan and Setoodeh (2011)	Optimization of the feeding policy in a fed-batch fermentation of *Saccharomyces cerevisiae* (exp.[1])	S	ANN	Maximisation of the final amount of ethanol

exp[1]: experimental
vir[2]: virtual

2.4.5 Model Reduction Approaches

Real processes have an overwhelming underlying complexity. In order to derive a general valid model, simplifications in form of assumptions are usually made. Simplifications might also be made in order to facilitate the analysis or to obtain a computational inexpensive solution (Qi & al., 1999; Safavi & al., 1999). In this respect hybrid modeling can be applied to correct for the unconsidered phenomena therefore maintaining a high degree of accuracy, while still being computationally efficient.

A model reduction approach based on residualization is proposed in Hahn & al. (2002). Therein the reduction is based upon the idea that the time derivatives of the less important states can be approximated by zero (resulting in algebraic equations) while the rest of the system remains unchanged. The behavior of the approximated algebraic state equations can then be learned by nonparametric techniques, which is attractive since the most important components of the system are contained in the remaining states and the neural network only corrects the system for the reduced states.

Chen & al. (2004) reduces a model, which was initially obtained by formulation of the balance equations, through the application of a singular perturbation technique that can be applied when a system has fast and slow reactions. The obtained reduced model is then decomposed into a dynamic linear and a non-linear static model which can conveniently be modeled by a hybrid parallel approach.

A different hybrid model based reduction approach is proposed by Romijn & al. (2008). Therein the computational efficient numerical solution of Partial Differential Equations (PDE) is addressed. At first, a transformation of the model equations to a reduced independent latent basis using proper orthogonal decomposition, i.e. Principal Component Analysis, is proposed. Then, in the reduced space, a hybrid model is applied such avoiding frequently encountered problems, such as (i) the loss of sparsity in the model formulation after reduction, (ii) computationally expensive function evaluations, (iii) the inability to cope with nonlinear uncertainty in the model and (iv) the extraction of relevant features for the specific purpose of the model is enabled.

The simplification of the dynamic FBA model through a hybrid model as proposed in Eslamloueyan and Setoodeh (2011) (see also the section on "Optimization"), is another example for the possibility of maintaining the model quality while reducing the computational effort.

2.4.6 Scale-up

A model developed on small scale, e.g. a pilot plant, cannot necessarily describe the same process on larger scale, since the dominating effects might differ with the scale. The situation might even become more precarious when nonparametric techniques are applied since (i) the therein captured interactions do no stem from mechanistic considerations; and (ii) the data used for their determination might contain scale specific information.

The development of a "scalable" hybrid model is carefully investigated by Braake & al. (1998). Beside some heuristics for the development of a scalable hybrid model, an emphasis is on the experiments (on both scales) that are necessary to develop a hybrid model with good range and frequency extrapolation properties. It is demonstrated on a simulation case study that the nonparametric model that is contained in the hybrid

model could be constructed solely on operational data, while the development of a pure nonparametric model requires data which are rich in information.

A different approach is investigated by Bollas & al. (2003), i.e. at first a mechanistic model is developed for a fluid catalytic cracking (FCC) pilot plant, which is then applied to describe the behavior of an industrial FCC plant. The differences noted between the data of the industrial plant and the mechanistic model predictions are, beside some variations in the feed and catalyst properties, assumed to be due to the scale-up factors. In order to account for these differences, several hybrid models are developed on the basis of the mechanistic model, and their predictions and extrapolation capabilities, are compared to those of the pilot plant mechanistic model and to a developed pure non-parametric model. The best performances, reaching the limitations of the experimental error, are observed for the hybrid models.

Similar approaches, to the one by Bollas & al. (2003), were used by Bellos & al. (2005); Simon & al. (2006). Three different industrial hydrotreaters are modeled with a serial hybrid model by Bellos & al. (2005). While the underlying mechanistic model was derived for the industrial scale, some of the kinetic parameters where determined at laboratory scale. It is shown that the application of the hybrid model allows to estimate the effect of the feed quality on the catalyst reactivity and the catalyst activity level, from few laboratory data and data from the operation units.

Simon & al. (2006) developed a mechanistic model for a small scale three-phase batch reactor, which then, for the large scale process, was complemented with neural network models to account for (i) the assumptions which do not hold true on this scale and (ii) other scale differences. The large scale hybrid model was carefully applied to derive process improvements and it was stated that for complex processes, for which detailed (thermodynamic) knowledge does not exist, hybrid modeling poses a valuable alternative.

2.5 Summary

Starting from the questions what hybrid modeling is and why it should be applied, the framework of hybrid modeling has been reviewed in light of all parts that contribute to the model performance. Various applications of hybrid modeling in several scientific areas have been discussed and also the utilization of hybrid approaches for monitoring,

control, optimization, scale-up and model reduction have been perused.

The following points have been striking:

i) Hybrid modeling found considerable attention during the last 20 years and the advantages with which such an approach awards are tremendous.

ii) Throughout the various applications, hybrid models are compared to pure nonparametric models or to pure phenomenological models. In almost all cases it was reported that the hybrid models performed better than the other two approaches. In control or optimization cases the better model performance translated usually into improved control and optimization results.

iii) The combination of different sources of knowledge into a hybrid modeling approach can, but must not necessarily, result into better system descriptions, than when compared to models that base on only one source of knowledge. This means that the application of hybrid approaches does not automatically result into improved models but that a differentiated perspective has to be kept and an analysis of the reason for eventual models shortcomings must be applied.

iv) The incorporation of additional phenomenological knowledge has been discussed, and it was concluded that the model performance can be enhanced when the incorporated structure has a low uncertainty. On the other hand it was stated that in cases of high structural uncertainty the application of parallel approaches is generally to prefer. However a rigorous comparison of the parallel structure to a serial structure **C** would be interesting.

v) The utilization of several nonparametric models in hybrid approaches has been reported. In this respect it can be stated that it is case dependent which nonparametric model is the best to be applied.

vi) Different identification procedures of the nonparametric models have been discussed and it might be said that the incremental approach together with the sensitivities approach can be expected to provide the best performance.

vii) Measures for extrapolative situations have been discussed and it was concluded that those methods mostly take range or dimensional extrapolation into account, while frequency extrapolation (the dynamics) is not considered. However in cases of control the dynamics are an important factor and should be taken into account. This

could for instance be accomplished by augmenting the inputs of the extrapolation measures by the derivatives.

viii) It was shown that hybrid models can be used for experimental design. The question whether it is better to systematically explore the process operational space by using e.g. a coverage approach or whether an iterative batch-to-batch optimization is used to plan the next experiment might depend on the case and the pursued objective.

2.6 Acknowledgment

Sincere thanks for financial support to the Fundação para a Ciência e a Tecnologia (reference scholarship no.: SFRH / BD / 36990 / 2007).

2.7 Nomenclature

Abbreviations	
AIC	Akaike Information Criterion
ANN	Artificial Neural Network
BHK	Baby Hamster Kidney
BIBO	Bounded Input Bounded Output
BIC	Bayesian Information Criterion
CER	Carbon dioxide Evolution Rate
CPR	Carbon dioxide Production Rate
CSTR	Continuous Stirred Tank Reactor
DD1	Divided Difference 1
EKF	Extended Kalman Filter
FBA	Flux Balance Analysis
FCC	Fluid Catalytic Cracking
FFNN	Feed Forward Neural Network
FLC	Feedback Linearizing Control
FLN	Functional Link Network
GMC	Generic Model Control
ICA	Independent Component Analysis
IVMC	Inverse Model Control
IMC	Internal Model Control
MARC	Model Adaptive Reference Control
MARS	Multivariate Adaptive Regression Splines
ME	Mixture of Experts
MLP	MultiLayer Perceptron
MPC	Model Predictive Control
NARX	Nonlinear AutoRegressive eXogenous
NLP	Nonlinear Programming
NN	Neural Network

NPLS	Nonlinear-PLS / Neural Network-PLS
ODE	Ordinary Differential Equation
PAT	Process Analytical Technology
PCA	Principal Component Analysis
PDE	Partial Differential Equation
PHB	Poly-β-HydroxyButyric acid
PID	Proportional Integral Differential
PLS	Partial Least Squares / Projection to Latent Structures
QPLS	Quadratic-PLS
RBFN	Radial Basis Function Network
RNN	Recurrent Neural Network
SERM	Semi-Empirical Regression Model
SVD	Singular Value Decomposition
SVM	Support Vector Machine
TSE	Taylor Series Extrapolation
TSK	Takagi, Sugeno Kang - type of Fuzzy model (Takagi and Sugeno, 1985)

Chapter 3

A novel identification method for hybrid (N)PLS dynamical systems with Application to bioprocesses

3.1 Abstract

This paper presents a method for the identification of Non-linear Partial Least Square (NPLS) models embedded in macroscopic material balance equations with application to bioprocess modeling. The proposed model belongs to the class of hybrid models and consists of a NPLS submodel, which mimics the cellular system, coupled to a set of material balance equations defining the reactor dynamics. The method presented is an analog to the NIPALS (Non-iterative Partial Least Square) algorithm where the PLS inner model is trained using the sensitivity method. This strategy avoids the estimation of the target fluxes from measurements of metabolite concentrations, which is rather unrealistic in the case of sparse and noisy off-line measurements.

The method is evaluated with a simulation case study on the fed-batch production of a recombinant protein, and an experimental case study of *Bordetella pertussis* batch cultivations. The results show that the proposed method leads to more consistent models with higher statistical confidence, better calibration properties and reinforced prediction power when compared to other dynamic (N)PLS structures.

3.2 Introduction

Partial Least Square (PLS) (also called Projection to Latent Structures) and Non-linear PLS (NPLS) have been shown to be powerful regression methods for static processes when the data is noisy and highly correlated. There are numerous applications of PLS and NPLS in biotechnology (Clementschitsch and Bayer, 2006; Henneke *&* al., 2005; Soons *&* al., 2008b). The difference between PLS and NPLS lies in the inner models which correlate the latent variables. In PLS the inner model is based on linear regression, whereas in most NPLS the inner model is non-linear, mimicked by quadratic functions (Wold *&* al., 1989), artificial neural networks (Qin and McAvoy, 1992), radial basis functions (Baffi *&* al., 2000) or support vector machines (Wang and Yu, 2004).
Many biotechnological processes are inherently dynamic and the PLS structure cannot be directly applied. Several attempts in the literature were made in order to extend the static PLS models for dynamical systems (Baffi *&* al., 2000; Lakshminarayanan *&* al., 1997; Ljung, 1991; Qin, 1993; Ricker, 1988). In most cases modeling of dynamic systems has been achieved through the augmentation of the inputs with lagged values of input and output data (Baffi *&* al., 2000; Ljung, 1991; Qin, 1993; Ricker, 1988). One-step-ahead prediction was developed inspired on the series–parallel identification scheme (Eykhoff, 1974) and recurrent training schemes (Qin and McAvoy, 1992; Werbos, 1988) or parallel identification schemes were used (Qin and McAvoy, 1996) for long term predictions. In the paper by Baffi *&* al. (2000) NPLS with different inner nonlinear models is successfully applied for modeling of nonlinear dynamical systems.
A bioprocess is ruled by a large number of complex physical, chemical and biological constraints, which are associated with both the cellular system and the bioreactor system. The above mentioned PLS models completely disregard such constraints since they are empirical data based techniques.
The dynamic nature of a bioprocess can be established by macroscopic material balances of the compounds with capacity to influence the physiological state of a cell. Thus an alternative way to add dynamics to a PLS model is to combine a static (N)PLS submodel with material balance equations in a hybrid dynamical structure. This type of strategy has been extensively reported in the literature for artificial neural networks (Lee *&* al., 2005; Oliveira, 2004; Peres *&* al., 2001; Preusting *&* al., 1996; Schubert *&* al., 1994a,b; Simutis *&* al., 1997) but very rarely for (N)PLS (Henneke *&* al., 2005; Lee *&* al., 2005). In this paper, a generic Non-linear dynamic PLS approach is developed within the hybrid modeling framework, i.e. by combining a (N)PLS submodel with material balance equations. There are two possible strategies to develop such a model. The probably simplest

way is to estimate the reaction rates from the material balance equations and from the concentrations' measurements and then to run a static NPLS model with the rates as target outputs and the state space vector as the inputs (Henneke & al., 2005; Lee & al., 2005). The difficulty of using this method arises when dealing with a limited number of observations and noisy measurements. The conjugation of these two factors is frequent in a real application, leading to very inaccurate estimation of the reaction rates. The second alternative, which is explored in this paper, follows the simultaneous parameter estimation strategy, using the well known sensitivity method (Oliveira, 2004; Peres & al., 2001; Preusting & al., 1996; Schubert & al., 1994a,b; Simutis & al., 1997).
The paper is organized as follows: in section 2 the proposed semi-parametric hybrid model, the parameter identification algorithm and model performance criteria are described; section 3 presents the application, results and discussion of the proposed method for two complementary case studies - one case with simulation data, specifically a model on protein synthesis, also known as the Park Ramirez model (Park and Ramirez, 1988), and another case comprising sparse, infrequent experimental data of *Bordetella pertussiss* cultures; then, in section 4, the conclusions are drawn.

3.3 The semi-parametric hybrid model

The semi-parametric hybrid structure here developed can also be referred to as an intrinsically dynamic NPLS model, which consists of two parts, namely material balances and a nonparametric/parametric submodel. The general hybrid model structure is described in the first subsection. The integration of the nonparametric model, a Nonlinear Partial Least Square model, is explained in the second subsection and a novel parameter identification algorithm is presented in the third subsection. The question of choosing the best model structure is finally addressed.

3.3.1 The general semi-parametric hybrid model structure

The general hybrid model structure proposed is depicted in Fig. 3.2. The concept is an evolution of the semi-parametric hybrid model proposed originally by Oliveira (2004). The structure is based on a bioreactor dynamic model, consisting of n material balances represented in vectorial terms as

$$\frac{dc}{dt} = f = r(L_x, w_A) - D \cdot c + u, \tag{3.1}$$

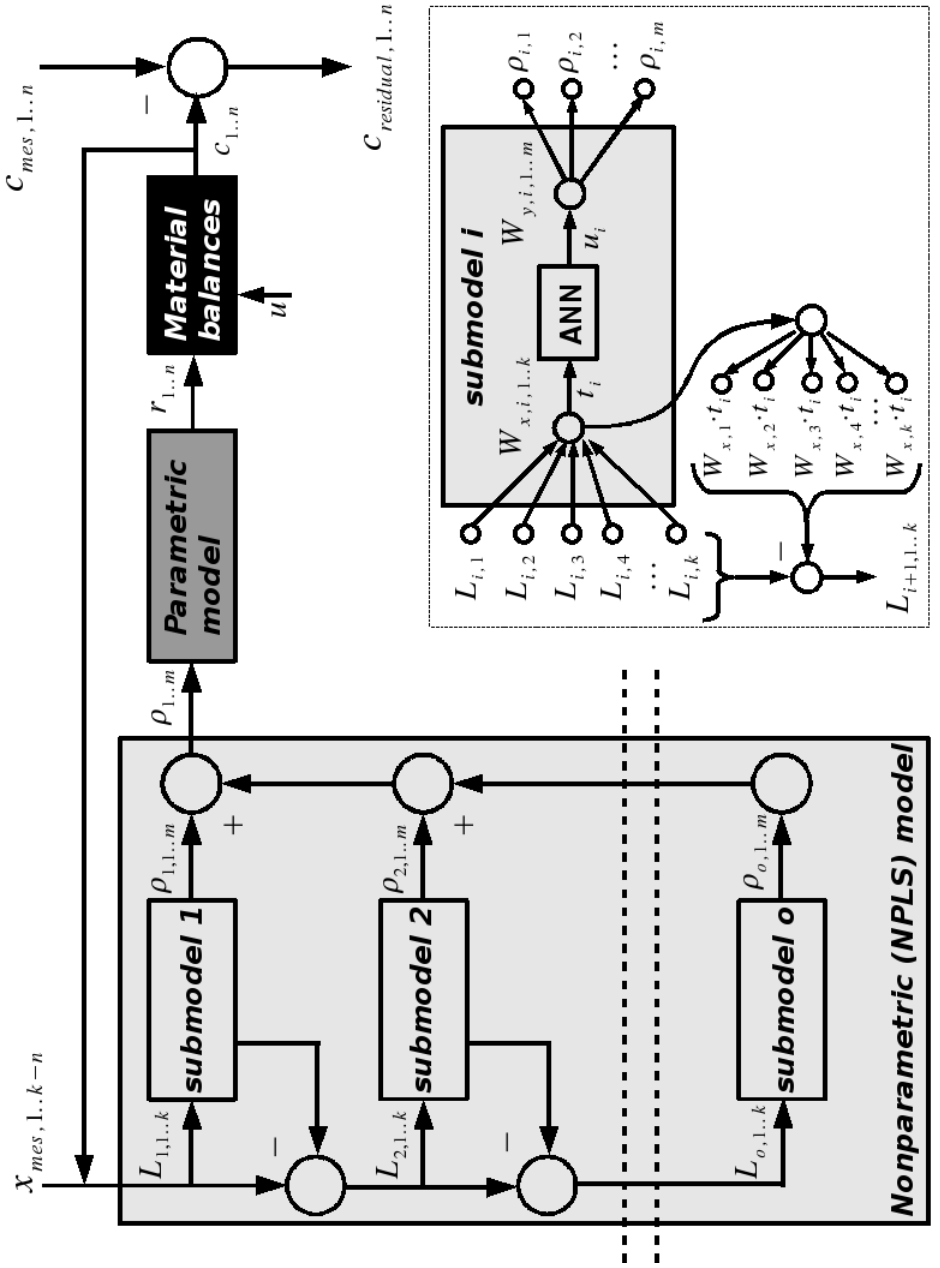

Figure 3.1 Diagram of the general semi-parametric hybrid model structure and of the incorporated submodels (mathematical symbols as in the text).

where c is the vector of concentrations, D is the dilution rate, u is a vector of volumetric control inputs and r is the vector of kinetic rates, i.e. the reaction term mimicking the cell system, which is modeled with a nonparametric/parametric submodel, using the vector of parameters w_A.

The nonparametric/parametric submodel reads as,

$$r(c, L_x, w_A) = K \cdot \langle \phi_j(c) \times \rho_j(L_x, w_A) \rangle_{j=1,..,m}, \tag{3.2}$$

with K being a $n \times m$ matrix of yield coefficients, ϕ being m kinetic functions and $\rho(L_x, w_A)$ being unknown kinetic functions which include w_A and the inputs L_x. These unknown kinetic functions are modeled with nonparametric techniques, such as Artificial Neural Networks (Oliveira, 2004; Peres & al., 2001; Preusting & al., 1996; Psichogios and Ungar, 1992; Schubert & al., 1994a,b; Simutis & al., 1997; Thompson and Kramer, 1994) or, as presented in the following, by a Nonlinear Partial Least Square alike model. The hybrid model can either be classified as a one-step or a multi-step ahead predictor. This is due to the unknown kinetic functions, $\rho(L_x, w_A)$, in Eq. (3.2), more precisely the inputs, L_x. When, the inputs cover only measured inputs at discrete time points, equivalent to a Finite Impulse Response (FIR) model, then the hybrid model functions behave as a one-step ahead predictor. When, alternatively, L_x comprises only the estimates of the model at discrete time points, equivalent to an AutoRegression (AR) model, then the hybrid model is a multi-step ahead predictor. It should be pointed out that the combination of measured and estimated data for L_x, equivalent to an AutoRegression eXogenous (ARX) model, results in a one-step ahead predictor.

3.3.2 The Nonparametric Model

The proposed nonparametric submodel, hereafter referred to as nonparametric model, is the key feature of the novel hybrid model. The structure is the one of a NPLS model, which is embedded into the hybrid framework, as reported to have been successfully applied in many areas, (Baffi & al., 2000; Henneke & al., 2005; Lee & al., 2005; Qin and McAvoy, 1996). In fact the structure exhibits all (N)PLS features, such as maximization of the covariance between input and output variables, minimization of redundant information of the inputs and identification of a minimal number of latent variable models. In the method here proposed the estimation of the unknown kinetic rates from noisy and sparse concentration measurement data is circumvented.

The Nonparametric Model Structure

The nonparametric model, for each component, j, of the vector of unknown kinetic functions, $\rho_j(L_x, w_A)$, is composed of o separate latent variable models (referred to as submodels $i = 1, .., o$, see Fig. 3.2), such that

$$\rho_j(L_x, w_A) = \sum_{i=1}^{o} \rho_{i,j}(L_x, w_A), j = 1..m, \tag{3.3}$$

where the index i denotes latent variable i. Note that in the following the term "latent variable model" is relaxed to latent variable.

Each submodel can further be divided into two parts, an outer and an inner model (Fig. 3.2): the outer model firstly linearly compresses the respective high dimensional input, by the use of input loadings, to one inner latent variable; the inner model then correlates, (non)linearly, the input latent variable, t_i, to the output latent variable, u_i; and subsequently the outer model decompresses the outer latent variable, u_i, through the use of the output loadings, into the respective outer vector $\rho_{i,j}(L_x, w_A)$ (for details see Baffi & al. (2000); Qin and McAvoy (1992)). In Baffi & al. (2000), ANN and RBF were used as inner models, which proved to be successful. In this approach an ANN model is applied. Mathematically this nonparametric model is expressed as follows:

$$\rho_{1..m,i}(L_x, w_A) = W_{y,i} \cdot (w_{2,i} \cdot g(w_{1,i} \cdot h(W_{x,i} \cdot L_{i,1..k}) + b_{1,i}) + b_{2,i}) \tag{3.4}$$

where $W_{x,i}$ and $W_{y,i}$ are the compression factors of the outer model, also called loadings, $w_{2,i}$ and $w_{1,i}$ are parameters of the ANN inner model, $b_{2,i}$ and $b_{1,i}$ are the biases of the ANN inner model, $h(\cdot)$ and $g(\cdot)$ are transfer functions, here linear and tangential, and $L_{i,1..k}$ comprises all inputs 1 to k to the model.

For $i = 1$ the vector of inputs comprises the estimated state variables or/and additional measured data $x_{mes,1..n}$, as illustrated in Fig. 3.2.

For $i > 1$, the vector of inputs, $L_{i,1..k}$ is the difference between the previous input vector and the information captured by the previous input latent variable, i.e. mathematically,

$$L_{i,1..k} = L_{i-1,1..k} - W_{x,i-1} \cdot L_{i-1,1..k} \cdot W_{x,i-1}^T \tag{3.5}$$

The arising advantage when compared to the so far used nonparametric model is that high numbers of redundant experimental data can be considered as inputs to the nonparametric model. In contrast to (N)PLS models the advantage for the identification of the involved parameters is that the kinetic rates do not need to be known explicitly, and that the hybrid structure is inherently dynamic. It should however be stressed that while the structure is a relevant prerequisite, the parameter identification method is essential for the success of the overall procedure.

Identification of the Nonparametric Model Parameters

The identification of the nonparametric model parameters proposed in this paper differs from the NIPALS identification procedure, but the general idea of this algorithm is kept. This idea is somehow identical to a twofold objective optimization, where both the covariance between inputs and outputs and the captured variance of the input are maximized. The maximizations are accomplished by the application of the sensitivity approach (Frank, 1978; Oliveira, 2004; Peres & al., 2001; Simutis & al., 1997), as it was shown to be preferable over the error-prone initial estimation of the kinetic rates with sequent parameter identification for ANNs (e.g. see Oliveira (2004)):

a) Maximization of the covariance between inputs and outputs

The maximization of the covariance between the inputs and outputs is analogous to the minimization of a weighted least-square error function of the state variables, c, which reads as

$$\min_{w_A} \left\{ E_1 = \frac{1}{P \cdot n} \sum^{P} \sum_{j=1}^{n} \frac{(c_{mes,j}(t) - c_j(t, w_A))^2}{c_{\sigma,j}} \right\}, \tag{3.6}$$

and where w_A are the model parameters, $c_{mes,1..n}$ is the vector of measured-known state variables, and $c_{\sigma,j}$ is the standard deviations of the experimentally measured concentration.

This objective function requires the determination of the number of latent variables prior to application, which is in contrast to (N)PLS models where consecutive latent variables are added till the desired level of abstraction is reached.

b) Maximization of the captured input variance

The first objective function E_1 serves only to maximize the covariance between the inputs and outputs, while the NIPALS algorithm also provides orthogonality of the latent variables that span the subspace (Baffi & al., 2000). This feature is important, because parameter identification problems arising from redundant input information are prevented and the dimension of the solution space is reduced. As for (N)PLS structures, redundant information is minimized on one hand by the compression of the input dimensions and on the other hand by subtraction of the information covered by the respective latent variable from the input information, Eq. (3.5), i.e. capturing the variance of the inputs. In analogy to this intrinsic feature of the NIPALS algorithm, the objective defined in the following seeks to account for such.

Capturing the variance of the inputs is analogous to the minimization of the residual of

the inputs, i.e. minimizing

$$L_{res,1..k} = L_{0,1..k} - \sum_{i}^{o} W_{x,i-1} \cdot L_{i-1,1..k} \cdot W_{x,i-1}. \tag{3.7}$$

A direct application of this equation for optimization is not feasible as uncorrelated inputs hinder the convergence of the optimization. In order to circumvent this problem the following procedure was developed:
i) The first step therein is to regress the matrix of inputs $L_{i,1..k}$ with the input scores, t_i, in order to obtain the input loadings in a PCA manner, i.e.:

$$W_{x,i,lin,un} = \frac{L_{i,1..k} \cdot t_i}{t_i^T \cdot t_i}. \tag{3.8}$$

The obtained solution is then normalized to unit length,

$$W_{x,i,lin} = \frac{W_{x,i,lin,un}}{\left\| W_{x,i,lin,un} \right\|}. \tag{3.9}$$

ii) The second step is the calculation of the residual between the input loading which is incorporated in the system of model equations, $W_{x,i}$, and the one obtained from eqs. (3.8) and (3.9), $W_{x,i,lin}$. The minimization of this residual is thought to be similar to the minimization of eq. (3.7). For the minimization a least square error function is adopted, i.e.:

$$\min_{w_A} \left\{ E_2 = \frac{1}{o \cdot k} \sum_{i=1}^{o} (W_{x,i,lin} - W_{x,i})^2 \right\}. \tag{3.10}$$

In such a way the inputs that are not correlated to other inputs or to the outputs are taken into account.
For the minimization of the error functions, E_1 and E_2, the sensitivity equations are employed. This means that the objective functions are differentiated with respect to the parameters w_A.:

Sensitivity equations for E_1
The sensitivity equations are obtained by differentiating eq. (3.6) with respect to w_A, which in general implies the derivation of eq. (3.1) with respect to w_A.

For the inner models, i.e. the ANN's, this reduces to the derivation of eq. (3.1) with respect to $w_{1,i}$ and $w_{2,i}$ (embodied in the following by w) resulting in:

$$\frac{d}{dt}\cdot\frac{dc}{dw}=\frac{\partial f}{\partial c}\cdot\frac{dc}{dw}+\frac{\partial f}{\partial w}, \tag{3.11}$$

where the first term on the right hand side of eq. (3.11) is due to the optional consideration of estimated state variables as inputs to the nonparametric model, as displayed in Fig.3.2.
For the outer models the sensitivity equations are similarly obtained by differentiating eq. (3.6) with respect to the input and output loadings $W_{x,i}$ and $W_{y,i}$ (which are in the following embodied by $W_{x/y,i}$). Not yet mentioned, but essential to report in this context is the normalization of the loadings $W_{x/y,i}$. This normalization carried out by analogy with the NIPALS algorithm facilitates mathematical operations since $W_{x/y,i}^{T}=W_{x/y,i}^{-1}$, where:

$$W_{x/y,i}=\frac{W_{x/y,i}^{up}}{\left\|W_{x/y,i}^{up}\right\|}, \tag{3.12}$$

with $W_{x/y,i}^{up}$ being the vector of parameters obtained from the optimization procedure. For the derivation of the sensitivity equation, eq. (3.12) is accounted for by the chain rule, i.e. the chain factor resulting from eq. (3.12) reads:

$$\frac{dW_{x/y,i}}{dW_{x/y,i}^{up}}=\begin{pmatrix}\frac{\left(\left\|W_{x/y,i}^{up}\right\|-W_{x/y,i,1,1}^{up}\cdot W_{x/y,i,1,1}^{up}\right)}{\left\|W_{x/y,i}^{up}\right\|^{2}} & \cdots & \frac{\left(-W_{x/y,i,1,1}^{up}\cdot W_{x/y,i,1,p}^{up}\right)}{\left\|W_{x/y,i}^{up}\right\|^{2}}\\ \vdots & \ddots & \vdots\\ \frac{\left(-W_{x/y,i,1,1}^{up}\cdot W_{x/y,i,q,1}^{up}\right)}{\left\|W_{x/y,i}^{up}\right\|^{2}} & \cdots & \frac{\left(\left\|W_{x/y,i}^{up}\right\|-W_{x/y,i,q,p}^{up}\cdot W_{x/y,i,q,p}^{up}\right)}{\left\|W_{x/y,i}^{up}\right\|^{2}}\end{pmatrix} \tag{3.13}$$

The sensitivity equations can then, similarly to Eq. (3.11), be obtained by differentiating eq. (3.1) with respect to $W_{x/y,i}^{up}$ which for the output loadings results in

$$\frac{d}{dt}\cdot\frac{dc}{dW_{y,i}^{up}}=\frac{\partial f}{\partial c}\cdot\frac{dc}{dW_{y,i}^{up}}+\frac{\partial f}{\partial W_{y,i}}\cdot\frac{dW_{y,i}}{dW_{y,i}^{up}}, \tag{3.14}$$

and for the input loadings gives:

$$\frac{d}{dt}\cdot\frac{dc}{dW_{x,i}^{up}}=\frac{\partial f}{\partial c}\cdot\frac{dc}{dW_{x,i}^{up}}+\frac{\partial f}{\partial W_{x,i}}\cdot\frac{dW_{x,i}}{dW_{x,i}^{up}}+\frac{\partial f}{\partial L_{i,1..k}}\cdot\frac{dL_{i,1..k}}{dW_{x,i}}\cdot\frac{dW_{x,i}}{dW_{x,i}^{up}}. \tag{3.15}$$

The sensitivity equations of the input loadings, eq. (3.15), bare the specialty that the subtraction of the covered information from the input information, namely eq. (3.5), must be taken into account, what is accomplished by the third term on the right hand side in eq. (3.15) (for details see the appendix). The derivation of $\partial f/\partial c$ and $\partial f/\partial w_A$ is straightforward and similar to the nonparametric structure given e.g. in (Oliveira, 2004) wherefore they are not described in detail.

Sensitivity equations for E_2

The sensitivity equations for the second objective function, E_2, are obtained by the differentiation of Eq. (3.10) with respect to w_A (i.e. the in-/output loadings and the ANN parameters). In the case of $W_{x,i}$, the derivative is obtained in a relatively straight forward way, resulting in Eq. (3.13) for the input loadings, while being zero for the output loadings and ANN parameters. In contrast, deriving the gradients of $W_{x,i,lin}$ with respect to w_A is operose. The chain rule can be applied using Eq. (3.13) to account for Eq. (3.9) and differentiating Eq. (3.12) with respect to w_A (i.e. the input loadings, output loadings and ANN parameters), as shown in the Appendix.

The least square problem functions, E_1 and E_2, are optimized simultaneously by using the "lsqnonlin" MATLAB function, which uses a subspace trust region method and is based on the interior-reflective Newton method (MATLAB Optimization toolbox), therefore gradient based, i.e. the sensitivity equations are required. However, when estimates of the state space variables are considered as inputs, then all parameters of the nonparametric model are also used to maximize the captured input variance, which is not desirable.

In order to account for this, first the simultaneous parameter identification is carried out and then, when the best parameter of the respective structure are identified as described below, only w and $W_{y,i}$ are further optimized subject only to the first objective function E_1.

In any case, the sensitivity equations are integrated along with the model equations, namely the system of equations comprised by eq. (3.1). In this study an Euler integration scheme is adapted. Initial values of sensitivity equations are zero, because the initial state variables are independent of the parameters.

Additional challenges for parameter identification

Parameter identification of nonparametric structures, especially when gradient based, exhibit a few additional challenges, namely restoring of the model generalization capabilities and avoiding local minima. The first challenge is usually overcome by (i) splitting

the data set into two partitions: the training set that contains about 2/3 of the data; and the validation set, which comprises about 1/3; and (ii) terminating the parameter optimization when a certain level of sophistication is reached (Bishop, 1995; Haykin, 1998; Oliveira, 2004)
The second challenge, namely local minima, arises from the shape of the solution space spanned by the objective functions and the parameters (Bishop, 1995; Haykin, 1998). The consistency of the minima obtained for various random initiations of the parameters (in this study at least four) is on one hand a measure of the quality of the solution obtained, and on the other hand a measure of the problem formulation quality. Notice that the larger the number of random initializations, the larger is the statistical confidence of the solution (Bishop, 1995; Haykin, 1998).

3.3.3 Model performance criteria

In order to identify the best hybrid model, both a measure of model performance must be defined, i.e. a model performance criteria and a suitable set of model structure variations must be considered. As outlined above, in section 3.3.2, the identification of the best hybrid model structure goes along with the identification of the number of latent variables. Besides this variation in the number of latent variables, the architecture of the ANN structure usually involves the variation of the number of layers and the number of nodes in these layers.
In this work, a number of decisions were taken, in order to downsize the degrees of freedom, namely: (i) a selection of three layers (input, hidden and output layer) was decided, which is usually sufficient if nonlinear continuous functions are sought to be modeled (Haykin, 1998); (ii) the number of nodes for the hidden layers of the ANN is fixed to be one; (iii) the number of nodes for the input and output layers for each submodel is fixed as one, as it results from the (N)PLS structure. What remains is then the evaluation of the variation of numbers of submodels, i.e. the variation of the number of latent variables, for each hybrid model set-up.
One criterion for model performance is the residual, also addressed as the goodness of fit of the model estimates regarding the data, which can be assessed through the Mean Square Error, MSE, where MSE is defined as:

$$MSE = \frac{1}{P \cdot n} \cdot \sum_{P} \sum_{j=1}^{n} (c_{mes,j}(t) - c_j(t, w_A))^2 \tag{3.16}$$

Evaluation and comparison of model concepts and structures cannot however be only

built up on the estimation error obtained for the training, validation or test set, in form of the residual (Bishop, 1995; Haykin, 1998). It is known that as model complexity grows, i.e. the number of parameters grows, the quality of fit may apparently improve, but often at the expense of robustness and generalization capabilities, (Bishop, 1995; Haykin, 1998). With respect to these issues the Akaike Information Criteria, AIC, is a suitable and widely applied criteria, but according to (Burnham and Anderson, 2004; Leonard and Hsu, 1999; Peres & al., 2008), the Bayesian Information Criteria, BIC, is more appropriate for the applications which this approach addresses. Therefore the BIC is applied for the model comparison and selection in this study.
The Bayesian Information Criteria, (BIC), is defined as:

$$BIC = \left(-\frac{n \cdot P}{2} \cdot \ln \left(\sum_{P} \sum_{j=1}^{n} [c_{mes,j}(t) - c_j(t, w_A)]^2 \right) \right) - \left(\frac{n_w}{2} \cdot \ln \left(\frac{n \cdot P}{2\Pi} \right) \right) \quad (3.17)$$

where the term in the first bracket is the logarithmic maximum likelihood and n_w is the total number of parameters/weights. In terms of the BIC, the model to be selected is the one that exhibits the larger BIC value for the validation set (Burnham and Anderson, 2004; Leonard and Hsu, 1999; Peres & al., 2008).

3.4 Application, Results & Discussion

In this section the application, results and discussion of the proposed hybrid model and of reference dynamic (N)PLS models are reported for two complementary case studies. The first study focuses on the process dynamics and the identification of the number of latent variables. The second study concentrates on the model identification from typical noisy, sparse and infrequent experimental data, a case which hinders the direct application of the reference dynamic (N)PLS models. The results obtained for the hybrid model are rigorously analyzed and benchmarked against reference dynamic (N)PLS models.

3.4.1 Case Studies

A protein synthesis, the Park Ramirez Model

The protein synthesis process

The method proposed in section 3.3 is evaluated in this subsection with simulation data of protein synthesis in a fed-batch reactor, also known as the Park-Ramirez model, as

originally proposed by Park and Ramirez (Park and Ramirez, 1988). This model found wide application, for similar model structures to the one proposed here, e.g. in (Kulkarni & al., 2004) for the evaluation of their Principal Component Analysis – General Regression Neural Network model, or in (Oliveira, 2004) for the evaluation of the traditional semi-parametric hybrid model. The reactor model comprises material balances of the secreted and total protein/product, the biomass, the substrate and the volume. The model dynamics, i.e. the offset between formation of secreted and total protein on the one side and biomass growth and substrate uptake on the other, poses some challenge, which is one reason for the application of this model in this study. Also, this model finds application because the number of latent variables therein is expected to be larger than one, but smaller than four as analytically at least two kinetic rates (substrate uptake and biomass growth) are linearly dependent and such accounts for the model capability of identifying the underlying latent variables.

In this paper the model equations, the feeding profile, the variation of the initial concentrations and the corruption of the generated simulation data with a Gaussian error of 5%, were applied for simulation case data generation, as described in (Kulkarni & al., 2004). Normal and abnormal (in the sense of initial data outside the usual range, as defined by (Kulkarni & al., 2004)) fed-batch data were generated, through variations in the initial values of concentrations, which significantly influence the concentrations dynamics. Three sets were defined, comprising 12 normal plus 4 abnormal fed-batches for the training data set, 2 plus 2 for the validation set and 2 plus 2 for the test set, respectively. After generation, the sets were corrupted with 5% Gaussian noise, except for the feeding and volume data which were corrupted with 1.5% Gaussian noise.

The reference models

As reference for comparison with the proposed dynamic hybrid models, (N)PLS models which account for the dynamics by the augmentation of the inputs in the sense of Finite Impulse Response (FIR) or AutoRegression (AR) are used (as in most cases: (Baffi & al., 2000; Ljung, 1991; Qin, 1993; Ricker, 1988)). The model structure identification of such dynamic (N)PLS models comprises the identification of inputs to the models, namely the number, type and time-points, in the sense of FIR or AR, and the identification of the number of latent variables, i.e. the structure is adapted in order to obtain the smallest mean square prediction error in the validation set. In the following (see Table 3.1) they will be referred to as FIR-(N)PLS and AR-(N)PLS, respectively. The NPLS models contain the same ANN inner model functions as the hybrid models, which are described in more detail in section 3.3.

The hybrid models

In this study four different hybrid models are investigated.

In the hybrid structures (A) and (B) no mechanistic knowledge of the process is considered. The model equations for concentrations of secreted protein, total protein, biomass and substrate, read,

$$\frac{d}{dt}\begin{bmatrix} P_{sec} \\ P_{tot} \\ X \\ S \end{bmatrix} = \begin{bmatrix} r_{P_{sec}}(L_x) \\ r_{P_{tot}}(L_x) \\ \mu(L_x) \\ r_S(L_x) \end{bmatrix} - D \cdot \begin{bmatrix} P_{sec} \\ P_{tot} \\ X \\ (S - S_{(F)}) \end{bmatrix}, \tag{3.18}$$

respectively. This corresponds to the bioreactor dynamic model structure represented by eqs. (3.1) and (3.2), where the matrices K and ϕ are identity matrices.

The hybrid structures C) and D) consider some basic knowledge about the process, and the system of equations is generally represented by,

$$\frac{d}{dt}\begin{bmatrix} P_{sec} \\ P_{tot} \\ X \\ S \end{bmatrix} = \begin{bmatrix} (P_{tot} - P_{sec}) & 0 & 0 & 0 \\ 0 & X & 0 & 0 \\ 0 & 0 & X & 0 \\ 0 & 0 & 0 & X \end{bmatrix} \cdot \begin{bmatrix} r_{P_{sec}}(L_x) \\ r_{P_{tot}}(L_x) \\ \mu(L_x) \\ r_S(L_x) \end{bmatrix} - D \cdot \begin{bmatrix} P_{sec} \\ P_{tot} \\ X \\ (S - S_{(F)}) \end{bmatrix}. \tag{3.19}$$

While structures A) and C) are one-step-ahead predictor models, structures B) and D) are multi-step-ahead predictor models, i.e. while the input vector L_x (see eq. 3.2) contains the measured values of substrate, biomass, total and secreted product concentrations for A) and C), it contains only estimated values of substrate, biomass, secreted and total product concentration for B) and D).

The only remaining undetermined structural feature is thus the number of latent variables. This was identified, in all cases, by an heuristic search of the number of latent variables that produces the best performance in terms of BIC (eq. 3.17) for the validation data.

These hybrid structures can directly be compared to their dynamic (N)PLS counterpart in terms of one-step or multi-step ahead prediction. By doing so, it is possible to evaluate the different structures regarding their statistical confidence, their calibration properties and the model estimation errors.

In advance it should be pointed out that the one-step ahead predictor hybrid models,

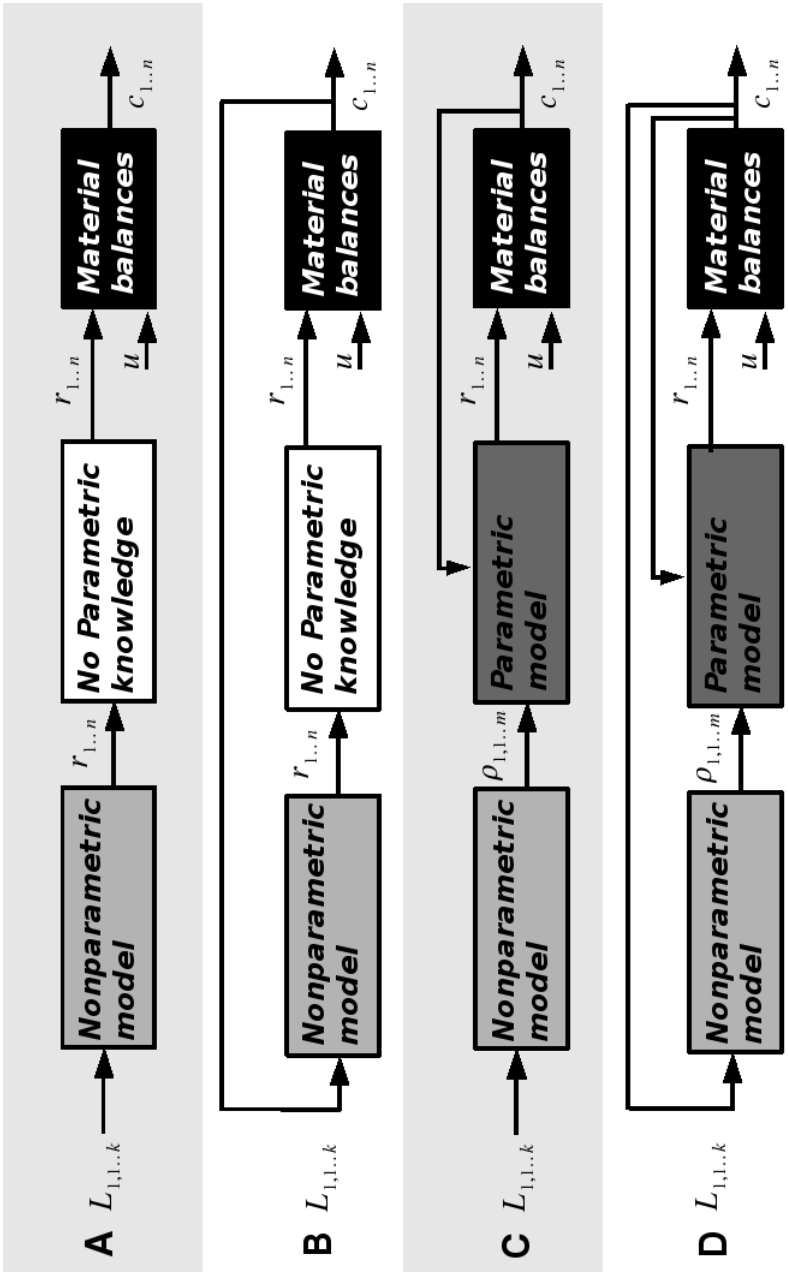

Figure 3.2 Diagrams of the four hybrid structures of the Park-Ramirez Case Study: A) one-step ahead predictor hybrid model structure, with no mechanistic knowledge incorporated. B) multi-step ahead predictor hybrid model, with no mechanistic knowledge incorporated; C) one-step ahead predictor hybrid model structure, with mechanistic knowledge incorporated; D) multi-step ahead predictor hybrid model with mechanistic knowledge incorporated. (mathematical symbols as in the text).

A) and C), are expected to perform worse than the multi-step ahead predictor hybrid models B) and D), because: (i) the uncertainty, i.e. noise, in the input data is directly passed to the estimates in the case of one-step ahead predictors; (ii) uncertainty in an estimate is passed to all future estimates due to the numerical integration; and (iii) the one-step ahead predictor hybrid models, A) and C), in contrast to the multi-step ahead predictor hybrid models, B) and D), have no feedback of the actual state estimates to the nonparametric model, wherefore the nonparametric model can neither identify nor correct for errors in the actual state estimates.

An experimental case study: Bordetella pertussis

The Bordetella pertussis process

The experimental study published by (Soons & al., 2008a,b) is the basis for the second case study of the present paper. The challenge here is to examine a dynamic process where only typically infrequent, sparse experimental data is available. Soons & al. (2008b) reported runs in batch mode and variations to the process conditions, such as in pH, Temperature and dissolved oxygen. Their measurements of the concentrations of lactate, glutamate and biomass over time for eight batches were reported as PAB0003, PAB0004, PAB0005, PAB0006-1, PAB0006-2, PAB0007, PAB0009-1, and PAB0009-2.

In order to identify and avoid bias from possible measurement errors, two sets of studies were carried out in the present paper:

In Set 1 - batches PAB0003, PAB0005, PAB0006-1, PAB0006-2 and PAB0009-2 were employed for training and batches PAB0007 and PAB0009-1 used for validation.

In Set 2 - batches PAB0003, PAB0005, PAB0006-1, PAB0006-2 and PAB0009-1 were employed for training, and PAB0007 and PAB0009-2 for validation.

It should be pointed out that batch PAB0007 is an "abnormal" batch, where a dissolved oxygen limitation and a lowered pH from 0-9 hours occurred, whereas batches PAB009-1 or PAB009-2 can be taken as "normal" (Soons & al., 2008b). By doing so, it is guaranteed that in both sets a "normal" and an "abnormal" batch were used in the validation step. The measured biomass concentration of batch PAB0004, was used as final test data, in order to provide a final assessment of the generalization capabilities of the models.

The reference models

The reference models in this case study are, as before described for the other case study, (N)PLS models which account for the dynamics by augmentation of the inputs.

Beside the augmentation of the inputs in the sense of FIR, the inputs here are also augmented using the AutoRegressive eXogenous (ARX) approach. As before the model structure identification of such dynamic (N)PLS models comprises the identification of the number of latent variables and of inputs to the models, namely the number, type and time-points, in the sense of FIR or ARX, i.e. the structure is adapted in order to obtain the smallest mean square prediction error in the validation set. In both schema a time lag of 1 hour and a maximum number of 4 equidistant lags for each input were investigated. In the context of sparse and infrequent measurements the application of these specifications requires that the measurements are pretreated, i.e. in this study the (N)PLS model inputs at the specific time instances were obtained through a cubic smoothing spline (MATLAB routine: csaps). However, this mandatory procedure must be accounted for when analysing the results, since on one hand the smoothing of the data can be expected to enhance the model performance while on the other hand the data interpolation might diminish the same. The NPLS models contain the same ANN inner model functions as the hybrid models, which are described in more detail in section 3.3.

The hybrid models

The hybrid model in this case contains mechanistic knowledge about the process, which was reported in (Soons & al., 2008a). This results in improved convergence of the parameter identification and into less random initiations for the parameters in order to obtain consistent results. The system of model equations reads,

$$\frac{d}{dt}\begin{bmatrix} Lac \\ Glu \\ X \end{bmatrix} = \begin{bmatrix} Lac \cdot X & 0 & 0 \\ 0 & Glu \cdot X & 0 \\ 0 & 0 & X \end{bmatrix} \cdot \begin{bmatrix} r_{Lac} \\ r_{Glu} \\ \mu \end{bmatrix} - D \cdot \begin{bmatrix} Lac \\ Glu \\ X \end{bmatrix}, \tag{3.20}$$

where Lac, Glu and X are the concentrations of Lactate, glutamate and biomass, respectively and r_{Lac}, r_{Glu} and μ are the respective unknown kinetic functions which are obtained by the nonparametric model.

The input vector L_x of the nonparametric model in this study contains the estimates of all concentrations, pH, temperature and the percentage of dissolved oxygen, as reported to be responsible for the process variations (Soons & al., 2008b). A gain, as reported in the previous case study, the only remaining undetermined structural feature is the number of latent variables. This was as well identified, in all cases, by an heuristic search of the number of latent variables that produces the best performance in terms of BIC (eq. 3.17).

3.4.2 Issues of hybrid model development and implementation

The proposed semi-parametric hybrid model might be understood as a dynamic NPLS model wherein the dynamics are modeled by material balances. In the following the dynamics and the performance of the hybrid model are rigorously analyzed.

Performance Criteria

Statistical Confidence – the BIC

In comparison to reference dynamic (N)PLS approaches, such as AR(X)- or FIR- (N)PLS models, it was observed that the hybrid methodologies possess way fewer model parameters, i.e. latent variables. This is a qualitative observation which is reflected in both presented simulation cases by the significantly larger BIC values obtained for the hybrid models when compared to the values obtained for the comparative dynamic (N)PLS models (see Tables 3.1 or 3.2). It should be pointed out that the dynamic (N)PLS approaches, namely the AR(X)- and FIR- (N)PLS models, are disadvantaged in terms of BIC, due to: (i) the higher number of latent variables; and (ii) the dynamic structure itself which increases the number of parameters on the input side.

Table 3.1 Values of model performance criteria over model types and structural parameters - simulation case study on the protein synthesis, also called the Park Ramirez Simulation Case.

	BIC			MSE		
Model type	Train	Valid	Test	Train	Valid	Test
FIR-PLS (lv[a] = 4; nt[b] = 1)	-1930	-477	-348	0.0334	0.0812	0.0295
AR-PLS (lv[a] = 4; nt[b] = 1)	-2074	-519	-433	0.0408	0.0893	0.0456
FIR-NPLS (lv[a] = 4; nt[b] = 1)	-1945	-458	-383	0.0343	0.0699	0.0388
AR-NPLS (lv[a] = 4; nt[b] = 1)	-1994	-486	-433	0.0349	0.0691	0.0456
Hybrid structure A	-1248	-219	-337	0.0134	0.0228	0.0610
Hybrid structure B	-1189	-140	-222	0.0119	0.0118	0.0235
Hybrid structure C	-1962	-379	-402	0.0595	0.0869	0.1055
Hybrid structure D	-1083	-135	-93	0.0095	0.0114	0.0080

[a] lv: number of latent variables

[b] nt: number of time series elements

From the BIC definition, Eq. (3.17), the model to prefer is the one with the larger BIC value, i.e. the one which for equal residual and number of data, has fewer parameters, in this way penalizing complex models (Bishop, 1995).

Table 3.2 Values of model performance criteria over model types and structural parameters - experimental case study on *Bordetella pertussis* cultivation data.

Model		Data	BIC			MSE		
Type	Structure	Set[c]	Train	Valid	Test	Train	Valid	Test
FIR-PLS	[lv[a] = 5, nt[b] = 3]	1	-486	-168	-	0.1486	0.0736	-
ARX-PLS	[lv[a] = 6, nt[b] = 3]	1	-549	-209	-55	0.1567	0.0703	0.3996
FIR-NPLS	[lv[a] = 5, nt[b] = 3]	1	-473	-173	-	0.1318	0.0831	-
ARX-NPLS	[lv[a] = 6, nt[b] = 3]	1	-516	-224	-41	0.1142	0.1028	0.0488
Hyb – NPLS	[lv[a]= 1]	1	-430	-103	9	0.2884	0.1397	0.0160
Hyb – NPLS	[lv[a] = 2]	1	-330	-88	10	0.1020	0.0836	0.0106
Hyb – NPLS	[lv[a] = 3]	1	-317	-96	9	0.0842	0.0910	0. 0094
FIR-PLS	[lv[a]= 5, nt[b] = 3]	2	-478	-163	-	0.1483	0.0568	-
ARX-PLS	[lv[a] = 6, nt[b] = 3]	2	-542	-204	-52	0.1573	0.0540	0.2361
FIR-NPLS	[lv[a] = 2, nt[b] = 3]	2	-402	-115	-	0.1509	0.0737	-
ARX-NPLS	[lv[a] = 2, nt[b] = 4]	2	-410	-134	-14	0.1328	0.0782	0.0511
Hyb – NPLS	[lv[a] = 1]	2	-416	-79	-3	0.2694	0.0687	0.0878
Hyb – NPLS	[lv[a] = 2]	2	-357	-79	13	0.1420	0.0609	0.0075
Hyb – NPLS	[lv[a] = 3]	2	-393	-82	16	0.1882	0.0582	0.0037

[a] lv: number of latent variables

[b] nt: number of time series elements

[c] Set: Set 1 or 2 refer to the grouping of batches the respective model has been trained and validated on.

In general, models with higher numbers of parameters are thought to be less robust and to exhibit worse generalization capabilities than models that offer similar residua, but with smaller number of parameters. Thus the BIC is a measure of the statistical confidence of the model performance and therefore the proposed hybrid models exhibit a higher statistical confidence than the comparative dynamic (N)PLS approaches.

Performance under the MSE criterion

The statistical confidence observed for the hybrid models is in agreement with the performance of such models observed and evaluated in terms of the MSE criteria, as shown in Tables 3.1 and 3.2.

It was observed that the proposed hybrid method most times exhibits significantly better and only rarely worse performance than the other evaluated dynamic (N)PLS models.

Cases in which the hybrid method exhibited a worse performance in terms of MSE values than the comparative methods were graphically analyzed. As example, for the

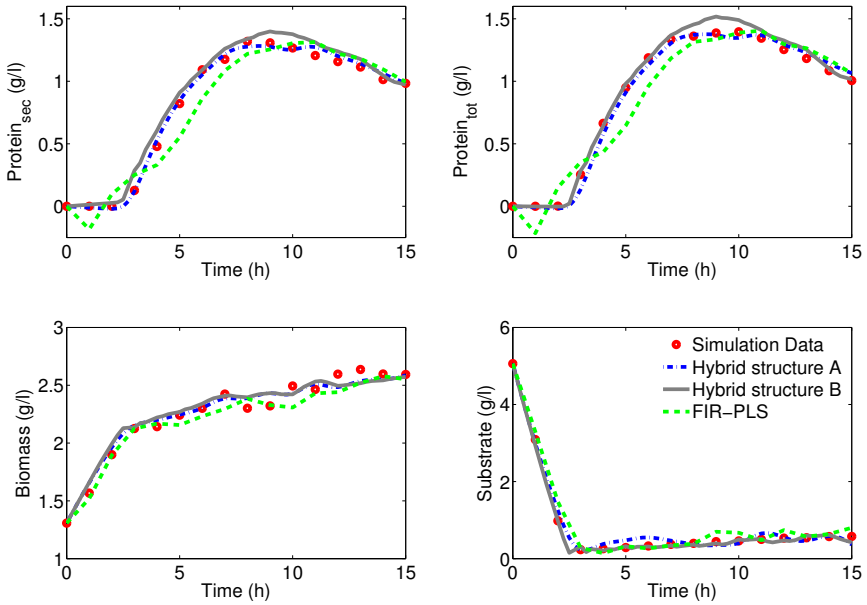

Figure 3.3 Park-Ramirez Case Study - plots of secreted protein, total protein, substrate and biomass concentrations, over time: predictions of hybrid structures A (dashed dotted blue line) and B (grey line), and of the best reference FIR-PLS model (dashed green line, Table 1) vs. the process simulation data (red dots), for a 'normal' validation run.

Park-Ramirez case study, Figs. 3.3 and 3.4, it was observed that the highest deviations are to be found in the substrate concentrations for hybrid structures A) and C).

When seeking for an explanation it must be kept in mind that: (i) both hybrid models, A) and C), are one-step ahead predictor models, in the sense of FIR; and (ii) the estimations by these hybrid models are sensitive to noise in the feeding rate data, as outlined in subsection 3.4.1.

In the case of the feeding rates, the hybrid model cannot account for the uncertainty therein, because neither the feeding rate data are inputs to the nonparametric model nor the state estimates are feedback to the nonparametric model. That those uncertainties can partially be accounted for when the state estimates are inputs to the nonparametric model, is demonstrated by the excellent performance of hybrid structures B) and D). However for the best performance by hybrid model D), those uncertainties are still observable in form of the slightly bumpy estimations of biomass and substrate and in form

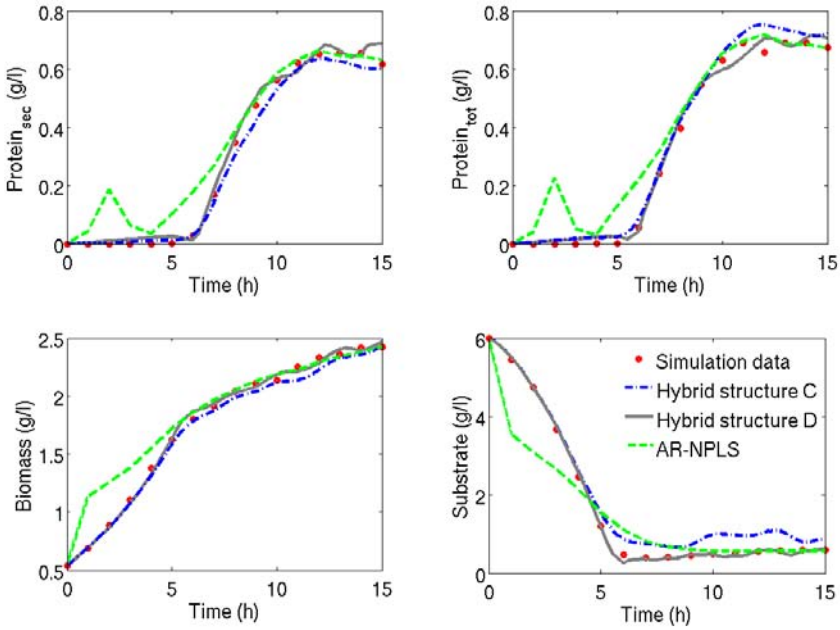

Figure 3.4 Park-Ramirez Case Study - plots of secreted protein, total protein, substrate and biomass concentrations, over time: predictions of hybrid structures C (dashed dotted blue line) and D (grey line), and of the best reference NPLS-AR model (dashed green line, Table 1) vs. the process simulation data (red dots), for an 'abnormal' test run.

of the bumpy estimations of secreted and total protein towards the end of the abnormal fed-batch, shown in Fig. 3.4.

For the experimental case study it is observed in Table 3.2, that the performance, in terms of MSE, for the hybrid models on data Sets 1 and 2, is non-coherent: using as example the hybrid model with 3 latent variables, the MSE values obtained for the training data of data Set 1, are half as big when compared to the MSE values on the training data of data Set 2. In order to identify the reason for this contradiction an additional analysis, reported below, was carried out on the influences which errors in the initial concentration values have on the whole dynamics. However, observations for the MSE values of the test data for both sets, wherein the performance of the hybrid models are found to be significantly better than the ones of dynamic (N)PLS models, show the excellent generalization capabilities of the hybrid models.

Model Structures and Error Propagation Issues

In the case studies presented several sources of errors can be identified, namely (i) noises in input measurements; (ii) errors inherent to model structures; (iii) errors in estimated inputs and/or in estimated parameters; and (iv) errors associated to 'defective' initial values. These are representative of essentially all experimental applications.
Leaving aside the trivial, though in practice often difficult, issues of error propagation due to the nature of numerical integration methods employed, it is relevant to analyze the issues associated to the nature of model structures chosen.

Error Propagation due to State feedback to the Nonparametric model
One way of propagation of the error in the estimates occurs in all those model structures in which the state estimate is a nonparametric model input, e.g. hybrid structures B) and D) in the Park Ramirez case study, or the ARX-(N)PLS model in the experimental case study. However, the form of the time evolution of the sensitivity equations, eqs. (3.11), (3.14) and (3.15), in hybrid structures where state feedback is embedded, tend to have a damping effect on such error propagation. This can be excellently seen by the enhanced performances, in terms of MSE, through hybrid structures B) and D) in contrast to the ones for A) and C) which are all-together shown in Table 3.1.

Error Propagation due to State feedback to the Parametric model
Another way in which the error is propagated arises when mechanistic knowledge, namely knowledge about the kinetics, in form of the model estimates, is incorporated, such as in hybrid structures C) and D).
The incorporation of the estimates is somewhat identical to the case when the inputs to the nonparametric model comprise the estimates, with the significant difference that an error in the estimation (e.g. from noisy feeding rates as in hybrid structure C)), depending on the arithmetic operator, (e.g. a multiplication sign for hybrid structures C)) might amplify the error (e.g. rather large deviations in the substrate concentrations, Fig. 3.4, and a rather large MSE value, Table 3.1, are obtained for hybrid structure C)). The excellent performance observed with hybrid structure D), whose mechanistic knowledge is equivalent to C), is explained by the damping qualities of the nonparametric model.

Errors in the Initial values, a special case
A relevant issue in all model analysis is that of the 'condition' of the model structures

proposed.
As addressed above, the results of the experimental case study in Table 3.2, of applying the hybrid models to data Sets 1 and 2, show some inconsistency (see section 3.4.1). In order to find the reason for such, an additional analysis of the experimental data was carried out, namely a PCA. It was observed that the correlations for the initial values of concentrations in some of the batches vary significantly from the correlations obtained for the whole data set, which is in line with the eye observations made. When (i) correcting the initial values in the validation and test batches of data Set 1 by using PCA and (ii) applying on these sets the hybrid model with two latent variables and the ARX-PLS, which both were prior trained on Set 1, then (iii) the results shown in Table 3.3 are obtained. Therein it can be seen that the performance in terms of both BIC and MSE values obtained for the hybrid model is significantly better than the performance of the ARX-PLS model. The performance of the hybrid model in Table 3.3 compared to the very same hybrid model in Table 3.2 led to more than 50% reduction in the MSE values of the validation and test batches.
This outlines the sensitivity of the proposed hybrid model to a high noise to signal ratio, which is in line to the observations made for the hybrid structures A) and C), i.e. the noise in the measurements enters directly the nonparametric model, leading to deviations of the estimations regarding the simulation data. In particular, defective initial values due to noise effects constitutes a special case, as those values are the base for the integration and as such are a significant source of misprediction.

3.4.3 Challenges of the Park Ramirez Case Study

The challenges offered by the Park Ramirez simulation case are on the model dynamics and on the identification of the number of latent variables.

The first challenge: The model dynamics
The dynamic delay between formation of secreted and total protein on one hand and biomass growth and substrate uptake on the other hand, varies depending on the initial values of concentrations. This dynamic feature was very well modeled by all applied hybrid structures, apart from the slightly "bumpy" shape of the trajectories, which were ascribed to the error propagation in the discussion above.
Small deviations between estimates and reference values of concentrations can be observed, especially for the one-step ahead predictor hybrid structures A) and C), but the general dynamic state behavior is well predicted, as can e.g. be seen in Fig. 3.4 for the

substrate concentration.

Even the dynamics of the abnormal fed-batches are very well predicted by the hybrid structures, in contrast to the observations made for the reference dynamic (N)PLS approaches, as illustrated in Fig. 3.4. For these special batches it can be concluded that the proposed hybrid models, in comparison to the other dynamic (N)PLS models, even when applied to "regions" where they have been poorly trained on, offer smaller deviations from the simulation data, which confirms the higher statistical confidence of the estimates from such models.

The preceding also means that even if the training set does not contain all possible variations, which can occur during the process, still the performances of the proposed hybrid model for different operating conditions, can be expected to be superior to the one of the comparative dynamic (N)PLS models. These conclusions are according with the findings reported by (Oliveira, 2004; Thompson and Kramer, 1994).

The second challenge: The number of latent variables

The second challenge of the Park Ramirez simulation case is the identification of the number of latent variables for both, the hybrid and the reference (N)PLS models.

Analytically, it is clear that at least two kinetic rates, namely the substrate and biomass rates, are linearly correlated. However, from observations made on the simulation data it might be concluded that also the rates of secreted and total protein are linearly correlated, which in total then sums up to two independent latent variables. This number is observed for the identified hybrid model structures, where it was found that the best hybrid structures always comprised only two latent variables.

In contrast, identification of the best model structure for reference dynamic (N)PLS models always revealed four latent variables. Partially this is due to the fact that linear correlations of the kinetic rates do not necessarily mean that the respective concentrations are linearly correlated in the same way, because the initial value of the concentrations poses a bias.

In such context it has to be kept in mind that prior to the application of the chemometric tools the data are, as usual, zero-mean-centred and scaled by the standard variance, which might also contribute to the bias. Hence, three latent variables would be justifiable in the identification of the reference (N)PLS models.

The extra latent variable in these structures might be thought to account for the dynamics, which however is for the cost of a higher number of parameters involved, with the subsequent cost of lower BIC value.

In general, it is worth noting that for the identification of the number of latent vari-

ables for the hybrid model the kinetic dimensions with or without mechanistic knowledge incorporation can be reduced to two independent rates, which might suggest that any additional kinetic rate of the simulation model may be redundant.

3.4.4 Challenges of the experimental case study

The challenge in this case study on *Bordetella pertussis*, arises mainly from the typical infrequent, sparse and noisy experimental concentration data available. The main objective was to show that the developed hybrid model is under these circumstances competitive with the reference dynamic (N)PLS models. The number of latent variables was unknown a priori and such was also subject of the study.

The "best" number of latent variables

The BIC values of the hybrid models were significantly better when compared to the ones of the reference dynamic (N)PLS approaches, this being mainly due to the smaller number of modeling parameters involved in the former.

For both data sets of this study (section 3.4.1) the BIC values obtained on the application of the hybrid models to the validation batches suggest the selection of two latent variables, which is partially in agreement with the reference (N)PLS structures identified (see Table 3.2). It has been seen that due to defective initial values, the MSE values obtained for the same validation batches were inconsistent among themselves, but the BIC values obtained for the corrected files nevertheless reinforce the selection of two latent variables.

For the case of the reference dynamic (N)PLS models, it was observed that in general five to six latent variables are necessary to obtain model performances which are, in terms of the MSE, in the same range than those of the hybrid models. Exceptions to this observation exhibit the performances of the dynamic NPLS models of Set 2 (Table 3.2), which both only comprise two latent variables. It seems that nonlinear inner functions are capable to account better for the general process dynamics than linear ones. This assumption is further supported by the observation that the MSE values of the test data obtained for the nonlinear models are significant smaller than the ones obtained for the linear models.

General Performance

The following observations hold concerning general model performance:

(i) The MSE values obtained for the hybrid structures are seen to be significantly better

than the ones obtained with the (N)PLS structures for the test set, as presented in Table 3.2. The correction of those initial values of the substrate concentrations in the test batch (initial data corrected as described in section 3.4.2), have lead to further improved performance in terms of MSE for the hybrid model, as expressed in Table 3.3. Considering that only the initial substrate concentrations were corrected, it can be further concluded that the hybrid model captures well the known fact that the estimation of the biomass concentration is sensitive to the initial substrate concentration.

Table 3.3 Values of model performance criteria over model types and structural parameters - corrected initial value data of data Set 1 of the experimental case study on the *Bordetella pertussis* cultivation.

Model type	Structure	BIC train	BIC valid	BIC test	MSE train	MSE valid	MSE test
ARX-PLS	[lv[a] = 6, nt[b] = 3]	-550	-213	-53	0.1582	0.0784	0.3047
Hyb – NPLS	[lv[a] = 2]	-329	-54	22	0.1019	0.0371	0.0018

[a] lv: number of latent variables

[b] nt: number of time series elements

(ii) The MSE values obtained with the application of the hybrid structure to the "corrected" validation batches (initial data corrected as described in section 3.4.2) were significantly better than those of the reference ARX-PLS model.
In the context of this analysis, it should be pointed out that the ARX-PLS reference model exhibit a rather low sensitivity to initial values. This can be observed in the difference between the MSE values of the corrected and uncorrected test batch, which is of the order 0.0088, about eleven percent of the respective MSE value.

(iii) The good performance of the hybrid models for the estimation of the biomass concentration is also observed in Fig. 3.5. The superior performance of the hybrid models is strengthened by comparing, in the same figure, the shape of the trajectories, which are rather bumpy for the (N)PLS model against rather smooth trajectories for the hybrid models (especially for the one with three latent variables). In the case of the dynamic (N)PLS models, the mandatory pretreatment of the data, i.e. the application of a cubic smoothing spline (section 3.4.1), therefore does not seem to act smoothing on the estimates, but instead the error introduced through the data interpolation seems to board the predictions.

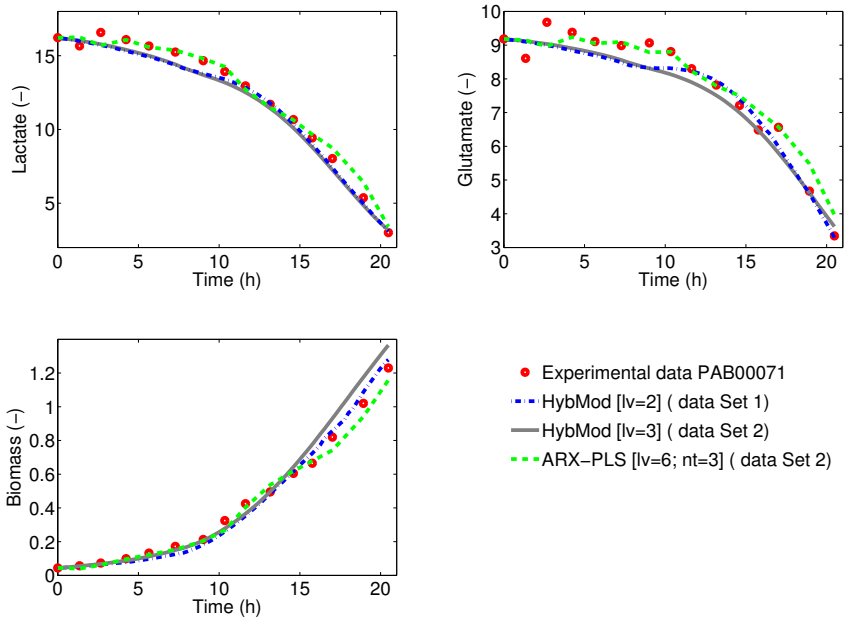

Figure 3.5 *Bordetella pertussis* experimental case study - plots of concentrations of lactate, glutamate and biomass concentrations over time for the validation batch PAB00071 (red dots): predictions of the NPLS hybrid model with 2 latent variables (dashed dotted blue line) and 3 latent variables (grey line), vs. estimates of a ARX-PLS, with 3 latent variables (dashed green line).

3.4.5 Complementary features of the hybrid model

In the Park Ramirez case study, it was observed that the identification of the non-parametric model parameters exhibited a faster convergence, a higher consistency of the results and an improved performance, e.g. in form of the MSE criteria in Table 3.1, when comparing between hybrid models with and without mechanistic knowledge, in favor of the former. Thus, the incorporation of mechanistic knowledge into the hybrid structure leads to a better model performance, which is in line with observations in (Oliveira, 2004; Psichogios and Ungar, 1992).

It is known that for (N)PLS models the analysis of the input and output scores represents a relevant source of information concerning characteristics and features of the processes and of model performance. This important feature of (N)PLS structures is present in the hybrid model developed in this study. For instance, and as illustration with the experimental case study, the plot of scores u_2 over t_2 (Fig. 3.6) shows a sin-

gularity of behavior for batch PAB0003 (see red crosses). This batch, employed in the training stage, distinguishes from the others by: (1) having the smallest initial values for all concentrations, namely lactate, glutamate and biomass; (2) exhibiting the highest concentration of biomass in the end of the batch; and (3) comprising a defect in the DO signal towards the end of the batch.

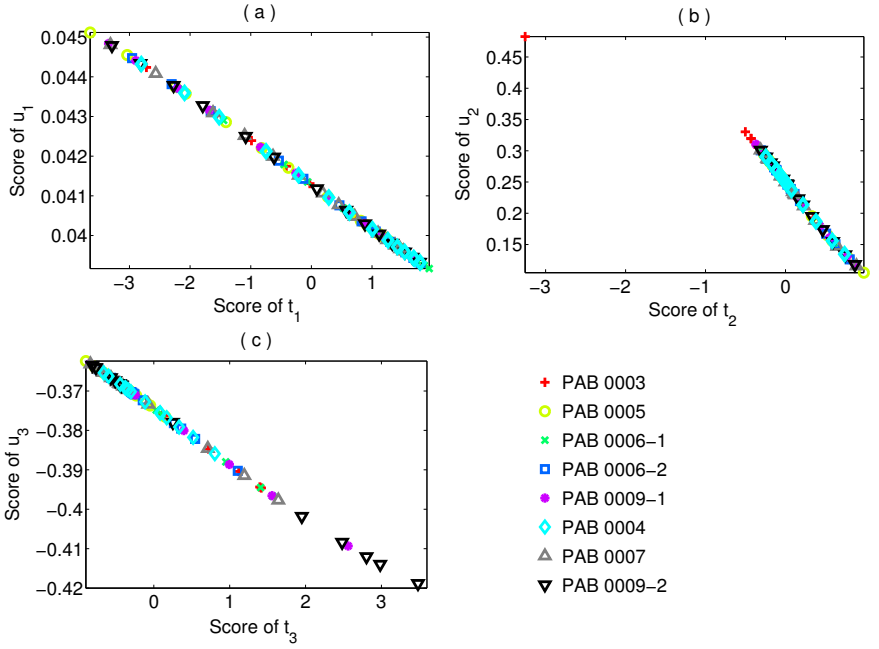

Figure 3.6 Bordetella pertussis experimental case study - phase plane plots of input scores, t_i, vs. output scores, u_i, ($i = 1, 2, 3$ in a, b, c, respectively) for the three latent variables of the inner model ANN functions of the hybrid structure comprising 3 latent variables - application to all batches (PAB0003 red crosses; PAB0005 light green circles; PAB0006-1 green x-es; PAB0006-2 blue boxes; PAB0009-1 purple filled squares; PAB0004 turquoise diamonds; PAB0007 gray upward-pointing triangles and PAB0009-2 black downward-pointing triangle).

Finally, and still for the experimental case study, the fairly linear inner model functions, which can be seen in Fig. 3.6, might explain for the fast convergence and the consistency of the hybrid model performance, as in general the optimal parameters of linear models are unique.

3.5 Conclusion

A novel methodology consisting of a hybrid dynamic (N)PLS model together with an algorithm for parameter identification is proposed for bioprocess modeling. The model consists of a set of macroscopic material balance equations in which the kinetic rates (the reaction terms) are mimicked by a NPLS (Nonlinear Partial Least Square) submodel and wherefore the global approach belongs to the class of hybrid models.
This methodology was benchmarked against reference dynamic (N)PLS models, (in which the dynamics are modeled by the augmentation of the inputs by lagged variables, such as FIR or AR(X)) through the application to two complementary case studies; (i) a simulation case study, also called the Park Ramirez simulation case after Park and Ramirez (1988); and (ii) an experimental case study of a *Bordetella pertussis* cultivation, as published by Soons & al. (2008a,b).
The following has been observed and can be stated:

(i) The novel approach, due to its inherent dynamics, exhibits, in general, fewer model parameters which results in a higher statistical confidence, observed in form of higher BIC values, when compared to the reference dynamic (N)PLS models.

(ii) In the application to validation data, the model performance, observed in terms of the MSE criterion, was generally significantly better.

(iii) Better calibration properties can be observed, expressed in terms of extrapolation capabilities to broader process conditions (e.g. predictions concerning the abnormal fed-batch data).

(iv) The application of the proposed model to typical infrequent, sparse and noisy experimental data leads to realistic, smooth trajectory estimations of the process states and does not require data interpolation as necessary in the reference dynamic (N)PLS methods.

(v) The integration of mechanistic knowledge into the proposed framework was identified to have a significant impact on the good results obtained, which is in line with the findings of (Oliveira, 2004; Psichogios and Ungar, 1992).

(vi) The novel proposed nonparametric structure and the related parameter identification algorithm exhibit PLS features such as dimension reduction and the opportunity to interpret the plot of scores:

(a) The Park Ramirez case study involves four kinetic rates, where two of which are linearly correlated. The hybrid model revealed that only two independent latent variables are already sufficient to model the process, in contrast to mostly four obtained by the reference (N)PLS models.
In general fewer latent variables were required regarding the same process than by the reference dynamic models.

(b) For the *Bordetella pertussis* case study, from the analysis of the score plots, it was shown that unusual variations in the process conditions could be identified.

(vii) Several sources of errors were identified: (a) noise in the input measurements to the nonparametric model; (b) noise in the measurements of the feeding rates (in the Park Ramirez case study); (c) errors inherent to the feedback nature of the models (where applicable); or (d) defective initial values.

(viii) For all sources of errors, except for the case of defective initial values, it was observed that state feedback to the nonparametric model had a damping effect on error propagation.

(ix) For cases of defective initial values, it was shown that corrective action on such errors has led to improved performance of the hybrid approach in comparison to the reference dynamic (N)PLS models (e.g. a more than twofold improvement of the MSE value in the experimental case study on a *Bordetella pertussis* cultivation).

In all, it can be stated that the application of a suitable hybrid (N)PLS model structure leads to significantly enhanced process estimations when compared to the reference dynamic (N)PLS models.

3.6 Appendices

3.6.1 The calculation of the Input & Output Scores

The input and output scores are an inherent component of the proposed nonparametric structure, eq. (3.5). The scores, in analogy to (N)PLS models, give an insight into the information captured by the respective submodel and are further suitable to identify "abnormal" process behavior.
The input scores, also called input latent variable, are directly obtained from multiplication of the input vector $L_{i,1..k}$ (see eq. 3.5) with the input loadings, $W_{x,i}$, i.e.:

$$t_i = W_{x,i} \cdot L_{i,1..k}. \tag{3.21}$$

The output scores are obtained by processing the input scores, t_i, with the ANN inner model, such that

$$u_i = (w_{2,i} \cdot g(w_{1,i} \cdot h(t_i) + b_{1,i}) + b_{2,i}), \tag{3.22}$$

using the weights, biases and functions defined in section 3.3.2.

3.6.2 The Sensitivity equations

The derivative of eq. 3.10 can be split into the derivative of $dW_{x,i,lin}/dw_A$ and in $dW_{x,i}/dw_A$, where w_A, the vector of parameters, comprises $W_{x,i}$, $W_{y,i}$ and w. The latter derivative is straight forward as described above, section 3.3.2. Considering eq. (3.9), the derivative $dW_{x,i,lin}/dw_A$ can be extended to

$$\frac{dW_{x,i,lin}}{dw_A} = \frac{dW_{x,i,lin}}{dW_{x,i,lin,un}} \cdot \frac{dW_{x,i,lin,un}}{dw_A} \tag{3.23}$$

making use of the chain rule. The first term on the right hand side is equivalent to Eq. (3.13). The second term on the right hand side is the derivative of Eq. (3.8) with respect to w_A. This term can be reformulated using the quotient rule to,

$$\frac{dW_{x,i,lin,un}}{dw_A} = \frac{(t_i^T \cdot t_i) \cdot \frac{d(L_{i,1..k} \cdot t_i)}{dw_A} - (L_{i,1..k} \cdot t_i) \cdot \frac{d(t_i^T \cdot t_i)}{dw_A}}{(t_i^T \cdot t_i)^2}. \tag{3.24}$$

The first derivative in the numerator can be split up, applying the chain rule again, to:

$$\frac{d(L_{i,1..k} \cdot t_i)}{dw_A} = t_i \cdot \frac{dL_{i,1..k}}{dw_A} + L_{i,1..k} \cdot \frac{dt_i}{dw_A}. \tag{3.25}$$

The second derivative can equivalently be treated, giving:

$$\frac{d(t_i^T \cdot t_i)}{dw_A} = 2 \cdot t_i \cdot \frac{dt_i}{dw_A}. \tag{3.26}$$

The derivative dt_i/dw_A emerges in (3.25) and (3.26), which, considering eq. (3.21) and applying the chain rule, can be formulated to:

$$\frac{dt_i}{dw_A} = W_{x,i} \cdot \frac{dL_{i,1..k}}{dw_A} + L_{i,1..k} \cdot \frac{dW_{x,i}}{dw_A}. \tag{3.27}$$

Noting that $w_A^T = [W_{x,i}, W_{y,i}, w]$, then the derivative corresponding to the second term on the right hand side is a matrix comprising the identity submatrix for the derivative of $W_{x,i}$ with respect to $W_{x,i}$ and zero elsewhere.

The derivative in the first term on the right side, namely $dL_{i,1..k}/dw_A$, also appears in eq. (3.25) and is reformulated using eq. (3.5) to:

$$\frac{dL_{i,1..k}}{dw_A} = \frac{dL_{i-1,1..k}}{dw_A} - \frac{d(W_{x,i-1} \cdot L_{i-1,1..k} \cdot W_{x,i-1})}{dw_A}, \tag{3.28}$$

where the second term on the right side can be simplified by using the chain rule, a straightforward solution and therefore not carried out here.

The only remaining derivative is $dL_{i-1,1..k}/dw_A$, which is calculated sequentially, starting with $dL_{1,1..k}/dw_A$. It should be noted that only the partition of entries of $L_{1,1..k}$, corresponding to the feedback of model estimates into the nonparametric model (Fig. 3.2) depend on w_A. As such those derivatives reduce to dc/dw_A which are nothing else than the derivatives given by eqs. (3.11), (3.14) and (3.15).

3.7 Acknowledgment

Sincere thanks for the provided data go to Mathieu Streefland, Zita I.T.A. Soons and the Netherlands Vaccine Institute and for financial support to the Fundação para a Ciência e a Tecnologia, where the reference number of the provided scholarship to Moritz von Stosch is: SFRH / BD / 36990 / 2007.

3.8 Nomenclature

Abbreviations	
AIC	Akaike Information Criterion
ANN	Artificial Neural Network
BIC	Bayesian Information Criterion
FIR	Finite Impulse Response
MSE	Mean Squared Error
NARX	Nonlinear AutoRegressive eXogenous
NPLS	Nonlinear-PLS / Neural Network-PLS
NIPALS	Non-iterative Partial Least Square
ODE	Ordinary Differential Equation
PCA	Principal Component Analysis

PLS	Partial Least Squares / Projection to Latent Structures
	Mathematical Symbols
$b_{1,i}$	Bias of the input layer in the latent variable submodel i
$b_{2,i}$	Bias of the hidden layer in the latent variable submodel i
c	Vector of concentrations
$c_{\sigma,j}$	Standard deviations
$c_{mes,j}$	Off-line measured concentration values
f	Function
$g(\cdot)$	Transfer function of the hidden layer in the latent variable submodel i
$h(\cdot)$	Transfer function of the input layer in the latent variable submodel i
i	Counter
j	Counter
k	Number of inputs
m	Number of kinetic functions
n	Number of components
o	Number of latent variables
r	Vector of kinetic rate functions
r_{Glu}	Specific Glutamate uptake rate
r_{Lac}	Specific Lactate uptake rate
$r_{P_{sec}}$	Specific secreted product formation rate
$r_{P_{tot}}$	Specific total product formation rate
r_S	Specific substrate uptake rate
t	Time
t_i	Input latent variable
u	Vector of control inputs
u_i	Output latent variable
w	Weights of all $i = 1..o$ ANNs
w_A	Vector of parameters
$w_{1,i}$	Weights of the input layer in the latent variable submodel i
$w_{2,i}$	Weights of the hidden layer in the latent variable submodel i
$x_{mes,1..n}$	Measurement data
D	Dilution rate
E_1	First Least Squared objective function
E_2	Second Least Squared objective function
Glu	Glutamate concentration
K	Matrix of stoichiometric coefficients
$L_{i,1..k}$	Inputs 1 to k for latent variable submodel i
$L_{res,1..k}$	Residual of the inputs
L_x	Vector of nonparametric model inputs
Lac	Lactate concentration
P	Number of samples
P_{sec}	Secreted Protein concentration
P_{tot}	Total Protein concentration
S	Substrate concentration
$S_{(F)}$	Feeding substrate concentration
$W^{up}_{x/y,i}$	Input/Output loadings obtained from the optimization procedure
$W_{x,i,lin,un}$	Unnormalized input loadings calculated by linear regression eq. (3.8)

$W_{x,i,lin}$	Normalized Input loadings
$W_{x,i}$	Input loadings of latent variable i
$W_{y,i}$	Output loadings of latent variable i
X	Biomass concentration
ϕ	Known kinetic functions
ρ / $\rho_{1..m,1..o}$	Unknown kinetic functions
μ	Specific biomass growth rate

Chapter 4

Modelling biochemical networks with intrinsic time delays: a hybrid semi-parametric approach

4.1 Abstract

This paper presents a method for modelling dynamical biochemical networks with intrinsic time delays. Since the fundamental mechanisms leading to such delays are many times unknown, non conventional modelling approaches become necessary. Herein, a hybrid semi-parametric identification methodology is proposed in which discrete time series are incorporated into fundamental material balance models. This integration results in hybrid delay differential equations which can be applied to identify unknown cellular dynamics.

The proposed hybrid modelling methodology was evaluated using two case studies. The first of these deals with dynamic modelling of transcriptional factor A in mammalian cells. The protein transport from the cytosol to the nucleus introduced a delay that was accounted for by discrete time series formulation. The second case study focused on a simple network with distributed time delays that demonstrated that the discrete time delay formalism has broad applicability to both discrete and distributed delay problems. Significantly better prediction qualities of the novel hybrid model were obtained when compared to dynamical structures without time delays, being the more distinctive the more significant the underlying system delay is. The identification of the system delays by studies of different discrete modelling delays was enabled by the proposed structure. Further, it was shown that the hybrid discrete delay methodology is not limited to dis-

crete delay systems. The proposed method is a powerful tool to identify time delays in ill-defined biochemical networks.

4.2 Background

Time delays play a very important role in genetic regulatory systems. Gene regulation and signal transduction as a whole involves the synthesis and maturation of complex proteins. Their synthesis and transport takes a considerable amount of time, which introduces delays in the overall regulation chain. At a process level, metabolic time delays can be observed macroscopically by recognizing a certain time delay between substrate uptake and the corresponding biomass growth or product formation as in cultivations of *Saccharomyces cerevisiae* (Daugulis *&* al., 1997) or *Pichia pastoris* (Ren *&* al., 2003). The nature of time delays in regulatory networks is twofold (Nikolov *&* al., 2008). They are either related to a process that takes an intrinsic time to be accomplished, i.e. some reactions, such as translational or transcriptional reactions, take a significant amount of time to be completed, or as a consequence of the modelling approach used, i.e. lumping a sequence of events might lead to an apparent time delay.
The bottom-up systems biology approach for building dynamic network models can be too cumbersome due to their complex nature and lack of fundamental knowledge (Teixeira *&* al., 2007b; Wang and Chen, 2010; Wang *&* al., 2010b; Yang *&* al., 2010). Typical limitations are the involvement of large scale kinetic models with poorly defined kinetic parameters, limited generalization capacity and their cost expansive development. In this paper we propose the use of mathematical hybrid semi-parametric systems as a cost effective alternative to model biochemical networks with intrinsic time delays, since it is not likely to know in advance which fundamental mechanisms cause such delays. Hybrid semi-parametric systems combine fundamental (parametric) biological constraints with more empirical data-based (nonparametric) constraints. Mechanistic and nonparametric models can therein be arranged in two possible ways: parallel or serial (Oliveira, 2004; Peres *&* al., 2008; Preusting *&* al., 1996; Psichogios and Ungar, 1992; Schubert *&* al., 1994b; Teixeira *&* al., 2007b). In the serial structure, which has been the one applied in this study, the biological system dynamics are described by time differentials of classifying variables, while the unknown metabolic functions with intrinsic delays are handled by a nonparametric structure.
A general mathematical representation of delayed dynamics is given by Retarded Functional Differential Equations (RFDE) (Bocharov and Rihan, 2000). After applying certain simplifications, some special concepts arise such as models considering either dis-

crete time delays, (Smolen & al., 1999; Tian & al., 2007; Wolkowicz & al., 1997), distributed time delays, (Daugulis & al., 1997; Rateitschak and Wolkenhauer, 2007; Wolkowicz and Xia, 1997) or ordinary differential equations (ODE) of kinetic rates (Ren & al., 2003). Although with varying performance, these models are shown to be capable of explaining the stability of biochemical networks (Bocharov and Rihan, 2000; Chen and Chang, 2008; Chen and Chen, 2009; Daugulis & al., 1997; Rateitschak and Wolkenhauer, 2007).

Similar simplifications of RFDE as reported for parametric models can also be applied to hybrid semi-parametric models, i.e. either discrete delays or distributed delays of state variables in the kinetics or differential equations of the kinetics. The latter is not well suited for hybrid modelling, because neither kinetic function nor the kinetic values are known a priori and thus a solution or estimation of the kinetics is not straightforward. Distributed delays are also rather unlikely to be used, because one would have to introduce some function accounting for the delay, which is generally not known. Furthermore, some mathematical postulation of arbitrarily large delays for unknown weighting functions of the delayed variable would have to be assumed and this mathematical convenience is in limit biologically unrealistic (see (Bocharov and Rihan, 2000)). Instead, the use of discrete delays in the inputs to the nonparametric structure is proposed herein. This is analogous to the application of discrete time series, namely AutoRegressive (eXogenous), (AR(X)), models. This presents no limitation for application, as it is mathematically clear that a weighted discrete time series is equivalent to the integration of a time delay weighting function and thus analogous to the application of the distributed delay framework.

Unfortunately, the theoretically endless number of time lagged values of one variable would in practice lead to high computational times and identification problems of the network structure and parameters (see (Bishop, 1995; Haykin, 1998)). In theory, an optimal number of time lagged values exists given by the ratio between redundancy and additional gain of information in the inputs. Several methods for identification of the optimal number of time lagged values such as autocorrelation, cross-correlation, or partial mutual information, have been proposed (Bishop, 1995; Haykin, 1998). However, they either assume that inputs are linearly correlated or are based on maximum information transfer and thus require known outputs, which unfortunately are not directly available in hybrid models. Hence these methods cannot be applied here and thus the number of lagged values is rather chosen by trial and error, as done by several other authors (Parlos & al., 1999; Qin and McAvoy, 1996). However, choice by trial and error is not a disadvantage, when (i) the delay is an important property of the system and when

(ii) series of delays are systematically studied, since it can be expected that models that account for time delays perform the best when the studied and the "true" delays are congruent.

In this paper hybrid delay differential equations with discrete time series was determined to be a powerful method to identify delayed dynamics of ill-defined biochemical networks. This technique is described in detail in the results section. The technique was applied to a typical gene regulatory system where the transport of macromolecules between the cytosol and nucleus introduce strong delay dynamic effects. In addition, heterologous protein expression by recombinant *Pichia pastoris* was studied by assuming a hypothetical network with distributed time delays.

4.3 Results & Discussion

4.3.1 Delay Differential Equation Hybrid Model (DDEHM)

Material balances over intracellular metabolites can be generically stated by the following dynamical equation

$$\frac{dc_{int}}{dt} = K_{int} \cdot r_{int} + b_{int} - \mu \cdot c_{int} \tag{4.1}$$

where c_{int} is a vector comprising the concentrations of intracellular metabolites, K_{int} a $m \times q$ stoichiometric matrix of m metabolites and q metabolic reactions, r_{int} a vector of q kinetic rates, b_{int} a vector of transport fluxes across the cellular membrane and μ the specific growth rate.

If a macroscopic bioreactor model is formulated accounting only for the unbalanced extracellular metabolites, a similar equation is obtained which accounts for the volume dilution term ($D \cdot c_{ext}$) in substitution of the cell growth dilution term ($\mu \cdot c_{int}$),

$$\frac{dc_{ext}}{dt} = K_{ext} \cdot r_{ext} - D \cdot c_{ext} + u_{ext}. \tag{4.2}$$

Here c_{ext} is a vector of concentrations of extracellular metabolites, K_{ext} a matrix of stoichiometric coefficients, D is the dilution rate, u_{ext} is a vector of volumetric feeding rates, and r_{ext} is the kinetic rate vector.

All the results, presented from this point forward, are derived from eq. (4.2), which can however be automatically extended to eq. (4.1).

Delayed reaction kinetics

As suggested by (Oliveira, 2004), the vector of kinetic rates can be described either mechanistically, statistically or as a mixture of both types of models depending on the a priori knowledge about the metabolic network. A general definition is to state every metabolic flux as the multiplication of a mechanistic term (ψ) with an unknown non-parametric term (ρ) representing the unknown phenomena that must be identified from data.:

$$r(X, w) = \psi(X) \cdot \rho(X, w), \tag{4.3}$$

with X a vector of input variables and w a vector of empirical parameters. When no a priori mechanistic knowledge is available then the ψ term is dropped and eq. (4.3) reduces to

$$r(X, w) = \rho(X, w). \tag{4.4}$$

As stated previously, the intrinsic causes of delays are the occurrence of several serial reaction steps with slow kinetics. To mimic this effect, and analogous to AR(X) models, both the ψ and ρ kinetic terms are modelled as a function of X, which includes discrete past values of metabolite concentrations, c,(that can be intracellular or extracellular, depending on the application of eqs. 4.1 or 4.2) and/or exogenous inputs:

$$X = \begin{bmatrix} c_i(t), c_i(t-\tau_i), c_i(t-2\cdot\tau_i), \ldots, c_i(t-N_i\cdot\tau_i), \\ s_j(t), s_j(t-\tau_j), s_j(t-2\cdot\tau_j), \ldots, s_j(t-M_j\cdot\tau_j) \end{bmatrix}. \tag{4.5}$$

Here c_i means value i of vector c, τ_i is the associated time lag, N_i defines the number of time lags assumed for each value c_i of vector c, s_j is the j^{th} exogenous input, τ_j the associated time lag and its lag number is defined by M_j. Note that the time lags and the numbers of time lags, τ_i, τ_j, N_i, and M_j can be chosen independently. However, it might be advantageous to model a time series around rough estimates of the "true" delays.

After considering eq. (4.2) - (4.5), it becomes clear that the model equations are Delay Differential Equations (DDE) in which the "retarded" or "lagged" phenomena are accounted by the reaction term, eq. (4.4).

Several linear or nonlinear regression methods can be used to formulate the unknown nonparametric kinetic function ρ. Here we adopted a three layer back propagation neural

network with hyperbolic tangential activation function for the sake of comparability with other hybrid modelling studies since this method is the most reported in the literature (Oliveira, 2004; Peres & al., 2008; Preusting & al., 1996; Psichogios and Ungar, 1992; Schubert & al., 1994b; Teixeira & al., 2007b):

$$\rho(X, w) = w_2 \cdot g(w_1 \cdot X + b_1) + b_2, \tag{4.6}$$

where w, the parameter vector, comprises the weights and biases, w_1, w_2, and b_1, b_2, respectively. The hyperbolic tangential activation function $g(\cdot)$ is,

$$g(y) = \frac{1 - e^{-2 \cdot y}}{1 + e^{-2 \cdot y}}. \tag{4.7}$$

Note that the incorporation of AR into the hybrid approach results in delay differential equations, which is why the proposed hybrid model is referred to as the Delay Differential Equation Hybrid Model.

Nonparametric structure identification

The identification of the best network architecture by means of a trade-off between residual minimization, quantity of data and quantity of parameters is a central question when nonparametric models find application. This trade-off is due to the fact that more parameters on one hand will improve the fitting of the model to the data, but on the other hand might result in parameter over-fitting, leading to a degradation of model robustness or/and, even worse, in the addition of synthetic noise to the estimates (Bishop, 1995; Haykin, 1998).
The architecture of the Artificial Neural Network (ANN) structure involves the variation of the number of layers and the number of nodes. This variability is in this study, prior to application, already reduced by the selection of three layers, namely input, hidden and output layer. The application of three layers is usually sufficient if nonlinear continuous functions are sought to be modelled (Bishop, 1995). Remaining in terms of structural variability is such the evaluation of the variation of numbers of nodes for each hybrid model set-up.

Parameter identification

For each nonparametric structure, the respective parameters w must be estimated from data. In this paper a weighted least squares criteria of model residuals in concen-

trations is adopted:

$$min\left\{E = \frac{1}{P \times n}\sum_{l=1}^{P}\sum_{i=1}^{n}\frac{(c_{mes,l,i}(t) - c_{l,i}(t,w))^2}{c_{\sigma,i}}\right\}, \tag{4.8}$$

where P is the number of samples, n is the number of state variables, $c_{mes,l,i}$ are measured state variables, $c_{l,i}(t,w)$ are calculated state variables and $c_{\sigma,i}$ are the standard deviations. The serial hybrid structure, consisting of an ANN and material balances, was shown to be trained best by using the sensitivity approach along with analytical gradients (Oliveira, 2004). Here we extended the sensitivity equations to the DDEHM case. The sensitivities equations are derived by differentiating eq. (4.2) with respect to w while taking into account the time lagged differential variables, which then reads as follows,

$$\frac{d}{dt}\cdot\frac{dc}{dw} = \sum_{k=0}^{N_i}\left\{\frac{\partial(K\cdot\rho\cdot\psi)}{\partial c(t-k\cdot\tau)}\cdot\frac{dc(t-k\cdot\tau)}{dw}\right\} + \frac{\partial K\cdot\psi\cdot\rho}{\partial w} - D\cdot I_n\cdot\frac{dc}{dw}, \tag{4.9}$$

where

$$\begin{aligned}&\sum_{k=0}^{N_i}\left\{\frac{\partial(K\cdot\rho\cdot\psi)}{\partial c(t-k\cdot\tau)}\cdot\frac{dc(t-k\cdot\tau)}{dw}\right\} = \\ &\quad K\cdot\rho\cdot\sum_{k=0}^{N_i}\left\{\frac{\partial\psi}{\partial c(t-k\cdot\tau)}\cdot\frac{dc(t-k\cdot\tau)}{dw}\right\}, \\ &+K\cdot\psi\cdot\sum_{k=0}^{N_i}\left\{\frac{\partial\rho}{\partial c(t-k\cdot\tau)}\cdot\frac{dc(t-k\cdot\tau)}{dw}\right\}\end{aligned} \tag{4.10}$$

with ρ and ψ depending on the time lagged concentrations and where

$$\frac{\partial K\cdot\psi\cdot\rho}{\partial w} = K\cdot\rho\cdot\frac{\partial\psi}{\partial w} + K\cdot\psi\cdot\frac{\partial\rho}{\partial w}; \tag{4.11}$$

For comparison of time-delay gradients for network training see (Haykin, 1998; Nikolov & al., 2008)

This least square problem is solved by using the "lsqnonlin" Matlab function which uses a subspace trust region method and is based on the interior-reflective Newton method (Matlab Optimization toolbox) (Coleman and Li, 0 01). The sensitivity equations are integrated along with the delay differential model equations. This can either be accomplished using the dde23 Matlab algorithm, which integrates the delay differential

equations with the explicit Runge-Kutta (2,3) pair and interpolant, or by using linear approximation of the differential equations for integration with storage of the respective delay values, which results in a time inexpensive algorithm. For the latter case, unfortunately, some error is introduced along with this simplification. However if average kinetic rates are estimated for each time step, the error is significantly diminished. Initial state values, $c(t_0)$, are problem dependent (for instance the initial concentration of biomass or substrate in a bioreactor). The initial values of the sensitivity equations are however zero ($(dc/dw)_{t_0} = 0$, $(dc/dw)_{t<t_0} = 0$), because the initial state values, $c(t_0)$, are independent of model parameters w. The residual gradients are then obtained using the corresponding sensitivity values. Notice that the lagged values of both state variables and exogenous inputs are assumed to be equal to the initial values $c(t_0)$ for all $t - N_i \cdot \tau < t_0$.

Identification is initialized from a random selection of weight values as usually done for ANNs. The solution space is spanned by these weights and the identification, i.e. the objective to reduce the model residual, is a nonlinear optimization problem. Therefore, one cannot expect to obtain the global minimum as the result of the model's residuals minimum found from one random weight initialization. Instead, several iterations of the same set-up with random initialization should be carried out. The greater the number of random initializations, the greater the statistical confidence of the solution (Bishop, 1995; Haykin, 1998).

However, parameter identification is an iterative process which should be stopped when the model exhibits the best generalization of the target functions (Bishop, 1995; Haykin, 1998). This is usually accomplished using two independent data sets: one for identification (also called training) which contains about 2/3 of all data points and another data set for validation with the remaining data. For these data sets some error criteria such as the Mean Least Square Error or the Bayesian Information Criterion (described in detail below) is calculated for the model residuals. Along the iterations, the best parameters are the ones where the selected criterion of the validation data set has its “best” value. A test data set can be used to additionally exploit the generalization capabilities.

Model performance criteria

The model residual, also addressed as the goodness of fit of the model estimates and the data, can be assessed with the Mean Square Error, MSE. The MSE decreases

the better the fit and is defined as:

$$MSE = \frac{1}{P \cdot n} \cdot \sum_{l=1}^{P} \sum_{i=1}^{n} (c_{mes,l,i}(t) - c_{l,i}(t, w))^2. \tag{4.12}$$

This criterion is directly linked to the least square error which is used for parameter identification.

Due to the reason mentioned above, the MSE criterion is not addressed when it comes to architecture, structure, model comparison or selection. Appropriate criteria are (i) the Akaike Information Criteria, AIC, which is wildly used or (ii) the Bayesian Information Criteria, BIC, which is more appropriate for datasets with more than 46 data points (Burnham and Anderson, 2004; Leonard and Hsu, 1999; Peres & al., 2008). Therefore the BIC is applied for model comparison and selection in this study. The BIC is defined as:

$$\begin{aligned} BIC = & \left(-\frac{n \cdot P}{2} \cdot \ln \left(\sum_{l=1}^{P} \sum_{i=1}^{n} [c_{mes,l,i}(t) - c_{l,i}(t, w)]^2 \right) \right) \\ & - \left(\frac{n_w}{2} \cdot \ln \left(\frac{n \cdot P}{2 \cdot \pi} \right) \right) \end{aligned} \tag{4.13}$$

where the term in the first bracket is the logarithmic maximum likelihood, π is the number "Pi" and n_w is the total number of parameters/weights. In terms of the BIC, the model to be selected is the one that exhibits the larger BIC value for the validation set, see (Burnham and Anderson, 2004; Leonard and Hsu, 1999; Peres & al., 2008).

4.3.2 Case Study I: Transcription Factor A (TF-A) dynamics with discrete time delay

Genetic regulatory systems are built on signal transduction pathways through which specific transcription factors (TF) are phosphorylated. The phosphorylated TFs are then able to bind to responsive DNA sequences thereby regulating the transcription of nearby genes. Herein we consider the example of the TF-A model reported by (Smolen & al., 1999) and (Tian & al., 2007) (see Fig. 4.1A). In this case, the TF activates its own transcription according to a typical positive feedback loop.

The translocation of macromolecules between cytosol and nucleus have a tremendous impact on gene regulation dynamics. Herein we consider a discrete delay for the

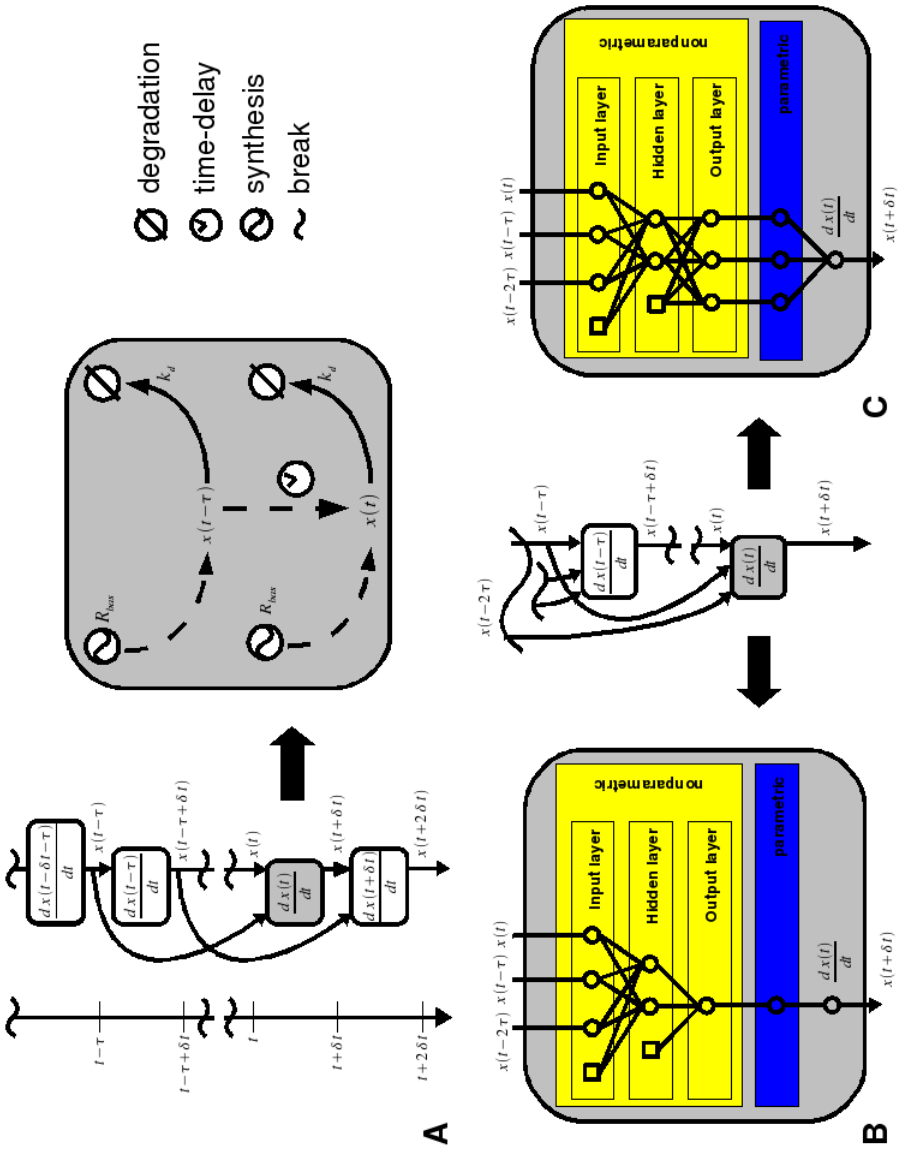

Figure 4.1 Network Structures:
Delay TF-A transcription model. (A) true network structure (B) DDEHM network without prior knowledge, (C) DDEHM network with some prior knowledge. In structures (B) and (C), the ANN comprises three layers. The nodes of the input and output layer have linear transition functions, except for the input node of the time which has a hyperbolic tangential transition function as do the nodes of the hidden layer.

translocation of TF-A as suggested by (Smolen & al., 1999) and (Tian & al., 2007), giving rise to the following single delay differential equation describing the dynamics of the TF-A monomeric concentration in the nucleus, x:

$$\frac{dx(t)}{dt} = \frac{(k_f \cdot x(t-\tau)^2)}{(x(t-\tau)^2 + K_d)} - x(t) \cdot k_d + R_{bas} \tag{4.14}$$

The first term on the right-hand side of eq. (4.14) is the rate of TF-A transcription in the cytosol which in the perspective of nucleus is affected by the translocation delay, $\tau = 120min$. The second term refers to TF-A dissociation in the nucleus and the third term to a basal transcription rate, R_{bas}, observed at very low TF-A concentrations.
Figure 4.2 shows the simulation of model eq. (4.14) with the parameters proposed by (Smolen & al., 1999). The TF-A dynamics are of a typical bistable system induced by the increase of the cytosol synthesis rate, k_f, at time $t = 200min$, forcing the system to jump to another state. The effect of the time delay can be assessed by comparing the full-line (with delay) with the dashed-line (without delay). The main consequence of the delay is that the TF-A concentration exhibits a "staircase" transition between the steady-states.
The main goal in this case study is to investigate if the TF-A delay dynamics, shown in Fig. 4.2, can be properly identified by the DDEHM framework proposed in this paper. With this goal in mind, 6 data sets of TF-A concentration in the nucleus over time with varying initial concentrations were generated (3 data sets with "clean" data, which were corrupted with white noise in order to obtain the training, validation and test set data).

Formulation and discrimination of a suitable DDEHM structure

The two DDEHM structures, shown in Fig. 4.1B and 4.1C, were identified from the simulation data. In the former structure, no prior knowledge about the TF-A network is incorporated while in the latter case some prior knowledge inspired in eq. (4.14) and in autoregulated systems is considered.)

(In preliminary studies, we concluded that structure (4.1C) leads to both a faster convergence and improved results than structure (4.1B) (results not shown). This observation is in line with the study reported by (Oliveira, 2004), where it was shown that including a priori knowledge in the hybrid structure generally improves their identification capacity.
A selection of results obtained with structure (4.1C) are presented in Table 4.1 showing model performance criteria for the training, validating and testing data sets (namely

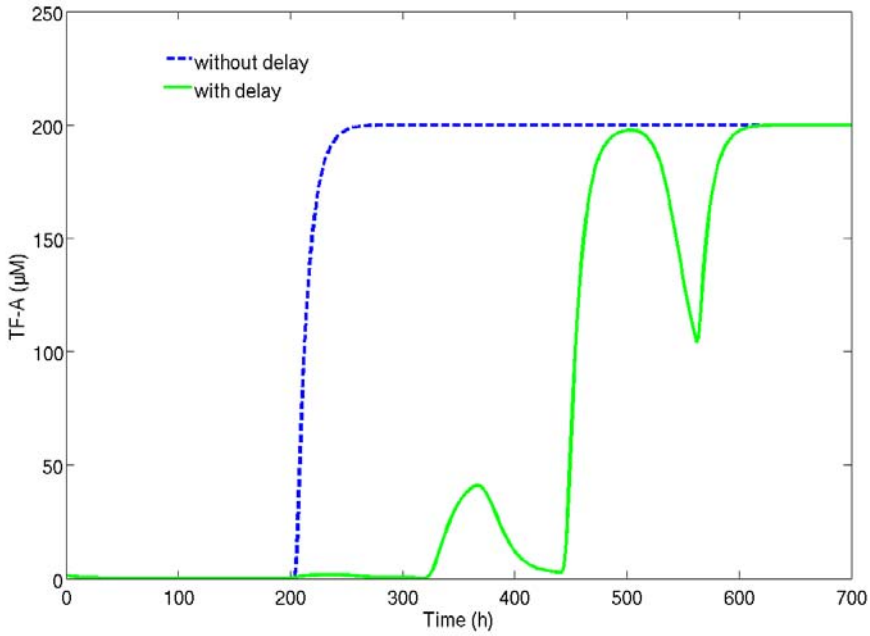

Figure 4.2 Impact of delays on the TF-A profile
Demonstration of the impact of the delay on the trajectory of TF-A transcription model over time. The TF-A model trajectory without delay is the blue dashed line while the TF-A trajectory with delay is the green continuous line.

MSE and BIC) over structure parameters. Overall, it can be observed that structures without time delays are in general outperformed by those containing time delays if one of the effectual delays is close to the "true" delay, i.e. a model with a delay mismatch as high as 10% still gives an improved performance in comparison to no delay at all (Table 4.1). It can also be noticed that the MSE values for the case of one delay tend to improve the closer the effectual delay gets to the "true" delay, peaking when the effectual is the true delay. Also, it strikes that the best models (highest BIC values for the validation set) are obtained mostly for 4-7 nodes in the hidden layer, an observation that reflects the complexity of the addressed system. Owed to this complexity, are also the strayed deviations in the overall consistent performance in terms of BIC. The consideration of series of delays also gives rise to consistent models, especially if only two delays are considered. When three delays are considered, model performance increases with decreasing number of nodes, which contrasts with the results obtained for one or two delays. Even so, the best values therewith are obtained with 4 numbers of

nodes. While the good model performances are due to the fact that the "true" delay is present in the applied models, the slightly worse performance when compared to the single delay models sources from the evitable, additional information. Evitable (correlated) information hampers the model structure identification (Bishop, 1995; Haykin, 1998), which explains why the model performance for three considered delays decreases with the increasing number of nodes.

The most consistent structure with highest predictive power has 5 nodes in the hidden layer and a single delay coincident to the "true" delay of 120 min. The respective BIC value was -5489 while the MSE value was 0.0071 for the test data set. The best structure without time delays, which had also 5 nodes in the hidden layer, showed a fourfold increase in the MSE value for the test data set (0.0210) and a considerably lower BIC value (-5997). This result clearly demonstrates the advantage of the delay hybrid modelling approach proposed in this paper.

Table 4.1 Results for Case Study I

Effect of structure parameters (number of nodes in the hidden layer, NN, and number and values of time delays) on the performance of the structure displayed in Fig. 1C. For every structure incorporating delays two random initial weight sets were investigated. For those without delays four different random initial weight changes were investigated. At least 25 iterations were carried out for each set of weights. The number of iterations was expanded if network learning was observed during the last iterations. Integration of the material balances along with the differential equations resulting from the sensitivity method for parameter identification is carried out for this simulation case with the dde23 MATLAB function for the studies with delays, and with the ode23 MATLAB function for the ones without delays. This results in higher simulation times, but as the dimension of the set of equations is rather small, the total simulation time is maintainable.

NN	τ_i	BIC			MSE			NN	τ_i	BIC			MSE		
		train	valid	test	train	valid	test			train	valid	test	train	valid	test
5	0	-12217	-5836	-5997	0.0141	0.0152	0.0210	6	0	-12220	-5869	-6039	0.0139	0.0157	0.0222
2	100	-13118	-6209	-6190	0.0368	0.0350	0.0337	2	110	-13058	-6150	-6157	0.0347	0.0310	0.0315
3	100	-13087	-6273	-6336	0.0350	0.0384	0.0437	3	110	-13043	-6269	-6275	0.0334	0.0381	0.0385
4	100	-11826	-5650	-5888	0.0096	0.0105	0.0170	4	110	-12273	-5805	-5832	0.0151	0.0144	0.0152
5	100	-11386	-5379	-5733	0.0060	0.0059	0.0120	5	110	-12302	-5864	-6008	0.0152	0.0156	0.0210
6	100	-12873	-6174	-6176	0.0265	0.0282	0.0284	6	110	-13162	-6336	-6329	0.0355	0.0392	0.0386
7	100	-13144	-6269	-6176	0.0342	0.0330	0.0273	7	110	-11516	-5572	-5731	0.0066	0.0081	0.0111
2	120	-13047	-6148	-6139	0.0343	0.0309	0.0303	2	130	-13242	-6332	-6371	0.0417	0.0449	0.0486
3	120	-12105	-5782	-5960	0.0130	0.0142	0.0204	3	130	-13076	-6173	-6203	0.0346	0.0314	0.0333
4	120	-11974	-5761	-5891	0.0111	0.0132	0.0171	4	130	-12652	-6087	-6090	0.0221	0.0254	0.0256

5	120	-11436	-5462	-5489	0.0062	0.0068	0.0071	5	130	-11823	-5604	-5676	0.0094	0.0092	0.0107	
6	120	-10820	-5170	-5714	0.0033	0.0036	0.0108	6	130	-12679	-6093	-6108	0.0218	0.0240	0.0247	
7	120	-12533	-6002	-5881	0.0184	0.0193	0.0151	7	130	-13269	-6384	-6393	0.0388	0.0417	0.0424	
2	140	-13069	-6155	-6167	0.0351	0.0313	0.0321	2	160	-13195	-6295	-6257	0.0398	0.0416	0.0385	
3	140	-12303	-5805	-5803	0.0158	0.0149	0.0149	3	160	-12252	-5823	-5771	0.0151	0.0155	0.0139	
4	140	-13288	-6375	-6384	0.0420	0.0455	0.0464	4	160	-13063	-6186	-6241	0.0334	0.0311	0.0347	
5	140	-12537	-6043	-6039	0.0193	0.0225	0.0223	5	160	-12022	-5716	-5909	0.0114	0.0116	0.0171	
6	140	-12564	-6067	-6078	0.0194	0.0228	0.0233	6	160	-12052	-5800	-5995	0.0116	0.0133	0.0197	
7	140	-11439	-5535	-5994	0.0061	0.0075	0.0189	7	160	-11466	-5431	-5441	0.0063	0.0061	0.0062	
2	80, 120	-13016	-6146	-6079	0.0330	0.0305	0.0266	2	120, 160	-12984	-6137	-6027	0.0320	0.0299	0.0240	
3	80, 120	-12334	-5860	-5968	0.0162	0.0164	0.0204	3	120, 160	-13115	-6296	-6163	0.0357	0.0397	0.0303	
4	80, 120	-11221	-5276	-5566	0.0051	0.0048	0.0087	4	120, 160	-12250	-5872	-5934	0.0145	0.0162	0.0183	
5	80, 120	-12780	-6221	-6207	0.0243	0.0314	0.0305	5	120, 160	-12293	-5872	-5984	0.0148	0.0155	0.0194	
6	80, 120	-12233	-5837	-5944	0.0136	0.0139	0.0172	6	120, 160	-11240	-5352	-7991	0.0050	0.0052	1.0762	
7	80, 120	-11688	-5663	-5630	0.0077	0.0094	0.0088	7	120, 160	-11703	-5623	-6004	0.0078	0.0086	0.0187	
2	80, 120, 160	-12994	-6144	-6034	0.0321	0.0300	0.0241	5	80, 120, 160	-12487	-5953	-6045	0.0178	0.0178	0.0215	
3	80, 120, 160	-11855	-5641	-5937	0.0099	0.0104	0.0189	6	80, 120, 160	-12824	-6193	-6213	0.0244	0.0276	0.0288	
4	80, 120, 160	-11879	-5605	-5734	0.0099	0.0092	0.0120	7	80, 120, 160	-12167	-5758	-5774	0.0122	0.0110	0.0113	

Comparison of best structures with and without time delays

Figure 4.3 compares the modelling results for the two best structures with delay $\tau = 120min$ and without delay for the validation and test data sets. It can be seen that the model without time delay provides a very smooth transition between the two steady-states. However, the "true" dynamics, i.e. the staircase transition function, of the measured data are not met. In contrast, the model with $\tau = 120min$ was able to capture these "staircase" dynamics. The curves both increase slightly in the beginning until a time value of about 320 min, where the first "stair" appears in the data. Thereafter, a significant increase can be noticed until a time of about 390 min, where the estimate has a first maximum peak. Then the concentration estimation decreases until a time point of about 450 min. Subsequently, in both cases, the data points are almost completely met by the estimates in the time interval from 450 until 650 min.

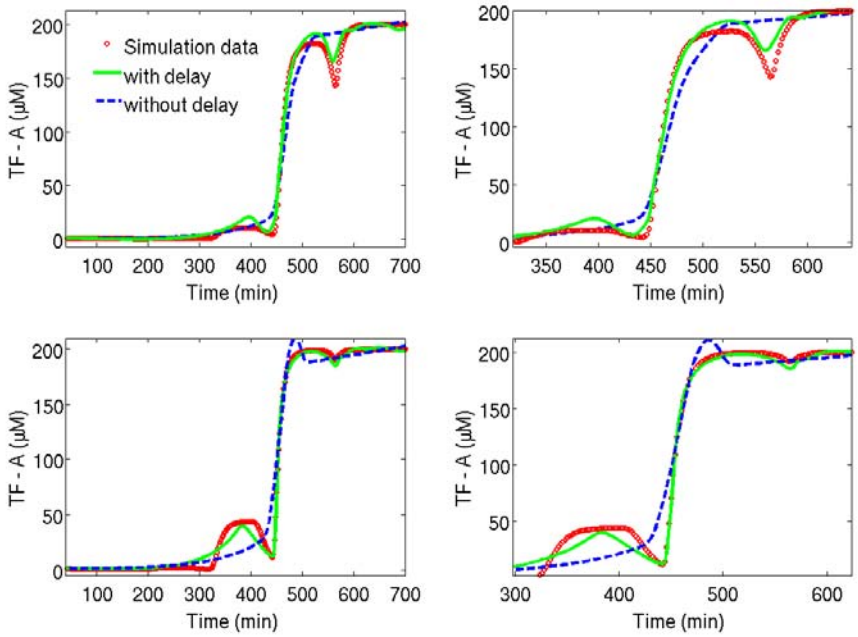

Figure 4.3 Qualitative results on the time curse of TF-A
TF-A modelling results for the two runs of test data. On the left side the whole simulation region of the data set is shown while on the right side the most interesting section of the respective data set is highlighted: red circles are "measured" TF-A data over time, green line are the identification results by model structure (4.1C) with 5 hidden nodes and $\tau = 120$ minutes; blue dashed line are the identification results by model (4.1C) with 5 hidden nodes and no time delay.

4.3.3 Case Study II: Heterologous protein expression by MUT+ Pichia pastoris

In methanol utilizing MUT+ *Pichia pastoris* strains, fast phenotype, foreign protein expression is controlled by the promoter of the alcohol oxidase gene 1 (AOX1). In typical culture conditions, the yeast cells are first grown on glycerol to reach a certain optimal cell density. Glycerol and most carbon sources other than methanol strongly repress AOX1, thus product is not formed in this phase. Then methanol feeding induces AOX1 over 1000-fold (Khatri and Hoffmann, 2006) thereby initiating foreign protein expression. The transition between glycerol and methanol phases can take between 1 to 4 hours depending on the strategy for methanol feeding. This transition phase corresponds to the time needed by the cells to express the alcohol oxidase enzyme, which is an essential enzyme for the cells to metabolize methanol. Apart from this delay in the transition phase, time delays between methanol uptake, biomass growth and product formation were also observed during the post-induction phase, (Ren *&* al., 2003). In the paper by Bellgardt and co-workers (Ren *&* al., 2003), a so called extended regulator model was adopted which is somewhat analogous to a linear distributed time delay kernel of the specific protein synthesis rate over the specific growth rate. The inclusion of such a delay model was essential to fit their data, although the underlying biological fundamentals are not clearly understood. The effects causing such time delays seem to be a principal part of the *Pichia pastoris* systems. However, they are poorly studied (Yamashita *&* al., 2009) and mechanistically not understood. Thus the nature of the apparent delays can mechanistically not be precisely defined (i.e. as a discrete delay model), wherefore a distributed delay model is the most appropriate representation.

The main goal in this section is to determine if the hybrid methodology proposed herein is able to effectively identify such unknown distributed time delay dynamics in *P. pastoris*. The *P. pastoris* network shown in Fig. 4.4A was used as a case study. This network includes a quadratic distributed delay kernel (Eqs. A5 and A6 of Table 4.2), which is considered as a strong delay kernel, see Fig. 4.5. In this case, the cell growth rate and the foreign protein expression rate are delayed in relation to the methanol uptake. The corresponding model equations are listed in Table 4.2. Note that the reactor balance equations are also listed since they are an important element to generate consistent simulation data.

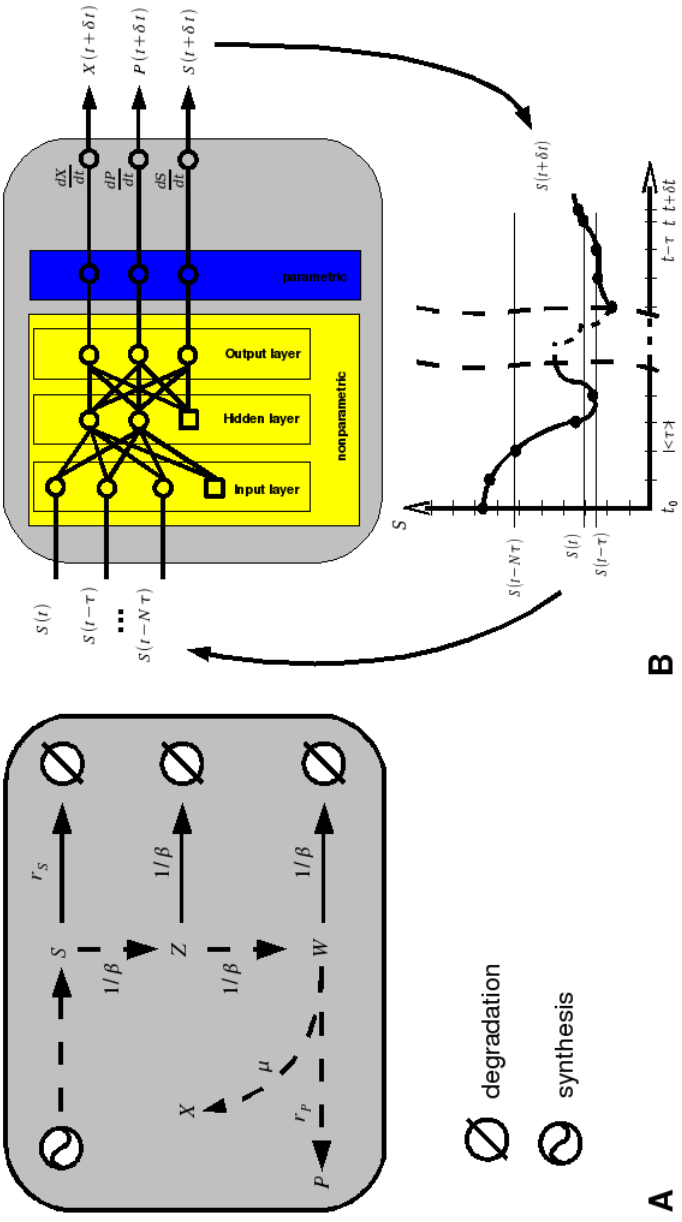

Figure 4.4 Pichia pastoris network with delay dynamics
(A) network with a quadratic distributed time delay kernel of cell growth and protein expression over methanol uptake. The respective equations are listed in Table 4.2. This network was used to generate simulation data (B) Approximation of network (A) by a hybrid network. Structure (B) was investigated to see if the novel framework is able to identify unknown distributed delay dynamics.

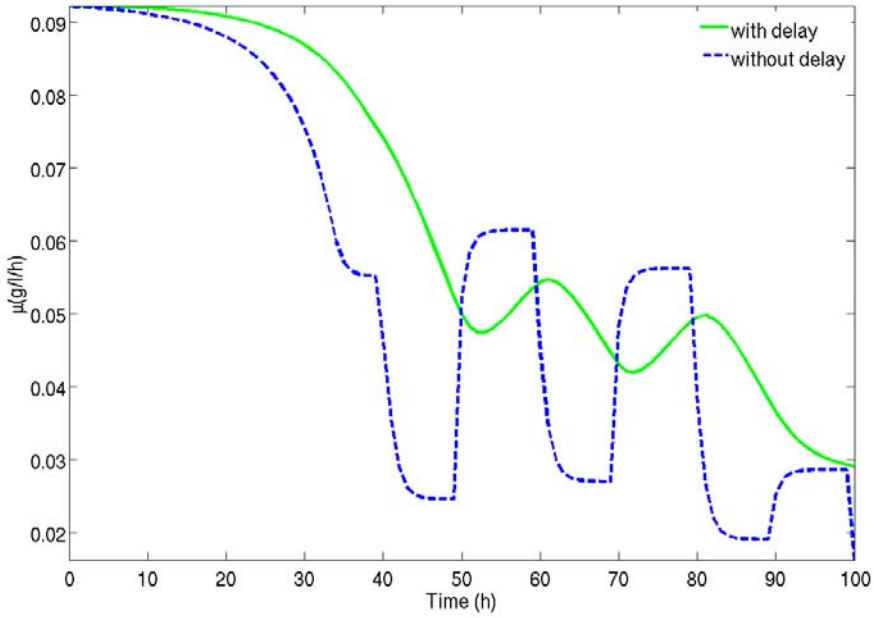

Figure 4.5 Impact of delays on the specific biomass growth rate
Green full line, is the specific growth rate when considering the network shown in Fig. 4.4A; blue dashed line is the specific growth rate when no delay between substrate uptake and biomass growth is considered.

Identification of distributed time delays by DDEHM

The hybrid model structure shown in Fig. 4.4B was adopted to identify the network, Fig. 4.4A. The neural network assumes no prior knowledge about Fig. 4.4B and uses as the external excitation signal present delayed methanol concentration values. The specific methanol uptake rate, $r_S(t)$, specific growth rate, $\mu(t)$, and specific product synthesis rate, $r_P(t)$, are the target kinetic variables that need to be identified. Note that in the real system $\mu(t)$ and $r_P(t)$ are delayed in relation to $r_S(t)$ according to a quadratic distributed delay kernel (see Eqs. A9-A12 in Table 4.2). These three kinetic rates are passed to the macroscopic reactor material balances for the calculation of the respective concentrations.

Table 4.2 Mathematical model for data generation

Mathematical model of MUT + *Pichia pastoris* expression with a quadratic distributed delay kernel. This model was used to generate six data sets. Three of which contain the clean, noise-free data and the other three the associated white noise corrupted data. One data set of the noise corrupted sets was used to train the hybrid model, one was used for validation and the third one for testing. Integration was performed with the ode45 MATLAB function which integrates the differential equation with a Runge-Kutta (4,5) integration scheme. The obtained state variables, namely concentrations of biomass, substrate and product, the reactor volume and as well the feed concentration are recorded and assumed as measured data for the evaluation. Variation in the data was obtained by application of varying initial values, i.e. the initial values were 5% Gaussian distributed. Note that model equations (A5 and A6) are derived from equation (A 12) using the linear chain trick (Rateitschak and Wolkenhauer, 2007; Wolkowicz and Xia, 1997) and that (A 12) is never used for model calculations.

Reactor model equations:

$\frac{dX(t)}{dt} = X(t) \cdot \mu(W(t)) + D \cdot X(t)$	(A 1)	$\frac{dS(t)}{dt} = -r_S(S(t)) \cdot X(t) - D \cdot (S(t) - S_F)$	(A 2)
$\frac{dP(t)}{dt} = r_P(W(t)) \cdot X(t) - D \cdot P(t)$	(A 3)	$\frac{dV(t)}{dt} = F(t)$	(A 4)
$\frac{dW}{dt} = \frac{Z-W}{\beta}$	(A 5)	$\frac{dZ}{dt} = \frac{S(t)-Z}{\beta}$	(A 6)
$F(t) = \left(\frac{V(t)}{S_F - S(t)}\right) \cdot \left(r_S \cdot X(t) + \frac{S_{set} - S(t)}{\tau_{set}}\right)$	(A 7)	$D = \frac{F}{V}$	(A 8)

Cell model equations:

$\mu = K_{B1} \cdot \frac{W(t)}{K_S + W(t)} - K_{B2} \cdot m_{ATP}$	(A 9)	$r_P = K_{p1} \cdot \mu + K_{p2}$	(A 10)
$r_S = r_{S,max} \cdot \frac{S(t)}{K_S + S(t)}$	(A 11)	$W(t) = \int_{-\infty}^{t} \frac{(S(t-\tau)}{\beta^2}) \cdot \tau \cdot \exp(\frac{-\tau}{\beta}) \cdot d\tau$	(A 12)

Parameters and initial Values:

D,-, $(1/h)$; F,-, (g/l); S_{set},10, (g/l); K_{B1},0.1184, $(1/h)$; K_{B2},4.7376, (g/mol); K_{p1},0.48, $(-)$; K_{p2},0.0008, $(1/h)$; K_s,10, (g/l); m_{ATP},0.0015, $(mol/(g \cdot h))$; P,0, (mg/l); $r_{S,max}$,0.19, $(1/h)$; S,40, (g/l); S_F,1260, (g/l); t,-, h; V,15, l; W, $W_0 = S_0$, (g/l); X,1, (g/l); Z, $Z_0 = S_0$, (g/l); τ_{set},1, h; β,5, h; μ,-, $(1/h)$;

Table 4.3 shows model performance parameters (BIC and MSE) for the hybrid model structure (Fig. 4.4B) with varying number of hidden nodes (between 2 and 8) and different series of lagged input variables (between 0 and 4 with intervals of 2 or 2.5 hours). In general, the BIC values for the two series of lagged variables are much better than when no lagged variables are considered. This again confirms the advantages of the DDEHM methodology proposed herein for identifying delayed dynamics.

The effect of discrete time delays

It can be further noticed that BIC values of the time series with a time delay of 2.5(h) are slightly better when compared to the ones for time series with a 2(h) time delay. This observation agrees with the results of the previous case study where the performance of

the hybrid model peaked the closer the model delay was to the true delay. In this case the maximal weighted delay is $5(h)$. Moreover, the BIC values tend to improve with increasing number of lagged variables. The increasing number of input lagged variables, which are weighted by the neural network, seem to result in more accurate discrete time approximations of the continuous distributed time delays. In contrast, it can also be observed in Table 4.3 that with increasing number of delays the best BIC value is more likely to be found for a lower number of nodes in the hidden layer of the ANN. However, this was expected as the BIC is constrained by the number of model parameters. Nevertheless, the same observation is made for the MSE values. Furthermore, it was observed that significantly more random changes of the parameter values were required when the numbers of incorporated delays increased in order to achieve results which were coherent with the ones obtained for smaller numbers of delays.

Table 4.3 Results for Case Study II

Results of the performance measures, BIC and MSE, over structure parameters, namely Numbers of Nodes in the hidden layer of the ANN, NN, Number of time lags, Nlag and the time lag, τ, for *Pichia pastoris* cells with distributed time delays using the structure of Fig. 4B. Integration of material balances along with the equations obtained from the sensitivity method is carried out using the linear approximation integration schema described in the Methods section. The times series were chosen such that one of the delays matched the maximum of the time delay of the weighting function of the simulation case (see Eqs. A 12).

NN	Nlag	τ	BIC			MSE			NN	Nlag	τ	BIC			MSE		
			train	valid	test	train	valid	test				train	valid	test	train	valid	test
3	0	0	-18561	-5147	-5000	0.0155	0.0187	0.0157	6	0	0	-18500	-5199	-5030	0.0148	0.0187	0.0154
4	0	0	-18514	-5164	-4994	0.0155	0.0187	0.0150	7	0	0	-18576	-5197	-4990	0.0153	0.0180	0.0144
5	0	0	-18531	-5186	-5011	0.0151	0.0189	0.0153	8	0	0	-18430	-5211	-5053	0.0146	0.0182	0.0158
2	1	2	-17144	-4697	-4475	0.0086	0.0098	0.0065	2	1	2.5	-16343	-4293	-4331	0.0077	0.0076	0.0067
3	1	2	-16981	-4588	-4470	0.0085	0.0108	0.0067	3	1	2.5	-17230	-4725	-4584	0.0186	0.0327	0.0183
4	1	2	-16877	-4581	-4484	0.0087	0.0117	0.0072	4	1	2.5	-16635	-4506	-4690	0.0154	0.0220	0.0170
5	1	2	-16792	-4544	-4490	0.0078	0.0093	0.0064	5	1	2.5	-16927	-4614	-4524	0.0086	0.0111	0.0072
6	1	2	-16911	-4643	-4544	0.0087	0.0119	0.0073	6	1	2.5	-16944	-4523	-4488	0.0082	0.0092	0.0066
7	1	2	-17197	-4632	-4505	0.0097	0.0151	0.0084	7	1	2.5	-17044	-4742	-4608	0.0082	0.0107	0.0075
8	1	2	-17181	-4530	-4496	0.0123	0.0155	0.0105	8	1	2.5	-16976	-4694	-4608	0.0084	0.0110	0.0076
2	2	2	-16466	-4512	-4525	0.0067	0.0079	0.0067	2	2	2.5	-17813	-4833	-4759	0.0130	0.0152	0.0129
3	2	2	-16856	-4656	-4535	0.0081	0.0116	0.0073	3	2	2.5	-16637	-4684	-4595	0.0079	0.0123	0.0079
4	2	2	-16788	-4616	-4525	0.0079	0.0111	0.0072	4	2	2.5	-16703	-4507	-4517	0.0073	0.0083	0.0064

5	2	2	-16734	-4430	-4446	0.0075	0.0088	0.0061	5	2	2.5	-16384	-4327	-4404	0.0068	0.0074	0.0063
6	2	2	-16573	-4271	-4353	0.0077	0.0081	0.0065	6	2	2.5	-16601	-4400	-4432	0.0071	0.0079	0.0062
7	2	2	-16704	-4569	-4541	0.0071	0.0089	0.0065	7	2	2.5	-16569	-4405	-4466	0.0068	0.0072	0.0061
8	2	2	-16921	-4728	-4632	0.0082	0.0132	0.0079	8	2	2.5	-16833	-4790	-4664	0.0080	0.0127	0.0079
2	3	2	-19006	-5136	-5080	0.0181	0.0177	0.0158	2	3	2.5	-16619	-4566	-4514	0.0077	0.0099	0.0071
3	3	2	-16811	-4692	-4549	0.0078	0.0119	0.0073	3	3	2.5	-16037	-4218	-4259	0.0064	0.0072	0.0058
4	3	2	-16737	-4474	-4466	0.0069	0.0089	0.0063	4	3	2.5	-16439	-4224	-4287	0.0068	0.0078	0.0057
5	3	2	-16519	-4357	-4408	0.0066	0.0076	0.0058	5	3	2.5	-16199	-4358	-4295	0.0063	0.0076	0.0056
6	3	2	-16832	-4506	-4415	0.0090	0.0107	0.0078	6	3	2.5	-16604	-4577	-4556	0.0072	0.0094	0.0067
7	3	2	-16565	-4385	-4439	0.0066	0.0072	0.0058	7	3	2.5	-16344	-4475	-4432	0.0064	0.0078	0.0060
8	3	2	-16758	-4672	-4569	0.0079	0.0117	0.0069	8	3	2.5	-16471	-4374	-4505	0.0066	0.0069	0.0061
2	4	2	-16655	-4365	-4504	0.0071	0.0107	0.0077	2	4	2.5	-16562	-4532	-4503	0.0079	0.0097	0.0073
3	4	2	-16377	-4301	-4431	0.0067	0.0078	0.0064	3	4	2.5	-16325	-4471	-4470	0.0072	0.0086	0.0066
4	4	2	-16215	-4183	-4316	0.0062	0.0067	0.0057	4	4	2.5	-16261	-4189	-4281	0.0064	0.0068	0.0058
5	4	2	-16611	-4481	-4484	0.0070	0.0101	0.0066	5	4	2.5	-15954	-4190	-4255	0.0056	0.0060	0.0052
6	4	2	-17503	-4762	-4597	0.0189	0.0374	0.0197	6	4	2.5	-16439	-4334	-4476	0.0064	0.0073	0.0060
7	4	2	-25835	-7171	-7226	10.160	9.2368	11.843	7	4	2.5	-25949	-7288	-7280	13.781	19.040	17.824
8	4	2	-16256	-4328	-4369	0.0058	0.0061	0.0052	8	4	2.5	-16296	-4326	-4381	0.0061	0.0065	0.0054

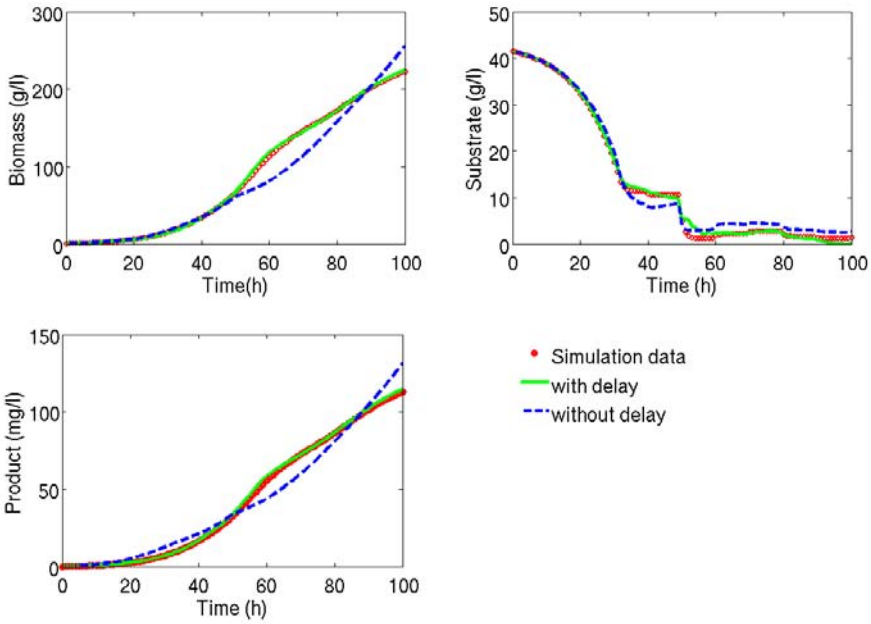

Figure 4.6 Qualitative results for trajectories of concentrations
Pichia Pastoris distributed delay modelling result for a fed-batch of the test data set: red circles are "measured" data over time; green line are the identification results by model structure (4.4B) with 5 hidden nodes and a series of 4 time lagged variables of 2.5 hours; blue dashed line are the identification results by model (4.4B) with 7 hidden nodes and no time delay.

Comparing standard and DDE hybrid models

Figure 4.6 shows non-noisy simulation data and the the best modelling results of hybrid models with and without delays for the concentrations of biomass, substrate and product in a fed-batch of the test set. In the figures of biomass and product concentrations, predictions of the hybrid models with delays are practically congruent to the true process behaviour. The intrinsic dynamics of the organism are perfectly met. In clear contrast are the predictions of the hybrid standard model without time delays. Biomass and product concentrations are under-predicted for a time span between 50 to 85h, then followed by over-prediction from 85h till the end. Before 50h only insignificant differences between predictions and data are visible. For the substrate concentration, the DDEHM model shows a significant amount of error in a short time window from about 48 to 55h, which is coincident to a fast decrease in the substrate concentration. It should

be noticed that such fast dynamics are rather challenging to integrate (see comments below). As for the standard hybrid model without delays, it predicts accurately substrate dynamics only at the beginning, i.e. from 0 to 30h. Thereafter, between 30 and 50h, the model under-predicts substrate concentration, and after 50h, it over-predicts the substrate concentration.

Note about numerical integration of DDEs

The integration of the hybrid model differential equations using the built-in MATLAB solvers (dde23, ode23) showed to be computationally intensive. A typical training run took 4 to 5 days. Moreover, convergence was sometimes not accomplished due to the limitation of the integration step size and accuracy. On the other hand, the integration of the hybrid model differential equations with the linear approximation only lead to small discrepancies for substrate concentrations if the step size was chosen adequately small, i.e. between 0.05-0.1h. Convergence was tested by decreasing even further the integration step without significant changes in the integration results. This approach lead to a reduction of computation of about 75% (i.e. 3 days) when compared to the MATLAB solvers.

4.4 Conclusions

Time-delays have a profound impact on cellular regulatory mechanisms. Therefore, their modelling is essential in metabolic engineering and process optimization studies. The detailed mechanisms behind observed time delays are often unknown. The required "omic" data for a fundamental mathematical modelling of such phenomena is generally unavailable at the required time resolution and accuracy. As a result, biochemical delayed dynamics are many times only "measurable" through their external consequences in terms of extracellular properties. We propose herein a hybrid semi-parametric modelling method to identify such delayed dynamics. The principle is probing from outside to understand the inner workings. The concept was applied to two illustrative case studies. The overall results show that significantly better prediction qualities of the novel hybrid model when compared to the traditional approach were obtained in all case studies, being the more distinctive the more significant the underlying system delay is. When system and model delay are identical the model quality peaked but even with a delay mismatch as high as 10% in the TF-A gene-regulatory network, modelling results were significantly enhanced in comparison to no delay at all. These results support a system delay identification strategy by studies of different discrete delays in the input variables. For the

studies on *Pichia pastoris* with intrinsic distributed time delays significant enhancements were introduced by the DDEHM model. This suggests that even though the proposed structure bases on discrete time delays directly of external excitation variables, it poses no limitation of applicability. In conclusion the method proposed herein is a powerful tool to identify time delays in ill-defined biochemical networks.

4.5 Acknowledgements

Sincere thanks for the financial support to the Fundacao para a Ciencia e a Tecnologia (scholarship reference no.: SFRH / BD / 36990 / 2007) and to Anthony Danko for his corrections on the manuscript.

4.6 Nomenclature

Abbreviations	
AIC	Akaike Information Criterion
ANN	Artificial Neural Network
AOX1	Alcohol OXidase gene 1
AR(X)	AutoRegressive (eXogenous)
BIC	Bayesian Information Criterion
DDE	Delay Differential Equation
DDEHM	Delay Differential Equation Hybrid Model
MSE	Mean Squared Error
ODE	Ordinary Differential Equation
RFDE	Retarded Functional Differential Equation
TF	Transcription factors
Mathematical Symbols	
b_1	Bias of the input layer
b_2	Bias of the hidden layer
b_{int}	Vector of transport fluxes
c	Vector of concentrations
c_{ext}	Concentrations of extracellular metabolites
c_{int}	Vector of intracellular metabolites
$c_{\sigma,j}$	Standard deviations
$c_{mes,l,i}$	Off-line measured concentration values
$g(\cdot)$	Transfer function of the hidden layer
$h(\cdot)$	Transfer function of the input layer
i	Counter
j	Counter
k_f	Cytosol synthesis rate
l	Counter
m	Number of intracellular metabolites
n	Number of components

o	Number of latent variables
q	Number of metabolic reactions
r	Vector of kinetic rate functions
r_{ext}	Kinetic rate vector
r_{int}	Vector of q kinetic rates
r_S	Specific methanol uptake rate
r_P	Specific product synthesis rate
s_j	Exogenous input j
t	Time
t_0	Initial time
u_{ext}	Vector of volumetric feeding rates
w	Vector of empirical parameters
w_1	Weights of the input layer
w_2	Weights of the hidden layer
x	TF-A monomeric concentration in the nucleus
y	Variables
D	Dilution rate
E	Weighted least squared criteria
F	Feeding rate
K_{ext}	Matrix of stoichiometric coefficients
K_{int}	Stoichiometric matrix of m metabolites and q metabolic reactions
M_j	Number of time delay associated with exogenous input s_j
N_i	Number of time delays associated with entry i of c
P	Number of samples
R_{bas}	Basal transcription rate
S	Substrate concentration
S_F	Substrate feeding rate concentration
S_{set}	Substrate set-point
V	Reactor volume
W	Delayed substrate concentration
X	Inputs to the nonparametric model / Biomass concentration
Z	Intermediated delayed substrate concentration
ψ	Mechanistic term
ρ	Unknown nonparametric terms
μ	Specific biomass growth rate
τ	Time delay
τ_i	Time delay associated with entry i of c
τ_j	Time delay associated with exogenous input s_j

Chapter 5

A hybrid modeling framework for PAT: Application to Bordetella pertussis cultures

5.1 Abstract

In the Process Analytical Technology (PAT) initiative the application of sensors technology and modeling methods are promoted. The emphasis is on Quality by Design (QbD), on-line monitoring and closed-loop control with the general aim of building in product quality into manufacturing operations. As a result, on-line high-throughput process analyzers find increasing application and therewith high amounts of highly correlated data becomes available on-line. In this study, an hybrid chemometric/mathematical modeling method is adopted for data analysis, which is shown to be advantageous over the commonly employed chemometric techniques in PAT applications. This methodology was applied to the analysis of process data of *Bordetella pertussis* cultivations, namely on-line data of Near-InfraRed (NIR), pH, temperature and dissolved oxygen, and off-line data of biomass, glutamate and lactate concentrations. The hybrid model structure consisted of macroscopic material balance equations in which the specific reactions rates are modeled by Nonlinear Partial Least Square. This methodology revealed a significant higher statistical confidence in comparison to Partial Least Squares (PLS), translated in a reduction of mean squared error of glutamate (~50%), lactate (~30%) and biomass (~5%) estimates. Moreover, the analysis of loadings and scores in the hybrid approach revealed that process features can, as for PLS, be extracted by the hybrid method.

5.2 Introduction

The Process Analytical Technology (PAT) initiative, published as a nonbinding guidance for industry by the U.S. Food and Drug Administration in 2004 (PAT, 2004), was recognized worldwide, because it offers the opportunity to cut down product trial time and thus costs. In PAT guidelines the use of system integrated approaches, throughout the different stages of product trial (from the development till the manufacture) is encouraged. The integration of different levels and sources of information requires a framework in which the integrated objects can be adequately linked in order to establish the desired synergy.
The idea is Quality by Design (QbD), which starts at the design stage of the manufacturing process but also addresses the need for improved on-line monitoring and control methods to maintain high product quality during manufacturing operations (Glassey *&* al., 2011). On the level of process development the intermeshing of process analyzers and adequate data evaluation tools is encouraged in PAT (PAT, 2004), in order to determine the process state at-time and to ultimately manipulate it. Many times the direct identification of the state is hindered by the fact that either the process key-variables are not at-time measurable, or (as undesirable from the process engineering perspective (Read *&* al., 2010b)) these measurements are invasive or destructive.

At-time knowledge about the key-variables can in principle be derived from non-invasive and nondestructive measurements of other quantities (Soons *&* al., 2008b). Devices fulfilling these requirements and that are able to provide information about the physiological state of cells, are for instance capacitance probes, InfraRed spectroscopy, Fluorescence Spectroscopy and others (Harms *&* al., 2002; Read *&* al., 2010b).

While it is many times difficult to calibrate the measured physiochemical properties to a meaningful process quantity, the huge amount of generated data poses an additional challenge. Solely for one spectroscopic device the dimensions might easily reach a number that is unfeasible to be analyzed without the support of very efficient mathematic tools. This is one of the reasons why Multivariate Data Analysis (MVDA) tools are frequently applied in the process analysis (Read *&* al., 2010b) and why they are expected to potentially play a central role in PAT (Glassey *&* al., 2011).

Multivariate regression, (nonlinear) partial least square, evolving factor analysis, support vector machines or principal component analysis are probably the ones which are most commonly applied and most successful (Schenk *&* al., 2007; van Sprang *&* al.,

2007). These methods are data driven and in most cases are applied on their own disregarding other valuable process knowledge. As recently highlighted in a review paper by (Glassey & al., 2011), the use of hybrid modeling tools that combine MVDA into a common (hybrid) modeling framework still presents a major challenge to the integration of different layers of information about cells and macroscopic processes.

Hybrid modeling that can link different types of process knowledge presents a suitable alternative to pure MVDA (Galvanauskas & al., 2004; Gnoth & al., 2008b; Oliveira, 2004; Psichogios and Ungar, 1992; Schubert & al., 1994a; Teixeira & al., 2007a; Thompson and Kramer, 1994). The linking of process information helps to understand the interplay between certain key quantities and it enhances the reliability of the process predictions. It is such an integrated systems framework where the "Process Understanding" and the "Principles and Tools", both defined in the PAT initiative (PAT, 2004), are brought together in order to manage the complexity while every time drawing a more complete process picture.

In principle, either parallel or serial hybrid topologies can be adopted. The latter is particularly suitable for complex systems where some internal mechanisms are poorly known, but for which large data sets are available without direct physical interpretation (Teixeira & al., 2007b). Since Artificial Neural Networks (ANNs), that are traditionally applied in serial hybrid structures (Galvanauskas & al., 2004; Gnoth & al., 2008b; Oliveira, 2004; Psichogios and Ungar, 1992; Schubert & al., 1994a; Thompson and Kramer, 1994), are unsuitable for knowledge extraction from large/highly-correlated data (Bishop, 1995) an alternative approach is applied, namely a (Nonlinear) Partial Least Square ((N)PLS) model.

In this study, the application of such a hybrid methodology is reported for the monitoring of target metabolites concentrations in a *Bordetella pertussis* cultivation, from online available Near InfraRed (NIR), pH, temperature and dissolved oxygen measurements, Fig. 5.1. This monitoring system provides critical online process knowledge that can be used for closed-loop control in order to maintain process quality or maximize its quantity. For comparison, the hybrid methodology is benchmarked against the standard chemometric tool, a static PLS method.

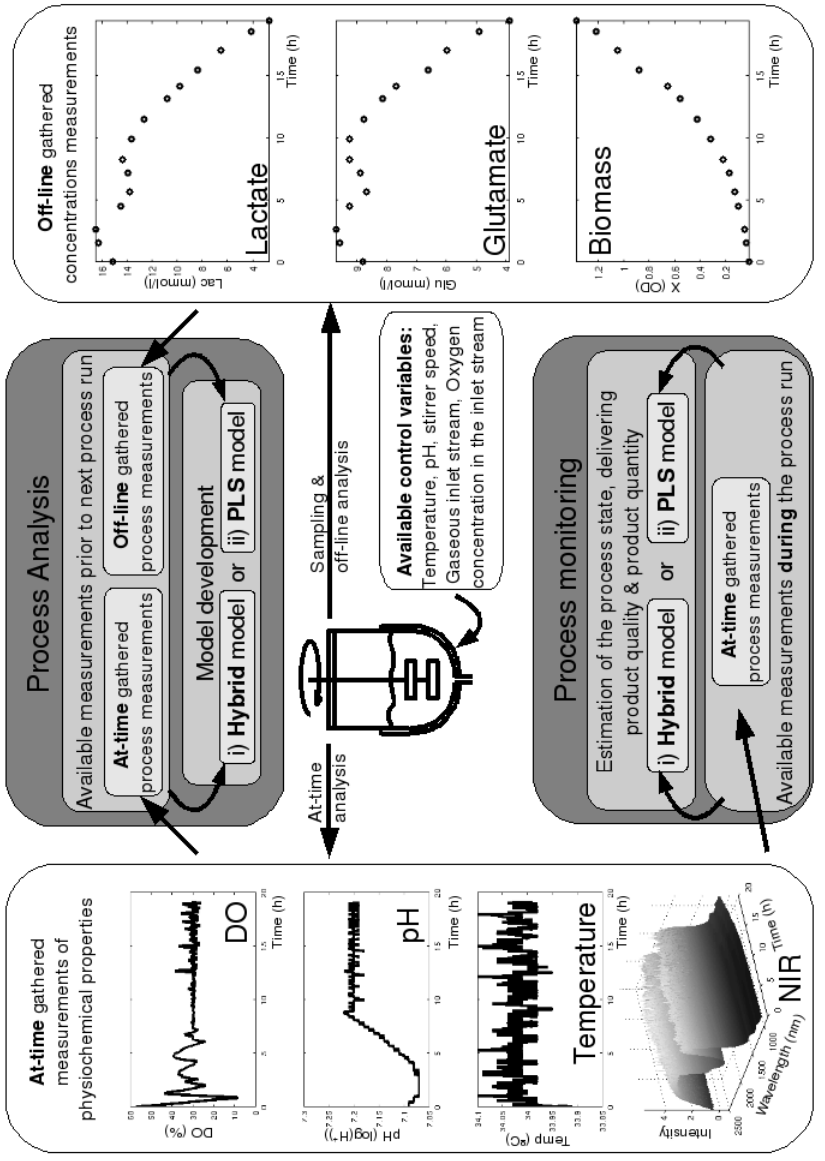

Figure 5.1 Schematic sketch of the present study

5.3 Materials and Method

5.3.1 The Process and Data

B. pertussis is cultivated for the production of a vaccine against whooping cough. Different cultivations strategies are reported (Licari & al., 1991; Rodriguez & al., 1994; Westdijk & al., 1997) which all seek to identify the optimal cultivations conditions ensuring vaccine quality and quantity. The key to ensure quality and quantity is the real-time control of biomass concentration and specific growth (Soons & al., 2008b, 2006). For the at-time identification of the biomass concentration and the specific growth during the process (Soons & al., 2008b) compared a methodology using a dissolved oxygen (DO) sensor to an approach which based on in situ NIR spectroscopy. The conclusion anent the NIR based model was however rather disillusioning, in the sense that the DO sensor based methodology in the situation of fixed path length and limited number of batches is to prefer (Soons & al., 2008b).

The number of samples along with robustness is a major concern for model calibration from NIR data, (ASTM, 2005; Brereton, 2000; Rhiel & al., 2002; Schenk & al., 2007). Process conditions and the component under study should be varied in such a way that a robust calibration model can be developed from the response in the spectra. While this is mostly a task for the experimental design prior to the experiment, it will be shown in the following that it is feasible to obtain enhanced prediction quality from the same data when incorporating additional information using hybrid modeling methods.

The experimental data of *B. pertussis* which find application in this study is the one reported by Soons & al. (2008b). The process was run in batch mode with the two main carbon sources for cell growth being glutamate and lactate. Variations to the process conditions were made as reported in Soons & al. (2008b), namely pH, temperature and DO varied considerably from 6.9-7.25 $\log(H^+)$, 33.8-34.1 ^{o}C and 0-100%, respectively.

The model input data are, as usually, auto-scaled, i.e. the inputs are shifted to zero mean and are scaled by the variance. Fluctuations of the wavelength intensities, i.e. noise, is one of the reported problems of NIR data. These fluctuations are especially problematic when for modeling purposes the event number of the spectral data is reduced to the time dimensions of the counterparts, which are usually infrequent off-line measured concentration data. Thus the NIR data were pretreated as in Soons & al. (2008b) by the application of a Savtisky-Golay smoothing with a 45-point window and

a second-order polynomial along the dimension of time.

In order to account for the natural deviations experienced during production runs in the calibration data, two sets of data (A and B) were designed. Each of these sets comprises 5 batches for the calibration and 2 batches for validation, wherein one of the batches, the one which exhibits a limitation in DO and a lower pH, was assigned in both sets for validation. A remaining batch, for which no substrate measurements were available, was applied in both sets for testing. Note that the training data span the space of process operating conditions in which the model will reliably work while the validation data are a measure of the performance of the spanned space, i.e. natural deviations of the process should be reflected in both.

5.3.2 Partial Least Square / Projection to Latent Structures (PLS)

PLS is commonly applied to correlate spectroscopic data to chemical compound concentration data (Wold & al., 2001). The correlation is established through maximization of the data covariances. The fundamentals for this maximization are provided by the model structure of PLS. Therein, in the so called "outer model", the matrix of input values and the matrix of output values are decomposed into loading matrices, score matrices and matrices of residua. Through the "inner" model, which is also referred to as "latent structure", the score matrices are then linked, see (Brereton, 2000; Wold & al., 2001). In the perspective of statistical process monitoring, quality prediction and fault diagnosis, these latent structure is of special interest since it can reveal important process information (MacGregor and Kourti, 1995; Read & al., 2010b; Undey & al., 2003).

Two settings of PLS model inputs are investigated in this study. Setting (a) comprises the on-line measured data of pH, temperature, percentage of dissolved oxygen and the complete wavelength spectra (833-2400 nm). Setting (b) contains the on-line measured data of pH, temperature, percentage of dissolved oxygen and a selection of wavelength of the spectra (1111-1397 and 1587-1852 nm). This wavelength selection originates from van Sprang & al. (2007), was applied in Soons & al. (2008b) and is chosen in this study because the excluded wavelengths correspond to saturated intensities due to water. Both settings, (a) and (b), are augmented by the initial components concentration values of every batch, in order to compare the PLS model to the hybrid model providing the exact same data. The PLS model outputs comprise the concentrations of lactate, glutamate

and biomass. This PLS structure implies that the identification of the correlations between the inputs and the outputs can only be established when measurements of the input as well as of the output are available for the same time instant, implying that the high number of sampled input data is significantly reduced namely to the sampling rate of the concentration data.

5.3.3 The Hybrid (Nonlinear)PLS model

The adopted serial semi-parametric hybrid model structure is schematized in Fig. 5.2. This structure consists of two major modules, namely a module assigned to the macroscopic material balances and another assigned to the biological fluxes. The formulation of material balances is straightforward, yet the balances take a central role, since (i) they build the frame of the system and (ii) they link the macroscopic reactor system to the microscopic cell factory. The material balances written for a batch reactor in the state space form are

$$\frac{dc}{dt} = r \tag{5.1}$$

Therein c is a vector of concentrations (in the present case comprising the concentrations of Lactate, Glutamate and Biomass, $c = [Lac, Glu, X]^T$) and r is the vector of kinetic rate functions.

The vector of kinetic rates is the link to the biological system, see Fig. 5.2, and it describes the rate of consumption or production of the particular compound. In the displayed hybrid model the biological system is mimicked by a semi-parametric model proposed by Oliveira (2004). For the present case, a set of kinetic constraints are assumed *a priori*, namely that (i) the substrate uptake is zero if substrate depletes and (ii) the uptake rates are proportional to the concentration of biomass, and thus the semi-parametric model reads as:

$$r = \underbrace{\begin{bmatrix} Lac \cdot X & 0 & 0 \\ 0 & Glu \cdot X & 0 \\ 0 & 0 & X \end{bmatrix}}_{\phi} \cdot \underbrace{\begin{bmatrix} r_{Lac} \\ r_{Glu} \\ \mu \end{bmatrix}}_{\rho}. \tag{5.2}$$

where ϕ comprises the *a priori* knowledge about the kinetics and ρ is a vector that comprises the unknown kinetic rates, i.e. the specific reaction rates. These rates are

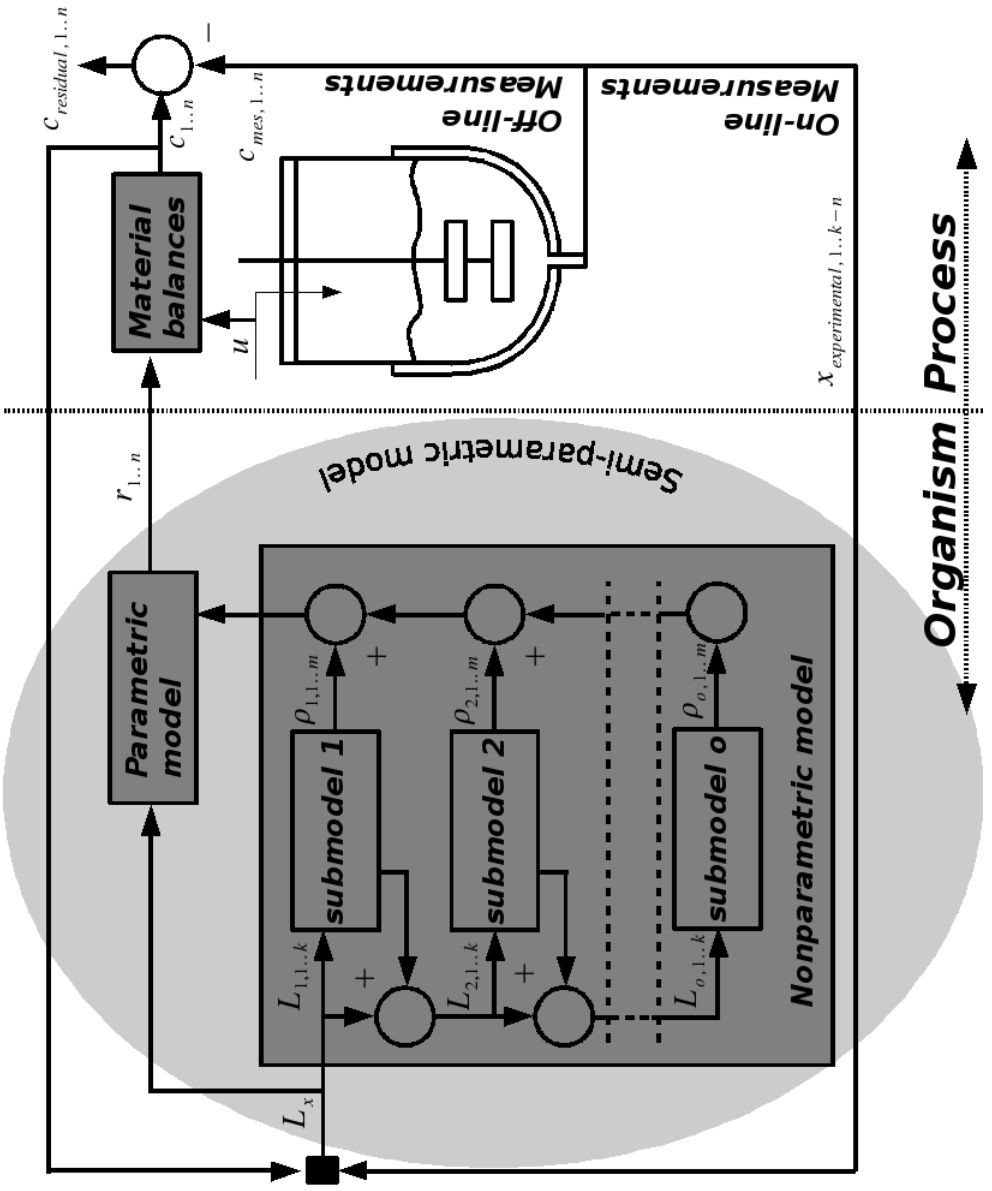

Figure 5.2 A schematic overview of the general, serial, semi-parametric hybrid model structure. Variables and abbreviations are according to the text.

functions of the inputs, L_x, and some parameters, w_A, i.e.: $\rho = \rho(L_x, w_A)$. The vector of inputs, L_x, may in general comprise (i) the concentrations, c, and/or (ii) at-time measurements $X_{experimental,1..k-n}$, see Fig. 5.2. Thus the rates $\rho = [r_{Lac}, r_{Glu}, \mu]^T$ might not only depend on the presently modeled concentrations, but also (i) on the physiochemical properties, such as pH or temperature and (ii) on the concentrations of metabolites which are not comprised by the model but whose traces are for instance contained in measured spectra. In the present case only at-time measurements are comprised in the inputs, i.e. settings (a) and (b) are used as defined above in the section on PLS. Due to the high numbers and the nature of the information comprised by L_x, the adoption of Artificial Neural Networks is infeasible, as (i) this would lead into a highly underdetermined system of equations and (ii) these ANNs are unsuitable for knowledge extraction from large/highly-correlated data (Bishop, 1995). Instead, a NPLS alike nonparametric model is adopted, (von Stosch & al., 2011b). The nonparametric model, as illustrated in Fig. 5.2, consists of o independent submodels, i.e.:

$$\rho_{1..m}(L_x, w_A) = \sum_{i=1}^{o} \rho_{i,1..m}(L_{i,1..k}, w_A). \tag{5.3}$$

Each submodel $\rho_{i,1..m}(L_{i,1..k}, w_A)$ can further be decomposed into an outer and an inner model. The outer model reduces the high dimension of the inputs

$$L_{i,1..k} = W_{x,i} \cdot t_i, \tag{5.4}$$

by the application of input loadings, $W_{x,i}$, to the input latent variable t_i and decompresses the output latent variable u_i through the application of output loadings, $W_{y,i}$, to:

$$\rho_{i,1..m} = W_{y,i} \cdot u_i. \tag{5.5}$$

The inner model links (non)linearly the input latent variable t_i with the output latent variable u_i, i.e. in this study a ANN representation is chosen:

$$u_i = w_{2,1} \cdot g(w_{1,i} \cdot h(t_i) + b_{1,i}) + b_{2,i}, \tag{5.6}$$

where $w_{1,i}$ and $w_{2,i}$ are weights, $b_{1,i}$ and $b_{2,i}$ are biases and $h(\cdot)$ and $g(\cdot)$ are transfer functions, which are in this study linear and hyperbolic tangential, respectively

($h(t_i) = t_i$; $g(x) = \tanh(x)$). The number of nodes in the hidden layer of the ANN are fixed in this study to be one, as they are shown to have only little influence on the quality of the estimates (Baffi & al., 1999, 2000; Qin and McAvoy, 1996). In this context it should be mentioned that the term "inner model" is also referred to as "latent variable model", where many times (as in the following) the term model is dropped and thus the expression relaxed to "latent variable".

The latent variables, t_i and u_i, comprise condensed information about the process state, wherefore they pose, as in the case of PLS, a valuable source of information about the process state and can e.g. be used for statistical process monitoring, quality prediction and fault diagnosis (Undey & al., 2003).

The parameters w_A which for latent variables $i = 1, .., o$ comprise the ANN parameters (i.e. the weights $w_{1,i}$, $w_{2,i}$ and biases $b_{1,i}$, $b_{2,i}$) and the input and output loadings, $W_{x,i}$ and $W_{y,i}$, respectively. Their identification can in principle be accomplished in two manners: (i) by estimation of the kinetic rates through the differentiation of c with respect to the time and the subsequent application of e.g. the NIPALS algorithm (Henneke & al., 2005) or (ii) by the sensitivity equation technique (von Stosch & al., 2011b). The sensitivity equation technique in the context of fluctuating or sparse or noisy concentration data definitely is to prefer (Oliveira, 2004; Psichogios and Ungar, 1992; Schubert & al., 1994a) and was therefore adopted.

The sensitivities equations have to be integrated along with the reactor material balances, wherefore a time inexpensive Euler integration scheme was applied. It is convenient to fit the time-steps of this scheme to the sampling rate of the on-line measurements, e.g. spectral measurements, in order to circumvent the interpolation between those.

5.3.4 Model Assessment Criteria

Model assessment criteria are required in order to assess the model performances objectively and in order to select an appropriate number of latent variables. For the latter cross-validation is applied, i.e. (i) in the case of PLS the number of latent variables is increased till the desired level of sophistication is reached, i.e. the best number of latent variables, is selected according to the lowest Mean Square Error (MSE) calculated for the validation data; and (ii) in the case of the hybrid approach the number of latent variables needs to be determined *a priori*, wherefore an heuristic search of numbers of latent variables that produce the best performing hybrid model in terms of the Bayesian

Information Criteria (BIC) value (defined below) obtained for the validation data, is performed.

The Mean Square Error is a qualitative measure of the model performance. Its calculation bases on the number of samples and the distance between the prediction and the measured data value:

$$MSE = \frac{1}{P \cdot n} \cdot \sum_{j=1}^{P} \sum^{n} \frac{(c_{j,mes}(t) - c_j(t, w_A))^2}{\sigma_j^2}, \tag{5.7}$$

where P signifies the number of samples, $c_{1\ldots n,mes}$ are the n off-line measured concentration values and $\sigma_{1\ldots n}$ are the standard deviations of the measured concentrations. However a criterion for model selection should not only base on the quality of the model estimates. It should also account for both the complexity of the structure in form of the number of parameters and the number of measured events.

Two criteria regarding these requirements find wide application, namely the Akaike Information Criteria (AIC) and the Bayesian Information Criteria, (BIC). In the context of the processes addressed the BIC is reported to be more appropriate (Burnham and Anderson, 2004; Peres & al., 2008), and therefore is adopted in this work for model comparison. The Bayesian Information Criteria, (BIC), is defined as:

$$BIC = \left(-\frac{n \cdot P}{2} \cdot \ln \left(\sum_{j=1}^{P} \sum^{n} \frac{[c_{j,mes}(t) - c_j(t, w_A)]^2}{\sigma_j^2} \right) \right) - \left(\frac{n_w}{2} \cdot \ln(\frac{n \cdot P}{2\pi}) \right) \tag{5.8}$$

where the term in the first bracket is the logarithmic maximum-likelihood and n_w is the total number of parameters/weights. In the sense of the BIC the model to prefer is the one that exhibits the larger BIC value for the validation set.

5.4 Results and Discussion

5.4.1 Comparing PLS and hybrid modeling

An overview of the best model performances in terms of MSE and BIC is compiled in Table 5.1.

The BIC values obtained by the PLS models for the validation and test data are therein found to be disproportionally high in comparison to the ones obtained for the hybrid

models. This significant difference is due to the much higher number of parameters in the PLS models, which is indicated by the respective higher number of latent variables. For the calculation of the BIC a model with a higher number of parameters is penalized, as this indicates a model structure which is more complex and less robust. Especially for control purposes, model robustness is important, since uncertainty and model-plant mismatch compromise the controller performance. Model robustness can be addressed through the statistical confidence of the estimates and thus the BIC is the measure of such. Therefore it is concluded that all PLS models presented in Table 5.1 have a lower statistical confidence than the hybrid models presented.

Table 5.1 Model performance criteria, the Bayesian Information Criteria (BIC) & the Mean Square Error (MSE) for training, validation & test data over model types, model inputs (see section 5.3.2), data sets (see section 5.3.1) and the number of latent variables.

Model		Data		BIC			MSE		
Type	Input setting	Set	lv[1]	training	validation	test	training	validation	test
HYB[2]	(a)	A	1	-369	-89	10	0.1596	0.0982	0.0136
HYB[2]	(a)	A	2	-330	-83	-1	0.1025	0.0754	0.0487
HYB[2]	(a)	A	3	-463	-92	-3	0.3461	0.0810	0.0560
HYB[2]	(b)	A	2	-321	-77	2	0.0941	0.0648	0.0341
HYB[2]	(b)	A	3	-312	-69	-3	0.0801	0.0466	0.0557
PLS	(a)	A	3	-11612	-8374	-3092	0.3223	0.4852	0.0070
PLS	(a)	A	7	-26363	-19287	-7253	0.1307	0.2147	0.0379
PLS	(b)	A	3	-4341	-3018	-1064	0.3596	0.6556	0.0134
PLS	(b)	A	7	-11828	-8776	-4304	0.1979	0.2416	0.0458
HYB[2]	(a)	B	1	-366	-99	10	0.1658	0.1103	0.0127
HYB[2]	(a)	B	2	-488	-99	4	0.5122	0.0971	0.0245
HYB[2]	(a)	B	3	-498	-92	-4	0.5281	0.0733	0.0616
HYB[2]	(b)	B	2	-336	-71	1	0.1150	0.0509	0.0389
HYB[2]	(b)	B	3	-311	-83	-4	0.0842	0.0598	0.0660
PLS	(a)	B	3	-11575	-8476	-3095	0.3642	0.3804	0.0104
PLS	(a)	B	7	-26261	-19521	-7257	0.1354	0.1403	0.0590
PLS	(b)	B	5	-6879	-4952	-1785	0.3595	0.3412	0.0204
PLS	(b)	B	7	-9384	-6810	-2511	0.2159	0.0870	0.0534

lv[1]: number of latent variables
HYB[2]: Hybrid Model

The analysis of performance in terms of MSE point in the direction just stated, i.e. the results in Table 5.1 show that the PLS model performance with the best number of latent variables is worse than that observed for hybrid models.

From another point of view, but also supporting the statistical confidence results, the PLS models exhibit performance inconsistencies between themselves, in that the best model structure on the basis of validation data is significantly different from the best structure that would be obtained on the basis of the test data (discussed below).

Finally, comparing results with different model input settings ((a) and (b)), it is seen that, as expected, excluding the NIR wavelength with saturated intensities due to water,

(cases (b)) lead to an increase in performance of all hybrid models.

5.4.2 Analysis of model structural differences

The observed discrepancy in the MSE performance raises the question about the possible structural reasons for the better hybrid model performance in comparison to the PLS models.

One main structural difference arises from the nature of the models - the input spectral information is linearly correlated to the concentrations in the PLS model whereas it is correlated to the kinetic rates in the hybrid model. This issue was subject of analysis, where it was concluded that, in the case of PLS, the estimation quality of the kinetic rates is poor for two main reasons: (1) the calculation of the kinetic rates is prone to error (Oliveira, 2004; Psichogios and Ungar, 1992; Schubert & al., 1994a) and consequently the identification of the correlation becomes more difficult; and (2) the correlation between the spectral intensities and the rates is most probably nonlinear.

A further issue analyzed was the effect that the noise in the input NIR intensities had on estimates of both hybrid and PLS model state variables and parameters.

Fluctuations in the hybrid model estimates are less distinct due to two main reasons: (1) In the serial hybrid modeling framework the estimated kinetic rates are integrated, this leading to a smoothening effect to the noise in the kinetic rate estimates; (2) the application of the sensitivities approach for parameter identification enables the utilization of input data at each integration time step, as such diminishing the impact of punctual fluctuations on the kinetic rate estimations. This is in contrast to the standard PLS, because for PLS, the identification of the correlations can only be established when both input and output data exist for the same instance of time, see section 5.3.2. This huge difference in the number of input data, that are used for the parameter identification, is exemplary shown in Fig. 5.3. Therein it can be seen that in the case of PLS the number of available 89 inputs samples decreases to 15, namely to the time instances for which both, input and output samples exist. Note that the number of output samples for both models, the hybrid and the PLS, is exactly the same.

5.4.3 Effect of latent variables

The observation, in this study, of a relatively high number of latent variables (mostly seven) in the case of PLS models, is in agreement with Soons & al. (2008b) and van Sprang & al. (2007). An additional Principal Component Analysis (PCA) on the in-

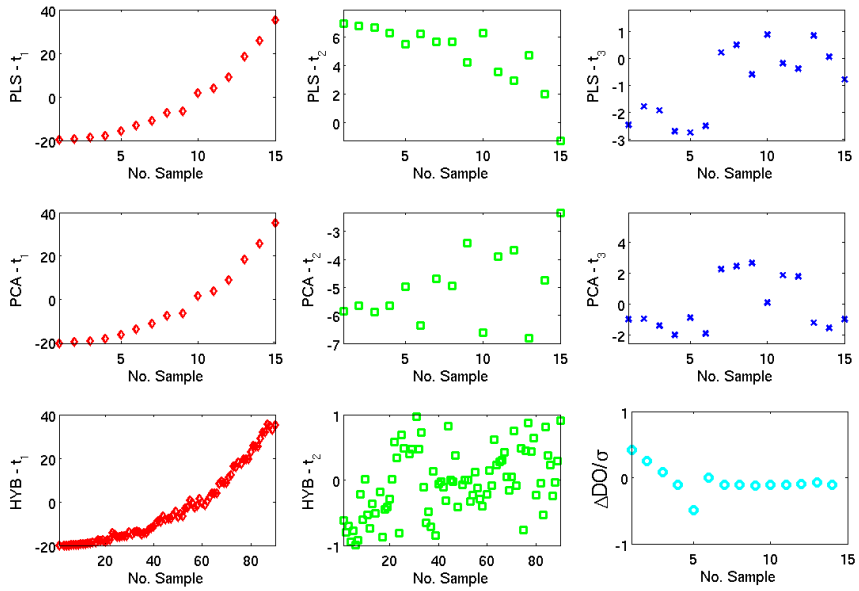

Figure 5.3 Input latent variables scores obtained for a validation batch of set A with inputs (b) from: the best Partial Least Square model (PLS t_1 - red diamond, PLS t_2 - green square and PLS t_3 - blue cross); Principal Component Analysis (PCA t_1 - red diamond, PCA t_2 - green square and PCA t_3 - blue cross); and the two latent variable hybrid model (HYB t_1 - red diamond, HYB t_2 - green square); and additionally auto-scaled dissolved oxygen measurements ($\Delta DO/\sigma$ - turquoise circles).

puts L_x revealed however that the variance in the inputs can be captured by only three latent variables (˜97.9% of the variance explained). Furthermore, it is observed that the input variance captured by PLS with three latent variables was ˜97%, whereas the corresponding captured output variance was ˜82%. This low number of latent variables was also observed for the PLS model performance with the test data, in terms of MSE (see Table 5.1).

The ˜97% of input variance captured in L_x with three latent variables, gives rise to another question, namely what these latent variables are due to and whether the correlations between inputs and outputs is biunique. It can be seen in Fig. 5.3 that the trajectory of the first input latent variable score t_1 of the PCA, PLS and the hybrid models are almost identical, and that, further, their shape is similar to the trajectories

of lactate, glutamate and especially biomass, which are shown in Fig. 5.4. The shapes of the trajectory of the second input latent variable, t_2, are observed to be influenced by the DO measurements, as: (1) the respective input loading value is usually high; and (2) the characteristics of the trajectories partially coincide, see Fig. 5.3. The third input latent variables scores, t_3, only shown in Fig. 5.3 for PCA and PLS, can however not be directly related to any specific input quantity. The observation that the shapes of

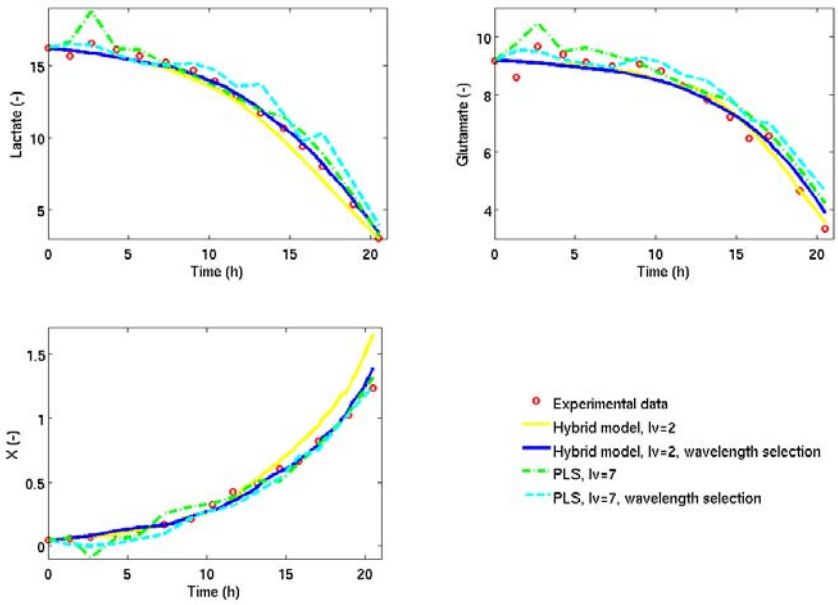

Figure 5.4 Concentrations of Lactate, Glutamate and Biomass over time for a validation batch of set A:, Experimental data - are red circles; Estimates from hybrid model, two latent variables, inputs (a) - continuous yellow line; Estimates from PLS model, seven latent variables, inputs (a) - dashed-dotted green line; Estimates from hybrid model, two latent variables, inputs (b) - continuous blue line; and Estimates from PLS, seven latent variables, inputs (b) - dashed turquoise line.

the trajectories of t_1 are similar to the ones of the concentrations is very interesting in connection to the observation that the spectral intensities increase towards the end of each batch. This points at a unique correlation of biomass and the spectral intensities. Therefore the correlations for glutamate and lactate concentrations would represent a stoichiometric relation to the biomass concentration. The observation further scrutinizes the wavelength selections for lactate, glutamate and biomass made by van Sprang & al. (2007), and questions the reason or need of any number of latent variables that

is greater than one.

In the case of the hybrid models, a number of two latent variables is justifiable on the basis of the observation that the second latent variable scores can be linked to the DO measurements. In contrast to the PLS models, the hybrid model therefore seems to profit from the known relation between dissolved oxygen and biomass production (Soons & al., 2006). Variations in the intensities of the NIR spectra which are due to other properties than concentrations, i.e. pH, temperature, DO and so on, do not seem to strongly effect the best hybrid model identification.

5.4.4 Qualitative Analysis of the Performance

A final qualitative analysis is presented based on the comparison of the individual prediction errors for lactate, glutamate and biomass, presented in Table 5.2, and the results presented in Fig. 5.4. Globally, it can be observed that the differences in results between PLS and hybrid models are the most significant for the lactate and glutamate concentrations, as (i) the respective prediction errors (Table 5.2) of both substrates, lactate and glutamate, obtained for the hybrid model cases are in general improved when compared to the ones obtained for PLS; and (ii) these improvements are also visually observable, Fig. 5.4. In the case of the biomass concentrations no such general trend is observable, however the partially worse performances for biomass in the hybrid model cases (Table 5.2) might be explained through the compromise between overall performance and individual performance.

The expected increase in performance, when excluding the wavelengths from the inputs whose intensities are saturated, due to water, can be observed for the estimates of lactate, glutamate and biomass in cases of both hybrid and PLS models (Table 5.2).

As can be exemplary seen in Fig. 5.4, the different hybrid models structures tested in general lead to smoother estimates than those obtained by corresponding PLS models; The relatively worse quality of the biomass estimates at the beginning of the batches, obtained from PLS models, may be explained with the relatively low NIR spectral intensities due to low biomass concentrations at the beginning of the batch (low signal to noise ratio).

Table 5.2 Individual prediction errors in form of MSEs (eq. (5.7) in which the standard deviation term is dropped) for lactate, glutamate and biomass concentrations obtained for training, validation & test data over model types, data sets (see section 5.3.1), model inputs (see section 5.3.2) and latent variables.

Model		Data		MSE								
Type	Input setting	Set	lv[1]	Lactate			Glutamate			Biomass		
				Train	Valid	Test	Train	Valid	Test	Train	Valid	Test
				(mmol²/l²)			(mmol²/l²)			(OD²)		
HYB[2]	(a)	A	2	0.7597	0.3044	-	0.2715	0.1929	-	0.0016	0.0247	0.0195
HYB[2]	(b)	A	2	0.5734	0.1304	-	0.3007	0.1201	-	0.0018	0.0420	0.0136
PLS	(a)	A	7	1.1267	1.6728	-	0.3193	0.6104	-	0.0049	0.0065	0.0172
PLS	(b)	A	7	1.5852	4.0445	-	0.4286	1.0447	-	0.0065	0.0269	0.0308
HYB[2]	(a)	B	2	4.8814	0.4231	-	0.7866	0.1443	-	0.0491	0.0537	0.0099
HYB[2]	(b)	B	2	0.7568	0.5169	-	0.3669	0.0851	-	0.0027	0.0008	0.0159
PLS	(a)	B	7	1.0978	1.5008	-	0.3888	0.2511	-	0.0056	0.0041	0.0273
PLS	(b)	B	5	3.2456	3.2692	-	0.8481	0.7145	-	0.0222	0.0217	0.0094
PLS	(b)	B	7	1.8357	1.0174	-	0.5818	0.1228	-	0.0090	0.0017	0.0247

lv[1]: number of latent variables
HYB[2]: Hybrid Model

The hybrid model does not suffer from this type of effect since more data are incorporated during the parameter identification and therefore fluctuations are damped. This is an important feature, since for instance for the control of specific biomass growth rate, a reliable estimation of it is required. In case that this estimate would be derived from fluctuating state estimates it is prone to error. In the hybrid model case the specific growth rate estimate is (i) directly accessible, as a result of the chosen structure; and (ii) in comparison to PLS less noisy, wherefore enhanced control performance would be enabled.

The slight overestimation for biomass by the hybrid model with input setting (a) at about 15 hours, which remains till the end of the batch, can be explained by error propagation.

5.4.5 Extracting process knowledge from latent variables scores

The structure of the NPLS submodel, described by eqs. (5.3) - (5.6), is similar to the structure of PLS models. For such PLS models the analysis of the input scores represents a relevant source of information concerning characteristics and features of the processes (MacGregor and Kourti, 1995; Read & al., 2010b; Undey & al., 2003). This important PLS feature is present in the applied hybrid NPLS models. Figure 5.5 shows the scores plot of the final hybrid model. By analyzing the relative position of the scores to each other as a measure of their similarity, one can extract important process information, on-line. For instance it is possible to detect outlying data samples,

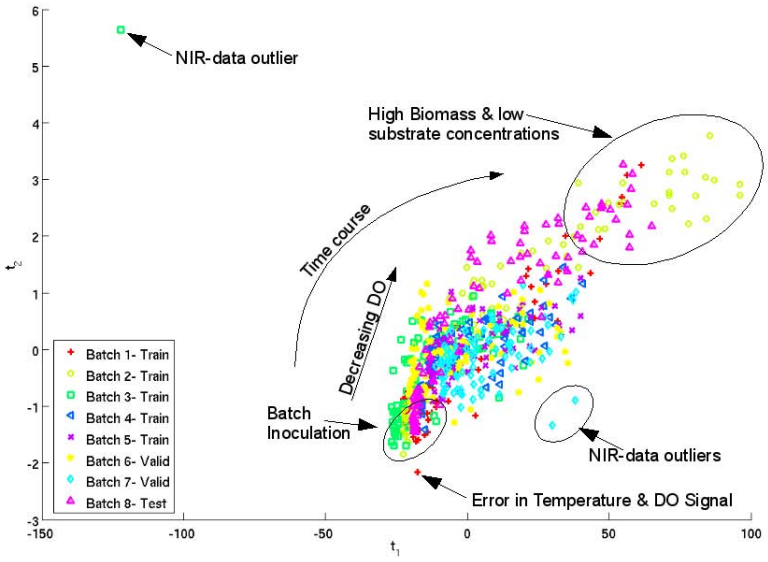

Figure 5.5 Input latent variables scores obtained for set A with inputs (b) by the hybrid model that comprises 2 latent variables.

which enables automatic fault detection. In the present case it was possible to pinpoint in a single plot NIR, temperature and DO data outliers. The latent variables time trajectories carry information about distinct process phases and also batch-to-batch variability. Certain process regions can be classified, i.e. a region of inoculation and a region of high biomass and low substrate concentrations. Thus from the time-course of the latent scores of complex spectral data for a certain batch, conclusions anent its performance can be effectively extracted using the hybrid modeling approach.

5.5 Conclusions

In what is called the PAT initiative (PAT, 2004), the Food and Drug Administration proposes an integrated systems approach. On the process level, the intermeshing use of process analyzers and adequate tools for the incoming data evaluation is recommended, in order to accurately determine the process state at-time and to ultimately manipulate

it.

Hybrid modeling provides an integrated systems approach whereby different sources of knowledge can be linked. When applied to chemical or biochemical processes, such a framework can be build on material balances wherein the specific reaction rates are modeled through the combination of both fundamentals and models, typically adopted in PAT, such as (N)PLS.

This methodology was applied to process data of a *B. pertussis* cultivation, in order to correlate on-line NIR, pH, temperature and dissolved oxygen measurements to off-line biomass, glutamate and lactate concentration measurements. Thus, during the process, the state identification would be feasible by using only the at-time available measurements. Benchmarking is provided by the classical PLS methodology.
The following was observed and can be stated:

(i) Results revealed that the statistical confidence in terms of the BIC of the hybrid method in comparison to the PLS method improved by several orders of magnitude (from ~ (-1000) to ~ (-10)), an evidence that was supported by the analysis of concentration trajectories, as shown in Fig. 5.4.

(ii) The higher statistical confidence traces back not only to a significantly lower number of latent variables (from 7 to 2), but also to enhanced quality of estimates, which is observed in the form of lower overall and individual mean square errors.

(iii) The lower number of latent variables results from the fact that the scheme proposed could incorporate the existing correlations between main state variables (in this case dissolved oxygen and specific growth), thus lowering dimensionality.

(iv) The improved quality of the state estimates was essentially the result of two factors: (1) the smoothing effect that the integration procedure had on noise contained in the kinetic rates estimates and (2) the incorporation of a wider range of input data, viz at each integration step, which was feasible only due to the applied parameter identification procedure.

(v) The extraction of valuable process information from the analysis of the latent scores is enabled, equivalently to the case of PLS.

All in all, the better performance of the hybrid model is worth the higher computational load, in comparison to the PLS method. Further, it provides a more consistent

interpretation of the process data in terms of fundamental mechanisms, thus enhancing the level of sophistication of knowledge generated. Finally, as a result of applying this hybrid structure, the trajectories of the estimated fluxes are directly accessible on-line, which allows for their control.

5.6 Acknowledgment

Sincere thanks for the provided data go to Mathieu Streefland, Zita I.T.A. Soons and the Netherlands Vaccine Institute and for financial support to the Fundação para a Ciência e a Tecnologia (reference scholarship no.: SFRH / BD / 36990 / 2007).

5.7 Nomenclature

Abbreviations	
AIC	Akaike Information Criterion
ANN	Artificial Neural Network
BIC	Bayesian Information Criterion
DO	Dissolved Oxygen
FIR	Finite Impulse Response
HYB	Hybrid Model
MSE	Mean Squared Error
MVDA	Multivariate Data Analysis
NIR	Near-InfraRed
NPLS	Nonlinear-PLS / Neural Network-PLS
PAT	Process Analytical Technology
PCA	Principal Component Analysis
PLS	Partial Least Squares / Projection to Latent Structures
QbD	Quality by Design
Mathematical Symbols	
$b_{1,i}$	Bias of the input layer in the latent variable submodel i
$b_{2,i}$	Bias of the hidden layer in the latent variable submodel i
c	Vector of concentrations
$c_{\sigma,j}$	Standard deviations
$c_{j,mes}$	Off-line measured concentration values
$g(\cdot)$	Transfer function of the hidden layer in the latent variable submodel i
$h(\cdot)$	Transfer function of the input layer in the latent variable submodel i
i	Counter
j	Counter
k	Number of inputs
n	Number of components
o	Number of latent variables
r	Vector of kinetic rate functions
r_{Glu}	Specific Glutamate uptake rate
r_{Lac}	Specific Lactate uptake rate

t	Time
t_i	Input latent variable
u_i	Output latent variable
w	Weights of all $i = 1..o$ ANNs
w_A	Vector of parameters
$w_{1,i}$	Weights of the input layer in the latent variable submodel i
$w_{2,i}$	Weights of the hidden layer in the latent variable submodel i
Glu	Glutamate concentration
$L_{i,1..k}$	Inputs 1 to k for latent variable submodel i
L_x	Vector of nonparametric model inputs
Lac	Lactate concentration
P	Number of samples
$W_{x,i}$	Input loadings of latent variable i
$W_{y,i}$	Output loadings of latent variable i
X	Biomass concentration
$X_{experimental,1..k-n}$	At-time measurements
ϕ	Known kinetic functions
ρ / $\rho_{1..o,1..m}$	Vector of unknown kinetic functions
μ	Specific biomass growth rate

Chapter 6

A general hybrid semi-parametric controller

6.1 Abstract

A general hybrid semi-parametric process control framework is proposed in this study. It is a general framework, because the presented control methodology can be customized for each application and hybrid semi-parametric because nonparametric and parametric models, founded on different knowledge sources, are merged into the control framework.

Tuning of the controller parameters is achieved through the application of a hybrid process model. Two identification schemata are presented, (i) an off-line process data-based scheme for simultaneous identification of the hybrid process model and the controller parameters; and (ii) an off-line identification scheme that can improve the controller characteristics beyond the characteristics learned from process data.

Several combinations of hybrid control schemata are applied to a bioprocess control problem, i.e. closing the loop for the control of the biomass concentration through manipulation of the substrate feeding rate and improving the control of the dissolved oxygen concentration through the stirring velocity. The results demonstrate that (i) due to the hybrid approach the control loop can be closed without any additional experiments; (ii) incorporation of different types of knowledge can enhance the controller performance, when compared to schema without integrated knowledge; (iii) knowledge incorporation facilitates the tuning of the controller; and (iv) the control action can be analyzed in relation to structural information incorporated into the hybrid controller.

6.2 Introduction

Proportional-Integral-Differential (PID) control is a widespread control strategy which is mainly due to its simplicity and/or its many times sufficient performance. There are, however, certain problems in which standard PID control has a limited performance (Astrom and Hagglund, 2001), two of which, namely process non-linearities and varying process conditions, are frequent in bioprocess applications. In order to enhance the performance in those two problems, the concept of PID control has been extended by using either gain-scheduling (Akesson & al., 2001; Kuprijanov & al., 2009; Levisauskas & al., 1996), on-line parameter adaptation (Chang & al., 2002; Chen and Huang, 2004; Kansha & al., 2008), or nonlinear PID approaches (Chang & al., 2003; Shu and Pi, 2000; Yeo and Kwon, 1999). The source of limitation might however be the rather simple structure of the controller, especially with respect to the first two named approaches.

Neural Networks can be seen as an universal extension in structural terms, constituting General Linear Controllers as a special case. There are numerous applications and manifold ways to apply Neural Networks to control (Azlan Hussain, 1999; Hagan and Demuth, 1999). Evidentially those methodologies are most times more advanced than PID control, which brings along different problems, namely that the controller many times suffers from over-parametrization and the controller parameters can usually not be intuitively tuned.

Over-parametrization reduces the controller robustness and can introduce uncertainties into the process. While it can be directly tackled with penalization approaches (Shao, 2009) or by restrictions on the network structure (Krishnapura and Jutan, 2000), it stems many times from the lack of a physical structure. In this respect it was observed for hybrid process models that combine first principles knowledge and data-driven techniques (such as ANN) can structure the process operation space (Fiedler and Schuppert, 2008; Mogk & al., 2002), resulting in better calibration properties, better extrapolation capabilities and enhanced statistical confidence of the model estimates when compared to pure data-driven techniques (von Stosch & al., 2011a). Thus it can be expected that the combination of phenomena founded knowledge and data-driven techniques for control purposes might enhance controller robustness and performance. There are two ways to exploit this idea:

i) Identifying a hybrid-process model in canonical form and reformulating it in such a way that the control variable is expressed as an explicit function. This can be

achieved as proposed e.g. in (Bazaei and Majd, 2003; Hussain and Ho, 2004; Hussain & al., 2001; Madar & al., 2005; Oliveira, 1998; van Can & al., 1996; Xiong and Jutan, 2002)

ii) Using the structure of a standard controller and complement or replace parameters of the controller by functions that are modeled through data-driven techniques. The Generalized PID would be one example in which the PID parameters are given by ANNs (this is similar to a serial hybrid structure). Another example would be the parallel combination of PID and ANN by (Andrasik & al., 2004; Hisbullah & al., 2002; Li & al., 2006)(which is similar to a parallel hybrid structure).

While for the first case many times an heuristic intuitive search provides quickly good controller parameters, in the latter case the ANN controller parameters can usually not be tuned in the same way. Instead, controller tuning through optimization techniques becomes necessary. Off-line tuning is usually based on process data. A disadvantage is that the controller can, in limit, perform only as good as the controller used for acquiring the process data, so nothing is won. The off-line tuning can, of course, be enhanced whenever a process model is available but (i) when ANNs find application for control it is usually not straightforward to derive a process model from first principles only; and (ii) the application of process models solely obtained through data-driven techniques, such as ANNs, is delicate since such models are not parsimonious (Krishnapura and Jutan, 2000) and have limited extrapolation capabilities to new process regions.

Hybrid modeling also in this aspect is an interesting alternative since it can exceed the limits of pure data-driven techniques and since it does not require "precise" first principle knowledge, wherefore it additionally is cost-effective (Psichogios and Ungar, 1992; Schubert & al., 1994a; Thompson and Kramer, 1994). Nevertheless only relatively few closed-loop control applications applying hybrid models are reported. Most of which directly capitalize the model structure, as mentioned above (Bazaei and Majd, 2003; Hussain and Ho, 2004; Hussain & al., 2001; Madar & al., 2005; Xiong and Jutan, 2002), while others use the capability of the hybrid process model to accurately capture the process dynamics, i.e. for Model Predictive Control (Cubillos and Acuna, 2007; Cubillos and Lima, 1998; Ibrehem & al., 2011; van Can & al., 1996) or on-line tuning of the controller (Andrasik & al., 2004).

In this study, a general hybrid control framework is presented, in that (i) any structural limitations (e.g. for the standard PID, posed through the small numbers of controller

parameters and through the three controller inputs, namely the three Error types) are sought to be overcome using a control structure that incorporates ANNs; and in that (ii) additional available knowledge about the control structure can be incorporated into the controller which limits over-parameterization and facilitates parameter identification by structuring the parameter space.

The associated controller parameters are tuned by (i) using parameter identification from process data; or/and (ii) the application of a process model, namely a semi-parametric hybrid model is used, such avoiding the costly development of a detailed mechanistic model and extending the limits of pure data-driven techniques.

In order to investigate how knowledge about the controller impacts on the controller performance and to what extend the identification schema can be applied, a bioprocess control problem is adopted as a simulation case study. In this simulation case it is desired to close the loop for the control of the biomass concentration through the manipulation of the substrate feeding rate and to improve the performance of the dissolved oxygen concentration control by manipulating the stirring velocity. This is challenging since the control of these two quantities is coupled, the underlying kinetic system is highly non-linear and the inherent dynamics have different time scales.

6.3 Methodology

6.3.1 State space process model

The process model equations are based on the macroscopic material balance, which are derived for a stirred-tank reactor, assuming that the reactor content is ideally mixed, i.e.:

$$\frac{dc}{dt} = r(c, u, w) - D \cdot c \tag{6.1}$$

where c is a vector of concentrations, V is the reaction volume, D is the dilution rate, w is a set of model parameters, u are the control inputs and $r(\cdot)$ is a vector of kinetic rates. The formulation of the kinetic functions $r(\cdot)$ is based on a semi-parametric approach, (Oliveira, 2004), i.e.:

$$r(c, u, w) = K \cdot \langle \phi(c) \times \rho(c, u, w) \rangle \tag{6.2}$$

where K is a matrix of stoichiometric coefficients, ϕ is a function that can comprise available knowledge about the kinetics and ρ is a function that accounts for the unknown parts.

6.3.2 Structure of the General Hybrid Controller

The structure of the controller, sketched in Fig. 6.1, is determine by (i) the outputs of the controller which are the control inputs u; (ii) the inputs to the controller L_X; (iii) the controller parameters θ; and (iv) the vector of functions that connect the inputs and outputs, $g(\cdot)$. Each control input j of the number of total control inputs, J, can be formulated either as an algebraic equations, i.e.:

$$u_j = g_j(L_x, \theta), \tag{6.3}$$

or as an Ordinary Differential Equation:

$$\frac{du}{dt}_j = g_j(L_x, \theta), \tag{6.4}$$

where in eq. (6.4) the control inputs are obtained through integration.

Controller Inputs

The inputs L_X of the controller can optionally contain c_X, u, E_P (the proportional errors given by $E_P = c_s - c_X$), E_I (the integral errors given by $E_I = \int_t (c_s - c_X) \cdot d\tau$), and E_D (the differential errors given by $E_D = dc_s/dt - dc_X/dt$) with c_s being a vector of concentration set-points. The variable c_X can be chosen to be either the model predicted concentrations, c, or to be the experimentally measured concentrations c_{mes}. This choice determines (i) whether the model structure is the one of a one-step ahead predictor or the one of a multi-step ahead predictor, and (ii) whether the controller must be used in conjunction with the state space model or can standalone. This input choice is however not strictly binding, meaning that the structure can e.g. be identified with the predicted concentrations, but then applied for control using the measured concentrations, as done in the following case study.

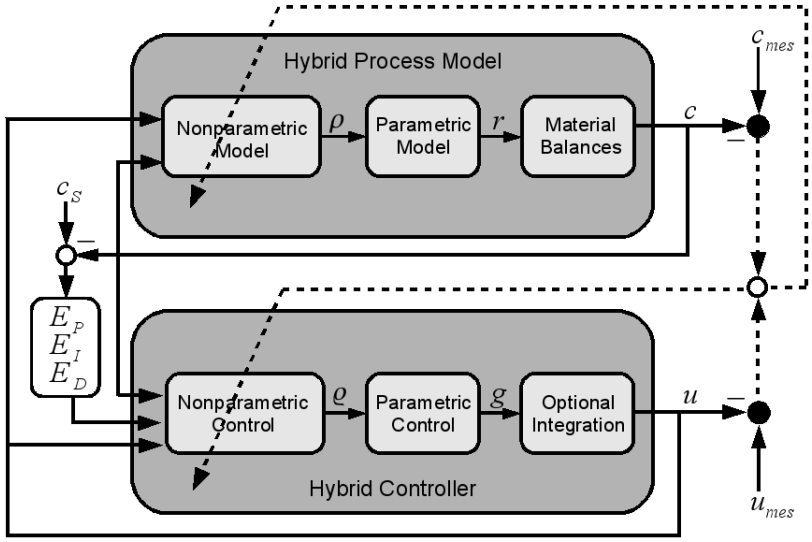

Figure 6.1 Schematic sketch of the hybrid controller structure and the hybrid process model structure in the context of the process data based identification procedure. Symbols as in the text.

Controller Functions

The vector of functions $g(\cdot)$ of the controller is here composed as:

$$g(c, u, \theta, E_p, E_i, E_d) = \Gamma \cdot \langle \varphi \times \varrho \rangle \,, \tag{6.5}$$

where Γ is a matrix that can contain ratio coefficients, $\varphi(\cdot)$ are functions that can contain available knowledge about the control structure, e.g. derived from the process model, and $\varrho(\cdot)$ are functions that account for unknown relations. It is evident that this structure is inspired by eq.(6.2).

6.3.3 Modeling of the unknown functions ρ and ϱ

The functions $\rho(\cdot)$ and $\varrho(\cdot)$ are modeled by three layer ANNs (which is sufficient for the modeling of nonlinear, continuous functions), that are represented by the following expressions:

$$\rho(c, u, w) = w_2 \cdot h(w_1 \cdot [c, u] + b_1) + b_2 \tag{6.6}$$

$$\varrho(c, u, \theta, E_p, E_i, E_d) = \theta_2 \cdot h(\theta_1 \cdot [c, u, E_p, E_i, E_d] + \varpi_1) + \varpi_2 \tag{6.7}$$

where the model parameters w comprise the weights matrices w_1 and w_2 and the bias vectors b_1 and b_2 ; and the control parameters θ comprise the weights matrices θ_1 and θ_2 and the bias vectors ϖ_1 and ϖ_2. The function $h(\cdot)$ can e.g. be chosen linear, sigmoidal, or hyperbolic tangential. In this study all nodes in the input and output layer were chosen to have a linear transfer function, while the nodes in the hidden layer were chosen to be hyperbolic tangential.

The identification of the neural network is a critical factor in order to achieve accurate and robust predictions (Bishop, 1995)[36]. In this study, the optimal ANN structure was chosen from a set of the possible candidates, as the one that produced the best performance in terms of the Bayesian Information Criteria (BIC), (Peres & al., 2008; von Stosch & al., 2011a).

6.3.4 Off-line Parameter Identification based on Process data

While the identification of the model parameters, w, from process data is generally feasible, regardless whether the process was run in open-loop or closed-loop, the off-line identification of the controller parameters, θ, from data is feasible only when the process was run in closed-loop. The following described identification procedure procures to the closed-loop case since for the application to open-loop process data simply the terms referring to the identification of the respective controller parameters can be dropped.

The sensitivities approach is applied in this study, due to its advantages (Oliveira, 2004; Psichogios and Ungar, 1992; Schubert & al., 1994a). The following weighted least-square error function is adopted to simultaneously identify the state space model and the controller parameters (as shown in Fig. 6.1), i.e.

$$\min_{w,\theta} \left\{ E_{LS} = \frac{1}{P_c} \sum_{l=1}^{P_c} \frac{(c_{mes,l}(t) - c_l(t))^2}{c_\sigma^2} + \frac{1}{P_u} \sum_{l=1}^{P_u} \frac{(u_{mes,l}(t) - u_l(t))^2}{u_\sigma^2} \right\}, \tag{6.8}$$

wherein P_c, P_u signify the respective number of events, c_{mes}, u_{mes} are the experimentally measured concentrations, feeding rates at time t, respectively, and c_σ, u_σ contain the respective measured data standard deviations. In order to minimize this objective a gradient based optimization scheme is used, namely the "lsqnonlin" MATLAB routine is

applied. The gradients are obtained through differentiation of eq. (6.8) with respect to w and θ, for details see appendix.

There exist two well known problems for the optimal parameter identification of ANNs: (i) over-fitting and; (ii) local minima in the shape of the error surface entangle the identification. The first problem is, as usually (Bishop, 1995; Schubert & al., 1994a), tackled with early-stopping wherefore two sets of data are required, a training and a validation set. The second problem is, in this study and as usual (Oliveira, 2004; von Stosch & al., 2011a), overcome with several random weight initializations for the same structure (at least 20 in all following cases). From the obtained candidates the best performing one, in terms of lowest prediction error obtained for the validation set, is selected. Additionally, a test data set can be used to explore the generalization capabilities of the identified model. Note that the simultaneous identification of w and θ was found to, in general, result into faster convergence and lower variances of the final parameter values for several random initializations than in the case that w and θ were identified sequentially.

6.3.5 Off-line Controller tuning based on the process model

Once the model parameters w are identified from experimental data, the controller parameters θ can be either identified in case that solely open-loop process data were available or further tuned off-line, as to meet certain specifications. The specifications for the tuning of the controller are expressed in requirements on (a) Set-point response; (b) Measurement noise response; (c) Load disturbance response; and (d) Robustness to model uncertainties (Astroem and Haegglund, 1995).

a) Set-point response: In order to obtain desired set-point tracking characteristics one can (i) build a time varying function that represents the desired set point tracking properties, $c_{desired}(t)$; and (ii) apply different available objective criteria in order to obtain varying desired characteristics of the set-point following (Astroem and Haegglund, 1995). In this study, exemplary, a normalized mean sum of squared error based objective is chosen, and smooth ramps are considered in lieu of abrupt set-point changes for $c_{desired}(t)$, such that.:

$$\min_{\theta} \left\{ E_{NMSSE} = \frac{1}{P_s} \cdot \sum_{l=1}^{P_s} \frac{(c_{desired,l}(t) - c_l(t))^2}{c_\sigma} \right\}. \tag{6.9}$$

This objective function is minimized applying the "lsqnonlin" MATLAB routine. The required gradients can be obtained as described in the appendix.

b) Measurement noise response: Measurement noise is usually simulated as an impulse function (Astroem and Haegglund, 1995). In this study the measurement noise is simulated by adding Gaussian noise to the inputs of the controller during the minimization, see Fig. 6.2. This has four desirable effects, namely that (I) the controller gain will not exceed a certain limit; (II) thus a penalty term in the objective that would aim at the reduction of differences in the control actions can be avoided; (III) the controller is more robust; and (IV) the controller tuning is not interrupted based on early stopping as in the case of the data-based identification, but when no further improvement of the E_{NMSSE} value is observed.

c) Load disturbance response: In order to account for load disturbances, the very also need to be integrated into the tuning. Therefore step functions are added to the control input, u, during the off-line tuning (see Fig. 6.2) as proposed in (Astroem and Haegglund, 1995).

d) Robustness to model uncertainties: It is obvious that the desired set-point tracking, i.e. the tuning of θ, depends strongly on the model quality, i.e. the prior identified parameters, w. Since these parameters have been derived from the training data, high model accuracy is limited to the vicinity of those regions of process variations that the model has been trained on. Off-line controller parameter tuning might however extrapolate from these regions. While one, of course, can simply limit the changes in the controller parameters, it might be superior to constraint the optimization such that extrapolations are restricted by some measure. Such a measure can, for instance, be provided by clustering (Teixeira & al., 2006) or convex-hull techniques (Kahrs and Marquardt, 2007). In this study, the off-line controller tuning is carried out in the vicinity to that process regions which the model has been identified on and therefore no constraints are incorporated.

In any case c_X (see the definitions of the PID errors) must be represented by c during the off-line tuning of the controller parameters, and thus c_{mes} in the controller inputs eventually replaced through c. This is due to the fact that when c_{mes} is used in the definitions of the PID errors, then the complete structure, consisting of the process model and the controller, is the one of a one-step ahead predictor, which is unsuitable for the off-line tuning of the controller parameters.

Needless to point out that in any case the model parameters remain unchanged.

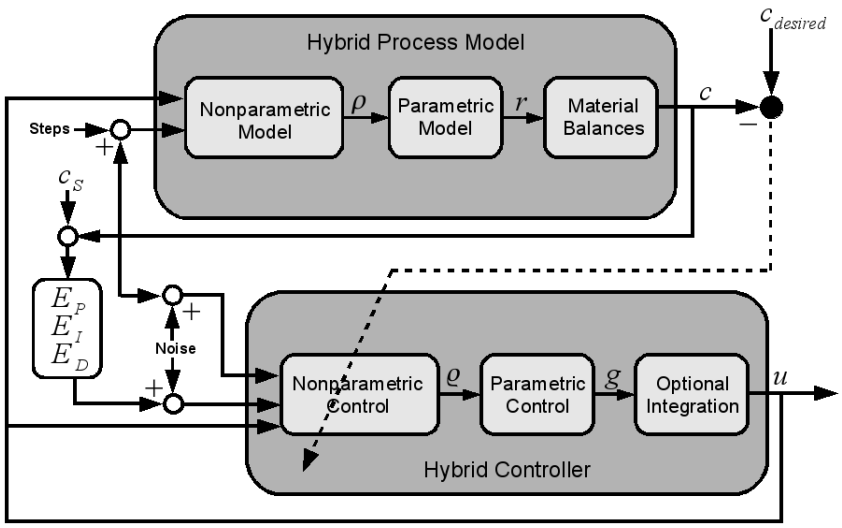

Figure 6.2 Schematic sketch of the hybrid controller structure and the hybrid process model structure in the context of the process model based identification procedure. Symbols as in the text.

6.3.6 Controller Performance Criteria

For the evaluation of the controller performance, a differentiation between different scenarios, must be accounted for by the controller performance evaluation criteria.

In case that load disturbances are applied, (Astroem and Haegglund, 1995) suggest to use the Integrated Error (IE) and the Integrated Absolute Error (IAE). These criteria therefore find application, but for their calculation, the tracking error is divided by the respective standard deviation in order of provide equal measures for different scales of state values.

In case that set-point changes are performed, (Astroem and Haegglund, 1995) suggest to use the Integral of Time and Error (ITE) and the Integral of Time and Absolute Error (ITAE). These criteria find application, but as before, the tracking error is divided by the standard deviation.

The closer the values of IE and ITE are to zero the better the controller performance. For IAE or ITAE the controller performance is the better the smaller the values.

6.4 Results & Discussion

6.4.1 The process

The performance of the different controller structures and the associated controller parameter identification is assessed in relation to a non-linear dynamic control problem. This problem is frequently addressed for bioprocess fed-batch reactor control, i.e.: the tight control for product quality and quantity. In the presented simulation case this is translated into (i) the control of biomass concentration through the substrate feeding rate manipulation (this goes to quantity, since either the specific biomass growth rate or the total amount of biomass is usually correlated to the total amount of product); and into (ii) the control of the dissolved oxygen concentration through the stirring velocity manipulation (this goes to quality, since variations in the dissolved oxygen concentration can limit growth or/and lead to the formation of toxic byproducts thus resulting into process variations). The simulation case, represented through equations (A.6.17) to (A.6.27), is an adaptation of the models by (Jahic *&* al., 2002) and (Cunha *&* al., 2004). Its control problem is challenging because of several reasons:

i) the kinetics of biomass growth and oxygen uptake are highly non-linear;

ii) the reaction kinetic term (as the transport term) is in general much larger than the accumulation term. The direct consequence is that, poor control leads to control saturation and min-max behavior;

iii) maximum productivity is often achieved at very low DO concentrations, close to the critical level where process is irreversibly lost;

iv) the methanol uptake rate is substrate inhibited (as reported for various *Pichia pastoris* fermentations);

v) the simultaneous control of biomass and dissolved oxygen concentrations is highly coupled;

The control problem anent the manipulation of the stirring velocity, is typically solved employing PI(D) control, while the control of the biomass concentration is usually either accomplished in open-loop following a pre-optimized substrate feeding profile or in closed-loop through the control of the specific biomass growth rate (Dabros *&* al.,

2010; Jenzsch & al., 2006a; Soons & al., 2006). However in this study, consciously, the biomass concentration is controlled because (i) the specific biomass growth rate cannot be directly measured while the biomass concentration, in principle, can (this enabling the application of standalone controllers); and (ii) the biomass concentration is an integrating quantity (which slightly complicates the tuning of a standard PID controller).

Further it is assumed that up to this point only the dissolved oxygen concentration was controlled in closed-loop, while the biomass concentration was controlled in open-loop subject to pre-optimized substrate feeding rate profiles. Process data of in total 8 runs were generated applying equations (A.6.17) to (A.6.27). The data of the process runs were then divided into three data sets, i.e. for training (4 fed-batches), validation (2 fed-batches) and test (2 fed-batches).

The tasks are: (i) to close the loop for the control of biomass concentration; and (ii) to investigate alternative approaches for the control of dissolved oxygen concentration (replacing eq. (A.6.27) by the controller equations presented in section 6.4.3).

6.4.2 The Hybrid process model

The hybrid process model comprises the material balance of the biomass and the dissolved oxygen concentrations, since it is assumed that these are the only accessible on-line measurements, i.e.:

$$\begin{aligned} \frac{dX}{dt} &= \mu \cdot X - D \cdot X \\ \frac{dC_O}{dt} &= r_T \cdot RPM - r_O \cdot X - D \cdot C_O \end{aligned}, \tag{6.10}$$

additionally the volume is integrated along, namely eq. (A.6.24). The specific rates of biomass growth, $\mu = \mu(X, C_O, u_{Met})$, oxygen uptake, $r_O = r_O(X, C_O, u_{Met})$, and a function of the oxygen transfer, $r_T = r_T(C_O)$ are each modeled by separate ANNs. The best performing hybrid process model comprises two nodes for each of the kinetic rates, i.e. μ, r_O, and r_T. In Fig. 6.3 the process data and the predictions for the hybrid process model and the hybrid controller, $u_{1,A}$, (which were trained together using the process data) are shown for a fed-batch of the test data. Therein it can be seen that very good predictions were obtained for both biomass and dissolved oxygen concentrations and that also the characteristics of the PI controller are very well mimicked by the $u_{1,A}$ controller. In case of the stirring velocity two facts are striking, i.e. (i) the step changes; and (ii)

the fluctuation in the control actions for the process data (the stirring velocity data were not corrupted with noise). The step changes are due to changes in the feeding rate at a constant dissolved oxygen set-point. The fluctuations are caused by the PI-controller that translates the simulated measurement error into control action. These fluctuations cannot be observed for the $u_{1,A}$ controller, since its inputs are based on the predicted dissolved oxygen concentration during the identification.

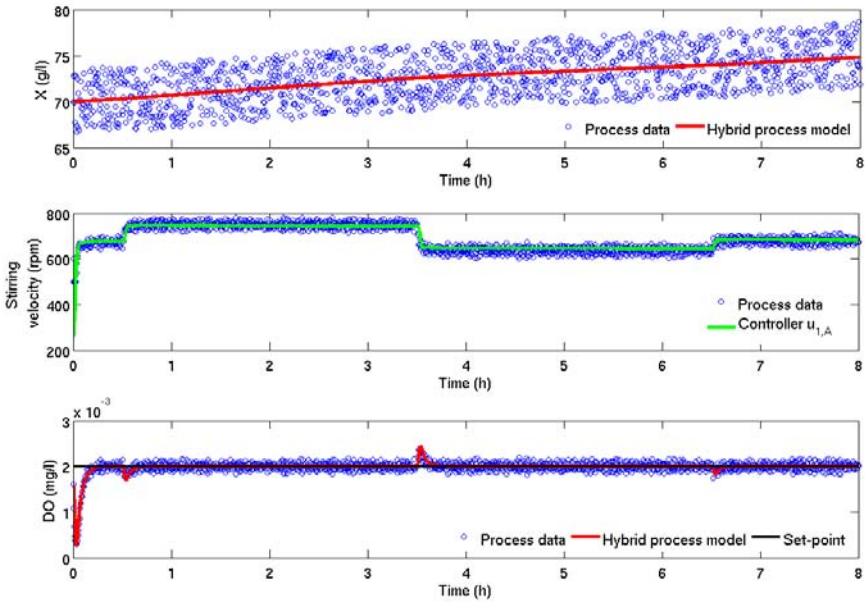

Figure 6.3 Plots of the biomass concentration, X, the dissolved oxygen concentration, DO, and the stirring velocity, RPM, over time, for one fed-batch of the test data set in case of: measured data, red circles; the hybrid-process model estimates, blue dashed line; and the control action of controller $u_{1,A}$, green dashed line. The set-point is also displayed, fine black continuous lines.

6.4.3 Hybrid controller structures

In the following a number of different control structures are proposed, which present a selection from various investigated controller designs.

Dissolved oxygen concentration control through manipulation of the stirring velocity

The stirring velocity, in this case, could be regulated by a simple PI approach, eq. (A.6.27). The tuning of the PI controller can e.g. be carried out with designs (e.g. Ziegler-Nichols or Tyreus-Luyben) that uses the knowledge of the ultimate gain and of the ultimate period. It might however be difficult to precisely determine the ultimate gain and period, e.g. when the sampling frequency is relatively low or/and when the sensor signal is relatively noisy. The step response method could then be used. However, the ultimate gain and ultimate period of the system might change during the process, leading to significant decreases in controller performance, for instance see(Zhang & al., 2002). Therefore in the following three alternative structures are investigated.

Controller 1A: A simple controller based on an ANN is proposed at first, i.e.:

$$u_{1,A} = rpm = ANN((C_{OS} - C_O), \int_t (C_{OS} - C_O) \cdot d\tau). \tag{6.11}$$

The data based identification scheme, in this case, can find application for the tuning of the controller, since the dissolved oxygen concentration "during the experiment" was controlled in closed-loop fashion. This means that the characteristics of the PI-controller, eq. (A.6.27) are mimicked at first, which can for instance be seen in Fig. 6.3. Thereafter the process model based identification schema is applied to tune the parameters regarding the specifications. Different network inputs were investigated, i.e. the network structure was prune. The best performing ANN has 2 hyperbolic tangential nodes in the hidden layer and only the proportional and integral error of the dissolved oxygen concentration as inputs. The biases in the input and hidden layers could be eliminated, since they had no significant effect on the controller performance.

Controller 1B: A second controller is based on a first order Model Reference Control (MRC) schema, i.e.: $C_{OS} = C_O + \tau_c \cdot \frac{dC_O}{dt}$, and using the hybrid process model, namely the balance for the dissolved oxygen concentration, the controller reads as:

$$u_{1,B} = rpm = \frac{1}{r_T} \cdot [(C_{OS} - C_O)/\tau_c + r_O \cdot X + D \cdot C_O], \tag{6.12}$$

where the rate expression stem from the hybrid process model (identified before) and thus the only control parameter that needs to be tuned is τ_c. This can either be accomplished by an heuristic search (which usually quickly results in good performance) or applying the proposed schema for parameter identification based on the hybrid process model. The gradients in the latter case are obtained by differentiating eq. (6.12) with respect to τ_c. It is interesting to note that this controller structure is similar to a proportional controller with steady state bias ($rpm = K_p \cdot (C_{OS} - C_O) + K_B$) (Seborg & al., 2010), just that in the given structure both the controller gain and the bias are not fixed values but variable in correspondents to eq. (6.12), i.e.: $K_p = \frac{1}{r_T \cdot \tau_c}$ and $K_B = \frac{r_O \cdot X + D \cdot C_O}{r_T}$.

Controller 1C: In a third control structure the concept of Generic Model Control (GMC) is applied, i.e.: $\frac{dC_O}{dt} = K_p \cdot (C_{OS} - C_O) + K_i \cdot \int_t (C_{OS} - C_O) \cdot d\tau$, which along with the mass balance equation of the dissolved oxygen concentration combines to:

$$u_{1,C} = rpm = \frac{1}{r_T} \cdot \left[K_p \cdot (C_{OS} - C_O) + K_i \cdot \int_t (C_{OS} - C_O) \cdot d\tau + r_O \cdot X + D \cdot C_O \right], \tag{6.13}$$

In this case the kinetic rates are the ones of the process model and therefore the only parameters left to tune are K_p and K_i. Tuning of both parameters can, as before, be carried out manually or with the process model based identification schema. In case that the latter is applied it was found that if the ratios $K_p = 2 \cdot \zeta / \vartheta$ and $K_i = 1/\vartheta^2$, proposed by (Lee and Sullivan, 1988), were respected, i.e. ζ and ϑ are identified instead, then the tuning converged faster and resulted into consistent performance for various random initiations of the parameters.

Biomass concentration control through substrate feed rate manipulation:

The control of the biomass concentration is many times accomplished in open-loop, which is due to the fact that (i) reliable biomass measurements are many times not available at-time (but due to the PAT initiative become increasingly available); (ii) simple control methods with reasonable performance are many times preferred over more complex control schema with greater performance (Alford, 2006). In the following three controllers are proposed to close the loop. All of which only use generic first-principle knowledge or/and process knowledge that was captured from readily available process

data (that were recorded when biomass was controlled in an open-loop fashion). Since the biomass concentration was controlled in open-loop, the process data identification schema could not be applied for the tuning of the controllers presented in the following but instead only the process model based controller tuning schema finds application.

Controller 2A: The first control structure is defined by an Ordinary Differential Equation, where the right hand side is modeled by a simple ANN, i.e.:

$$u_{2,A} = \frac{du_{Met}}{dt} = ANN\left(u_{Met}, (X_S - X), \int_t (X_S - X)\cdot d\tau, \left(\frac{dX_S}{dt} - \frac{dX}{dt}\right)\right). \quad (6.14)$$

As before, the network structure was prune concerning the inputs. The best performing ANN has 2 hyperbolic tangential nodes in the hidden layer and, as inputs, the methanol feeding rate and the proportional, integral and differential errors. The biases in the input layer could be eliminated.

Controller 2B: The second controller proposed is based on a second order MRC schema, i.e.: $X_S = X + \beta_1 \cdot \frac{dX}{dt} + \beta_2 \cdot \frac{d^2X}{dt^2}$, which yields when using the hybrid process model, namely the balance for biomass, into:

$$u_{2,B} = \frac{du_{Met}}{dt} = \frac{1}{\beta_2 \cdot A}\cdot\left[(X_S - X) - \frac{dX}{dt}\cdot(\beta_1 + \beta_2 \cdot B) - \beta_2 \cdot X \cdot C\right], \quad (6.15)$$

where $A = \left(X \cdot \frac{d\mu}{du_{Met}} - \frac{X}{V}\right)$; $B = \left(X \cdot \frac{d\mu}{dX} - D + \mu\right)$ and $C = \left(\frac{d\mu}{dC_O}\cdot\frac{dC_O}{dt} + \frac{u_{Met}^2}{V^2}\right)$ and β_1 and β_2 are the MRC controller parameters. These controller parameters can either be tuned heuristically as e.g. in (Oliveira & al., 2004) or, as before, with the proposed method for process model based parameter identification. The required gradients are obtained by differentiating eq. (6.15) with respect to β_1 and β_2.

As a matter of fact, this schema is similar to the one proposed by (Soons & al., 2006), which in turn is similar to a GMC design. In case that the prior mentioned ratios proposed by (Lee and Sullivan, 1988) for GMC were applied, i.e. $\beta_1 = 1/\gamma_1^2$; $\beta_2 = 2\cdot\gamma_1/\gamma_2$ and identification of γ_1 and γ_2 instead, also in this case the tuning converged faster (than in the case the ratios were not used) and resulted into consistent performance for various random initializations. It must be pointed out that due to the derivatives, $d\mu/du_{Met}$, $d\mu/dX$ and $d\mu/dC_O$, the Hessian is required for the gradient based identification, which can, however, be relatively easy obtained through symbolic manipulations.

Controller 2C: A third control structure is investigated in which it is assumed that the dynamics can be captured by a so called general PID. The controller equations read as:

$$u_{2,C} = \frac{du_{Met}}{dt} = K_{p,fun} \cdot (X_S - X) + K_{i,fun} \cdot \int_t (X_S - X) \cdot d\tau + K_{d,fun} \cdot \left(\frac{dX_S}{dt} - \frac{dX}{dt} \right), \tag{6.16}$$

where $K_{p,fun}$, $K_{i,fun}$ and $K_{d,fun}$ are not parameters but functions that are modeled by an ANN. The associated network weights, the controller parameters, are tuned through the process model based schema. The ANN was prune concerning the inputs. The best performing ANN has 3 hyperbolic tangential nodes in the hidden layer and, as inputs, the methanol feeding rate, the biomass concentration, the dissolve oxygen concentration and the volume.

General remarks regarding the controller structures and their tuning:

During investigations on other than the here presented controller structures it was observed that for those biomass controllers which consisted of an algebraic equation, it was either infeasible to tune the parameters or the controller candidates in comparison to the proposed structure performed poorly.

Note further that for the parameter identification predicted concentration values were used for the calculation of the tracking errors, but that for the simulation of the process under control, the measured concentration values are used.

6.4.4 Process under Control

The controllers of the dissolved oxygen concentration and the biomass concentration can be pairwise arranged in nine possible combinations. Additionally, the biomass controller were paired with the dissolved oxygen PI-controller (eq. A.6.27), such providing a baseline for comparison. In order to evaluate these combinations under control, they are applied to the process simulation described in the appendix. For the evaluation four cases are considered, namely two cases in which changes were applied to the set-points and two other in which the process was disturbed.

Step changes in the dissolved oxygen set-point:

Step changes (a positive step and a negative step, both with a difference of $0.4 \cdot 10^{-3}$ (mg/l)) were applied to the set-point of the dissolved oxygen concentration, the most

insightful results are compiled in Table 6.1. Therein it can be seen that the performance of the dissolved oxygen concentration controllers, irrespective of the biomass controller (according to the ITAE criteria), seem to be ranked as follows, $PI < u_{1,A} < u_{1,B} < u_{1,C}$, where the latter shows the best performance. The better performance of the controllers $u_{1,B}$ and $u_{1,C}$ when compared to the $u_{1,A}$ could be expected since (i) in this control schema the coupling between the biomass and dissolved oxygen control is taken into consideration; and (ii) the most process knowledge regarding the dissolved oxygen concentration is integrated into the structure. The same reasoning can be used to explain the good performance of the $u_{2,B}$ controller. However in comparison with the $u_{2,C}$ controller significantly more knowledge is incorporated into the structure of the $u_{2,B}$, but the performance is not significantly better. As a matter of fact, the second decimal digit (values of ITAE) in this case is rather due to the random measurement noise. In case of the ITE, relatively low absolute values are obtained for all controller combinations, which is due to the error canceling. Thus for none of the controller combinations, an error bias can be observed, which is interesting to note since the formulation of the controller $u_{1,B}$, as mentioned above, is similar to the one of a proportional controller with constant bias.

Table 6.1 Controller performance criteria values, namely ITAE and ITE, obtained for the dissolved oxygen concentration with all possible combinations of the controller, in case that a step change was applied to the set-point of the dissolved oxygen concentration.

	$u_{1,A}$		$u_{1,B}$		$u_{1,C}$		PI	
$u_{2,A}$	4.0160	0.0574	1.9207	-0.0109	1.8830	-0.0018	4.8169	0.0321
$u_{2,B}$	1.0984	-0.0492	0.7427	-0.0226	0.7255	-0.0227	2.0687	-0.0310
$u_{2,C}$	1.1423	-0.0095	0.7398	-0.0156	0.7414	-0.0189	1.5749	-0.0183
	ITAE	ITE	ITAE	ITE	ITAE	ITE	ITAE	ITE

In Fig. 6.4 it can be seen that all displayed controller combinations manage equally well to follow the biomass set-point closely. In case of the dissolved oxygen it is observable that the controllers can follow the positive and the negative set-point step changes, but that those combinations that contain the $u_{2,A}$ controller exhibit significantly larger variations. This becomes even more evident, when comparing the results obtained for the combinations ($u_{1,A}$, $u_{2,A}$), ($u_{1,A}$, $u_{2,C}$) and ($u_{1,C}$, $u_{2,A}$), where it can be concluded that fluctuations in the control action of the methanol feeding lead to fluctuations in the dissolved oxygen concentration which then cause changes in the stirring speed. Further, it is interesting to note that in the case of the combination ($u_{1,C}$, $u_{2,B}$), the methanol feeding rate is increased at the time instance on which the negative step in the

dissolved oxygen set-point occurs, while the stirring velocity is kept relatively constant and while no significant deviations can be observed for the biomass concentration. The explanation for this behavior can be found in the simulation case equations, eq. (A6.21), i.e. when the dissolved oxygen concentration is lower then the methanol concentration must be greater in order to maintain the substrate uptake rate at the same level, and thus the specific biomass growth (A6.20). Therefore it can be stated that the hybrid process model can perfectly capture the "true"underlying transient behavior, and that this knowledge can be translated into adequate control action.

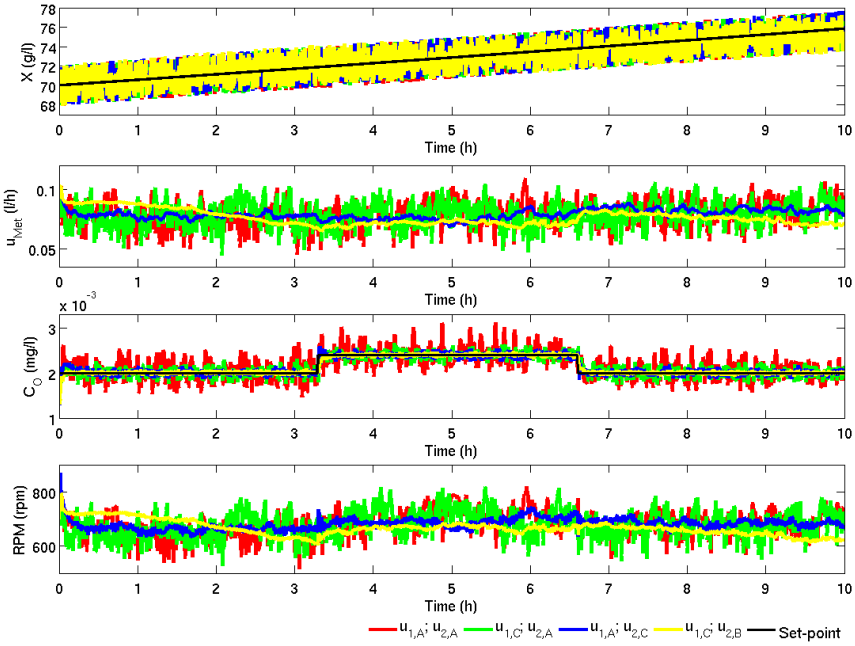

Figure 6.4 Simulated biomass concentration, X, methanol feeding rate, u_{Met}, simulated dissolved oxygen concentration, DO, and stirring velocity, RPM, over time in case of dissolved oxygen set-point changes, under control of: ($u_{1,A}$, $u_{2,A}$) red line; ($u_{1,A}$, $u_{2,C}$) blue line; ($u_{1,C}$, $u_{2,A}$) green line; and ($u_{1,C}$, $u_{2,B}$) yellow line. The set-point is also displayed, fine black continuous lines.

Slope changes in the biomass set-point:

Changes to the slope (an increase in the slope at 3.33 (h) followed by setting the slope to zero at 6.66 (h), the latter constituting a situation that was not covered during the identification phases) of the biomass set-point were applied, the results obtained in

from of the ITAE and ITE values, are presented in Table 6.2. Therein it can be seen that the best performance (in terms of ITAE) is obtained for those controller combinations that contain the $u_{2,B}$, and the same ranking of the dissolved oxygen controllers as before can be observed, i.e. $PI < u_{1,A} < u_{1,B} < u_{1,C}$, where the latter performs the best.

Table 6.2 Controller performance criteria, namely ITAE and ITE, obtained for the dissolved oxygen concentration with all possible combinations of the controller, in case that changes in the set-point slopes of the biomass set-points were applied.

	$u_{1,A}$		$u_{1,B}$		$u_{1,C}$		PI	
$u_{2,A}$	5.8858	-0.2707	2.5503	-0.0360	2.4851	-0.0370	9.5185	-0.1781
$u_{2,B}$	1.8747	0.0446	0.7005	-0.0200	0.6810	-0.0196	1.6927	-0.1172
$u_{2,C}$	17.4738	-0.3931	1.9825	-0.0326	1.8528	-0.0357	18.1136	-0.3067
	ITAE	ITE	ITAE	ITE	ITAE	ITE	ITAE	ITE

The worst performance in terms of the ITAE (apart from those combinations with the PI controller), is this time observed for the combination ($u_{1,A}$, $u_{2,C}$) followed by ($u_{1,A}$, $u_{2,A}$), which already performed poor before. This observation is supported by the respective, rather large, ITE values, which indicate that the tracking error is lopsided.

These findings are also supported, when looking at Fig. 6.5. Therein, bang-bang behavior in the dissolved oxygen concentration trajectory can be observed in case of the ($u_{1,A}$, $u_{2,C}$) when the slope of the biomass set-point is zero. The same would be observed with the PI-controller, as indicated by the performance criteria see Table 6.2. Similar behavior, but less distinct can also be noted for the ($u_{1,C}$, $u_{2,C}$). In case of both combinations (($u_{1,A}$, $u_{2,C}$) and ($u_{1,C}$, $u_{2,C}$)), it can be seen that during the bang-bang phase the feeding rate is almost zero, wherefore the accumulation term in the dissolved oxygen balance becomes dominant since oxygen is only up-taken for biomass maintenance. All dissolved oxygen controller were however tuned for the opposite case, and thus the performance of the $u_{1,C}$ is rather good, which demonstrates its extrapolation capabilities. Also the dissolved oxygen set-point tracking performance of the $u_{1,C}$ in combination with the $u_{2,B}$ is outstanding.

The same can be stated for the $u_{2,C}$ controller (when paired with the $u_{1,C}$ or $u_{1,b}$, the latter not shown) for which a prompt decrease in the feeding rate can be observed, which directly translates into the limitation of the specific biomass growth. Also for the profile of the methanol feeding rate of this controller, it is visible that the controller promptly acts when the slope of the biomass set-point is increased. In contrast, the feeding rate trajectories obtained by the other two controller, namely $u_{2,A}$ and $u_{2,B}$,

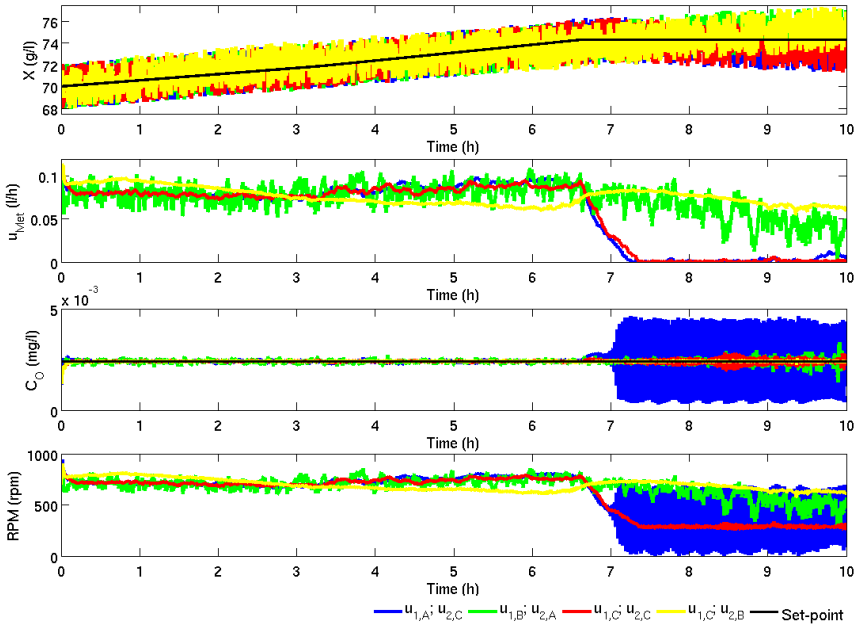

Figure 6.5 Simulated biomass concentration, X, methanol feeding rate, u_{Met}, simulated dissolved oxygen concentration, DO, and stirring velocity, RPM, over time in case of biomass set-point changes, under control of: ($u_{1,A}$, $u_{2,C}$) blue line; ($u_{1,C}$, $u_{2,C}$) red line; ($u_{1,B}$, $u_{2,A}$) green line; and ($u_{1,C}$, $u_{2,B}$) yellow line. The set-point is also displayed, fine black lines.

do not clearly exhibit these set-point changes. In case of the $u_{2,A}$ this is due to the fact that the control action is, as before when changing the dissolved oxygen set-point, constantly varying which is not desirable and which translates into fluctuations in the dissolved oxygen concentration profile. The feeding rate profile in the case of the $u_{2,B}$ is due to the rather poor biomass set-point tracking. This behavior is owed to the identified constants $\beta_1 = 1/(12.5^2)$ and $\beta_2 = 2 \cdot 15/12.5$ which lead to good noise rejection properties, but result in rather slow adaption of the control action.

Disturbances due to variations in the oxygen transfer and the cell characteristics:

Variations in the oxygen transfer and the cell characteristics are typically encountered during bioprocesses. Variations in the oxygen transfer were simulated by setting the value of 0.82 (1/h/rpm) in eq. (A.6.23) to 0.7 (1/h/rpm) at 3.33 hours and to 0.75 (1/h/rpm) at 6.66 hours. In order to enhance the effect on the dissolved oxygen

concentration, at the same time instance firstly an increase in both anaerobic and energetic metabolism yields were considered (the value of $Y_{OS,en}$ was increased to 1.7; and the value of $Y_{OS,an}$ to 0.7), which leads to a rise in the demand of dissolved oxygen, and then, at the second time instance, the yields were slightly decreased (the value of $Y_{OS,en}$ was decreased to 1.6; and the value of $Y_{OS,an}$ to 0.6). Additionally, the yield of biomass on substrate was lowered, such that more substrate needs to be taken up to meet the biomass set-point specifications and following slightly back increased (the value for Y_X was lowered from 0.36 to 0.30 at 3.33 hours and increased to 0.32 at 6.66 hours).

Table 6.3 Controller performance criteria, namely IAE and IE, obtained for the dissolved oxygen concentration with all possible controller combinations for variations in the oxygen transfer and the cell characteristics.

	$u_{1,A}$		$u_{1,B}$		$u_{1,C}$		PI	
$u_{2,A}$	1.6347	1.1449	1.3813	1.2128	1.7573	1.5755	1.6263	1.1128
$u_{2,B}$	0.6926	0.5424	1.1515	1.0347	1.8192	1.7210	1.6051	1.4169
$u_{2,C}$	1.1577	0.9724	1.1686	1.0427	1.1508	1.0309	1.2660	1.0935
	IAE	IE	IAE	IE	IAE	IE	IAE	IE

The obtained performance in terms of IAE and IE criteria calculated for the dissolved oxygen concentration are for this case compiled in Table 6.3. Therein it can be seen that all IE values are positive, which stems for lopsided tracking errors that are due to the disturbances. Further, the performances in terms of IE and IAE are consistent. It strikes that the best and the worst performances (in terms of IAE) seem to be obtained with controller combinations that comprise the $u_{2,B}$, i.e. ($u_{1,A}$, $u_{2,B}$) and ($u_{1,C}$, $u_{2,B}$), respectively. This is unexpected since the combinations with the $u_{1,A}$ usually demonstrate significantly worse performances than those with the $u_{1,C}$. However it can be seen that the $u_{1,A}$ also in combination with other controllers performs better or at least equally, when compared to the respective performance obtained for combinations with the $u_{1,C}$. Further, the good performance of the $u_{2,B}$ in this case would be unexpected because the $u_{2,B}$ is based on the hybrid process model that was identified from process data that did not contain changes in the model parameters. When analyzed visually, see Fig. 6.6, it can be seen that those combinations with the $u_{2,B}$ perform well only for about 0-9h and thereafter produce poor results, due to increases in the methanol feeding. These increases result from the fact that the controller underlying model was, as already mentioned, not trained for parameter changes. It can however be expected that the application of an on-line parameter adaptation schema would lead to an increase in performance in that case.

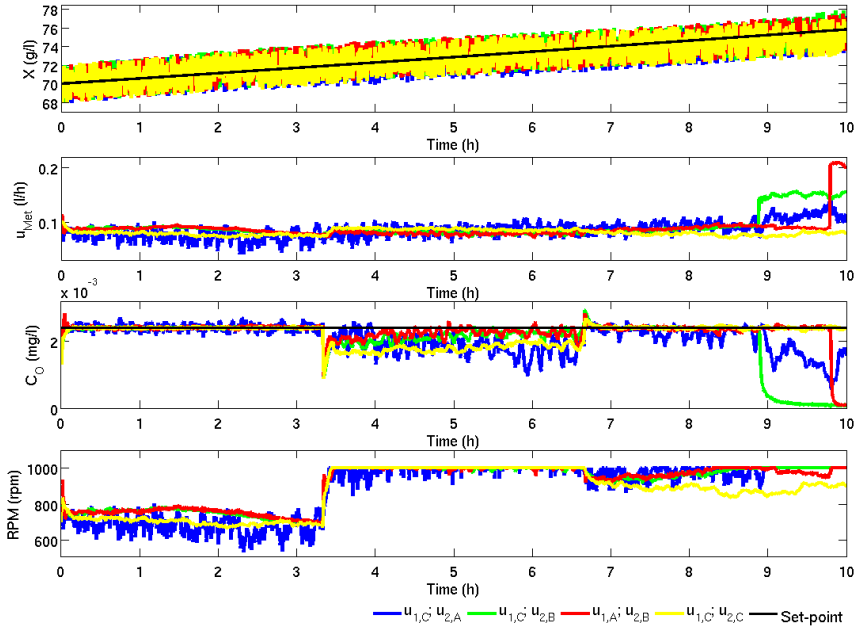

Figure 6.6 Simulated biomass concentration, X, methanol feeding rate, u_{Met}, simulated dissolved oxygen concentration, DO, and stirring velocity, RPM, over time in case of variations in the oxygen transfer and the cell characteristics under control of: ($u_{1,C}$, $u_{2,A}$) blue line; ($u_{1,A}$, $u_{2,B}$) red line; ($u_{1,C}$, $u_{2,B}$) green line; and ($u_{1,C}$, $u_{2,C}$) yellow line. The set-point is also displayed, fine black continuous lines.

Further, it can be observed in Fig. 6.6, that the methanol feeding rate increases when the first disturbance occurs. This is due to the fact that one part of the simulated disturbance was to lower the biomass yield on methanol and thus the biomass controller need to increase the feeding rate in order to meet the biomass set-point specifications. It can also be seen that after the occurrence of the first disturbance, all controllers are limited by the maximum possible stirring velocity of 1000 (rpm) (The integrator windup is accounted for by fixing the integral error value, when the maximal or minimal feasible actuator value is reached). This fact together with the increased feeding rate are the reasons why the dissolved oxygen set-point specifications are not met, which in turn results in the lopsided IE values.

Else, it can be seen that the difference in the stirring velocity of the $u_{1,A}$ is greater than those differences observed for the other controllers. Therefore it can be concluded that the controller has a greater gain than the others, explaining for its sensitivity to

measurement noise, which in turn is in agreement with the observations made before. During the controller tuning this effect was less distinct and since the interpretation of ANNs is relatively difficult, it was not clear, prior to application, how the controller would behave.

Disturbances due to methanol accumulations in the reactor:

Accumulations of substrates is not desirable but frequent in fed-batch bioprocess, partly due to non-ideal mixing. Methanol accumulations were simulated by setting the methanol concentration at 4.5 hours to 5 (g/l) (under control the usual ranges is 0-0.15 g/l), the performances in terms of the IAE and IE calculated for the dissolved oxygen concentration are comprised in Table 6.4. Therein, the values of IE can be observed, again, to be all positive which points at lopsided tracking errors. Further the performances in terms of IAE and IE seem consistent apart from the performances of the combination ($u_{1,A}$, $u_{2,A}$) and (PI, $u_{2,A}$), where the obtained IE value is disproportionately low, which indicates oscillatory behavior. This indeed can be observed for the ($u_{1,A}$, $u_{2,A}$) in Fig. 6.7, i.e. almost a bang-bang situation in case of the dissolved oxygen concentration, which origins from the relative low methanol feeding rate, as described before.

Table 6.4 Controller performance criteria, namely IAE and IE, obtained for the dissolved oxygen concentration with all possible combinations of the controller in case of methanol accumulations in the reactor.

	$u_{1,A}$		$u_{1,B}$		$u_{1,C}$		PI	
$u_{2,A}$	3.2353	0.9616	1.6433	0.8913	1.7327	0.9614	3.7404	1.0483
$u_{2,B}$	3.6400	3.2521	1.2885	0.8926	1.2669	0.9456	2.3317	1.0968
$u_{2,C}$	1.3656	1.0830	1.3218	1.0690	1.3142	1.0646	2.0356	1.1580
	IAE	IE	IAE	IE	IAE	IE	IAE	IE

In general, the dissolved oxygen controller performance (in terms of IAE) seems to rank as before, i.e. $PI < u_{1,A} < u_{1,B} < u_{1,C}$, where the latter performs the best. These observations are reinforced by the visual inspections made for Fig. 6.7.

Again, the best and the worst performance in terms of IAE, Table 6.4, seem to be encounter for the $u_{2,B}$. This time the ($u_{1,C}$, $u_{2,B}$) seems to perform better than the ($u_{1,A}$, $u_{2,B}$), which is in agreement with the results displayed in Fig. 6.7.

In fact, the combination ($u_{1,A}$, $u_{2,B}$) at the moment in which the disturbance occurs performs adequately, i.e. the feeding rate is promptly lowered and the stirring velocity

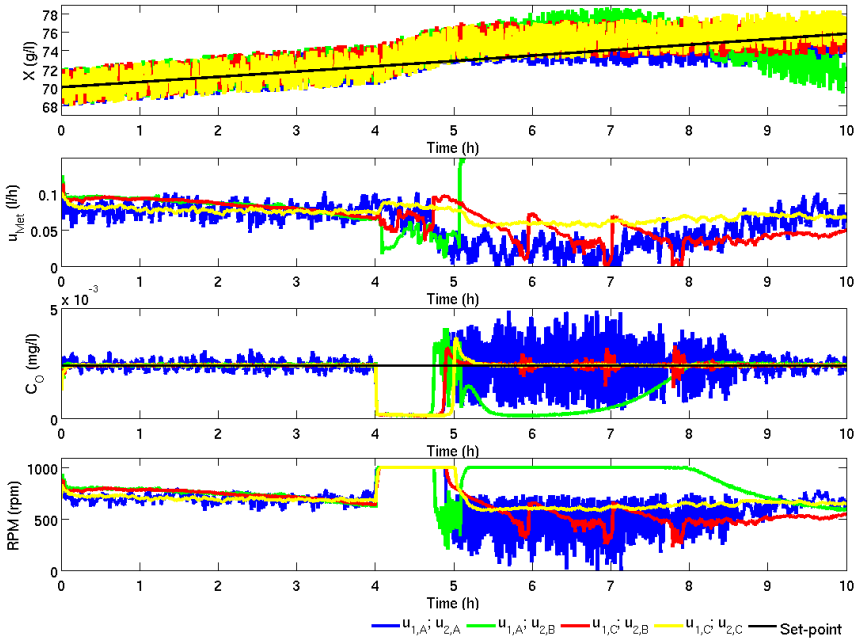

Figure 6.7 Simulated biomass concentration, X, methanol feeding rate, u_{Met}, simulated dissolved oxygen concentration, DO, and stirring velocity, RPM, over time in case of unaccounted methanol accumulations in the reactor, under control of: ($u_{1,A}$, $u_{2,A}$) blue line; ($u_{1,C}$, $u_{2,B}$) red line; ($u_{1,A}$, $u_{2,B}$), green line; and ($u_{1,C}$, $u_{2,C}$) yellow line. The set-point is also displayed, fine black continuous lines.

increased to the limit. Also in the following 3/4 hour the performance is superior when compared to other combinations. Then, at about 5 (h), a number of circumstances accrue, namely (i) the till this point greatest difference between the measured biomass concentration and the set-point; plus (ii) high values of the gradient, dC_O/dt as a result of the fluctuations in the dissolved oxygen concentration; plus (iii) high sensitivity of the biomass growth with respect to the dissolved oxygen concentration, $d\mu/dC_O$; and as a result the value of the feeding rate increases instantaneously to about 0.2 (l/h) (which is outside of the plot, since else the characteristics of the other feeding rate trajectories would not be visible, however the qualitative behavior can be seen in Fig. 6.8). From thereon the situation gets worse, since biomass is growing further, limited only by the dissolved oxygen concentration. The $u_{2,B}$ controller, which was not trained for such a situation, outruns the specific growth rate by increasing the feeding rate (and thus the dilution term), which seems to normalize the situation (decreasing of the biomass

concentration and rise in the dissolved oxygen concentration). However, this leads to the accumulation of an enormous amount of methanol in the reactor, see Fig. 6.8, and thus into substrate inhibited biomass growth, eq. (A6.21). In reality, the process would be aborted.

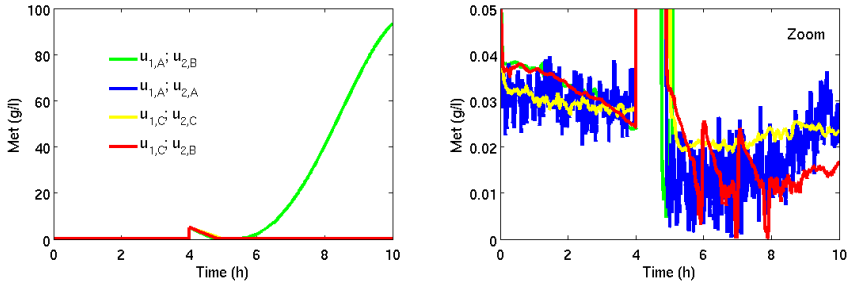

Figure 6.8 Plots of the methanol concentration, *Met*, (without the simulated noise) over time in case of methanol accumulations in the reactor under control of: ($u_{1,A}$, $u_{2,A}$) blue line; ($u_{1,C}$, $u_{2,B}$) red line; ($u_{1,A}$, $u_{2,B}$), green line; and ($u_{1,C}$, $u_{2,C}$) yellow line. The set-point is also displayed, fine black continuous lines.

As a matter of fact, the good performances observed in this case for the other $u_{2,B}$ combinations, are on the edge of losing the control in the way just described. The unstable behavior of the controller ($u_{2,B}$) in this case is due to the fact that the process model does not adequately account for the disturbance. Thus the gradients of the specific biomass rate with respect to the feeding rate, the biomass concentration or the dissolved oxygen concentration, which are part of the controller equation do not represent the real situation. The uncertainty in these gradients is then further amplified through multiplications by high time derivative values and by $\beta_2 = 2 \cdot 15/12.5$, see eq. (6.15). An on-line parameter identification schema could put things right. The model based controllers $u_{1,B}$ and $u_{1,C}$ are due to their structure less prone to these kind of uncertainty amplifications (better conditioned), since the governing factors in case of disturbances are the tracking error based terms.

In case of the $u_{2,C}$ controller it can be seen in Fig. 6.7, that when the disturbance occurs the controller does not react adequately, i.e. the feeding rate is slightly increased. This results into a slightly longer period of high methanol concentrations (Fig. 6.8) and thus into a longer phase of dissolved oxygen limited biomass growth (Fig. 6.7). However, this controller maintained the control over the biomass concentration in any case. Worthwhile to mention is the fact that the methanol concentration along time,

for almost all combinations that include the $u_{2,C}$ and for all of the presented cases, remained almost constant, as can e.g. be seen in Fig. 6.8. This is very desirable, since the maximum product formation is usually assumed to occur at a distinct small band of methanol concentrations.

6.5 Conclusions

A general hybrid control structure was proposed, that seeks to overcome nescience by using Artificial Neural Networks (ANNs), while limiting the typical shortcomings of ANNs through the integration of other available process knowledge, therefore being hybrid. This control structure is general since it can be customized to each application, e.g. General Linear Controllers or the traditional PID controller constitute special cases.

Along with the hybrid controller, the application of hybrid process models, namely semi-parametric hybrid models, is proposed since

i) Full advantage can be taken of the hybrid model qualities, i.e. its excellent prediction capabilities and its good extrapolation properties, for the tuning of the controller parameters with respect to desired specifications.

ii) The application of control designs that allow to use the nonlinear dynamic hybrid process model to its full capacity, such as Model Reference Control (MRC) or Generic Model Control (GMC), is enabled.

Two parameter identification schema are proposed: One with which the parameters of the process model and the controller are, on the bases of process data, simultaneously identified (resulting in synergy effects regarding their identification); and another for the tuning of the controller parameters beyond the characteristics that can be learned from the process data.

A control problem is taken from the area of bioprocesses, namely the control of biomass through the manipulation of the substrate feeding rate and the control of dissolved oxygen concentration through the manipulation of the stirring velocity is considered. In this particular case, the loop for the control of one quantity, namely the biomass concentration, is closed (for the gathering of the process data this control loop was open), wherefore coupling effects come into play. This case study further awards with highly non-linear kinetics and different time scales of the inherent dynamics, which

can result in bang-bang situations. Several possible combinations of hybrid control structures were investigated for this case study, the following could be observed:

i) The complex dynamics of the process could, as expected be captured by the hybrid process model.

ii) Even so no closed-loop process data were available for one of the quantities, identification and tuning of controllers could be accomplished utilizing the hybrid process model, which allows to close the loop. When closing the loop, i.e. considering that the case study is under control, good performance for the derived controllers was demonstrated.

iii) The coupling effects that came into play when closing the loop could be well accounted for by the tuning procedure.

iv) For those control structures which hold "true" structural knowledge, the parameter identification converged faster and the consistency of the minima obtained for various random initiations of the parameters was greater than in cases that the controller was purely based on ANNs, which is in agreement to findings made for hybrid models (Fiedler and Schuppert, 2008; Mogk *&* al., 2002; von Stosch *&* al., 2011a).

v) When set-point changes were applied, those control designs which incorporate the hybrid process model performed significantly better than other controllers. Controllers that incorporate structural knowledge still performed better than purely ANN based control.

vi) In cases that the set-point specifications reached into process regions that had not been explored during the tuning, bang-bang situations could be observed for some controllers, most distinctively for a pure ANN based controller.

vii) When severe disturbances were applied to the simulated study, it could be observed that of the hybrid process model based controllers, which had performed well before, exhibited a rather poor performance. The reason was identified to be due to controller structure. However it can be expected that the application of an on-line adaption schema (such as the one proposed in (Hussain *&* al., 2001)) would significantly improve the performance of this controller.
Those controllers that incorporated structural knowledge still showed a better overall performance than those that are based on ANNs only.

viii) In situations that were not covered during the controller tuning, the structured hybrid controller performed better when compared to pure ANN based controller. This is due to the better extrapolation properties of structured approaches (Fiedler and Schuppert, 2008; Mogk & al., 2002), and thus in direct analogy to the findings for hybrid models.

ix) As a result of the incorporation of structure into the controller, it is possible to assess, understand and interpret the controller functioning, wherefore prior to application the controller characteristics can be assessed.

6.6 Appendix

6.6.1 The simulation case – A Fed-batch Pichia pastoris cultivation

A fed-batch cultivation of *Pichia pastoris* is considered, that comprises the macroscopic material balance equations of biomass, X^m, substrate, S^m, and dissolved oxygen, C_O^m, derived for an ideally mixed fed-batch reactor. The set of balance equation reads:

$$\frac{dX^m}{dt} = \mu^m(S^m, C_O^m) \cdot X^m - D \cdot X^m, \tag{6.17}$$

where μ^m is the specific biomass growth and D is the dilution rate;

$$\frac{dS^m}{dt} = -qs(S^m, C_O^m) \cdot X^m - D \cdot (S^m - S_F), \tag{6.18}$$

with q_S being the specific substrate uptake rate, and S_F being the methanol concentration in the feeding, 791.8 (g/l);

$$\frac{dC_O^m}{dt} = kLa \cdot (C_O^* - C_O^m) - q_O \cdot X^m - D \cdot C_O^m, \tag{6.19}$$

where C_O^* is the saturated dissolved oxygen concentration,0.008 (mg/l), kLa is the oxygen transfer coefficient defined below (1/h) and q_O is the specific oxygen uptake rate.

The specific biomass growth is in this study modeled as:

$$\mu^m = (q_S - q_M) \cdot Y_X, \tag{6.20}$$

with q_M being a term assigned to the biomass maintenance, 0.013 (1/h), Y_X is the biomass yield on substrate, 0.36 (-), and q_S is the substrate uptake rate which is defined as:

$$q_S = q_{S,Max} \cdot \frac{S^m}{K_S + S^m + S^{m2}/K_I} \cdot \frac{C_O^m}{K_O + C_O^m}. \tag{6.21}$$

Therein $q_{S,Max}$ is the maximum possible uptake rate, 0.8 (1/h), K_s is 0.4 (g/l), K_I is 5 (g/l) and the last term ensures that substrate uptake can only be accomplished if sufficient dissolved oxygen is present in the broth where the constant related to the limiting concentration, K_O, is 0.0004 (mg/l). The oxygen uptake rate is according to (Jahic & al., 2002) governed by two factors, i.e. the biomass growth rate (anabolic metabolism) and the energy metabolism given by the expression;

$$q_O = (Y_{OS,an} - Y_{OS,en}) \cdot \mu^m \cdot 0.96/0.375 + Y_{OS,en} \cdot q_S, \tag{6.22}$$

where $Y_{OS,an}$ is the methanol consumption per methanol in the anabolism 0.5 ($g_{O2}/g_{methanol}$) and $Y_{OS,en}$ is the conversion coefficient for the methanol flux to the energy metabolism, 1.5 ($g_{O2}/g_{methanol}$). The *kla* in the oxygen transfer rate is modeled through a linear relation with the stirring speed *RPM*, according to (Cunha & al., 2004) and such:

$$kla = 0.82 \cdot RPM - 74.2. \tag{6.23}$$

The dilution rate is $D = u/V$, where u is the manipulated methanol feeding rate and V is the volume of the broth in the reactor. The volume is obtained through:

$$\frac{dV}{dt} = u; \tag{6.24}$$

and the feeding rate is, in the case of data generation, obtained from open-loop control, while obtained from the hybrid control methodologies proposed in this study for the investigation of the simulated process under control.

For the control it is assumed that the only on-line measurable state variables are the dissolved oxygen concentration and the biomass concentration. Additionally, a 20

seconds time-lag of the sensor signal was incorporated in order to account for the sensor dynamics, see e.g. (Soons & al., 2006). The sensor dynamics are given by:

$$\frac{dC_O^{Sensor}}{dt} = \frac{C_O^m - C_O^{Sensor}}{\tau_{C_O}^{Sensor}}, \tag{6.25}$$

where C_O^{Sensor} is the dissolved oxygen concentration sensor value, and $\tau_{C_O}^{Sensor}$ is the time constant of the sensor lag, 20 seconds. The stirring velocity is then obtained through a PI controller:

$$RPM = K_P \cdot (C_{OS} - C_O^{Sensor}) + K_I \cdot \int_t (C_{OS} - C_O^{Sensor}) \cdot d\tau, \tag{6.26}$$

which has been tuned using the more conservative Tyreus-Luyben ratios for ultimate gain and frequency, instead of the the Ziegler-Nichols ratios (Seborg & al., 2010), resulting in slight overshoots only when confronted with step changes of the set-point.

The sensor dynamics for the biomass concentration sensor are modeled similarly to the dissolved oxygen concentration sensor, i.e.:

$$\frac{dX^{Sensor}}{dt} = \frac{X^m - X^{Sensor}}{\tau_X^{Sensor}}, \tag{6.27}$$

where X^{Sensor} is the biomass concentration sensor value, and τ_X^{Sensor} is the time constant of the sensor lag, 20 seconds. The initial values, marked with the index "0", are $X_0 = 70$ (g/l), $S_0 = 0.5$ (g/l), $C_{O,0} = 0016.$ (mg/l), $V_0 = 18$ (l), $u_0 = 0.1$ (l/h), $C_{O,0}^{Sensor} = C_{O,0}$ and $X_0^{Sensor} = X_0$. In order to have batch-to-batch variations in the case of data generation the initial values were randomly varied by 10%. The data were corrupted with 7% Gaussian noise..

6.6.2 Sensitivities Equations

The gradients are obtained when differentiating eq. (6.8), with respect to w,

$$\frac{dE_{LS}}{dw} = \frac{1}{P_c} \sum_{P_c} \frac{2 \cdot (c_{mes}(t) - c(t))}{c_\sigma^2} \cdot \frac{dc}{dw} + \frac{1}{P_u} \sum_{P_u} \frac{2 \cdot (u_{mes}(t) - u(t))}{u_\sigma^2} \cdot \frac{du}{dw}, \tag{6.28}$$

and accordingly differentiating eq. (6.8), with respect to θ,

$$\frac{dE_{LS}}{d\theta} = \frac{1}{P_c}\sum_{P_c}\frac{2\cdot(c_{mes}(t)-c(t))}{c_\sigma^2}\cdot\frac{dc}{d\theta} + \frac{1}{P_u}\sum_{P_u}\frac{2\cdot(u_{mes}(t)-u(t))}{u_\sigma^2}\cdot\frac{du}{d\theta}. \quad (6.29)$$

The gradients, dc/dw, $dc/d\theta$, du/dw and $du/d\theta$, in eqs. (A6.28) and (A6.29) are the sensitivities equations which are obtained in two steps:

(i) At first, eqs (6.1) and (6.4) are differentiated with respect to w and θ, making use of the total derivatives, i.e.:

$$\frac{d}{dt}\frac{dc}{dw} = \frac{\partial r}{\partial c}\cdot\frac{dc}{dw} + \frac{\partial r}{\partial u}\cdot\frac{du}{dw} + \frac{\partial r}{\partial w} - D\cdot\frac{dc}{dw} - c\cdot\frac{\partial D}{\partial u}\cdot\frac{du}{dw}; \quad (6.30)$$

$$\frac{d}{dt}\frac{dc}{d\theta} = \frac{\partial r}{\partial c}\cdot\frac{dc}{d\theta} + \frac{\partial r}{\partial u}\cdot\frac{du}{d\theta} - D\cdot\frac{dc}{d\theta} - c\cdot\frac{\partial D}{\partial u}\cdot\frac{du}{d\theta} \quad (6.31)$$

$$\frac{d}{dt}\cdot\frac{du_j}{dw} = \frac{\partial g_j}{\partial c}\cdot\frac{dc}{dw} + \frac{\partial g_j}{\partial u}\cdot\frac{du}{dw} + \frac{\partial g_j}{\partial E_P}\cdot\frac{dE_P}{dw} + \frac{\partial g_j}{\partial E_I}\cdot\frac{dE_I}{dw} + \frac{\partial g_j}{\partial E_D}\cdot\frac{dE_D}{dw}; \quad (6.32)$$

$$\frac{d}{dt}\cdot\frac{du_j}{d\theta} = \frac{\partial g_j}{\partial c}\cdot\frac{dc}{d\theta} + \frac{\partial g_j}{\partial u}\cdot\frac{du}{d\theta} + \frac{\partial g_j}{\partial E_P}\cdot\frac{dE_P}{d\theta} + \frac{\partial g_j}{\partial E_I}\cdot\frac{dE_I}{d\theta} + \frac{\partial g_j}{\partial E_D}\cdot\frac{dE_D}{d\theta} + \frac{\partial g_j}{\partial \theta}; \quad (6.33)$$

where

$$\frac{dE_D}{dw} = -\frac{d}{dt}\frac{dc}{dw}; \quad (6.34)$$

$$\frac{dE_P}{dw} = -\frac{dc}{dw}; \quad (6.35)$$

$$\frac{dE_I}{dw} = \int_t -\frac{dc}{dw}\cdot d\tau; \quad (6.36)$$

when (a) assuming that the set-point is independent of the model and the controller parameters; and (b) choosing c_X to be c. In the case that c_X is chosen to be c_{mes},

the gradients in eqs. (A6.34), (A6.35) and (A6.36) become zero. In case that for the controller eq. (6.3) is used instead of eq. (6.4), the derivative with respect to time in eqs. (A6.32) and (A6.33) is neglected.

(ii) Secondly, the derived eqs. (A6.30), (A6.31), (A6.32) and (A6.33) are numerically integrated along with the model and controller eqs. (6.1) and (6.4). The initial values of eqs. (A6.30), (A6.31), (A6.32) and (A6.33) are assumed to be zero, since the initial values of the concentrations and the initial values of the feeding rates are independent of the model and controller parameters. The integration of eqs. (6.1), (6.4), (A6.30), (A6.31), (A6.32) and (A6.33) is in this study carried out with a time-inexpensive Euler integration scheme that can be easily fitted to the sampling frequency, while the associate numerical integration error is sufficiently small not to affect the predictions.

6.7 Acknowledgment

The financial support of the Fundação para a Ciência e a Tecnologia (reference scholarship no.: SFRH / BD / 36990 / 2007) is most gratefully acknowledged.

6.8 Nomenclature

Abbreviations	
AIC	Akaike Information Criterion
ANN	Artificial Neural Network
BIC	Bayesian Information Criterion
DO	Dissolved Oxygen concentration
GMC	Generic Model Control
IE	Integrated Error
IAE	Integrated Absolute Error
ITE	Integral Time and Error
ITAE	Integral Time and Absolute Error
MRC	Model Reference Control
MSE	Mean Squared Error
ODE	Ordinary Differential Equation
PID	Proportional-Integral-Differential
Mathematical Symbols	
b_1	Bias of the input layer
b_2	Bias of the hidden layer
c	Vector of concentrations
$c_{desired}(t)$	Desired controller response for controller tuning
c_s	Vectoe of set-points
$c_{\sigma,j}$	Standard deviations

c_{mes}	Off-line measured concentration values
c_X	Concentration inputs to the controller
$g(\cdot)$	Vector of controller functions
$h(\cdot)$	Transfer function of the hidden layer
i	Counter
j	Counter
kLa	Oxygen Transfer Coefficient - Simulation model
q_S	Specific methanol uptake rate - Simulation model
$q_{S,Max}$	Maximum possible methanol uptake rate
q_O	Specific oxygen uptake rate - Simulation model
q_M	Biomass maintenance- Simulation model
r	Vector of kinetic rates
r_O	Oxygen uptake rate
r_T	Oxygen transfer rate
t	Time
u	Control inputs
$u_{mes,1..P_u}$	Measured control input
u_{Met}	Methanol feeding rate
u_σ	Standard deviation of the control input
$u_{1,A}$	ANN-based controller
$u_{2,A}$	Dynamic ANN-based controller
$u_{1,B}$	First order Model Reference Controller
$u_{2,B}$	Second order Model Reference Controller
$u_{1,C}$	Generic Model Controller
$u_{2,C}$	Generalized PID Controller
w	Vector of ANN parameters
w_1	Weights of the input layer
w_2	Weights of the hidden layer
C_O	Dissolved Oxygen concentration
C_O^m	Dissolved Oxygen concentration - Simulation model
C_O^*	Saturated dissolved oxygen concentration - Simulation model
C_{OS}	Dissolved Oxygen concentration set-point
C_O^{Sensor}	Dissolved oxygen concentration sensor value - Simulation model
D	Dilution rate
E_{LS}	Weighted least squared criteria
E_{NMSSE}	Normalized Mean Sum of Squared Error
E_D	Derrivative error
E_I	Integral error
E_P	Proportional error
J	Total number of control inputs
K	Matrix of stoichiometric coefficients
K_i	GMC integral parameter
K_p	GMC proportional parameter
K_s	Simulation model parameter
K_O	Simulation model parameter
K_I	Simulation model parameter
L_X	Inputs to the controller

P_c	Number of concentration samples
P_s	Number of data points considered for controller tuning
P_u	Number of input control samples
PI	Proportional Integral Controller
RPM	Stir velocity
S^m	Substrate concentration - Simulation model
S_F	Substrate feeding rate concentration - Simulation model
V	Reactor volume
X	Biomass concentration
X_S	Biomass concentration set-point
X^m	Biomass concentration - Simulation model
X^{Sensor}	Biomass concentration sensor value- Simulation model
$Y_{OS,an}$	Methanol consumption per methanol in the anabolism - Simulation model
$Y_{OS,en}$	Conversion coefficient for the methanol flux to the energy metabolism- Simulation model
Y_X	Biomass yield on substrate
β_1	MRC parameter
β_2	MRC parameter
γ_1	MRC parameter
γ_2	MRC parameter
ϖ_1	Bias of the linear layer
ϖ_2	Bias of the hidden layer
θ	Controller Parameters
θ_1	Weights of the linear layer
θ_2	Weights of the hidden layer
ϑ	GMC tuning parameter
Γ	Ratio coefficients
ϕ	Mechanistic term
φ	Structural knowledge terms
ϱ	Unknown controller terms
ρ	Unknown nonparametric terms
μ	Specific biomass growth rate
μ^m	Specific biomass growth rate - Simulation model
τ_c	MRC parameter
$\tau_{C_O}^{Sensor}$	Time constant of the sensor lag- Simulation model
τ_X^{Sensor}	Time constant of the sensor lag- Simulation model
ζ	GMC tuning parameter

Chapter 7

Conclusion and Prospects

7.1 Conclusion

In this thesis, it was attempted to develop the topic of hybrid modeling of biological systems and to develop new strategies for process monitoring and control based on hybrid models. The strategy pursued in this thesis consisted first the development of novel hybrid structures that go deeper in the description of complex intracellular dynamics, aligned with the new developments in systems biology. Also novel hybrid model structures particularly suitable for modeling highly dimensional multivariate data sets were developed as an effort to answer to the challenges posed by emergent measurement technologies based on optical probes for on-line spectral data acquisition. Then the application of such hybrid structures for process monitoring and closed-loop control was pursued under a practical viewpoint, wherein the focus has been on improving widespread monitoring and control strategies, such as PID, by the incorporation of hybrid models, resulting for instance in improved hybrid PID algorithms.
As first note, throughout this thesis it was observed that the careful combination of different sources of knowledge can lead to outstanding hybrid modeling results, but it was also observed that the utilization of a hybrid approach is not an automatic guaranty for good model performance. Regarding this matter, the following considerations about model structure are worthwhile to notice :

i) When the true system can be decomposed into a static nonlinear and a dynamic part, then the formulation of a parallel hybrid model, whose structure is equivalent to the one of the system, is, in principle, to prefer over a serial structure.

ii) The parallel approach seems also preferable when the structural uncertainty of the phenomenological/mechanistic model is high, since the nonlinear parallel model can, at least partially, account for the structural mismatch.

iii) In case of the serial approach, it was observed that when changing the structure of the model formulation, the performance peaks when the model structure is the closest to the one of the true system (Corazza & al., 2005). This also concerns the consideration of inherent dynamic effects by the structure, which results into a better coherence with the process underlying cell system (von Stosch & al., 2010). Also the partition of the reaction rate function by a mixture of experts approach relies on this principle (Peres & al., 2008).

iv) The model structure can, however, also be a source for error propagation. Two scenarios were distinguished, namely (1) errors inherent to the feedback nature of the models (where applicable); (von Stosch & al., 2011b) or (2) defective initial values (Vande Wouwer & al., 2004; von Stosch & al., 2011b).

v) Other sources for estimation errors were identified to be due to (1) noise in the input measurements to the nonparametric model; (2) noise in the measurements of the feeding rates (Chabbi & al., 2008; Schubert & al., 1994a; von Stosch & al., 2011b);

vi) The integration of phenomenological/ mechanistic knowledge can structure the space of operation and it can reduce the curse of dimension (Fiedler and Schuppert, 2008; Mogk & al., 2002).

vii) A result of the structured operation space is that the hybrid models have better calibration properties than pure nonparametric models, i.e. when training a serial hybrid model and a pure nonparametric with the same data, then the hybrid model will, in principle, perform better, meaning that the estimations are more accurate and the intra- and extrapolation properties are better.
The better calibration properties of hybrid models translate directly into lower requirements on the experimental data.

viii) The utilization of the "true" structure and integration of further knowledge can moreover result into better extrapolation properties (Fiedler and Schuppert, 2008) than when no knowledge at all is integrated. The extrapolation properties are especially enhanced in those cases where the model extrapolation relies on the mechanistic parts (van Can & al., 1996, 1998).

ix) In certain cases of knowledge integration even Bounded Input Bounded Output stability can be guaranteed (Karama & al., 2010; Oliveira, 2004). One prerequisite therefore is e.g. the consideration that a reaction can only occur when all educts are present.

x) Another positive consequence of structural knowledge incorporation is that the model is more transparent and therefore can be analyzed, which is especially interesting in cases such as the optimization of the control policy.

These points served to properly address the general properties of hybrid models, especially the ones of serial hybrid models.
When developing a bioprocess hybrid model it is important to realize that the source of complexity of observed dynamics lies mostly in the cellular system. It is common to observe very complex dynamic patterns, such as delay dynamics, without any apparent rational to explain such observations. Thus this thesis has put particular emphasis on the development of hybrid structures that display better coherence with the nature of cell systems. Regarding the hybrid methodologies which were developed to display better coherence with the underlying cell-system, the following main conclusions can be stated:

i) Intrinsic cell dynamics can mathematically be accounted for by the incorporation of time delays into the modeling framework. Therefore the integration of a discrete delay approach into the hybrid modeling framework was proposed. It was observed that the delay constitutes an important property of the cell systems and that the dynamics could only be captured when it was accounted for the delay. The coherence between the experimental data and the model estimations are found to be the better the closer the model delay is to the "true" system's delay. The model performance peaked when both delays coincided, which theoretically renders possible the identification of the system's underlying delay. Further, it could be seen that the hybrid approach, even so basing on a discrete delay formulation, is not limited to discrete delay dynamics, but also can deal with e.g. distributed delay dynamics.

ii) The cellular activity is highly sensitive to its environment. The environmental conditions are tried to be captured by several sensors, e.g. pH, temperature, dissolved oxygen concentration sensors or spectroscopic devices such as Near InfraRed. In this regard, the high number of correlated data that become available at each time instant have to be processed by adequate tools. For this purpose a Nonlinear Partial Least Square (NPLS) alike model was incorporated into a serial hybrid model. When evaluating the proposed methodology against a static Partial Least Square (PLS)

model using experimental data of a *Bordetella pertussis* batch cultivation it was observed that, in general, better estimations were obtained with lower numbers of involved parameters, in favor of the former. When comparing the hybrid methodology, which is inherently dynamic, to standard dynamic (N)PLS formulations similar observations were made. An additional feature of PLS, which is restored by the hybrid NPLS model, is the opportunity to analyze the score values, e.g. for fault diagnosis.

Of course the quest for models and hybrid model structures that display better consistency with the cell-system is not over, if it ever will be, but with these hybrid methodologies two dynamic modeling solutions are provided, which address eminent problems, namely intrinsic cell dynamics and the cell environment interaction.
The better consistency of the model with the system is, in general, of advantage also for model-based process control. It was already mentioned that there exist two ways to profit from a process model, namely to use a control schema that is based on the model directly or to exploit the model for the tuning of the controller parameters. Both scenarios have been studied. Additionally, it was investigated whether the incorporation of structure into the controller equation gives similar advantages as those observed for hybrid models. From this study, it was observed that:

i) For those control structures which hold structural knowledge, the parameter identification converged faster and the consistency of the minima obtained for various random initiations of the parameters was greater than in cases that the controller was purely based on ANNs.

ii) The controllers with incorporated structural knowledge performed better than purely ANN based control, when e.g. confronted with set-point changes.

iii) In situations that were not covered during the controller tuning, the structured hybrid controller performed better when compared to a pure ANN based controller.

iv) As a result of the incorporation of structure into the controller, it is possible to understand the controller functioning, wherefore prior to application the controller characteristics can be assessed.

Other general observations that could be made when exploiting hybrid models for process control are:

i) Through the application of the hybrid process model it was possible to close the loop for the control of biomass. The closing of the loop resulted into coupling effects

with the closed-loop control of the dissolved oxygen concentration. Those coupling effects could *a priori* be accounted for by the controller tuning.

ii) In case of model based control structures, in general good control performance could be achieved, but one of the controllers was observed to be less suitable. In this particular case, it was concluded that the model mismatch due to the controller structure was amplified, leading to poor performances when model uncertainty is high.

All in all, it can be said that both the controller and the model performance depend highly on the availability of quantitative measurement values, at best with a, relative to the process, high sampling frequency. The fusion of the data and other process knowledge, in a hybrid sense, can then be used to maximally exploit these data and in turn provide an optimal base for process control and process optimization. However, and as mentioned above, the application of an hybrid approach is not a guarantee for better performance than when each knowledge source is exploited on there own, but the advantages that can be achieved through careful knowledge fusion are, in general, manifold.

7.2 Prospects

The advantages that hybrid modeling can offer for process monitoring and control are striking, however in comparison to other modeling approaches, it still lives in the shadows. One reason might be that, at a first sight, the development of a hybrid model is rather unappealing, since several details have to be considered before a more or less readily first modeling approach is obtained. A commercial software that actively supports the user during the whole modeling procedure, while enabling a maximum of flexibility to interact with other software and different kind of sources, would for sure push things forward.

The amount of process data that is available both, at-time and *a posteriori* has significantly risen in the last decades (Schuegerl, 2001). Further, mechanistic/ phenomenological knowledge about the underlying physical system becomes increasingly available since various areas of science focus on the investigation of specific details, e.g. Systems Biology or Molecular Thermodynamics. Those evolutions can be expected to continue and thus the amount of knowledge that is available will increase dramatically in the upcoming years. Therefore it is important to develop and evaluate interfaces and fusion techniques to obtain a hybrid framework in which all the knowledge sources can

optimally be linked. This also addresses the combination of knowledge form different scales/ levels (Parrott, 2011). It was for instance shown by Teixeira & al. (2007b) that through the use of hybrid modeling, process engineering and systems biology can be linked. When these links are established then information might become redundant and reconciliation techniques can find application to evaluate the information. Modeling approaches that critically question themselves (a form of intelligence that is usually associated with scientists) might be developed. This can help the user to understand the limitations of the overall system's representation and to overcome those where desired. Ultimately, scenarios where the hybrid model is self-evolving in order to be optimal at all time could be imagined, as e.g. envisaged by Patnaik (2009). By doing so it might become more and more possible to manage the complexity of all kind of systems.

The application of hybrid models is promoted in several areas. Just recently the value added by hybrid modeling to the PAT framework was outlined by Gernaey and Gani (2010); Glassey & al. (2011); Teixeira & al. (2009). As a matter of fact, the requirements on the "PAT tools" read as the list of hybrid model advantages. Further, the pharmaceutical industry could immensely profit from the integration of hybrid modeling at all development stages of the pharmaceutical process, and also of utilization of hybrid modeling plant-wide (up-stream and down-stream).

Another emerging area for the application of hybrid models is systems biology, see for instances Carinhas & al. (2011). Hybrid modeling is attractive in this area since it can help to link the different scales of cell modeling and can account for unknown or uncertain parts.

The integration of hybrid models into complex flowsheets for (bio)chemical processes and the resulting overall representation of the plant, is envisaged as a consequence of the publication by Fiedler and Schuppert (2008). The advantages, such as the possibility to (further) optimize the plant set-points or the opportunity to achieve better closed-loop control performance, are obvious.

In general the application of hybrid models is advisable, whenever different sources of knowledge exist for a given process. In order to achieve the best possible results, a software tool should be developed that actively supports the user, while offering a maximum on flexibility for source and structure integration.

.

References

Abonyi, J., Chovan, T., Nagy, L. and Szeifert, F. (1999), 'Hybrid convolution model and its application in predictive ph control', *Comput. Chem. Eng.*, **23**(Supplement 1), S227–S230, URL http://www.sciencedirect.com/science/article/pii/S0098135499800568.

Abonyi, J., Madar, J. and Szeifert, F. (2007), 'Combining first principles models and neural networks for generic model control', .

Acuna, G., Cubillos, F., Thibault, J. and Latrille, E. (1999), 'Comparison of methods for training grey-box neural network models', *Comput. Chem. Eng.*, **23**(Supplement 1), S561–S564, URL http://www.sciencedirect.com/science/article/B6TFT-50KFCYJ-4X/2/6c130940ffe03f93082b616cd504d694.

Agarwal, M. (1997), 'Combining neural and conventional paradigms for modelling,prediction and control', *International Journal of Systems Science*, **28**(1), 65–81, URL http://www.informaworld.com/10.1080/00207729708929364.

Aguiar, H. C. and Filho, R. M. (2001), 'Neural network and hybrid model: a discussion about different modeling techniques to predict pulping degree with industrial data', *Chemical Engineering Science*, **56**(2), 565–570, URL http://www.sciencedirect.com/science/article/pii/S000925090000261X.

Akesson, M., Hagander, P. and Axelsson, J. (2001), 'Probing control of fed-batch cultivations: analysis and tuning', *Control Engineering Practice*, **9**(7), 709–723, URL http://www.sciencedirect.com/science/article/pii/S096706610100020X.

Al-Yemni, M. (2003), *Hybrid neural networks models for a membrane reactor*, Master Thesis, West Virginia University, United States – West Virginia.

Al-Yemni, M. and Yang, R. Y. K. (2005), 'Hybrid neural-networks modeling of an enzymatic membrane reactor', *Journal of the Chinese Institute of Engineers*, **28**(7), 1061–1067, URL http://www.informaworld.com/10.1080/02533839.2005.9671083.

Alford, J. S. (2006), 'Bioprocess control: Advances and challenges', *Comput. Chem. Eng.*, **30**(10-12), 1464–1475, URL http://www.sciencedirect.com/science/article/B6TFT-4KJ0SRH-3/2/693c961c2f1775cfb8ed02de8b31401e.

Anderson, J. S., McAvoy, T. J. and Hao, O. J. (2000), 'Use of hybrid models in wastewater systems', *Industrial & Engineering Chemistry Research*, **39**(6), 1694–1704, URL http://dx.doi.org/10.1021/ie990557r.

Andrasik, A., Meszaros, A. and de Azevedo, S. F. (2004), 'On-line tuning of a neural pid controller based on plant hybrid modeling', *Computers & Chemical Engineering*, **28**(8), 1499–1509, URL http://www.sciencedirect.com/science/article/B6TFT-4BGW2V0-2/2/7c1d1eabee766f1277907d2a1db48207.

Arahal, M. R., Cirre, C. M. and Berenguel, M. (2008), 'Serial grey-box model of a stratified thermal tank for hierarchical control of a solar plant', *Solar Energy*, **82**(5), 441–451, URL http://www.sciencedirect.com/science/article/pii/S0038092X07002186.

ASTM (2005), *ASTM E1655 - 05 Standard Practices for Infrared Multivariate Quantitative Analysis*, ASTM International.

Astroem, K. J. and Haegglund, T. (1995), *PID Controllers: Theory, Design, and Tuning, 2nd Edition*, Instrument Society of America.

Astrom, K. J. and Hagglund, T. (2001), 'The future of pid control', *Control Engineering Practice*, **9**(11), 1163–1175, URL http://www.sciencedirect.com/science/article/B6V2H-44GH62H-2/2/20fa533ef25d09610d1e2a00c42971af.

Azlan Hussain, M. (1999), 'Review of the applications of neural networks in chemical process control – simulation and online implementation', *Artificial Intelligence in Engineering*, **13**(1), 55–68, URL http://www.sciencedirect.com/science/article/B6V1X-3VGTB2B-5/2/754be2af43f9cc7f3926aaff01bb4d64.

Baffi, G., Martin, E. B. and Morris, A. J. (1999), 'Non-linear projection to latent structures revisited (the neural network pls algorithm)', *Comput. Chem. Eng.*, **23**(9), 1293–1307, URL http://www.sciencedirect.com/science/article/B6TFT-3Y0RFPH-F/2/a9c5c529234cf3892b5a525b65511998.

Baffi, G., Martin, E. B. and Morris, A. J. (2000), 'Non-linear dynamic projection to latent structures modelling', *Chemom. Intell. Lab. Syst.*, **52**(1), 5–22, URL http://www.sciencedirect.com/science/article/B6TFP-40WDSNB-2/2/e76e08c21f9d57ba498cc3c209a88239.

Baldi, P. and Chauvin, Y. (1996), 'Hybrid modeling, hmm/nn architectures, and protein applications', *Neural Computation*, **8**(7), 1541–1565, URL http://dx.doi.org/10.1162/neco.1996.8.7.1541.

Bastin, G. and Dochain, D. (1990), *On line estimation and adaptive control of bioreactors.*, Elsevier, Amsterdam.

Bazaei, A. and Majd, V. J. (2003), 'Feedback linearization of discrete-time nonlinear uncertain plants via first-principles-based serial neuro-gray-box models', *Journal of Process Control*, **13**(8), 819–830, URL http://www.sciencedirect.com/science/article/pii/S0959152403000271.

Becker, T., Hitzmann, B., Muffler, K., PÃČÂűrtner, R., Reardon, K., Stahl, F. and Ulber, R. (2007), 'White biotechnology', in R. Ulber and D. Sell (Editors), *Advances in Biochemical Engineering/Biotechnology*, Springer Berlin / Heidelberg, volume 105, pages 249–293, URL http://dx.doi.org/10.1007/10_2006_036.

Bellos, G., Kallinikos, L., Gounaris, C. and Papayannakos, N. (2005), 'Modelling of the performance of industrial hds reactors using a hybrid neural network approach', *Chemical Engineering and Processing*, **44**(5), 505–515, URL http://www.sciencedirect.com/science/article/pii/S0255270104001503.

Beluhan, D. and Beluhan, S. (2000), 'Hybrid modeling approach to on-line estimation of yeast biomass concentration in industrial bioreactor', *Biotechnology Letters*, **22**(8), 631–635, URL http://dx.doi.org/10.1023/A:1005604926063.

Bhutani, N., Rangaiah, G. P. and Ray, A. K. (2006), 'First-principles, data-based, and hybrid modeling and optimization of an industrial hydrocracking unit', *Industrial & Engineering Chemistry Research*, **45**(23), 7807–7816, URL http://dx.doi.org/10.1021/ie060247q.

Bishop, C. (1995), *Neural Networks for Pattern Recognition*, Oxford University Press Inc, New York.

Boareto, Ã. J. M., De Souza, M. B., Valero, F. and Valdman, B. (2007), 'A hybrid neural model (hnm) for the on-line monitoring of lipase production by candida rugosa', *J. Chem. Technol. Biotechnol.*, **82**(3), 319–327, URL http://dx.doi.org/10.1002/jctb.1678.

Bocharov, G. A. and Rihan, F. A. (2000), 'Numerical modelling in biosciences using delay differential equations', *J. Comput. Appl. Math.*, **125**(1-2), 183–199, URL http://www.sciencedirect.com/science/article/B6TYH-41TMSP6-H/2/f3292364e686cbfe8d6bf220527f1abd.

Bollas, G. M., Papadokonstadakis, S., Michalopoulos, J., Arampatzis, G., Lappas, A. A., Vasalos, I. A. and Lygeros, A. (2003), 'Using hybrid neural networks in scaling up an fcc model from a pilot plant to an industrial unit', *Chemical Engineering and Processing*, **42**(8-9), 697–713, URL http://www.sciencedirect.com/science/article/pii/S0255270102002064.

Bonvin, D. (1998), 'Optimal operation of batch reactorsÃćÂĂÂŤa personal view', *Journal of Process Control*, **8**(5-6), 355–368, URL http://www.sciencedirect.com/science/article/pii/S0959152498000109.

Bonvin, D. and Rippin, D. (1990), 'Target factor analysis for the identification of stoichiometric models', *Chemical Engineering Science*, **45**(12), 3417–3426, URL http://www.sciencedirect.com/science/article/pii/000925099087147K.

Braake, H. A. B. t., van Can, H. J. L. and Verbruggen, H. B. (1998), 'Semi-mechanistic modeling of chemical processes with neural networks', *Engineering Applications of Artificial Intelligence*, **11**(4), 507–515, URL http://www.sciencedirect.com/science/article/B6V2M-3WD6F2C-6/2/59da156eaf4ea5c7aa4582428b2e799c.

Brendel, M. and Marquardt, W. (2008), 'Experimental design for the identification of hybrid reaction models from transient data', *Chemical Engineering Journal*, **141**(1-3), 264–277, URL http://www.sciencedirect.com/science/article/pii/S1385894708000089.

Brereton, R. (2000), 'Introduction to multivariate calibration in analytical chemistry', *Analyst*, **125**, 2125–2154.

Burnham, K. and Anderson, D. (2004), 'Multimodel inference: Understanding aic and bic in model selection', *Sociol. Method Res.*, **33**(2), 261–304, URL http://www.scopus.com/inward/record.url?eid=2-s2.0-8744307994&partnerID=40&md5=63d4973884615ad6f9eb281946369a96.

Cao, M., Wang, K., Fujii, Y. and Tobler, W. (2004a), 'Advanced hybrid neural network automotive friction component model for powertrain system dynamic analysis. part 1: model development', *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, **218**(8), 831–843, URL http://dx.doi.org/10.1243/0954407041581147.

Cao, M., Wang, K. W., Fujii, Y. and Tobler, W. E. (2004b), 'A hybrid neural network approach for the development of friction component dynamic model', *J. Dyn. Sys., Meas., Control*, **126**(1), 144–153, URL http://link.aip.org/link/?JDS/126/144/1.

Caracotsios, M. and Stewart, W. E. (1985), 'Sensitivity analysis of initial value problems with mixed odes and algebraic equations', *Comput. Chem. Eng.*, **9**(4), 359–365, URL http://www.sciencedirect.com/science/article/pii/0098135485850146.

Carinhas, N., Bernal, V., Teixeira, A., Carrondo, M., Alves, P. and Oliveira, R. (2011), 'Hybrid metabolic flux analysis: combining stoichiometric and statistical constraints to model the formation of complex recombinant products', *BMC Systems Biology*, **5**(1), 34–, URL http://www.biomedcentral.com/1752-0509/5/34.

Ccopa Rivera, E., Mantovaneli, I., da Costa, A. C. and Maciel Filho, R. (2006), 'Hybrid modeling for continuous production of bioethanol', in W. Marquardt and C. Pantelides (Editors), *16th European Symposium on Computer Aided Process Engineering and 9th International Symposium on Process Systems Engineering*, Elsevier, volume Volume 21, pages 613–618, URL http://www.sciencedirect.com/science/article/pii/S1570794606801136.

Chabbi, C., Taibi, M. and Khier, B. (2008), 'Neural and hybrid neural modeling of a yeast fermentation process', *International Journal of Computational Cognition*, **6**(3), 42–47.

Chang, J.-S., Lu, S.-C. and Chiu, Y.-L. (2007), 'Dynamic modeling of batch polymerization reactors via the hybrid neural-network rate-function approach', *Chemical Engineering Journal*, **130**(1), 19–28, URL http://www.sciencedirect.com/science/article/pii/S1385894706005055.

Chang, W.-D., Hwang, R.-C. and Hsieh, J.-G. (2002), 'A self-tuning pid control for a class of nonlinear systems based on the lyapunov approach', *Journal of Process Control*, **12**(2), 233–242, URL http://www.sciencedirect.com/science/article/pii/S0959152401000415.

Chang, W.-D., Hwang, R.-C. and Hsieh, J.-G. (2003), 'A multivariable on-line adaptive pid controller using auto-tuning neurons', *Engineering Applications of Artificial Intelligence*, **16**(1), 57–63, URL http://www.sciencedirect.com/science/article/B6V2M-4899VGH-1/2/8795e29ca770cd7ffb1cb0467b09d9db.

Chen, B.-S. and Chang, Y.-T. (2008), 'A systematic molecular circuit design method for gene networks under biochemical time delays and molecular noises', *BMC Systems Biology*, **2**(1), 103, URL http://www.biomedcentral.com/1752-0509/2/103.

Chen, B.-S. and Chen, P.-W. (2009), 'On the estimation of robustness and filtering ability of dynamic biochemical networks under process delays, internal parametric perturbations and external disturbances', *Math. Biosci.*, **222**(2), 92–108, URL http://www.sciencedirect.com/science/article/B6VHX-4XB1TDV-1/2/48b72d47f410ce6bf9afd52d40a3d6d7.

Chen, J. and Huang, T.-C. (2004), 'Applying neural networks to on-line updated pid controllers for nonlinear process control', *Journal of Process Control*, **14**(2), 211–230, URL http://www.sciencedirect.com/science/article/B6V4N-49BY1WP-3/2/edb7917a3efe4157bd0ca8003c041646.

Chen, L., Bernard, O., Bastin, G. and Angelov, P. (2000), 'Hybrid modelling of biotechnological processes using neural networks', *Control Engineering Practice*, **8**(7), 821–827, URL http://www.sciencedirect.com/science/article/pii/S0967066100000368.

Chen, L., Hontoir, Y., Huang, D., Zhang, J. and Morris, A. J. (2004), 'Combining first principles with black-box techniques for reaction systems', *Control Engineering Practice*, **12**(7), 819–826, URL http://www.sciencedirect.com/science/article/pii/S0967066103002119.

Chorukova, E. and Simeonov, I. (2008), 'Neural and hybrid modelling of biotechnological process', *Studies In Informatics and Control*, **17**(3), 305–314.

Chungui, Z., Xinong, Z., Shilin, X., Tong, Z. and Changchun, Z. (2009), 'Hybrid modeling of wire cable vibration isolation system through neural network', *Mathematics and Computers in Simulation*, **79**(10), 3160–3173, URL http://www.sciencedirect.com/science/article/pii/S0378475409000767.

Clementschitsch, F. and Bayer, K. (2006), 'Improvement of bioprocess monitoring: development of novel concepts', *Microbial Cell Factories*, **5**(1), 19, URL http://www.microbialcellfactories.com/content/5/1/19.

Clifford, G. (2005), *Singular Value Decomposition & Independent Component Analysis for Blind Source Separation*, Technical Report, HST582J/6.555J/16.456J Biomedical Signal and Image Processing Spring (MIT).

Coleman, T. F. and Li, Y. (1994-10-01), 'On the convergence of interior-reflective newton methods for nonlinear minimization subject to bounds', *Math. Prog.*, **67**(1), 189–224–224, URL http://dx.doi.org/10.1007/BF01582221.

Conlin, J., Peel, C. and Montague, G. A. (1997), 'Modelling pressure drop in water treatment', *Artificial Intelligence in Engineering*, **11**(4), 393–400, URL http://www.sciencedirect.com/science/article/pii/S0954181096000581.

Corazza, F. C., Calsavara, L. P. V., Moraes, F. F., Zanin, G. M. and Neitzel, I. (2005), 'Determination of inhibition in the enzymatic hydrolysis of cellobiose using hybrid neural modeling', *Brazilian Journal of Chemical Engineering*, **22**, 19 – 29, URL http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0104-66322005000100003&nrm=iso.

Costa, A., Henriques, A., Alves, T., Maciel Filho, R. and Lima, E. (1999), 'A hybrid neural model for the optimization of fed-batch fermentations', *Brazilian Journal of Chemical Engineering*, **16**, 53 – 63, URL http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0104-66321999000100006&nrm=iso.

Costa, A. C., Alves, T. L. M., Henriques, A. W. S., Filho, R. M. and Lima, E. L. (1998), 'An adaptive optimal control scheme based on hybrid neural modelling', *Comput. Chem. Eng.*, **22**(Supplement 1), S859âĂŞS862, URL http://www.sciencedirect.com/science/article/B6TFT-4164VH5-9J/2/405c8cc29c5a28723e5f0e4edb3d0045.

Costa, R. S., Machado, D., Rocha, I. and Ferreira, E. C. (2010), 'Hybrid dynamic modeling of escherichia coli central metabolic network combining michaelis-menten and approximate kinetic equations', *Biosystems*, **100**(2), 150–157, URL http://www.sciencedirect.com/science/article/pii/S0303264710000377.

Cote, M., Grandjean, B. P. A., Lessard, P. and Thibault, J. (1995), 'Dynamic modelling of the activated sludge process: Improving prediction using neural networks', *Water Research*, **29**(4), 995–1004, URL http://www.sciencedirect.com/science/article/pii/004313549593250W.

Crowley, T., Harrison, C. and Doyle, I., F.J. (2001), 'Batch-to-batch optimization of psd in emulsion polymerization using a hybrid model', in *American Control Conference, 2001. Proceedings of the 2001 DOI - 10.1109/ACC.2001.945848*, volume 2, pages 981–986 vol.2.

Cubillos, F. and Acuna, G. (2007), 'Adaptive control using a grey box neural model: An experimental application', *Advances in Neural NetworksâĂŞLecture Notes in Computer Science*, **4491**, 311–318âĂŞ318, URL http://dx.doi.org/10.1007/978-3-540-72383-7_37.

Cubillos, F., Callejas, H., Lima, E. and Vega, M. (2001), 'Adaptive control using a hybrid-neural model: application to a polymerisation reactor', *Brazilian Journal of Chemical Engineering*, **18**, 113–120.

Cubillos, F. A., Alvarez, P. I., Pinto, J. C. and Lima, E. L. (1996), 'Hybrid-neural modeling for particulate solid drying processes', *Powder Technology*, **87**(2), 153–160, URL http://www.sciencedirect.com/science/article/pii/0032591095030832.

Cubillos, F. A. and Lima, E. L. (1997), 'Identification and optimizing control of a rougher flotation circuit using an adaptable hybrid-neural model', *Minerals Engineering*, **10**(7), 707–721, URL http://www.sciencedirect.com/science/article/pii/S0892687597000502.

Cubillos, F. A. and Lima, E. L. (1998), 'Adaptive hybrid neural models for process control', *Computers & Chemical Engineering*, **22**(Supplement 1), S989–S992, URL http://www.sciencedirect.com/science/article/pii/S0098135498001975.

Cunha, A. E., Clemente, J. J., Gomes, R., Pinto, F., Thomaz, M., Miranda, S., Pinto, R., Moosmayer, D., Donner, P. and Carrondo, M. J. T. (2004), 'Methanol induction optimization for scfv antibody fragment production in pichia pastoris', *Biotechnol. Bioeng.*, **86**(4), 458–467, URL http://dx.doi.org/10.1002/bit.20051.

Dabros, M., Schuler, M. and Marison, I. (2010), 'Simple control of specific growth rate in biotechnological fed-batch processes based on enhanced online measurements of biomass', *Bioprocess and Biosystems Engineering*, **33**(9), 1109–1118, URL http://dx.doi.org/10.1007/s00449-010-0438-2.

Dadhe, K., Rossmann, V., Durmus, K. and Engell, S. (2001), 'Neural networks as a tool for gray box modelling in reactive distillation', *Computational Intelligence. Theory and Applications*, **2206**, 576–588, URL http://dx.doi.org/10.1007/3-540-45493-4_58.

Daugulis, A. J., McLellan, P. J. and Li, J. (1997), 'Experimental investigation and modeling of oscillatory behavior in the continuous culture of zymomonas mobilis', *Biotechnol. Bioeng.*,

56(1), 99–105, URL http://dx.doi.org/10.1002/(SICI)1097-0290(19971005)56:1<99::AID-BIT11>3.0.CO;2-5.

Dochain, D. (2003), 'State and parameter estimation in chemical and biochemical processes: a tutorial', *Journal of Process Control*, **13**(8), 801–818, URL http://www.sciencedirect.com/science/article/pii/S095915240300026X.

Dors, M., Simutis, R. and Luebbert, A. (1995), 'Advanced supervision of mammarian cell cultures using hybrid process models', in A. Munack and K. Schugerl (Editors), *Preprints of the 6th International Conference on Computer Applications in Biotechnology*, pages 72–77.

Dors, M., Simutis, R. and Luebbert, A. (1996), 'Hybrid process modeling for advanced process state estimation, prediction, and control exemplified in a production-scale mammalian cell culture', in *ACS Symposium Series*, American Chemical Society, volume 613, pages 144–154–, URL http://dx.doi.org/10.1021/bk-1995-0613.ch014.

Doyle, F. J., Harrison, C. A. and Crowley, T. J. (2003), 'Hybrid model-based approach to batch-to-batch control of particle size distribution in emulsion polymerization', *Computers & Chemical Engineering*, **27**(8-9), 1153–1163, URL http://www.sciencedirect.com/science/article/pii/S0098135403000437.

Duarte, B., Saraiva, P. M. and Pantelides, C. C. (2004), 'Combined mechanistic and empirical modelling', *International Journal of Chemical Reactor Engineering*, **2**, A3, URL http://www.bepress.com/ijcre/vol2/A3.

Duarte, B. P. M. and Saraiva, P. M. (2003), 'Hybrid models combining mechanistic models with adaptive regression splines and local stepwise regression', *Industrial & Engineering Chemistry Research*, **42**(1), 99–107, URL http://dx.doi.org/10.1021/ie0107744.

Emmanuel, A. N., David, A., Benjamin, Y. K. and Yannick, E. T. (2009), 'Modelling fed-batch fermentation processes: An approach based on artificial neural networks', *Australian Journal of Basic and Applied Sciences*, **3**(4), 3930–3936.

Eslamloueyan, R. and Setoodeh, P. (2011), 'Optimization of fed-batch recombinant yeast fermentation for ethanol production using a reduced dynamic flux balance model based on artificial neural networks', *Chemical Engineering Communications*, **198**(11), 1309–1338, URL http://dx.doi.org/10.1080/00986445.2011.560512.

Eykhoff, ., Pieter (1974), *System identification:Parameter and State estimation*, Wiley-Interscience, London ; New York.

Feil, B., Abonyi, J., Nemeth, S., Nemeth, O., Arva, P., Nemeth, M. and Nagy, G. (2004), 'Semi-mechanistic models for state-estimation - soft sensor for polymer melt index prediction', in *Index Prediction, 7th International Conference on Artificial Intelligence and Soft Computing*, Springer,

volume 3070, pages 1111–1117, URL http://dx.doi.org/10.1007/978-3-540-24844-6_174.

Fellner, Delgado and Becker (2003), 'Functional nodes in dynamic neural networks for bioprocess modelling', *Bioprocess and Biosystems Engineering*, **25**(5), 263–270, URL http://dx.doi.org/10.1007/s00449-002-0297-6.

Ferreira, A., Dias, J., Clemente, J., Cunha, A. and Oliveira, R. (2011), 'A systems biology method for top-down gene-to-enviroment culture media engineering: Application to pichia pastoris cultures', in *11th international Chemical and Biological Engineering Conference, Lisbon, Portugal, 5-7 September*.

Feyo de Azevedo, S., Dahm, B. and Oliveira, F. (1997), 'Hybrid modelling of biochemical processes: A comparison with the conventional approach', *Computers & Chemical Engineering*, **21**(Supplement 1), S751–S756, URL http://www.sciencedirect.com/science/article/pii/S009813549787593X.

Feyo de Azevedo, S., Oliveira, R. and Sonnleitner, B. (2001), *Multiphase Bioreactor Design*, Taylor & Francis, chapter New Methodologies for Multiphase Bioreactors: Data Aquisition, Modeling and Control, pages 53–83.

Fiedler, B. and Schuppert, A. (2008), 'Local identification of scalar hybrid models with tree structure', *IMA Journal of Applied Mathematics*, **73**(3), 449–476, URL http://imamat.oxfordjournals.org/content/73/3/449.abstract.

Frank, P. M. (1978), *Introduction to system sensitivity theory*, Academic Press (New York).

Fu, P. C. and Barford, J. P. (1995a), 'A hybrid neural network–first principles approach for modelling of cell metabolism', *Comput. Chem. Eng.*, **20**(6-7), 951–958, URL http://www.sciencedirect.com/science/article/B6TFT-3TKMDRH-16/2/fc2ad74c3c41e03eb283cbcc03d51356.

Fu, P. C. and Barford, J. P. (1995b), 'Integration of mathematical modelling and knowledge-based systems for simulations of biochemical processes', *Expert Systems with Applications*, **9**(3), 295–307, URL http://www.sciencedirect.com/science/article/pii/095741749500006U.

Galvanauskas, V., Simutis, R. and Luebbert, A. (2004), 'Hybrid process models for process optimisation,monitoring and control', *Bioprocess Biosyst. Eng.*, **26**(6), 393âĂŞ400, URL http://dx.doi.org/10.1007/s00449-004-0385-x.

Genc, O. S. (2006), *Modelling and control of bioprocesses by using artificial neural networks and hybrid model*, Master Thesis, Graduate School of Engineering and Science of Izmir Institute of Technology.

Georgieva, P. and de Azevedo, S. (2009), 'Novel computational methods for modeling and control in chemical and biochemical process systems', URL http://dx.doi.org/10.1007/978-3-642-01888-6_4.

Georgieva, P. and Feyo de Azevedo, S. (2007), 'Neural network-based control strategies. applied to a fed-batch crystallization process', *World Academy of Science, Engineering and Technology*, **36**.

Georgieva, P., Meireles, M. and Feyo de Azevedo, S. (2003), 'Knowledge-based hybrid modelling of a batch crystallisation when accounting for nucleation, growth and agglomeration phenomena', *Chemical Engineering Science*, **58**(16), 3699–3713, URL http://www.sciencedirect.com/science/article/pii/S0009250903002604.

Gernaey, K. V. and Gani, R. (2010), 'A model-based systems approach to pharmaceutical product-process design and analysis', *Chemical Engineering Science*, **65**(21), 5757–5769, URL http://www.sciencedirect.com/science/article/pii/S000925091000285X.

Ghosh, P., Sundaram, A., Venkatasubramanian, V. and Caruthers, J. M. (2000), 'Integrated product engineering: a hybrid evolutionary framework', *Computers & Chemical Engineering*, **24**(2-7), 685–691, URL http://www.sciencedirect.com/science/article/pii/S0098135400004208.

Glassey, J., Gernaey, K. V., Clemens, C., Schulz, T. W., Oliveira, R., Striedner, G. and Mandenius, C.-F. (2011), 'Process analytical technology (pat) for biopharmaceuticals', *Biotechnology Journal*, **6**(4), 369–377, URL http://dx.doi.org/10.1002/biot.201000356.

Gnoth, S., Jenzsch, M., Simutis, R. and Luebbert, A. (2008a), 'Control of cultivation processes for recombinant protein production: a review', *Bioprocess Biosyst. Eng.*, **31**(1), 21–39, URL http://dx.doi.org/10.1007/s00449-007-0163-7.

Gnoth, S., Jenzsch, M., Simutis, R. and Luebbert, A. (2008b), 'Product formation kinetics in genetically modified e. coli bacteria: inclusion body formation', *Bioprocess Biosyst. Eng.*, **31**(1), 41–46–46, URL http://dx.doi.org/10.1007/s00449-007-0161-9.

Goncalves, L. R., Sousa, R., Fernandez-Lafuente, R., Guisan, J. M., Giordano, R. L. and Giordano, R. C. (2002), 'Enzymatic synthesis of amoxicillin: Avoiding limitations of the mechanistic approach for reaction kinetics', *Biotechnol. Bioeng.*, **80**(6), 622–631, URL http://dx.doi.org/10.1002/bit.10417.

Graefe, J., Bogaerts, P., Castillo, J., Cherlet, M., Werenne, J., Marenbach, P. and Hanus, R. (1999), 'A new training method for hybrid models of bioprocesses', *Bioprocess and Biosystems Engineering*, **21**(5), 423–429, URL http://dx.doi.org/10.1007/s004490050697.

Guclu, D. and Dursun, S. (2008), 'Amelioration of carbon removal prediction for an activated sludge process using an artificial neural network (ann)', *Clean Soil Air Water*, **36**(9), 781–787, URL http://dx.doi.org/10.1002/clen.200700155.

Guo, B., Li, D., Cheng, C., LÃij, Z.-a. and Shen, Y. (2001), 'Simulation of biomass gasification with a hybrid neural network model', *Bioresource Technology*, **76**(2), 77–83, URL http://www.sciencedirect.com/science/article/pii/S0960852400001061.

Guo, B., Shen, Y., Li, D. and Zhao, F. (1997), 'Modelling coal gasification with a hybrid neural network', *Fuel*, **76**(12), 1159–1164, URL http://www.sciencedirect.com/science/article/pii/S0016236197001221.

Gupta, S., Liu, P.-H., Svoronos, S. A., Sharma, R., Abdel-Khalek, N. A., Cheng, Y. and El-Shall, H. (1999), 'Hybrid first-principles/neural networks model for column flotation', *AIChE J.*, **45**(3), 557–566, URL http://dx.doi.org/10.1002/aic.690450312.

Gurden, S. P., Westerhuis, J. A., Bijlsma, S. and Smilde, A. K. (2001), 'Modelling of spectroscopic batch process data using grey models to incorporate external information', *J. Chemometrics*, **15**(2), 101–121, URL http://dx.doi.org/10.1002/1099-128X(200102)15:2<101::AID-CEM602>3.0.CO;2-V.

Hagan, M. and Demuth, H. (1999), 'Neural networks for control', in *American Control Conference, San Diego, June, 1642-1656.*.

Hahn, J., Lextrait, S. and Edgar, T. F. (2002), 'Nonlinear balanced model residualization via neural networks', *AIChE J.*, **48**(6), 1353–1357, URL http://dx.doi.org/10.1002/aic.690480621.

Hanomolo, A., Bogaerts, P., Graefe, J., Cherlet, M., Werenne, J. and Hanus, R. (2000), 'Maximum likelihood parameter estimation of a hybrid neural-classical structure for the simulation of bioprocesses', *Mathematics and Computers in Simulation*, **51**(3-4), 375–385, URL http://www.sciencedirect.com/science/article/pii/S0378475499001305.

Harada, L., da Costa, A. and Filho, R. (2002), 'Hybrid neural modeling of bioprocesses using functional link networks', *Applied Biochemistry and Biotechnology*, **98-100**(1), 1009–1023, URL http://dx.doi.org/10.1385/ABAB:98-100:1-9:1009.

Harms, P., Kostov, Y. and Rao, G. (2002), 'Bioprocess monitoring', *Curr. Opin. Biotech.*, **13**(2), 124–127, URL http://www.sciencedirect.com/science/article/B6VRV-45SHW2J-9/2/686a758e5b6cecd5160fcaa800009fd5.

Haykin, S. (1998), *Neural Networks - A Comprehensive Foundation (2nd. ed.).*, Prentice-Hall, Upper Saddle River, NJ.

Henneke, D., Hagedorn, A., Budman, H. and Legge, R. (2005), 'Application of spectrofluorometry to the prediction of phb concentrations in a fed-batch process', *Bioprocess Biosyst. Eng.*, **27**(6), 359–364, URL http://dx.doi.org/10.1007/s00449-004-0375-z.

Henriques, A. W. S., Costa, A. C., Alves, T. L. M. and Lima, E. L. (1999), 'Optimization of fed-batch processes : Challenges and solutions', *Brazilian Journal of Chemical Engineering*, **16**, 171–177.

Hinchliffe, M., Montague, G., Willis, M. and Burke, A. (2003), 'Hybrid approach to modeling an industrial polyethylene process', *AIChE J.*, **49**(12), 3127–3137, URL http://dx.doi.org/10.1002/aic.690491213.

Hinz, D. (2006), 'Process analytical technologies in the pharmaceutical industry: the fdas pat initiative', *Analytical and Bioanalytical Chemistry*, **384**(5), 1036–1042, URL http://dx.doi.org/10.1007/s00216-005-3394-y.

Hisbullah, Hussain and Ramachandran (2002), 'Comparative evaluation of various control schemes for fed-batch fermentation', *Bioprocess and Biosystems Engineering*, **24**(5), 309–318, URL http://dx.doi.org/10.1007/s00449-001-0272-7.

Hu, B.-G., Qu, H.-B., Wang, Y. and Yang, S.-H. (2009), 'A generalized-constraint neural network model: Associating partially known relationships for nonlinear regressions', *Information Sciences*, **179**(12), 1929–1943, URL http://www.sciencedirect.com/science/article/pii/S0020025509000875.

Hu, G., Mao, Z., He, D. and Yang, F. (2011), 'Hybrid modeling for the prediction of leaching rate in leaching process based on negative correlation learning bagging ensemble algorithm', *Computers & Chemical Engineering*, **In Press, Corrected Proof**, –, URL http://www.sciencedirect.com/science/article/pii/S0098135411000743.

Hussain, M. A. and Ho, P. Y. (2004), 'Adaptive sliding mode control with neural network based hybrid models', *Journal of Process Control*, **14**(2), 157âĂŞ176, URL http://www.sciencedirect.com/science/article/B6V4N-490H0YS-2/2/179c711a9aa529735dace307e27e2398.

Hussain, M. A., Ho, P. Y. and Allwright, J. C. (2001), 'Adaptive linearizing control with neural-network-based hybrid models', *Industrial & Engineering Chemistry Research*, **40**(23), 5604–5620, URL http://dx.doi.org/10.1021/ie000919r.

Hussain, M. A., Shafiur Rahman, M. and Ng, C. W. (2002), 'Prediction of pores formation (porosity) in foods during drying: generic models by the use of hybrid neural network', *Journal of Food Engineering*, **51**(3), 239–248, URL http://www.sciencedirect.com/science/article/pii/S0260877401000632.

Hwang, T.-M., Oh, H., Choi, Y.-J., Nam, S.-H., Lee, S. and Choung, Y.-K. (2009), 'Development of a statistical and mathematical hybrid model to predict membrane fouling and performance', *Desalination*, **247**(1-3), 210–221, URL http://www.sciencedirect.com/science/article/pii/S0011916409005013.

Ibrehem, A. S., Hussain, M. A. and Ghasem, N. M. (2011), 'Hybrid mathematical model and advanced control of a fluidized bed using a model-predictive controller', *Journal of Petroleum and Gas Engineering*, **2**(2), 25âĂŞ44.

Ignova, M., Paul, G. C., Kent, C. A., Thomas, C. R., Montague, G. A., Glassey, J. and Ward, A. C. (2002), 'Hybrid modelling for on-line penicillin fermentation optimisation', in *Proceedings of the 15th IFAC World Congress*.

Istadi and Amin, N. A. S. (2006), 'Hybrid artificial neural network - genetic algorithm technique for modeling and optimization of plasma reactor', *Industrial & Engineering Chemistry Research*, **45**(20), 6655–6664, URL http://dx.doi.org/10.1021/ie060562c.

Jahic, Jahic, M., Rotticci-Mulder, Rotticci-Mulder, J., Martinelle, Martinelle, M., Hult, Hult, K., Enfors and Enfors, S.-O. (2002), 'Modeling of growth and energy metabolism of pichia pastoris producing a fusion protein', *Bioprocess Biosyst. Eng.*, **24**(6), 385–393, URL http://dx.doi.org/10.1007/s00449-001-0274-5.

James, S., Legge, R. and Budman, H. (2002), 'Comparative study of black-box and hybrid estimation methods in fed-batch fermentation', *Journal of Process Control*, **12**(1), 113–121, URL http://www.sciencedirect.com/science/article/pii/S0959152400000652.

Jenzsch, M., Gnoth, S., Kleinschmidt, M., Simutis, R. and Luebbert, A. (2007), 'Improving the batch-to-batch reproducibility of microbial cultures during recombinant protein production by regulation of the total carbon dioxide production', *Journal of Biotechnology*, **128**(4), 858–867, URL http://www.sciencedirect.com/science/article/pii/S0168165607000326.

Jenzsch, M., Simutis, R. and Luebbert, A. (2006a), 'Generic model control of the specific growth rate in recombinant escherichia coli cultivations', *Journal of Biotechnology*, **122**(4), 483–493, URL http://www.sciencedirect.com/science/article/pii/S016816560500622X.

Jenzsch, M., Simutis, R. and Luebbert, A. (2006b), 'Optimization and control of industrial microbial cultivation processes', *Eng. Life Sci.*, **6**(2), 117–124, URL http://dx.doi.org/10.1002/elsc.200620901.

Jia, R.-d., Mao, Z.-z., Chang, Y.-q. and Zhao, L.-p. (2011), 'Soft-sensor for copper extraction process in cobalt hydrometallurgy based on adaptive hybrid model', *Chemical Engineering Research and Design*, **89**(6), 722–728, URL http://www.sciencedirect.com/science/article/pii/S0263876210002820.

Johansen, T. and Foss, B. (1992a), 'Nonlinear local model representation for adaptive systems', in *Intelligent Control and Instrumentation, 1992. SICICI '92. Proceedings., Singapore International Conference on DOI - 10.1109/SICICI.1992.637617*, volume 2, pages 677–682.

Johansen, T. A. and Foss, B. A. (1992b), 'Representing and learning unmodeled dynamics with neural network memories', in *American Control Conference, 1992 DOI -*, pages 3037–3043.

Jones, D. M., Watton, J. and Brown, K. J. (2007), 'Comparison of black-, white-, and grey-box models to predict ultimate tensile strength of high-strength hot rolled coils at the port talbot hot strip mill', *Proceedings of the Institution of Mechanical Engineers, Part L: Journal of Materials Design and Applications*, **221**(1), 1–9.

Junker, B. and Wang, H. (2006), 'Bioprocess monitoring and computer control: Key roots of the current pat initiative', *Biotechnol. Bioeng.*, **95**(2), 226–261, URL `http://dx.doi.org/10.1002/bit.21087`.

Kadlec, P., Gabrys, B. and Strandt, S. (2009), 'Data-driven soft sensors in the process industry', *Computers & Chemical Engineering*, **33**(4), 795–814, URL `http://www.sciencedirect.com/science/article/pii/S0098135409000076`.

Kahrs, O. and Marquardt, W. (2007), 'The validity domain of hybrid models and its application in process optimization', *Chemical Engineering and Processing: Process Intensification*, **46**(11), 1054–1066, URL `http://www.sciencedirect.com/science/article/pii/S0255270107001134`.

Kahrs, O. and Marquardt, W. (2008), 'Incremental identification of hybrid process models', *Computers & Chemical Engineering*, **32**(4-5), 694–705, URL `http://www.sciencedirect.com/science/article/pii/S009813540700049X`.

Kamali, M. and Mousavi, M. (2008), 'Analytic, neural network, and hybrid modeling of supercritical extraction of [alpha]-pinene', *The Journal of Supercritical Fluids*, **47**(2), 168–173, URL `http://www.sciencedirect.com/science/article/pii/S0896844608002714`.

Kansha, Y., Jia, L. and Chiu, M.-S. (2008), 'Self-tuning pid controllers based on the lyapunov approach', *Chemical Engineering Science*, **63**(10), 2732–2740, URL `http://www.sciencedirect.com/science/article/pii/S0009250908000997`.

Karama, A., Bernard, O., Genovesi, A., Dochain, D., Benhammou, A. and JP, S. (2001a), 'Hybrid modelling of anaerobic wastewater treatment processes', *Water Sci Technol*, **43**(1), 43–50.

Karama, A., Bernard, O., Gouze, J. L., Benhammou, A. and Dochain, D. (2001b), 'Hybrid neural modelling of an anaerobic digester with respect to biological constraints', *Water Sci Technol.*, **43**(7), 1–8.

Karama, A., Bernard, O. and GouzÃl', J.-L. (2010), 'Constrained hybrid neural modelling of biotechnological processes', *International Journal of Chemical Reactor Engineering*, **8**(A21), URL http://www.bepress.com/ijcre/vol8/A21.

Kasprow, R. K. (2000), *Hybrid Modeling (Neural Networks and First Principles) of Fermentation: Combining Biochemical Engineering Fundamentals and Process data*, PhD Thesis, University of Virginia, VA, USA.

Khatri, N. and Hoffmann, F. (2006), 'Oxygen-limited control of methanol uptake for improved production of a single-chain antibody fragment with recombinant pichia pastoris', *Appl. Microbiol. Biotechnol.*, **72**(3), 492–498, URL http://dx.doi.org/10.1007/s00253-005-0306-1.

Kim, B. and Kim, I. (2005), 'An application of hybrid least squares support vector machine to environmental process modeling', *Lecture Notes in Computer Science*, **3320**, 57–85, URL http://dx.doi.org/10.1007/978-3-540-30501-9_42.

Kim, B. and Kim, I. (2006), 'A novel hybrid system for dynamic control', *Lecture Notes in Computer Science*, **4247**, 609–616, URL http://dx.doi.org/10.1007/11903697_77.

Kim, H. and Chang, K. (2000), 'Hybrid neural network approach in description and prediction of dynamic behavior of chaotic chemical reaction systems', *Korean Journal of Chemical Engineering*, **17**(6), 696–703, URL http://dx.doi.org/10.1007/BF02699120.

Klimasauskas, C. C. (1998), 'Hybrid modeling for robust nonlinear multivariable control', *ISA Transactions*, **37**(4), 291âĂŞ297, URL http://www.sciencedirect.com/science/article/B6V3P-3W482N0-6/2/87ad71b83699a7081a24f7abf3c7e280.

Kramer, M. A., Thompson, M. L. and Bhagat, P. M. (1992), 'Embedding theoretical models in neural networks', in *American Control Conference, 1992*, pages 475 –479.

Krishnapura, V. G. and Jutan, A. (2000), 'A neural adaptive controller', *Chemical Engineering Science*, **55**(18), 3803–3812, URL http://www.sciencedirect.com/science/article/B6TFK-40GJ2MB-M/2/6b6db7615c6b8244736bd03868f2a481.

Kulkarni, S. G., Chaudhary, A. K., Nandi, S., Tambe, S. S. and Kulkarni, B. D. (2004), 'Modeling and monitoring of batch processes using principal component analysis (pca) assisted generalized regression neural networks (grnn)', *Biochem. Eng. J.*, **18**(3), 193–210, URL http://www.sciencedirect.com/science/article/B6V5N-4B5C2MW-1/2/ea127ebd70baa02f6f431ccf00be264f.

Kumar Akkisetty, P., Lee, U., Reklaitis, G. and Venkatasubramanian, V. (2010), 'Population balance model-based hybrid neural network for a pharmaceutical milling process', *Journal of Pharmaceutical Innovation*, **5**(4), 161–168, URL http://dx.doi.org/10.1007/s12247-010-9090-2.

Kumar Akkisetty, P., Reklaitis, G. V. and Venkatasubramanian, V. (2009), 'Ontological informatics based decision support system for pharmaceutical product development: Milling as a case study', in J. Jezowski and J. Thullie (Editors), *19th European Symposium on Computer Aided Process Engineering*, Elsevier, volume Volume 26, pages 159–164, URL http://www.sciencedirect.com/science/article/pii/S1570794609700276.

Kuprijanov, A., Gnoth, S., Simutis, R. and Luebbert, A. (2009), 'Advanced control of dissolved oxygen concentration in fed batch cultures during recombinant protein production', *Applied Microbiology and Biotechnology*, **82**(2), 221–229, URL http://dx.doi.org/10.1007/s00253-008-1765-y.

Lakshminarayanan, S., Shah, S. L. and Nandakumar, K. (1997), 'Modeling and control of multivariable processes: Dynamic pls approach', *AIChE J.*, **43**(9), 2307–2322, URL http://dx.doi.org/10.1002/aic.690430916.

Lauret, P., Boyer, H. and Gatina, J. C. (2000), 'Hybrid modelling of a sugar boiling process', *Control Engineering Practice*, **8**(3), 299–310, URL http://www.sciencedirect.com/science/article/pii/S0967066199001513.

Laursen, S. O., Webb, D. and Ramirez, W. F. (2007), 'Dynamic hybrid neural network model of an industrial fed-batch fermentation process to produce foreign protein', *Computers & Chemical Engineering*, **31**(3), 163–170, URL http://www.sciencedirect.com/science/article/pii/S0098135406001311.

Lee, D. S., Jeon, C. O., Park, J. M. and Chang, K. S. (2002), 'Hybrid neural network modeling of a full-scale industrial wastewater treatment process', *Biotechnol. Bioeng.*, **78**(6), 670–682, URL http://dx.doi.org/10.1002/bit.10247.

Lee, D. S., Vanrolleghem, P. A. and Park, J. M. (2005), 'Parallel hybrid modeling methods for a full-scale cokes wastewater treatment plant', *J. Biotechnol.*, **115**(3), 317–328, URL http://www.sciencedirect.com/science/article/B6T3C-4DTSKTT-1/2/4c7bd55729519fef16020bb04ed4ea18.

Lee, P. L. and Sullivan, G. R. (1988), 'Generic model control (gmc)', *Computers & Chemical Engineering*, **12**(6), 573–580, URL http://www.sciencedirect.com/science/article/pii/0098135488870066.

Leifsson, L., Saevarsdottir, H., Sigurosson, S. and Vesteinsson, A. (2008), 'Grey-box modeling of an ocean vessel for operational optimization', *Simulation Modelling Practice and Theory*, **16**(8), 923–932, URL http://www.sciencedirect.com/science/article/pii/S1569190X08000488.

Leonard, T. and Hsu, J. (1999), *Bayesian Methods*, Cambridge University Press, New York.

Levisauskas, D., Simutis, R., Borvitz, D. and Luebbert, A. (1996), 'Automatic control of the specific growth rate in fed-batch cultivation processes based on an exhaust gas analysis', *Bioprocess and Biosystems Engineering*, **15**(3), 145–150, URL http://dx.doi.org/10.1007/BF00369618.

Li, H., Guo, C. and Jin, H. (2006), 'Hybrid control of inverse model wavelet neural network and pid and its application to fin stabilizer', in *Intelligent Control and Automation, 2006. WCICA 2006. The Sixth World Congress*, volume 1, pages 2663–2667.

Licari, P., Siber, G. R. and Swartz, R. (1991), 'Production of cell mass and pertussis toxin by bordetella pertussis', *J. Biotechnol.*, **20**(2), 117–129, URL http://www.sciencedirect.com/science/article/B6T3C-4C9YYFG-K/2/2766516fb63b75d589cc14d35830240d.

Lima, P. V. and Saraiva, P. M. (2007), 'A semi-mechanistic model building framework based on selective and localized model extensions', *Computers & Chemical Engineering*, **31**(4), 361–373, URL http://www.sciencedirect.com/science/article/pii/S0098135406002043.

Linker, R. and Seginer, I. (2004), 'Greenhouse temperature modeling: a comparison between sigmoid neural networks and hybrid models', *Mathematics and Computers in Simulation*, **65**(1-2), 19–29, URL http://www.sciencedirect.com/science/article/pii/S037847540300137X.

Ljung, L. (1991), 'Issues in system identification', *Control Systems Magazine, IEEE*, **11**(1), 25–29.

Luebbert, A. and Bay Jorgensen, S. (2001), 'Bioreactor performance: a more scientific approach for practice', *Journal of Biotechnology*, **85**(2), 187–212, URL http://www.sciencedirect.com/science/article/pii/S0168165600003667.

MacGregor, J. F. and Kourti, T. (1995), 'Statistical process control of multivariate processes', *Control Eng. Prac.*, **3**(3), 403–414, URL http://www.sciencedirect.com/science/article/B6V2H-3YK00PN-5G/2/8472f367c50b8b1abfe0f0a12ff07e99.

Madar, J., Abonyi, J. and Szeifert, F. (2004), 'New approaches to the identification of semi-mechanistic process models', *Acta Agraria Kaposvariensis*, **8**, 1–9.

Madar, J., Abonyi, J. and Szeifert, F. (2005), 'Feedback linearizing control using hybrid neural networks identified by sensitivity approach', *Engineering Applications of Artificial Intelligence*, **18**(3), 343âĂŞ351, URL http://www.sciencedirect.com/science/article/B6V2M-4DS7VT1-2/2/f3466429504cb06aabd96976a905b227.

Mantovanelli, I. C. C., Rivera, E. C., Costa, A. C. and Filho, R. M. (2007), 'Hybrid neural network model of an industrial ethanol fermentation process considering the effect of temperature', in J. R. Mielenz, K. T. Klasson, W. S. Adney and J. D. McMillan (Editors), *ABAB Symposium*, Humana Press, pages 817–833–833, URL http://dx.doi.org/10.1007/978-1-60327-181-3_67.

Martinez, E. C. and Wilson, J. A. (1998), 'A hybrid neural network-first principles approach to batch unit optimisation', *Computers & Chemical Engineering*, **22**(Supplement 1), S893–S896, URL http://www.sciencedirect.com/science/article/pii/S0098135498001744.

Masri, S. F. (1994), 'A hybrid parametric/nonparametric approach for the identification of nonlinear systems', *Probabilistic Engineering Mechanics*, **9**(1-2), 47–57, URL http://www.sciencedirect.com/science/article/pii/0266892094900299.

Mazutti, M. A., Corazza, M. L., Maugeri, F., Rodrigues, M. I., Oliveira, J. V., Treichel, H. and Corazza, F. C. (2010), 'Hybrid modeling of inulinase bio-production process', *J. Chem. Technol. Biotechnol.*, **85**(4), 512–519, URL http://dx.doi.org/10.1002/jctb.2319.

McKay, B., Sanderson, C. S., Willis, M. J., Barford, J. P. and Barton, G. W. (1998), 'Evolving a hybrid model of a fed-batch fermentation process', *Transactions of the Institute of Measurement and Control*, **20**(1), 4–10, URL http://tim.sagepub.com/content/20/1/4.abstract.

Milanic, S., Strmcnik, S., Sel, D., Hvala, N. and Karba, R. (2004), 'Incorporating prior knowledge into artificial neural networks–an industrial case study', *Neurocomputing*, **62**, 131–151, URL http://www.sciencedirect.com/science/article/pii/S0925231204001948.

Mogk, G., Mrziglod, T. and Schuppert, A. (2002), 'Application of hybrid models in chemical industry', in J. Grievink and J. van Schijndel (Editors), *European Symposium on Computer Aided Process Engineering-12, 35th European Symposium of the Working Party on Computer Aided Process Engineering*, Elsevier, volume Volume 10, pages 931–936, URL http://www.sciencedirect.com/science/article/pii/S1570794602801833.

Molga, E. and Cherbanski, R. (1999), 'Hybrid first-principle-neural-network approach to modelling of the liquid-liquid reacting system', *Chemical Engineering Science*, **54**(13-14), 2467–2473, URL http://www.sciencedirect.com/science/article/pii/S0009250998005065.

Molga, E. and Westerterp, K. (1997), 'Neural network based model of the kinetics of catalytic hydrogenation reactions', in G. Froment and K. Waugh (Editors), *Dynamics of Surfaces and Reaction Kinetics in Heterogeneous Catalysis, Proceedings of the International Symposium*, Elsevier, volume Volume 109, pages 379–388, URL http://www.sciencedirect.com/science/article/pii/S0167299197804259.

Molga, E. J. (2003), 'Neural network approach to support modelling of chemical reactors: problems, resolutions, criteria of application', *Chemical Engineering and Processing*, **42**(8-9), 675–695, URL http://www.sciencedirect.com/science/article/pii/S0255270102002052.

Molga, E. J., van Woezik, B. A. A. and Westerterp, K. R. (2000), 'Neural networks for modelling of chemical reaction systems with complex kinetics: oxidation of 2-octanol with nitric acid', *Chemical Engineering and Processing*, **39**(4), 323–334, URL http://www.sciencedirect.com/science/article/pii/S0255270199000938.

Mouton, N., de Juan, A., Sliwa, M. and Ruckebusch, C. (2011), 'Hybrid hard- and soft-modeling approach for the resolution of convoluted femtosecond spectrokinetic data', *Chemometrics and Intelligent Laboratory Systems*, **105**(1), 74–82, URL http://www.sciencedirect.com/science/article/pii/S0169743910002200.

Narendra, K. S. . and Parthasarathy, K. (1990), 'Identification and control of dynamical systems using neural networks', *IEEE Transactlons On Neural Networks*, **1**(1), 1–27.

Nascimento, C. A. O., Giudici, R. and Scherbakoff, N. (1999), 'Modeling of industrial nylon-6,6 polymerization process in a twin-screw extruder reactor. ii. neural networks and hybrid models', *J. Appl. Polym. Sci.*, **72**(7), 905–912, URL http://dx.doi.org/10.1002/(SICI)1097-4628(19990516)72:7<905::AID-APP6>3.0.CO;2-7.

Ng, C. W. and Hussain, M. A. (2004), 'Hybrid neural network and prior knowledge model in temperature control of a semi-batch polymerization process', *Chemical Engineering and Processing*, **43**(4), 559âĂŞ570, URL http://www.sciencedirect.com/science/article/B6TFH-49SNCG6-1/2/0171b7c6da344cff4c4ebd013e0b5399.

Nikolov, S., Vera, J., Kotev, V., Wolkenhauer, O. and Petrov, V. (2008), 'Dynamic properties of a delayed protein cross talk model', *Biosystems*, **91**(1), 51–68, URL http://www.sciencedirect.com/science/article/B6T2K-4P718T0-1/2/f8fdbc1cbee4c3fada9829ef3b955749.

Oliveira, R. (1998), *Supervision, Control and Optimization of Biotechnological Processes based on Hybrid Models*, PhD Thesis, Martin-Luther-Universitaet Halle-Wittenberg.

Oliveira, R. (2004), 'Combining first principles modelling and artificial neural networks: a general framework', *Comput. Chem. Eng.*, **28**(5), 755–766, URL http://www.sciencedirect.com/science/article/B6TFT-4BYR3RN-3/2/01a70dc0724ad0547f1151dabfbe33a7.

Oliveira, R., Simutis, R. and Feyo de Azevedo, S. (2004), 'Design of a stable adaptive controller for driving aerobic fermentation processes near maximum oxygen transfer capacity', *Journal of Process Control*, **14**(6), 617–626, URL http://www.sciencedirect.com/science/article/pii/S0959152404000046.

Park, S. and Ramirez, W. F. (1988), 'Optimal production of secreted protein in fed-batch reactors', *AIChE J.*, **34**(9), 1550–1558, URL http://dx.doi.org/10.1002/aic.690340917.

Parlos, A., Rais, O. and Atiya, A. (1999), 'Multi-step-ahead prediction using dynamic recurrent neural networks', in *Neural Networks, 1999. IJCNN '99. International Joint Conference on*, volume 1, pages 349 –352 vol.1.

Parrott, L. (2011), 'Hybrid modelling of complex ecological systems for decision support: Recent successes and future perspectives', *Ecological Informatics*, **6**(1), 44–49, URL http://www.sciencedirect.com/science/article/pii/S1574954110000750.

PAT (2004), *Guidance for Industry PAT - A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance*, Technical Report, U.S. Department of Health and Human Services, Food and Drug Administration, URL http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm070305.pdf.

Patnaik (2001), 'Hybrid neural simulation of a fed-batch bioreactor for a nonideal recombinant fermentation', *Bioprocess and Biosystems Engineering*, **24**(3), 151–161, URL http://dx.doi.org/10.1007/s004490100246.

Patnaik, P. (2009), 'Intelligent models of the quantitative behavior of microbial systems', *Food and Bioprocess Technology*, **2**(2), 122–137, URL http://dx.doi.org/10.1007/s11947-008-0112-8.

Patnaik, P. (2010), 'Design considerations in hybrid neural optimization of fed-batch fermentation for phb production by <i>ralstonia eutropha</i>', *Food and Bioprocess Technology*, **3**(2), 213–225, URL http://dx.doi.org/10.1007/s11947-008-0086-6.

Patnaik, P. R. (2003), 'An integrated hybrid neural system for noise filtering,simulation and control of a fed-batch recombinant fermentation', *Biochemical Engineering Journal*, **15**(3), 165âĂŞ175, URL http://www.sciencedirect.com/science/article/B6V5N-47RJWFD-1/2/c4aa862b96b13f9a10b833c1c7d6c3af.

Patnaik, P. R. (2004), 'Neural and hybrid neural modeling and control of fed-batch fermentation for streptokinase: Comparative evaluation under nonideal conditions', *Can. J. Chem. Eng.*, **82**(3), 599–606, URL http://dx.doi.org/10.1002/cjce.5450820320.

Patnaik, P. R. (2008), 'Neural and hybrid optimizations of the fed-batch synthesis of poly-ĂŐÂš-hydroxybutyrate by ralstonia eutropha in a nonideal bioreactor', *Bioremediation Journal*, **12**(3), 117–130, URL http://dx.doi.org/10.1080/10889860802261687.

Pearson, R. K. and Pottmann, M. (2000), 'Gray-box identification of block-oriented nonlinear models', *Journal of Process Control*, **10**(4), 301–315, URL http://www.sciencedirect.com/science/article/pii/S0959152499000554.

Peres, J., Oliveira, R. and de Azevedo, S. F. (2008), 'Bioprocess hybrid parametric/nonparametric modelling based on the concept of mixture of experts', *Biochem. Eng. J.*, **39**(1), 190–206, URL http://www.sciencedirect.com/science/article/B6V5N-4PPW78D-2/2/e4a6a393f512bc38b38d6a9ca82bb1d2.

Peres, J., Oliveira, R. and Feyo de Azevedo, S. (2000), 'Knowledge based modular networks for process modelling and control', in S. Pierucci (Editor), *European Symposium on Computer Aided Process Engineering-10*, Elsevier, volume Volume 8, pages 247–252, URL http://www.sciencedirect.com/science/article/pii/S1570794600800437.

Peres, J., Oliveira, R. and Feyo de Azevedo, S. (2001), 'Knowledge based modular networks for process modelling and control', *Comput. Chem. Eng.*, **25**(4-6), 783–791, URL http://www.sciencedirect.com/science/article/B6TFT-454DTSG-14/2/928400eadd647504247038a7a6a1ad52.

Peres, J., Oliveira, R. and Feyo de Azevedo, S. (2003), 'Modelling cells reaction kinetics with artificial neural networks: A comparison of three network architectures', in A. Kraslawski and I. Turunen (Editors), *European Symposium on Computer Aided Process Engineering-13, 36th European Symposium of the Working Party on Computer Aided Process Engineering*, Elsevier, volume Volume 14, pages 839–844, URL http://www.sciencedirect.com/science/article/pii/S1570794603802213.

Peres, J., Oliveira, R., Serafim, L., Lemos, P., Reis, M. and Feyo de Azevedo, S. (2004), 'Hybrid modelling of a pha production process using modular neural networks', in A. Barbosa-PÃşvoa and H. Matos (Editors), *European Symposium on Computer-Aided Process Engineering-14, 37th European Symposium of the Working Party on Computer-Aided Process Engineering*, Elsevier, volume Volume 18, pages 733–738, URL http://www.sciencedirect.com/science/article/pii/S1570794604801883.

Piron, E., Latrille, E. and Rene, F. (1997), 'Application of artificial neural networks for crossflow microfiltration modelling: "black-box" and semi-physical approaches', *Computers & Chemical Engineering*, **21**(9), 1021–1030, URL http://www.sciencedirect.com/science/article/pii/S0098135496003328.

Ploemen, I. (1996), *A hybrid modeling methodology for non-linear mechanical systems using neural networks*, Master Thesis, Eindhoven University of Technology.

Ponton, J. and Klemes, J. (1993), 'Alternatives to neural networks for inferential measurement', *Computers & Chemical Engineering*, **17**(10), 991–1000, URL http://www.sciencedirect.com/science/article/pii/0098135493800807.

Porru, G., Aragonese, C., Baratti, R. and Servida, A. (2000), 'Monitoring of a co oxidation reactor through a grey model-based ekf observer', *Chemical Engineering Science*, **55**(2), 331–338, URL http://www.sciencedirect.com/science/article/pii/S0009250999003280.

Potocnik, P. and Grabec, I. (1999), 'Empirical modeling of antibiotic fermentation process using neural networks and genetic algorithms', *Mathematics and Computers in Simulation*, **49**(4-5), 363–379, URL http://www.sciencedirect.com/science/article/pii/S0378475499000452.

Potocnik, P., Grabec, I., Setinc, M. and Levec, J. (2000), 'Neural net based hybrid modeling of the methanol synthesis process', *Neural Processing Letters*, **11**(3), 219–228, URL http://dx.doi.org/10.1023/A:1009615710515.

Preusting, H., Noordover, J., Simutis, R. and Luebbert, A. (1996), 'The use of hybrid modelling for the optimization of the penicillin fermentation process', *CHIMIA*, **50**(9), 416–417.

Psichogios, D. C. and Ungar, L. H. (1992), 'A hybrid neural network-first principles approach to process modeling', *AIChE J.*, **38**(10), 1499–1511, URL http://dx.doi.org/10.1002/aic.690381003.

Qi, H., Zhou, X.-G., Liu, L.-H. and Yuan, W.-K. (1999), 'A hybrid neural network-first principles model for fixed-bed reactor', *Chemical Engineering Science*, **54**(13-14), 2521–2526, URL http://www.sciencedirect.com/science/article/pii/S0009250998005235.

Qin, S. (1993), 'Partial least squares regression for recursive system identification', volume 3, pages 2617–2622.

Qin, S. and McAvoy, T. (1992), 'Nonlinear pls modeling using neural networks', *Comput. Chem. Eng.*, **16**(4), 379–391, URL http://www.sciencedirect.com/science/article/B6TFT-43NW9V6-2R/2/10bec49918f60f2deceb400562bdb4f4.

Qin, S. J. and McAvoy, T. J. (1996), 'Nonlinear fir modeling via a neural net pls approach', *Comput. Chem. Eng.*, **20**(2), 147–159, URL http://www.sciencedirect.com/science/article/B6TFT-3Y0Y78N-3/2/2fd7d48a234c5709416203a4706edae5.

Ramaker, H., van Sprang, E., Gurden, S., Westerhuis, J. and Smilde, A. (2002), 'Improved monitoring of batch processes by incorporating external information', *Journal of Process Control*, **12**(4), 569–576, URL http://www.sciencedirect.com/science/article/pii/S0959152401000221.

Rateitschak, K. and Wolkenhauer, O. (2007), 'Intracellular delay limits cyclic changes in gene expression', *Math. Biosci.*, **205**(2), 163–179, URL http://www.sciencedirect.com/science/article/B6VHX-4KR4C22-2/2/b04f8254a44e27d3eb8d1113367afc7b.

Rathore, A., Bhambure, R. and Ghare, V. (2010), 'Process analytical technology (pat) for biopharmaceutical products', *Analytical and Bioanalytical Chemistry*, **398**(1), 137–154, URL http://dx.doi.org/10.1007/s00216-010-3781-x.

Read, E., Park, J., Shah, R., Riley, B., Brorson, K. and Rathore, A. (2010a), 'Process analytical technology (pat) for biopharmaceutical products: Part i. concepts and applications', *Biotechnol. Bioeng.*, **105**(2), 276–284, URL http://dx.doi.org/10.1002/bit.22528.

Read, E., Shah, R., Riley, B., Park, J., Brorson, K. and Rathore, A. (2010b), 'Process analytical technology (pat) for biopharmaceutical products: Part ii. concepts and applications', *Biotechnol. Bioeng.*, **105**(2), 285–295, URL http://dx.doi.org/10.1002/bit.22529.

Ren, H. T., Yuan, J. Q. and Bellgardt, K. H. (2003), 'Macrokinetic model for methylotrophic pichia pastoris based on stoichiometric balance', *J. Biotechnol.*, **106**(1), 53–68, URL http://www.sciencedirect.com/science/article/B6T3C-49V7CYW-3/2/c7fc6ac4410c1696ba4ce9cb712e2047.

Reuter, M., Van Deventer, J. and Van Der Walt, T. (1993), 'A generalized neural-net kinetic rate equation', *Chemical Engineering Science*, **48**(7), 1281–1297, URL http://www.sciencedirect.com/science/article/pii/000925099381009K.

Rhiel, M. H., Amrhein, M. I., Marison, I. W. and von Stockar, U. (2002), 'The influence of correlated calibration samples on the prediction performance of multivariate models based on mid-infrared spectra of animal cell cultures', *Anal. Chem.*, **74**(20), 5227–5236, URL http://dx.doi.org/10.1021/ac020165l.

Ricker, N. (1988), 'The use of biased least-squares estimators for parameters in discrete-time pulse-response models', *Ind. Eng. Chem. Res.*, **27**(2), 343–350.

Rodriguez, M. E., Hozbor, D. F., Samo, A. L., Ertola, R. and Yantorno, O. M. (1994), 'Effect of dilution rate on the release of pertussis toxin and lipopolysaccharide ofbordetella pertussis', *J. Ind. Microbiol. Biot.*, **13**(5), 273–278, URL http://dx.doi.org/10.1007/BF01569728.

Romijn, R., Ozkan, L., Weiland, S., Ludlage, J. and Marquardt, W. (2008), 'A grey-box modeling approach for the reduction of nonlinear systems', *J. Process Control*, **18**(9), 906–914.

Roubos, H. (2002), *Bioprocess Modelling and Optimization Fed-Batch Clavulanic Acid Production by Streptomyces Clavuligerus*, PhD Thesis, Technical University Delft.

Roubos, J., Krabben, P., Setnes, M., Babuska, R., Heijnen, J. and Verbruggen, H. (2000), 'Hybrid model development for fed-batch bioprocesses; combining physical equations with the metabolic network and black-box kinetics', *J. A - Benelux Quarterly Journal on Automatic Control*, **41**(3), 12–23.

Ruckebusch, C., Sliwa, M., RÃľhault, J., Naumov, P., Huvenne, J. and Buntinx, G. (2009), 'Hybrid hard- and soft-modelling applied to analyze ultrafast processes by femtosecond transient absorption spectroscopy: Study of the photochromism of salicylidene anilines', *Analytica Chimica Acta*, **642**(1-2), 228–234, URL http://www.sciencedirect.com/science/article/pii/S0003267008019399.

Rusinowski, H. and Stanek, W. (2010), 'Hybrid model of steam boiler', *Energy*, **35**(2), 1107–1113, URL http://www.sciencedirect.com/science/article/pii/S0360544209002291.

Safavi, A. A., Nooraii, A. and Romagnoli, J. A. (1999), 'A hybrid model formulation for a distillation column and the on-line optimisation study', *Journal of Process Control*, **9**(2), 125–134, URL http://www.sciencedirect.com/science/article/pii/S0959152498000419.

Saraceno, A., Aversa, M. and Curcio, S. (2010a), 'Advanced modeling of food convective drying: A comparison between artificial neural networks and hybrid approaches', *Food and Bioprocess Technology*, 1–12–12, URL http://dx.doi.org/10.1007/s11947-010-0477-3.

Saraceno, A., Curcio, S., Calabro, V. and Iorio, G. (2010b), 'A hybrid neural approach to model batch fermentation of "ricotta cheese whey" to ethanol', *Computers & Chemical Engineering*, **34**(10), 1590–1596, URL http://www.sciencedirect.com/science/article/pii/S0098135409002828.

Saraceno, A., Sansonetti, S., Curcio, S., CalabrÃš, V. and Iorio, G. (2010c), 'A hybrid neural approach to model batch fermentation of dairy industry wastes', in S. Pierucci and G. B. Ferraris (Editors), *20th European Symposium on Computer Aided Process Engineering*, Elsevier, volume Volume 28, pages 739–744, URL http://www.sciencedirect.com/science/article/pii/S1570794610281245.

Saxen, B. and Saxen, H. (1996), 'A neural-network based model of bioreaction kinetics', *Can. J. Chem. Eng.*, **74**(1), 124–131, URL http://dx.doi.org/10.1002/cjce.5450740116.

Schenk, J., Marison, I. W. and von Stockar, U. (2007), 'A simple method to monitor and control methanol feeding of pichia pastoris fermentations using mid-ir spectroscopy', *J. Biotechnol.*, **128**(2), 344–353, URL http://www.sciencedirect.com/science/article/B6T3C-4M69J9D-1/2/1644abe9b0cca9a67b9b4ef66d6d37c5.

Schenker, B. and Agarwal, M. (2000), 'Online-optimized feed switching in semi-batch reactors using semi-empirical dynamic models', *Control Engineering Practice*, **8**(12), 1393–1403, URL http://www.sciencedirect.com/science/article/pii/S0967066100000770.

Scheper, T., Hitzmann, B., StÃďrk, E., Ulber, R., Faurie, R., Sosnitza, P. and Reardon, K. F. (1999), 'Bioanalytics: detailed insight into bioprocesses', *Analytica Chimica Acta*, **400**(1-3), 121–134, URL http://www.sciencedirect.com/science/article/B6TF4-3XX6871-F/2/2874cbab7770ca9d2fdc3af85529db17.

Schubert, J., Simutis, R., Dors, M., Havlik, I. and Luebbert, A. (1994a), 'Bioprocess optimization and control: Application of hybrid modelling', *J. Biotechnol.*, **35**(1), 51–68, URL http://www.sciencedirect.com/science/article/B6T3C-47DKVW2-8K/2/53a811903775c5093cfee3c453abdaef.

Schubert, J., Simutis, R., Dors, M., Havlik, I. and Luebbert, A. (1994b), 'Hybrid modelling of yeast production processes - combination of a priori knowledge on different levels of sophistication', *Chem.Eng. Technol.*, **17**(1), 10–20, URL http://dx.doi.org/10.1002/ceat.270170103.

Schuegerl, K. (2001), 'Progress in monitoring, modeling and control of bioprocesses during the last 20 years', *Journal of Biotechnology*, **85**(2), 149–173, URL http://www.sciencedirect.com/science/article/pii/S0168165600003618.

Schuppert, A. (1999), 'Extrapolability of structured hybrid models: a key to the optimization of complex processes', in *Proceedings of the International Conference on Differential Equations (Equadiff)*.

Seborg, D., Edgar, T., Mellichamp, D. and III, F. D. (2010), *Process Dynamics and Control, 3rd edition,*, John Wiley & Sons, New York.

Shao, H. (2009), 'Convergence of a gradient algorithm with penalty for training two-layer neural networks', in L. Liu and G. Zheng (Editors), *Computer Science and Information Technology, International Conference on*, volume 0, pages 76–79, URL http://doi.ieeecomputersociety.org/10.1109/ICCSIT.2009.5234616.

Shu, H. and Pi, Y. (2000), 'Pid neural networks for time-delay systems', *Computers & Chemical Engineering*, **24**(2-7), 859–862, URL http://www.sciencedirect.com/science/article/B6TFT-448HNR0-3P/2/f7371ac3b53a736386b3b4a5f7df6ad6.

Silva, R., Cruz, A., Hokka, C., Giordano, R. and Giordano, R. (2000), 'A hybrid feedforward neural network model for the cephalosporin c production process', *Brazilian Journal of Chemical Engineering*, **17**, 587–598.

Silva, R., Cruz, A., Hokka, C., Giordano, R. and Giordano, R. (2001), 'A hybrid neuroal network algorithm for on-line state inference that accounts for differences in inoculum of <i>cephalosporium acremonium</i> in fed-batch fermentors', *Applied Biochemistry and Biotechnology*, **91-93**(1), 341–352, URL http://dx.doi.org/10.1385/ABAB:91-93:1-9:341.

Simon, L. L., Fischer, U. and Hungerbuehler, K. (2006), 'Modeling of a three-phase industrial batch reactor using a hybrid first-principles neural-network model', *Industrial & Engineering Chemistry Research*, **45**(21), 7336–7343, URL http://dx.doi.org/10.1021/ie060181z.

Simutis, R., Havlik, I., Schneider, F., Dors, M. and Luebbert, A. (1995), 'Artificial neural networks of improved reliability for industrial process supervision', in *Preprints of the 6th International Conference on Computer Applications in Biotechnology*.

Simutis, R. and Luebbert, A. (1997), 'Exploratory analysis of bioprocesses using artificial neural network-based methods', *Biotechnol Progress*, **13**(4), 479–487, URL http://dx.doi.org/10.1021/bp9700364.

Simutis, R., Oliveira, R., Manikowski, M., de Azevedo, S. F. and Luebbert, A. (1997), 'How to increase the performance of models for process optimization and control', *J. Biotechnol.*, **59**(1-2), 73–89, URL http://www.sciencedirect.com/science/article/B6T3C-3S3M282-9/2/245e3db28ed6259fa3322a542e8d155c.

Smolen, P., Baxter, D. A. and Byrne, J. (1999), 'Effects of macromolecular transport and stochastic fluctuations on dynamics of genetic regulatory systems', *J Physiol Cell Physiol*, **277**(4), C777–C790.

Sohlberg, B. (2005), 'Hybrid grey box modelling of a pickling process', *Control Engineering Practice*, **13**(9), 1093–1102, URL http://www.sciencedirect.com/science/article/B6V2H-4F2V50J-2/2/ef98aea312d6e4fc813d0eb96563aabd.

Sohn, J.-H., Lee, S.-K. and Yoo, W.-S. (2008), 'Hybrid neural network bushing model for vehicle dynamics simulation', *Journal of Mechanical Science and Technology*, **22**(12), 2365–2374, URL http://dx.doi.org/10.1007/s12206-008-0712-2.

Soons, Z. I. T. A., Shi, J., Stigter, J. D., van der Pol, L. A., van Straten, G. and Van Boxtel, A. J. B. (2008a), 'Observer design and tuning for biomass growth and k(l)a using online and offline measurements', *J. Process Contr.*, **18**(7-8), 621–631.

Soons, Z. I. T. A., Streefland, M., van Straten, G. and van Boxtel, A. J. B. (2008b), 'Assessment of near infrared and "software sensor" for biomass monitoring and control', *Chemom. Intell. Lab. Syst.*, **94**(2), 166–174.

Soons, Z. I. T. A., Voogt, J. A., van Straten, G. and van Boxtel, A. J. B. (2006), 'Constant specific growth rate in fed-batch cultivation of bordetella pertussis using adaptive control', *J. Biotechnol.*, **125**(2), 252–268.

Stelling, J. (2004), 'Mathematical models in microbial systems biology', *Current Opinion in Microbiology*, **7**(5), 513–518, URL http://www.sciencedirect.com/science/article/pii/S1369527404000980.

Su, H. T., Bhat, N., Minderman, A. and McAvoy, T. J. (1992), 'Integrating neural networks with first principles models for dynamic modeling', in *IFAC Symposium on Dynamics and Control of Chemical Reactors Distillation Columns and Batch Processes*, IFAC, Maryland.

Su, H. T. and McAvoy, T. J. (1993), 'Integration of multilayer perceptron networks and linear dynamic models: a hammerstein modeling approach', *Industrial & Engineering Chemistry Research*, **32**(9), 1927–1936, URL http://dx.doi.org/10.1021/ie00021a017.

Takagi, T. and Sugeno, M. (1985), 'Fuzzy identification of systems and its applications to modeling and control', *IEEE Transactions on Systems, Man, and Cybernetics*, **15**(1), 116–132, URL http://www.hi.cs.meiji.ac.jp/~takagi/paper/TS-MODEL.tar.gz.

Teissier, P., Perret, B., Latrille, E., Barillere, J. M. and Corrieu, G. (1997), 'A hybrid recurrent neural network model for yeast production monitoring and control in a wine base medium', *Journal of Biotechnology*, **55**(3), 157–169, URL http://www.sciencedirect.com/science/article/pii/S0168165697000655.

Teixeira, A., Cunha, A., Clemente, J., Alves, P., Carrondo, M. and Oliveira, R. (2005a), 'Dynamic modelling and optimisation of a mammalian cells process using hybrid grey-box systems', URL http://dx.doi.org/10.1007/3-211-27389-1_71.

Teixeira, A., Cunha, A. E., Clemente, J. J., Moreira, J. L., Cruz, H. J., Alves, P. M., Carrondo, M. J. T. and Oliveira, R. (2005b), 'Modelling and optimization of a recombinant bhk-21 cultivation process using hybrid grey-box systems', *J. Biotechnol.*, **118**(3), 290–303.

Teixeira, A. P., Alves, C., Alves, P. M., Carrondo, M. J. T. and Oliveira, R. (2007a), 'Hybrid elementary flux analysis/nonparametric modeling: application for bioprocess control', *Bmc Bioinformatics*, **8**.

Teixeira, A. P., Carinhas, N., Dias, J. M., Cruz, P., Alves, P. M., Carrondo, M. J. and Oliveira, R. (2007b), 'Hybrid semi-parametric mathematical systems: Bridging the gap between systems biology and process engineering', *J. Biotechnol.*, **132**(4), 418–425, URL http://www.sciencedirect.com/science/article/B6T3C-4PF1W7F-1/2/cdb9cb9c0d03c549617476f77ca1040b.

Teixeira, A. P., Clemente, J. J., Cunha, A. E., Carrondo, M. J. T. and Oliveira, R. (2006), 'Bioprocess iterative batch-to-batch optimization based on hybrid parametric/nonparametric models', *Biotechnol. Prog.*, **22**(1), 247–258.

Teixeira, A. P., Oliveira, R., Alves, P. M. and Carrondo, M. J. T. (2009), 'Advances in on-line monitoring and control of mammalian cell cultures: Supporting the pat initiative', *Biotechnol. Adv.*, **27**(6), 726–732.

Thibault, J., Acuna, G., Perez-Correa, R., Jorquera, H., Molin, P. and Agosin, E. (2000), 'A hybrid representation approach for modelling complex dynamic bioprocesses', *Bioprocess and Biosystems Engineering*, **22**(6), 547–556, URL http://dx.doi.org/10.1007/s004499900110.

Tholudur, A. and Ramirez, W. F. (1996), 'Optimization of fed-batch bioreactors using neural network parameter function models', *Biotechnol Progress*, **12**(3), 302–309, URL http://dx.doi.org/10.1021/bp960012h.

Thompson, M. L. and Kramer, M. A. (1994), 'Modeling chemical processes using prior knowledge and neural networks', *AIChE J.*, **40**(8), 1328–1340, URL http://dx.doi.org/10.1002/aic.690400806.

Tian, T., Burrage, K., Burrage, P. M. and Carletti, M. (2007), 'Stochastic delay differential equations for genetic regulatory networks', *J. Comput. Appl. Math.*, **205**(2), 696–707, URL http://www.sciencedirect.com/science/article/B6TYH-4KKFP86-1/2/f3bf54ea22230e58749b896c0e2ea9c5.

Tian, Y., Zhang, J. and Morris, J. (2001), 'Modeling and optimal control of a batch polymerization reactor using a hybrid stacked recurrent neural network model', *Industrial & Engineering Chemistry Research*, **40**(21), 4525–4535, URL http://dx.doi.org/10.1021/ie0010565.

Tsen, A. Y.-D., Jang, S. S., Wong, D. S. H. and Joseph, B. (1996), 'Predictive control of quality in batch polymerization using hybrid ann models', *AIChE J.*, **42**(2), 455–465, URL http://dx.doi.org/10.1002/aic.690420215.

Undey, C., Ertunc, S. and Cinar, A. (2003), 'Online batch/fed-batch process performance monitoring, quality prediction, and variable-contribution analysis for diagnosis', *Ind. Eng. Chem. Res.*, **42**(20), 4645–4658, URL http://dx.doi.org/10.1021/ie0208218.

van Can, H. J. L., Hellinga, C., Luyben, K. C. A. M., Heijnen, J. J. and Te Braake, H. A. B. (1996), 'Strategy for dynamic process modeling based on neural networks in macroscopic balances', *AIChE J.*, **42**(12), 3403–3418, URL http://dx.doi.org/10.1002/aic.690421211.

van Can, H. J. L., te Braake, H. A. B., Bijman, A., Hellinga, C., Luyben, K. C. A. M. and Heijnen, J. J. (1999), 'An efficient model development strategy for bioprocesses based on neural networks in macroscopic balances: Part ii', *Biotechnol. Bioeng.*, **62**(6), 666–680, URL http://dx.doi.org/10.1002/(SICI)1097-0290(19990320)62:6<666::AID-BIT6>3.0.CO;2-S.

van Can, H. J. L., Te Braake, H. A. B., Dubbelman, S., Hellinga, C., Luyben, K. C. A. M. and Heijnen, J. J. (1998), 'Understanding and applying the extrapolation properties of serial gray-box models', *AIChE J.*, **44**(5), 1071–1089, URL http://dx.doi.org/10.1002/aic.690440507.

van Can, H. J. L., te Braake, H. A. B., Hellinga, C. and Luyben, K. C. A. M. (1997), 'An efficient model development strategy for bioprocesses based on neural networks in macroscopic balances', *Biotechnol. Bioeng.*, **54**(6), 549–566, URL http://dx.doi.org/10.1002/(SICI)1097-0290(19970620)54:6<549::AID-BIT6>3.0.CO;2-J.

van de Merbel, N. C., Lingeman, H. and Brinkman, U. A. T. (1996), 'Sampling and analytical strategies in on-line bioprocess monitoring and control', *Journal of Chromatography A*, **725**(1), 13–27, URL http://www.sciencedirect.com/science/article/B6TG8-3TKMJ84-41/2/90cd1e5d46147c60cb57f09134c6151f.

van Deventer, J. S., Kam, K. M. and van der Walt, T. J. (2004), 'Dynamic modelling of a carbon-in-leach process with the regression network', *Chemical Engineering Science*, **59**(21), 4575–4589, URL http://www.sciencedirect.com/science/article/pii/S0009250904004063.

van Lith, P. F., Betlem, B. H. L. and Roffel, B. (2002), 'A structured modeling approach for dynamic hybrid fuzzy-first principles models', *Journal of Process Control*, **12**(5), 605–615, URL http://www.sciencedirect.com/science/article/pii/S0959152401000543.

van Lith, P. F., Betlem, B. H. L. and Roffel, B. (2003), 'Combining prior knowledge with data driven modeling of a batch distillation column including start-up', *Computers & Chemical Engineering*, **27**(7), 1021–1030, URL http://www.sciencedirect.com/science/article/pii/S009813540300067X.

van Sprang, E. N. M., Streefland, M., Ramaker, H.-J., van der Pol, L. A., Beuvery, E. C. and Smilde, A. K. (2007), 'Manufacturing vaccines: An illustration of using pat tools for controlling the cultivation of bordetella pertussis', *Quality Engineering*, **19**(4), 373–384, URL http://www.informaworld.com/10.1080/08982110701630461.

Vande Wouwer, A., Renotte, C. and Bogaerts, P. (2004), 'Biological reaction modeling using radial basis function networks', *Computers & Chemical Engineering*, **28**(11), 2157–2164, URL http://www.sciencedirect.com/science/article/pii/S0098135404000651.

Vieira, J., Dias, F. and Mota, A. (2005), 'Hybrid neuro-fuzzy network-priori knowledge model in temperature control of a gas water heater system', in *Hybrid Intelligent Systems, 2005. HIS '05. Fifth International Conference on DOI - 10.1109/ICHIS.2005.60*, pages 6 pp.–.

Vieira, J. and Mota, A. (2005), 'Combining first principles with grey-box approaches for modelling a water gas heater system', in *Intelligent Control, 2005. Proceedings of the 2005 IEEE International Symposium on, Mediterrean Conference on Control and Automation DOI - 10.1109/.2005.1467174*, pages 1131–1136.

Vilim, R. B., Garcia, H. E. and Chen, F. W. (2001), 'An identification scheme combining first principle knowledge, neural networks, and the likelihood function', in *IEEE Transactions On Control Systems Technology*, IEEE, volume 9.

Vojinovic, V., Cabral, J. and Fonseca, L. (2006), 'Real-time bioprocess monitoring: Part i: In situ sensors', *Sensors and Actuators B: Chemical*, **114**(2), 1083–1091, URL http://www.sciencedirect.com/science/article/B6THH-4H7THFS-1/2/e2bcaecdbda6dea9b418f5d182c65431.

Vojinovic, V., Cabral, J. and Fonseca, L. (2007), 'Ex situ bioprocess monitoring techniques', *Chemical Industry and Chemical Engineering Quarterly*, **13**(2), 103–116.

von Stosch, M., Oliveira, R., Peres, J. and Feyo de Azevedo, S. (2011a), 'A hybrid modeling framework for pat: Application to bordetella pertussis cultures', *Accepted for publication by the Journal of Biotechnolgy*.

von Stosch, M., Oliveira, R., Peres, J. and Feyo de Azevedo, S. (2011b), 'A novel identification method for hybrid (n)pls dynamical systems with application to bioprocesses', *Expert Syst. Appl.*, **In Press, Accepted Manuscript**, –, URL http://www.sciencedirect.com/science/article/B6V03-528909R-2/2/e6dfd8ffba1de4c8d497160e67bfa931.

von Stosch, M., Peres, J., de Azevedo, S. and Oliveira, R. (2010), 'Modelling biochemical networks with intrinsic time delays: a hybrid semi-parametric approach', *BMC Systems Biology*, **4**(1), 131, URL http://www.biomedcentral.com/1752-0509/4/131.

Wang, H. and Yu, J. (2004), 'Application study on nonlinear dynamic fir modeling using hybrid svm-pls method', volume 4, pages 3479–3482.

Wang, X., Chen, J., Liu, C. and Pan, F. (2010a), 'Hybrid modeling of penicillin fermentation process based on least square support vector machine', *Chemical Engineering Research and Design*, **88**(4), 415–420, URL http://www.sciencedirect.com/science/article/pii/S0263876209002160.

Wang, Y.-C. and Chen, B.-S. (2010), 'Integrated cellular network of transcription regulations and protein-protein interactions', *BMC Systems Biology*, **4**(1), 20, URL http://www.biomedcentral.com/1752-0509/4/20.

Wang, Y.-C., Lan, C.-Y., Hsieh, W.-P., Murillo, L., Agabian, N. and Chen, B.-S. (2010b), 'Global screening of potential candida albicans biofilm-related transcription factors via network comparison', *BMC Bioinformatics*, **11**(1), 53, URL http://www.biomedcentral.com/1471-2105/11/53.

Wei, N. C., Hussain, M. A. and Wahab, A. K. A. (2007), 'Control of a batch polymerization system using hybrid neural network - first principle model', *Can. J. Chem. Eng.*, **85**(6), 936–945, URL http://dx.doi.org/10.1002/cjce.5450850616.

Werbos, P. (1974), *Beyond Regression: New Tools for Prediction and Analysis in Behavioral Sciences*, PhD Thesis, Harvard University.

Werbos, P. (1988), 'Backpropagation: past and future', volume 1, pages 343–353.

Westdijk, J., Ijssel, J. v. d., Thalen, M., Beuvery, C. and Jiskoot, W. (1997), 'Quantification of cell-associated and free antigens in bordetella pertussis suspensions by antigen binding elisa', *J. Immunoassay*, **18**(3), 267–284, URL http://www.informaworld.com/10.1080/01971529708005819.

Wilson, J. A. and Zorzetto, L. F. M. (1997), 'A generalised approach to process state estimation using hybrid artificial neural network/mechanistic models', *Comput. Chem. Eng.*, **21**(9), 951–963, URL http://www.sciencedirect.com/science/article/B6TFT-3S9TDFC-3/2/f51e139e790f88727032bfd23dba3a97.

Wold, S., Kettaneh-Wold, N. and Skagerberg, B. (1989), 'Nonlinear pls modeling', *Chemom. Intell. Lab. Syst.*, **7**(1-2), 53–65, URL http://www.sciencedirect.com/science/article/B6TFP-44J0R0J-CM/2/94e2e450e7e44e081cfd41b26ef3b1f2.

Wold, S., Sjostrom, M. and Eriksson, L. (2001), 'Pls-regression: a basic tool of chemometrics', *Chemom. Intell. Lab. Syst.*, **58**(2), 109–130, URL http://www.sciencedirect.com/science/article/B6TFP-44B4XN8-6/2/902049f55bd33375bb5ae90aac740e74.

Wolkowicz, G. S. and Xia, H. (1997), 'Global asymptotic behavior of a chemostat model with discrete delays', *SIAM J. Appl. Math*, **57**, 411–422.

Wolkowicz, G. S. K., Xia, H. and Ruan, S. (1997), 'Competition in the chemostat: A distributed delay model and its global asymptotic behavior', *SIAM J. Appl. Math*, **57**, 1281–1310.

Xiong, Q. and Jutan, A. (2002), 'Grey-box modelling and control of chemical processes', *Chemical Engineering Science*, **57**(6), 1027âĂŞ1039, URL http://www.sciencedirect.com/science/article/B6TFK-44VWWNB-2/2/75f0bd7e308f171bb54a243b2a3a08e2.

Yamashita, S.-i., Yurimoto, H., Murakami, D., Yoshikawa, M., Oku, M. and Sakai, Y. (2009), 'Lag-phase autophagy in the methylotrophic yeast pichia pastoris', *Genes to Cells*, **14**(7), 861–870, URL http://dx.doi.org/10.1111/j.1365-2443.2009.01316.x.

Yamuna Rani, K. and Ramachandra Rao, V. S. (1999), 'Control of fermenters - a review', *Bioprocess Biosyst. Eng.*, **21**(1), 77–88, URL http://dx.doi.org/10.1007/PL00009066.

Yang, A., Martin, E. and Morris, J. (2011), 'Identification of semi-parametric hybrid process models', *Computers & Chemical Engineering*, **35**(1), 63–70, URL http://www.sciencedirect.com/science/article/pii/S0098135410001626.

Yang, S.-K., Wang, Y.-C., Chao, C.-C., Chuang, Y.-J., Lan, C.-Y. and Chen, B.-S. (2010), 'Dynamic cross-talk analysis among tnf-r, tlr-4 and il-1r signalings in tnfalpha-induced inflammatory responses', *BMC Medical Genomics*, **3**(1), 19, URL http://www.biomedcentral.com/1755-8794/3/19.

Yeo, Y.-K. and Kwon, T.-I. (1999), 'A neural pid controller for the ph neutralization process', *Industrial & Engineering Chemistry Research*, **38**(3), 978–987, URL http://dx.doi.org/10.1021/ie9805133.

Zabot, G., Mecca, J., Mesomo, M., Silva, M., Pra, V., de Oliveira, D., Oliveira, J., Castilhos, F., Treichel, H. and Mazutti, M. (2011), 'Hybrid modeling of xanthan gum bioproduction in batch bioreactor', *Bioprocess and Biosystems Engineering*, 1–12, URL http://dx.doi.org/10.1007/s00449-011-0548-5.

Zahedi, G., Elkamel, A., Lohi, A., Jahanmiri, A. and Rahimpor, M. (2005), 'Hybrid artificial neural network–first principle model formulation for the unsteady state simulation and analysis of a packed bed reactor for co2 hydrogenation to methanol', *Chemical Engineering Journal*, **115**(1-2), 113–120, URL http://www.sciencedirect.com/science/article/pii/S1385894705003104.

Zahedi, G., Lohi, A. and Mahdi, K. (2011), 'Hybrid modeling of ethylene to ethylene oxide heterogeneous reactor', *Fuel Processing Technology*, **92**(9), 1725–1732, URL http://www.sciencedirect.com/science/article/pii/S0378382011001433.

Zander, H.-J., Dittmeyer, R. and Wagenhuber, J. (1999), 'Dynamic modeling of chemical reaction systems with neural networks and hybrid models', *Chem. Eng. Technol.*, **22**(7), 571–574, URL http://dx.doi.org/10.1002/(SICI)1521-4125(199907)22:7<571::AID-CEAT571>3.0.CO;2-5.

Zbicinski, I., Strumillo, P. and Kaminski, W. (1996), 'Hybrid neural model of thermal drying in a fluidized bed', *Computers & Chemical Engineering*, **20**(Supplement 1), S695–S700, URL http://www.sciencedirect.com/science/article/pii/0098135496001251.

Zhang, L., Pan, M., Quan, S., Chen, Q. and Shi, Y. (2006), 'Adaptive neural control based on pemfc hybrid modeling', in *Intelligent Control and Automation, 2006. WCICA 2006. The Sixth World Congress on DOI - 10.1109/WCICA.2006.1713598*, volume 2, pages 8319–8323.

Zhang, W., Smith, L. A., Plantz, B. A., Schlegel, V. L. and Meagher, M. M. (2002), 'Design of methanol feed control in pichia pastoris fermentations based upon a growth model', *Biotechnol Progress*, **18**(6), 1392–1399, URL http://dx.doi.org/10.1021/bp025516w.

Zorzetto, L., Filho, R. and Wolf-Maciel, M. (2000), 'Processing modelling development through artificial neural networks and hybrid models', *Computers & Chemical Engineering*, **24**(2-7), 1355–1360, URL http://www.sciencedirect.com/science/article/pii/S0098135400004191.

Zorzetto, L. and Wilson, J. (1996), 'Monitoring bioprocesses using hybrid models and an extended kalman filter', *Comput. Chem. Eng.*, **20**(Supplement 1), S689–S694, URL http://www.sciencedirect.com/science/article/B6TFT-48JC24K-44/2/2ee3c3b8de769c80595cea09c4379e70.

Zuo, K., Cheng, H.-P., Wu, S.-C. and Wu, W.-T. (2006), 'A hybrid model combining hydrodynamic and biological effects for production of bacterial cellulose with a pilot scale airlift reactor', *Biochemical Engineering Journal*, **29**(1-2), 81–90, URL http://www.sciencedirect.com/science/article/pii/S1369703X05001683.

Zuo, K. and Wu, W. (2000), 'Semi-realtime optimization and control of a fed-batch fermentation system', *Computers & Chemical Engineering*, **24**(2-7), 1105–1109, URL http://www.sciencedirect.com/science/article/pii/S0098135400004907.