This article was published in Computers & Chemical Engineering, 60, 86-101, 2014 http://dx.doi.org/10.1016/j.compchemeng.2013.08.008

Hybrid semi-parametric modeling in process systems engineering: Past, present and future

Moritz von Stosch^b, Rui Oliveira^b, Joana Peres^a, Sebastião Feyo de Azevedo^a,*

^a LEPAE, Departamento de Engenharia Quimica, Faculdade de Engenharia, Universidade do Porto, Rua Dr. Roberto Frias s/n, 4200-465 Porto, Portugal

^b REQUIMTE, Departamento de Quimica, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal

Abstract

Hybrid semi-parametric models consist of model structures that combine parametric and nonparametric submodels based on different knowledge sources. The development of a hybrid semi-parametric model can offer several advantages over traditional mechanistic or data-driven modeling, as reviewed in this paper. These advantages, such as broader knowledge base, transparency of the modeling approach and cost-effective model development, have been widely recognized, not only in academia but also in the industry.

In this paper, the most common hybrid semi-parametric modeling and parameter identification techniques are revisited. Applications in the areas of (bio)chemical engineering for process monitoring, control, optimization, scaleup and model-reduction are reviewed. It is outlined that the application of hybrid semi-parametric techniques does not automatically lead into better results but that rational knowledge integration has potential to significantly improve model-based process operation and design.

1. Introduction

In process systems engineering process modeling takes a central role (Cameron & Hangos, 2001). In its essence, process modeling is an exercise of translation of knowledge about the process into an abstract mathematical representation (Cameron & Hangos, 2001). The nature of knowledge is diverse and thus modeling methods can naturally be segmented according to the nature of the knowledge. First-Principles, mechanistic or phenomenological models represent a broad class of more transparent (white box) models. In relation thereto, data-driven modeling represents a less transparent (black-box) modeling framework

based exclusively on process data.

A closely related mathematical classification can be done with respect to the form of model parameterization. Parametric models are determined a priori on the basis of knowledge about the process (Thompson & Kramer, 1994; Walter, Pronzato, & Norton, 1997). Their number of parameters is fixed and they might have a physical or empirical interpretation depending on the level of knowledge sophistication. White-box models naturally fall in the category of parametric models. On the contrary, nonparametric models are determined exclusively from data (Haerdle, Mueller, Sperlich, & Werwatz, 2004; Thompson & Kramer, 1994). The term nonparametric is not meant to imply that these models completely lack parameters but that the number and nature of the parameters are flexible and not fixed in advance by knowledge. In between these two extremes lies hybrid semi-parametric modeling, which is the focus of this review (Fig. 1).

Hybrid semi-parametric models can thus be defined as model structures that combine parametric and nonparametric submodels (Thompson & Kramer, 1994). Their application to process modeling has evolved from the field of neural networks, being first reported in 1992 by Psichogios and Ungar (1992), Kramer, Thompson, and Bhagat (1992), Johansen and Foss (1992a), and Su, Bhat, Minderman, and McAvoy (1992). 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 process data, the hybrid semi-parametric model was capable to predict the process states better, was able to interpolate and extrapolate mostly more accurately and was easier to interpret than models based solely on neural networks.

Several other modeling methods exist that combine different types of knowledge or/and submodels. The term grey-box modeling appeared in the 1990-s in systems and control theory describing the incorporation of prior information (mainly structural information derived from first-principles, i.e. white-box models) into empirical (black-box) models (Bohlin & Graebe, 1995; Jorgensen & Hangos, 1995; Tulleken, 1993). According to Braake, van Can, and Verbruggen (1998) a grey-box model is based on the same unstructured nature as a black-box model. The term has however evolved to designate all types of models that combine white-box and black-box submodels. For instance, the grey-box models in Akkari, Chevallier, and Boillereaux (2005), Estrada-Flores, Merts, De Ketelaere, and Lammertyn (2006), and Worden et al. (2007) combine first-principle based "white-box" models and empirical "black-box" models. In these cases both, white- and black-box models, are parametric models. According to the definition in this paper they are therefore not hybrid semi-parametric models. Hybrid semi-parametric models may however be viewed as a class of grey- box models in that the parametric and nonparametric submodels have different levels of transparency.

Block oriented models are another class of separable models consisting of linear

dynamic and static nonlinear elements (Haber & Keviczky, 1999; Pearson & Pottmann, 2000), e.g. the Hammerstein model or the Wiener model (Haber & Keviczky, 1999). Block oriented models have some resemblance with hybrid semiparametric models in that the nonlinear element could e.g. be represented by a neural network with a standard linear time invariant form as the linear dynamic model, such as in Su and McAvoy (1993). However, block oriented models are not necessarily hybrid semi-parametric models if the blocks do not explicitly combine parametric and non- parametric submodels.

Multiscale models are also compositions of two or more sub models that describe phenomena at different scales (Ingram, Cameron, & Hangos, 2004). In the vast majority of cases multiscale models are mechanistic parametric models (Ingram et al., 2004). However, multiscale models could also be hybrid semiparametric models if parametric and nonparametric submodels target different scales (Teixeira, Alves, Alves, Carrondo, & Oliveira, 2007).

It should be noted that the term "hybrid modeling" has been frequently used as an equivalent to "hybrid semi-parametric modeling" in the literature, which is however a rather ambiguous definition as it can embrace many other types of modeling methods such as grey-box, block-oriented or multiscale model- ing approaches referred to above. For coherence, we keep the term "hybrid semiparametric" throughout this review.

1.1. Why hybrid semi-parametric modeling? What is the gain?

Mechanistic modeling and data-driven modeling constitute two approaches which are different in their traits. While the development of a mechanistic model is many times cumber- some/laborious and requires detailed knowledge about the process, data-driven approaches are rather quickly applicable and require less knowledge. In comparison to mechanistic 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. Hybrid semi-parametric modeling can balance the advantages and disadvantages of strictly mechanistic and nonparametric modeling. In relation to those approaches it can award with several benefits, such as higher estimation/prediction accuracy, better calibration properties, enhanced extrapolation properties, more efficient model development or better interpretability (for details see – supplementary material – section 1). The main advantage is a higher benefit/cost ratio to solve complex problems, which is a key factor for process systems engineering.

Problems in the application of hybrid semi-parametric models mostly concern the model implementation and especially the implementation of the algorithms for the parameter identification is error prone and laborious. However, once a general hybrid semi- parametric modeling tool is implemented, it can easily be reused. It should also be noted that the limitations of mechanistic or nonparametric models may pertain if the hybrid semi-parametric model is not carefully developed or/and the experiments are not carefully designed.

2. Hybrid semi-parametric modeling: the framework

Hybrid semi-parametric models combine nonparametric and parametric models that are based on different types of knowledge. Questions about model configuration, integration of various knowledge types, representation of unknown parts and their identification, best model set-up and requirements on experimental data will be addressed in detail below.

2.1. How to arrange the models? Hybrid semi-parametric model structures

Two models can be arranged in three ways, see Fig. 2, 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 mechanistic information, and the black box consists of a nonparametric model. However, in the serial case, the order of the black and the white model might not be inter- changeable. This is for instance the case when the white box in the serial structure **B** represents a material balance equation in which the kinetic rate term is an input variable that is always computed first by a nonparametric model (e.g. Psichogios & Ungar, 1992). However, the information flow between two serial connected sub- models can be bidirectional.

2.1.1. The parallel structure

The parallel structure **A** usually finds application if a process model (white box) is available, but its performance, due to whatever reasons (e.g. unmodeled effects, nonlinearities, dynamic behavior) is limited. The parallel arrangement of a nonparametric model can lead to significantly improved estimations. Of course the prediction power of the nonparametric model remains poor on input constellations that have not been trained. The parallel approach is especially interesting if certain effects in the system can be uncoupled (e.g. a static nonlinear and dynamic linear behavior as in block-oriented models) and thus each effect can be represented by a separate model (Abonyi, Chovan, Nagy, & Szeifert, 1999; Chen, Hontoir, Huang, Zhang, & Morris, 2004; Klimasauskas, 1998; Masri, 1994; Narendra & Parthasarathy, 1990; Potocnik & Grabec, 1999; Su & McAvoy, 1993).

There exist several possible manners to combine the outputs of the two models, as reviewed in more detail in Section 2.2.2. However, pure superposition is the most frequently applied, i.e. the summation of the outputs, in which case the nonparametric model predicts the residual between the white-box model and the experimental data (Su et al., 1992; Thompson & Kramer, 1994).

2.1.2. The serial structure

The most popular serial combination is structure **B**, Fig. 2. In this structure the white box usually represents a model derived from first-principles such as the conservation laws, namely material, momentum, impulse, population or energy balances derived for the process at hand. The black-box usually represents the under-lying kinetic or transport terms, because it is much more difficult to establish a generally valid model representation at an acceptable cost.

This serial structure **B** is especially suitable when no precise knowledge about the underlying mechanisms is available, but sufficient process data exists to infer the unknown patterns. Also large data sets, rich in information about the process state but without direct physical interpretation, can be exploited by using this data as inputs to the nonparametric model, which might improve the estimations of the kinetics (Teixeira, Carinhas, et al., 2007; von Stosch, Oliveira, Peres, & Feyo de Azevedo, 2011a).

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 & Filho, 2001; Hwang et al., 2009; Mahalec & Sanchez, 2012; Nascimento, Giudici, & Scherbakoff, 1999; Quiza, Lopez-Armas, & Davim, 2012; Schenker & Agarwal, 2000; Zhang, Pan, Quan, Chen, & Shi, 2006). In general, hybrid structure **C** models did not find much application in chemical or biochemical engineer- ing, but in mechanical engineering (Quiza et al., 2012).

2.1.3. Parallel or serial?

Whether a serial or a parallel hybrid semi-parametric model structure is more suitable for a given application, depends strongly on the structure of the whitebox model, since it imposes an inductive bias on the final model (Psichogios & Ungar, 1992), i.e. the assumptions included in the white-box model constrain the possible solution space. When the structure of the white-box model is not accurate, the parallel arrangement can perform better than the serial one, since the parallel nonparametric model can partially compensate for the white-box model structural mismatch, as for instance in (Bhutani, Rangaiah, & Ray, 2006; Lee, Jeon, Park, & Chang, 2002). Due to the fact that extrapolation properties are mostly determined by the underlying model structure (Braake et al., 1998; Fiedler & Schuppert, 2008; Schuppert, 1999; van Can et al., 1998, 1999; van Can, Hellinga, Luyben, Heijnen, & Te Braake, 1996; van Can, te Braake, Hellinga, & Luyben, 1997), the serial structure in such a case (structural mismatch) cannot be expected to perform well, i.e. a suitbale nonparametric model will probably perform better, e.g. Bhutani et al. (2006), and Lee et al. (2002). When the white-box model structure is accurate, then the prediction quality of the serial model can be expected to be considerably better than that of the parallel model, see e.g. Conlin,

Peel, and Montague (1997) and also the extrapolation properties of the serial model will be significantly better, see e.g. van Can et al. (1996). In Corazza, Calsavara, Moraes, Zanin, and Neitzel (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 mechanistic knowledge.

2.1.4. One-step or multi-step ahead prediction

The structure of dynamic hybrid semi-parametric models can either enable onestep ahead or multi-step ahead prediction, Fig. 3, regardless whether the structure is serial or parallel. When measured quantities are used as inputs, then 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 nature of the problem and the availability of measurements which structure is to prefer or can at all be applied. In some cases both one-step and multi-step structures are feasible. In van Can et al. (1998) the hybrid semi-parametric 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 analyzed for a serial hybrid semi-parametric model in von Stosch et al. (2011a). They observed that those models that feedback the predictions provide, in general, enhanced predictions when compared to strict feed-forward models.

2.2. What kind and in what way can information be integrated in hybrid semiparametric models?

While the overall hybrid structure is usually assessed to categorize the hybrid approach into parallel or serial, the substructures can be versatile according to the nature of the incorporated knowledge. Additional knowledge can reduce and structure the space spanned by variables and the parameters of the nonparametric model (Fiedler & Schuppert, 2008; Thompson & Kramer, 1994). 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) can be obtained for the hybrid semi-parametric model.

2.2.1. Mechanistic knowledge incorporation

Mechanistic knowledge inherits a high degree of knowledge abstraction. The domain in which this knowledge describes the system accurately is relatively large. Incorporation of mechanistic knowledge into hybrid semi-parametric models was observed to have the potential to improve model performance considerably (Al-Yemni & Yang, 2005; Vande Wouwer, Renotte, & Bogaerts, 2004; von Stosch et al., 2011a). A differentiation might be drawn between structuring knowledge, in the sense that the interplay between different components is decoupled, and forming

knowledge which describes the form of an interaction. However, such a differentiation is vague, since the effect that the incorporation of the knowledge has, might be indistinguishable. The following short example provides an intuitive understanding of this differentiation.

In a biochemical fed-batch process, the concentration of biomass *X* usually increases along time while substrate *S* is taken up. For the modeling of the biomass growth, *rBiomass*, or of the substrate consumption, *rSubstrate*, biomass is taken to be a catalyst wherefore the rates are formulated as a product of the biomass concentration *X* with the specific growth rate μ or with the specific substrate uptake rate *vs*, i.e. *rBiomass* = *X* μ or *rSubstrate*=*X vs*, respectively. This type of integrated knowledge, which describes the interaction of two varying variables, is hereby classified as forming knowledge. The specific rates μ and *vs* can, in addition, be assumed to be coupled by the biomass yield on substrate *Ysx*, i.e. μ = *Ysx vs*. This type of knowledge, which describes the relationship of two variables, is hereby classified as structural knowledge.

2.2.1.1. Forming knowledge. By the incorporation of forming knowledge into a hybrid semi-parametric model, the extrapolation properties can be shaped and the function that the nonparametric model has to learn might be simplified. In the example above, the incorporation of the assumption that biomass is a catalyst (e.g. Oliveira, 2004; Psichogios & Ungar, 1992; Schubert, Simutis, Dors, Havlik, & Luebbert, 1994a; Vande Wouwer et al., 2004) facilitates the learning of the rate functions μ or vs in relation to *rBiomass* or *rSubstrate*. Further, the impact of X on the rates is explicitly captured, which allows to predict the rates for X values beyond those which the model is trained on, i.e. extrapolation. Similarly, Reuter, Van Deventer, and Van Der Walt (1993), who proposed a general reaction schema for batch and continuous mineral and metallurgical processes, observed that the proposed schema could represent the reactor performance in various situations. More forming knowledge can also be incorporated using "standard" formulations of the kinetic rates, representing the contained kinetic parameters by nonparametric model expressions (Al-Yemni, 2003; Bellos, Kallinikos, Gounaris, & Papayannakos, 2005; Kasprow, 2000; Mazutti et al., 2010). In addition, forming knowledge can also lead to Bounded Input Bounded Output (BIBO) stability properties of the model (Karama, Bernard, & Gouz, 2010; Oliveira, 2004), e.g. a reaction can only occur when all reactants are present.

2.2.1.2. Structural knowledge. The incorporation of structural knowledge into a hybrid semi-parametric model can facilitate the identification of the parameters and it might reduce the number of rate expression which are modeled by nonparametric techniques. In the given example the number of rates could be reduced from two (μ and vs) to one (either μ or vs) by the incorporation of the

yield coefficient Ysx. In addition, the identification of the remaining rate is facilitated, since (i) less parameters have to be identified while the number of data points stays constant; and (ii) the redundancy of the nonparametric model structure, which generally poses a problem for identification (Bishop, 1995), is reduced. Further, as for instance demonstrated in Mogk, Mrziglod, and Schuppert (2002), the incorporation of structural information can improve the prediction and modeling accuracy. Stoichiometry or yield coefficients (Brendel & Marquardt, 2008; Chen, Bernard, Bastin, & Angelov, 2000; Georgieva & de Azevedo, 2009; Vande Wouwer et al., 2004) present structural information. For the identification or adaption of the stoichiometric coefficients a state transformation technique (Georgieva & de Azevedo, 2009; Vande Wouwer et al., 2004) or a target factor analysis (Brendel & Marquardt, 2008) can be applied. In contrast, the problem for the integration of metabolic networks into the hybrid semi-parametric model, is not the stoichiometry but the large number of under-determined reactions fluxes. Teixeira, Alves, et al. (2007) computed Elementary Flux Modes (EFM) for a simple metabolic network and integrated the gained structural information about the most important EFMs along with the stoichiometry into a hybrid semi-parametric model. Fiedler and Schuppert (2008) addressed the integration of knowledge into a tree-structured scalar hybrid semiparametric model, in which several parametric and nonparametric models can be integrated (Identifiability is also addressed). It is theoretically assessed that such a structure can avoid the curse of dimensionality of strictly nonparametric structures and that it can inherit better extrapolation capabilities.

2.2.2. Combination of incorporated information

Two general ways to fuse the model outputs are superposition, Fig. 4a, (as in most parallel structures) and multiplication, Fig. 4b, (as proposed in Oliveira, 2004). If however the same quantity is predicted by two different techniques, Fig. 4c, then other fusion approaches must be considered.

Weighting methods can be used, as e.g. for parallel structures (Fellner, Delgado, & Becker, 2003; Johansen & Foss, 1992a, 1992b; Klimasauskas, 1998; Su & McAvoy, 1993). Dors, Simutis, and Luebbert (1995, 1996) applied a weighting function in a serial hybrid semi-parametric 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, Simutis, & Luebbert, 2004), where more weight is given to the nonparametric model in regions where process data are available, while restricting it when extrapolating. This weighting method was also applied by Patnaik (2010), who however determined the weighting iteratively. As an extension to this the Mixture of Experts framework proposed by Peres, Oliveira, and Feyo de Azevedo (2001) can be understood. This framework consists of several

parallel submodels, 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 which the gating function has its analogy in the rules (attendance part) and the submodel in the Fuzzy con- sequent part. However, the identification of the parameters in the mixture of experts approach is considerably 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, see e.g. Peres et al. (2001). Another option for weighting different predictions of the same quantity, Fig. 4c, is to use a nonparametric model, where all predictions are inputs to the nonparametric model and only the final prediction is the output (Bollas et al., 2003; Cao, Wang, Fujii, & Tobler, 2004). The options shown in Figs. 4d or e, which are somewhat similar to the parallel schema presented in Fellner et al. (2003) or in Su and McAvoy (1993), respectively, have so far not been applied for combining the submodels in serial hybrid semi-parametric models. However, the kinetic rate predictions of the nonparametric model may be constrained in order to pertain to physical limits. This combination could be interpreted to be the one shown in Fig. 4e.

2.2.3. Operational knowledge - rule based information

Fuzzy systems make use of a logic structure to describe certain rule-alike procedures, e.g. if glucose concentration is high, then biomass growth rate is high; else if glucose concentration is low, then biomass growth rate is low. The expressions (low or high) are associated to parameters that can either be determined manually, through the experience of an operator, or can be fitted to exper- imental data (Roubos, 2002; Roubos et al., 2000; Schubert et al., 1994a; van Lith, Betlem, & Roffel, 2002, 2003).

One of the most popular Fuzzy models is the Takagi, Sugeno Kang type (Takagi & Sugeno, 1985), in which the consequent part of each rule consists of a linear equation, (van Lith et al., 2002). There- fore 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, and therefore they can be used instead of e.g. neural networks. The biggest advantage of Fuzzy models, when compared to more data-driven techniques, is that they are interpretable, wherefore they can offer transparency in situations where physical models are difficult to derive (van Lith et al., 2002, 2003). However, for their derivation considerable more knowledge is required than for other data-driven models.

The integration of Fuzzy models along with first-principles knowledge can, as before, be accomplished in parallel (Abonyi et al., 1999; Fu & Barford, 1995b) or in series (van Lith et al., 2002, 2003; Vieira, Dias, & Mota, 2005). Moreover they can be complementarily combined into an existing hybrid approach, e.g. in parallel to

a non- parametric model (Dors et al., 1995, 1996; Peres et al., 2001) where a gating function decides 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 & Beluhan, 2000; Preusting, Noordover, Simutis, & Luebbert, 1996; Schubert et al., 1994a; Simutis, Havlik, Schneider, Dors, & Luebbert, 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 et al., 1996; Roubos, 2002; Roubos et al., 2000; Schubert et al., 1994a; Schubert, Simutis, Dors, Havlik, & Luebbert, 1994b; van Lith et al., 2003) for examples.

2.3. How can unknown parts be represented? - 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 semi-parametric model its flexibility,

e.g. to model systems with partially unknown underlying effects. The most frequently applied nonparametric models, are the MultiLayer Perceptron (MLP) and the Radial Basis Function Network (RBFN) (see supplementary material Table 1). Both provide equally good predictions (in favor of the former (James, Legge, & Budman, 2002)), but due to differing standard training methods, the training takes considerably longer for MLPs than for RBFNs. The advantage of MLPs is that the outputs (model) do not need to be known explicitly for the training. This is especially important for the serial structure **B**, since e.g. the kinetic rates are not directly measured and their calculation from sparse, infrequent noisy concentration measurements is error prone. The advantage of RBFNs is that they have certain, inherent stability characteristics, which make them suitable for control and monitoring, (James et al., 2002).

2.3.1. Nonparametric models for specific problems

Different situations call for the incorporation of different non- parametric approaches into the hybrid semi-parametric model. Some authors proposed to use more than one nonparametric model. Tian, Zhang, and Morris (2001) used stacked neural net- works in a parallel hybrid structure and they found that those stacked networks provide better predictions than a single neural network. In a similar manner Bollas et 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.

The concept of using more than one neural network was also explored in serial hybrid semi-parametric models (Cao et al., 2004; Gnoth, Jenzsch, Simutis, & Luebbert, 2008; Gupta et al., 1999; Patnaik, 2001, 2003, 2010; Piron, Latrille, & Rene, 1997; Preusting et al., 1996; Reuter et al., 1993; Silva, Cruz, Hokka,

Giordano, & Giordano, 2000, 2001). Preusting et al. (1996) used two ANNs in parallel to model separate phenomena, i.e. one ANN to model the kinetics another to model the viscosity. Gupta et al. (1999) applied two parallel ANNs, each of which inferring a variable value, in series with three other parallel ANNs, each of which estimating a quantity that enters as an input to the mechanistic model. In Gnoth et al. (2008), Silva et 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 semi-parametric model with one individual nonparametric model, as e.g. done by Patnaik (2001, 2003, 2010), Piron et al. (1997), Saraceno, Curcio, Calabro, and Iorio (2010) can help to make the model structure more transparent, and increase the accuracy of each predicted quantity. Differently Cao et al. (2004) applied two individual non-parametric models to predict the same quantity with each model relying on different phenomena, i.e. the inputs are different. The performance of various other nonparametric techniques has also been tested in hybrid semi-parametric models, as can be seen in Table 1 (supplementary material).

2.3.2. Comparison of nonparametric models

Comparisons between different nonparametric models that were embedded in the same hybrid semi-parametric model structure have been carried out (see supplementary material – section 2), but the findings are sometimes contradicting. This might be due to the fact that the performance of the nonparametric model is highly problem dependent (what kind of function should be approximated, 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.

2.4. How can unknown parts be identified? - Methods for model identification

The identification of the "unknown" parts of the hybrid semi- parametric model most times comprises only the identification of the nonparametric model (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, Mao, He, & Yang, 2011; Kahrs & Marquardt, 2008; Vande Wouwer et al., 2004). While, in principle the same identification schema can be applied when also other parameters are unknown, e.g. yield/stoichiometric coefficients, it might,

in this case, be beneficial to decompose the identification since e.g. the initial values of the parameters might be known, which can simplify the identification. Approaches explicitly dealing with this scenario are given in Vande Wouwer et al. (2004), Kahrs and Marquardt (2008), Yang, Martin, and Morris (2011) the latter is shortly presented in Section 2.4.2.

In case of the serial structure **C**, Fig. 2, or the parallel structure **A** the identification of the nonparametric models can be carried out with standard techniques (e.g. back-propagation for MLPs (Werbos, 1974)).

In case of the serial hybrid structure **B**, Fig. 2, 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 et 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.

2.4.1. 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 & Ramirez, 1996) or through smoothing spline approx- imations (Schubert et al., 1994a). With these calculated outputs and the available inputs, standard techniques can be used for the parameter identification. However, a fact that has found little 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. The calculated kinetic data might be biased and when using these data for parameter identification, the bias might be passed on to the model.

2.4.2. 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, thereby reducing the curse of dimensionality. During this phase standard training techniques for the identification of the nonparametric model can be used. Once the four sub-identifications are accomplished, 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 approach described in Chen et al. (2000) is similar, in that the problem is decomposed, but not to the entirety of the incremental approach.

2.4.3. The indirect approach – the sensitivities equations

Right from the beginning of serial hybrid semi-parametric modeling, a method for the identification of the neural network weights was required. Psichogios and Ungar (1992) adapted the well- known error back-propagation technique (Werbos, 1974) using the sensitivities equations. Schubert et al. (1994a) and Oliveira (2004) 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 in the direct approach 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, which can result in better noise rejection properties (von Stosch, Oliveira, Peres, & Feyo de Azevedo, 2011b).

2.4.4. Other alternative approaches

Gradient free parameter identification procedures, requiring only the model residual (between the data and model estimate) but in turn increasing the computational costs (Roubos et al., 2000), were applied (Madar, Abonyi, & Szeifert, 2004; McKay, Sanderson, Willis, Barford, & Barton, 1998; Roubos et al., 2000). However, with the ever increasing computation power and the fact that several random initiation of the parameters might not be required, this is an attractive solution for relatively small systems.

An approach seeking to identify the "optimal parameters" by removing data that are not rich in information from the sample space, was proposed for parallel structures by Potocnik and Grabec (1999). However, while the model fit might be good locally, e.g. for certain fermentation phases, the overall process representation might suffer.

All above mentioned identification schema are batch-learning techniques, i.e. all training data are used at the same time to infer the parameter values. An alternative is incremental learning, which can be used to adapt the network weights on-line, given that the state measurements become available on-line or are other- wise observable (Dochain, 2003). For closed-loop control, on-line parameter adaptation can increase the performance, due to bet- ter local approximations. However, for parameter identification on a entire process operation region, i.e. for global approximations, batch learning is usually preferred.

2.4.5. General remarks about the identification

Two well-known identification problems are over-fitting and local minima. The former is normally addressed with early- stopping, cross-validation or with 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, choosing the parameter set which performs best on additional data, which have not been used for parameter identification (Simutis & Luebbert, 1997; van Can et al., 1996, 1997; Vande Wouwer et al., 2004).

Convergence and success of the identification depend on the initialized parameter values (Kahrs & Marquardt, 2008; von Stosch, Oliveira, Peres, & Feyo de Azevedo, 2012). Relatively small weight values are preferential (because generally inheriting better generalization capabilities), wherefore the initialization values of the parameters are usually constrained, e.g. smaller than one, greater than minus one. Additionally, in case that only few experimental values are available and 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 et al., 2004; Graefe et al., 1999; Henriques, Costa, Alves, & Lima, 1999; Tsen, Jang, Wong, & Joseph, 1996).

Whenever the conservation laws are posed in the form of Ordi- nary Differential Equations (ODEs), then 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 et al., 2011a). 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 et al. (2004) proposed to include the initial values into the set of parameters (after those have been optimized to a certain threshold) and to, thereafter, optimize all those values together.

2.5. What model is performing best? – Model discrimination and extrapolation capabilities

The model structure, its generalization and 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 dimension.

2.5.1. Nonparametric model discrimination

The discrimination of the nonparametric model structure (e.g. for MLP the number of hidden layers and the therein covered numbers of nodes, or in case of Partial Least Square/Projection to Latent Structures (PLS) the number of latent variables) can be addressed with the Akaike Information Criterion or Bayesian Information Criteria, the latter being more suitable for models with large numbers of parameters (Lee, Vanrolleghem, & Park, 2005; Peres, Oliveira, & de Azevedo, 2008; von Stosch, Peres, de Azevedo, & Oliveira, 2010). Also other statistical criteria can be applied (Bollas et al., 2003; Kim & Chang, 2000) to evaluate the estimations 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. The number of parameters and the identified "optimal" parameter values impact on the quality of prediction, generalization and extrapolation. Integration of knowledge can significantly reduce the size of the nonparametric model, while enhancing the extrapolation properties Mogk et al. (2002). In any case, it is advisable to manually assess at least the model properties of the best candidate structures (Braake et al., 1998).

2.5.2. Hybrid semi-parametric model structures and extrapolation

A systematic investigation on the hybrid semi-parametric model extrapolation properties was conducted by van Can et al. (1996, 1997, 1998, 1999), distinguishing between four scenarios, as shown in Fig. 5:

· dimensional extrapolation (A variable, which was kept constant during identification, varies during the application of the model van Can et al., 1998),

· range extrapolation (A variable is applied outside the range

within which it was varied during identification van Can et al., 1998),

 \cdot interpolation (A variable is constant during the identification and

application of the model and its amplitude during application is between the highest and lowest amplitude during the identification van Can et al., 1998)

· and frequency extrapolation (A variable is used at a frequency

that is lower or higher than the lowest or highest frequency in the identification experiments van Can et al., 1998).

When testing serial and parallel hybrid semi-parametric models, through their incorporation into a model predictive control scheme, experimentally for their dimensional extrapolation properties, van Can et al. (1996) observed that the serial hybrid semi-parametric model showed good dimensional extrapolation properties. These properties were found to be due to the accurately known terms in the balances. The parallel hybrid semi-parametric models, in contrast, did not show any advantage compared to strictly nonparametric models. Studying different levels of incorporated mechanistic knowledge, van Can et al. (1998) found that due to the accurately known terms in the balances, good range, dimensional and reliable frequency extrapolation properties were yield; (ii) the unknown terms could relatively easily be identified from the available data; and (iii) in comparison to more data-driven models, the serial hybrid structures have better extrapolation properties. Thus with the same identification data, the model can be applied to a much wider range of conditions, which also means that a smaller domain of identification data is required for serial hybrid models, limiting the experimental effort. Ergo a strong connection between the model properties and the identification data exists, which will be the subject of the experimental data section.

2.5.3. Measures for model extrapolation

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

Klimasauskas (1998) proposed to apply some measure, i.e. a confidence module, to restrict the influence of the nonlinear model on the prediction when extrapolating (although the details are not provided). In Simutis et 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 et al. (2005) clustering of the nonparametric model inputs is carried out. Then, the optimization is constrained by a user defined risk (typically 25%) taking the minimal distance between the inputs obtained during the optimization and the closest cluster mean into account. Mahalec and Sanchez (2012) propose to constrain the optimization by two measures, one accounting for the distance of the current inputs to historical ones and a second ensuring that the residual and bias of the predictions in relation to the model plane do not exceed a certain threshold. Similarly, in Kahrs and Marquardt (2007) two complementary criteria to check the validity domain of hybrid semi-parametric models are proposed: (1) A convex-hull criteria to check whether each empirical model part only interpolates the data encountered during model identification; and (2) a confidence interval criterion with which the confidence intervals for the hybrid semi-parametric model are calculated. In comparison to 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 nonlinear 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.

A shortcoming of all these criteria is their focus on the distribution of points in the space while they do not account for the transient behavior (frequency extrapolation). Investigations in this respect are especially interesting in cases where the transient behavior is of importance such as for controller tuning (von Stosch et al., 2012).

2.6. What is the influence of the data? – Experimental data and data pre-treatment

Data are necessary to identify the structure and the parameters of the hybrid semi-parametric model and basically all model properties (prediction quality,

model operation range, interpolation capabilities) depend strongly not only on the quantity but also on the quality of the data.

2.6.1. Design of experiments

In industrial settings the attitude tellingly described by Sohlberg (2005), i.e. "you have to take what you can get" is dominant, but to yield high quality data the design of experiments should correspond to the objectives (Simutis, Oliveira, Manikowski, de Azevedo, & Luebbert, 1997). In van Can et 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, though it is rather unrealistic to know these in advance. 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 highly valuable (Chang, Lu, & Chiu, 2007; Gupta et al., 1999; Saraceno et al., 2010; Thibault et al., 2000; Tholudur & Ramirez, 1999; Tholudur, Ramirez, & McMillan, 2000). Another option, if at least some knowledge or data are available from which a first hybrid semi-parametric model can be derived, is to apply the coverage approach proposed by Brendel and Marquardt (2008), which proved to be better than a factorial design.

A different strategy is iterative batch-to-batch optimization, where the experiments are performed in such a way, i.e. the degrees of freedom are controlled in such a manner, as to meet some objective, e.g. Doyle, Harrison, and Crowley (2003) and Teixeira, Clemente, Cunha, Carrondo, and Oliveira (2006) or Section 3.4). It is of course rational 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 et al., 2006).

2.6.2. Experimental data pretreatment

Experimental data can enter into the hybrid semi-parametric model in two ways: (1) as inputs to the nonparametric submodel; and (2) 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 Schubert et al. (1994a) and Chabbi, Taibi, and Khier (2008) that variances in the feeding concentration can cause big errors in the estimation of the respective substrate concentrations. Similar observations were made by vonStosch et al. (2011a). Studies on the impact of different levels of experimental noise on the identification results, performed by Yang et 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. Laursen, Webb, and Ramirez (2007), for instance, proposed to use a smoothing cubic spline function to account for the noise in the feeding rate data. There are, however, 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. The pretreatment of those data that are inputs to the nonparametric model, was found to improve the nonparametric model performance (Bishop, 1995). In any case the nonparametric model input values should be scaled to a range between zero and one, e.g. by subtracting the mean and dividing by the standard deviation (Bishop, 1995).

3. Application of hybrid semi-parametric modeling

3.1. Modeling

Modeling provides the ground for process operation and design, such as monitoring, control, optimization or scale-up. The focus in this section is on modeling applications that deal with experimental data rather than data from virtual, simulated experiments, because simulation cases are typically applied to validate a pro- posed methodology (methodologies have been discussed above), while experimental studies are much more practically oriented. It can be seen in Fig. 6 that about half of the applications in the area of chemical and biochemical engineering, which are the areas with the most applications, are built upon experimental data. According to the number of representations, the section is in the following divided into applications in chemical engineering and biochemical engineering. Applications on other areas (e.g. water treatment processes Anderson, McAvoy, & Hao, 2000; Conlin et al., 1997; Karama et al., 2010; Lee et al., 2002, 2005 or food and beverages Simutis et al., 1995; Teissier, Perret, Latrille, Barillere, & Corrieu, 1997) can be found in the supplementary material, Tables 4-6. Several applications exist also in mechanical engineering, e.g. Masri (1994), Cao et al. (2004) and Quiza et al. (2012), but those are out of the scope of this review.

3.1.1. Chemical engineering

Hybrid semi-parametric modeling applications in chemical engineering deal for instance with the chemical reactor (Bellos et al., 2005; Bollas et al., 2003; Gupta et al., 1999; Luo, Du, Ye, & Qian, 2012; Molga & Cherbanski, 1999; Porru, Aragonese, Baratti, & Servida, 2000; Qi, Zhou, Liu, & Yuan, 1999; Simon, Fischer, & Hungerbuehler, 2006; Xiong & Jutan, 2002; Zahedi, Lohi, & Mahdi, 2011), polymerization processes (Bhutani et al., 2006; Feil et al., 2004; Fiedler & Schuppert, 2008; Hinchliffe, Montague, Willis, & Burke, 2003; Mogk et al., 2002; Tian et al., 2001; Tsen et al., 1996; Vega, Lima, & Pinto, 2000), crystallization (Georgieva & de Azevedo, 2009; Georgieva, Meireles, & Feyo de Azevedo, 2003; Lauret, Boyer, & Gatina, 2000), metallurgic processes (Hu et al., 2011; Jia, Mao, Chang, & Zhao, 2011; Reuter et al., 1993; Sohlberg, 2005), distilla- tion columns (Chen et al., 2004; Mahalec & Sanchez, 2012; Safavi, Nooraii, & Romagnoli, 1999), drying processes (Cubillos & Acuna, 2007), thermal devices (Arahal, Cirre, & Berenguel, 2008), mechan- ical reactors (Nascimento et al., 1999) or milling (Aguiar & Filho, 2001; Kumar Akkisetty, Lee, Reklaitis, & Venkatasubramanian, 2010), for more details and references see Table2 – supplementary material. Since the number of applications is relatively large and most of which either use the standard serial approach consisting of material and/or energy balances in which the kinetics are represented by a nonparametric model or a parallel approach, only some approaches are discussed here, namely those that present solutions of more complex problems.

Particle size distribution, which is of major interest in many processes, can be modeled with population balances, as for instance in crystallization (Georgieva et al., 2003; Hermanto, Braatz, & Chiu, 2011; Lauret et al., 2000; Zhang, Wang, He, & Jia, 2012), milling (Kumar Akkisetty et al., 2010) or polymerization (Doyle et al., 2003). The application of a complementary nonparametric model, e.g. in order to enhance the prediction quality, can also be beneficial in this context. While a parallel set-up (Doyle et al., 2003; Hermanto et al., 2011; Zhang et al., 2012) is relatively easy to apply and might be sufficient in many cases, a serial approach can help to under- stand the complex interactions. For example, Georgieva et al. (2003) model the most uncertain parts in a set of material, energy and population balances, namely the agglomeration kernel, the nucleation and growth rate, through nonparametric techniques. 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 et al., 2010).

In certain situations it might be necessary or desired to account for gradients in the temperature or concentration distribution along a spatial component. In Gupta et al. (1999) the material balances are formulated for the phosphate particles along the height of a flotation column, resulting into Partial Differential Equations. The reaction rate parameters in those balances, namely the flotation rate constants, 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 et al. (2011). In Dadhe, Rossmann, Durmus, and Engell (2001) the distillation column is divided into several 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 ODEs. The vapor-liquid equilibrium is in this approach described by a RBFN. Similar approaches are also pro- posed by Mahalec and Sanchez (2012), Hinchliffe et al. (2003) and Arahal et al. (2008), where Hinchliffe et al. (2003) divides the polymerization reactor into several stages whereas Arahal et al. (2008) uses discrete volume and wall segments. 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 represented by a serial hybrid semi-parametric model based on overall material

and energy balances. Similar observations were also made by Qi et al. (1999) who compared hybrid semi-parametric models to detailed mechanistic twodimensional models, finding that the hybrid is simpler in model structure, has lower computational costs and provides about the same prediction quality.

3.1.1. Biochemical engineering

Hybrid semi-parametric modeling is frequently applied in bio- chemical engineering, e.g. for the modeling of yeast fermentations (Beluhan & Beluhan, 2000; Boareto, De Souza, Valero, & Valdman, 2007; Eslamloueyan & Setoodeh, 2011; Mazutti et al., 2010; Peres et al., 2001; Saraceno et al., 2010; Saxen & Saxen, 1996; Schubert et al., 1994a, 1994b), for modeling of fungi cultivations (Chen et al., 2000; Ignova et al., 2002; Preusting et al., 1996; Silva et al., 2000, 2001; Thibault et al., 2000; van Can et al., 1997, 1998; Wang, Chen, Liu, & Pan, 2010), for modeling of bacteria cultivations (Costa, Alves, Henriques, Filho, & Lima, 1998; Gnoth et al., 2008; Henneke, Hagedorn, Budman, & Legge, 2005; Henriques et al., 1999; James et al., 2002; Jenzsch, Gnoth, Kleinschmidt, Simutis, & Luebbert, 2007; Laursen et al., 2007; Roubos et al., 2000; Simutis & Luebbert, 1997; Thibault et al., 2000; Tholudur & Ramirez, 1999; Zuo, Cheng, Wu, & Wu, 2006; Zuo & Wu, 2000), for modeling of mammalian cell cultivations (Dors et al., 1995, 1996; Simutis et al., 1997; Teixeira, Alves, et al., 2007; Teixeira et al., 2005; Vande Wouwer et al., 2004), for modeling of insect cell cultivations (Carinhas et al., 2011), for modeling of hybridoma cell cultivations (Fu & Barford, 1995a, 1995b) or for modeling the counter-ion fluxes across an ion-exchange membrane in a membrane-supported biofilm reactor (Ricardo, Oliveira, Velizarov, Reis, & Crespo, 2012); more details on these models can be found in Table 3 - supplementary mate- rial. Most of these approaches follow the original approaches of (Psichogios & Ungar, 1992; Schubert et al., 1994a). The under-lying biological system - the cell, which houses highly complex chemical reaction networks and transport mechanisms, is usually modeled assuming lumped kinetics. Along with these, biomass is considered to be a catalyst to the reactions and then either specific kinetic rates are directly modeled by nonparametric techniques or after some knowledge has been incorporated only the "miss- ing" parts are represented by nonparametric models (Al-Yemni, 2003; Corazza et al., 2005; Fu & Barford, 1995a; Kasprow, 2000; Mazutti et al., 2010). The incorporation of additional knowledge can, as described above, improve the model properties, especially when the knowledge structures the model such as stoichiometric coefficients, wherefore current efforts reach out to integrate knowledge from systems biology (Teixeira, Alves, et al., 2007). However, besides the modeling of the reactor system with ODEs also other approaches can be found.

A crossflow microfiltration process of a suspensions of baker's yeast using a serial hybrid semi-parametric model is considered in Piron et al. (1997). They derived a physical model for the microfiltration process, wherein those

parameters that are unknown (cake resistance, cake diffusion interface and the concentration gradient) are represented by ANNs.

In Thibault et 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 FNN.

The production process of bacterial cellulose with a pilot scale airlift reactor is, in Zuo et al. (2006), decomposed into two models, which are separately identified. The first is a standard serial hybrid semi-parametric ODE model 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 tubes, taking into consideration the hydrodynamic effects. Good results are obtained with both models and so the whole airlift reactor cultivation is appropriately represented.

3.2. Monitoring

For monitoring hybrid semi-parametric modeling has been applied in two ways, as schematically depicted in Fig. 7. One way is to predict certain quantities from available on-line measurements and/or model's own predictions, which is referred to as Predictor or Soft-sensor. The other way applies the hybrid semiparametric model along with a corrector method to correct the state predictions and eventually to adapt the model parameters (Corrector).

3.2.1. Soft-sensor – predictor method

The application of a model to estimate a certain quantity based on at-time available measurement is referred to as soft-sensor. The application of hybrid semi-parametric models in form of a soft- senor is very attractive for monitoring and both parallel (Lee et al., 2005) and serial (Boareto et al., 2007; Gnoth et al., 2008; Henneke et al., 2005; James et al., 2002; Jenzsch et al., 2007; Psichogios & Ungar, 1992; Schubert et al., 1994a; Silva et al., 2000, 2001; von Stosch et al., 2011b) hybrid semi-parametric models find application. It was shown that the performance of a model in which the states and parameters were estimated by Nonlinear Programming (NLP) optimization or Extended Kalman Filter (EKF) approaches was inferior to the performance of a model in which the variable parameters were estimated using neural networks (Psichogios & Ungar, 1992), namely a hybrid semi-parametric model. Similar to modeling applications, the hybrid semi-parametric model might outperform Linear models, FNNs or RBFNs for monitoring (James et al., 2002). In addition the hybrid semiparametric model might not only be used to monitor the process, but also to derive the set- points for the control (Jenzsch et al., 2007). The requirements for the application of hybrid semi-parametric models that are based on the dynamic formulation of material or energy balances are that (i) the sampling rate of the attime available measurements is more or less constant (a requirement that stems

from the numerical integration); (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 interpolation method is required. When these requirements are met then a hybrid semi-parametric model can in principle provide better pre- dictions than other models, since (a) either fewer parameters are required to achieve similar prediction gualities (when compared to strictly nonparametric models) which reduces the statistical uncertainty or, when compared to strictly mechanistic models, the model can benefit from the actual process conditions, reflected through sets of at-time available measurements; and (b) the hybrid semi- parametric model has better calibration properties. For the serial hybrid semi-parametric model it can further be stated that (c) the numerical 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 et al., 2011b); 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 semi-parametric models' sensitivity to noise, (von Stosch et al., 2011b).

Table 7, supplementary material, comprises a list of hybrid semi- parametric soft-sensor applications.

3.2.2. Corrector scheme

The corrector scheme finds application when the state variables (which are considered in the material or energy balances) are measured at some instances during the process, since the predictions can be corrected and/or the model parameters can be adapted. However, the corrector method is subject to certain restrictions regarding the state observability (Dochain, 2003). The underlying hybrid semi-parametric model can either rely 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 semi-parametric model is serial and uses at-time available measurements the same requirements formulated above for the soft-sensor case hold.

Standard corrector schema that found application comprise the EKF (Porru et al., 2000; van Lith et al., 2002; Wilson & Zorzetto, 1997) or a DDI filter (Feil et al., 2004). For a certain class of serial hybrid semi-parametric models, a state transformation technique can be applied for (i) inference of unmeasured states, (ii) on-line state correction and (iii) ANN weight adaptation (Georgieva & de Azevedo, 2009).

The correction of the hybrid semi-parametric model (consisting of a material balance based model in parallel to a block-wise PLS scheme) predictions applying a rectification method, which can utilize the off-line, time-lagged measured samples, was proposed by Jia et al. (2011). They state that the predictions of this adaptive hybrid semi-parametric model are more accurate and

efficient than the same model without adaption or a recursive PLS model.

3.3. Control

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

For closed-loop control, there are two possible ways to exploit the hybrid process model (von Stosch et al., 2012), namely (i) by employing a control structure that directly uses the hybrid process model equations for the calculation of the control action; or

(i) by the application of the hybrid semi-parametric model for the controller tuning.

3.3.1. Hybrid semi-parametric model based controller structures

The way in which the control inputs appear in the hybrid semi- parametric model equations determines which control methods can find application. Whenever the process model equations are invertible, i.e. an analytical explicit expression can be obtained through manipulation, direct Feedback Linearizing Control (FLC) (Bazaei & Majd, 2003), Generic Model Control (Abonyi, Madar, & Szeifert, 2007; von Stosch et al., 2012; Xiong & Jutan, 2002) or Model (Adaptive) Reference Control (von Stosch et al., 2012) schema can be applied. These methods can account for nonlinearities and are relatively computationally inexpensive. In case that the process model equations are not invertible, FLC (Bazaei & Majd, 2003; Hussain, Ho, & Allwright, 2001; Madar, Abonyi, & Szeifert, 2005), sliding mode control (Hussain & Ho, 2004), Model Predictive Con- trol (MPC) (Abonyi et al., 1999; Cubillos, Callejas, Lima, & Vega, 2001; Hermanto et al., 2011; Ibrehem, Hussain, & Ghasem, 2011; Klimasauskas, 1998; Tsen et al., 1996; van Can et al., 1996; Vega et al., 2000; Vega, Lima, & Pinto, 1997), predictive or optimal control (Anderson et al., 2000; Costa et al., 1998; Costa, Henriques, Alves, Maciel Filho, & Lima, 1999; Cubillos & Lima, 1997, 1998; Schenker & Agarwal, 2000; Vieira et al., 2005) schema can be employed, where FLC and sliding mode control are computational less expensive while MPC or optimal control may provide better performance.

When comparing the performances of control schema that utilize hybrid semiparametric models to those using either traditional control methods (such as a self-tuning PID (Xiong & Jutan, 2002), a generalized minimum variance controller (Xiong & Jutan, 2002), a FLC based on a linear model (Hussain et al., 2001) or a MPC based on a linearized model (Anderson et al., 2000)) or to non- parametric model based controllers (Cubillos et al., 2001; Hussain et al., 2001; Ibrehem et al., 2011; Schenker & Agarwal, 2000), it was mostly observed that the hybrid semiparametric model based control schema performed significantly better. However,

the controller performance depends on the limitations of the underlying model. For instance, Anderson et al. (2000) observed that the control performance utilizing the parallel hybrid semi-parametric model was inferior to the one using a linearized model, because the control situation considered had an extrapolative character. That parallel hybrid structures can have poor extrapolation properties, if not restricted by some measure (Klimasauskas, 1998), was already reported in van Can et al. (1996). However, given that the model estimates can be compared to measurements at-time, the model parameters, i.e. mainly the network weights, might be adapted, thus reducing or eliminating the model-plant mismatch (Costa et al., 1998, 1999; Cubillos & Acuna, 2007; Cubillos & Lima, 1997, 1998; Hermanto et al., 2011). In such a case, it might be argued that hybrid semi-parametric models bear no advantage over non-parametric models, since those can be adapted in the same way, but (i) the hybrid semi-parametric model is easier to interpret, wherefore the control action can be scrutinized and (ii) the hybrid semi-parametric model might be easier to adapt, e.g. if the number of parameters is lower (Cubillos et al., 2001).

A list of hybrid semi-parametric model based control applications can be found in Table 8 – supplementary material.

3.3.2. Hybrid semi-parametric model based controller tuning

The hybrid process model can also be exploited to tune any kind of standard controller both, prior to application (off-line) (Georgieva & de Azevedo, 2009; Georgieva & Feyo de Azevedo, 2007; Schubert et al., 1994a; von Stosch et al., 2012); and under control (on-line) (Andrasik, Meszaros, & de Azevedo, 2004; Chen et al., 2004; Patnaik, 2003, 2004, 2008, 2010; Schubert et al., 1994a; Wei, Hussain, & Wahab, 2007; Zhang et al., 2006). The utilization of a hybrid semiparametric model for controller tuning bears some advantages, because (i) the hybrid semi-parametric model might be identified from standard operational data, but the controllers can be tuned considering set-point changes and load disturbances, both scenarios in which the model might have to extrapolate (von Stosch et al., 2012); and (ii) the coupling of different control inputs and/or variables can be captured with little effort by the hybrid semi-parametric model and subsequently considered during the tuning (multivariate control) (von Stosch et al., 2012). Controllers which are tuned in such a way comprise, for instance, neural net- work controllers (Patnaik, 2003, 2004, 2008, 2010; Schubert et al., 1994a; von Stosch et al., 2012; Zhang et al., 2006), the ANN models in a MPC schema (Georgieva & de Azevedo, 2009; Georgieva & Feyo de Azevedo, 2007) or hybrid controllers (Andrasik et al., 2004; Ng & Hussain, 2004; von Stosch et al., 2012; Wei et al., 2007), where the term hybrid semi-parametric controllers is an analogy to hybrid semi-parametric models. Frequently applied standard approaches which incorporate hybrid semi-parametric models comprise Internal Model Control (IMC) (Chen et al., 2004; Schubert et al., 1994a; Wei et al., 2007;

Zhang et al., 2006) and Inverse Model Control (IVMC) (Ng & Hussain, 2004; Wei et al., 2007), see supplementary material – Table 9. In comparison to standard PID control (Ng & Hussain, 2004; Schubert et al., 1994a; Wei et al., 2007) or in comparison to I(V)MC schema which utilize a neural network process model (Ng & Hussain, 2004; Wei et al., 2007), the hybrid semi-parametric model based approaches were reported to per- form better. For an industrial reactive distillation column Chen et al. (2004) show that the application of a hybrid semi-parametric model based closed-loop IMC schema can reduce the process variability significantly in comparison to an open-loop schema.

3.4. Optimization

Hybrid semi-parametric models have been used to optimize the control policy either to maximize some quantity (Dors et al., 1995; Eslamloueyan & Setoodeh, 2011; Henriques et al., 1999; Ignova et al., 2002; Kahrs & Marquardt, 2007; Mahalec & Sanchez, 2012; Preusting et al., 1996; Psichogios & Ungar, 1992; Schubert et al., 1994a; Teixeira, Alves, et al., 2007; Teixeira et al., 2005, 2006; Tholudur & Ramirez, 1996, 1999; Zuo & Wu, 2000) or to meet specific quality specifications (Doyle et al., 2003; Hermanto et al., 2011; Safavi et al., 1999; Tian et al., 2001; Zhang et al., 2012) (see Table 10 - supplementary material for a complete list). Theoretically, the control policy can be optimized off-line or on-line. Online optimization, which essentially devolves to closed-loop (sub)optimal control, can be expected to achieve better performance than off- line optimization (implemented as open-loop control) (Hermanto et al., 2011), since e.g. process variations can be taken into account. However, on-line optimization is not always feasible due to the lack of reliable at-time available measurements or due to high computational costs of the optimization. In order to flee from these problems, a possible strategy is to perform an off-line optimization of the control inputs followed by an on-line re-optimization whenever new state measurements become available (Ignova et al., 2002; Zuo & Wu, 2000) or to adapt the pre-optimized control inputs at each sampling instance aiming to achieve the quality specifications (Hermanto et al., 2011). Nevertheless, most processes are run in open-loop, subject to optimized control policies, as e.g. in pharmaceutical industry where "approved process recipes" is tightly followed.

Hybrid semi-parametric modeling provides a valuable alternative for process optimization, because operational variables that impact on the product can easily be incorporated into the model (due to the nonparametric model) and yet good model extrapolation properties can be yield (due to the parametric model), which is essential for the identification of optima beyond the conditions covered in the data (Mogk et al., 2002). However, the quality of the model predictions will deteriorate when the non- parametric model is confronted with input constellations which it had not been trained on, wherefore measures have been

pro- posed to constrain the optimization (Kahrs & Marquardt, 2007; Mahalec & Sanchez, 2012; Teixeira et al., 2006), see Section 2.5.3. The optimized predictions can also be assessed with the calculated confidence interval (Kahrs & Marquardt, 2007; Tian et al., 2001). It is obvious that due to these restrictions the optimal solution will usually not be encountered during the first optimization. Hence, batch-to-batch methodologies have been pro- posed for quantity maximization (Teixeira, Alves, et al., 2007; Teixeira et al., 2006) or to meet quality specifications (Doyle et al., 2003; Hermanto et al., 2011; Zhang et al., 2012). It was observed that the predicted performance and the one experimentally obtained converged with increasing number of experiments (Hermanto et al., 2011; Teixeira et al., 2006), where the convergence rate was improved when preoptimized control inputs were on-line adapted (Hermanto et al., 2011). Even though the mechanistic knowledge impact considerably on the optimization results (Teixeira et al., 2005) (via extrapolation) it is important to design the excitation experiments (the data of which are utilized for the nonparametric model identification) in such a manner that the process region of interest is covered, see Section 2.6.1. It would be interesting to study whether it is generally better to explore the design space first and to perform an optimization then or whether iterative batch-to-batch optimization might converge faster to the optimal conditions. Maybe the best strategy is even a mixture of both.

3.5. Model-reduction

Real processes are many times overwhelmingly complex. In order to derive a workable 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 e.g. for control purposes. In this respect hybrid semi-parametric modeling can be applied to correct for the unconsidered or simplified phenomena therefore maintaining a high degree of accuracy, while still being computationally efficient (Chen et al., 2004; Eslamloueyan & Setoodeh, 2011; Madar et al., 2005; Qi et al., 1999; Safavi et al., 1999; Vega et al., 1997). Similarly, after the reduction of a set of equations through e.g. singular perturbation (Chen et al., 2004), residualization (Hahn, Lextrait, & Edgar, 2002) or orthogonal decomposition (Romijn, Ozkan, Weiland, Ludlage, & Marquardt, 2008), it might be desirable to approximate a subset of equations by nonparametric models, whereby a hybrid semi-parametric model is yielded. A solution can be obtained in a computationally efficient way while most of the properties of the original system of equations can be retained.

3.6. 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. Despite this fact, scale-up confronts the model developed on the small scale, with process conditions on larger scale that the model has not been developed on i.e. extrapolation. As outlined above, the model extrapolation properties are determined by the incorporated mechanistic knowledge. Ergo the application of strictly nonparametric models, which have very limited extrapolation properties, is not expedient, a frotiori considering that the data used for their determination on small scale might contain scale specific information. Hybrid semi-parametric models can have good extrapolation properties, while at the same time being less laborious to develop than strictly mechanistic models. Thus, hybrid semi-parametric modeling presents a very efficient approach for scale-up (Braake et al., 1998). The probably most efficient way is to aim from the beginning at the development of a hybrid semi-parametric model, because it can be tented to develop the model in such a way that good inter- and extrapolation properties are yielded with a small number of specifically designed experiments at both scales, e.g. Braake et al. (1998). Another strategy (Bollas et al., 2003; Simon et al., 2006) is to complement a mechanistic model, developed at small-scale, with nonparametric techniques that represent specific parts of the model on the large-scale. In this case, the nonparametric techniques account for the scale-up factors (Bollas et al., 2003; Simon et al., 2006) or other assumptions incorporated into the small scale mechanistic model, which do not hold true on the larger scale (Simon et al., 2006). Similarly, Bellos et al. (2005) show that a mechanistic model developed on the industrial scale while supported by non- parametric models for modeling of specific parts, allows to estimate the effect of the feed quality on the catalyst reactivity and the cat- alyst activity level, using only few laboratory and unit operation data. In fact, the lower requirements on data from both scales is, besides the extrapolation properties, the major advantage of hybrid semi-parametric models over nonlinear nonparametric techniques (Braake et al., 1998). Comparisons of the predictions and extrapolation capabilities of several hybrid semi-parametric models to those of the pilot plant mechanistic model and to a nonparametric model by Bollas et al. (2003) show that the best performances, reaching the limitations of the experimental error, are obtained by the hybrid semi-parametric models.

4. Concluding remarks and outlook

The hybrid semi-parametric modeling framework is reviewed covering the most critical aspects for structure definition, identification and their impact on model performance. Various applications of hybrid semi-parametric modeling in different areas were reviewed such as process monitoring, control, optimization, scale-up and model reduction. From this revision the following main points can be highlighted:

(i) Hybrid semi-parametric modeling found considerable attention during the last 20 years and the advantages of this approach are significant.

(ii) Throughout the applications, hybrid semi-parametric models are compared to nonparametric models or mechanistic models. In almost all cases it was reported that the hybrid semi-parametric models performed better than either of the other. For control or optimization, the better model performance translated usually into improved control and optimization results.

(iii) The interlinking of different knowledge sources into a hybrid semiparametric modeling approach can, although not necessarily, result into better system descriptions, than when compared to models that are based on a single source of knowledge. This means that the application of hybrid semi- parametric approaches does not automatically result into improved models but that a differentiated view has to be kept and an analysis of the reason for eventual models shortcomings must be applied.

(iv) The incorporation of additional mechanistic/phenomenological knowledge was discussed, and it was concluded that the model performance can be enhanced when the incorporated structure is accurate. On the other hand it was stated that in cases of inaccurate model structure the application of parallel approaches is, in general, to prefer. A rigorous comparison of the parallel structure to a serial structure \mathbf{C} is still lacking.

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

(vi) Different identification procedures of the nonparametric models have been reviewed suggesting that the incremental approach together with the sensitivities approach are the best identification methods.

(vii) Measures for extrapolative situations have been discussed and it was concluded that those methods mostly take the range or the dimensional extrapolation into account, while frequency extrapolation (the dynamics) is not considered. However, in cases of control the transient behavior is 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.

It was shown that hybrid semi-parametric 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 objectives.

In future, hybrid semi-parametric models can be expected to consist of several parametric and nonparametric sub-models. Besides the challenges that arise when incorporating e.g. different time scales, especially structuring the models, model discrimination and parameter identification for these complex, large-

scale systems will be challenging. Novel methods will probably seek to overcome those by decomposing the system into sub-systems (similar to the incremental approach Kahrs & Marquardt, 2008) and/or by employing tailored design of experiments. Also semi- quantitative data and qualitative information might be increasingly incorporated for this purpose. The calculation of confidence intervals for each of the sub-models is another challenge.

Perspectively, the application of hybrid semi-parametric models seems promising in several areas. Recently, the value added by hybrid semi-parametric modeling to the PAT initiative was outlined by Gernaey and Gani (2010) and Glassey et al. (2011). Interestingly, the requirements on the "PAT tools" reads as the list of hybrid semiparametric model properties. The pharmaceutical industry could profit immensely from the adaptation of hybrid semi-parametric modeling methodologies at several development stages of pharmaceutical processes (Gernaey, Cervera-Padrell, & Woodley, 2012).

Another emerging area for the application of hybrid semi- parametric models is systems biology see for instances (Carinhas et al., 2011; von Stosch et al., 2010). Hybrid semi-parametric 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. In systems biology middle-out approaches, which seek to combine top-down and bottom-up approaches, are expected to find increasing application in future (Rollie, Mangold, & Sundmacher, 2012). Hybrid semiparametric modeling is especially suited for the development of middle-out approaches, since it can support the bridging of top-down (data-driven) and bottom-up (mechanism- driven) approaches.

Hybrid semi-parametric modeling also appears promising for (bio)medical research, where challenges are multi-scale ranging from the sub-cellular level up to patients response and where the integration of data from several scales along with mechanistic models seems necessary in order to provide affordable modeling solutions that support rational drug development (Schuppert, 2011).

Likewise, in synthetic biology rational design of biological sys- tems could be enabled through hybrid semi-parametric modeling as opposed to current designs, which mostly are based on trial- and-error. In particular, hybrid modular parts could be created that integrate standard mathematical formulations describing biological parts (e.g. Cooling et al., 2010; Rodrigo, Carrera, & Jaramillo, 2007) with nonparametric approaches that model the contained coefficients based on e.g. the DNA sequence. The combination of various hybrid modular parts can subsequently be used to either describe given systems or to design synthetic systems. For the design an iterative schema could be applied, in which the model is used to gradually improve the performance of the system and in which the experimental data are used alongside to improve the model. With this it should be feasible to reduce the experimental load and therefore to enable much more efficient development of synthetic biologic systems.

The integration of hybrid semi-parametric models into complex flowsheets for (bio)chemical processes and the resulting overall representation of the plant, is envisaged as a consequence of the publications by Fiedler and Schuppert (2008), Schweiger, Sayyar- Rodsari, Bartee, and Axelrud (2010). An integrated plantwide modeling approach provides several advantages, such as the possibility to plant-wide optimize the set-points or the opportunity to achieve better closed-loop control performance.

All in all, this paper shows that the application of hybrid semi-parametric models can provide significant advantages in several areas. Hybrid semi-parametric modeling ultimately enables a rational management of multiple knowledge sources and there- with improves decision-making. In order to further close the gap between theory and practice, software tools should be developed that enable flexible integration of different sources of knowledge into hybrid semi-parametric model structures.

Acknowledgment

Sincere thanks for financial support to the Fundação para a Ciência e a Tecnologia (References of the scholarship provided to Moritz von Stosch: SFRH/BD/36990/2007, and of the funded project: POCI/BIO/56571/2004).

Appendix A. Supplementary Data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.compchemeng. 2013.08.008.

References

Abonyi, J., Chovan, T., Nagy, L., & Szeifert, F. (1999). Hybrid convolution model and its application in predictive pH control. *Computers and Chemical Engineering*, *23*, S227–S230.

Abonyi, J., Madar, J., & Szeifert, F. (2007). *Combining first principles models* and neural networks for generic model control.

Agarwal, M. (1997). Combining neural and conventional paradigms for modelling, prediction and control. *International Journal of Systems Science*, *28*, 65–81.

Aguiar, H. C., & 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*, 565–570.

Akkari, E., Chevallier, S., & Boillereaux, L. (2005). A 2d non-linear "greybox" model dedicated to microwave thawing: Theoretical and experimental investigation. *Computers and Chemical Engineering*, *30*, 321–328. Al-Yemni, M. (2003). *Hybrid neural networks models for a membrane reactor*. United States/West Virginia: West Virginia University [Master's Thesis].

Al-Yemni, M., & Yang, R. Y. K. (2005). Hybrid neural-networks modeling of an enzymatic membrane reactor. *Journal of the Chinese Institute of Engineers*, 28, 1061–1067.

Anderson, J. S., McAvoy, T. J., & Hao, O. J. (2000). Use of hybrid models in wastewater systems. *Industrial and Engineering Chemistry Research*, *39*, 1694–1704.

Andrasik, A., Meszaros, A., & de Azevedo, S. F. (2004). On-line tuning of a neural pid controller based on plant hybrid modeling. *Computers and Chemical Engineering*, *28*, 1499–1509.

Arahal, M. R., Cirre, C. M., & Berenguel, M. (2008). Serial grey-box model of a stratified thermal tank for hierarchical control of a solar plant. *Solar Energy*, *82*, 441–451. Bazaei, A., & 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*, 819–830.

Bellos, G., Kallinikos, L., Gounaris, C., & Papayannakos, N. (2005). Modelling of the performance of industrial hds reactors using a hybrid neural network approach. *Chemical Engineering and Processing*, *44*, 505–515.

Beluhan, D., & Beluhan, S. (2000). Hybrid modeling approach to on-line estimation of yeast biomass concentration in industrial bioreactor. *Biotechnology Letters*, *22*, 631–635.

Bhutani, N., Rangaiah, G. P., & Ray, A. K. (2006). First-principles, databased, and hybrid modeling and optimization of an industrial hydrocracking unit. *Industrial and Engineering Chemistry Research*, *45*, 7807–7816.

Bishop, C. (1995). *Neural networks for pattern recognition*. New York: Oxford University Press Inc.

Boareto, A. J. M., De Souza, M. B., Valero, F., & Valdman, B. (2007). A hybrid neural model (hnm) for the on-line monitoring of lipase production by candida rugosa. *Journal of Chemical Technology and Biotechnology*, *82*, 319–327.

Bohlin, T., & Graebe, S. F. (1995). Issues in nonlinear stochastic grey box identification. *International Journal of Adaptive Control and Signal Processing*, *9*, 465–490.

Bollas, G. M., Papadokonstadakis, S., Michalopoulos, J., Arampatzis, G., Lappas, A. A., Vasalos, I. A., & 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*, 697–713.

Braake, H. A. B. t., van Can, H. J. L., & Verbruggen, H. B. (1998). Semimechanistic modeling of chemical processes with neural networks. *Engineering Applications of Artificial Intelligence*, *11*, 507–515.

Brendel, M., & Marquardt, W. (2008). Experimental design for the

identification of hybrid reaction models from transient data. *Chemical Engineering Journal, 141, 264–277.*

Cameron, Ian. T., & Hangos, Katalin. (2001). *Process Modelling and Model Analysis*. Academic Press, Technology & Engineering.

Cao, M., Wang, K. W., Fujii, Y., & Tobler, W. E. (2004). A hybrid neural network approach for the development of friction component dynamic model. *Journal of Dynamic Systems, Measurement, and Control, 126*, 144–153.

Carinhas, N., Bernal, V., Teixeira, A., Carrondo, M., Alves, P., & Oliveira, R. (2011). Hybrid metabolic flux analysis: Combining stoichiometric and statistical con- straints to model the formation of complex recombinant products. *BMC Systems Biology*, *5*, 34.

Chabbi, C., Taibi, M., & Khier, B. (2008). Neural and hybrid neural modeling of a yeast fermentation process. *International Journal of Computational Cognition*, *6*, 42–47. Chang, J.-S., Lu, S.-C., & Chiu, Y.-L. (2007). Dynamic modeling of batch polymerization reactors via the hybrid neural-network rate-function approach. *Chemical Engineering Journal*, *130*, 19–28.

Chen, L., Bernard, O., Bastin, G., & Angelov, P. (2000). Hybrid modelling of biotechnological processes using neural networks. *Control Engineering Practice*, *8*, 821–827. Chen, L., Hontoir, Y., Huang, D., Zhang, J., & Morris, A. J. (2004). Combining first principles with black-box techniques for reaction systems. *Control Engineering Practice*, *12*, 819–826.

Conlin, J., Peel, C., & Montague, G. A. (1997). Modelling pressure drop in water treatment. *Artificial Intelligence in Engineering*, *11*, 393–400.

Cooling, M. T., Rouilly, V., Misirli, G., Lawson, J., Yu, T., Hallinan, J., et al. (2010). Standard virtual biological parts: A repository of modular modeling components for synthetic biology. *Bioinformatics*, *26*, 925–931.

Corazza, F. C., Calsavara, L. P. V., Moraes, F. F., Zanin, G. M., & Neitzel, I. (2005). Determination of inhibition in the enzymatic hydrolysis of cellobiose using hybrid neural modeling. *Brazilian Journal of Chemical Engineering*, *22*, 19–29.

Costa, A. C., Alves, T. L. M., Henriques, A. W. S., Filho, R. M., & Lima, E. L. (1998). An adaptive optimal control scheme based on hybrid neural modelling. *Computers and Chemical Engineering*, *22*, S859–S862.

Costa, A., Henriques, A., Alves, T., Maciel Filho, R., & Lima, E. (1999). A hybrid neural model for the optimization of fed-batch fermentations. *Brazilian Journal of Chemical Engineering*, *16*, 53–63.

Cubillos, F., & Acuna, G. (2007). Adaptive control using a grey box neural model: An experimental application. *Advances in Neural Networks-Lecture Notes in Com- puter Science*, 4491, 311-318-318.

Cubillos, F. A., & Lima, E. L. (1997). Identification and optimizing control of a rougher flotation circuit using an adaptable hybrid-neural model. *Minerals Engineering*, *10*, 707–721.

Cubillos, F. A., & Lima, E. L. (1998). Adaptive hybrid neural models for process control. *Computers and Chemical Engineering*, *22*, S989–S992.

Cubillos, F., Callejas, H., Lima, E., & Vega, M. (2001). Adaptive control using a hybrid-neural model: Application to a polymerisation reactor. *Brazilian Journal of Chemical Engineering*, *18*, 113–120.

Dadhe, K., Rossmann, V., Durmus, K., & Engell, S. (2001). Neural networks as a tool for gray box modelling in reactive distillation. *Computational Intelligence. Theory and Applications*, 2206, 576–588.

Dochain, D. (2003). State and parameter estimation in chemical and biochemical processes: A tutorial. *Journal of Process Control*, *13*, 801–818.

Dors, M., Simutis, R., & 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, volume 613* American Chemical Society, (pp. 144–154).

Dors, M., Simutis, R., & Luebbert, A. (1995). Advanced supervision of mammarian cell cultures using hybrid process models. In A. Munack, & K. Schugerl (Eds.), *Preprints of the 6th international conference on computer applications in biotechnology* (pp. 72–77).

Doyle, F. J., Harrison, C. A., & Crowley, T. J. (2003). Hybrid model-based approach to batch-to-batch control of particle size distribution in emulsion polymerization. *Computers and Chemical Engineering*, *27*, 1153–1163.

Eslamloueyan, R., & 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,* 1309–1338.

Estrada-Flores, S., Merts, I., De Ketelaere, B., & Lammertyn, J. (2006). Development and validation of "grey-box" models for refrigeration applications: A review of key concepts. *International Journal of Refrigeration*, *29*, 931–946.

Feil, B., Abonyi, J., Nemeth, S., Nemeth, O., Arva, P., Nemeth, M., et al. (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, volume 3070, Springer* (pp. 1111–1117).

Fellner, Delgado, & Becker. (2003). Functional nodes in dynamic neural networks for bioprocess modelling. *Bioprocess and Biosystems Engineering*, *25*, 263–270.

Fiedler, B., & Schuppert, A. (2008). Local identification of scalar hybrid models with tree structure. *IMA Journal of Applied Mathematics*, *73*, 449–476. Fu, P. C., & Barford, J. P. (1995a). A hybrid neural network-first principles approach for modelling of cell metabolism. *Computers and Chemical Engineering*, *20*, 951–958.

Fu, P. C., & Barford, J. P. (1995b). Integration of mathematical modelling and knowledge-based systems for simulations of biochemical processes. *Expert Sys- tems with Applications*, *9*, 295–307.

Galvanauskas, V., Simutis, R., & Luebbert, A. (2004). Hybrid process models for process optimisation, monitoring and control. *Bioprocess and Biosystems Engineering*, *26*, 393–400.

Georgieva, P., & de Azevedo, S. (2009). *Computational intelligence techniques for bioprocess modelling, supervision and control, Volume 218*. Berlin/Heidelberg: Springer.

Georgieva, P., & 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., & Feyo de Azevedo, S. (2003). Knowledgebased hybrid modelling of a batch crystallisation when accounting for nucleation, growth and agglomeration phenomena. *Chemical Engineering Science*, *58*, 3699–3713.

Gernaey, K. V., & Gani, R. (2010). A model-based systems approach to pharmaceutical product-process design and analysis. *Chemical Engineering Science*, *65*, 5757–5769.

Gernaey, K. V., Cervera-Padrell, A. E., & Woodley, J. M. (2012). A perspective on pse in pharmaceutical process development and innovation. *Computers and Chemical Engineering*, *42*, 15–29.

Glassey, J., Gernaey, K. V., Clemens, C., Schulz, T. W., Oliveira, R., Striedner, G., et al. (2011). Process analytical technology (pat) for biopharmaceuticals. *Biotechnology Journal*, *6*, 369–377.

Gnoth, S., Jenzsch, M., Simutis, R., & Luebbert, A. (2008). Product formation kinetics in genetically modified *E. coli* bacteria: Inclusion body formation. *Bioprocess and Biosystems Engineering*, *31*, 41–46.

Graefe, J., Bogaerts, P., Castillo, J., Cherlet, M., Werenne, J., Marenbach, P., et al. (1999). A new training method for hybrid models of bioprocesses. *Bioprocess and Biosystems Engineering*, *21*, 423–429.

Gupta, S., Liu, P.-H., Svoronos, S. A., Sharma, R., Abdel-Khalek, N. A., Cheng, Y., et al. (1999). Hybrid first-principles/neural networks model for column flotation. *AIChE Journal*, *45*, 557–566.

Haber, Robert, & Keviczky, L. (1999). Nonlinear System Identification — Input-Output Modeling Approach Volume 1: Nonlinear System Parameter Identification. Mathematical Modelling: Theory and Applications, (7). XXXI, 802 pp.

Haerdle, Wolfgang, Mueller, Marlene, Sperlich, Stefan, & Werwatz, Axel. (2004). *Nonparametric and semiparametric models*. Springer., 299 pp.

Hahn, J., Lextrait, S., & Edgar, T. F. (2002). Nonlinear balanced model residualization via neural networks. *AIChE Journal*, *48*, 1353–1357.

Henneke, D., Hagedorn, A., Budman, H., & Legge, R. (2005). Application of spectrofluorometry to the prediction of phb concentrations in a fed-batch process. *Bioprocess and Biosystems Engineering*, *27*, 359–364.

Henriques, A. W. S., Costa, A. C., Alves, T. L. M., & Lima, E. L. (1999). Optimization of fed-batch processes: Challenges and solutions. *Brazilian Journal of Chemical Engineering*, *16*, 171–177.

Hermanto, M. W., Braatz, R. D., & Chiu, M.-S. (2011). Integrated batch-tobatch and nonlinear model predictive control for polymorphic transformation in pharmaceutical crystallization. *AIChE Journal*, *57*, 1008– 1019.

Hinchliffe, M., Montague, G., Willis, M., & Burke, A. (2003). Hybrid approach to modeling an industrial polyethylene process. *AIChE Journal*, *49*, 3127–3137.

Hu, G., Mao, Z., He, D., & Yang, F. (2011). Hybrid modeling for the prediction of leaching rate in leaching process based on negative correlation learning bagging ensemble algorithm. *Computers and Chemical Engineering*.

Hussain, M. A., & Ho, P. Y. (2004). Adaptive sliding mode control with neural network based hybrid models. *Journal of Process Control*, *14*, 157–176.

Hussain, M. A., Ho, P. Y., & Allwright, J. C. (2001). Adaptive linearizing control with neural-network-based hybrid models. *Industrial and Engineering Chemistry Research*, *40*, 5604–5620.

Hwang, T.-M., Oh, H., Choi, Y.-J., Nam, S.-H., Lee, S., & Choung, Y.-K. (2009). Devel- opment of a statistical and mathematical hybrid model to predict membrane fouling and performance. *Desalination*, *247*, 210–221.

Ibrehem, A. S., Hussain, M. A., & 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*, 25–44.

Ignova, M., Paul, G. C., Kent, C. A., Thomas, C. R., Montague, G. A., Glassey, J., et al. (2002). Hybrid modelling for on-line penicillin fermentation optimisation. In *Proceedings of the 15th IFAC world congress*.

Ingram, G., Cameron, I., & Hangos, K. (2004). Classification and analysis of integrating frameworks in multiscale modelling. *Chemical Engineering Science*, *59*, 2171–2187.

James, S., Legge, R., & Budman, H. (2002). Comparative study of black-box and hybrid estimation methods in fed-batch fermentation. *Journal of Process Control*, *12*, 113–121.

Jenzsch, M., Gnoth, S., Kleinschmidt, M., Simutis, R., & 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*, 858–867.

Jia, R.-d., Mao, Z.-z., Chang, Y.-q., & Zhao, L.-p. (2011). Soft-sensor for copper extraction process in cobalt hydrometallurgy based on adaptive

hybrid model. Chemical Engineering Research and Design, 89, 722-728.

Johansen, T. A., & Foss, B. A. (1992). Representing and learning unmodeled dynamics with neural network memories. In *American Control Conference* (pp. 3037–3043). Johansen, T., & Foss, B. (1992). 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

(pp. 677-682).

Jorgensen, S. B., & Hangos, K. M. (1995). Grey box modelling for control: Qualitative models as a unifying framework. *International Journal of Adaptive Control and Signal Processing*, 9, 547–562.

Kahrs, O., & Marquardt, W. (2007). The validity domain of hybrid models and its application in process optimization. *Chemical Engineering and Processing: Process Intensification*, *46*, 1054–1066.

Kahrs, O., & Marquardt, W. (2008). Incremental identification of hybrid process models. *Computers and Chemical Engineering*, *32*, 694–705.

Karama, A., Bernard, O., & Gouz, J.-L. (2010). Constrained hybrid neural modelling of biotechnological processes. *International Journal of Chemical Reactor Engineering*, *8*.

Kasprow, R. K. (2000). *Hybrid modeling (neural networks and first principles) of fermentation: Combining biochemical engineering fundamentals and process data*. USA: University of Virginia, VA [Ph.D. Thesis].

Kim, H., & 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*, 696–703.

Klimasauskas, C. C. (1998). Hybrid modeling for robust nonlinear multivariable con- trol. *ISA Transactions*, *37*, 291–297.

Kramer, M. A., Thompson, M. L., & Bhagat, P. M. (1992). Embedding theoretical models in neural networks. In *American Control Conference* (pp. 475–479).

Kumar Akkisetty, P., Lee, U., Reklaitis, G., & Venkatasubramanian, V. (2010). Population balance model-based hybrid neural network for a pharmaceutical milling process. *Journal of Pharmaceutical Innovation*, *5*, 161–168.

Lauret, P., Boyer, H., & Gatina, J. C. (2000). Hybrid modelling of a sugar boiling process. *Control Engineering Practice*, *8*, 299–310.

Laursen, S. O., Webb, D., & Ramirez, W. F. (2007). Dynamic hybrid neural network model of an industrial fed-batch fermentation process to produce foreign protein. *Computers and Chemical Engineering*, *31*, 163–170.

Lee, D. S., Jeon, C. O., Park, J. M., & Chang, K. S. (2002). Hybrid neural network modeling of a full-scale industrial wastewater treatment process. *Biotechnology and Bioengineering*, *78*, 670–682.

Lee, D. S., Vanrolleghem, P. A., & Park, J. M. (2005). Parallel hybrid modeling methods for a full-scale cokes wastewater treatment plant. *Journal of*

Biotechnology, 115, 317-328.

Luo, N., Du, W., Ye, Z., & Qian, F. (2012). Development of a hybrid model for industrial ethylene oxide reactor. *Industrial and Engineering Chemistry Research*, *51*, 6926–6932.

Madar, J., Abonyi, J., & Szeifert, F. (2004). New approaches to the identification of semi-mechanistic process models. *Acta Agraria Kaposvariensis*, *8*, 1–9.

Madar, J., Abonyi, J., & Szeifert, F. (2005). Feedback linearizing control using hybrid neural networks identified by sensitivity approach. *Engineering Applications of Artificial Intelligence*, *18*, 343–351.

Mahalec, V., & Sanchez, Y. (2012). Inferential monitoring and optimization of crude separation units via hybrid models. *Computers and Chemical Engineering*, 45, 15–26.

Masri, S. F. (1994). A hybrid parametric/nonparametric approach for the identification of nonlinear systems. *Probabilistic Engineering Mechanics*, *9*, 47–57.

Mazutti, M. A., Corazza, M. L., Maugeri, F., Rodrigues, M. I., Oliveira, J. V., Treichel, H., et al. (2010). Hybrid modeling of inulinase bio-production process. *Journal of Chemical Technology and Biotechnology*, *85*, 512–519.

McKay, B., Sanderson, C. S., Willis, M. J., Barford, J. P., & Barton, G. W. (1998). Evolving a hybrid model of a fed-batch fermentation process. *Transactions of the Institute of Measurement and Control*, *20*, 4–10.

Mogk, G., Mrziglod, T., & Schuppert, A. (2002). Application of hybrid model in chemical industry. In J. Grievink, & J. van Schijndel (Eds.), *European symposium on computer aided process engineering-12, 35th European symposium of the working party on computer aided process engineering, Volume 10* (pp. 931–936). Elsevier.

Molga, E., & Cherbanski, R. (1999). Hybrid first-principle-neural-network approach to modelling of the liquid-liquid reacting system. *Chemical Engineering Science*, *54*, 2467–2473.

Narendra, K. S., & Parthasarathy, K. (1990). Identification and control of dynamical systems using neural networks. *IEEE Transactions on Neural Networks*, *1*, 1–27.

Nascimento, C. A. O., Giudici, R., & Scherbakoff, N. (1999). Modeling of industrial nylon-6,6 polymerization process in a twin-screw extruder reactor. ii. neural networks and hybrid models. *Journal of Applied Polymer Science*, *72*, 905–912.

Ng, C. W., & 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*, 559–570.

Oliveira, R. (2004). Combining first principles modelling and artificial neural net- works: A general framework. *Computers and Chemical*

Engineering, *28*, 755–766. Patnaik. (2001). Hybrid neural simulation of a fedbatch bioreactor for a nonideal recombinant fermentation. *Bioprocess and Biosystems Engineering*, *24*, 151–161.

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*, 165–175.

Patnaik, P. R. (2004). Neural and hybrid neural modeling and control of fedbatch fermentation for streptokinase: Comparative evaluation under nonideal conditions. *Canadian Journal of Chemical Engineering*, *82*, 599–606.

Patnaik, P. R. (2008). Neural and hybrid optimizations of the fed-batch synthesis of poly-hydroxybutyrate by ralstonia eutropha in a nonideal bioreactor. *Bioremediation Journal*, *12*, 117–130.

Patnaik, P. (2010). Design considerations in hybrid neural optimization of fed-batch fermentation for phb production by ralstonia eutropha. *Food and Bioprocess Technology*, *3*, 213–225.

Pearson, R. K., & Pottmann, M. (2000). Gray-box identification of blockoriented nonlinear models. *Journal of Process Control*, 10, 301–315.

Peres, J., Oliveira, R., & Feyo de Azevedo, S. (2001). Knowledge based modular net- works for process modelling and control. *Computers and Chemical Engineering*, *25*, 783–791.

Peres, J., Oliveira, R., & de Azevedo, S. F. (2008). Bioprocess hybrid parametric/nonparametric modelling based on the concept of mixture of experts. *Biochemical Engineering Journal*, *39*, 190–206.

Piron, E., Latrille, E., & Rene, F. (1997). Application of artificial neural networks for crossflow microfiltration modelling: "Black-box" and semi-physical approaches. *Computers and Chemical Engineering*, *21*, 1021–1030.

Porru, G., Aragonese, C., Baratti, R., & Servida, A. (2000). Monitoring of a co oxidation reactor through a grey model-based ekf observer. *Chemical Engineering Science*, *55*, 331–338.

Potocnik, P., & Grabec, I. (1999). Empirical modeling of antibiotic fermentation pro- cess using neural networks and genetic algorithms. *Mathematics and Computers in Simulation*, *49*, 363–379.

Preusting, H., Noordover, J., Simutis, R., & Luebbert, A. (1996). The use of hybrid modelling for the optimization of the penicillin fermentation process. *CHIMIA*, *50*, 416–417.

Psichogios, D. C., & Ungar, L. H. (1992). A hybrid neural network-first principles approach to process modeling. *AIChE Journal*, *38*, 1499–1511.

Qi, H., Zhou, X.-G., Liu, L.-H., & Yuan, W.-K. (1999). A hybrid neural network- first principles model for fixed-bed reactor. *Chemical Engineering Science*, *54*, 2521–2526.

Quiza, Ram>n, Lopez-Armas, Omar, & Davim, J. Paulo. (2012). *Hybrid* modeling and optimization of manufacturing. Combining Artificial Intelligence

and Finite Element Method (Vol. VIII) Springer., 95 pp.

Reuter, M., Van Deventer, J., & Van Der Walt, T. (1993). A generalized neural-net kinetic rate equation. *Chemical Engineering Science*, 48, 1281–1297.

Ricardo, A. R., Oliveira, R., Velizarov, S., Reis, M. A., & Crespo, J. G. (2012). Hybrid modeling of counterion mass transfer in a membrane-supported biofilm reactor. *Biochemical Engineering Journal*, *62*, 22–33.

Rodrigo, G., Carrera, J., & Jaramillo, A. (2007). Asmparts: Assembly of biological model parts. *Systems and Synthetic Biology*, *1*, 167–170.

Rollie, S., Mangold, M., & Sundmacher, K. (2012). Designing biological systems: Systems engineering meets synthetic biology. *Chemical Engineering Science*, *69*, 1–29.

Romijn, R., Ozkan, L., Weiland, S., Ludlage, J., & Marquardt, W. (2008). A greybox modeling approach for the reduction of nonlinear systems. *Journal of Process Control*, *18*, 906–914.

Roubos, H. (2002). *Bioprocess modelling and optimization fed-batch clavulanic acid production by streptomyces clavuligerus*. Technical University Delft [Ph.D. thesis]. Roubos, J., Krabben, P., Setnes, M., Babuska, R., Heijnen, J., & Verbruggen, H. (2000). Hybrid model development for fed-batch bioprocesses; combining physical equations with the metabolic network and black-box kinetics. *Journal A – Benelux Quarterly Journal on Automatic Control*, *41*, 12–23.

Safavi, A. A., Nooraii, A., & Romagnoli, J. A. (1999). A hybrid model formulation for a distillation column and the on-line optimisation study. *Journal of Process Control*, *9*, 125–134.

Saraceno, A., Curcio, S., Calabro, V., & Iorio, G. (2010). A hybrid neural approach to model batch fermentation of "ricotta cheese whey" to ethanol. *Computers and Chemical Engineering*, *34*, 1590–1596.

Saxen, B., & Saxen, H. (1996). A neural-network based model of bioreaction kinetics. *Canadian Journal of Chemical Engineering*, *74*, 124–131.

Schenker, B., & Agarwal, M. (2000). Online-optimized feed switching in semi-batch reactors using semi-empirical dynamic models. *Control Engineering Practice*, *8*, 1393–1403.

Schubert, J., Simutis, R., Dors, M., Havlik, I., & Luebbert, A. (1994a). Bioprocess opti- mization and control: Application of hybrid modelling. *Journal of Biotechnology*, *35*, 51–68.

Schubert, J., Simutis, R., Dors, M., Havlik, I., & Luebbert, A. (1994b). Hybrid modelling of yeast production processes – Combination of a priori knowledge on different levels of sophistication. *Chemical Engineering & Technology*, *17*, 10–20.

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)*.

Schuppert, A. (2011). Efficient reengineering of meso-scale topologies for functional networks in biomedical applications. *Journal of Mathematics in Industry*, *1*, 6.

Schweiger, C., Sayyar-Rodsari, B., Bartee, J., & Axelrud, C. (2010). Plant-wide optimization of an ethanol plant using parametric hybrid models. In *49th IEEE Conference on Decision and Control*.

Silva, R., Cruz, A., Hokka, C., Giordano, R., & Giordano, R. (2000). A hybrid feed forward 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., & Giordano, R. (2001). A hybrid neuroal network algorithm for on-line state inference that accounts for differences in inoculum of cephalosporium acremonium in fed-batch fermentors. *Appl. Biochem. Biotech.*, *91-93*, 341–352.

Simon, L. L., Fischer, U., & Hungerbuehler, K. (2006). Modeling of a threephase industrial batch reactor using a hybrid first-principles neural-network model. *Industrial and Engineering Chemistry Research*, *45*, 7336–7343.

Simutis, R., & Luebbert, A. (1997). Exploratory analysis of bioprocesses using artificial neural network-based methods. *Biotechnol Progress, 13,* 479–487.

Simutis, R., Havlik, I., Schneider, F., Dors, M., & 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., Oliveira, R., Manikowski, M., de Azevedo, S. F., & Luebbert, A. (1997). How to increase the performance of models for process optimization and control. *Journal of Biotechnology*, *59*, 73–89.

Sohlberg, B. (2005). Hybrid grey box modelling of a pickling process. *Control Engineering Practice*, *13*, 1093–1102.

Su, H. T., & McAvoy, T. J. (1993). Integration of multilayer perceptron networks and linear dynamic models: A hammerstein modeling approach. *Industrial and Engineering Chemistry Research*, *32*, 1927–1936.

Su, H. T., Bhat, N., Minderman, A., & McAvoy, T. J. (1992). Integrating neural net- works with first principles models for dynamic modeling. In *IFAC symposium on dynamics and control of chemical reactors distillation columns and batch processes* IFAC, Maryland.

Takagi, T., & Sugeno, M. (1985). Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man, and Cybernetics*, *15*, 116–132.

Teissier, P., Perret, B., Latrille, E., Barillere, J. M., & Corrieu, G. (1997). A hybrid recur- rent neural network model for yeast production monitoring and control in a wine base medium. *Journal of Biotechnology*, *55*, 157–169.

Teixeira, A., Cunha, A. E., Clemente, J. J., Moreira, J. L., Cruz, H. J., Alves, P. M., et al. (2005). Modelling and optimization of a recombinant bhk-21 cultivation

process using hybrid grey-box systems. *Journal of Biotechnology*, *118*, 290–303.

Teixeira, A. P., Clemente, J. J., Cunha, A. E., Carrondo, M. J. T., & Oliveira, R. (2006). Bioprocess iterative batch-to-batch optimization based on hybrid paramet-ric/nonparametric models. *Biotechnology Progress, 22*, 247–258.

Teixeira, A. P., Carinhas, N., Dias, J. M., Cruz, P., Alves, P. M., Carrondo, M. J., et al. (2007). Hybrid semi-parametric mathematical systems: Bridging the gap between systems biology and process engineering. *Journal of Biotechnology*, *132*, 418–425.

Teixeira, A. P., Alves, C., Alves, P. M., Carrondo, M. J. T., & Oliveira, R. (2007). Hybrid elementary flux analysis/nonparametric modeling: Application for bioprocess control. *Bmc Bioinformatics*, *8*.

Thibault, J., Acuna, G., Perez-Correa, R., Jorquera, H., Molin, P., & Agosin, E. (2000). A hybrid representation approach for modelling complex dynamic bioprocesses. *Bioprocess and Biosystems Engineering*, *22*, 547–556.

Tholudur, A., & Ramirez, W. F. (1996). Optimization of fed-batch bioreactors using neural network parameter function models. *Biotechnology Progress*, *12*, 302–309.

Tholudur, A., & Ramirez, W. F. (1999). Neural-network modeling and optimization of induced foreign protein production. *AIChE Journal*, *45*, 1660–1670.

Tholudur, A., Ramirez, W., & McMillan, J. D. (2000). Interpolated parameter functions for neural network models. *Computers and Chemical Engineering*, *24*, 2545–2553.

Thompson, M. L., & Kramer, M. A. (1994). Modeling chemical processes using prior knowledge and neural networks. *AIChE Journal*, *40*, 1328–1340.

Tian, Y., Zhang, J., & Morris, J. (2001). Modeling and optimal control of a batch polymerization reactor using a hybrid stacked recurrent neural network model. *Industrial and Engineering Chemistry Research*, *40*, 4525–4535.

Tsen, A. Y.-D., Jang, S. S., Wong, D. S. H., & Joseph, B. (1996). Predictive control of quality in batch polymerization using hybrid ann models. *AIChE Journal*, *42*, 455–465.

Tulleken, H. J. (1993). Grey-box modelling and identification using physical knowl- edge and Bayesian techniques. *Automatica*, *29*, 285–308.

van Can, H. J. L., Hellinga, C., Luyben, K. C. A. M., Heijnen, J. J., & Te Braake, H. A.

B. (1996). Strategy for dynamic process modeling based on neural networks in macroscopic balances. *AIChE Journal*, *42*, 3403–3418.

van Can, H. J. L., te Braake, H. A. B., Hellinga, C., & Luyben, K. C. A. M. (1997). An effi- cient model development strategy for bioprocesses based on neural networks in macroscopic balances. *Biotechnology and Bioengineering*, *54*, 549–566. van Can, H. J. L., Te Braake, H. A. B., Dubbelman, S., Hellinga, C., Luyben, K. C. A. M., & Heijnen, J. J. (1998). Understanding and applying the extrapolation properties of serial gray-box models. *AIChE Journal*, *44*, 1071–1089.

van Can, H. J. L., te Braake, H. A. B., Bijman, A., Hellinga, C., Luyben, K. C. A. M., & Heijnen, J. J. (1999). An efficient model development strategy for bioprocesses based on neural networks in macroscopic balances: Part II. *Biotechnology and Bioengineering*, *62*, 666–680.

van Lith, P. F., Betlem, B. H. L., & Roffel, B. (2002). A structured modeling approach for dynamic hybrid fuzzy-first principles models. *Journal of Process Control*, *12*, 605–615.

van Lith, P. F., Betlem, B. H. L., & Roffel, B. (2003). Combining prior knowledge with data driven modeling of a batch distillation column including start-up. *Computers and Chemical Engineering*, *27*, 1021–1030.

Vande Wouwer, A., Renotte, C., & Bogaerts, P. (2004). Biological reaction modeling using radial basis function networks. *Computers and Chemical Engineering*, *28*, 2157–2164.

Vega, M., Lima, E., & Pinto, J. (1997). Modeling and control of tubular solution polymerization reactors. *Computers & Chemical Engineering*, *21*(Supplement), S1049–S1054.

Vega, M., Lima, E., & Pinto, J. (2000). Control of a loop polymerization reactor using neural networks. *Brazilian Journal of Chemical Engineering*, *17*, 471–482.

Vieira, J., Dias, F., & 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*.

von Stosch, M., Peres, J., de Azevedo, S., & Oliveira, R. (2010). Modelling biochemical networks with intrinsic time delays: A hybrid semi-parametric approach. *BMC Systems Biology*, *4*, 131.

von Stosch, M., Oliveira, R., Peres, J. S., & Feyo de Azevedo, S. (2011). A novel identification method for hybrid (n)pls dynamical systems with application to bioprocesses. *Expert Systems with Applications, 38*, 10862–10874.

von Stosch, M., Oliveira, R., Peres, J., & Feyo de Azevedo, S. (2011). A hybrid mod- eling framework for pat: Application to bordetella pertussis cultures. *Journal of Biotechnology Progress, 28,* 284–291.

von Stosch, M., Oliveira, R., Peres, J., & Feyo de Azevedo, S. (2012). A general hybrid semi-parametric process control framework. *Journal of Process Control*, *22*, 1171–1181.

Walter, E., Pronzato, L., & Norton, J. (1997). *Identification of parametric models: From experimental data*. Paris: Springer, Original French edition published by MAS-SON., 1994.

Wang, X., Chen, J., Liu, C., & Pan, F. (2010). Hybrid modeling of penicillin

fermentation process based on least square support vector machine. *Chemical Engineering Research and Design*, *88*, 415–420.

Wei, N. C., Hussain, M. A., & Wahab, A. K. A. (2007). Control of a batch polymerization system using hybrid neural network—First principle model. *Canadian Journal of Chemical Engineering*, *85*, 936–945.

Werbos, P. (1974). *Beyond regression new tools for prediction and analysis in behavioral sciences*. Harvard University (Ph.D. thesis).

Wilson, J. A., & Zorzetto, L. F. M. (1997). A generalised approach to process state esti- mation using hybrid artificial neural network/mechanistic models. *Computers and Chemical Engineering*, *21*, 951–963.

Worden, K., Wong, C., Parlitz, U., Hornstein, A., Engster, D., Tjahjowidodo, T., et al. (2007). Identification of pre-sliding and sliding friction dynamics: Grey box and black-box models. *Mechanical Systems and Signal Processing*, *21*, 514–534.

Xiong, Q., & Jutan, A. (2002). Grey-box modelling and control of chemical processes. *Chemical Engineering Science*, *57*, 1027–1039.

Yang, A., Martin, E., & Morris, J. (2011). Identification of semi-parametric hybrid process models. *Computers and Chemical Engineering*, *35*, 63–70.

Zahedi, G., Lohi, A., & Mahdi, K. (2011). Hybrid modeling of ethylene to ethylene oxide heterogeneous reactor. *Fuel Processing Technology*, *92*, 1725–1732.

Zhang, L., Pan, M., Quan, S., Chen, Q., & 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*

(pp. 8319-8323).

Zhang, S., Wang, F., He, D., & Jia, R. (2012). Batch-to-batch control of particle size distribution in cobalt oxalate synthesis process based on hybrid model. *Powder Technology*, *224*, 253–259.

Zuo, K., & Wu, W. (2000). Semi-realtime optimization and control of a fedbatch fermentation system. *Computers and Chemical Engineering*, *24*, 1105– 1109.

Zuo, K., Cheng, H.-P., Wu, S.-C., & Wu, W.-T. (2006). A hybrid model combining hydro- dynamic and biological effects for production of bacterial cellulose with a pilot scale airlift reactor. *Biochemical Engineering Journal*, *29*, 81–90.



Fig. 1. Parametric, nonparametric and hybrid semi-parametric modeling and the types of knowledge they are based on.



Fig. 2. Schematic sketch of the three ways to combine two models (represented by a white and a black box). **A** shows a parallel configuration. **B** and **C** present serial structures.



Fig. 3. Schematic sketches of the model structure for one-step and multistep ahead predictors.



Fig. 4. Schematic representation of white box and black box model combination schema. (a) Superposition; (b) multiplication; (c) weighting of the predictions of the same quantity by using either a black or a white box model; (d) weighting of the white box model predictions using a black box model; (e) weighting of the black box model predictions using a white box model.



Fig. 5. Schematic sketches for dimensional extrapolation, range extrapolation, interpolation and frequency extrapolation, adapted from van Can et al. (1998).



Fig. 6. Number of publications on hybrid semi-parametric modeling over the area of applications and with respect to the type of data used from 1992 to 2012.



Fig. 7. Diagram of two possibilities to use hybrid semi-parametric modeling for monitoring.