Computers and Chemical Engineering 151 (2021) 107365

Contents lists available at ScienceDirect

Computers and Chemical Engineering

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

Review

Recent trends on hybrid modeling for Industry 4.0

Joel Sansana[a], Mark N. Joswiak[b], Ivan Castillo[b], Zhenyu Wang[b], Ricardo Rendall[c], Leo H. Chiang[b], Marco S. Reis[a,*]

[a] Univ Coimbra, CIEPQPF, Department of Chemical Engineering, Rua Sílvio Lima, Pólo II - Pinhal de Marrocos, Coimbra 3030-790, Portugal
[b] Continuous Improvement Center of Excellence, Dow Inc., Lake Jackson, TX, United States of America
[c] Continuous Improvement Center of Excellence, Dow Inc., Herbert H. Dowweg 5, NM Hoek 4542, the Netherlands

ARTICLE INFO

Article history:
Received 10 September 2020
Revised 1 March 2021
Accepted 7 May 2021
Available online 11 May 2021

Keywords:
Hybrid modeling
Gray-box modeling
Semi-parametric modeling
Metamodeling
Physics-informed machine learning
Industrial process data analytics

ABSTRACT

The chemical processing industry has relied on modeling techniques for process monitoring, control, diagnosis, optimization, and design, especially since the third industrial revolution and the emergence of Process Systems Engineering. The fourth industrial revolution, connected to massive digitization, made it possible to collect and process large volumes of data triggering the development of data-driven frameworks for knowledge extraction. However, one must not leave behind the successful solutions developed over decades based on first principle mechanistic modeling approaches. At present, both industry and researchers are realizing the need for new ways to incorporate process and phenomenological knowledge in big data and machine learning frameworks, leading to more robust and intelligible artificial intelligence solutions, capable of assisting the target stakeholders in their activities and decision processes. In this article, we review hybrid modeling techniques, associated system identification methodologies and model assessment criteria. Applications in chemical and biochemical processes are also referred.

Contents

* Corresponding author.
E-mail address: marco@eq.uc.pt (M.S. Reis).
URL: https://www.eq.uc.pt/marco/ (M.S. Reis)

https://doi.org/10.1016/j.compchemeng.2021.107365
0098-1354/

1. Introduction

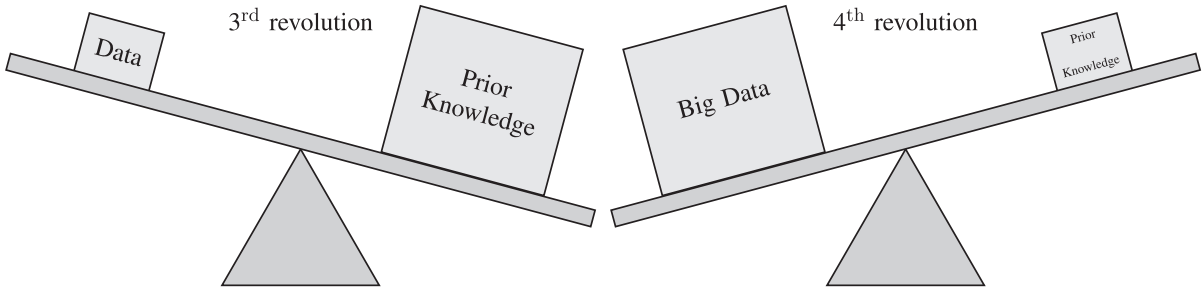
Ever since the publication of the celebrated Bird, Stewart and Lightfoot's book, *Transport Phenomena* (Bird et al., 1960), mechanistic process modeling has been a major part of Chemical Engineering and a central activity of Process Systems Engineering (PSE) - a discipline that capitalized on such rigorously formulated Chemical Engineering Science, together with the existence of analytical methods to solve the formulated problems (numerical methods and mathematical programming) and the existence of technological resources (computers) to run them. Mechanistic models, also known as first-principles, white-box or phenomenological models, have since then been extensively used for process monitoring, control, diagnosis, optimization and design. They require detailed knowledge about the system, including relevant chemical (kinetics) and transport (heat, mass, momentum) phenomena, materials, composition, boundary conditions, thermodynamics, properties, etc. They are therefore knowledge intensive, even though a gradient of modeling rigor is available to the engineer, who may consider a coarser approximation of the system behavior using lumped parameter modeling approaches based on macroscopic balances over the extensive quantities, or finer descriptions of it, using distributed parameter modeling based on microscopic balances. The former approach requires the specification of fewer parameters and usually involves the solution of sets of ordinary-algebraic equations. The later approach implies considerably more parameters and gives rise to partial differential equations (PDEs). This approach is also deductive by nature, as the model formulation is obtained by the application of a few fundamental and general principles (laws) of nature, which are then applied to the particular case under analysis to deduce the specific model for predicting the behaviour for the set of intensive variables.

In general terms, the lumped parameter description is simpler from the implementation and modeling complexity perspectives, whereas the distributed description is more complex from both perspectives. From a statistical estimation point of view, it can be said that the first approach tend to present more bias (due the more extensive simplifications and assumptions made) and less variance (due to the existence of fewer parameters, easier to specify, to know or to estimate), while the second present less bias (the description is more accurate in principle) but higher variance (there are many more parameters whose values are not known with enough rigor and their uncertainty propagates to the model predictions, or that need to be estimated from data, constituting modeling degrees of freedom that add extra-flexibility and therefore more variance to the estimates, see Fig. 1a).

On the other hand, data-driven, black-box, statistical or empirical models, do not require extensive *a priori* knowledge about the system, but strongly rely on the existence of data collected from the process, in sufficient quantity and with the necessary information content and quality, in order to be properly estimated, as shown in Fig. 1b. They are inductive approaches, as they infer patterns and knowledge from data, with minimum assumptions (if any) about the nature of the models. These models first entered the chemical engineering mainstream during the 1980s, with some of the initial contributions arising from the field of chemometrics (Geladi, 1988). Nevertheless, the use of statistical methods for process monitoring and fault detection can be traced back to the 1920s (Shewhart, 1931). Also here, the bias-variance trade-off is a guiding principle for defining the complexity of the empirical model (Hastie et al., 2016). For instance, in partial least squares (PLS) modeling, the optimal model complexity (number of latent variables retained) is usually decided upon the analysis of the cross-validation errors, as a measure of the optimal parsimony level representing the best bias-variance trade-off for predictive purposes (Martens and Næs, 1992).

When addressing real world problems, all sources of information available (*a priori* knowledge and historical data) should be brought together to support the analysis. However, combining data-driven and model-based methods is far from being a trivial task. The potential benefits of integrating both the deductive and inductive sources of knowledge has motivated the emergence of hybrid modeling during the 1990s (Psichogios and Ungar, 1992). The initial contributions consisted on coupling mechanistic models with black-box artificial neural networks (ANNs) to yield more reliable and intelligible models. From then on and through the early 2000s, a large number of approaches were developed and papers published on hybrid modeling applied to chemical engineering problems (von Stosch et al., 2014).

The interest on hybrid modeling reemerged in recent years due to the new Big Data environment, the development of machine learning techniques and the advent of Industry 4.0 (Glassey and von Stosch, 2018; Asprion, 2020). In the context of this work, Big Data refers to a disruptive scenario where the easy access to massive amounts of data, advanced analytical tools and computational resources, provide new solutions to old problems and open new opportunities for improvement not addressed in the past (Reis et al., 2016; Qin, 2014). This new scenario, co-occurrent with the emergence of Industry 4.0, led some authors to take their expectations even further, claiming that Big Data advances may replace theory, although innovation often implies extrapolation and out-of-the-box ideas, that are typically not in the data (Anderson, 2008). Industry 4.0, also known as the fourth indus-



(a) Modeling approach in the classical paradigm.

(b) Modeling approach in the Big Data paradigm.

Fig. 1. Representation of the weight put on types of information in different modeling paradigms.

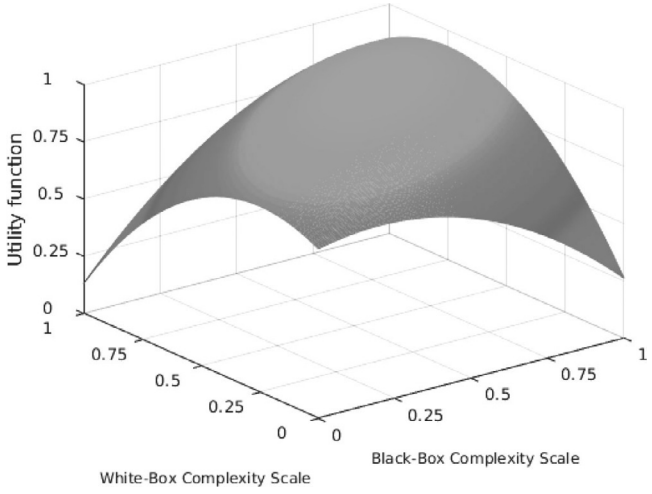


Fig. 2. Scheme representing the bidimensional complexity plane of hybrid modeling (established by the complexity of the white-box and black-box sub-models) and the utility function (representing the usefulness of a model given an application). The authors believe the utility function can be maximized by finding the right balance between prior knowledge (white-box) and knowledge in data (black-box).

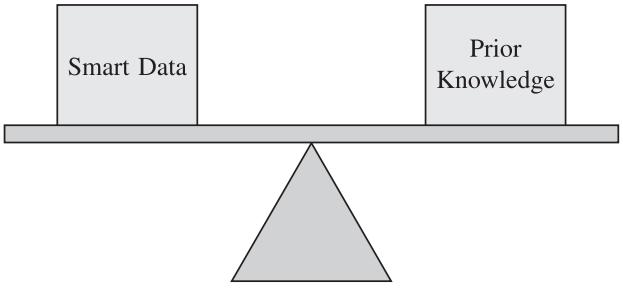


Fig. 3. Modeling approach aimed by hybrid modeling in recent years.

trial revolution, is a paradigm shift in manufacturing where the internet of things combined with generalized advanced digitalization, business analytics and additive manufacturing will drive industry to a market more decentralized, flexible and customizable, focused on client-by-client solutions and with reduced time-to-market due to rapid disruptive innovation (Lasi et al., 2014; Reis and Gins, 2017; Reis and Saraiva, 2019). With the exponential increase of the amount of data available, hybrid modeling requires new frameworks and tools for combining first-principles with data-induced knowledge, including for instance the ability to incorporate new types of data (Bhosekar and Ierapetritou, 2018) and for dealing with the 5 Vs of Big Data (Reis et al., 2016). A mechanistic modeling structure can also facilitate the integration of data-driven models into pre-existing control structures (Zhu and Zabaras, 2018). More than before, in this new stage of development of hybrid modeling, special focus should be devoted to finding the optimal parsimony/model complexity/bias-variance trade-off, in the bi-dimensional domain of white-box and black-box modeling approaches (Figure 2), for a given purpose (modeling, implementation or computation complexity), i.e., the one maximizing the quality of information generated in the empirical study (Reis and Kenett, 2018).

However, balancing the bias-variance trade-off in the bidimensional hybrid modeling plane, is a problem that has not been systematically addressed so far. Therefore, in this critical review we aim to bring forward the relevant challenges of implementing hybrid modeling in Big Data environments by reviewing the publications made so far in this area, identifying the scientific gaps and suggesting opportunities for future research directions. We also discuss how to convert big data into smart data to be integrated with first-principles models and other forms of prior knowledge, in the search for the optimal complexity for the model, as shown in Fig. 3.

The review is organized as follows. In the next section, the terminology and definitions on the topics discussed in this paper are summarized and systematized. Then, the mechanistic and data-driven methodologies are first presented isolated to set the ground for introducing the integrated structures for hybrid modeling and their identification techniques. Afterwards, hybrid modeling applications on chemical and industrial processes are reviewed. The paper ends with a discussion of the scientific gaps identified in the literature.

2. Terminology and definitions

Research on the development of modeling approaches that combine mechanistic and data-driven elements started in the beginning of the 1990s and continue today. From the early 2000s onward, the number of publications raised steeply with contributions arising from different fields. The diversity of backgrounds led to different terminologies for describing the combination of mechanistic and data-driven models (hybrid, gray-box, etc.). Furthermore, different understandings exist of what is really an hybrid modeling approach – from broader definitions, where data may just be used to adjust mechanistic parameters, to stricter definitions, where two dissimilar sub-models (mechanistic and data-driven) are required to be simultaneously present. Therefore, in the current section, we systematize the different perspectives and define the one adopted in this review.

2.1. Terminology and definitions found in the published literature

The idea of integrating mechanistic with data-driven models was first introduced in 1992 and consisted of combining ANNs or radial basis functions (RBFs) with equations expressing energy and mass balances (Psichogios and Ungar, 1992; Su et al., 1992; Johansen and Foss, 1992; Kramer et al., 1992). White and black-box models can be organized in different hybrid structures, serial and parallel, which will be described and discussed in more detail further ahead. More than just combining different modeling techniques, the general goal is to gather all knowledge available on a system and create synergies that yield better prediction accuracy and/or stability in the final model. In chemical systems, one can distinguish three types of knowledge about the processes: mechanistic knowledge, heuristic knowledge and knowledge within process data (Schubert et al., 1994; Oliveira, 1998).

Mechanistic or phenomenological knowledge is the more sophisticated type of knowledge as it involves deep understanding of established laws of physics, phenomena explained by physical/chemical causes or relations between empirical observations of phenomena. It is often described by mathematical models. These are called first-principles, mechanistic, phenomenological or white-box models. Generally used interchangeably, the aforementioned terms actually refer to sightly different modeling approaches: (i) first-principles models come entirely from the established laws of nature; (ii) mechanistic models describe systems by referencing the laws of nature, but their parameters, although with physical meaning, must be determined from data; (iii) phenomenological models represent empirical relations and their parameters determined from data do not have physical meaning; (iv) finally, white-box models refer to situations where all relevant interactions are known and are made explicit, in a transparent way. For the sake of simplicity, we will adopt the term “mechanistic model” to address to any of the citations described above (von Stosch et al. (2011); Solle et al. (2017)).

Heuristic knowledge is a kind of knowledge that is often overlooked in most modeling methodologies. This knowledge is drawn from common sense or previous experience and is generally abundant among plant operators. Attempts to model this knowledge have been implemented using fuzzy theory and expert systems (Zadeh, 1973; Kosko, 1992; Wang, 1993).

Finally, knowledge within process data, that was largely ignored until the 1980s, has shown a gradual increase afterwards and boomed in the transition to the 21st century, with the emergence of all the conditions that converged to what is now called as Industry 4.0 and Big Data (Reis et al., 2016; Reis and Gins, 2017). Models of this kind are called data-driven, statistical, black-box, data analytics, chemometric, etc., and infer relevant information from large databases (Hastie et al., 2016; Rato and Reis, 2018).

Research on combining mechanistic and data-driven models have diverged on the designations adopted. Early contributions from the PSE arena called it hybrid modeling due to the fact that two contrasting modeling techniques were being combined, such as mass balances and an ANN (Psichogios and Ungar, 1992; Kramer et al., 1992; Thompson and Kramer, 1994). The term hybrid modeling caught up and became the unofficial, *de facto* standard in PSE, although Su et al. (1992) used the terminology “integrated neural network” to designate a similar modeling approach.

Some variations to the hybrid terminology were introduced over time such as hybrid semi-parametric modeling (von Stosch et al., 2011; Von Stosch, 2011; Yang et al., 2011; von Stosch et al., 2014) and knowledge-based hybrid modeling (Georgieva et al., 2003). The first was meant to emphasize the presence of a parametric, *i.e.* with a finite dimensional number of parameters, and a nonparametric submodel. The latter terminology aims to highlight the fact that the new modeling approach integrates all kinds of knowledge about the system.

It is also important to recognize that while hybrid modeling is very often mentioned in literature, authors do not always agree on how broad its scope can be. While Tahir et al. (2019) claimed to have developed a hybrid soft sensor because parameters from a transfer function (with a mechanistic origin) were identified using experimental data, Solle et al. (2017) believed that *the only hybrid modeling approach that actually leads to a hybrid model is the composition of submodels of different types*.

In parallel, works related to process optimization began to use the designation of grey/gray-box modeling (Tulleken, 1993; Van Can et al., 1996). This practice is a consequence of the pre-existing well-established terminology of black box models to designate situations where the way inputs are mapped onto outputs is unknown (as if the model was inside a black box) and the opposite case where one has access to all mechanistic details connecting inputs to outputs, called white box models (also by opposition to black box case). Therefore, in models where data is available alongside some partial phenomenological knowledge, the term grey/gray modeling was adopted (Tan and Li, 2002; van Sprang et al., 2005), to reflect such middle-ground situation.

As for the hybrid modeling designation, the precise definition of the widely used term grey/gray-box may also vary. For instance, in process optimization applications, when a mechanistic model is particularly cumbersome to evaluate, a surrogate model is often built to overcome the high computational burden, yielding what is also called, a gray-box model (Asprion et al., 2019). This term is also often used interchangeably with hybrid modeling (Bonvin et al., 2016; Beykal et al., 2018a; 2018b). Furthermore, some authors from the classical side of optimization call gray-box to a model that has no data-driven part (the case of condensed transshipment models in heat transfer network synthesis, where it is only relevant to model the relationship between inputs and outputs from system boundaries based on energy balances defined in a superstructure, without detailing the streams that are actually exchanging heat) (Papoulias and Grossmann, 1983a; 1983b; 1983c).

In the initial incursions of PSE to hybrid modeling, the goal was to complement mechanistic models with information inferred from data regarding aspects that were missing (e.g., kinetics). Thus, the starting point was a mechanistic representation and the idea is to improve it by working out the parts where knowledge is lacking. By the turn of the century, data-driven domains also start using the term hybrid modeling in the context of inferential sensors. The goal was to bring different dimensions to the model, such as causality, interpretability, physicochemical consistency, among others (Vedam and Venkatasubramanian, 1999). One approach focused in the agglomeration of all sources of knowledge was proposed by Peres et al. (2001), who called it, knowledge based modular network. Further examples can be found in statistical monitoring of batch processes, where methods where developed that introduced external mechanistic information in multiway data-driven methodologies, leading to superior fault detection performances (Gurden et al., 2001; Yoon and MacGregor, 2001; Ramaker et al., 2002). Lima and Saraiva (2007) also presented a novel framework to build semi-mechanistic model structures that combine mechanistic and empirical models.

With the emergence of the big data paradigm, data-driven methods grew in relevance and many applications have been proposed for handling a variety of problems. It would be fair to say that most of them disregard mechanistic and heuristic knowledge in favor of machine learning, artificial intelligence and data mining tools (Venkatasubramanian, 2019). However, some authors begin to realize the limitations of big data, especially when collected from plants operating most of the time under normal conditions, with limited variability patterns and therefore low information content,

Table 1
Summary of terminology in literature used for the combination of mechanistic and data-driven models.

Terminology	Publications
Hybrid modeling	(Kramer et al., 1992; Psichogios and Ungar, 1992; Thompson and Kramer, 1994; Tsen et al., 1996; Martinez and Wilson, 1998; Oliveira, 1998; Nascimento et al., 1999; Aguiar and Filho, 2001; Duarte and Saraiva, 2001; Tian et al., 2001; Yoon and MacGregor, 2001; Bollas et al., 2003; Duarte and Saraiva, 2003; Georgieva et al., 2003; van Lith et al., 2003; Chen et al., 2004; Duarte et al., 2004; Oliveira, 2004; Henneke et al., 2005; Lee et al., 2005; Kahrs and Marquardt, 2007; Brendel and Marquardt, 2008; Kahrs and Marquardt, 2008; Kahrs et al., 2009; Kumar et al., 2010; Wang et al., 2010; Carinhas et al., 2011; von Stosch et al., 2011; Von Stosch, 2011; Yang et al., 2011; Safari et al., 2014; von Stosch et al., 2014; Bonvin et al., 2016; Azarpour et al., 2017; Solle et al., 2017; Hanachi et al., 2019; Tahir et al., 2019; Venkatasubramanian, 2019)
Grey/Gray(-box)	(Tulleken, 1993; Van Can et al., 1996; Tan and Li, 2002; Xiong and Jutan, 2002; Bazaei and Majd, 2003; Sohlberg, 2003; van Sprang et al., 2005; Bonvin et al., 2016; Romijn et al., 2008; Beykal et al., 2018b; 2018a; Asprion et al., 2019)
Knowledge based modular networks	(Peres et al., 2001)
Incorporating external information	(Gurden et al., 2001; Yoon and MacGregor, 2001; Ramaker et al., 2002)
Semi-mechanistic model structures	(Lima and Saraiva, 2007)
Theory-guided data science	(Karpatne et al., 2017a)
Physics-guided neural networks	(Karpatne et al., 2017b)
Physics-informed neural networks	(Raissi et al., 2017; 2019)
Knowledge-integrated sparse matrix	(Luo and Bao, 2018)
Process-specific structure	(Reis et al., 2019; Wu et al., 2020)

despite their large volume. To complement the limitations of industrial data, efforts have been devoted to include process knowledge into machine learning and other data-driven methods of the big data era. Climate change researchers introduced the terminology "theory-guided data science" and "physics-guided neural networks" to emphasize a new kind of data-driven methodologies that are constrained to comply with physical or phenomenological constrains (Karpatne et al., 2017a; 2017b). Luo and Bao (2018) observed that in an industrial environment, sensors can fail or be out of commission originating sparse data matrices. To address this issue, the knowledge-integrated sparse matrix was proposed, where prior knowledge attempts to compensate for missing data. Fault diagnosis is a process engineering problem that requires a model of the causality relationships between observed variables, a characteristic frequently lacking in pure data-driven models. Various approaches were proposed using graph theory (Vedam and Venkatasubramanian, 1999), gray box modeling (van Sprang et al., 2005) or parity relations (Yoon and MacGregor, 2000). These were collectively classified as incorporating process-specific structure in a recent review paper (Reis et al., 2019). A summary of all the terminologies found in literature is presented in Table 1.

3. Modeling methodologies

The implementation of a modeling framework falling under the scope of hybrid modeling requires a careful consideration of the spectrum of modeling techniques (mechanistic and data-driven), model structures, system identification methodologies and model selection and validation methods. In this section, a critical overview is provided on these topics.

3.1. Mechanistic modeling

The mathematization of Chemical Engineering (the so-called Chemical Engineering Science), emerged in the 1960s with the works of several seminal authors (Bird, Stewart and Lightfoot; Rutherford Aris; Jacques Villermaux, etc.) and since then models have been in the core of PSE (Bird et al., 1960; Solle et al., 2017). It is important to point out that modeling is not just about writing the equations for the system behaviour. It also includes a broader view of the problem at hand, a clear statement of model hypotheses and assumptions, and the validation of the model (Bequette, 1998; Cameron et al., 2001).

The construction of a mechanistic model for hybrid modelling frameworks, depends on the available prior knowledge. Usually this knowledge is in the form of balances of extensive properties, such as mass and energy balances (Psichogios and Ungar, 1992; Oliveira, 2004; Azarpour et al., 2017) or physical constraints (Su et al., 1992; Thompson and Kramer, 1994; Gurden et al., 2001). These mathematical statements can be expressed more simply as algebraic equations or, with increasing complexity, as ordinary differential equations (ODEs) (for lumped parameter system modeling), differential algebraic equations (DAEs) or PDEs (for distributed parameter system modeling). By increasing the complexity of the mechanistic model, one reduces the structural mismatch between the model and the chemical process, but the larger number of parameters can lead to an unsuitable model if there is not a way to reasonably estimate said parameters (this is the bias-variance trade-off for mechanistic models).

In most cases found in the literature, process knowledge is encoded in equations such as the above, but in some situations equations are not available and this does not mean the absence of process knowledge. Process flow diagrams and even engineering expertise can provide enough information to improve a data-driven model. Vedam and Venkatasubramanian (1999) used signed directed graphs (SDGs) to encode variable causal relations between process variables, increasing the readability and performance of a principal component analysis (PCA) based fault detection and diagnosis methodology. The SDG was designed recurring to physical models or expert knowledge and when a fault is detected, the contributions of each variable are evaluated against a threshold. Those with higher values are signalized in the SDG to uncover the root cause. This concept was successfully applied to the Tennessee-Eastman benchmark by Wan et al. (2013) and was further extended by He et al. (2014) to include fuzzy systems.

A statistical process monitoring methodology that incorporates prior knowledge and aims for a more robust fault detection and diagnosis was proposed by (Yoon and MacGregor, 2001). The information incorporated combines variable relations, independent manipulated variables, controller information and process variations. Using orthogonal projections (or generalized least squares), the observation matrix is decomposed into a term that reflects prior knowledge and a term including the remaining effects present in the data.

Prior knowledge on a batch process was divided into batch-run specific (reactor conditions, cleaning & maintenance, reactor number, upstream conditions, batch duration) and process specific information (dosing profiles, ideal trajectories, process stages, variable grouping, instrumental characteristics, spectra of pure components, kinetics) by Ramaker et al. (2002). Multiway regression to

product quality was used along with orthogonal signal correction to remove information uncorrelated with product quality in batch process. Still on batch process monitoring, Luo and Bao (2018) proposed the knowledge-integrated sparse (KDIS) modeling approach. The KDIS begins with designing the knowledge-integrated sparse projection matrix by classifying process variables into ordered categories of correlation: control, reaction and location, and type and location correlations. Furthermore, if *a priori* information is present, the correlations nature (positive or negative) can be coded into the KDIS projection matrix. This matrix has the advantage of being sparse and easy to interpret, besides its usage being akin to a loadings matrix in multiway PCA.

More recent works, expand the possibilities for encoding prior information. Regularization techniques and smoothing can be used to input process knowledge in datasets. Orthogonal collocation on finite elements allowed the continuous approximation of sparse kinetic data through the introduction of shape constraints (Vertis et al., 2016) and physically impossible relations between variables can be blacklisted in a Gaussian Bayesian Network probabilistic description of the system structure (Yang et al., 2018; 2020).

Furthermore, Venkatasubramanian (2019) enumerated some insufficiently explored methodologies in chemical engineering, stating that there is a need to develop domain-specific representations and languages for hybrid modeling to include all available knowledge. To accomplish this, ontologies (Bohács and Rinkács, 2017), Petri networks and even a return to rule-based systems and agent-based models used in the expert systems of the 1980s are suggested.

From the literature review on mechanistic models used in hybrid modeling, it is important to highlight that equations are by far the most widely used form for expressing prior knowledge. This is expected, since it was the dominant modeling paradigm in Chemical Engineering during the last 50 years, yielding many solutions to a wide variety of problems. Their wide success and relevance should be taken into account, now that the use of data-driven methods is fast increasing and more focus is placed on the exploitation of the immense data resources. This was the reason of the success of hybrid models and we argue that there still exists a vast uncharted territory to explore, namely from translating plant floor experience and process flowsheets into workable models, to the use of more complex mechanistic model structures through data augmentation and self-supervised Artificial Intelligence methods. To accomplish this, graph theory is a foreseeable useful tool, as previous works have shown (Cheng et al., 2008; Chiang et al., 2015).

3.2. Data-driven modeling

The growing availability of data since the turn of the century, due to continuous development of faster and more informative sensors, larger databases and more computational power, have pushed engineers to start looking at data as a valuable source of information (Kresta et al., 1991; MacGregor et al., 1994; MacGregor and Kourti, 1995; Kourti and MacGregor, 1995; Reis et al., 2019). Data science emerged as an area of knowledge that involves statistics, mathematics and computer science, to handle the challenges raised by the new data intensive environment called big data. Research on this topic in the context of PSE has driven the development of techniques to deal with several data-related issues such as high-dimensionality, collinearity, sparsity, nonlinearity, non-stationarity, the existence of noise, outliers and missing data, among others (Rendall and Reis, 2018). Next, we report some of the most used data-driven models in the context of hybrid modeling approaches.

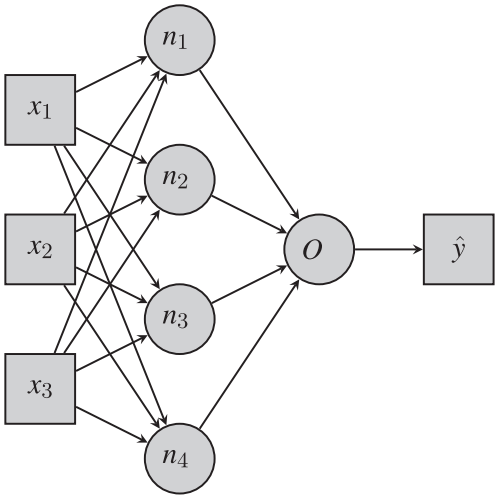


Fig. 4. Schematic representation of an artificial neural network with one hidden layer (x_i represents inputs, n_i are hidden nodes, O is the output node and $\hat{y}$ is the network's estimate).

3.2.1. Neural networks

In the scope of hybrid modeling, several different inferential modeling techniques can be used to incorporate data-driven knowledge into mechanistic descriptions. The most commonly used in the literature are ANNs and RBF networks, due to their ability to model nonlinearities and the fact they do not require structural knowledge of the system being modeled (Psichogios and Ungar, 1992; Thompson and Kramer, 1994; Oliveira, 2004).

ANNs consist of a series of data processing units, called neurons, connected by flows of information. There are many different types of neural networks but the most adopted for hybrid modeling is the multilayer perceptron, as shown in Fig. 4. An ANN with three layers is considered an universal approximator (Oliveira, 2004). Each neuron computes the weighted sum of all signals received from its connections (S_j) from a previous layer, plus a bias term. S_j are known as activations, linear combinations of the input variables (Bishop, 2006),

$$S_j = \sum_{i=1}^{n} \omega_{i,j} X_i + \theta_j \,, \tag{1}$$

where $\omega_{i,j}$ are the weights of neuron i in layer j, X_i are the inputs to neuron i and θ_j is the bias term. An output is then generated through an activation function,

$$\hat{y}_k^{(m)} = f(S_j) \,, \tag{2}$$

The activation function can be linear, an hyperbolic tangent, the logistic function or a rectified linear unit (ReLU). In the learning process (or parameter estimation), the most widely used method is the backpropagation algorithm (Hecht-Nielsen, 1992). This method minimizes the mean squared error (MSE) between the training data and p model estimates for the r output variables,

$$\min(MSE) = \min\left(\sum_{m=1}^{r} \sum_{k=1}^{p} \left(y_k^{(m)} - \hat{y}_k^{(m)} \right)^2 \right) \text{ Nascimento et al. (1999)} \,. \tag{3}$$

It is worth to note the outputs should be weighted/scaled. Generally, many of predicted outputs can have different magnitudes, and without proper scaling, the MSE will give more importance to outputs with larger values. Even though most ANNs used in hybrid modeling are composed of three layers, many more layers can be used with deep learning techniques and stochastic gradient descent for the learning process (Schmidhuber, 2015).

Another possibility often found in the literature consists in using a Gaussian radial basis function (RBF) (Bhosekar and Ierapetritou, 2018) as the activation function (2),

$$f(S_j) = e^{-(\gamma S_j)^2}, \tag{4}$$

yielding what is known as a RBF network (Hastie et al., 2016).

A concern when dealing with ANNs, is the need to avoid overfitting. Ghosh et al. (2019) argue that subspace identification models allow for a more direct handling of the overfitting problem and have passed over ANNs in favor of subspace identification. Others authors use Bayesian regularization to cope with this problem (Hirschen and Schäfer, 2006; Burden and Winkler, 2009; Ticknor, 2013).

3.2.2. Support vector machines

Support vector machines (SVM) have been used as a data-driven method to build the nonparametric submodel in hybrid models. This methodology was chosen for the simplicity of the training algorithm and faster computations (Wang et al., 2010; Yang et al., 2011). SVM maps input vectors x into a high-dimensional feature space through a possibly nonlinear mapping, chosen *a priori*. The model is defined as

$$f(x) = \sum_{i=1}^{P} a_i K(x, x_i), \tag{5}$$

where a_i is an element of the parameter vector, x_i is a vector of regressors, K is a function referred to as the kernel and P is the number of parameters. The kernel used can be a RBF, a polynomial or a two-layer neural network (Vapnik, 1999).

3.2.3. Multivariate adaptive regression splines

Multivariate adaptive regression splines (MARS) are nonparametric estimators develop by Friedman (1991) that model a dataset through localized piece-wise submodels (splines). The algorithm establishes a procedure for selecting local spline functions and partitioning the original data domain, and implements a backward procedure to prune unnecessary splits to achieve the most parsimonious set of spline functions and a smoothing procedure to provide the final model with continuity at the knots. The model takes the following form:

$$\hat{y} = \sum_{n=1}^{M} a_m B_m(\mathbf{X}), \tag{6}$$

where $\hat{y}$ is the predicted estimate, M is the number of basis functions, a_m are adjustment parameters, $\mathbf{X}$ is the vector of input variables and $B_m(.)$ are the basis functions,

$$B_m(\mathbf{X}) = \prod_{k=1}^{K_m} H\left[s_{k,m}\left(\mathbf{X}_{\upsilon(k,m)} - \mathbf{T}_{k,m}\right)\right], \tag{7}$$

where K_m is the number of domain splits, $H[.]$ is the step function, $s_{k,m}$ is ± 1 and $\mathbf{T}_{k,m}$ are the knots of the splits. Although this method has a data-driven character, MARS is locally composed by parametric functions which can be easier to understand and interpret as the method is based on recursive partitioning (Duarte and Saraiva, 2001; 2003; Duarte et al., 2004).

3.2.4. Latent variable methods

Latent variable methods are modeling approaches used to overcome issues related to high-dimensional correlated predictors. These methods assume the existence of an underlying set of unmeasured variables that are responsible for the observed variability in the predictors and response spaces. Latent variables are estimated using linear combinations of the original variables, characterizing the latent variable model subspace (Rendall and Reis, 2018).

Among the most popular latent variable methods are PCA, principal component regression (PCR) and PLS. PCA extracts the eigenvectors of the predictors' covariance matrix and arranges them in decreasing order of their eigenvalues magnitude. The eigenvectors associated to higher eigenvalues form linear combinations of variables (called principal components) that explain more of the original variability in the data, while eigenvectors associated to eigenvalues closer to zero mainly explain noise (Bro and Smilde, 2014). The principal components are uncorrelated and therefore the covariance matrix of the most important ones (those with high eigenvalues), can be easily inverted. With PCA, the predictor matrix ($\mathbf{X}$) can be rewritten using the PCA scores matrix ($\mathbf{T}$), loading matrix ($\mathbf{P}$) and a residual error matrix ($\mathbf{E}$). PCR consists in relating a response with the principal component scores through ordinary least squares (OLS) (Jackson, 1991; Jollife, 2002; Rendall and Reis, 2018).

PLS, on the other hand, finds the linear combinations of predictors with maximal covariance with the response. These linear combinations are called latent variables. They are fewer than the number of predictor variables and besides explaining the response, they also capture a large amount of the predictors variability (Wold et al., 2001). In this method, the predictor matrix ($\mathbf{X}$) and the response matrix ($\mathbf{Y}$) are described by score matrices ($\mathbf{T}$ and $\mathbf{U}$), loading matrices ($\mathbf{P}$ and $\mathbf{Q}$) and residual error matrices ($\mathbf{E}$ and $\mathbf{F}$). In the inner model, $\mathbf{Y}$ is linearly related to $\mathbf{X}$ by the vector β. There are several algorithms available to compute PLS parameters but the most widely used are the non-linear iterative partial least squares (NIPALS) (Wold, 1975; Andersson, 2009) and SIMPLS (de Jong, 1993).

PLS extensions are also available in order to capture nonlinear associations between the predictor and the response, namely through the incorporation of polynomial functions, neural networks or RBF (Lee et al., 2005; Von Stosch, 2011).

In recent years, autoencoders have been increasingly more adopted as a method to extract latent variables. An autoencoder is a feed-forward ANN that outputs a reconstructed version of the input. It does this by simultaneously learning a mapping that converts the input into a reduced set ϕ and learning the associated inverse mapping that converts ϕ into the output. ϕ is often called "code" and its dimensionality is predetermined by the networks structure. The undercomplete autoencoder forces ϕ to have lower dimension than the input, yielding a reduced set of latent variables. (Goodfellow et al., 2016; Sun et al., 2020; Tsay and Baldea, 2020)

3.2.5. Summary

It is imperative to recognize the importance of data-driven models in a hybrid structure, specially in the increasingly data intensive industrial settings. The first wave of scientific papers on hybrid modeling, mainly used data-driven techniques to fill in the gaps of knowledge in mechanistic models, such as unknown nonlinear behavior (e.g., kinetics) or unknown parameters suspected to have a complex dependency on the process variables behavior. To achieve this, authors recurred mainly to the integration of ANNs and RBF networks with first principles models. One notable exception came from the work of Bikmukhametov and Jäschke (2020) that presented a framework to enhance data-driven models (such as gradient boosting and long short-term memory networks) with engineering knowledge. With the increasing proliferation of data-driven methods in the analysis of data from the chemical processing industry, other methods were explored but always with the same goal. In Table 2, a summary list of the data-driven and mechanistic techniques found in the literature during the initial years of hybrid modeling is presented bellow.

Currently, as data-driven modeling becomes more widely accepted as a legitimate and useful technology for analysing process

Table 2
Compendium of modeling frameworks.

Data-driven	First-principles
Ordinary least squares	(non)linear algebraic equations
Stepwise regression	(non)linear ODEs
Error in variables models	DAEs
Principal component analysis	PDEs
Neural networks	Probabilistic graphical models
Support vector machine	Bayesian networks
Multivariate adaptive regression splines	Agent-based models
Fuzzy systems	Expert systems
Genetic algorithms	Rule-based systems

data, an opportunity arises to apply them not just as a complement to mechanistic models, but as a way to model and analyze a chemical process, which may benefit from the incorporation of prior knowledge and process understanding. At the present time, a lot of engineering knowledge is included in an *ad-hoc* way by inputting mass balance calculations and heat transfer coefficients as training data to augment the data-driven model's training set. Domain knowledge is also used to interpret results from data-driven models. Therefore, the dominance of data-driven or process knowledge in the modeling framework to adopt are not limiting prerequisites, but something that should be naturally accommodated and flexibly tuned case by case, depending upon the information sources available, their relative quality and the specific goal of the analysis.

3.3. Hybrid modeling structures

Finding the right structure for a hybrid model is of paramount importance for the achievement of the goal for which a model is destined. The type and quality of information available will play a defining role in the subject. In the classical hybrid modelling literature, the prevalent way to combine the two modeling approaches (mechanistic and data-driven) starts from an analysis of the structure of mechanistic model and its assumptions. A parallel configuration can compensate for mechanistic structural mismatch, but if the mechanistic structure is accurate enough, then a serial configuration is usually a better choice (von Stosch et al., 2014).

3.3.1. Serial

The serial structure is fairly popular among hybrid modeling researchers with most of the published work focusing on the first configuration presented on Fig. 5. In this approach, the white box model is usually a mechanistic model that expresses conservation laws and transport phenomena, like mass, energy and/or momentum balances. The black box model is then used to account for the part of the chemical phenomenon for which there is no available model or it is too complex to formulate and parameterize (e.g., reaction kinetics).

Psichogios and Ungar (1992) designed an hybrid model structure comprised of a mechanistic model describing the systems transport phenomena and of an ANN capturing unknown reaction kinetics. The network's output is fed as a parameter to the mass and energies balances, making this an example of the first serial configuration presented in Fig. 5. This approach is more flexible than a single mechanistic model and extrapolates better than a pure data-driven model, thus increasing its reliability and interpretability. The authors argue that the use of a mechanistic model reduces the complexity of ANN leading at the same time to a model presenting higher accuracy than standard ANN, which is able to substitute parts of the mathematical model that would otherwise be difficult to obtain. The results of the simulation using hybrid modeling were compared against extended Kalman filtering and nonlinear programming (NLP) optimization. Hybrid modeling outperformed both methods in estimating the unobserved process parameter, *i.e.* reaction kinetics. Furthermore, it is theorized that when modeling a continuous process operating around a setpoint in a small region, hybrid modeling may not offer so clear advantages when compared to data-driven modeling. However, for systems undergoing larger changes or time-variant process parameters hybrid modeling attains superior performance.

The same base framework was employed by numerous authors with some modifications of the goal, identification methodology or sometimes the data-driven model adopted (See Table 3). The simplicity of the serial structure, the convenience of substituting an unknown parameter by a black box model and the higher predictive performance achieved explain the success of this structure and its wide adoption in reported studies.

In a data sparse environment, a serial framework can take a second configuration of Fig. 5. In this configuration, the estimates from the mechanistic model along with process variables are fed to a data-driven model which uses that information to improve the estimate, an approach similar to the parallel configuration discussed ahead (Nascimento et al., 1999). Aguiar and Filho (2001) followed this approach using a kinetic model as mechanistic model and an ANN as data-driven model. The predictions were compared with a pure ANN and the hybrid model performed better and was 50% faster in training. A similar reasoning can be followed by data augmentation: using a mechanistic model to augment the historical dataset at disposal and then fed the aug-

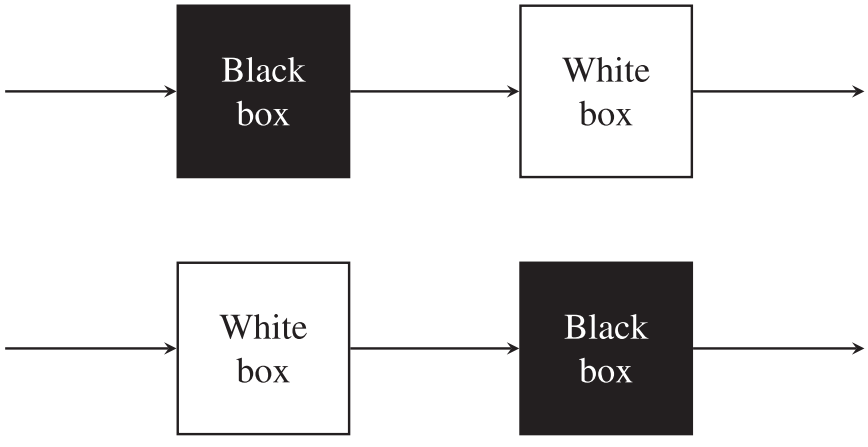


Fig. 5. Schematic representation of serial structure for hybrid modeling contemplating the two possible alternatives.

Table 3
Summary of data-driven models used for hybrid modeling and their structure.

Data-driven model	Structure	Publications
ANN	Serial	Psichogios and Ungar (1992); Tsen et al. (1996); Martinez and Wilson (1998); Nascimento et al. (1999); Aguiar and Filho (2001); Tian et al. (2001); Bazaei and Majd (2003); Bollas et al. (2003); Georgieva et al. (2003); Oliveira (2004); Kahrs and Marquardt (2008); Kahrs et al. (2009); Azarpour et al. (2017); Karpatne et al. (2017b); Chen and Ierapetritou (2020)
Neural network	Parallel	Su et al. (1992); Thompson and Kramer (1994); Garcia and Vilim (1998); Oliveira (1998); Peres et al. (2001); Xiong and Jutan (2002); Bollas et al. (2003); Chen et al. (2004); Safari et al. (2014); Lee et al. (2005); Chen and Ierapetritou (2020)
SVM	Serial	Wang et al. (2010); Yang et al. (2011); Chen and Ierapetritou (2020)
Padé expansion	Parallel	Tan and Li (2002)
Extended Kalman filter	Serial	Sohlberg (2003)
MARS	Parallel	Duarte and Saraiva (2001, 2003); Duarte et al. (2004)
ARMAX	Serial	Tulleken (1993); Kumar et al. (2010)
PCA	Serial	Destro et al. (2020); Hassanpour et al. (2020)
(N)PLS	Serial	Henneke et al. (2005); Carinhas et al. (2011); Von Stosch (2011); von Stosch et al. (2011)
(N)PLS	Parallel	Lee et al. (2005)
Fuzzy system	Serial	van Lith et al. (2003)

mented dataset to train the data-driven model (Tsen et al., 1996; Martinez and Wilson, 1998). A comparative study conducted by Tsen et al. (1996) shows that this structure attains superior results when compared to other serial or parallel configurations and the theoretical model itself. Nevertheless, this approach has not yet found much application in practice.

Batch polymerization reactors present great flexibility which is important in market-driven multi-product manufacturing settings. However, this added flexibility often goes hand in hand with a lack of mechanistic information when compared to continuous processes. To mitigate this issue, a hybrid model was built with a simplified mechanistic model that includes mass and energy balances and three stacked recurrent neural networks, in a serial configuration (Tian et al., 2001). The networks are trained with data generated from a cubic spline function adjusted from sparsely collected measurements from the batch reactor. The goal is to predict how temperature will affect the polymerization in terms of conversion, molecular weight and polydispersity to incorporate in a temperature optimal control. The results were compared with the best-single-network-based hybrid model and the stacked recurrent neural networks showed to be more reliable and accurate for controlling the polymerization processes (Tian et al., 2001).

3.3.2. Parallel

The parallel structure has also found a wide adoption among researchers on hybrid modeling. It usually finds good use in situations where a mechanistic model is available but its prediction power is limited due to limitation in describing some effects, nonlinearities or dynamic behavior. There are two schemes through which the parallel structure can be deployed: the cooperative and the competitive.

In the former, a data-driven model is set in place to learn, during training, the mismatch between the mechanistic model and historical data, in order to capture the unmodeled effects, nonlinearities or dynamic behavior. The predictions from the mechanistic model and the data-driven model are then fused, using one of the several available approaches (addition, multiplication, Kalman filtering), to yield the final response prediction (See Fig. 6). In the second scheme, both the mechanistic and the data-driven model are trained to yield the same prediction. These predictions are then combined by a weighting function (Dors et al., 1996; Peres et al., 2001; Galvanauskas et al., 2004; Ghosh et al., 2019) in order to provide a fused estimate of the target variable. The parallel structure is more limited in terms of extrapolation ability from the training domain, but the performance can be improved when certain effects in the system are uncoupled, *i.e.* when a system is decomposed for instance into a static non-linear component and a dynamic linear component (Chen et al., 2004; Su et al., 1992; Kumar et al., 2010; Wills et al., 2013).

Thompson and Kramer (1994) reviewed design and training approaches for hybrid modeling, presenting some of their advantages and disadvantages. They argue that a parallel semi-parametric approach yields better results. The reason is that the nonparametric model can compensate for uncertainties arising from process complexity. The authors proposed a structured approach to integrate RBF/ANNs with mechanistic models, maximizing the value of domain-specific knowledge. The structure takes a default parametric model in parallel with the nonparametric model in a Kalman filter where the default model provides a first estimate to be updated by the ANN. The predictions of both models are fused by simple superposition, *i.e.* added to give the output. Then, a parametric output model establishes equality and inequality constraints. Guidelines for the best way to include prior knowledge according to its nature are discussed. The hybrid model achieved better prediction performance compared to pure data-driven and pure mechanistic model. Combining prior knowledge helps providing accurate, consistent and reliable predictions when facing sparse and noisy data. Besides, hybrid modeling is flexible enough to provide customized models.

The parallel structure, despite being adequate for handling model mismatch, still relies on a robust mechanistic model. An example is the coupling a data-driven submodel, comprising a time domain partitioning procedure, with a mechanistic submodel. The data-driven submodel delimits zones where residuals are described by piecewise linear polynomials. These polynomials are obtained by performing univariate adaptive regression splines and the construction of local regression models considering only exogenous or exogenous and auto-regressive terms through a forward stepwise regression procedure. Results indicate that the inclusion of auto-regressive terms leads to better predictive capabilities. The hybrid approach proved to always outperform purely mechanistic models, but their superiority cannot be guaranteed against purely empirical models particularly if the mechanistic submodel does not make a good representation of the physical behavior of the process (Duarte and Saraiva, 2003; Duarte et al., 2004).

3.3.3. Surrogate models

Surrogate models, also known as substitute models (Romijn et al., 2008), metamodels (Jin et al., 2001) or response surface models (Wang and Georgakis, 2019), are simpler mathematical representations of more complex models. They require less computational effort to be run than the more rigorous representations, and have been extensively used in process modeling and optimization (Bhosekar and Ierapetritou, 2018).

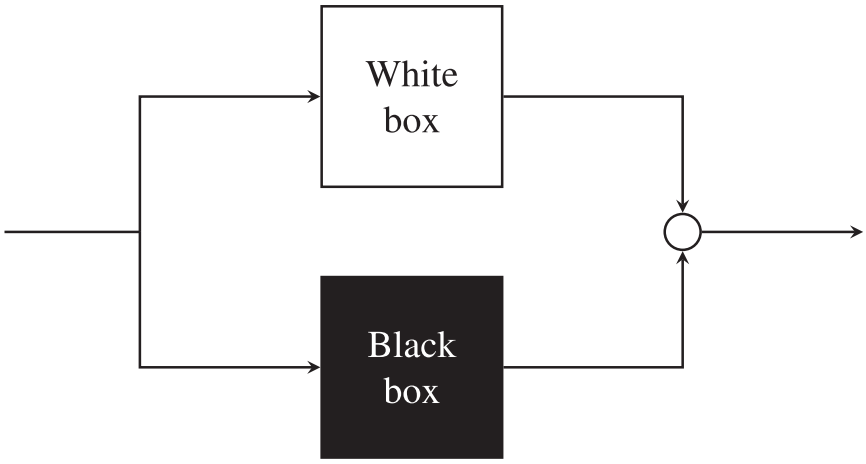


Fig. 6. Schematic representation of parallel structure for hybrid modeling.

These models are designed to yield unbiased predictions of sampled or simulated data (Bhosekar and Ierapetritou, 2018) which is useful to generate regular measurements from a irregularly sampled variable (Wang and Chiang, 2019). Another way to design a surrogate model is to generate data points from a complex mechanistic model and to use them to train a data-driven model (Solle et al., 2017). However, these model evaluations may be very computationally expensive, which has motivated research efforts towards improving the sampling efficiency. Eason and Cremaschi (2014) proposed two sampling methods to generate surrogate models: pure adaptive sampling algorithm and mixed adaptive sampling algorithm. The methods were compared to incremental Latin hypercube sampling.

Tsay et al. (2017) developed an adaptive simultaneous model identification that uses a superstructure for recursive multi-step design of experiments (DoE). This approach can be applied to a domain-restricted model. DoE is an approach that enables the correlation of output variables and input variables by designing a set of the minimum required experiments to obtain the desired information (Atkinson and Bailey, 2001; Box et al., 2005; Box and Draper, 2007; Montgomery, 2012; Wu and Hamada, 2009). Several expansions of the original DoE framework have been proposed over time, such as: Model-based DoE (Lundberg et al., 2015), optimal DoE (Atkinson and Donev, 1992; Atkinson and Bogacka, 2002; Cappuyns et al., 2006; Atkinson, 2008), robust DoE (Asprey and Macchietto, 2000; 2002) and, more recently, dynamic DoE (Georgakis, 2013).

Over the years, software has been develop to create efficient and robust surrogate models capable of being used to solve complex optimization problems (McBride and Sundmacher, 2019). Caballero and Grossmann proposed Kriging interpolation to generate surrogate models in nonlinear programming. Romijn et al. (2008) presented a hybrid methodology to find computationally efficient substitute models for nonlinear dynamical systems. The methodology takes a PDE and approximates its nonlinear part to an empirical model in a serial structure. The empirical model was trained through proper orthogonal decomposition or PCA. The results were compared with a pure ANN and showed that the hybrid model takes significantly lower computational effort for parameter estimation and is also more interpretable.

Henao and Maravelias (2011) proposed a methodology to generate surrogate models based on Aspen Plus simulations using ANNs. The surrogate models are used to approximate the more complex systems with the objective of superstructure optimization for process synthesis. Cozad et al. (2014, 2015) created the automated learning of algebraic models for optimization (ALAMO) which was later on extended by Wilson and Sahinidis (2017). More recently, another algorithm for global optimization of constrained grey-box computational problems (ARGONAUT) was published (Boukouvala and Floudas, 2017; Boukouvala et al., 2017). These software solutions will be brought again to discussion in Section 4.

3.3.4. Alternative structures

A comparative study of various modeling techniques was conducted using for benchmark a nonlinear dynamic system, namely the free radical polymerization of methylmethacrylate in continuous stirred tank reactor (CSTR). Five data-driven techniques (backpropagation, real-time recurrent learning, multi-layer recurrent network and the NMAX versions of the former two) were compared and the nonlinear moving-average with exogenous inputs (NMAX) techniques were found to performed better. Then, prior knowledge was included in the form of isotime lines to limit the networks dynamics, which improved predicting performance for all methods (Meert and Rijckaert, 1998).

For monitoring industrial batch processes, van Sprang et al. (2005) presented a strategy based on hybrid/gray models. The models, gray Tucker1 and gray Tucker3, are arranged in a parallel structure where the variation not explained by the mechanistic part is modeled by a latent variable approach. Both models have unique model parameters due to imposed constraints (reaction kinetics and Beer-Lambert's law). Gray Tucker1 led to a better fit and was found to be well-suited for online monitoring as it focus on within-batch variation. Gray Tucker3, on the other hand, has a simple interpretation for between-batch variation, being therefore more adequate for post-batch monitoring and analysis. A similar approach was employed by Gurden et al. (2001).

In a publication, Lima and Saraiva (2007) developed a symbolic framework for deriving hybrid mechanistic-empirical models, called semi-mechanistic model structures. This strategy results in a set of atomic equations that allow for empirical elements to be added selectively and locally. Through combinatorial optimization, the best hybrid model is selected based on available process data. Two case studies were studied: a simulated non-ideal CSTR and the Otto-Williams benchmark optimization problem (Williams and Otto, 1960). Results showed an order of magnitude improvement in model performance criteria over the initial mechanistic model performance. Furthermore, the model predictions in extrapolation indicate very good performance and stability, enabling good model-based optimization.

3.3.5. Summary

The choice of the hybrid structure is a critical decision to make when developing a hybrid modeling approach. In the literature reviewed (See Table 3), this decision is strongly influenced by the mechanistic model available. If mechanistic model has a high structural mismatch with the real process, the parallel structure (Fig. 6) is often favored, with a few exceptions where the preference went to the second serial structure (Fig. 5, bottom). This kind of structure allows engineers to use linear and static mass and energy balances even in nonlinear and dynamic systems, because the data-driven model will, in principle, account for those effects and accommodate them in the hybrid model. The disadvantage of the parallel structure is the lack of extrapolation properties when compared to the serial structure. That said, when the structural uncertainty on a mechanistic model is low, the adopted structure should be the first serial one (Fig. 5, top) due to its better extrapolation properties and prediction quality.

Surrogate models and other hybrid structures are also used, but it is harder to find a general set of rules for their adoption. Thus, there may be room for developing more generalized approaches for handling irregular measurements using surrogate models due to their interpolating capabilities, robustness and previous successes in control applications. Furthermore, the application of hybrid modeling techniques to monitoring batch processes was not extensively explored yet, but the opportunities are many. Given that this kind of processes usually have complex kinetics (*e.g.* polymerization, fermentation, etc.) and changes occur between batches (*e.g.* feedstock, pH fluctuations), the combination of prior knowledge with information inferred from inter/intra-batch data, can greatly improve batch tracking models.

3.4. System identification methodologies

Once the structure of the hybrid model is defined, it is necessary to estimate the models' parameters. Parameter estimation/identification is a well-known activity for mechanistic and data-driven models. In mechanistic modeling, identification serves the purpose of estimating equation parameters, whereas in data-driven it usually regards the definition of the number of parameters as well as the estimation of their values. In statistical learning, identification is also known as training (Hastie et al., 2016).

The identification activity aims at optimizing an objective function, usually a measure of fit between experimental data and model predictions. In hybrid structures, it is often the case that mechanistic and data-driven models are identified separately. Usually, priority is given to the mechanistic model and once this model is set, the data-driven model is identified with standard techniques (*e.g.* backpropagation for ANN (Hecht-Nielsen, 1992), NIPALS for PLS (Wold, 1975)). Besides this direct approach, researchers also proposed an incremental approach, a sensitivity approach and an evolutionary computing approach for environments where data is sparse, infrequent and noisy (von Stosch et al., 2011). These methods are briefly reviewed in the rest of this section.

3.4.1. Direct approach

In the direct approach, the priority in identification is given to the mechanistic model. Since it is a parametric model with a predefined number of parameters, these are identified without considering the data-driven model. Once the mechanistic parameters are known, the workflow then proceeds to the identification of the data-driven structure and parameters. The procedure can be repeated for several iterations until convergence, but it does not guarantee optimality in the end, and sometimes it is not stable either.

Yang et al. (2011) follows the identification of a hybrid model of the toluene nitration process using SVM to model reaction kinetics, while mass and energy balances constitute the mechanistic part. The method employs simultaneous estimation of unknown mechanistic parameters and the identification of the data-driven model, in a two-step methodology: (i) the regularized identification of the black-box component of the entire model and (ii) the optimal tuning of the regularization parameters and the mechanistic parameters, simultaneously. Cross-validation error was used as the cost function and the results showed accurate estimation of mechanistic parameters if the measurement error in data was small; otherwise the estimation accuracy was diminished. Nevertheless, the hybrid model outperforms a pure data-driven model in extrapolation circumstances, especially when the extrapolation has little influence on the data-driven component of the hybrid model.

A generic model-based framework for hybrid dynamic modeling is presented by Azarpour et al. (2017). This framework is fairly comprehensive, emphasizing process identification, problem and objective definition and data collection in the first steps. The hybrid model development comprises the elicitation of conservation equations, the classification of dominant process phenomena and the modeling of reaction and deactivation rates with an ANN. The final steps of the framework comprise model validation and simulation development.

3.4.2. Incremental approach

The incremental approach decomposes a large problem into smaller subproblems using a decomposition algorithm based on the Dulmage-Mendelsohn algorithm (Dulmage and Mendelsohn, 1963). The hybrid model, for example, describing an ethylene glycol process and comprising a mechanistic model and ANNs for reaction kinetics (Kahrs and Marquardt, 2008), is decomposed into four parts: an unknown part comprising all equations with unknown parameters, an undetermined part without unknown parameters, a determined part without unknown parameters and an overdetermined part without unknown parameters. The overdetermined part is treated as a data reconciliation problem, while the rest are solved through parameter estimation and neural network training. Since this incremental method does not guarantee optimal parameter estimation, a simultaneous parameter estimation problem is solved to correct the incremental estimates (Kahrs et al., 2009).

A design criterion has been derived for the identification of data-driven model parts from dynamic data. Applied to identification of reaction kinetics using neural networks in a hybrid process model structure, the approach achieved a good performance. The design criterion proposed is not limited to the identification of reaction kinetics but universally applicable to model identification problems from dynamic data (Brendel and Marquardt, 2008). This identification technique was included in a software modeling package, called CHEMASIM, developed by a major chemical company (Asprion et al., 2019).

3.4.3. Sensitivity approach

The sensitivity approach was proposed by Psichogios and Ungar (1992). It consists in training the hybrid model by backpropagating the errors through the ANN and the associated mechanistic model. It involves the computation of gradients in the hybrid model through nonlinear algorithms like sequential quadratic programming. The gradients are provided by information taken from the sensitivity equations. This method was also used by Georgieva et al. (2003) to identify the parameters of a hybrid model in an industrial fed-batch evaporative crystallization process for sugar cane refining with the purpose of predicting the size distribution parameters of the crystals. The mechanistic submodel results from a nonlinear set of DAEs reflecting general mass and energy balances while the data-driven submodel is an ANN

describing the growth rate, nucleation and agglomeration kinetic parameters. The hybrid structure is serial and the results showed a better agreement between experimental data and hybrid model predictions than what observed with the complete mechanistic model.

Oliveira (2004) proposed that reaction kinetics can be broken up into a mechanistic part and a data-driven part, running in parallel and fused by multiplication. In this mathematical formalism, the priority is in the mechanistic knowledge. In two bioreactor case studies (fed-batch production of a recombinant protein and baker's yeast production process), two identification strategies are compared: (i) direct minimization of errors in concentrations and, (ii) direct minimization of errors in reaction rates. The best strategy proved to be the direct minimization errors in concentration in the sense of least squares employing SQP with gradients computed by the sensitivities equations. This method can be used in a noisy and sparse data environment. Furthermore, stability conditions (BIBO) were assured by imposing constraints during optimization.

The sensitivity method was also used for parameter identification of a hybrid dynamic nonlinear PLS model. Macroscopic material balances are combined with a nonlinear PLS modeling kinetic rates. The method was benchmarked against reference dynamic nonlinear PLS models and applied to the Park-Ramirez simulation case (Park and Ramirez, 1988) and to an experimental case study of *B. pertussis* cultivation. The new model had fewer parameters, yielding higher statistical significance (translated into higher BIC values), significantly better prediction performance, better calibration and better extrapolation properties. The model can be applied to noisy and sparse data without the need for interpolation or smoothing, retaining PLS features such as dimension reduction and the opportunity to interpret scores. It also presents a damping effect on error propagation (Von Stosch, 2011).

The main disadvantage of the sensitivity approach is the computation of the sensitivities equation which is computationally hard and puts a constraint on the complexity of the mechanistic model chosen for the hybrid model. It is however a stable approach that has the guarantee of optimality under certain assumptions.

3.4.4. Evolutionary computing approach

Evolutionary computation algorithms are used to build empirical models for chemical processes. These techniques do not require explicit specification of a model structure but explore candidate models assembled from sets of variables, parameters and simple mathematical operators. When applying these methods to a leaching process, the authors (Greeff and Aldrich, 1998) concluded that the models obtained were either of comparable accuracy or significantly more accurate than those developed through standard least-squares methods, although they are difficult to interpret. Furthermore, search time spent was consumed by evolution of appropriate parameter values for the models, while model structures could be identified more rapidly.

These evolutionary algorithms have also been used for hybrid model identification. In an example of this methodology (Tan and Li, 2002), a mechanistic submodel consisted on a balance of momentum and the data-driven model is a Padé approximation in the form of a regression model, although the author claims that a power series polynomial, fuzzy logic or ANN can be used as well. The structure is parallel with the data-driven model representing non-explained variation by the mechanistic model. The evolutionary technique is capable of accommodating multiple objectives to examine different trade-offs between the model complexity and fitting accuracy, while the introduction of a hybrid model structure elevates the algorithm to a more cost-efficient technique.

3.5. Model selection and validation

Assessment of the model's performance is extremely important when it comes to decide if the model is appropriate for the intended goal. In cases where it is not clear the best hybrid structure to adopt *a priori*, for instance if there is no way to assess the model mismatch, one might need to compare a parallel structure with a serial structure recurring to data. For such, metrics are needed for model selection and to quantify the quality of the model under analysis (Hastie et al., 2016).

In any model development process, one should find the model complexity that provides the best compromise between bias and variance for a given purpose. Some common metrics establishing the bias-variance trade-off (Wilson and Sahinidis, 2017) were compiled by Bhosekar and Ierapetritou (2018), namely: the mean square error (MSE), mean absolute error (MAE), Bayesian information criterion (BIC), Akaike information criterion (AIC), Hannan-Quinn information criterion, risk inflation criterion and Mallow's Cp. However, many more are available in the literature, such as cross-validation metrics of the prediction error and the use of validation and independent test sets.

Schweidtmann and Mitsos (2019) presented a method for deterministic global optimization of optimization problems with ANNs embedded, where the problem is formulated in a reduced space and lower bounds are computed using McCormick relaxations propagated through the network equations (Mitsos et al., 2009). This methodology hides ANNs from the optimization problem, thus increasing computational efficiency. Even though the method was develop for optimization with data-driven models (ANNs), the authors believe it can be further extended to bring different data sources for ANN training. With some additional research, the method has the potential to be used in the selection and validation of a hybrid scheme, if the data-driven sub-model is an ANN. The objective function should be chosen among the metrics listed in the previous paragraph (Hastie et al., 2016).

Furthermore, Kahrs and Marquardt (2007) proposed two complementary methods for accessing the validity of a hybrid model: (i) the convex hull criterion and, (ii) the confidence interval criterion. The first one checks whether the input variables to all empirical model parts are within their respective validity domains (where the model was trained). The latter computes confidence intervals for the hybrid model predictions, quantifying the model's extrapolation uncertainty.

4. Applications

In this section, an overview is presented on the use of hybrid modeling in different application scenarios, including: process monitoring, control, optimization, model-plant mismatch, model transfer, automatic knowledge discovery, environmental science, predictive maintenance and sensor fusion. A summary of all the topics covered is presented in Table 4.

4.1. Process monitoring

It has been argued that hybrid methodologies can improve monitoring of batch processes through incorporation of external information. The main advantages are a better fault detection and diagnosis. Two types of information can be incorporated: batch-run specific (reactor conditions, cleaning & maintenance, reactor number, upstream conditions, batch duration) and process specific (dosing profiles, ideal trajectories, process stages, variable grouping, instrumental characteristics, spectra of pure components, kinetics). To incorporate the former, multiway regression to product quality can be used along with orthogonal signal correction to remove information in batch process data that are uncorrelated with

Table 4
Summary of hybrid modeling used in the different application scenarios.

References	Applications
Process monitoring	
(Yoon and MacGregor, 2001)	Monitoring process conditions of the dehydrogenation section of a styrene mononer plant
(Ramaker et al., 2002)	Prediction of the pH level in a batch reactor
(Vedam and Venkatasubramanian, 1999)	Root cause analysis of process faults
(Hassanpour et al., 2020)	Fault detection and identification
(Krippl et al., 2020)	Flux and duration analysis of ultrafiltration processes
(Bayer et al., 2020)	*Escherichia coli* fedatch cultivation characterization
Process control	
(Zhao et al., 2001)	Model predictive control with parallel hybrid model
(Xiong and Jutan, 2002)	Development and application of a general model-based control algorithm and its application to a simulated batch reactor and a real continuous stirred tank
(Chen et al., 2004)	Implementation of an online internal model control scheme for an industrial continuous reactive distillation column producing epichlorhydrin
(Su et al., 1992)	Model predictive control in a continuous copolymerization reactor where the reaction is initiated by a catalyst
Tsen et al. (1996)	Model predictive control scheme for batch processes using an augmented data hybrid ANN applied to the latex polymerization process
Sohlberg (2003)	Continuous-discrete model predictive controller for a heating process (the continuous model is used for time prediction and the minimization of a loss function is made in discrete time)
Bazaei and Majd (2003)	Control the pressure on a fermentor using an exact feedback linearization scheme
Ghosh et al. (2021)	Model predictive control scheme for seeded batch crystallization
Process optimization	
Henao and Maravelias (2011); Henao (2012); Pedrozo et al. (2020)	Use of surrogate models in superstructure optimization for process design
Fahmi and Cremaschi (2012)	Superstructure optimization to synthesize a biodiesel production plant
Graciano and Le Roux (2013)	Process synthesis and optimization of a water purification network
Boukouvala and Ierapetritou (2013)	Flowsheet optimization of continuous tablet manufacturing
Eason and Cremaschi (2014)	Optimization of carbon dioxide capture process with aqueous amines
Chu and You (2014); Shi and You (2015)	Planning, scheduling and dynamic optimization for sequential batch processes
Kumar et al. (2015, 2016)	Optimization of an industrial steam methane reformer
Pistikopoulos and Diangelakis (2016); Diangelakis et al. (2017)	Optimization and control of tank, CSTR, distillation column and CHP unit
Pattison et al. (2016a, 2016b, 2017)	Scheduling framework for a cryogenic air separation unit
Boukouvala et al. (2017)	General constrained optimization of a pressure swing adsorption unit
Beykal et al. (2018a)	Maximization of the net present value, over a five-year time horizon, of an oilfield using water-flooding
Tsay et al. (2019)	Power cost minimization over a four-day horizon in a cryogenic air separation unit
Wu et al. (2019a, 2019b, 2020)	MPC for closed loop stability in a CSTR with irreversible second order exothermic reaction
Schäfer et al. (2019b,a); Tsay and Baldea (2020)	Scheduling and process control integration in a cryogenic air separation unit
Santos et al. (2021)	MPC and hybrid RTO applied to the Williams-Otto reactor scheme benchmark
Feasibility analysis	
Boukouvala and Ierapetritou (2012)	Kriging-based feasibility analysis of black-box processes
Zhang et al. (2016)	Approximation of feasible region and cost function indicated by a set of data points
Analysis of model-plant mismatch	
Meneghetti et al. (2014)	Reducing model-plant mismatch in a jacket-cooled continuous stirred tank reactor and a solids milling unit
Kumar et al. (2018)	Reducing model-plant mismatch to improve fault detection in simulated production of acetoacetylation of pyrrole in batch, semi-batch and continuous stirred tank reactors
Hille et al. (2017)	Batch-to-batch optimization of a fed-batch penicillin process in the presence of model-plant mismatch and input uncertainty
Hille and Budman (2018)	Identification and optimization of penicillin production and chinese hamster ovary cell cultivation process under model-plant mismatch
Hille and Budman (2019)	Experimental design for Batch-to-batch optimization of the chinese hamster ovary cell cultivation process
Chen and Ierapetritou (2020)	Reduce model-pant mismatch to improve tracking of pharmaceutical unit operations
Model transfer	
Tomba et al. (2012)	Monitoring model transfer of a pharmaceutical spray-drying process from a pilot-scale unit to a production-scale unit
Automatic knowledge discovery	
Raissi et al. (2017, 2019)	Encoding of underlying physical laws (Korteweg-de Vries equation, Navier–Stokes equation and Schrodinger equation) that govern a given process from which data is available
Environmental science	
Fienen et al. (2013)	Forecasting mean depth to water in a groundwater model of Assateague Island in Virginia and Maryland, USA
Predictive maintenance	
Hanachi et al. (2019)	Prediction of wear in a cutting tool
Sensor fusion	
Safari et al. (2014)	Tracking problem with sensors that have different and asynchronous sampling rates

product quality. Also, a multiblock model can be built segmenting feedstock, process and quality data. Regarding process specific information, first-principles knowledge can be used to estimate key quality variables from measured ones (soft sensor). Additionally, hybrid approaches combining a simple mechanistic model with an ANN model for compensating mismatch are used for batch monitoring, but one needs to take into consideration that the sequential nature of the fitting implies that the overall fit is not necessary optimal in a least-squares sense; thus, an iterative fitting procedure should be conducted (Ramaker et al., 2002).

Yoon and MacGregor (2001) proposed a hybrid correlation model to support statistical process monitoring, that incorporates prior knowledge to achieve more robust fault detection and isolation. The information available includes known variable relations, independent manipulated variables, controller information and known process variations. Using orthogonal projections (or generalized least squares), the observation matrix is decomposed into a term that reflects prior knowledge and a term including the remaining effects. The hybrid correlation model achieves better performance than classic statistical process monitoring in a industrial case study (dehydrogenation section of a styrene monomer plant).

Process monitoring and diagnosis can also be improved by an automated framework for the interpretation of PCA-based measured variable contributions, using SDG (Vedam and Venkatasubramanian, 1999). A PCA model is built using historical data and its respective control charts are monitored. When a fault is detected, the contributions of each variable are evaluated against a threshold and those with higher values are signed in the SDG to uncover the root cause. The SDG is constructed recurring to physical models or expert knowledge. The framework was tested in a simulated example. The authors suggested the use of a probability density function approach developed by Rengaswamy (1995) where fuzzy clustering identifies the number of Gaussians and their centers, and the use of fuzzy signed digraphs to overcome some of the SDGs limitations (Vedam and Venkatasubramanian, 1999).

Sun et al. (2020) defined a comprehensive state space descriptive system which consists of stacked autoencoders to produce what the authors call, deep process features. These features are used to classify the system in different subsystems when different processing modes are present. Additionally, a parallel hybrid modeling framework is proposed to help monitor the cobalt removal process performed in a unit within zinc hydrometallurgy.

On bioprocesses, hybrid models have been used to predict flux evolution and duration of batch and fed batch ultrafiltration process states. In a serial scheme, an ANNs outputs were fed to a mass balance and achieved better performance than the model benchmark based on film theory. Results suggest that with further research the model can be used as a one-step-ahead predictor and, as such, can be considered a digital twin of the real process (Krippl et al., 2020). A similar hybrid scheme, where an ANN approximated reaction kinetics in a model of a fed batch cultivation of *Escherichia coli*, outperformed the state of the art response surface models on the prediction of products produced (Bayer et al., 2020).

Furthermore, data augmentation is also a form of hybrid modeling approach for fault detection in process monitoring. In a recent article, Rato et al. proposed the concept of First Principles Statistical Process Monitoring, which is based on a data augmentation approach based on structured simulations using mechanistic models for the “common causes” variation, solving the problem of excessive false alarms and promoting interpretation of results during fault diagnosis activities. This methodology was successfully applied to an industrial Surface Mount Technology (SMT) production line (Rato et al., 2020). In a different application, residuals between the mechanistic model and process data are added as variable to the historical data matrix to improve the PCA-MSPC fault detection capability in a heating ventilation and air conditioning systems. The authors compared their results with two benchmark approaches: residual analysis and traditional PCA-multivariate statistical process control model (MSPC) and found superior performance for the new method (Hassanpour et al., 2020).

4.2. Process control

Several authors have been trying to take advantage of hybrid models in process control applications. Zhao et al. (2001) developed a nonlinear industrial model predictive control (MPC) that uses a state space model in parallel with PLS and ANNs. Xiong and Jutan (2002) investigated a hybrid model based control strategy using the generic model control algorithm for a simulated exothermic batch reactor and a real-time continuous stirred tank. The model presents a parallel structure where the broad process dynamics are first modeled using an approximate mechanistic model and an ANN is used to compensate model mismatch. This methodology maximizes prior knowledge, reduces the ANN size, requires less training data and converges faster.

For modeling nonlinear dynamics of reaction systems, model reduction techniques together with some assumptions (such as neglecting fast reaction kinetics) can turn a nonlinear dynamic system into a linear one with parameters that can be estimated using a training dataset. This control-oriented hybrid model structure is tested with data from an industrial reactive distillation column and successfully implemented in a online internal model control scheme as a way of making a trade-off between model complexity and performance of the controller (Chen et al., 2004).

A hybrid model with a parallel structure was incorporated into an MPC structure. PLS regression was used for weights’ initialization (Su et al., 1992). On another application, an MPC scheme for batch processes using an augmented data hybrid ANN has been developed and applied to the latex polymerization process. Augmented data hybrid ANN is basically a surrogate model trained with data interpolated and extrapolated from a rigorous mechanistic model fitted to experimental measurements. The ANN is used to predict batch end time, molecular weight and dispersity of the final polymer product. The presented framework showed to be superior in terms of its correlative capability to other hybrid ANN approaches and to the theoretical model (Tsen et al., 1996). Furthermore, Sohlberg (2003) developed a hybrid model where the base model consists of a system of continuous nonlinear ordinary equations, where the unknown parameters are estimated from measured data, by optimizing of the likelihood function. Unmeasured variables are estimated by an extended Kalman filter. The procedure is included in a MPC. The author claims that a hybrid model is a suitable option when state variables and inputs are limited by hard constraints.

On discrete control schemes, exact feedback linearization provides desirable global performance but requires exact process knowledge. For partially known nonlinear processes this is not possible. Thus, Bazaei and Majd (2003) suggested the use of a serial neuro-gray-box model, formed by a mechanistic model and an ANN for modeling unknown dynamics (this option was found to be more accurate than a data-driven model while it took shorter training time and allowed that some variables or parameters of the plant, which have been fixed during model training, be allowed to change during the use of the model without need for re-training).

Wu et al. (2019a,b) proposed a Lyapunov MPC coupled with an ensemble of recurrent neural networks to predict system dynamics. The control structure presented several numerical challenges that were addressed by the authors. The controller is ap-

plied to maintain closed-loop stability in a well-mixed nonisothermal CSTR where an irreversible second order exothermic reaction takes place. In a subsequent manuscript, Wu et al. (2020) addressed ways to insert physical knowledge in data-driven models. Three different methods were proposed: the first is a parallel hybrid scheme where a recurrent neural network learns the error between process data and mechanistic model predictions; the second is a partially connected recurrent neural network where the network is modified to translate the causal relationships between load variables and state variables; the third, takes the form of a weight constraint recurrent neural network where constraints are added to the neural network as a vector of weights that represent the dynamic effects of load variables on state variables. The hybrid models are then integrated in an MPC formulation for regulating two well-mixed nonisothermal CSTRs in series, where an irreversible second order exothermic reaction takes place.

In a recent work, Ghosh et al. (2021) integrated hybrid modeling in closed loop with MPC. The mechanistic nonlinear model is built but since process data is available the authors enhance this model with a data-driven part. They used two hybrid approaches: a parallel scheme, where the data-driven model learns the residuals from the mechanistic model; and a second approach where the mechanistic nonlinear model was approximated by surrogate linear model and the data-driven model learned the residuals between the surrogate in the process. This strategy is applied to control a seeded batch crystallization. Santos et al. (2021) proposed an MPC coupled with real-time optimization where an unscented Kalman filter continually estimates model parameters using process data. The methodology was tested on the Williams-Otto reactor scheme benchmark (Williams and Otto, 1960).

4.3. Process optimization

Henao and Maravelias (2011) and Henao (2012) developed a methodology for surrogate modeling in a superstructure optimization environment. It uses mixed-integer linear programming to choose a set of dependent and independent variables and build a linear regression model or an ANN, for non-linear problems. These surrogate models are then used for simulation and optimization instead of the rigorous models and, in combination with the mechanistic models, they constitute a hybrid modeling approach to superstructure optimization. The goal of this methodology is superstructure optimization for process synthesis. Four different case studies are presented ranging in complexity from the continuous stirred tank reactor to a small unit with process reaction and separation.

The use of surrogate models for process synthesis is popular due to its increased efficiency. Boukouvala and Ierapetritou (2013) used Kriging-based optimization of computationally expensive flowsheet models in continuous tablet manufacturing. Expanding on the use of surrogate models for general constrained optimization problems, Kriging, quadratic functions or sigmoid functions, were adopted to approximate explicitly unknown equations such as partial differential equations for pressure swing adsorption (Boukouvala et al., 2017). ANNs or polynomial functions are used to approximate process units, and then integrated in optimization formulations that generally minimize net present sink or maximize net present value (Fahmi and Cremaschi, 2012; Graciano and Le Roux, 2013). Pedrozo et al. (2020) leaned on piecewise linear surrogate models for process synthesis. Data are drawn from Aspen Plus simulations and these models are in turn used in a superstructure to optimize an ethane-based ethylene plant design, maximizing the net present value.

The ALAMO is a methodology that constructively learns a basic surrogate model using design of experiments and an adaptive sampling technique, known as error maximization sampling (EMS), to update the training set and refine the model. This process incorporates a criterion for finding the trade-off between model accuracy and optimization flexibility (Cozad et al., 2014). The ALAMO is able to train data-driven models with its model-building methods using adaptive sampling. The software builds a low-complexity surrogate model using a best subset method optimized by a mixed-integer programming formulation. The goal is to build constrained regression models using a mixed-integer quadratic program (Cozad et al., 2015). The ALAMO has some extensions, including the use of derivative-free optimization and the aforementioned EMS that makes efficient use of small datasets outperforming even space-filling designs with larger training sets. Constrained regression forces the model to conform to physical and chemical constraints which allow for extrapolation beyond the training data domain, a characteristic that is important in optimization applications (Wilson and Sahinidis, 2017).

Beykal et al. (2018b) implemented a data-driven hybrid methodology that integrates surrogate model identification and deterministic global optimization through the ARGONAUT algorithm for global optimization of general constrained grey-box multi-objective optimization problems. The ARGONAUT is a hybrid optimization framework designed to globally optimize general constrained grey-box problems by identifying an appropriate surrogate representation in order to form a nonconvex NLP problem. The framework is an adaptive sampling and optimization method which iterates until it converges. The authors report results for three benchmark problems and a detailed case study of an energy market design problem for a commercial building. The ϵ-constant method was used to convert multi-objective optimization into series of single-objective optimization subproblems, while the p-ARGONAUT globally optimizes the resulting constrained grey-box problems. Results showed that p-ARGONAUT outperforms other available derivative-free optimization (DFO) algorithms by providing consistent feasible solutions for the energy systems design case study, involving numerous equality constants which is typically challenging for general DFO algorithms.

On another case study, the goal was to maximize the net present value (NPV) over a 5-year period by manipulating oil well pressures, constrained by water-cut (content in oil) limitations, water handling and storage capacities (Beykal et al., 2018a). These problems are described by highly nonlinear equations and some unknown constraints. The optimization of such problems was tackled through surrogate modeling and derivative-free optimization. Surrogate functions are assessed in two steps of this work: the reduction of the dimensionality of the water-flooding optimization problem via parameterization of the control domain, and the optimization of simulation-based grey-box problems within the ARGONAUT framework (Beykal et al., 2018a).

Integration of planning, scheduling and dynamic optimization for sequential batch processes creates a mixed-integer dynamic optimization (MIDO) problem which can be discretized into a large-scale mixed-integer nonlinear programing (MINLP) problem. As this is a computationally cumbersome problem, two methods were developed that efficiently separate the sub-problems and apply surrogate models to represent the linking functions (Chu and You, 2014; Shi and You, 2015). This approach reformulates the original MINLP into an MILP. It was applied to a multi-product batch processing plant with eight units leading to a 4-fold reduction in computational time. Hybrid models were also used by Zhao and You (2019) to achieve a good trade-off between optimality and robustness in operational decisions under uncertainty conditions.

A surrogate-based optmization of an industrial steam methane reformer was proposed by Kumar et al. (2015, 2016). They approximate the profile of burners which have an egg crate shape. This model aims to substitute rigorous mechanistic models such

as computational fluid dynamics, reducing the computational burden.

Pistikopoulos and Diangelakis (2016) developed the PARametric Optimization and Control (PAROC) methodology for receding horizon optimization and decision making. In this methodology, the process "high fidelity" dynamic model is approximated using system identification techniques and/or model reduction approaches. The methodology was applied to a tank, a CSTR, a distillation column and a combined heat and power (CHP) unit Diangelakis et al. (2017).

A scheduling framework to deal mainly with changing electricity prices was also developed, that consists of two stages. The first one is dedicated to the identification of critical process and quality variables, while in the second system identification is performed in order to get the low order dynamic models, called scale bridge models (i.e., models that connect the high resolution production time-scale with the low resolution scheduling time-scale (Baldea and Daoutidis, 2012; Baldea et al., 2015)). The methodology is applied to a cryogenic air separation unit (Cao et al., 2015; Pattison et al., 2016b). It was also further expanded to address moving horizon closed loop scheduling (Pattison et al., 2016a; 2017). Scale bridge models are also used as low order data driven models for closed loop process dynamics and can take the form of Hammerstein-Wiener models or finite step response models. These models can be linearized, turning an optimal production scheduling problem into a mixed integer linear programming optimization. This makes the optimization process more efficient (Kelley et al., 2018).

Schäfer et al. (2019b,a) proposed solving a dynamic optimization problem in an economic nonlinear MPC layer where the complex mechanistic differential algebraic equations are surrogated by ANNs arranged as a compartment model. This approach is tested on an air separation unit, for establishing the optimal operation in real time electricity markets.

To optimize and integrate production scheduling in an efficient way, Tsay et al. (2019) utilized scale bridge models that learn the schedule relevant dynamics of the industrial process using historical data. As in Kelley et al. (2018), models took the form of low order dynamic Hammerstein-Wiener models. The proposed scheduling framework achieves relevant power cost savings over a four-day horizon in a cryogenic air separation unit. More recently, Tsay and Baldea (2020) used autoencoders to reduce a high dimensional feature space to a low dimensional one, and find the subset of variables relevant to scheduling. This approach contrasted with previous works that relied solely on engineering knowledge (Pattison et al., 2016b; Tsay et al., 2018; 2019). Scale bridge models are built using the obtained latent variables to optimize scheduling in a cryogenic air separation unit. Furthermore, the autoencoders performance is compared with different mapping functions and PCA.

4.4. Feasibility analysis

Hybrid modeling can also be useful for performing feasibility analysis in high dimensional feature spaces. The scenario considered is the absence of a process model or the high cost of evaluation of the model. In these conditions, feasibility analysis is carried out using Kriging. Furthermore, Boukouvala and Ierapetritou (2012) developed an adaptive sampling strategy that minimizes sampling costs without sacrificing accuracy.

Zhang et al. (2016) formulated an optimization algorithm for generating piecewise linear surrogate models able to approximate the feasible region and the cost function indicated by a given set of data points. These surrogates are designated convex region surrogates and are formulated as set of mixed integer linear constraints that can be embedded in scheduling optimization. The methodology is applied in an Praxair case study where product space and power consumption are the cost function. Notwithstanding the methodology's good performance, the authors identify computational efficiency as a weak point and recommend further research.

4.5. Analysis of model-plant mismatch

Meneghetti et al. (2014) proposed a methodology to identify model-plant mismatch that may arise when a mechanistic model is compared with historical process data. Auxiliary variables were defined as nonlinear combinations of the model variables, parameters and process variables. These auxiliary variables were arranged in two matrices, one for the model and another for the process, and their correlation structure was compared with PCA. Results from two simulated case studies show that the methodology generally succeeds in diagnosing the root cause of the mismatch, provided that there is enough engineering know-how available. The authors suggest further investigation to define appropriate confidence limits for the residual distribution and to generalize the methodology for a wider range of mismatch cases.

A three step methodology to model-plant mismatch identification and model update for reaction systems using the reaction extents was developed by Kumar et al. (2018). It formulates a model-plant mismatch problem as a fault detection and diagnosis problem, making use of extended Kalman filters and fault interaction graphs. Batch optimization and design frameworks under model-plant mismatch conditions have also been reported in the literature (Hille et al., 2017; Hille and Budman, 2018; 2019).

Furthermore, partial correlation coefficient analysis was used to identify mechanistic model mismatch against plant data. Once the source of the mismatch has been identified, it is replaced with appropriate data-driven models. The methodology was applied to two pharmaceutical case studies where three different hybrid schemes are tested (serial, parallel and both combined). The black box models used are artificial neural networks and support vector regression (Chen and Ierapetritou, 2020).

4.6. Model transfer

Tomba et al. (2012) proposed an approach for transferring process monitoring models between plants that combines fundamental knowledge with latent variable techniques. The model transfer solution finds applications in scale-up or startup of a new process plant or unit. In this framework, common variables must be identified. Common variables are measured quantities that have the same physical meaning in both plants. With these variables and a mechanistic model, one can derive a plant-independent factor, since the underlying physical phenomena are the same. Then by concatenating matching data from the two plants, one can build a standard PCA-MSPC to monitor process faults. The authors also proposed the use of Joint Y-PLS, where all variables can be used and not just the common ones. The two methods were applied to a simulated spray-drying process.

4.7. Automatic knowledge discovery

Hybrid modeling has also been used in a deep learning framework to address two classes of problems: solving and encoding PDEs from available data. The method uses ANNs to attain these goals while respecting fundamental laws. Its need is justified by environments where data acquisition is cost-prohibitive and, therefore, state-of-art machine learning techniques lack the necessary robustness. On the other hand, there are considerable amounts of prior knowledge about these systems that can improve the model and its extrapolation capability. It is argued that Gaussian processes lack robustness and prediction accuracy in strongly non-

linear problems, despite their flexibility and elegance in encoding prior knowledge. The core technology of the proposed framework is automatic differentiation. The cost function to be minimized is a combination of the model error and the constraint error (Raissi et al., 2017; 2019).

4.8. Environmental science

In environmental research, specially in what concerns to climate change, it is important that the model used be able to represent causality, something that passively collected data alone (also know as "observational data") cannot guarantee (Pearl and Mackenzie, 2018) (the same requirements are found in process control and optimization). In designing, a decision support system for natural resource management in future environmental conditions, namely sea level rise, the models involve a certain level of uncertainty and propagating that uncertainty across the entire system raises some challenges. Fienen et al. (2013) propose the use of surrogate groundwater models using Bayesian networks, that are quite suitable for rigorous uncertainty propagation. To avoid overfitting k-fold cross-validation is used to train the Bayesian network.

4.9. Predictive maintenance

The management of modern manufacturing process are increasingly based on the use of information collected in real time from industrial processes to monitor machine and tool condition. Physics-based predictive models for maintenance can include large uncertainties, and pure data-diven (or sensor-based) inference techniques are affected by sensor noise and measurement model uncertainties. To work around these limitations, an hybrid data-driven physics-based model fusion framework was proposed, that fuses the outputs of the physics-based model and of the measurement-based model (adaptive neuro-fuzzy inference systems) using a particle filter. Results show significant improvements when compared to each of the submodels alone, namely the prediction errors were reduced to approximately half) (Hanachi et al., 2019).

4.10. Sensor fusion

An asynchronous multirate multisensor data fusion linear system was studied to handle the prevalent issue of asynchronous sampling rates. Parallel Kalman filters were used for state estimation while an ANN fused the Kalman filters outputs. Simulation results have shown that this scheme is better than the use of federated Kalman filters (See Carlson, 1990; Safari et al., 2014).

5. Conclusions and future challenges

From terminology and definitions, through modeling methodologies, structure, identification, assessment and applications, the contribution of scientific publications to hybrid modeling in the last 28 years were briefly reviewed. The following ten bullet points refer some of the main facts and trends identified in this review work:

(i) Hybrid modeling caught the interest of the scientific community for almost 30 years now and their advantages have been demonstrated and tested;
(ii) The amount of publications and the various backgrounds from researchers led to the existence of different terminologies and definitions, which we have addressed and clarified;
(iii) Hybrid modeling approaches have been systematically benchmarked against pure data-driven or pure mechanistic and shown to consistently achieve better performance;
(iv) Prior knowledge, although dominated by equations, has not been fully incorporated in all its configurations, leaving room to explore alternative ways to incorporate plant floor experience and process flowsheets;
(v) Data-driven models have been used to improve mechanistic models, but the opposite path remains vastly underexplored;
(vi) Solving engineering problems in model-plant mismatch conditions is still an open problem where hybrid modeling may find good application;
(vii) Multimode process modeling can benefit from the incorporation of prior knowledge, because some of the differences between modes are known to process operators;
(viii) Batch process monitoring is currently based on data-driven statistical models, but some works point out the advantages of incorporating prior knowledge to improve the performance of batch tracking;
(ix) Fault detection and diagnosis benefit of the introduction of causality brought by process knowledge;
(x) There is a need for the development of software tools that facilitate the incorporation of the various sources of knowledge in a holistic to come up with the best hybrid model for a given application.

In this article, we have identified and systematized the state of the art approaches on hybrid modeling, the various applications where they can be applied and the existing gaps in scientific knowledge. From the extensive analysis of all the contributions considered and the current data intensive context, we believe to be both important and opportune to develop a platform that could be deployed as an accessible software tool for hybrid modeling development, fulfilling the necessities identified in chemical process industry applications.

While moving forward, it is interesting to remember and acknowledge the words written more than 45 years ago, by Wold (1974), and recognize that integrating deductive and inductive knowledge has been a central concern in science, and will remain so:

> This is not a dualism, but rather a wide range of diversifications, with many intermediate cases, and with several partings of the ways at fundamental level. Speaking generally, a model gains in operative use and efficiency if prior information is available and is incorporated in the model. On the other hand the model can be overambitious in using information that is not available or is mainly hypothetical.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

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. Chem. Eng. Sc. 6 10/c9mph7.
Anderson, C., 2008. The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. Wired.
Andersson, M., 2009. A comparison of nine PLS1 algorithms. J. Chemom. 23 (10), 518–529 10/bq2pkh.
Asprey, S.P., Macchietto, S., 2000. Statistical tools for optimal dynamic model building. Comput. Chem. Eng. 24 (2), 1261–1267. doi:10.1016/S0098-1354(00)00328-8.
Asprey, S.P., Macchietto, S., 2002. Designing robust optimal dynamic experiments. J. Process Control 12 (4), 545–556. doi:10.1016/S0959-1524(01)00020-8.
Asprion, N., 2020. Modeling, simulation, and optimization 4.0 for a distillation column. Chem. Ingenieur Tech. 92 (7), 879–889 10/gg9jgz.
Asprion, N., Böttcher, R., Pack, R., Stavrou, M.-E., Höller, J., Schwientek, J., Bortz, M., 2019. Gray-box modeling for the optimization of chemical processes. Chem. Ingenieur Tech. 91 (3), 305–313 10/gf5twq.

Atkinson, A.C., 2008. DT-optimum designs for model discrimination and parameter estimation. J. Stat. Plan. Inference 138 (1), 56–64. doi:10.1016/j.jspi.2007.05.024.

Atkinson, A.C., Bailey, R.A., 2001. One hundred years of the design of experiments on and off the pages of Biometrika. Biometrika 88 (1), 53–97. doi:10.1093/biomet/88.1.53.

Atkinson, A.C., Bogacka, B., 2002. Compound and other optimum designs for systems of nonlinear differential equations arising in chemical kinetics. Chemom. Intell. Lab. Syst. 61 (1), 17–33. doi:10.1016/S0169-7439(01)00173-3.

Atkinson, A.C., Donev, A.N., 1992. Optimum Experimental Designs. Oxford University Press, Oxford, New York.

Azarpour, A., N.G. Borhani, T., R. Wan Alwi, S., A. Manan, Z., I. Abdul Mutalib, M., 2017. A generic hybrid model development for process analysis of industrial fixed-bed catalytic reactors. Chem. Eng. Res. Des. 117, 149–167. 10/f9qc8d.

Baldea, M., Daoutidis, P., 2012. Dynamics and Nonlinear Control of Integrated Process Systems. Cambridge University Press.

Baldea, M., Du, J., Park, J., Harjunkoski, I., 2015. Integrated production scheduling and model predictive control of continuous processes. AIChE J. 61 (12), 4179–4190. 10.1002/aic.14951.

Bayer, B., von Stosch, M., Striedner, G., Duerkop, M., 2020. Comparison of modeling methods for DoE-based holistic upstream process characterization. Biotechnol. J. 15 (5), 1900551. doi:10.1002/biot.201900551.

Bazaei, A., Majd, V.J., 2003. Feedback linearization of discrete-time nonlinear uncertain plants via first-principles-based serial neuro-gray-box models. J. Process Control 13 (8), 819–830. 10/fd36md.

Bequette, B.W., 1998. Process Dynamics: Modeling, Analysis and Simulation, first ed. Prentice Hall, Upper Saddle River, N.J.

Beykal, B., Boukouvala, F., Floudas, C.A., Pistikopoulos, E.N., 2018. Optimal design of energy systems using constrained grey-box multi-objective optimization. Comput. Chem. Eng. 116, 488–502. 10/gfmqgr.

Beykal, B., Boukouvala, F., Floudas, C.A., Sorek, N., Zalavadia, H., Gildin, E., 2018. Global optimization of grey-box computational systems using surrogate functions and application to highly constrained oil-field operations. Comput. Chem. Eng. 114, 99–110. 10/gf4kg3.

Bhosekar, A., Ierapetritou, M., 2018. Advances in surrogate based modeling, feasibility analysis, and optimization: a review. Comput. Chem. Eng. 108, 250–267. 10/gfxhv7.

Bikmukhametov, T., Jäschke, J., 2020. Combining machine learning and process engineering physics towards enhanced accuracy and explainability of data-driven models. Comput. Chem. Eng. 138, 106834. doi:10.1016/j.compchemeng.2020.106834.

Bird, R.B., Stewart, W.E., Lightfoot, E.N., 1960. Transport Phenomena, 1st John Wiley & Sons, New York.

Bishop, C., 2006. Pattern Recognition and Machine Learning. Springer-Verlag, New York.

Bohács, G., Rinkács, A., 2017. Development of an ontology-driven, component based framework for the implementation of adaptiveness in a Jellyfish-Type simulation model. J. Ambient Intell. Smart Environ. 9 (3), 361–374. 10/f98psg.

Bollas, G., Papadokonstadakis, S., Michalopoulos, J., Arampatzis, G., Lappas, A., Vasalos, I., Lygeros, A., 2003. Using hybrid neural networks in scaling up an FCC model from a pilot plant to an industrial unit. Chem. Eng. Process. 42 (8–9), 697–713. 10/bhfp5w.

Bonvin, D., Georgakis, C., Pantelides, C.C., Barolo, M., Grover, M.A., Rodrigues, D., Schneider, R., Dochain, D., 2016. Linking models and experiments. Ind. Eng. Chem. Res. 55 (25), 6891–6903. 10/gc3hn2.

Boukouvala, F., Floudas, C.A., 2017. argonaut: algorithms for global optimization of constrained grey-box computational problems. Optim. Lett. 11 (5), 895–913. doi:10.1007/s11590-016-1028-2.

Boukouvala, F., Hasan, M.M.F., Floudas, C.A., 2017. Global optimization of general constrained grey-box models: New method and its application to constrained PDEs for pressure swing adsorption. J. Glob. Optim. 67 (1–2), 3–42. doi:10.1007/s10898-015-0376-2.

Boukouvala, F., Ierapetritou, M.G., 2012. Feasibility analysis of black-box processes using an adaptive sampling Kriging-based method. Comput. Chem. Eng. 36, 358–368. doi:10.1016/j.compchemeng.2011.06.005.

Boukouvala, F., Ierapetritou, M.G., 2013. Surrogate-based optimization of expensive flowsheet modeling for continuous pharmaceutical manufacturing. J. Pharm. Innov. 8 (2), 131–145.

Box, G.E.P., Draper, N.R., 2007. Response Surfaces, Mixtures, and Ridge Analyses. John Wiley & Sons.

Box, G.E.P., Hunter, J.S., Hunter, W.G., 2005. Statistics for Experimenters: Design, Innovation, and Discovery, second ed. Wiley-Interscience, Hoboken, N.J.

Brendel, M., Marquardt, W., 2008. Experimental design for the identification of hybrid reaction models from transient data. Chem. Eng. J. 141 (1-3), 264–277. 10/dnr2s5.

Bro, R., Smilde, A.K., 2014. Principal component analysis. Anal. Methods 6 (9), 2812–2831. 10/gcx742.

Burden, F., Winkler, D., 2009. Bayesian regularization of neural networks. In: Livingstone, D.J. (Ed.), Artificial Neural Networks: Methods and Applications. Humana Press, Totowa, NJ, pp. 23–42. doi:10.1007/978-1-60327-101-1_3.

Cameron, I.T., Hangos, K., Perkins, J., Stephanopoulos, G., 2001. Process Modelling and Model Analysis, Volume 4, first ed. Academic Press, San Diego.

Cao, Y., Swartz, C.L.E., Baldea, M., Blouin, S., 2015. Optimization-based assessment of design limitations to air separation plant agility in demand response scenarios. J. Process Control 33, 37–48. doi:10.1016/j.jprocont.2015.05.002.

Cappuyns, A.M., Bernaerts, K., Smets, I.Y., Ona, O., Prinsen, E., Vanderleyden, J., Van Impe, J.F., 2006. Optimal experiment design in bioprocess modeling: from theory to practice. IFAC Proc. Vol. 39 (2), 535–540. doi:10.3182/20060402-4-BR-2902.00535.

Carinhas, N., Bernal, V., Teixeira, A.P., Carrondo, M.J., Alves, P.M., Oliveira, R., 2011. Hybrid metabolic flux analysis: Combining stoichiometric and statistical constraints to model the formation of complex recombinant products. BMC Syst. Biol. 5 (1), 34. 10/b5q9f9.

Carlson, N.A., 1990. Federated square root filter for decentralized parallel processors. IEEE Trans. Aerosp. Electron.Syst. 26 (3), 517–525. 10/cpw6pz.

Chen, L., Hontoir, Y., Huang, D., Zhang, J., Morris, A., 2004. Combining first principles with black-box techniques for reaction systems. Control Eng. Pract. 12 (7), 819–826. 10/bm8ss9.

Chen, Y., Ierapetritou, M., 2020. A framework of hybrid model development with identification of plant-model mismatch. AIChE J. n/a (n/a), e16996. 10/gg9jj4.

Cheng, H., Nikus, M., Jämsä-Jounela, S.-L., 2008. Fault diagnosis of the paper machine short circulation process using novel dynamic causal digraph reasoning. J. Process Control 18 (7-8), 676–691. 10/c4zr2m.

Chiang, L.H., Jiang, B., Zhu, X., Huang, D., Braatz, R.D., 2015. Diagnosis of multiple and unknown faults using the causal map and multivariate statistics. J. Process Control 28, 27–39. 10/f697b4.

Chu, Y., You, F., 2014. Integrated planning, scheduling, and dynamic optimization for batch processes: minlp model formulation and efficient solution methods via surrogate modeling. Ind. Eng. Chem. Res. 53 (34), 13391–13411. doi:10.1021/ie501986d.

Cozad, A., Sahinidis, N.V., Miller, D.C., 2014. Learning surrogate models for simulation-based optimization. AIChE J. 60 (6), 2211–2227. 10/f52rqm.

Cozad, A., Sahinidis, N.V., Miller, D.C., 2015. A combined first-principles and data–driven approach to model building. Comput. Chem. Eng. 73, 116–127. 10/f63nfk.

de Jong, S., 1993. SIMPLS: an alternative approach to partial least squares regression. Chemom. Intell. Lab. Syst. 18 (3), 251–263. 10/dmk73j.

Destro, F., Facco, P., Muñoz, S.G., Bezzo, F., Barolo, M., 2020. A hybrid framework for process monitoring: Enhancing data-driven methodologies with state and parameter estimation. J. Process Control 92, 333–351. 10/gg9jg2.

Diangelakis, N.A., Burnak, B., Katz, J., Pistikopoulos, E.N., 2017. Process design and control optimization: asimultaneous approach by multi-parametric programming. AIChE J. 63 (11), 4827–4846. 10.1002/aic.15825.

Dors, M., Simutis, R., Lübbert, A., 1996. Hybrid process modeling for advanced process state estimation, prediction, and control exemplified in a production-scale mammalian cell culture. In: Biosensor and Chemical Sensor Technology. In: ACS Symposium Series, 613. American Chemical Society, pp. 144–154. doi:10.1021/bk-1995-0613.ch014.

Duarte, B., Saraiva, P.M., 2001. Hybrid modeling through the combination of mechanistic models and MARS. IFAC Proc. Vol. 34 (25), 699–704. 10/gfxhv8.

Duarte, B., Saraiva, P.M., Pantelides, C.C., 2004. Combined Mechanistic and Empirical Modelling. Int. J. Chem. React.Eng. 2 (1). 10/fbx6mz.

Duarte, B.P.M., Saraiva, P.M., 2003. Hybrid Models Combining Mechanistic Models with Adaptive Regression Splines and Local Stepwise Regression. Ind. Eng. Chem. Res. 42 (1), 99–107. 10/drp9tk.

Dulmage, A., Mendelsohn, N., 1963. Two algorithms for bipartite graphs. J. Soc. Ind. Appl.Math. 11 (1), 183–194. 10/c46pgx.

Eason, J., Cremaschi, S., 2014. Adaptive sequential sampling for surrogate model generation with artificial neural networks. Comput. Chem. Eng. 68, 220–232. doi:10.1016/j.compchemeng.2014.05.021.

Fahmi, I., Cremaschi, S., 2012. Process synthesis of biodiesel production plant using artificial neural networks as the surrogate models. Comput. Chem. Eng. 46, 105–123. doi:10.1016/j.compchemeng.2012.06.006.

Fienen, M.N., Masterson, J.P., Plant, N.G., Gutierrez, B.T., Thieler, E.R., 2013. Bridging groundwater models and decision support with a Bayesian network: Bayesian network model emulation. Water Resour. Res. 49 (10), 6459–6473. 10/f5kxdf.

Friedman, J.H., 1991. Multivariate adaptive regression splines. Ann. Stat. 19 (1), 1–67. 10/ctgh88.

Galvanauskas, V., Simutis, R., Lübbert, A., 2004. Hybrid process models for process optimisation, monitoring and control. Bioprocess Biosyst. Eng. 26 (6), 393–400. 10/cbxt8p.

Garcia, H.E., Vilim, R.B., 1998. Combining physical modeling, neural processing, and likelihood testing for online process monitoring. In: SMC'98 Conference Proceedings. 1998 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No.98CH36218), 1, pp. 806–810vol.1. 10/c94mzk.

Geladi, P., 1988. Notes on the history and nature of partial least squares (PLS) modelling. J. Chemom. 2 (4), 231–246. 10/b5gqvk.

Georgakis, C., 2013. Design of dynamic experiments: a data-driven methodology for the optimization of time-varying processes. Ind. Eng. Chem. Res. 52 (35), 12369–12382. doi:10.1021/ie3035114.

Georgieva, P., Meireles, M., Feyo de Azevedo, S., 2003. Knowledge-based hybrid modelling of a batch crystallisation when accounting for nucleation, growth and agglomeration phenomena. Chem. Eng. Sc. 58 (16), 3699–3713. 10/c9wnw7.

Ghosh, D., Hermonat, E., Mhaskar, P., Snowling, S., Goel, R., 2019. Hybrid modeling approach integrating first-principles models with subspace identification. Ind. Eng. Chem. Res. 58 (30), 13533–13543. 10/gf5qmv.

Ghosh, D., Moreira, J., Mhaskar, P., 2021. Model predictive control embedding a parallel hybrid modeling strategy. Ind. Eng. Chem. Res. 60 (6), 2547–2562. doi:10.1021/acs.iecr.0c05208.

Glassey, J., von Stosch, M. (Eds.), 2018, first ed. CRC Press, Milton Hybrid Modeling in Process Industries.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press.

Graciano, J.E.A., Le Roux, G.A.C., 2013. Improvements in surrogate models for process synthesis. Application to water network system design. Comput. Chem. Eng. 59, 197–210. doi:10.1016/j.compchemeng.2013.05.024.

Greeff, D., Aldrich, C., 1998. Empirical modelling of chemical process systems with evolutionary programming. Comput. Chem. Eng. 22 (7-8), 995–1005. 10/b58qkm.

Gurden, S.P., Westerhuis, J.A., Bijlsma, S., Smilde, A.K., 2001. Modelling of spectroscopic batch process data using grey models to incorporate external information. J. Chemom. 15 (2), 101–121. 10/d4bvxk.

Hanachi, H., Yu, W., Kim, I.Y., Liu, J., Mechefske, C.K., 2019. Hybrid data-driven physics-based model fusion framework for tool wear prediction. Int. J. Adv. Manuf. Technol. 101 (9–12), 2861–2872. 10/gf4kfs.

Hassanpour, H., Mhaskar, P., House, J.M., Salsbury, T.I., 2020. A hybrid modeling approach integrating first-principles knowledge with statistical methods for fault detection in HVAC systems. Comput. Chem. Eng. 142, 107022. 10/gg9jkd.

Hastie, T., Tibshirani, R., Friedman, J., 2016. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, second ed. Springer, New York, NY.

He, B., Chen, T., Yang, X., 2014. Root cause analysis in multivariate statistical process monitoring: Integrating reconstruction-based multivariate contribution analysis with fuzzy-signed directed graphs. Comput. Chem. Eng. 64, 167–177. 10/f5w6bv.

Hecht-Nielsen, R., 1992. III.3 – theory of the backpropagation neural network. In: Wechsler, H. (Ed.), Neural Networks for Perception. Academic Press, pp. 65–93. doi:10.1016/B978-0-12-741252-8.50010-8.

Henao, C.A., 2012. A Superstructure Modeling Framework for Process Synthesis Using Surrogate Models. University of Wisconsin-Madison.

Henao, C.A., Maravelias, C.T., 2011. Surrogate-based superstructure optimization framework. AIChE J. 57 (5), 1216–1232. doi:10.1002/aic.12341.

Henneke, D., Hagedorn, A., Budman, H.M., Legge, R.L., 2005. Application of spectrofluorometry to the prediction of PHB concentrations in a fed-batch process. Bioprocess Biosyst. Eng. 27 (6), 359–364. 10/dpcdms.

Hille, R., Budman, H.M., 2018. Simultaneous identification and optimization of biochemical processes under model-plant mismatch using output uncertainty bounds. Comput. Chem. Eng. 113, 125–138. 10/gdd36t.

Hille, R., Budman, H.M., 2019. Experimental Design for Batch-to-Batch Optimization under Model-Plant Mismatch. Ind. Eng. Chem. Res. 58 (30), 13599–13610. 10/gf7fj5.

Hille, R., Mandur, J., Budman, H.M., 2017. Robust batch-to-batch optimization in the presence of model-plant mismatch and input uncertainty. AIChE J. 63 (7), 2660–2670. 10/gbjqnc.

Hirschen, K., Schäfer, M., 2006. Bayesian regularization neural networks for optimizing fluid flow processes. Comput. Methods Appl. Mech. Eng. 195 (7), 481–500. doi:10.1016/j.cma.2005.01.015.

Jackson, J.E., 1991. A User's Guide To Principal Components, first ed. John Wiley & Sons, Inc, New York, USA.

Jin, R., Chen, W., Simpson, T., 2001. Comparative studies of metamodelling techniques under multiple modelling criteria. Struct. Multidiscip. Optim. 23 (1), 1–13. 10/fm4gp8.

Johansen, T.A., Foss, B.A., 1992. Representing and learning unmodeled dynamics with neural network memories. In: 1992 American Control Conference, pp. 3037–3043. 10/gf5bv7.

Jollife, I.T., 2002. Principal Component Analysis, second ed. Springer, New York, USA.

Kahrs, O., Brendel, M., Michalik, C., Marquardt, W., 2009. Incremental identification of hybrid models of dynamic process systems. In: Hof, P.M., Scherer, C., Heuberger, P.S. (Eds.), Model-Based Control: Bridging Rigorous Theory and Advanced Technology. Springer US, Boston, MA, pp. 185–202. doi:10.1007/978-1-4419-0895-7_11.

Kahrs, O., Marquardt, W., 2007. The validity domain of hybrid models and its application in process optimization. Chem. Eng. Process. 46 (11), 1054–1066. 10/fqbpcj.

Kahrs, O., Marquardt, W., 2008. Incremental identification of hybrid process models. Comput. Chem. Eng. 32 (4-5), 694–705. 10/c4zbrz.

Karpatne, A., Atluri, G., Faghmous, J.H., Steinbach, M., Banerjee, A., Ganguly, A., Shekhar, S., Samatova, N., Kumar, V., 2017. Theory-guided data science: a new paradigm for scientific discovery from data. IEEE Trans. Knowl. Data Eng. 29 (10), 2318–2331. 10/gbzdt4.

Karpatne, A., Watkins, W., Read, J., Kumar, V., 2017. Physics-guided Neural Networks (PGNN): an application in lake temperature modeling. Phys. Stat. arXiv:1710.11431.

Kelley, M.T., Pattison, R.C., Baldick, R., Baldea, M., 2018. An efficient MILP framework for integrating nonlinear process dynamics and control in optimal production scheduling calculations. Comput. Chem. Eng. 110, 35–52. doi:10.1016/j.compchemeng.2017.11.021.

Kosko, B., 1992. Neural Networks and Fuzzy Systems: a Dynamical Systems Approach to Machine Intelligence, first ed. Prentice-Hall, Englewood Cliffs, New Jersey.

Kourti, T., MacGregor, J., 1995. Process analysis, monitoring and diagnosis, using multivariate projection methods. Chemom. Intell. Lab. Syst. 28, 3–21. 10/b45kmt.

Kramer, M.A., Thompson, M.L., Bhagat, P.M., 1992. Embedding theoretical models in neural networks. In: 1992 American Control Conference, pp. 475–479. 10/gf5bv9.

Kresta, J.V., Macgregor, J.F., Marlin, T.E., 1991. Multivariate statistical monitoring of process operating performance. Can. J. Chem. Eng. 69, 35–47. 10/dpcjn4.

Krippl, M., Dürauer, A., Duerkop, M., 2020. Hybrid modeling of cross-flow filtration: predicting the flux evolution and duration of ultrafiltration processes. Sep. Purif. Technol. 248, 117064. doi:10.1016/j.seppur.2020.117064.

Kumar, A., Baldea, M., Edgar, T.F., 2016. Real-time optimization of an industrial steam-methane reformer under distributed sensing. Control Eng. Pract. 54, 140–153. doi:10.1016/j.conengprac.2016.05.010.

Kumar, A., Baldea, M., Edgar, T.F., Ezekoye, O.A., 2015. Smart manufacturing approach for efficient operation of industrial steam-methane reformers. Ind. Eng. Chem. Res. 54 (16), 4360–4370. doi:10.1021/ie504087z.

Kumar, D.D., Narasimhan, S., Bhatt, N., 2018. Detection of model-plant mismatch and model update for reaction systems using concept of extents. J. Process Control 72, 17–29. 10/gf7fj7.

Kumar, P., Potluri, C., Sebastian, A., Chiu, S., Urfer, A., Naidu, D.S., Schoen, M.P., 2010. An adaptive multi sensor data fusion with hybrid nonlinear ARX and Wiener-Hammerstein models for skeletal muscle force estimation. In: The 14th World Scientific and Engineering Academy and Society (WSEAS) International Conference on Systems, Corfu Island, Greece.

Lasi, H., Fettke, P., Kemper, H.-G., Feld, T., Hoffmann, M., 2014. Industrie 4.0. Wirtschaftsinformatik 56 (4), 261–264. doi:10.1007/s11576-014-0424-4.

Lee, D.S., Vanrolleghem, P.A., Park, J.M., 2005. Parallel hybrid modeling methods for a full-scale cokes wastewater treatment plant. J. Biotechnol. 115 (3), 317–328. 10/bnp8n6.

Lima, P.V., Saraiva, P.M., 2007. A semi-mechanistic model building framework based on selective and localized model extensions. Comput. Chem. Eng. 31 (4), 361–373. 10/dk46vv.

Lundberg, B., Sjöblom, J., Johansson, Å., Westerberg, B., Creaser, D., 2015. Model-based experimental screening for DOC parameter estimation. Comput. Chem. Eng. 74, 144–157. doi:10.1016/j.compchemeng.2015.01.004.

Luo, L., Bao, S., 2018. Knowledge-data-integrated sparse modeling for batch process monitoring. Chem. Eng. Sc. 189, 221–232. 10/gftv3c.

MacGregor, J.F., Jaeckle, C., Kiparissides, C., Koutoudi, M., 1994. Process monitoring and diagnosis by multiblock PLS methods. AIChE J. 40 (5), 826–838. 10/bv9bc4.

MacGregor, J.F., Kourti, T., 1995. Statistical process control of multivariate processes. Control Eng. Pract. 3 (3), 403–414. 10/bg9znw.

Martens, H., Næs, T., 1992. Multivariate Calibration, first ed. Wiley, Chichester.

Martinez, E.C., Wilson, J.A., 1998. A hybrid neural network-first principles approach to batch unit optimisation. Comput. Chem. Eng. 22, S893–S896. 10/bftk8x.

McBride, K., Sundmacher, K., 2019. Overview of surrogate modeling in chemical process engineering. Chem. Ingenieur Tech. 91 (3), 228–239. 10/gg972f.

Meert, K., Rijckaert, M., 1998. Intelligent modelling in the chemical process industry with neural networks: a case study. Comput. Chem. Eng. 22, S587–S593. 10/cmd3s4.

Meneghetti, N., Facco, P., Bezzo, F., Barolo, M., 2014. A methodology to diagnose process/model mismatch in first-principles models. Ind. Eng. Chem. Res. 53 (36), 14002–14013. 10/f6hsgn.

Mitsos, A., Chachuat, B., Barton, P.I., 2009. mccormick-based relaxations of algorithms. SIAM J. Optim. 20 (2), 573–601. doi:10.1137/080717341.

Montgomery, D.C., 2012. Design and Analysis of Experiments, 8th edition international student version Wiley, Singapore.

Nascimento, C.A.O., Giudici, R., Scherbakoff, N.L., 1999. Modeling of industrial nylon-6,6 polymerization process in a twin-screw extruder reactor. II. Neural networks and hybrid models. J. Appl. Polym. Sci. 72, 905–912. 10/bbx8sk.

Oliveira, R., 1998. Supervision, Control and Optimization of Biotechnological Processes Based on Hybrid Models. Ph. D. Dissertation, Martin-Luther Universität, Halle-Wittenberg, Germany.

Oliveira, R., 2004. Combining first principles modelling and artificial neural networks: a general framework. Comput. Chem. Eng. 28 (5), 755–766. 10/bsqhf5.

Papoulias, S.A., Grossmann, I.E., 1983. A structural optimization approach in process synthesis—I. Comput. Chem. Eng. 7 (6), 695–706. 10/bs9w4m.

Papoulias, S.A., Grossmann, I.E., 1983. A structural optimization approach in process synthesis—II. Comput. Chem. Eng. 7 (6), 707–721. 10/drfwbv.

Papoulias, S.A., Grossmann, I.E., 1983. A structural optimization approach in process synthesis—III. Comput. Chem. Eng. 7 (6), 723–734. 10/b9qh95.

Park, S., Ramirez, W.F., 1988. Optimal production of secreted protein in fed-batch reactors. AIChE J. 34 (9), 1550–1558. 10/fq5b8t.

Pattison, R., Touretzky, C.R., Johansson, T., Baldea, M., Harjunkoski, I., 2016. Moving horizon scheduling of an air separation unit under fast-changing energy prices. IFAC-PapersOnLine 49 (7), 681–686. doi:10.1016/j.ifacol.2016.07.257.

Pattison, R.C., Touretzky, C.R., Harjunkoski, I., Baldea, M., 2017. Moving horizon closed-loop production scheduling using dynamic process models. AIChE J. 63 (2), 639–651. doi:10.1002/aic.15408.

Pattison, R.C., Touretzky, C.R., Johansson, T., Harjunkoski, I., Baldea, M., 2016. Optimal process operations in fast-changing electricity markets: framework for scheduling with low-order dynamic models and an air separation application. Ind. Eng. Chem. Res. 55 (16), 4562–4584. doi:10.1021/acs.iecr.5b03499.

Pearl, J., Mackenzie, D., 2018. The Book of Why: The New Science of Cause and Effect, first ed. Basic Books, New York.

Pedrozo, H.A., Rodriguez Reartes, S.B., Chen, Q., Diaz, M.S., Grossmann, I.E., 2020. Surrogate-model based MILP for the optimal design of ethylene production from shale gas. Comput. Chem. Eng. 141, 107015. doi:10.1016/j.compchemeng.2020.107015.

Peres, J., Oliveira, R., Feyo de Azevedo, S., 2001. Knowledge based modular networks for process modelling and control. Comput. Chem. Eng. 9. 10/d3csp5.

Pistikopoulos, E.N., Diangelakis, N.A., 2016. Towards the integration of process design, control and scheduling: Are we getting closer? Comput. Chem. Eng. 91, 85–92. doi:10.1016/j.compchemeng.2015.11.002.

Psichogios, D.C., Ungar, L.H., 1992. A hybrid neural network-first principles approach to process modeling. AIChE J. 38 (10), 1499–1511. 10/fcgvd7.

Qin, S.J., 2014. Process data analytics in the era of big data. AIChE J. 60 (9), 3092–3100. 10/gf69mr.
Raissi, M., Perdikaris, P., Karniadakis, G., 2019. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. J. Comput. Phys. 378, 686–707. 10/gfzbvx.
Raissi, M., Perdikaris, P., Karniadakis, G.E., 2017. Machine learning of linear differential equations using Gaussian processes. J. Comput. Phys. 348, 683–693. 10/gbzfgr.
Ramaker, H., van Sprang, E., Gurden, S., Westerhuis, J., Smilde, A., 2002. Improved monitoring of batch processes by incorporating external information. J. Process Control 12 (4), 569–576. 10/chh9jk.
Rato, T.J., Delgado, P., Martins, C., Reis, M.S., 2020. First principles statistical process monitoring of high-dimensional industrial microelectronics assembly processes. Processes 8 (11), 1520. 10.3390/pr8111520.
Rato, T.J., Reis, M.S., 2018. Building optimal multiresolution soft sensors for continuous processes. Ind. Eng. Chem. Res. 57 (30), 9750–9765. 10/gd35g8.
Reis, M.S., Braatz, R.D., Chiang, L.H., 2016. Challenges and future research directions. Chem. Eng. Prog. 112 (3), 46–50.
Reis, M.S., Gins, G., 2017. Industrial process monitoring in the Big Data/Industry 4.0 era: from detection, to diagnosis, to prognosis. Processes 5 (3), 35. 10/gf5dt5.
Reis, M.S., Gins, G., Rato, T.J., 2019. Incorporation of process-specific structure in statistical process monitoring: a review. J. Qual. Technol. 1–15. 10/gf4kf8.
Reis, M.S., Kenett, R., 2018. Assessing the value of information of data-centric activities in the chemical processing industry 4.0. AIChE J. 64 (11), 3868–3881. 10/gff327.
Reis, M.S., Saraiva, P.M., 2019. Data-centric process systems engineering for the chemical industry 4.0. In: Systems Engineering in the Fourth Industrial Revolution. John Wiley & Sons, Ltd, pp. 137–159. doi:10.1002/9781119513957.ch6.
Rendall, R., Reis, M.S., 2018. Which regression method to use? Making informed decisions in “Data-Rich/Knowledge poor” scenarios – The Predictive Analytics Comparison framework (PAC). Chemom. Intell. Lab. Syst. 181, 52–63. 10/gfc6c8.
Rengaswamy, R., 1995. A Framework for Integrating Process Monitoring, Diagnosis and Supervisory Control. Purdue University.
Romijn, R., Özkan, L., Weiland, S., Ludlage, J., Marquardt, W., 2008. A grey-box modeling approach for the reduction of nonlinear systems. J. Process Control 18 (9), 906–914. 10/cfc4sn.
Safari, S., Shabani, F., Simon, D., 2014. Multirate multisensor data fusion for linear systems using Kalman filters and a neural network. Aerosp. Sci. Technol. 39, 465–471. 10/gfzbgk.
Santos, J.E.W., Trierweiler, J.O., Farenzena, M., 2021. Model update based on transient measurements for model predictive control and hybrid real-time optimization. Ind. Eng. Chem. Res. doi:10.1021/acs.iecr.1c00212.
Schäfer, P., Caspari, A., Kleinhans, K., Mhamdi, A., Mitsos, A., 2019. Reduced dynamic modeling approach for rectification columns based on compartmentalization and artificial neural networks. AIChE J. 65 (5), e16568. doi:10.1002/aic.16568.
Schäfer, P., Caspari, A., Mhamdi, A., Mitsos, A., 2019. Economic nonlinear model predictive control using hybrid mechanistic data-driven models for optimal operation in real-time electricity markets: in-silico application to air separation processes. J. Process Control 84, 171–181. doi:10.1016/j.jprocont.2019.10.008.
Schmidhuber, J., 2015. Deep learning in neural networks: an overview. Neural Netw. 61, 85–117. 10/f6v78n.
Schubert, J., Simutis, R., Dors, M., Havlík, I., Lübbert, A., 1994. Hybrid modelling of yeast production processes – combination of a priori knowledge on different levels of sophistication. Chem. Eng. Technol. 17 (1), 10–20. 10/bwhv7q.
Schweidtmann, A.M., Mitsos, A., 2019. Deterministic global optimization with artificial neural networks embedded. J. Optim. Theory Appl. 180 (3), 925–948. doi:10.1007/s10957-018-1396-0.
Shewhart, W.A., 1931. Economic Control of Quality of Manufactured Product. Van Nostrand Company, Inc., Princeton, NJ.
Shi, H., You, F., 2015. A novel adaptive surrogate modeling-based algorithm for simultaneous optimization of sequential batch process scheduling and dynamic operations. AIChE J. 61 (12), 4191–4209. doi:10.1002/aic.14974.
Sohlberg, B., 2003. Grey box modelling for model predictive control of a heating process. J. Process Control 13 (3), 225–238. 10/bzkwss.
Solle, D., Hitzmann, B., Herwig, C., Pereira Remelhe, M., Ulonska, S., Wuerth, L., Prata, A., Steckenreiter, T., 2017. Between the poles of data-driven and mechanistic modeling for process operation. Chem. Ingenieur Tech. 89 (5), 542–561. 10/gfxhv9.
Su, H.-T., Bhat, N., Minderman, P., McAvoy, T., 1992. Integrating neural networks with first principles models for dynamic modeling. In: Dynamics and Control of Chemical Reactors, Distillation Columns and Batch Processes. Elsevier, pp. 327–332. doi:10.1016/B978-0-08-041711-0.50054-4.
Sun, B., Yang, C., Wang, Y., Gui, W., Craig, I., Olivier, L., 2020. A comprehensive hybrid first principles/machine learning modeling framework for complex industrial processes. J. Process Control 86, 30–43. doi:10.1016/j.jprocont.2019.11.012.
Tahir, F., Islam, M.T., Mack, J., Robertson, J., Lovett, D., 2019. Process monitoring and fault detection on a hot-melt extrusion process using in-line Raman spectroscopy and a hybrid soft sensor. Comput. Chem. Eng. 125, 400–414. 10/gf4kf9.
Tan, K., Li, Y., 2002. Grey-box model identification via evolutionary computing. Control Eng. Pract. 10 (7), 673–684. 10/dn5sqd.
Thompson, M.L., Kramer, M.A., 1994. Modeling chemical processes using prior knowledge and neural networks. AIChE J. 40 (8), 1328–1340. 10/dhpdjz.
Tian, Y., Zhang, J., Morris, J., 2001. Modeling and optimal control of a batch polymerization reactor using a hybrid stacked recurrent neural network model. Ind. Eng. Chem. Res. 40 (21), 4525–4535. 10/fv8sjv.
Ticknor, J.L., 2013. A Bayesian regularized artificial neural network for stock market forecasting. Expert Syst. Appl. 40 (14), 5501–5506. doi:10.1016/j.eswa.2013.04.013.
Tomba, E., Barolo, M., García-Muñoz, S., 2012. General framework for latent variable model inversion for the design and manufacturing of new products. Ind. Eng. Chem. Res. 51 (39), 12886–12900. 10/gf22z9.
Tsay, C., Baldea, M., 2020. Integrating production scheduling and process control using latent variable dynamic models. Control Eng. Pract. 94, 104201. doi:10.1016/j.conengprac.2019.104201.
Tsay, C., Baldea, M., Shi, J., Kumar, A., Flores-Cerrillo, J., 2018. Data-Driven Models and Algorithms for Demand Response Scheduling of Air Separation Units. In: Eden, M.R., Ierapetritou, M.G., Towler, G.P. (Eds.), Computer Aided Chemical Engineering. In: 13 International Symposium on Process Systems Engineering (PSE 2018), 44. Elsevier, pp. 1273–1278. doi:10.1016/B978-0-444-64241-7.50207-X.
Tsay, C., Kumar, A., Flores-Cerrillo, J., Baldea, M., 2019. Optimal demand response scheduling of an industrial air separation unit using data-driven dynamic models. Comput. Chem. Eng. 126, 22–34. doi:10.1016/j.compchemeng.2019.03.022.
Tsay, C., Pattison, R.C., Baldea, M., Weinstein, B., Hodson, S.J., Johnson, R.D., 2017. A superstructure-based design of experiments framework for simultaneous domain-restricted model identification and parameter estimation. Comput. Chem. Eng. 107, 408–426. doi:10.1016/j.compchemeng.2017.02.014.
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 J. 42 (2), 455–465. 10/c8rxvp.
Tulleken, H.J., 1993. Grey-box modelling and identification using physical knowledge and Bayesian techniques. Automatica 29 (2), 285–308. 10/db2r6t.
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 J. 42 (12), 3403–3418. 10/d5jdrb.
van Lith, P.F., Betlem, B.H., Roffel, B., 2003. Combining prior knowledge with data driven modeling of a batch distillation column including start-up. Comput. Chem. Eng. 27 (7), 1021–1030. 10/brh8r8.
van Sprang, E.N.M., Ramaker, H.-J., Westerhuis, J.A., Smilde, A.K., Wienke, D., 2005. Statistical batch process monitoring using gray models. AIChE J. 51 (3), 931–945. 10/bbkqfc.
Vapnik, V., 1999. The Nature of Statistical Learning Theory, second ed. Springer, New York.
Vedam, H., Venkatasubramanian, V., 1999. PCA-SDG based process monitoring and fault diagnosis. Control Eng. Pract. 7 (7), 903–917. 10/cb89bp.
Venkatasubramanian, V., 2019. The promise of artificial intelligence in chemical engineering: is it here, finally? AIChE J. 65 (2), 466–478. 10/gfpbzf.
Vertis, C.S., Oliveira, N.M., Bernardo, F.P., 2016. Constrained smoothing of experimental data in the identification of kinetic models. In: Computer Aided Chemical Engineering, 38. Elsevier, pp. 2121–2126. doi:10.1016/B978-0-444-63428-3.50358-1.
Von Stosch, M., 2011. Novel Strategies for Process Control Based on Hybrid Semi-Parametric Mathematical Systems. Universidade do Porto, Porto, Portugal.
von Stosch, M., Oliveira, R., Peres, J., Feyo de Azevedo, S., 2011. A novel identification method for hybrid (N)PLS dynamical systems with application to bioprocesses. Expert Syst. Appl. 38 (9), 10862–10874. 10/cm8tbk.
von Stosch, M., Oliveira, R., Peres, J., Feyo de Azevedo, S., 2014. Hybrid semi-parametric modeling in process systems engineering: past, present and future. Comput. Chem. Eng. 60, 86–101. 10/f5n73g.
Wan, Y., Yang, F., Lv, N., Xu, H., Ye, H., Li, W., Xu, P., Song, L., Usadi, A.K., 2013. Statistical root cause analysis of novel faults based on digraph models. Chem. Eng. Res. Des. 91 (1), 87–99. 10/f4pcxd.
Wang, L., 1993. Stable adaptive fuzzy control of nonlinear systems. IEEE Trans. Fuzzy Syst. 1 (2), 146–155. 10/dkg8cm.
Wang, X., Chen, J., Liu, C., Pan, F., 2010. Hybrid modeling of penicillin fermentation process based on least square support vector machine. Chem. Eng. Res. Des. 88 (4), 415–420. 10/dwxkrp.
Wang, Z., Chiang, L., 2019. Monitoring chemical processes using judicious fusion of multi-rate sensor data. Sensors 19 (10), 2240. 10/gf69mm.
Wang, Z., Georgakis, C., 2019. A Dynamic Response Surface Model for Polymer Grade Transitions in Industrial Plants. Ind. Eng. Chem. Res. 58 (26), 11187–11198. 10/gf7h9t.
Williams, T.J., Otto, R.E., 1960. A generalized chemical processing model for the investigation of computer control. Trans. Am. Inst. Electr. Eng. Part I 79 (5), 458–473. 10/gf7h7d.
Wills, A., Schön, T.B., Ljung, L., Ninness, B., 2013. Identification of Hammerstein–Wiener models. Automatica 49 (1), 70–81. 10/f22xr2.
Wilson, Z.T., Sahinidis, N.V., 2017. The ALAMO approach to machine learning. Comput. Chem. Eng. 106, 785–795. 10/gb4qc9.
Wold, H., 1974. Causal flows with latent variables: partings of the ways in the light of NIPALS modelling. Eur. Econ. Rev. 5 (1), 67–86. 10/bd6j8s.
Wold, H., 1975. Soft modelling by latent variables: the non-linear iterative partial least squares (NIPALS) approach. J. Appl. Probab. 12 (S1), 117–142. 10/gf7fbk.
Wold, S., Sjöström, M., Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. Chemom. Intell. Lab. Syst. 58 (2), 109–130. 10/cv6jxq.
Wu, C.F.J., Hamada, M.S., 2009. Experiments: Planning, Analysis, and Optimization, second ed. Wiley, Hoboken, N.J.
Wu, Z., Rincon, D., Christofides, P.D., 2020. Process structure-based recurrent neural network modeling for model predictive control of nonlinear processes. J. Process Control 89, 74–84. doi:10.1016/j.jprocont.2020.03.013.

Wu, Z., Tran, A., Rincon, D., Christofides, P.D., 2019. Machine learning-based predictive control of nonlinear processes. Part I: theory. AIChE J. 65 (11), e16729. doi:10.1002/aic.16729.

Wu, Z., Tran, A., Rincon, D., Christofides, P.D., 2019. Machine-learning-based predictive control of nonlinear processes. Part II: Computational implementation. AIChE J. 65 (11), e16734. doi:10.1002/aic.16734.

Xiong, Q., Jutan, A., 2002. Grey-box modelling and control of chemical processes. Chem. Eng. Sc. 57 (6), 1027–1039. 10/ckfhrm.

Yang, A., Martin, E., Morris, J., 2011. Identification of semi-parametric hybrid process models. Comput. Chem. Eng. 35 (1), 63–70. 10/bkt9s4.

Yang, W.-T., Blue, J., Roussy, A., Pinaton, J., Reis, M.S., 2020. A physics-informed Run-to-Run control framework for semiconductor manufacturing. Expert Syst. Appl. 155, 113424. doi:10.1016/j.eswa.2020.113424.

Yang, W.-T., Blue, J., Roussy, A., Reis, M.S., Pinaton, J., 2018. Virtual metrology modeling based on Gaussian Bayesian network. In: 2018 Winter Simulation Conference (WSC). IEEE, Gothenburg, Sweden, pp. 3574–3582. 10/gf4kh3.

Yoon, S., MacGregor, J.F., 2000. Statistical and causal model-based approaches to fault detection and isolation. AIChE J. 46 (9), 1813–1824. 10/dzwb2q.

Yoon, S., MacGregor, J.F., 2001. Incorporation of external information into multivariate pca/PLS models. IFAC Proc. Vol. 34 (27), 105–110. 10/gfxhwc.

Zadeh, L.A., 1973. Outline of a new approach to the analysis of complex systems and decision processes. IEEE Trans. Syst. Man Cybern. SMC-3 (1), 28–44. 10/b4j9t6.

Zhang, Q., Grossmann, I.E., Sundaramoorthy, A., Pinto, J.M., 2016. Data-driven construction of Convex Region Surrogate models. Optim. Eng. 17 (2), 289–332. doi:10.1007/s11081-015-9288-8.

Zhao, H., Guiver, J., Neelakantan, R., Biegler, L.T., 2001. A nonlinear industrial model predictive controller using integrated PLS and neural net state-space model. Control Eng. Pract. 9 (2), 125–133. doi:10.1016/S0967-0661(00)00101-5.

Zhao, L., You, F., 2019. A data-driven approach for industrial utility systems optimization under uncertainty. Energy 182, 559–569. doi:10.1016/j.energy.2019.06.086.

Zhu, Y., Zabaras, N., 2018. Bayesian deep convolutional encoder–decoder networks for surrogate modeling and uncertainty quantification. J. Comput. Phys. 366, 415–447. 10/gdt3jp.