



Review Article

Hybrid Modeling in Chemical Process Intensification: Architectures, Applications, and Critical Challenges

Pakiza Iman¹ and Noor Fatima²

¹School of Water and Environment, Chang'an University, China

²School of Biochemistry, Bahauddin Zakariya University, Pakistan

Article Information

Article History

Received: 16 June 2026

Revised: 28 July 2026

Accepted: 2 September 2026

Published online: 20 September 2026

Keywords

Hybrid Modeling

Process Intensification

Physics-Informed Machine Learning

Gray-Box Models

Reaction Engineering

Digital Twins

Correspondence*

pakizaiman98@gmail.com

ORCID

Pakiza Iman

<https://orcid.org/0009-0009-0091-4957>

Noor Fatima

<https://orcid.org/0009-0006-5482-4998>

Abstract

The hybrid models that use both first principles knowledge and machine learning: the first represents what is known about a process, while the second represents what is not. This review is based on the four domains of chemical process intensification (PI): structure, energy, synergy and time and discusses hybrid modelling for PI. Data demand, interpretability, extrapolation and training cost are used to define and compare serial, parallel and physics-constrained architectures. Soft sensing and control, CO₂ capture and power to X, design of intensified equipment and integrated reaction-separation units are considered in the applications. The four reaction-engineering case studies are then revisited using their original sources: methanol synthesis, coupled syngas and methanol synthesis and water-gas shift network, catalyzed polymerization, and oxidation of iso-octanol. The hybrid methanol model with the optimally designed experiments was found to have a test error of 8.49%, which is just slightly higher than the 9.06% of the recalibrated mechanistic model, and the mechanistic model with the experiments deleted had a test error of 13.20%. A thermodynamically consistent power-law backbone was found to have a lower test error (6.7%) than a Langmuir–Hinshelwood backbone (9.5%). The errors in polymerization predictions were decreased by 84%, 86% for the two dominant reagents for the two main cases, and for the surrogate-based hybrids the errors were reduced by more than 400 times when compared with simulation-based optimization and were run approximately 45 times faster than CFD. The key open challenges are the lack of data, particularly scarce and unbenchmarked data, coupled models trained together, guaranteed physical consistency, multiscale coupling, and uncertainty quantification. Type of synergy units have been the least researched. Priorities for future work include hybrid DT, closed-loop experimental design, and multiscale modelling.

© 2026 Centre for Research and Innovation (CRI). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

I. INTRODUCTION

The goal of process intensification (PI) is to find chemical processes that are significantly smaller, cleaner and more energy efficient than conventional plants [1]. Van Gerven and Stankiewicz defined PI in the following four aspects: structure (spatial), energy (thermodynamic), synergy (functional), and time (temporal) [2]. Reactive distillation, membrane reactors, microreactors, rotating packed beds, and hybrid separations are examples of technologies based on these concepts, and process systems engineering tools for the design of these technologies have been comprehensively discussed [3]. The incentive is large. Distillation is a thermally driven separation that consumes 10–15% of the world's energy [4] and, in fact, more than 150 industrial distillation units with capacities from 100 to 3000 kt/y have

been built for reactive distillation alone [5]. The integration that results in these benefits complicates the modelling of intensified units - as the reaction, transport and separation processes are integrated in one unit [6].

Throughout chemical engineering, models are used to design processes, optimize the operation of the process, and control the quality of the product. The construction of mechanistic models is derived from conservation laws, thermodynamics and rate expressions. Their terms are physical, but they are based on simplifying assumptions and have uncertain parameters [7]. For example, in catalytic reactors, the state of the active sites and its slow deactivation are still under debate even for a well-established process like the methanol synthesis over Cu/ZnO/Al₂O₃ [8,9] and errors in these sub-models compound over long prediction horizons. Data-

driven models are models that automatically learn the input–output relation from measurements. These are strong nonlinear and require large and informative datasets, violate mass and energy balances, and are hard to interpret [10, 11]. However, data of chemical engineering plants are generally limited and noisy, and deep learning is a technique that works well in the presence of much data. Chemical engineering data are often very limited in quantity and noisy, and deep learning is a technique that works well with a lot of data [12].

There is a mixture of both, which is known as hybrid modelling or gray box / semi-parametric modelling [7], [13]. A mechanistic part describes the known physics, while the data-driven part learns the unknown or too expensive to model parts [11]. The idea originated in the early nineties with the implementation of neural networks in process models [14] and [15]. Only recently, however, has the use of wider spreads become practical, due to the advent of better learning algorithms, cheaper computation and plant-wide data historians [16, 17]. Even though the data size and accuracy of the well-built hybrid model can be significantly less than that of the purely data driven model, as shown in Section IV, the magnitude of the benefit can vary widely when using this model [11], [18]. This method is not restricted to reaction engineering. For instance, in drilling engineering, physics-based hydraulic simulation has been adopted together with the response surface methodology to model the ECD of an extended-reach well [19].

Hybrid models are a logical place to apply intensified processes. The phenomena they exhibit are coupled and cannot be described by a simple mechanism; experimental data on the new equipment is too sparse to be used for a black-box model. There are general reviews of hybrid modeling [7, 13, 16, 20, 21] and a step-by-step implementation tutorial [11] as well as a review focused on PI [6]. The following is a third review of the present document. It is a structure that groups applications into the four PI domains. It is a re-examination of four case studies of reactions, using their main sources, and side by side comparisons of the quantitative results. And it relates the open methodological problems to the specific needs of intensified processes. Definitions, history and architectures are explored in Section II. Section III summarizes applications in the four domains, Section IV outlines case studies, Section V discusses unsolved challenges and Section VI suggests directions for the future. Section VII concludes.

II. FUNDAMENTALS OF HYBRID MODELING

A. Definitions and Terminology

Mechanistic (or white box) models are deductive models that can be generated from fundamental laws, where each term in

the model can be related to a physical cause. Data-driven, or black-box, models are constructed from observations and their predictions are as valid as the data they are fed with. Between the two are hybrid also known as gray-box models [7] and [13]. The predictions are close to physically plausible values in the mechanistic part, and more flexible, albeit a bit opaque, in the learned part, especially beyond the training domain. There are several terms that overlap. In hybrid semi-parametric modeling, both parametric (mechanistic) and nonparametric (learned) components are emphasized [7]. Typically, physics-informed machine learning requires that the governing equations are included in the training objective or the network architecture [22, 23]. The term physics-constrained machine learning is more general and applies to any technique that imposes any physical constraint, either softly as a penalty term, or exactly by enforcing the model structure [12]. The same family of methods have been applied with science-guided machine learning [20] and universal differential equations have embedded neural networks within differential equation models and trained them via the numerical solution [24]. All these are referred to as hybrid modeling in this review.

B. Historical Development

The mass balances in a fed-batch bioreactor were coupled with a neural network in order to estimate unmeasured process parameters by Psychosis and Ungar. They fed their hybrid interpolated and extrapolated more accurately than a black-box network and required fewer training examples [14]. Thompson and Kramer then introduced a methodology for using approximate first-principles models and known constraints within neural network models of chemical processes [15] and Oliveira extended the concepts into a methodology for integrating first principles with artificial neural networks [25]. Brendel and Marquardt discussed the design of experiments for the identification of hybrid reaction models from transient data [26].

However, for approximately 20 years, hybrid modelling was a niche area, and many applications were found in bioprocess engineering [7]. There has been very fast growth during the past decade [13] and [16]. It was a confluence of three things. Large plant datasets were provided by Distributed Control Systems and Data Historians. New tools were provided by both deep learning [22] and physics-informed neural networks [24]. Also, the increasing complexity of processes, such as intensified processes and multiscale processes, led to the need for models which none of the paradigms alone could provide [6], [17]. The most important contributions are summarized in Figure 1. Recent reviews and tutorials [6, 11, 13] and open-source tools like HybridML [27] have made it more accessible.

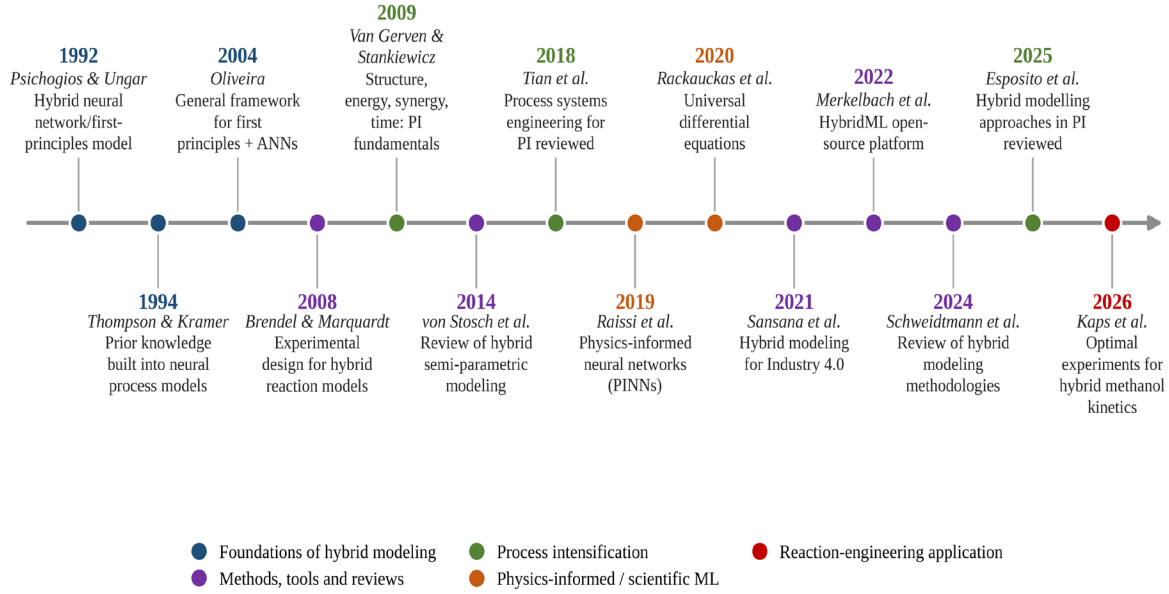


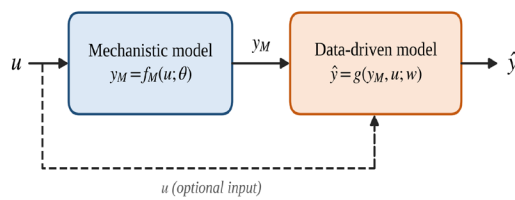
Fig.1 Selected Milestones in Hybrid Modeling for Chemical Engineering and Process Intensification, Compiled from the Primary Sources [2], [3], [6], [7], [13]–[16], [18], [22], [24]–[27]

C. Architectural Categories

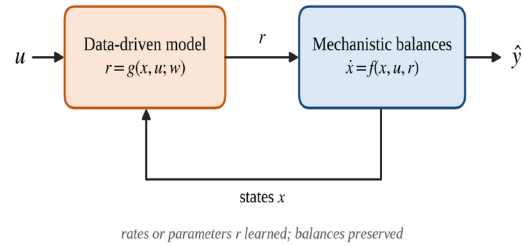
Typically, the hybrid models are divided depending on their combination of two parts [11, 13, 20]. The four structures

shown in Figure 2 are discussed below, where u are the inputs, x the states, θ the parameters of the mechanism, w the weights of the network, and y the outputs.

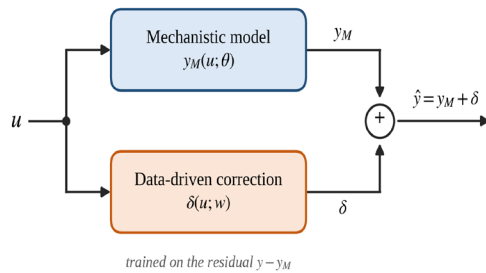
(a) Serial: mechanistic $\rightarrow$ data-driven



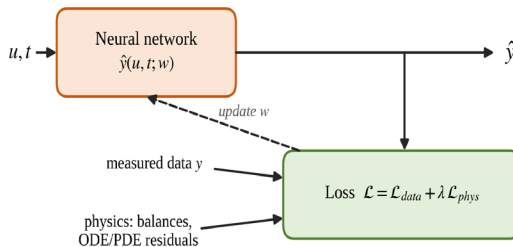
(b) Serial (nested): data-driven $\rightarrow$ mechanistic



(c) Parallel: residual correction



(d) Physics-constrained learning



Legend: Mechanistic component (blue), Data-driven component (orange), Physics-based loss (green).

Fig.2 Hybrid model architectures. (a) Serial structure in which the mechanistic output feeds a data-driven model, see (1). (b) Nested serial structure in which a data-driven model supplies rates or parameters to the mechanistic balances, see (2). (c) Parallel residual structure, see (3). (d) Physics-constrained learning, in which physics residuals enter the training loss, see (4).

1. Serial Architectures: One component supplies energy to the next one, also known as a serial structure. In the mechanistic model first case (Figure 2a), the output of the model is used as another input, or feature, to a data-driven model that corrects or extends the mechanistic model: comes an input, or feature, for a data-driven model that corrects or extends it:

$$\hat{y} = g(y_M, u; w), \quad y_M = f_M(u; \theta) \quad (1)$$

It is more frequently seen in reaction engineering in the reverse order (Figure 2b). Here the network provides quantities that are required by the mechanistic model but are not well predicted by it, such as rate constants, reaction rates, transport coefficients etc., and the balances are then integrated:

$$dx/dt = f(x, u, r), \quad r = g(x, u; w) \quad (2)$$

Mass conservation is satisfied by construction since the balances in f are maintained. This nested form dates to the first hybrid bioreactor models [14] and it is still the most popular one, when the physics is believed to be certain but one sub-model is not. For example, Kaps *et al.* replaced only the functions that correlate the catalyst oxidation state with each of the reaction rates in a methanol synthesis model [18] and Sorourifar *et al.* applied physics-enhanced neural ordinary differential equations (PODEs) to industrial reaction systems [28]. The gradients are required through the numerical solution of (2) which is more challenging than the fitting of a static correction (Section V-B).

2. Parallel Architectures: In a parallel structure the two components run side by side and their outputs are combined, usually by addition (Figure 2c):

$$\hat{y} = y_M(u; \theta) + \delta(u; w) \quad (3)$$

This form of network is trained with the difference between plant data and mechanistic prediction and is therefore referred to as a residual or delta model. It is appropriate for those situations in which the dominant behavior is accounted for by the mechanistic model, but the model contains systematic errors. A subspace identification approach for modelling this residual was employed by Ghosh *et al.* for a continuous stirred tank reactor [29] and a residual hybrid was used by Sharma and Liu for the prediction of melt-index in an industrial polyethylene process [20]. It is easy to build parallel models since the two parts can be recognised separately. The additive correction is not constrained by the balance equations, however, and can predict deviations from the balance equations outside of the training range.

3. Physics-Constrained Architectures: Physics-constrained models embed physical laws in the learning problem itself (Figure 2d). In physics-informed neural networks (PINNs), residuals of the governing equations, evaluated at collocation points, are added to the data misfit [22].

$$L(w) = L_{\text{data}}(w) + \lambda L_{\text{phys}}(w) \quad (4)$$

where the physics term is weighted by λ . The architecture can also be imposed, for instance by a learned soft plus output layer which enforces positive learned rate factors [18] or by simulated data augmentation [23]. Hard constraints of the form (4) ensure exact conservation but come at the cost of a more complicated training problem [12]. PINNs are appealing when information in the dataset is limited. Under data-scarce conditions, Nasiruddin *et al.* employed one PINN in forward and inverse modeling to predict concentration profiles in a flow reactor and estimate the rate constants, respectively, with errors below 2% [30]. Velioglu *et al.* demonstrated the ability of PINNs to enable dynamic process operation in the cases of limited physical knowledge and data [31] and Wu *et al.* summarized physics-informed learning for multiphysics problems in the field of chemical engineering [32].

TABLE I QUALITATIVE COMPARISON OF HYBRID MODEL ARCHITECTURES

Architecture	How physics enters	Data demand	Interpretability / extrapolation	Training cost	Typical use and examples
Serial: mechanistic → data-driven	Mechanistic output is an input to a network that corrects or extends it	Medium	Medium / moderate; the correction is not bound by the balances	Low	Correcting a trusted but biased model output [11]
Serial (nested): data-driven → mechanistic	Network supplies rates, parameters, or closures inside the balance equations	Low to medium	High / good; balances hold by construction	Medium to high (training through ODE/DAE solvers)	Uncertain kinetics and catalyst dynamics [14], [18], [33]
Parallel (residual)	Network adds a correction δ to the mechanistic prediction	Medium to high	Medium / limited outside the training range	Low (parts identified separately)	Systematic plant-model mismatch [20], [29]
Physics-constrained (PINN, hard-constrained networks)	Equation residuals in the loss (soft) or in the architecture (hard)	Low (labeled data)	Medium to high / good within the validity of the embedded physics	High (many collocation points)	Scarce data, inverse problems, safety-critical use [12], [22], [30]

4. *Comparison of Architectures*: A comparison of the four structures is given in Table I. The ratings are merely tendencies that are based on the studies cited and are not rules; the actual performance depends on the amount of the system that the mechanistic part explains, and on the quality of the data [11, 13]. Generally, serial models require less data than parallel models since the mechanistic part determines the dominant structure. Physics-constrained models learn from few labeled data, as the physics residual regularizes training, but are more expensive to train because the physics residual has to be evaluated at many collocation points [12, 22]. The choice should be based on the modeling question: in a modeling question, should it be: if there is a known sub-model, but it is still uncertain, then it should be a nested serial model; if there is an existing mechanical model, but it is still biased, then it should be a parallel model; if there is little data and physical consistency is important, then it should be a physics-constrained model.

III. HYBRID MODELING ACROSS THE FOUR DOMAINS OF PROCESS INTENSIFICATION

Figure 3 maps representative applications onto the four domains defined by Van Gerven and Stankiewicz [2]. The evidence base is uneven: it is broad for time- and energy-related applications and thin for synergy.

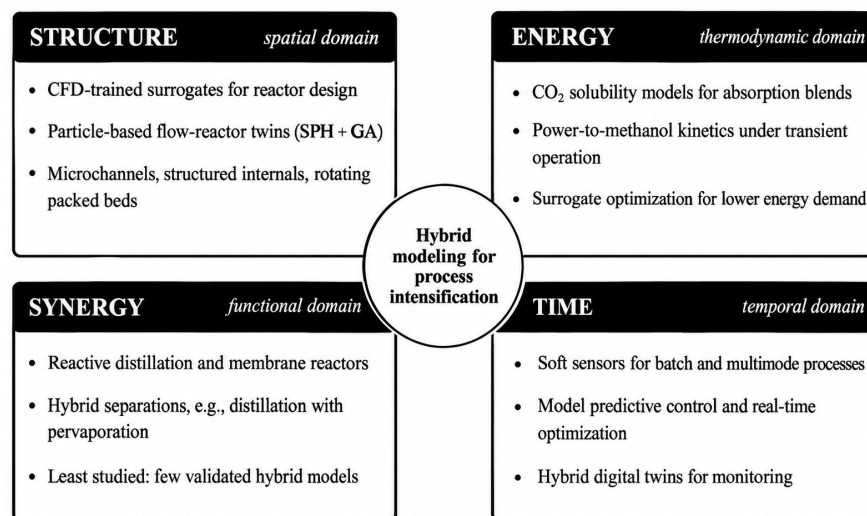


Fig.3 The Four Domains of Process Intensification Defined by Van Gerven and Stankiewicz [2], with Representative Hybrid Modeling Applications Reviewed in Section III

A present example is power-to-methanol reactors, which are used to follow a varying renewable hydrogen supply. Keßler *et al.* have developed a controller for a multistage power-to-methanol reactor with two degrees of freedom [38] which is discussed at length in Section IV-A in the context of the hybrid kinetic model [18] required to operate this type of reactor. The same applies to equipment degradation—there may be a mechanistic model of, say, heat exchanger fouling, and a learned estimate of the fouling rate based on the periodic measurements to assist with the maintenance planning.

A. Time

The time domain relates to the time scales of processes, such as, faster dynamics, shorter batch and residence times, and better utilization of operating time [2]. Hybrid models can make a significant contribution in the area of soft sensors, model predictive control (MPC) and real-time optimization (RTO). Soft sensors are able to estimate variables that cannot be measured online, like product quality, from variables that are measured online, regularly and can be built using data from the plant. Data-driven soft sensors are widely used in industry because they can be developed from the data, which are routinely recorded from the plant [34]. However, conventional soft sensors are prone to failure in batch processes when having to deal with nonlinearity and changes between and within batches [35]. When surveying industrial applications, both limited interpretability and limited efficiency were found to be common pitfalls of data-driven approaches, and a unifying approach was suggested to be "hybrid modeling" incorporating domain knowledge [36]. The model should be correct over the operating range, and efficiently executable to support multiple optimizations runs. In this setting, Wu *et al.* developed an ML model of the process for MPC [37] and hybrid models that maintained the balances and learned the uncertain rates were suitable [11].

B. Energy

The energy domain is related to the energy supplied to and utilized in a process [2]. The largest target is the separations, which represent 10–15% of world energy consumption, mainly thermally driven separations. Numerous thermodynamic models fail to be precise for strong non-ideal mixtures, and here, hybrid models come in handy. Capture of CO₂ by absorption is a key case. Since 1930 CO₂ separation from natural gas and hydrogen has been achieved using amine scrubbing. The minimum energy requirement for separation and compression of CO₂ to 150 bar for coal flue

gas is 0.11 MWh/t of CO₂, and further process and solvent optimization is anticipated to reduce actual consumption to approximately 0.2 MWh/t of CO₂ [39]. Solvent screening relies on solubility models that can be enhanced with the use of hybrid methods. Di Caprio *et al.* studied a fully data-driven model against a physics-informed hybrid model that splits the absorption from the physics part and a physics part with Henry's law. The hybrid separated well the absorbent mixtures and remained accurate even with those that were not well represented in the data set [40]. The same group developed a hybrid model of a batch separation process and compared the serial and parallel training methods [41]. Power to X processes use renewable electricity for production of fuels and chemicals. The synthesis reactor works in dynamic mode in case of intermittent hydrogen supply where methanol from CO₂, CO, and green hydrogen is a leading route. These are among the least certain aspects of the kinetics and the commercial {Cu} / {ZnO} / {Al}_2\text{O}_3 catalyst is known to undergo reversible oxidation [8]. The learning of this part by transient data will be explored in hybrid models (Section IV-A). Process-level surrogates can also be used to support energy-oriented design. Zhou *et al.* created a database using an Aspen Plus model, trained its deep learning surrogates and applied them to multi-objective optimization for the oxidation of iso-octanol to iso-octanal. The surrogate-based optimization was more than 400 times faster than the simulation-based optimization and the optimized conditions reduced both primary energy demand and greenhouse gas emissions [43].

C. Structure

The structure domain is related to geometric shapes and the spatial arrangement of equipment such as structured catalysts and microchannels or rotating packed beds [2]. These types of devices are typically designed using computational fluid dynamics (CFD), which is too slow to use for a large design space exploration or for on-line use. Hybrid surrogates fill this gap by being trained using CFD data. In Zhou *et al.*, CFD simulations were used as high-fidelity training data for deep learning models in an iso-octanol oxidation reactor. The hybrid predicted isooctane conversion is approximately 360 times that of instrumental analysis and its prediction of the concentration distribution of the predicted product, sodium is octanoate, is approximately 45 times more than CFD [44].

In a series hybrid, Nasruddin *et al.* integrated the SPH and genetic algorithm tuning and obtained the R² value up to 0.91 for a flow reactor to resolve the behavior of spatial mixing [30]. In a similar vein, Boy *et al.* explored the potential for supporting the intensification of equipment structures using artificial intelligence [45]. Rotating packed beds increase interphase mass transfer and now, complex reactor internals with additive manufacturing. Their hydrodynamics are both difficult to describe mechanistically, a clear sign that they are good candidates to be CFD-trained hybrids. In all cases the limitation is the same: a surrogate can only be relied upon within its training domain in which its predictions are

accurate, and predictions for new geometries require uncertainty estimates (Section V-E).

D. Synergy

The synergy domain is the domain when two or more functions are combined in a single unit; for instance, reaction and separation [2]. The most common example is Reactive Distillation, there are over 150 industrial units in operation, which were scaled up from pilot plants by either simply replicating the conditions, or by modelling [5]. This also applies to membrane reactors, where a product is removed to influence the direction of an equilibrium-limited reaction, and hybrid separation processes, like distillation/pervaporation [3]. These units cannot be modeled easily because a combination of kinetics, phase equilibrium and mass transfer occur in the same apparatus. A hybrid model would retain the description of the well-known parts of the model (e.g., phase equilibrium) and parameterize the coupled and poorly known parts (e.g., effective interfacial mass transfer, catalyst activity under reactive conditions). Synergy is the least advanced field in practice. The few studies found that validated hybrid models on integrated reaction–separation units exist, and there is no consensus regarding which part of the model needs to be mechanistic versus data-driven when phenomena are closely coupled. This is the most obvious methodological shift point of hybrid modeling and PI.

IV. CASE STUDIES IN REACTION ENGINEERING

Four case studies show how the architectures of Section II behave on reaction systems relevant to PI. Each was checked against its primary source. Table II and Figure 4 compare the quantitative outcomes.

A. Methanol Synthesis Under Dynamic Catalyst States

This involves the synthesis of methanol under dynamic catalyst states. This includes methanol synthesis under dynamic catalyst states. Flexible power-to-methanol operation requires changing reversible catalyst state [8] and [18]. Based on a kinetic model for CO hydrogenation and the reverse water–gas shift reaction, which incorporates a state variable for the degree of catalyst reduction [18, 42] Kaps *et al.* set up their model. The experiments were conducted in a Berty reactor, which was equipped with a commercial CuO/ZnO/Al₂O₃ catalyst (4.3 g) up to 260 °C and 60 bar with an online micro-gas chromatography every ~90 s. The authors first calibrated the mechanistic model with optimal experimental design (OED), which was used for conducting four informative experiments. They then substituted the heuristic functions that correlate the state of the catalyst with each of the reaction rates with a feed-forward neural network. Only these scaling factors were learned, with the mass balances remaining intact; it is a nested serial hybrid (Figure 2b) in terms of Section II. A total of 90 candidate networks were screened; the two selected had 88 trainable parameters; and an additional experiment was designed specifically for

the hybrid network [18]. The data set consisted of 44 experiments with a ratio of training and testing of 80/20. The mean absolute percentage error (MAPE) for the test-set averaged over species was 9.06% for the mechanistic model and 8.49% for the hybrid model, respectively, when optimally designed experiments were considered. Without them, the errors rose to 13.20% and 11.30% (Figure 4a). For methanol alone, the mechanistic test MAPE increased from 12.7% to 32.9% when those experiments were left out [18].

Two lessons follow. The informative data was more important than the class of the model, and the hybrid model was found to be beneficial for the testing of this mechanism: the learned network was able to mostly reproduce the existing heuristics for CO₂ hydrogenation and for the reverse water–gas shift. The authors conclude that the slight increment in accuracy is due to the fact that they only learned a small sub-model with the mechanistic parameters held fixed, and one of the two networks tested was found to be numerically unstable during training [18]. The choice of the mechanistic backbone covers the two reactions in which syngas undergoes an indirect reaction

B. Syngas Methanol and Water–Gas Shift: Choice of Mechanistic Backbone

Kay *et al.* studied the effect of the mechanistic backbone on a hybrid kinetic model [33] and [46]. The case of methanol synthesis from syngas and reverse water–gas shift over Cu/ZnO/Al₂O₃ in fixed-bed integral microreactors was studied, using the data reported by Wilkinson *et al.* [47]. The data consisted of 38 experiments with five species measured at four locations down the reactor, and at CO:CO₂ feed ratios of 3:3 to 3:0.3, reactor pressures of 15–35 bar and temperatures of 453–493 K [46]. Two hybrids were constructed: one Langmuir–Hinshelwood backbone and one power-law backbone with the modification to ensure thermodynamic equilibrium. The first step in parameter estimation was to divide the kinetic parameters into two groups: condition-invariant and condition-dependent. The operating conditions were then used in a neural network which was tuned using Bayesian optimization to predict the condition-dependent parameters, with uncertainty bands obtained through bootstrapping [46].

The Langmuir–Hinshelwood hybrid provided training and test errors of 3.6% and 9.5% and an average relative percentage uncertainty of 17% for 16 training and 22 test experiments. The power-law hybrid gave 3.7% and 6.7%, with an uncertainty of 13.0% (Figure 4b) [46]. In addition to that, the power-law version was more physical. Its learned rate constants obeyed the Arrhenius law without being imposed on, while the apparent rate constants of the Langmuir–Hinshelwood version did not and did not obey the equilibrium relation. Since the excess of hydrogen, the Langmuir–Hinshelwood model simplified to a power-law form to help explain the result [46]. A simpler backbone which honors thermodynamics can then outperform a more detailed, but less identifiable, backbone.

C. Catalyzed Polymerization Reactors

The first-principles models are inaccurate in capturing the strong nonlinear behaviour of polymerization reactors. In the early work, the gel effect of batch methyl methacrylate polymerization was modeled using stacked recurrent neural networks along with a simplified mechanistic model to control the polymerization process optimally [48]. In the more recent literature, Sobhani *et al.* proposed a model based on a first principles reactor model, a multilayer perceptron (MLP) and a linear regression (LR) model for predicting the concentration of main reagents and side products during the polymerization process in a catalyzed process [49]. In cases where the results were far from the first principles prediction, the hybrid decreased the mean absolute error (MAE) for the two major reagents by 84% and 86% (Figure 4d) and maximum error of prediction for the main and side products was 3.5%. Predictions remained acceptable when the data were corrupted with noise from 20% to the reaction rates and 5% to the temperatures, and the authors suggest that the method was facilitated to be transferred from one reactor to another with the help of a simple mechanistic part [49]. The selection of network structure, as well as the extent to which the learned correction can be relied upon beyond the circumstances of the training, remain questions.

D. Iso-Octanol Oxidation: Simulation-Trained Surrogates

Experiments are slow, and the data for iso-octanol are limited by hydrogen safety, making selective oxidation of alcohols an important process for chemical production [43]. Deep learning models were trained with the reaction mechanism-based CFD simulation of the oxidation reactor to produce training data for the models. This hybrid predicted isooctanol conversion was ~360 times faster than instrumental analysis and the concentration field of sodium isooctanoate was ~45 times faster than CFD (Figure 4c) [44]. When conducting a follow-up study on oxidation to iso-octanal, the same group trained surrogates on an Aspen Plus database that was constructed using a mechanistic model of mass, heat and momentum transfer and reaction and used them for multi-objective optimization which was more than 400 times faster than using the simulation to perform optimization [43]. The mechanistic model in both studies serves as a data generator, a typical approach when experiments cost a lot of money. The common problem is they're both only accurate within the sampled domain.

E. Cross-Case Comparison

The studies are summarized in Table II and in Figure 4. Three patterns are evident. In the cases where the mechanistic model was already well characterized with informative data (methanol), the hybrid model was not much more accurate and was primarily used for testing one uncertain sub-model. In the polymerization case, the learned correction was able to substantially reduce the errors where the mechanistic model had large structural errors. Mechanistic model accuracy did not seem to be advantageous, however, in cases where the

model was correct but slow, such as CFD and flowsheet simulation for iso-octanol oxidation, the primary advantage being speed, ranging from approximately 45 to over 400 times faster. Only some of the studies have explicitly

quantified the uncertainty, such as by bootstrapping [46] and the evidence is mainly from laboratory experiments or from simulation-generated data. There are still not many demonstrations of industrial scale intensified units.

TABLE II REACTION-ENGINEERING CASE STUDIES RE-EXAMINED AGAINST THEIR PRIMARY SOURCES

Case Study	Hybrid Structure	Data Basis	Main Quantitative Result	Main Limitation	Ref.
Methanol synthesis under dynamic catalyst states (Cu/ZnO/Al ₂ O ₃)	Nested serial: network replaces the catalyst-state effect on rates; balances kept	44 transient experiments, Berty reactor, ≤ 260 °C, ≤ 60 bar; 80/20 split	Test MAPE 8.49% (hybrid, OED) vs. 9.06% (mechanistic, OED); 13.20% for mechanistic without OED	Small gain over the recalibrated model; one candidate network unstable in training	[18]
Syngas methanol with reverse water–gas shift	Serial: ANN predicts condition-dependent kinetic parameters; LH or power-law backbone; bootstrapping	38 experiments, 453–493 K, 15–35 bar; 16 training / 22 test	Test error 6.7% (power law) vs. 9.5% (LH); mean relative uncertainty 13.0% vs. 17%	Microreactor data only; LH constants physically inconsistent	[33], [46]
Catalyzed polymerization	First-principles model combined with MLP and linear regression	Reactor experiments, including strongly deviating runs	MAE for the two main reagents reduced by 84% and 86%; maximum error 3.5%	Network selection and extrapolation not settled	[49]
Iso-octanol oxidation reactor	CFD results used to train a deep learning surrogate	CFD simulations based on the reaction mechanism	$\approx 360\times$ faster than instrumental analysis (conversion); $\approx 45\times$ faster than CFD (concentration field)	Valid only inside the sampled domain	[44]
Iso-octanol to iso-octanal process	Process model generates data for deep learning surrogates used in multi-objective optimization	Aspen Plus-generated database	$>400\times$ faster than simulation-based optimization; lower primary energy demand and GHG emissions	Accuracy bounded by the process model	[43]
Flow-reactor digital twin	Parallel (PINN) and series (SPH + GA) hybrids compared	Empirical plug-flow reactor data	PINN errors below 2% under data scarcity; series model R ² up to 0.91	Complementary strengths; neither dominates	[30]

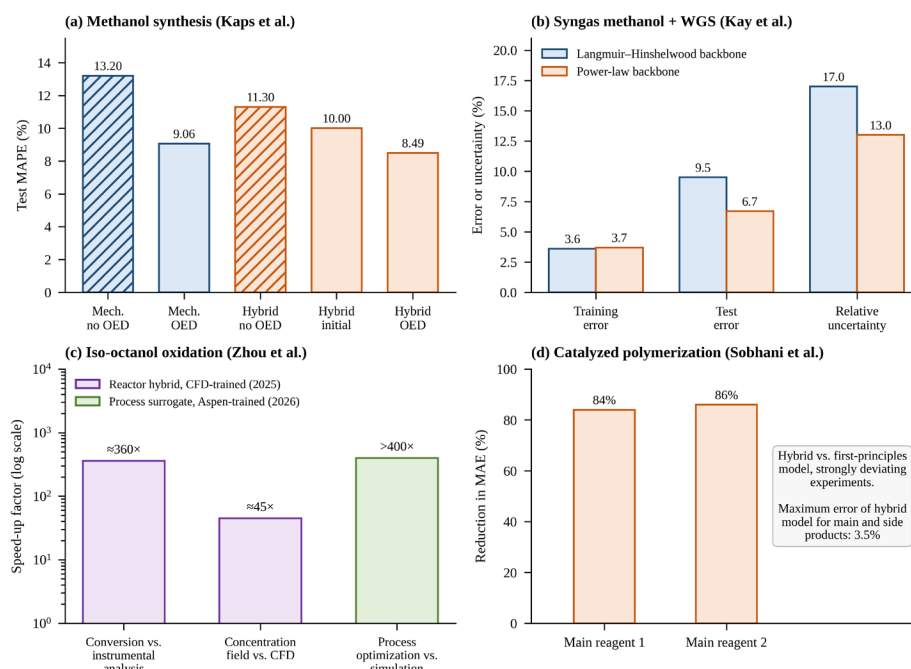


Fig.4 Quantitative outcomes of the reaction-engineering case studies, re-plotted from the primary sources. (a) Test-set MAPE averaged over species for methanol synthesis models trained with and without optimally designed experiments (OED); “initial” is the best hybrid architecture before the hybrid-specific design [18]. (b) Training error, test error, and average relative percentage uncertainty for hybrid models with Langmuir–Hinshelwood and power-law backbones [46]. (c) Speed-up of iso-octanol oxidation hybrids relative to instrumental analysis, CFD, and simulation-based optimization [43], [44]. (d) Reduction in MAE for the two main reagents in catalyzed polymerization relative to the first-principles model, in experiments with strongly deviating outcomes [49].

V. CRITICAL CHALLENGES AND RESEARCH GAPS

The case studies point to five recurring problems. Table III summarizes the evidence, the main gaps, and the directions suggested by the literature.

A. Data Quality and Quantity

Some models are hybrid models, which require less data than black-box models, but the data is still useful. Plant data are frequently sparse, noisy, and clustered at a small number of operating points, and therefore only excite a small portion of the plant dynamics [10, 12]. This is illustrated by the methanol study: when the optimal experimental design was not included, the test MAPE of the mechanistic model increased from 9.06% to 13.20% [18].

Conducting experiments for hybrid models is difficult compared to mechanistic models, since there are many weights in the network and these weights cannot be uniquely identified and the Fisher information matrix is ill-conditioned. Kaps *et al.* addressed this by constructing experiments in a reduced parameter space based on a singular value decomposition (SVD) [18, 50] to earlier experimental design research on reaction models in a hybrid manner [26]. There is a greater issue of a lack of shared benchmarks. There were few studies that would report on the amount of data required to obtain a certain level of accuracy, making comparison difficult.

B. Training and Validation

Training coupled models is more difficult than training either of the individual models. The gradients should have to cross the numerical solution of the mechanistic equations, the two components should be of different order of magnitude and the optimization problem may have poor conditioning properties. Popular techniques include sequential estimation of the mechanistic parameters followed by their fixation and simultaneous estimation of all the parameters with regularization [13, 18].

Incorporating deeper networks in the hybrid bioreactor models have led to enhanced predictive capabilities [51], and a hybrid model has been more effective than an established kinetic model in the batch fermentation bioreactor [52] but stable training is not guaranteed. Even when using adaptive DAE solvers using L1 regularization, Kaps *et al.* found that one of the two candidate networks became unstable on training [18] and propose several remedies. Validation must also be carefully prepared. Standard cross-validation is performed with independent samples, which is not true for time series data in dynamic processes, and it does not provide any indication of extrapolation. This means that data should be segmented by experiment/campaign, as done in the stratified experiment-wise split in Kaps *et al.* [18] and test sets should contain deliberately out-of-range conditions.

C. Physical Consistency

If it is used in a safety critical application, the hybrid model should be based on conservation laws and physical limits. Unlike classical PINNs, it is possible for a trained model to violate mass balance or have negative concentrations, because only violations are penalized by soft constraints. There are hard constraints that require exact satisfaction: they include structures which enforce mass conservation at construction time, e.g., the nested hybrids that learn only rate factors inside the balances [18]; output transformations that maintain quantities positive; hard constrained network formulations [12]. Thermodynamic consistency is also important. Kay *et al.* modified their power-law backbone to be consistent with the condition of equilibrium between the forward and reverse rates, and the Langmuir–Hinshelwood hybrid that did not satisfy this condition yielded physically inconsistent constants [46]. There is still a need for methods that ensure, not necessarily promote, consistency, better inequality constraint handling and tests to verify consistency across the entire operating domain.

D. Scalability and Multiscale Coupling

The training cost can rapidly increase as a function of the size of the mechanistic model, and as a function of the number of learned parameters; expensive full CFD or flowsheet models are sometimes too costly to be embedded in a training loop. The challenge becomes more complex for multiscale systems as each scale might require a different modeling method, and the couplings between scales are also not known [12]. Hybrid approaches have been suggested in which the expensive material properties are learned while the process level properties remain mechanistic [53] for integrated material and process design. Another answer is by using simulators (Section IV-D) trained by simulation data, and another is to use simulators or reduced-order models (an ROM) that are differentiable. The methods of coupling hybrid models across scales and of training them over distributed computing systems are still in a fledgling stage of development. As for equipment, scale-up of intensified equipment is itself a known hurdle in deployment [54] and models that scale-up from lab to plant scale are a part of that issue.

E. Uncertainty Quantification

There are three sources of uncertainty in a hybrid model: uncertainty of the mechanistic parameters, uncertainty of the learned component, and noise and bias in the data. It is still an open problem to combine them rigorously [12]. Bayesian neural networks, ensembles and bootstrapping, and Gaussian processes are the main tools. Kay *et al.* developed bootstrapped networks to generate prediction bands and found average relative uncertainties of 13 – 17 % on test data [46]. Lacking are techniques to propagate uncertainty through the coupled model, consensus-proven metrics to evaluate uncertainty estimates, and rules to apply the estimates in decision making, such as determining when the digital twin's predictions are no longer reliable.

TABLE III OPEN CHALLENGES, EVIDENCE FROM THE REVIEWED STUDIES, AND SUGGESTED DIRECTIONS

Challenge	Evidence from reviewed studies	Main gap	Suggested direction
Data quality and quantity	Excluding optimally designed experiments raised the mechanistic test MAPE from 9.06% to 13.20% [18]	No benchmarks for data efficiency; OED for many network weights is ill-conditioned	Report data requirements; reduced-space OED [50]; shared benchmark datasets
Training and validation	One of two candidate hybrid networks became unstable when trained through a DAE solver [18]	Stable joint training; validation that detects extrapolation	Sequential then joint training; multiple shooting or collocation; experiment-wise data splits
Physical consistency	A Langmuir–Hinshelwood hybrid produced rate constants inconsistent with equilibrium [46]	Guaranteed rather than penalized constraints; inequality constraints	Hard-constrained architectures; nested structures that keep the balances [12]
Scalability and multiscale coupling	Surrogates replace CFD and flowsheet models only within sampled domains [43], [44]	Coupling hybrids across scales; cost of embedding large simulators	Multiscale hybrid frameworks [53]; differentiable and reduced-order simulators
Uncertainty quantification	Bootstrapped hybrids gave mean relative uncertainties of 13–17% [46]	Combining parametric, structural, and data uncertainty; agreed metrics	Bayesian and ensemble methods linked to decisions in digital twins

VI. FUTURE OUTLOOK

A. Hybrid Digital Twins

A digital twin is a virtual version of a physical asset that is continually updated with data from the plant to monitor, optimize and make decisions. Digital twins are well suited to hybrid models, as the mechanistic portion of the model is interpretable and can offer the ability to extrapolate, and the learned portion is able to adjust to drift [17]. For a twin flow reactor, Nasruddin *et al.* have compared parallel and series hybrid. The two models were complementary: The parallel, PINN-based model had high accuracy with low computation cost, and the particle-based model is suitable to solve spatial dynamics and to characterize the system [30]. Practical barriers remain. There is the need for efficient learning algorithms for online updating of hybrid models, integration with existing control systems is not trivial, and connected twins will pose cybersecurity and data-governance questions.

B. Closed-Loop Experimental Design

Closed-loop experimental design: The model suggests an experiment; the outcome is fed back into the model and further experiments are suggested. Mukherjee and Zavala state this is one of the primary opportunities for physics constrained learning [12] and Ureel *et al.* wrote a review on active machine learning for chemical engineers [55]. The methanol study is an early example: an experiment designed from the trained hybrid model enhanced the predictive performance [18]. Cheap estimates of information gain for a large number of parameters in a model, algorithms that select experiments under safety and cost constraints, and integration with automated laboratories would be needed on wider use. Natural test beds include high-throughput catalyst screening platforms and continuous-flow platforms.

C. Multiscale Modeling

Multiscale models connect molecular kinetics, such as from density functional theory, and reactor/plant behavior. Learned closures can be provided by hybrid models in the

scales where mechanistic descriptions are weak, like effective transport in complex geometries, without compromising the balances at reactor scale [32, 53]. There is a need for methods to couple across scales, to propagate uncertainty across scales, and to solve the coupled models efficiently. The most convincing test would be carried out on actual catalytic reactors in industry, or on a polymerization process.

VII. CONCLUSION

In this review, the four domains of PI were explored using hybrid modeling. The choice of serial, parallel and physics-constrained architectures should be determined by the known information about the process and by the availability of data. Soft sensing, control and energy related prediction of units' properties have reached the most mature application stage, with synergy-type integrated units being the least studied. Four of the reaction-engineering case studies are examined and compared to their primary sources, resulting in a more tempered view than is often found in the literature. The hybrid model, based on an optimized methanol kinetic model, showed a slight improvement of the test accuracy (MAPE 8.49% compared to MAPE 9.06%), while the optimally designed experiments both with the well-calibrated methanol model and the optimized model showed a significant improvement of both types of model. A thermodynamically consistent power-law backbone yielded a better test error (6.7%) than a Langmuir–Hinshelwood backbone (9.5%) and yielded physically meaningful rate constants. These learned corrections decreased the prediction errors of polymerization by 84% and 86% for the main reagents, respectively, and training a surrogate using simulations resulted in a speedup of the models of 45–400 times for the main reagents. Key challenges for further progress include common benchmarks for data efficiency, robust training of coupled models, hard physical constraints, multiscale coupling, and practical uncertainty quantification. These methods can be tested in operation in closed-loop experimental design and hybrid digital twins. The integrated reaction–separation units are of special interest since the

coupling of phenomena is strongest there and validated hybrid models are least.

Declaration of Conflicting Interests

The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Use of Artificial Intelligence (AI)-Assisted Technology for Manuscript Preparation

The author(s) utilised generative AI technologies solely to improve language clarity and refine grammar. No images were generated, manipulated, or altered using AI-assisted technologies.

REFERENCES

- [1] A. I. Stankiewicz and J. A. Moulijn, "Process intensification: Transforming chemical engineering," *Chem. Eng. Prog.*, vol. 96, no. 1, pp. 22–34, 2000.
- [2] T. Van Gerven and A. Stankiewicz, "Structure, energy, synergy, time the fundamentals of process intensification," *Ind. Eng. Chem. Res.*, vol. 48, no. 5, pp. 2465–2474, 2009, doi: 10.1021/ie801501y.
- [3] Y. Tian, S. E. Demirel, M. M. F. Hasan, and E. N. Pistikopoulos, "An overview of process systems engineering approaches for process intensification: State of the art," *Chem. Eng. Process. – Process Intensif.*, vol. 133, pp. 160–210, 2018, doi: 10.1016/j.cep.2018.07.014.
- [4] D. S. Sholl and R. P. Lively, "Seven chemical separations to change the world," *Nature*, vol. 532, no. 7600, pp. 435–437, 2016, doi: 10.1038/532435a.
- [5] G. J. Harmsen, "Reactive distillation: The front-runner of industrial process intensification. A full review of commercial applications, research, scale-up, design and operation," *Chem. Eng. Process. – Process Intensif.*, vol. 46, no. 9, pp. 774–780, 2007, doi: 10.1016/j.cep.2007.06.005.
- [6] F. Esposito, U. Di Caprio, S. Buzzi, F. Vermeire, and M. E. Leblebici, "Hybrid modelling approaches in process intensification: A thorough review," *Chem. Eng. Process. – Process Intensif.*, vol. 217, Art. no. 110496, 2025, doi: 10.1016/j.cep.2025.110496.
- [7] M. von Stosch, R. Oliveira, J. Peres, and S. Feyer de Azevedo, "Hybrid semi-parametric modeling in process systems engineering: Past, present and future," *Comput. Chem. Eng.*, vol. 60, pp. 86–101, 2014, doi: 10.1016/j.compchemeng.2013.08.008.
- [8] A. Beck, M. A. Newton, L. G. A. van de Water, and J. A. van Bokhoven, "The enigma of methanol synthesis by Cu/ZnO/Al₂O₃-based catalysts," *Chem. Rev.*, vol. 124, no. 8, pp. 4543–4678, 2024, doi: 10.1021/acs.chemrev.3c00148.
- [9] M. B. Fichtl *et al.*, "Kinetics of deactivation on Cu/ZnO/Al₂O₃ methanol synthesis catalysts," *Appl. Catal. A: Gen.*, vol. 502, pp. 262–270, 2015, doi: 10.1016/j.apcata.2015.06.014.
- [10] M. R. Dobbelaere, P. P. Plehiers, R. Van de Vijver, C. V. Stevens, and K. M. Van Geem, "Machine learning in chemical engineering: Strengths, weaknesses, opportunities, and threats," *Engineering*, vol. 7, no. 9, pp. 1201–1211, 2021, doi: 10.1016/j.eng.2021.03.019.
- [11] P. Shah, S. Pahari, R. Bhavsar, and J. S.-I. Kwon, "Hybrid modelling of first principles and machine learning: A step-by-step tutorial review for practical implementation," *Comput. Chem. Eng.*, vol. 194, Art. no. 108926, 2025, doi: 10.1016/j.compchemeng.2024.108926.
- [12] A. Mukherjee and V. M. Zavala, "Physics-constrained machine learning for chemical engineering," *Curr. Opin. Chem. Eng.*, vol. 51, Art. no. 101228, Mar. 2026; preprint: arXiv:2508.20649.
- [13] A. M. Schweidtmann, D. Zhang, and M. von Stosch, "A review and perspective on hybrid modeling methodologies," *Digit. Chem. Eng.*, vol. 10, Art. no. 100136, 2024, doi: 10.1016/j.dche.2023.100136.
- [14] D. C. Psychogios and L. H. Ungar, "A hybrid neural network-first principles approach to process modeling," *AIChE J.*, vol. 38, no. 10, pp. 1499–1511, 1992, doi: 10.1002/aic.690381003.
- [15] M. L. Thompson and M. A. Kramer, "Modeling chemical processes using prior knowledge and neural networks," *AIChE J.*, vol. 40, no. 8, pp. 1328–1340, 1994, doi: 10.1002/aic.690400806.
- [16] J. Sansana, M. N. Joswiak, I. Castillo, Z. Wang, R. Rendall, L. H. Chiang, and M. S. Reis, "Recent trends on hybrid modeling for Industry 4.0," *Comput. Chem. Eng.*, vol. 151, Art. no. 107365, 2021, doi: 10.1016/j.compchemeng.2021.107365.
- [17] C. L. Gargalo *et al.*, "Navigating Industry 4.0 and 5.0: The role of hybrid modelling in (bio)chemical engineering's digital transition," *Front. Chem. Eng.*, vol. 6, Art. no. 1494244, 2024, doi: 10.3389/fceng.2024.1494244.
- [18] L. Kaps *et al.*, "Optimal experiments for hybrid modeling of methanol synthesis kinetics," *Comput. Chem. Eng.*, vol. 204, Art. no. 109387, 2026, doi: 10.1016/j.compchemeng.2025.109387.
- [19] E. Oguta and I. E. Ihua-Maduenyi, "Modelling and performance analyses of annular equivalent circulation density in extended reach wells using response surface methodology," *Asian J. Eng. Appl. Technol.*, vol. 14, no. 2, pp. 1–17, 2025, doi: 10.70112/ajeat-2025.14.2.4298.
- [20] N. Sharma and Y. A. Liu, "A hybrid science-guided machine learning approach for modeling chemical processes: A review," *AIChE J.*, vol. 68, no. 5, Art. no. e17609, 2022, doi: 10.1002/aic.17609.
- [21] W. Bradley *et al.*, "Perspectives on the integration between first principles and data-driven modelling," *Comput. Chem. Eng.*, vol. 166, Art. no. 107898, 2022, doi: 10.1016/j.compchemeng.2022.107898.
- [22] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *J. Comput. Phys.*, vol. 378, pp. 686–707, 2019, doi: 10.1016/j.jcp.2018.10.045.
- [23] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed machine learning," *Nat. Rev. Phys.*, vol. 3, no. 6, pp. 422–440, 2021, doi: 10.1038/s42254-021-00314-5.
- [24] C. Rackauckas *et al.*, "Universal differential equations for scientific machine learning," arXiv:2001.04385, 2020, doi: 10.48550/arXiv.2001.04385.
- [25] R. Oliveira, "Combining first principles modelling and artificial neural networks: A general framework," *Comput. Chem. Eng.*, vol. 28, no. 5, pp. 755–766, 2004.
- [26] M. Brendel and W. Marquardt, "Experimental design for the identification of hybrid reaction models from transient data," *Chem. Eng. J.*, vol. 141, no. 1–3, pp. 264–277, 2008, doi: 10.1016/j.cej.2007.12.027.
- [27] K. Merkelbach *et al.*, "HybridML: Open source platform for hybrid modelling," *Comput. Chem. Eng.*, vol. 160, Art. no. 107736, 2022, doi: 10.1016/j.compchemeng.2022.107736.
- [28] F. Sorourifar, Y. Peng, I. Castillo, L. Bui, J. Venegas, and J. A. Paulson, "Physics-enhanced neural ordinary differential equations: Application to industrial chemical reaction systems," *Ind. Eng. Chem. Res.*, vol. 62, no. 38, pp. 15563–15577, 2023, doi: 10.1021/acs.iecr.3c01471.
- [29] D. Ghosh, E. Hermonat, P. Mhaskar, S. Snowling, and R. Goel, "Hybrid modeling approach integrating first-principles models with subspace identification," *Ind. Eng. Chem. Res.*, vol. 58, no. 30, pp. 13533–13543, 2019, doi: 10.1021/acs.iecr.9b00900.
- [30] N. A. Nasruddin, N. Islam, S. Vernuccio, and J. Oyekan, "Hybridised mechanistic and machine learning digital twins for modelling and optimising chemical processes in flow: A comparative analysis of parallel and series-based hybridisation," *Chem. Eng. J. Adv.*, vol. 23, Art. no. 100775, 2025, doi: 10.1016/j.cej.2025.100775.
- [31] M. Velioglu, S. Zhai, S. Rupprecht, A. Mitsos, A. Jupke, and M. Dahmen, "Physics-informed neural networks for dynamic process operations with limited physical knowledge and data," *Comput. Chem. Eng.*, vol. 192, Art. no. 108899, 2025, doi: 10.1016/j.compchemeng.2024.108899.
- [32] Z. Wu, H. Wang, C. He, B. Zhang, T. Xu, and Q. Chen, "The application of physics-informed machine learning in multiphysics modeling in chemical engineering," *Ind. Eng. Chem. Res.*, vol. 62, no. 44, pp. 18178–18204, 2023, doi: 10.1021/acs.iecr.3c02383.
- [33] H. Kay, F. Vega-Ramon, R. Gallen, E. H. Stitt, and D. Zhang, "Developing a hybrid modeling framework for enhanced prediction in chemical reaction kinetics," *Ind. Eng. Chem. Res.*, vol. 64, no. 33, pp. 16027–16038, 2025, doi: 10.1021/acs.iecr.5c01597.
- [34] P. Kadlec, B. Gabrys, and S. Strandt, "Data-driven soft sensors in the process industry," *Comput. Chem. Eng.*, vol. 33, no. 4, pp. 795–814, 2009, doi: 10.1016/j.compchemeng.2008.12.012.
- [35] H. Jin, X. Chen, L. Wang, K. Yang, and L. Wu, "Adaptive soft sensor development based on online ensemble Gaussian process regression for

- nonlinear time-varying batch processes,” *Ind. Eng. Chem. Res.*, vol. 54, no. 30, pp. 7320–7345, 2015.
- [36] N. P. Lawrence *et al.*, “Machine learning for industrial sensing and control: A survey and practical perspective,” *Control Eng. Pract.*, vol. 145, Art. no. 105841, 2024, doi: [10.1016/j.conengprac.2024.105841](https://doi.org/10.1016/j.conengprac.2024.105841).
- [37] Z. Wu, A. Tran, D. Rincon, and P. D. Christofides, “Machine-learning-based predictive control of nonlinear processes. Part I: Theory,” *AIChE J.*, vol. 65, no. 11, Art. no. e16729, 2019, doi: [10.1002/aic.16729](https://doi.org/10.1002/aic.16729).
- [38] T. Keßler *et al.*, “Two degrees of freedom control of a multistage power to methanol reactor,” *Comput. Chem. Eng.*, vol. 192, Art. no. 108893, 2024, doi: [10.1016/j.compchemeng.2024.108893](https://doi.org/10.1016/j.compchemeng.2024.108893).
- [39] G. T. Rochelle, “Amine scrubbing for CO₂ capture,” *Science*, vol. 325, no. 5948, pp. 1652–1654, 2009, doi: [10.1126/science.1176731](https://doi.org/10.1126/science.1176731).
- [40] U. Di Caprio, F. Vermeire, T. Van Gerven, and M. E. Leblebici, “Physics-informed machine learning predicting CO₂ capture performances of organic mixtures,” *Chem. Eng. Process.–Process Intensif.*, vol. 216, Art. no. 110410, 2025, doi: [10.1016/j.cep.2025.110410](https://doi.org/10.1016/j.cep.2025.110410).
- [41] U. Di Caprio *et al.*, “Hybrid modelling of a batch separation process,” *Comput. Chem. Eng.*, vol. 177, Art. no. 108319, 2023, doi: [10.1016/j.compchemeng.2023.108319](https://doi.org/10.1016/j.compchemeng.2023.108319).
- [42] C. Seidel, A. Jörke, B. Vollbrecht, A. Seidel-Morgenstern, and A. Kienle, “Kinetic modeling of methanol synthesis from renewable resources,” *Chem. Eng. Sci.*, vol. 175, pp. 130–138, 2018.
- [43] X. Zhou, Z. Zhang, M. Zhu, H. Zhao, H. Yan, and C. Yang, “A hybrid modeling strategy based on deep learning surrogate models for accurate process multi-objective optimization of iso-octanol oxidation,” *ENG. Chem. Eng.*, vol. 20, Art. no. 10, 2026, doi: [10.1007/s11705-026-2630-1](https://doi.org/10.1007/s11705-026-2630-1).
- [44] X. Zhou *et al.*, “Hybrid modelling incorporating reaction and mechanistic data for accelerating the development of isooctanol oxidation,” *Chin. J. Chem. Eng.*, vol. 80, pp. 166–183, 2025, doi: [10.1016/j.cjche.2025.02.003](https://doi.org/10.1016/j.cjche.2025.02.003).
- [45] O. C. Boy, I. B. R. Nogueira, U. Di Caprio, and M. E. Leblebici, “AI-aided process intensification of structures,” *Chem. Eng. Process.–Process Intensif.*, vol. 216, Art. no. 110406, 2025, doi: [10.1016/j.cep.2025.110406](https://doi.org/10.1016/j.cep.2025.110406).
- [46] H. Kay, F. Vega-Ramon, and D. Zhang, “Hybrid modelling for reaction network simulation in syngas methanol production,” *Syst. Control Trans.*, vol. 4, pp. 111–116, 2025, doi: [10.69997/sct.172680](https://doi.org/10.69997/sct.172680).
- [47] S. K. Wilkinson, L. G. A. van de Water, B. Miller, M. J. H. Simmons, E. H. Stitt, and M. J. Watson, “Understanding the generation of methanol synthesis and water gas shift activity over copper-based catalysts – A spatially resolved experimental kinetic study using steady and non-steady state operation under CO/CO₂/H₂ feeds,” *J. Catal.*, vol. 337, pp. 208–220, 2016, doi: [10.1016/j.jcat.2016.01.025](https://doi.org/10.1016/j.jcat.2016.01.025).
- [48] Y. Tian, J. Zhang, and J. Morris, “Modeling and optimal control of a batch polymerization reactor using a hybrid stacked recurrent neural network model,” *Ind. Eng. Chem. Res.*, vol. 40, no. 21, pp. 4525–4535, 2001, doi: [10.1021/ie0010565](https://doi.org/10.1021/ie0010565).
- [49] O. Sobhani *et al.*, “A hybrid predictive modeling approach for catalyzed polymerization reactors,” *Chem. Eng. J. Adv.*, vol. 20, Art. no. 100662, 2024, doi: [10.1016/j.cej.2024.100662](https://doi.org/10.1016/j.cej.2024.100662).
- [50] C. Plate, C. J. Martensen, and S. Sager, “Optimal experimental design for universal differential equations,” *IEEE Trans. Autom. Control*, 2025 (early access), doi: [10.1109/TAC.2025.3609533](https://doi.org/10.1109/TAC.2025.3609533).
- [51] J. Pinto, M. Mestre, J. Ramos, R. S. Costa, G. Striedner, and R. Oliveira, “A general deep hybrid model for bioreactor systems: Combining first principles with deep neural networks,” *Comput. Chem. Eng.*, vol. 165, Art. no. 107952, 2022, doi: [10.1016/j.compchemeng.2022.107952](https://doi.org/10.1016/j.compchemeng.2022.107952).
- [52] M. S. F. Bangi, K. Kao, and J. S.-I. Kwon, “Physics-informed neural networks for hybrid modeling of lab-scale batch fermentation for β -carotene production using *Saccharomyces cerevisiae*,” *Chem. Eng. Res. Des.*, vol. 179, pp. 415–423, 2022, doi: [10.1016/j.cherd.2022.01.041](https://doi.org/10.1016/j.cherd.2022.01.041).
- [53] T. Zhou, R. Gani, and K. Sundmacher, “Hybrid data-driven and mechanistic modeling approaches for multiscale material and process design,” *Engineering*, vol. 7, no. 9, pp. 1231–1238, 2021.
- [54] D. C. Boffito, “Scaling process intensification technologies: What does it take to deploy?,” *Chem. Eng. Process.–Process Intensif.*, vol. 212, Art. no. 110275, 2025, doi: [10.1016/j.cep.2025.110275](https://doi.org/10.1016/j.cep.2025.110275).
- [55] Y. Ureel *et al.*, “Active machine learning for chemical engineers: A bright future lies ahead,” *Engineering*, vol. 27, pp. 23–30, 2023, doi: [10.1016/j.eng.2023.02.019](https://doi.org/10.1016/j.eng.2023.02.019).