

Modern Machine Learning Tools for Monitoring and Control of Industrial Processes: A Survey

R. Bhushan Gopaluni* Aditya Tulsyan** Benoit Chachuat***
Biao Huang** Jong Min Lee**** Faraz Amjad**
Seshu Kumar Damarla** Jong Woo Kim****
Nathan P. Lawrence†

* *Department of Chemical and Biological Engineering, University of British Columbia, Canada*

** *Department of Chemical and Materials Engineering, University of Alberta, Canada*

*** *Department of Chemical Engineering, Imperial College, UK*

**** *Department of Chemical Engineering, Seoul National University, South Korea*

† *Department of Mathematics, University of British Columbia, Canada*

Abstract: Over the last ten years, we have seen a significant increase in industrial data, tremendous improvement in computational power, and major theoretical advances in machine learning. This opens up an opportunity to use modern machine learning tools on large-scale nonlinear monitoring and control problems. This article provides a survey of recent results with applications in the process industry.

Keywords: statistical machine learning, deep learning, reinforcement learning, monitoring, control.

1. MOTIVATION

Recent advances in Artificial Intelligence (AI) have led to major breakthroughs in image processing, voice recognition and generation, facial recognition, natural language translation, autonomous driving, and software that can play complex games such as chess and Go (Jackson, 2019; Ertel, 2018). Currently, there is a tremendous interest in developing AI-based solutions for a variety of longstanding problems in many scientific and engineering disciplines (Nilsson, 2014; Russell and Norvig, 2016). While some of the promises of AI are potentially hype or misplaced, there are genuine reasons for us to be excited about AI and its impact on process industries (Tzafestas and Verbruggen, 2012). This optimism is driven by the realization that we are at an exciting moment in history with the availability of large volumes of data, enormous computing power, and significant algorithmic advances. Together, we, in the process systems engineering (PSE) community, can develop a vision to design advanced AI-based solutions under the Industry 4.0 paradigm.

We are at the beginning of the fourth industrial revolution that is going to be powered by data (Ustundag and Cevikcan, 2017; Li et al., 2017). This revolution is expected to sweep a wide range of industries, including Chemical and Biological industries such as Oil & Gas, Petrochemicals, Pharmaceuticals, Biotechnology, Pulp & Paper, and other manufacturing industries. A recent white

paper by the consulting firm McKinsey¹ concluded that the manufacturing industries could generate information worth hundreds of billions of dollars out of existing data and potentially generate much higher value with additional data (Brown et al., 2011). With the advent of the Internet of Things, these industries are poised to generate large volumes of data. This presents an exciting opportunity to exploit data, computers, and algorithms to solve traditional problems surrounding monitoring and control and provide new insights.

Data analytics and machine learning (ML) ideas are not new to the process industries². The review paper by Venkatasubramanian (2019) provides an excellent overview of the history, the successes and the failures of various attempts, over more than three decades, to use ideas from AI in the industry. In particular, statistical techniques such as principal component analysis (PCA), partial least squares (PLS), canonical variate analysis (CVA), time series methods for modeling such as maximum likelihood estimation and prediction error methods have been extensively used in industry (Chiang et al. (2000)). Several classification and clustering algorithms, such as k-means, Support Vector Machines (SVMs), Fisher Discriminant Analysis (FDA), are also widely used in the industry (Qin and Chiang (2019)). Also, several nonlinear approaches, such as kernel methods, Gaussian Processes (GPs), and adaptive control algorithms, such as Reinforce-

¹ www.mckinsey.com

² For simplicity; we refer to chemical and biological industries as process industries or just industries

* Email: bhushan.gopaluni@ubc.ca

ment Learning (RL), have been applied in some niche applications (Badgwell et al. (2018); Spielberg et al. (2019); Tulsyan et al. (2019b, 2020)). Most of these algorithms do not scale well with data, and their performance does not necessarily improve with data.

Through this survey, we are summarizing some of the recent advancements in ML with a focus on its applications in the process industries. We predominantly focus on three application areas – soft sensing, process monitoring, and process control. For brevity, we only survey four classes of ML methods commonly used in process industries, namely: (1) statistical learning methods; (2) deep learning and its variants; (3) hybrid modeling; and (4) reinforcement learning. Statistical learning methods and deep learning methods are discussed for process modeling, soft-sensing, and process monitoring applications, while hybrid modeling and reinforcement learning are discussed for process control applications. This is by no means an exhaustive survey of the recent research on these topics; however, we have tried our best to include some of the most critical developments of ML tools in process industries. The rest of the article is divided as follows: Section 2 surveys statistical and deep learning methods for soft sensor development in process industries; Section 3 discusses deep learning methods for process modeling and process monitoring; Section 4 surveys reinforcement learning methods for process control applications; Section 5 discusses hybrid modeling techniques for process optimization and control applications; and Section 6 surveys applications of ML tools in applications with limited (or small) data sets.

2. STATISTICAL AND DEEP LEARNING METHODS FOR SOFT SENSOR APPLICATIONS

Soft sensors have been used in process industries for over two decades to predict product quality and to estimate key process variables (Khatibisepehr et al. (2013); Fortuna et al. (2007)). Soft sensors can be categorized into model-driven and data-driven soft sensors. Model-driven soft sensors (or white box models), such as Kalman filters (and its variants and observers), are based on the first principles models (FPM) that describe the physical and chemical laws that govern a process (e.g., mass and energy balance equations). In contrast, data-driven soft sensors (or black box models) have no information about the process and are based on empirical observations (historical process data). In between white box and black box models, there is a third type of soft sensors called grey-box models in which a model-driven soft sensor uses a data-driven method to estimate the parameters of an FPM. The statistics in Figure 1 indicate that in the last five years, the research in soft-sensing was predominantly focused on data-based models³. This is not surprising as data-driven soft sensors are often able to capture complex (and unexplained) process dynamics much more succinctly. In contrast, model-driven soft sensors require a lot of process expert knowledge, which is not always available. Further, model-driven soft sensors are difficult to calibrate, especially for complex nonlinear processes. Note that the grey box model-based soft sensors received the least research attention. The data-

³ Supplementary material explaining the sources of the data used in this section is available at <http://dais.chbe.ubc.ca/publications/>

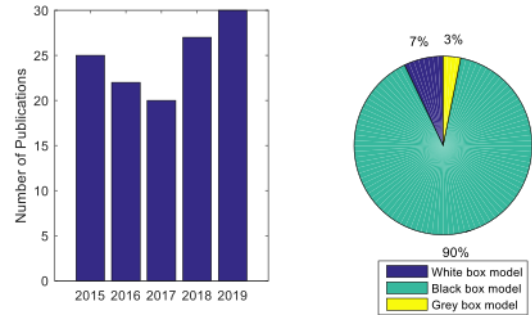


Fig. 1. Research publications in soft sensors from 2015 to 2019

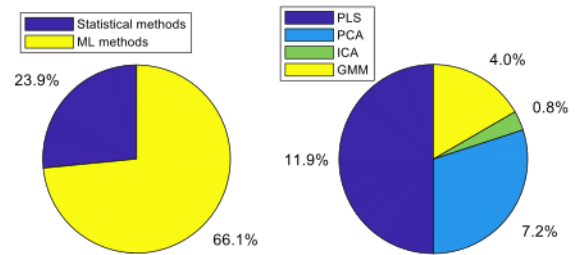


Fig. 2. Distribution of data-based modeling techniques (left figure), and statistical methods (right figure)

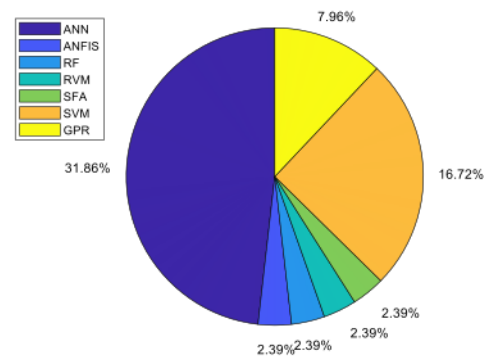


Fig. 3. Distribution of ML methods in soft sensing

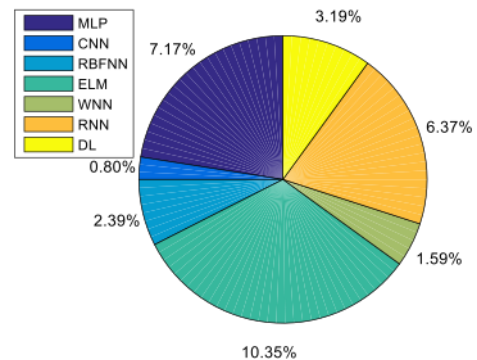


Fig. 4. Distribution of various types of ANN in soft sensing driven soft sensors can be further categorized based on the learning technique employed for modeling.

Figures 2 to 4 show the current trend in the data-driven soft sensing. Figure 5 contains full forms for the acronyms used in Figures 2 to 4. The research in soft sensing has dramatically shifted from traditional methods to ML meth-

ICA-Independent component analysis	SFA-Slow feature analysis
GMM-Gaussian mixture model	MLP-Multilayer perceptron
GPR-Gaussian process regression	CNN-Convolution neural network
ANN-Artificial neural networks	RBFN-Radial basis function network
ANFIS-Adaptive neuro fuzzy inference system	ELM-Extreme learning machine
RF-Random forest	RNN-Recurrent neural network
RVM-Relevant vector machine	WNN-Wavelet neural network

Fig. 5. Full forms for acronyms

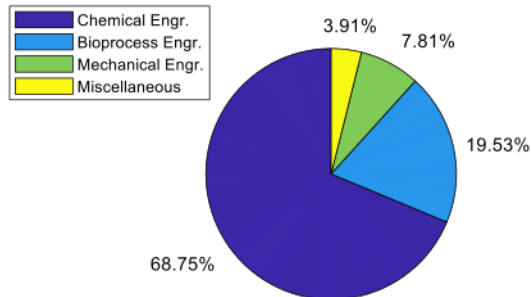


Fig. 6. Distribution of soft sensor applications The chemical industries such as refineries, steel plants, polymer industries, cement industries, etc. remain the dominant users of soft sensors.

ods. Among the ML methods, ANNs received the greatest attention due to their better generalization performance across several different application areas. The class of feed-forward single hidden layer neural networks encompassing MLP, ELM, RBFNN, WNN, etc. has more soft sensing applications than recurrent and deep neural networks. There has been a single research publication reporting the application of Transfer Learning (TL) in inferential measurement. TL may be more suitable for predictive modeling problems that use image data as inputs. TL has not yet been applied for online prediction of process variables. Aside from ANN, SVM is the second most widely used ML method for developing inferential models.

2.1 Image-based Soft Sensors

With the development of better camera technology as well as novel image processing techniques, image processing has found increasing use in an industrial setting, sometimes even replacing conventional sensors. There has been extensive research in the application of image processing to the mineral industry (Horn et al. (2017); Poliakov and Donskoi (2014); Patel and Chatterjee (2016); Kistner et al. (2013)). In Horn et al. (2017), for example, online estimation of ore composition on conveyor belts was done. Features and textures were extracted using PCA and Wavelet Texture Analysis (WTA), and then class boundaries were established using SVM. Recently, Convolutional Neural Networks (CNNs) have become popular due to training and hardware breakthroughs, and their use is steadily increasing. In Horn et al. (2017), CNN was used in the separation of Platinum from other impurities in a froth flotation cell. Images of the froth in the cell were captured using a camera, and CNN was used to extract features from this image. A feature in image processing refers to characteristics, such as edges, shapes, colors, etc., which can be used to distinguish between different objects. Classical ML techniques such as k-NN (k-Nearest Neighbor) and LDA (Linear Discriminant Analysis) were then trained on these features and used to predict Platinum

separation. The mineral industry has quickly adapted to using deep learning for image processing (Fu and Aldrich (2018); Carvalho et al. (2019); Iglesias et al. (2019); Li et al. (2019); Kholief et al. (2017)). In the last five years, there have been well over 200 papers on the application of both classical computer vision techniques and CNN in the minerals and ore industry. Computer vision has also been widely used in product fault detection (Maschi et al. (2012); Kong and Li (2018); Hocenski et al. (2016); Rodríguez-Pulido et al. (2013)). In Maschi et al. (2012), for example, CNN is trained to detect steel production defects directly on the production line.

Image processing has also found use in the food industry (Sarkate et al. (2013); Siswantoro et al. (2013); O'Brien and Cooper (2015)). For example, in Yu et al. (2003), image processing is used to find the amount of coating on chips particles, automating the coating process downstream. Features are extracted from images of chips by first obtaining the 3-channel histogram of the RGB color image, carrying out PCA to reduce the number of features, and then lumping/classifying pixels into L classes/bins. Once the features are obtained, PLS regression is used to find the relationship between feature data and the amount of coating on chips. Like the mineral industry, there have been over 200 papers in the field of food quality management in the past five years. Chemical process industries, on the other hand, have been a little slower in adopting computer vision and deep learning for their applications. This is, however, changing. For example, CNN has been applied to thermal cameras installed in furnaces to get the temperature inside the furnace (Zhu et al. (2018)). A combination of the famous ResNet-50 and U-Net CNN architectures were used for first segmenting (separating/detecting) tube regions in an image from the furnace, and then used to detect the temperature inside the furnace in normal operation. In Bruno et al. (2018), a CNN is used for classifying results of crystallization trials into four classes (crystals, precipitate, clear, other). The widely used Inception-v3 CNN architecture was used to train on nearly 350,000 such images, and at test time, it achieved an accuracy of 94%. In Vicente et al. (2019), image processing was used to track the interface level in a vessel with a sight glass in an oil sands process industry. Repeated frame differencing was used to find areas of maximum movement, which corresponds to the interface level. In Aissa et al. (2010), RGB color analysis is carried out to study and predict powder mixing efficiency. The gist of the procedure is that first, PCA is carried out to map the image's pixels into a lower-dimensional plane. Then secondly, pixel color classification is done using the pixel score scatter plot. Finally, a total count of the number of pixels in each class gives a measure of the area occupied by each color powder. This can be used to calculate the fraction of the area of each powder in a selected region of the image, which gives us a measure of the degree of homogeneity of the powder mixing process. In Eppel and Kachman (2014), computer vision is applied to separate a continuous liquid-liquid flow. Another similar application has been done in Li et al. (2018) in finding the level of a liquid in a process vessel. Overall, there has been comparatively less research in the application of computer vision to process industries, compared to the minerals and food industry, with roughly around 50 papers in the past 5

years, mostly for vessel rust detection and pipeline failure, etc.

3. DEEP LEARNING FOR MODELING AND FAULT DETECTION

Many recent breakthroughs in computer science and automation can arguably be attributed to a single breakthrough at the heart of all the other advances. That breakthrough is *Deep Learning*. Deep Learning refers to Deep Neural Networks (DNN), which are mostly Artificial Neural Networks (ANN) with many layers (typically more than 10) (Goodfellow et al. (2016)). A simplistic view of a DNN is that of a mathematical mapping between specific inputs and outputs. As such, at least in theory, it is possible to use DNNs to approximate a wide range of complex nonlinear functions. Universal Approximation Theorem for Feed-Forward ANNs states that a single layer ANN with a sufficiently large number of neurons can approximate a continuous function on a compact subset of \mathbb{R}^n with arbitrary accuracy (with certain mild conditions - please refer to Csáji (2001) for technical details). Moreover, with the currently available large-scale optimization tools, such as stochastic gradient descent, it is possible to use DNNs to approximate large-dimensional nonlinear functions. There are two fundamental problems in the process industry where the function approximation power of DNNs is immediately useful: (1) Large-scale nonlinear modeling and (2) Fault Detection and Diagnosis.

3.1 Large-Scale Modeling

There are several ways in which models can be used; however, the most common application of DNNs is in the soft-sensing area in industrial processes. One of the first articles on this topic appeared in Shang et al. (2014). In this article, it was argued that the depth of the network in DNNs provides a high level of flexibility for approximating multi-layered processes. In particular, they demonstrate, on a crude-oil distillation column, that a DNN is superior to other nonlinear modeling approaches. Since this initial work by Shang et al. (2014), several other authors have pursued this line of research. For instance, in Yan et al. (2016) the performance of a standard DNN was enhanced by combining it with a de-noising auto-encoder. In Gopakumar et al. (2018); Wang et al. (2018); Villalba-Diez et al. (2019) and the references therein various enhancements of DNN were used to model different processes.

There are different variants of the basic DNN that are more suitable for time-series data common in industrial processes. For instance, standard Recurrent Neural Networks (RNN) and their variant Long-Short-Term Memory (LSTM) networks are two popular categories of DNN. In Wang et al. (2018), LSTMs were used to model different phases of a batch process. More recently, there has also been work on using Deep Auto-Encoders (Yuan et al. (2018)) for soft-sensing. The deep auto-encoders allow us to identify a set of noise-less features that can, in turn, be used for modeling and prediction.

The initial success of deep learning was in image processing. In particular, a variant of DNNs called CNNs had been

widely used in image classification, voice recognition, etc. Building on the success of CNNs, a few authors have tried to convert multivariate time-series data into images and make large-scale CNN models for soft-sensing (Wang et al. (2019b)). In situations where there is limited data, the idea of transfer learning (where a pre-built DNN is modified slightly) has also been used for soft-sensor modeling (Liu et al. (2019); Patwardhan et al. (2019)).

DNNs provide an effective nonlinear alternative to traditional soft-sensing methods such as time-series models, PLS, PCA, etc. However, DNNs require large volumes of data with sufficient information in them. As of now, it is not clear what is “Good Data” and when it is sufficient to estimate a robust DNN. Moreover, DNNs are black-box models, and hence it is difficult to relate them to the physical properties of a process. However, some recent studies have shown that, despite their complexity, it is possible to study properties such as stability (Bonassi et al. (2019)).

3.2 Fault Detection and Diagnosis

The flexibility of DNNs in approximating complex nonlinear functions makes them an obvious choice for fault detection and diagnosis. In this context, DNNs have been used in two different ways: (1) as simple classifiers and (2) as tools for generating appropriate lower-dimensional latent space, which in turn is used in process monitoring and fault detection. Recently, several researchers in a wide range of application areas have published work on using DNNs for fault detection and diagnosis.

The vast majority of articles in this area fall under the category of classifiers. The basic idea is to use a DNN classifier to distinguish between normal and faulty operating regions. For instance, in Mireles Gonzalez (2018) and Lv et al. (2016), an RNN was used to build a classifier that identified faults in the well-known Tennessee Eastman benchmark problem. They show that the DNNs perform far better than other commonly used algorithms for fault detection.

The second class of algorithms uses variational auto-encoders to generate a lower-dimensional latent space and use it in conjunction with a statistic to monitor a process and detect any anomalies. An example of this approach is in Wang et al. (2019a). More recently, Generative Adversarial Networks have also been used to perform fault detection (Li et al. (2018)).

4. REINFORCEMENT LEARNING

RL refers to the class of numerical methods for the data-driven sequential decision-making problem (Sutton and Barto, 2018). The RL agent (algorithm) aims to find an optimal policy, or controller, through interactions with its environment. Finding such a policy requires solving the Bellman equation using the principle of optimality; however, it is intractable as it requires a solution in the entire state domain (Bertsekas, 2005). Recent advances in ML enable feature analysis of the raw sensory-level data using deep neural networks (DNNs) and the implementation of various information-theoretic techniques. As a result, Deep RL (DRL) is a general framework that has shown

remarkable performances in specific applications such as robotics, autonomous driving, and video games, and board games. (Levine et al., 2016; Williams et al., 2016; OpenAI et al., 2019; Silver et al., 2017).

In this paper, we classify algorithms based on the use of model knowledge; that is, model-free and model-based approaches. Model-free is a category of RL in which dynamic models, either first-principle or data-driven forms, are not utilized. Instead, only the reinforcement signal (i.e., reward or cost) is used without model knowledge. The only requirement for underlying system models in the RL framework is a Markov property (Puterman, 2014). Models are often challenging to build, so the ability to bypass the modeling exercise is a considerable advantage for designing the controller. A model-based method, on the other hand, utilizes the state-space model explicitly. Although the process modeling requires extensive domain knowledge, and the results are inaccurate in most cases, it provides a theoretical background of optimality under the specific conditions.

4.1 Process Control

With the demands on the performance of process systems, efficient optimization is becoming increasingly essential. The ultimate dream-goal of any process control system will be to develop a controller capable of attaining such optimality in large-scale, nonlinear, and hybrid models with constraints, fast online calculation, and adaptation. This ideal controller should be amenable to a closed-loop solution using a limited number of observations and should be robust to online disturbances. While the mathematical programming (MP) based approaches such as model predictive control and direct optimization have been prevalent, several researchers have also focused on the RL framework, as a complementary method (Buşoniu et al., 2018; Shin et al., 2019). This comes from RL’s advantages, which allow for designing the feedback controller to interact with its environment and reflect the massive raw data collected throughout the whole industrial process.

Several pioneering pieces of work in the area of PSE applications of RL were first reported in (Kaisare et al., 2003; Lee et al., 2006; Lee and Lee, 2009). The model-free RL was adopted for solving the optimal control, dual adaptive control, and scheduling problems. It was shown that the approximation of the value function could provide robust control despite the presence of process noise and model changes. These results were later extended to the dynamic optimization and robust optimal control problems (Nosair et al., 2010; Yang and Lee, 2010, 2013). The recent applications of the RL-based process control approach mainly rely on the linear approximator in both model-based and model-free methods (Cui et al., 2018; Sun et al., 2018; Ge et al., 2018; Kim et al., 2018a). In the meantime, a few DRL applications are studied in Kim et al. (2018b); Spielberg et al. (2019).

4.2 Challenges and Advances in DRL

We highlight several recent advances in the DRL literature with an emphasis on model-free methods. Model-free RL consists of value-based and policy-based methods. In the

former, the objective is to find the optimal value function parameter, which minimizes the Bellman error. The policy-based method focuses on the direct optimization of the original objective with respect to the parametrized policy function (Deisenroth et al., 2013). Classical algorithms for value-based methods, such as Q-learning and SARSA, and policy-based methods, such as REINFORCE, enjoy theoretical convergence. However, convergence can be slow, due to high variance in value estimates, or limited to the tabular setting or linear function approximation (Sutton and Barto, 2018). Nonetheless, these methods provide the foundation for DRL algorithms.

Deep Q Networks (DQNs) were introduced by Mnih et al. (2015) as a synthesis of deep learning and Q-learning. DQNs are limited to discrete action spaces but showed impressive results in tasks with high-dimensional sensory input data such as Atari games. More recent algorithms inspired by DQNs fall into the actor-critic (AC) framework (Konda and Tsitsiklis, 2000). The goal of AC methods is to jointly optimize the value-based and policy-based objectives. This strategy leverages a parameterized value function to optimize the policy parameters with policy gradient update steps in an online, off-policy fashion. Notably, the Deep Deterministic Policy Gradient (DDPG) algorithm was introduced by Lillicrap et al. (2015). DDPG extends DQN with the use of the deterministic policy gradient theorem (Silver et al., 2014). Despite the advances made by DDPG, it is notoriously difficult to use, for example, due to sensitivity to hyperparameters and overestimation of Q-function values (Henderson et al., 2018; Fujimoto et al., 2018). This limits the viability of DDPG for real-world applications such as process control.

More recent DRL algorithms have been geared towards addressing these issues. Twin Delayed DDPG (TD3) addressed the overestimation bias of Q-values as well as variance in target values through the use of two DQNs, delayed actor updates, and target policy smoothing (Fujimoto et al., 2018). Concurrently, Soft Actor-Critic (SAC) was developed with similar enhancements from TD3 (such as double DQNs) but is distinct from previous DRL algorithms discussed here (Haarnoja et al., 2018b,c). In particular, SAC considers a stochastic actor and optimizes it by maximizing the weighted entropy in an objective similar to the traditional policy-based objective. Such an objective naturally balances *exploitation* and *exploration*.

SAC has been shown to outperform DDPG in sample complexity, robustness, hyperparameters, and performance. In a follow-up paper, Haarnoja et al. (2018a) demonstrate this in benchmark simulated locomotion tasks. These are important measures for any RL algorithm, as a physical system cannot be extensively probed to find suitable hyperparameters. To this end, the authors show that SAC can learn a high-performance policy for a physical quadrupedal robot in around two hours. They demonstrate the robustness of the policy to new terrains as well as to perturbations at the base of the robot.

Despite these advances, model-free RL algorithms are not sufficiently data-efficient and, therefore, not yet useful in real industrial applications (Recht, 2019). Model-based methods, on the other hand, are often more data-efficient than model-free methods. Several model-based RL

algorithms have been developed to solve the Hamilton-Jacobi-Bellman equation adaptively, hence referred to as adaptive DP (Prokhorov and Wunsch, 1997; Lewis and Vrabie, 2009; Jiang and Jiang, 2014). The model-based local optimal policy can be parametrized to design the end-to-end controller (Levine et al., 2016). Another extension is to study the connection between the stochastic optimal control framework (Theodorou et al., 2010a,b). However, it is difficult to build accurate models of industrial processes due to the complexity of these processes and also due to a variety of unknown disturbances (Lee et al., 2018). There has been recent work focused on unifying model-free and model-based approaches (Janner et al., 2019; Du and Narasimhan, 2019). This is a useful endeavor, as model-free algorithms can often achieve superior final (asymptotic) performance over model-based approaches, but suffer from relatively weak sample complexity. It is crucial to understand the trade-offs between model-free and model-based methods and their respective data efficiencies from the perspectives of RL and PSE fields. This is an area with tremendous potential for applications that can redefine automation in the process industries.

5. HYBRID MODELING METHODOLOGY AND APPLICATIONS

In essence, *knowledge-driven* (mechanistic) modeling based on first principles and *data-driven* modeling constitutes two opposite strategies. Developing a mechanistic model requires a deep understanding of the processes at play and is often labor-intensive. In contrast, a data-driven model requires little physical knowledge, flexible structure, and fast to deploy or maintain. But a larger data set is also typically needed to construct a data-driven model, and its validity may not extend far beyond the conditions under which it was trained. A closely related classification of models concerns their parameterization. The structure of a *parametric* model is determined a priori based on mechanistic knowledge, which leads to a fixed number of parameters often with physical or empirical interpretation. By contrast, the nature and number of parameters in a so-called *black-box* model are not determined by a priori knowledge but tailored to the data at hand. In between these two extremes lies hybrid *semi-parametric* modeling. This approach—also referred to as gray-box or block-oriented modeling in the literature—aims precisely at balancing the pros and cons of pure knowledge- and data-driven methods.

Hybrid modeling has been investigated for over 25 years in chemical and biological process engineering (Psichogios and Ungar, 1992; Su et al., 1992; Thompson and Kramer, 1994; Chen et al., 2000). The claimed benefits of hybrid modeling in these application domains include a faster prediction capability, better extrapolation capability, better calibration properties, easier model life-cycle management, and higher benefit/cost ratio to solve complex problems—see recent survey papers on the development and applications of hybrid models by von Stosch et al. (2014); Solle et al. (2017); Zendehboudi et al. (2018). The rest of this section focuses on current trends and challenges in hybrid model formulations and their applications in model-based optimization and control.

5.1 Hybrid Model Formulations

A usual classification of hybrid model structures is either as *serial* or *parallel* (Agarwal, 1997). In the serial approach, the data-driven model is most commonly used as an input to the mechanistic model—for instance, a material balance equation with a kinetic rate expressed using a black-box model. This structure is especially suited to situations whereby precise knowledge about some underlying mechanisms is lacking, but sufficient process data exists to infer the corresponding relationship (Psichogios and Ungar, 1992; Chen et al., 2000). But when the mechanistic part of the model presents a structural mismatch, one should not expect the serial approach to perform better than a purely mechanistic approach. In the parallel approach, instead, the output of the data-driven model is typically superimposed to that of the mechanistic model (Su et al., 1992; Thompson and Kramer, 1994). This structure can significantly improve the prediction accuracy of a mechanistic model when the data-driven component is trained on the residuals between process observations and mechanistic model predictions. Of course, this accuracy may be no better than that of the sole mechanistic model when the process conditions differ drastically from those in the training set.

By far, the most common black-box models used in this context are multilayer perceptrons (MLP) and radial basis function (RBF) based regression techniques (von Stosch et al., 2014). For instance, Chaffart and Ricardez-Sandoval (2018) developed a serial hybrid model of the thin film growth process, which couples a macroscopic gas-phase model described by PDEs to a microscopic thin-film model described by stochastic PDEs via an MLP. But alternative statistical and ML techniques have also been investigated in hybrid modeling. Ghosh et al. (2019) recently proposed subspace identification as the data-driven component in a parallel hybrid model and demonstrated the approach on a batch polymerization reactor. The use of the GP regression, which provides an estimate of the prediction error, is also showing promises in bioprocess engineering applications (Zhang et al., 2019a). Further applications of hybrid modeling in the chemical industry are surveyed by Schuppert and Mrziglod (2018).

The use of parallel hybrid models can significantly alleviate the issue of maintaining or recalibrating a complex mechanistic model by performing the update on the black-box model only. By contrast, training or updating a serial hybrid model is more involved because the outputs of the data-driven component may not be directly measured. Thus, assessing the performance of the black-box model requires simulating the full serial hybrid model and comparing its outputs to the available observations. The classical Akaike Information Criterion or Bayesian Information Criteria can be applied to discriminate among multiple black-box model structures. Willis and von Stosch (2017) recently presented an approach based on sparse regression to simultaneously decide the structure and identify the parameters for a class of rational functions embedded in a serial hybrid model. More generally, one could envision the extension of sparse regression techniques such as ALAMO (Wilson and Sahinidis, 2017) to enable the construction of serial hybrid models.

Looking beyond hybrid semi-parametric models, Venkatasubramanian (2019) argues for the development of hybrid AI systems that would combine not only mechanistic with data-driven models but also causal model-based explanatory systems or domain-specific knowledge engines. Likewise, the mechanistic model could be replaced by a graph-theoretical model (e.g., signed digraphs) or a production system model (e.g., rule-based representations), creating entirely new fields of research.

5.2 Model-based Optimization and Control

A large number of hybrid semi-parametric modeling applications have been for open-loop process optimization. There, a hybrid model is appealing because key operational variables in terms of process performance may be included in the mechanistic part of the model—e.g., to retain sufficient generalization—while capturing other parts of the process using data-driven techniques—e.g., to reduce the computational burden. Either local (gradient-based) or stochastic search techniques have been applied to solve the resulting model-based optimization problems. A notable exception is a work by Schweidtmann and Mitsos (2019) applying complete search techniques to overcome convergence to a local optimum and guarantee global optimality in problems with MLP embedded.

It should be noted that developing a data-driven or hybrid model to speed up the optimization of a more fundamental model is akin to conducting a surrogate-based optimization. The latter constitutes active research in process flowsheet optimization, where existing approaches can be broadly classified into two categories. Global approaches proceed by constructing a surrogate model based on an ensemble of flowsheet simulations before optimizing it, often within an iteration where the surrogate is progressively refined. A number of successful implementations rely on MLP (Henao and Maravelias, 2011), GP (Caballero and Grossmann, 2008; Quirante et al., 2015; Keßler et al., 2019), or a combination of various basis functions (Cozad et al., 2015; Boukouvala and Floudas, 2017) for the surrogate modeling. By contrast, local approaches maintain an accurate representation of the flowsheet (or separate units thereof) within a trust region, whose position and size are adapted iteratively. This procedure entails reconstructing the surrogate model as the trust-region moves around, but it can offer convergence guarantees under mild conditions. Applications of this approach to flowsheet optimization include the work by Eason and Biegler (2016, 2018); Bajaj et al. (2018).

The real-time optimization (RTO) and nonlinear/economic model predictive control (MPC) methodologies use a process model at their core. A majority of the RTO and MPC successful implementations have so far relied on mechanistic models, but there has been a renewal of interest in data-driven approaches; see, for instance, Je et al. (2018); Yang and Kelly (2019); Wu et al. (2019). In this context, too, hybrid models can play a crucial role in reducing the dependency on data and infusing physical knowledge for better extrapolation capability (Klimasauskas, 1998; Noor et al., 2010). Recently, Zhang et al. (2019b) took the extra step of using the same hybrid model simultaneously in the RTO and MPC layers. Notice that the vast majority

of these applications consider serial hybrid models with embedded MLP to approximate complex nonlinearities in the system.

An RTO methodology that exploits the parallel approach of hybrid semi-parametric modeling is modifier adaptation (Chachuat et al., 2009). Unlike classical RTO, modifier adaptation does not adapt the mechanistic model but adds correction terms—the modifiers—to the cost and constraint functions in the optimization model. Original work used the process measurement to estimate linear (gradient-based) corrections (Marchetti et al., 2009). Gao et al. (2016) later explored the use of data-driven approaches based on quadratic surrogates trained with current and past process measurements. More recently, Ferreira et al. (2018) proposed to use GPs as the modifiers for the cost and constraint functions, then del Rio-Chanona et al. (2019) combined this approach with a trust-region algorithm. These developments share many similarities with surrogate-based optimization techniques, with the added complexity that the process data are noisy, and the process optimum might change over time.

6. ML WITH LIMITED DATA

The success of ML can be partly attributed to the breakthrough advancements in computation and computing technology and the availability of large amounts of data. ML continues to make further inroads into sciences (e.g., biomedical, pharmaceutical, and material sciences). However, there is a growing interest in extending the application of ML under Low-N (or small data) settings that define most of these chemical and biological systems. The two critical problems in applying ML under Low-N settings are inefficient model training and model generalization (see Figure 7). It is well-known that in the presence of small datasets, ML models (regression or classification) are nontrivial to train and exhibit limited validity. Next, we present a brief survey of traditional and current methods that make ML robust under Low-N settings.

6.1 Improving Model Generalization

Several methods based on data generation (e.g., generative models), model regularization (e.g., Ridge, LASSO, Elastic Nets), and ensemble methods (e.g., Bagging, Boosting) have been proposed to improve generalization capabilities of ML under Low-N settings. For brevity, only methods based on data generation are presented here. Data generators are an essential class of methods that improve model generalization by complementing original small data set with a large number of *in silico* (or computer-generated) data sets. The idea of data generators was first introduced by Poggio and Vetter (1992) for pattern recognition. There are two categories of data generators. The first category of generators extracts prior process knowledge to generate data. For example, Poggio and Vetter (1992) used the 3D view of the object to generate images from any other arbitrary angle. Knowledge-based generators have also been applied in handwritten number recognition (Schölkopf et al., 1998), text recognition (Su et al., 2006), and noise source recognition (Xu et al., 2008). In process industries, knowledge-based process simulators and first-principle models have also been extensively used to gener-

ate process data under Low-N settings. While knowledge-based generators ensure that the generated samples are representative of the application in-hand, such generators are often nontrivial to construct.

The second class of methods generates *in silico* data by perturbing the original small data set. Lee (2000) proposed a perturbation method to generate data by adding a small Gaussian noise to the original samples. Later Wang and Yang (2008) introduced a generator that adds a small constant to every dimension of the n_x -dimensional training sample, thereby generating n_x *in silico* samples for every training sample. Zhang and Gong-he (2006) then introduced another method that first divides the training samples into n_x groups using k -nearest neighbor algorithm, and then generates virtual samples by averaging every two samples in each group keeping the labels unchanged. Other methods based on probabilistic programming (Tulyan et al., 2019a, 2018), Jackknife, and bootstrapping resampling have also been proposed (Tsai and Li, 2008). Recently, there has been a growing buzz around a new

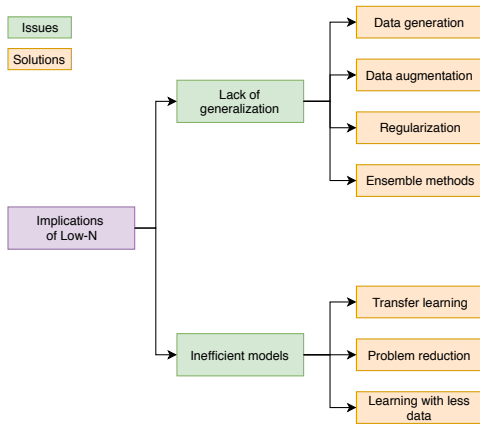


Fig. 7. Implications of Low-N in ML (Maheswari, 2018).

class of data generators, referred to as generative models. Generative models aim at learning the distribution of the original data set to generate new data points with some variations. Two of the most commonly used and efficient approaches are Variational Autoencoders (VAE) (Kingma and Welling, 2013) and Generative Adversarial Networks (GAN) (Goodfellow et al., 2014). Both VAEs and GANs are based on the deep-learning framework. Just in the last five years, several applications of GANs and VAEs have emerged in multiple disciplines encompassing image editing (Perarnau et al., 2016), medical image analysis (Kazemini et al., 2018), anomaly detection (Li et al., 2018), and financial analysis (Wiese et al., 2019). Despite the broad appeal of generative models, they require substantially large amounts of training data and also suffer from network instability, which warrants further research. However, there has been some interesting recent work on self-validating classifiers that allow a user to determine if the available limited data is informative enough .

6.2 Improving Model Training

It is well-known that ML requires access to large amounts of data for reliable model training (regression or classification). In the presence of Low-N, model training (i.e., model

parameter estimation) becomes a critical issue. Several methods have been proposed to address this problem, including model reduction techniques, novel training methods, and transfer learning. Model reduction techniques primarily aim at reducing the number of model parameters such that the reduced models can be adequately trained with small data sets, often at the cost of model accuracy.

Several attempts have been made to improve model training under Low-N through novel training strategies. For example, Byrne (1993) proposed a maximum-likelihood criterion to avoid model over-fitting under the Low-N scenario. A similar Bayesian approach was later introduced by Mao et al. (2006). Bayesian methods have proven to be useful in identifying robust process models. For example, Jang and Gopaluni (2011) proposed a Markov-Chain Monte Carlo (MCMC) method to identify a batch process, while Oniško et al. (2001) proposed a Bayesian network to identify a manufacturing process. Similarly, SVM and neural network (NN) models have also been studied under the Low-N scenario (Ingrassia and Morlini, 2005). Another approach to improve model training under Low-N is to use adaptive training (or online learning) (Zhao et al., 2019; Barazandegan et al., 2015), where information from new data is incorporated into the model, as and when they become available (Zhao, 2014). For example, in the context of batch process monitoring under Low-N: Li et al. (2000) proposed two adaptive PCA process monitoring algorithms to update the correlation matrix recursively and Flores-Cerrillo and MacGregor (2004) extended the multi-block PCA/PLS to incorporate explicitly batch-to-batch trajectory information. Zhao et al. (2005) proposed a double moving window PCA technique for adaptive monitoring, and Lu et al. (2004) proposed a two-dimensional dynamic PCA, which captures both within-batch and batch-to-batch dynamics simultaneously.

Recently, transfer learning has garnered a lot of interest from researchers as a tool of choice to make model training effective under Low-N settings (Pan and Yang, 2009). Transfer learning makes use of the knowledge gained while solving one problem and applying it to a different but related problem. For example, it is now well established that the initial layers of a neural net model, such as ResNet trained on ImageNet data, learn to identify edges and corners in the image, and later layers build on top of these features to learn more complicated structures. For any new problem whose data looks similar to ImageNet, we can start with pre-trained ImageNet models, change the final layers, and fine-tune it to our dataset. Since the lower layers feature remains relevant, this process makes the optimization process fast and reduces the amount of data required to train new models. While transfer learning is still an emerging area of research, some of the early successful applications, especially in computer vision tasks (Oquab et al., 2014; Shin et al., 2016), already make transfer learning an attractive tool under Low-N settings. One step further is meta-learning, or “learning to learn”. Meta-learning algorithms fine-tune some aspect of the underlying learning algorithm; for example, Finn et al. (2017); Nichol et al. (2018) develop simple algorithms for any neural network architecture that directly optimizes for initial parameters such that they can quickly be adapted to new tasks with a small amount of data, showing

superior performance over standard transfer learning in classification and RL tasks.

7. CONCLUSIONS

There have been several recent successes of modern ML algorithms in a variety of engineering and non-engineering areas. Several of these successes can potentially be ported to process industries. This gives us renewed optimism for developing a new level of enhanced automation in the process industries. We believe that these successes can be translated to build better models, better monitoring and fault diagnosis algorithms, and better controllers using a multitude of data types that include time-series data, text messages, images, and videos. In this survey, we made an effort to provide a vision of this emerging future.

8. ACKNOWLEDGEMENTS

We are truly grateful to the anonymous reviewers for their detailed and constructive feedback; however, we apologize for not being able to incorporate all of their suggestions.

REFERENCES

- Agarwal, M. (1997). Combining neural and conventional paradigms for modelling, prediction and control. *International Journal of Systems Science*, 28(1), 65–81.
- Aissa, A.A., Duchesne, C., and Rodrigue, D. (2010). Polymer powders mixing part II: Multi-component mixing dynamics using RGB color analysis. *Chemical Engineering Science*, 65(12), 3729–3738.
- Badgwell, T.A., Lee, J.H., and Liu, K.H. (2018). Reinforcement learning—overview of recent progress and implications for process control. 44, 71–85.
- Bajaj, I., Iyer, S.S., and Hasan, M.M.F. (2018). A trust region-based two phase algorithm for constrained black-box and grey-box optimization with infeasible initial point. *Computers & Chemical Engineering*, 116, 306–321.
- Barazandegan, M., Ekram, F., Kwok, E., Gopaluni, B., and Tulsyan, A. (2015). Assessment of type II diabetes mellitus using irregularly sampled measurements with missing data. *Bioprocess and biosystems engineering*, 38(4), 615–629.
- Bertsekas, D.P. (2005). *Dynamic programming and optimal control*, volume 1. Athena Scientific Belmont, MA.
- Bonassi, F., Terzi, E., Farina, M., and Scattolini, R. (2019). LSTM neural networks: Input to state stability and probabilistic safety verification. *arXiv preprint arXiv:1912.04377*.
- Boukouvala, F. and Floudas, C.A. (2017). ARGONAUT: Algorithms for Global Optimization of coNstrAined grey-box compUTational problems. *Optimization Letters*, 11(5), 895–913.
- Brown, B., Chui, M., and Manyika, J. (2011). Are you ready for the era of big data. *McKinsey Quarterly*, 4(1), 24–35.
- Bruno, A.E., Charbonneau, P., Newman, J., Snell, E.H., So, D.R., Vanhoucke, V., Watkins, C.J., Williams, S., and Wilson, J. (2018). Classification of crystallization outcomes using deep convolutional neural networks. *PLOS ONE*, 13(6).
- Buşoniu, L., de Bruin, T., Tolić, D., Kober, J., and Palunko, I. (2018). Reinforcement learning for control: Performance, stability, and deep approximators. *Annual Reviews in Control*, 8–28.
- Byrne, W. (1993). Generalization and maximum likelihood from small data sets. In *Proceedings of the IEEE Signal Processing Workshop*, 197–206.
- Caballero, J.A. and Grossmann, I.E. (2008). An algorithm for the use of surrogate models in modular flowsheet optimization. *AIChE Journal*, 54(10), 2633–2650.
- Carvalho, B.W.S.R., Civitarese, D., Szwarcman, D., Cavalin, P., Zadrozny, B., Moreno, M., and Marsden, S. (2019). Ore content estimation based on spatial geological data through 3D convolutional neural networks. In *81st EAGE Conference and Exhibition 2019 Workshop Programme*. EAGE Publications BV.
- Chachuat, B., Srinivasan, B., and Bonvin, D. (2009). Adaptation strategies for real-time optimization. *Computers & Chemical Engineering*, 33(10), 1557–1567.
- Chaffart, D. and Ricardez-Sandoval, L.A. (2018). Optimization and control of a thin film growth process: A hybrid first principles/artificial neural network based multiscale modelling approach. *Computers & Chemical Engineering*, 119, 465–479.
- Chen, L., Bernard, O., Bastin, G., and Angelov, P. (2000). Hybrid modelling of biotechnological processes using neural networks. *Control Engineering Practice*, 8(7), 821–827.
- Chiang, L.H., Russell, E.L., and Braatz, R.D. (2000). *Fault detection and diagnosis in industrial systems*. Springer Science & Business Media.
- Cozad, A., Sahinidis, N.V., and Miller, D.C. (2015). A combined first-principles and data-driven approach to model building. *Computers & Chemical Engineering*, 73, 116–127.
- Csáji, B.C. (2001). Approximation with artificial neural networks. *Faculty of Sciences, Eötvös Lornd University, Hungary*, 24, 48.
- Cui, Y., Zhu, L., Fujisaki, M., Kanokogi, H., and Matsubara, T. (2018). Factorial kernel dynamic policy programming for Vinyl Acetate monomer plant model control. In *2018 IEEE 14th International Conference on Automation Science and Engineering (CASE)*, 304–309.
- Deisenroth, M.P., Neumann, G., Peters, J., et al. (2013). A survey on policy search for robotics. *Foundations and Trends® in Robotics*, 2(1–2), 1–142.
- del Rio-Chanona, E.A., Graciano, J.E.A., Bradford, E., and Chachuat, B. (2019). Modifier-adaptation schemes employing Gaussian processes and trust regions for real-time optimization. *IFAC-PapersOnLine*, 52(1), 52–57.
- Du, Y. and Narasimhan, K. (2019). Task-agnostic dynamics priors for deep reinforcement learning. *arXiv preprint arXiv:1905.04819*.
- Eason, J.P. and Biegler, L.T. (2016). A trust region filter method for glass-box/black-box optimization. *AIChE Journal*, 62(9), 3124–3136.
- Eason, J.P. and Biegler, L.T. (2018). Advanced trust region optimization strategies for glass box/black box models. *AIChE Journal*, 64(11), 3934–3943.
- Eppel, S. and Kachman, T. (2014). Computer vision-based recognition of liquid surfaces and phase boundaries in transparent vessels, with emphasis on chemistry applications. *arXiv preprint arXiv:1404.7174*.
- Ertel, W. (2018). *Introduction to artificial intelligence*. Springer.
- Ferreira, T.D.A., Shukla, H.A., Faulwasser, T., Jones, C.N., and Bonvin, D. (2018). Real-time optimization of uncertain process systems via modifier adaptation and Gaussian processes. In *2018 European Control Conference (ECC)*, 465–470.
- Finn, C., Abbeel, P., and Levine, S. (2017). Model-agnostic meta-learning for fast adaptation of deep networks. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, 1126–1135.
- Flores-Cerrillo, J. and MacGregor, J.F. (2004). Multivariate monitoring of batch processes using batch-to-batch information. *AIChE Journal*, 50(6), 1219–1228.
- Fortuna, L., Graziani, S., Rizzo, A., and Xibilia, M.G. (2007). *Soft sensors for monitoring and control of industrial processes*. Springer Science & Business Media.
- Fu, Y. and Aldrich, C. (2018). Froth image analysis by use of transfer learning and convolutional neural networks. *Minerals Engineering*, 115, 68–78.
- Fujimoto, S., Van Hoof, H., and Meger, D. (2018). Addressing function approximation error in actor-critic methods. *arXiv preprint arXiv:1802.09477*.
- Gao, W., Wenzel, S., and Engell, S. (2016). A reliable modifier-adaptation strategy for real-time optimization. *Computers & Chemical Engineering*, 91, 318–328.

- Ge, Y., Li, S., and Chang, P. (2018). An approximate dynamic programming method for the optimal control of Alkali-Surfactant-Polymer flooding. *Journal of Process Control*, 64, 15–26.
- Ghosh, D., Hermonat, E., Mhaskar, P., Snowling, S., and Goel, R. (2019). Hybrid modeling approach integrating first-principles models with subspace identification. *Industrial & Engineering Chemistry Research*, 58(30), 13533–13543.
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. MIT press.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems*, 2672–2680.
- Gopakumar, V., Tiwari, S., and Rahman, I. (2018). A deep learning based data driven soft sensor for bioprocesses. *Biochemical engineering journal*, 136, 28–39.
- Haarnoja, T., Ha, S., Zhou, A., Tan, J., Tucker, G., and Levine, S. (2018a). Learning to walk via deep reinforcement learning. *arXiv preprint arXiv:1812.11103*.
- Haarnoja, T., Zhou, A., Abbeel, P., and Levine, S. (2018b). Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. *arXiv preprint arXiv:1801.01290*.
- Haarnoja, T., Zhou, A., Hartikainen, K., Tucker, G., Ha, S., Tan, J., Kumar, V., Zhu, H., Gupta, A., Abbeel, P., et al. (2018c). Soft actor-critic algorithms and applications. *arXiv preprint arXiv:1812.05905*.
- Henao, C.A. and Maravelias, C.T. (2011). Surrogate-based superstructure optimization framework. *AIChE Journal*, 57(5), 1216–1232.
- Henderson, P., Islam, R., Bachman, P., Pineau, J., Precup, d., and Meger, D. (2018). Deep reinforcement learning that matters. In *Thirty-Second AAAI Conference on Artificial Intelligence*.
- Hocenski, Z., Matic, T., and Vidovic, I. (2016). Technology transfer of computer vision defect detection to ceramic tiles industry. In *2016 IEEE International Conference on Smart Systems and Technologies (SST)*.
- Horn, Z., Auret, L., McCoy, J., Aldrich, C., and Herbst, B. (2017). Performance of convolutional neural networks for feature extraction in froth flotation sensing. *IFAC-PapersOnLine*, 50(2), 13–18.
- Iglesias, J.C.Á., Santos, R.B.M., and Paciornik, S. (2019). Deep learning discrimination of quartz and resin in optical microscopy images of minerals. *Minerals Engineering*, 138, 79–85.
- Ingrassia, S. and Morlini, I. (2005). Neural network modeling for small datasets. *Technometrics*, 47(3), 297–311.
- Jackson, P.C. (2019). *Introduction to artificial intelligence*. Courier Dover Publications.
- Jang, S.S. and Gopaluni, R.B. (2011). Parameter estimation in nonlinear chemical and biological processes with unmeasured variables from small data sets. *Chemical engineering science*, 66(12), 2774–2787.
- Janner, M., Fu, J., Zhang, M., and Levine, S. (2019). When to trust your model: Model-based policy optimization. In *Advances in Neural Information Processing Systems*, 12498–12509.
- Je, L.W., Na, J., Kim, K., Lee, C.J., Lee, Y., and Lee, J.M. (2018). NARX modeling for real-time optimization of air and gas compression systems in chemical processes. *Computers & Chemical Engineering*, 115, 262–274.
- Jiang, Y. and Jiang, Z.P. (2014). Robust adaptive dynamic programming and feedback stabilization of nonlinear systems. *IEEE Transactions on Neural Networks and Learning Systems*, 25(5), 882–893.
- Kaisare, N.S., Lee, J.M., and Lee, J.H. (2003). Simulation based strategy for nonlinear optimal control: Application to a microbial cell reactor. *International Journal of Robust and Nonlinear Control: IFAC-Affiliated Journal*, 13(3–4), 347–363.
- Kazemina, S., Baur, C., Kuijper, A., van Ginneken, B., Navab, N., Albarqouni, S., and Mukhopadhyay, A. (2018). GANs for medical image analysis. *arXiv preprint arXiv:1809.06222*.
- Keßler, T., Kunde, C., McBride, K., Mertens, N., Michaels, D., Sundmacher, K., and Kienle, A. (2019). Global optimization of distillation columns using explicit and implicit surrogate models. *Chemical Engineering Science*, 197, 235 – 245.
- Khatibisepehr, S., Huang, B., and Khare, S. (2013). Design of inferential sensors in the process industry: A review of bayesian methods. *Journal of Process Control*, 23(10), 1575–1596.
- Kholief, E.A., Darwish, S.H., and Fors, N. (2017). Detection of steel surface defect based on machine learning using deep auto-encoder network. *Industrial engineering and operations management*, 218–229.
- Kim, J.W., Choi, G.B., and Lee, J.M. (2018a). A POMDP framework for integrated scheduling of infrastructure maintenance and inspection. *Computers & Chemical Engineering*, 112, 239–252.
- Kim, J.W., Park, B.J., Yoo, H., Lee, J.H., and Lee, J.M. (2018b). Deep reinforcement learning based finite-horizon optimal tracking control for nonlinear system. *IFAC-PapersOnLine*, 51(25), 257–262.
- Kingma, D.P. and Welling, M. (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.
- Kistner, M., Jemwa, G.T., and Aldrich, C. (2013). Monitoring of mineral processing systems by using textural image analysis. *Minerals Engineering*, 52, 169–177.
- Klimasauskas, C.C. (1998). Hybrid modeling for robust nonlinear multivariable control. *ISA Transactions*, 37(4), 291–297.
- Konda, V.R. and Tsitsiklis, J.N. (2000). Actor-critic algorithms. In *Advances in neural information processing systems*, 1008–1014.
- Kong, X. and Li, J. (2018). Vision-based fatigue crack detection of steel structures using video feature tracking. *Computer-Aided Civil and Infrastructure Engineering*, 33(9), 783–799.
- Lee, J.H., Shin, J., and Realf, M.J. (2018). Machine learning: Overview of the recent progresses and implications for the process systems engineering field. *Computers & Chemical Engineering*, 114, 111–121.
- Lee, J.M., Kaisare, N.S., and Lee, J.H. (2006). Choice of approximator and design of penalty function for an approximate dynamic programming based control approach. *Journal of Process Control*, 16(2), 135–156.
- Lee, J.M. and Lee, J.H. (2009). An approximate dynamic programming based approach to dual adaptive control. *Journal of process control*, 19(5), 859–864.
- Lee, S.S. (2000). Noisy replication in skewed binary classification. *Computational Statistics & Data Analysis*, 34(2), 165–191.
- Levine, S., Finn, C., Darrell, T., and Abbeel, P. (2016). End-to-end training of deep visuomotor policies. *The Journal of Machine Learning Research*, 17(1), 1334–1373.
- Lewis, F.L. and Vrabie, D. (2009). Reinforcement learning and adaptive dynamic programming for feedback control. *IEEE Circuits and Systems Magazine*, 9(3).
- Li, B.h., Hou, B.c., Yu, W.t., Lu, X.b., and Yang, C.w. (2017). Applications of artificial intelligence in intelligent manufacturing: a review. *Frontiers of Information Technology & Electronic Engineering*, 18(1), 86–96.
- Li, D., Chen, D., Goh, J., and Ng, S.K. (2018). Anomaly detection with generative adversarial networks for multivariate time series. *arXiv preprint arXiv:1809.04758*.
- Li, S., Chen, J., and Xiang, J. (2019). Applications of deep convolutional neural networks in prospecting prediction based on two-dimensional geological big data. *Neural Computing and Applications*.
- Li, W., Yue, H.H., Valle-Cervantes, S., and Qin, S.J. (2000). Recursive PCA for adaptive process monitoring. *Journal of Process Control*, 10(5), 471–486.
- Lillicrap, T.P., Hunt, J.J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., and Wierstra, D. (2015). Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971*.
- Liu, Y., Yang, C., Liu, K., Chen, B., and Yao, Y. (2019). Domain adaptation transfer learning soft sensor for product quality prediction. *Chemometrics and Intelligent Laboratory Systems*, 192, 103813.
- Lu, N., Yang, Y., Wang, F., and Gao, F. (2004). A stage-based monitoring method for batch processes with limited reference

- data. *IFAC Proceedings Volumes*, 37(9), 787–792.
- Ly, F., Wen, C., Bao, Z., and Liu, M. (2016). Fault diagnosis based on deep learning. In *2016 American Control Conference (ACC)*, 6851–6856. IEEE.
- Maheswari, J.P. (2018). Breaking the curse of small datasets in machine learning: Part 1. *Towardsdatascience.com*.
- Mao, R., Zhu, H., Zhang, L., and Chen, A. (2006). A new method to assist small data set neural network learning. In *Proceedings of the 6th International Conference on Intelligent Systems Design and Applications*, 17–22.
- Marchetti, A., Chachuat, B., and Bonvin, D. (2009). Modifier-adaptation methodology for real-time optimization. *Industrial & Engineering Chemistry Research*, 48(13), 6022–6033.
- Masci, J., Meier, U., Ciresan, D., Schmidhuber, J., and Fricout, G. (2012). Steel defect classification with max-pooling convolutional neural networks. In *The 2012 IEEE International Joint Conference on Neural Networks (IJCNN)*.
- Mireles Gonzalez, J.I. (2018). *Deep Recurrent Neural Networks for Fault Detection and Classification*. Master’s thesis, University of Waterloo.
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., and Ostrovski, G. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529.
- Nichol, A., Achiam, J., and Schulman, J. (2018). On first-order meta-learning algorithms. *arXiv preprint arXiv:1803.02999*.
- Nilsson, N.J. (2014). *Principles of artificial intelligence*. Morgan Kaufmann.
- Noor, R.A.M., Ahmad, Z., Don, M.M., and Uzir, M.H. (2010). Modelling and control of different types of polymerization processes using neural networks technique: A review. *The Canadian Journal of Chemical Engineering*, 88(6), 1065–1084.
- Nosair, H., Yang, Y., and Lee, J.M. (2010). Min-max control using parametric approximate dynamic programming. *Control Engineering Practice*, 18(2), 190–197.
- O’Brien, M. and Cooper, D. (2015). Continuous flow liquid-liquid separation using a computer-vision control system: The bromination of enamines with n-bromosuccinimide. *Synlett*, 27(01), 164–168.
- Oniśko, A., Druzdzel, M.J., and Wasyluk, H. (2001). Learning Bayesian network parameters from small data sets: application of Noisy-OR gates. *International Journal of Approximate Reasoning*, 27(2), 165–182.
- OpenAI, Berner, C., Brockman, G., Chan, B., Cheung, V., Dbiak, P., Dennison, C., Farhi, D., Fischer, Q., Hashme, S., Hesse, C., Jzefowicz, R., Gray, S., Olsson, C., Pachoeki, J., Petrov, M., de Oliveira Pinto, H.P., Raiman, J., Salimans, T., Schlatter, J., Schneider, J., Sidor, S., Sutskever, I., Tang, J., Wolski, F., and Zhang, S. (2019). Dota 2 with large scale deep reinforcement learning. URL <https://arxiv.org/abs/1912.06680>.
- Oquab, M., Bottou, L., Laptev, I., and Sivic, J. (2014). Learning and transferring mid-level image representations using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1717–1724.
- Pan, S.J. and Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10), 1345–1359.
- Patel, A.K. and Chatterjee, S. (2016). Computer vision-based limestone rock-type classification using probabilistic neural network. *Geoscience Frontiers*, 7(1), 53–60.
- Patwardhan, R.S., Hamadah, H.A., Patel, K.M., Hafiz, R.H., and Al-Gwaiz, M.M. (2019). Applications of advanced analytics at Saudi Aramco: A practitioners perspective. *Industrial & Engineering Chemistry Research*.
- Perarnau, G., Van De Weijer, J., Raducanu, B., and Álvarez, J.M. (2016). Invertible conditional GANs for image editing. *arXiv preprint arXiv:1611.06355*.
- Poggio, T. and Vetter, T. (1992). Recognition and structure from one 2D model view: Observations on prototypes, object classes and symmetries. Technical report, DTIC Document.
- Poliakov, A. and Donskoi, E. (2014). Automated relief-based discrimination of non-opaque minerals in optical image analysis. *Minerals Engineering*, 55, 111–124.
- Prokhorov, D.V. and Wunsch, D.C. (1997). Adaptive critic designs. *IEEE transactions on Neural Networks*, 8(5), 997–1007.
- Psichogios, D.C. and Ungar, L.H. (1992). A hybrid neural network-first principles approach to process modeling. *AIChE Journal*, 38(10), 1499–1511.
- Puterman, M.L. (2014). *Markov Decision Processes.: Discrete Stochastic Dynamic Programming*. John Wiley & Sons.
- Qin, S.J. and Chiang, L.H. (2019). Advances and opportunities in machine learning for process data analytics. *Computers & Chemical Engineering*, 126, 465–473.
- Quirante, N., Javaloyes, J., and Caballero, J. (2015). Rigorous design of distillation columns using surrogate models based on kriging interpolation. *AIChE Journal*, 61(7), 2169–2187.
- Recht, B. (2019). A tour of reinforcement learning: The view from continuous control. *Annual Review of Control, Robotics, and Autonomous Systems*, 2, 253–279.
- Rodríguez-Pulido, F.J., Gordillo, B., González-Miret, M.L., and Heredia, F.J. (2013). Analysis of food appearance properties by computer vision applying ellipsoids to colour data. *Computers and Electronics in Agriculture*, 99, 108–115.
- Russell, S.J. and Norvig, P. (2016). *Artificial intelligence: A modern approach*. Malaysia; Pearson Education Limited,.
- Sarkate, R.S., Kalyankar, N.V., and Khanale, P.B. (2013). Application of computer vision and color image segmentation for yield prediction precision. In *IEEE International Conference on Information Systems and Computer Networks*.
- Schölkopf, B., Simard, P., Smola, A.J., and Vapnik, V. (1998). Prior knowledge in support vector kernels. 640–646.
- Schuppert, A. and Mrziglod, T. (2018). Hybrid model identification and discrimination with practical examples from the chemical industry. In J. Glassey and M. von Stosch (eds.), *Hybrid Modeling in Process Industries*, 63–88. CRC Press, Boca Raton.
- Schweidtmann, A.M. and Mitsos, A. (2019). Deterministic global optimization with artificial neural networks embedded. *Journal of Optimization Theory & Applications*, 180(3), 925–948.
- Shang, C., Yang, F., Huang, D., and Lyu, W. (2014). Data-driven soft sensor development based on deep learning technique. *Journal of Process Control*, 24(3), 223–233.
- Shin, H.C., Roth, H.R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., and Summers, R.M. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE transactions on medical imaging*, 35(5), 1285–1298.
- Shin, J., Badgwell, T.A., Liu, K.H., and Lee, J.H. (2019). Reinforcement Learning—Overview of recent progress and implications for process control. *Computers & Chemical Engineering*, 127, 282–294.
- Silver, D., Lever, G., Heess, N., Degris, T., Wierstra, D., and Riedmiller, M. (2014). Deterministic policy gradient algorithms.
- Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., et al. (2017). Mastering the game of Go without human knowledge. *Nature*, 550(7676), 354.
- Siswanto, J., Prabuwo, A.S., and Abdulah, A. (2013). Volume measurement of food product with irregular shape using computer vision and monte carlo method: A framework. *Procedia Technology*, 11, 764–770.
- Solle, D., Hitzmann, B., Herwig, C., Pereira Remelhe, M., Ulonska, S., Wuerth, L., Prata, A., and Steckenreiter, T. (2017). Between the poles of data-driven and mechanistic modeling for process operation. *Chemie Ingenieur Technik*, 89(5), 542–561.
- Spielberg, S., Tulsyan, A., Lawrence, N.P., Loewen, P.D., and Gopaluni, R.B. (2019). Towards self-driving processes: A deep reinforcement learning approach to control. *AIChE Journal*, 65(6), 1–20.
- Su, H.T., Bhat, N., Minderman, P.A., and McAvoy, T.J. (1992). Integrating neural networks with first principles models for dynamic

- modeling. *IFAC Proceedings Volumes*, 25(5), 327–332.
- Su, J.S., Zhang, B.F., and Xu, X. (2006). Advances in machine learning based text categorization. *Ruan Jian Xue Bao (Journal of Software)*, 17(9), 1848–1859.
- Sun, B., He, M., Wang, Y., Gui, W., Yang, C., and Zhu, Q. (2018). A data-driven optimal control approach for solution purification process. *Journal of Process Control*, 68, 171–185.
- Sutton, R.S. and Barto, A.G. (2018). *Reinforcement learning: An introduction*. MIT press.
- Theodorou, E., Buchli, J., and Schaal, S. (2010a). A generalized path integral control approach to reinforcement learning. *Journal of machine learning research*, 11(Nov), 3137–3181.
- Theodorou, E., Tassa, Y., and Todorov, E. (2010b). Stochastic differential dynamic programming. In *Proceedings of the 2010 American Control Conference*, 1125–1132. IEEE.
- Thompson, M.L. and Kramer, M.A. (1994). Modeling chemical processes using prior knowledge and neural networks. *AIChE Journal*, 40(8), 1328–1340.
- Tsai, T.I. and Li, D.C. (2008). Utilize bootstrap in small data set learning for pilot run modeling of manufacturing systems. *Expert Systems with Applications*, 35(3), 1293–1300.
- Tulsyan, A., Garvin, C., and Ündey, C. (2018). Advances in industrial biopharmaceutical batch process monitoring: Machine-learning methods for small data problems. *Biotechnology and Bioengineering*, 115(8), 1915–1924.
- Tulsyan, A., Garvin, C., and Undey, C. (2019a). Industrial batch process monitoring with limited data. *Journal of Process Control*, 77, 114–133.
- Tulsyan, A., Schorner, G., Khodabandehlou, H., Wang, T., Coufal, M., and Undey, C. (2019b). A machine learning approach to calibrate generic raman models for real-time monitoring of cell culture processes. *Biotechnology and Bioengineering*, 116(10), 2575–2586.
- Tulsyan, A., Wang, T., Schorner, G., Khodabandehlou, H., Coufal, M., and Undey, C. (2020). Automatic real-time calibration, assessment, and maintenance of generic raman models for online monitoring of cell culture processes. *Biotechnology and Bioengineering*, 117(2), 404–416.
- Tzafestas, S.G. and Verbruggen, H.B. (2012). *Artificial intelligence in industrial decision making, control and automation*, volume 14. Springer Science & Business Media.
- Ustundag, A. and Cevikcan, E. (2017). *Industry 4.0: managing the digital transformation*. Springer.
- Venkatasubramanian, V. (2019). The promise of artificial intelligence in chemical engineering: Is it here, finally? *AIChE Journal*, 65(2), 466–478.
- Vicente, A., Raveendran, R., Huang, B., Sedghi, S., Narang, A., Jiang, H., and Mitchell, W. (2019). Computer vision system for froth-middlings interface level detection in the primary separation vessels. *Computers & Chemical Engineering*, 123, 357–370.
- Villalba-Diez, J., Schmidt, D., Gevers, R., Ordieres-Meré, J., Buchwitz, M., and Wellbrock, W. (2019). Deep learning for industrial computer vision quality control in the printing industry 4.0. *Sensors*, 19(18), 3987.
- von Stosch, M., Oliveira, R., Peres, J., and Feyo de Azevedo, S. (2014). Hybrid semi-parametric modeling in process systems engineering: Past, present and future. *Computers & Chemical Engineering*, 60, 86 – 101.
- Wang, K., Forbes, M.G., Gopaluni, B., Chen, J., and Song, Z. (2019a). Systematic development of a new variational autoencoder model based on uncertain data for monitoring nonlinear processes. *IEEE Access*, 7, 22554–22565.
- Wang, K., Gopaluni, B., Chen, J., and Song, Z. (2018). Deep learning of complex batch process data and its application on quality prediction. *IEEE Transactions on Industrial Informatics*.
- Wang, K., Shang, C., Liu, L., Jiang, Y., Huang, D., and Yang, F. (2019b). Dynamic soft sensor development based on convolutional neural networks. *Industrial & Engineering Chemistry Research*.
- Wang, W.D. and Yang, J.Y. (2008). Quadratic discriminant analysis method based on virtual training samples. *Acta Automatica Sinica*, 34(34), 400–407.
- Wiese, M., Knobloch, R., Korn, R., and Kretschmer, P. (2019). Quant gans: Deep generation of financial time series. *arXiv preprint arXiv:1907.06673*.
- Williams, G., Drews, P., Goldfain, B., Rehg, J.M., and Theodorou, E.A. (2016). Aggressive driving with model predictive path integral control. In *IEEE International Conference on Robotics and Automation (ICRA)*, 1433–1440.
- Willis, M.J. and von Stosch, M. (2017). Simultaneous parameter identification and discrimination of the nonparametric structure of hybrid semi-parametric models. *Computers & Chemical Engineering*, 104, 366–376.
- Wilson, Z.T. and Sahinidis, N.V. (2017). The ALAMO approach to machine learning. *Computers & Chemical Engineering*, 106, 785–795.
- Wu, Z., Tran, A., Rincon, D., and Christofides, P.D. (2019). Machine learning-based predictive control of nonlinear processes. Part I: Theory. *AIChE Journal*, 65(11), e16729.
- Xu, R.w., He, L., Zhang, L.k., Tang, Z.y., and Tu, S. (2008). Research on virtual sample based identification of noise sources in ribbed cylindrical double-shells. *Journal of Vibration and Shock*, 5, 32–35.
- Yan, W., Tang, D., and Lin, Y. (2016). A data-driven soft sensor modeling method based on deep learning and its application. *IEEE Transactions on Industrial Electronics*, 64(5), 4237–4245.
- Yang, Y. and Kelly, J. (2019). Efficient real time optimization using a data-driven piecewise affine model. *Computers & Chemical Engineering*, 125, 545–557.
- Yang, Y. and Lee, J.M. (2010). Probabilistic modeling and dynamic optimization for performance improvement and risk management of plant-wide operation. *Computers & Chemical Engineering*, 34(4), 567–579.
- Yang, Y. and Lee, J.M. (2013). A switching robust model predictive control approach for nonlinear systems. *Journal of Process Control*, 23(6), 852–860.
- Yu, H., MacGregor, J.F., Haarsma, G., and Bourg, W. (2003). Digital imaging for online monitoring and control of industrial snack food processes. *Industrial & Engineering Chemistry Research*, 42(13), 3036–3044.
- Yuan, X., Huang, B., Wang, Y., Yang, C., and Gui, W. (2018). Deep learning-based feature representation and its application for soft sensor modeling with variable-wise weighted SAE. *IEEE Transactions on Industrial Informatics*, 14(7), 3235–3243.
- Zendehboudi, S., Rezaei, N., and Lohi, A. (2018). Applications of hybrid models in chemical, petroleum, and energy systems: A systematic review. *Applied Energy*, 228, 2539–2566.
- Zhang, D., Del Rio-Chanona, E.A., Petsagkourakis, P., and Wagner, J. (2019a). Hybrid physics-based and data-driven modeling for bioprocess online simulation and optimization. *Biotechnology & Bioengineering*, 116(11), 2919–2930.
- Zhang, L. and Gong-he, C. (2006). Method for constructing training data set in intrusion detection system. *Computer Engineering and Applications*, 28, 145–146.
- Zhang, Z., Wu, Z., Rincon, D., and Christofides, P.D. (2019b). Real-time optimization and control of nonlinear processes using machine learning. *Mathematics*, 7(10), 890.
- Zhao, C. (2014). Phase analysis and statistical modeling with limited batches for multimode and multiphase process monitoring. *Journal of Process Control*, 24(6), 856–870.
- Zhao, L., Chai, T., Wang, G., et al. (2005). Double moving window MPCA for online adaptive batch monitoring. *Chinese Journal of Chemical Engineering*, 13(5), 649.
- Zhao, Z., Tulsyan, A., Huang, B., and Liu, F. (2019). Estimation and identification in batch processes with particle filters. *Journal of Process Control*, 81, 1–14.
- Zhu, W., Ma, Y., Benton, M.G., Romagnoli, J.A., and Zhan, Y. (2018). Deep learning for pyrolysis reactor monitoring: From thermal imaging toward smart monitoring system. *AIChE Journal*, 65(2), 582–591.