Chemical Engineering Journal Advances 20 (2024) 100662

Contents lists available at ScienceDirect

Chemical Engineering Journal Advances

journal homepage: www.sciencedirect.com/journal/chemical-engineering-journal-advances

A hybrid predictive modeling approach for catalyzed polymerization reactors

Omid Sobhani [a,*], Hamid Toliati [a], Furkan Elmaz [a], Shahab Pormoradi Gerdposhteh [b], Benedict Carius [b], Kevin Mets [a], Siegfried Mercelis [a]

[a] University of Antwerp - imec, IDLab - Faculty of Applied Engineering, Sint-Pietersvliet 7, 2000 Antwerp, Belgium
[b] allnex R&D, allnex, Anderlechtstraat 33, 1620 Drogenbos, Belgium

ARTICLE INFO

Keywords:
Machine learning
Hybrid modeling
Polymerization process
Fed-batch chemical process

ABSTRACT

Polymerization reactions are characterized by complex, nonlinear behaviors that pose significant challenges for conventional modeling techniques. Accurate and reliable models are crucial for advancing material science and enabling technological innovations across various industries. Conventional first-principles models often fall short in capturing the intricate dynamics of polymeric systems, leading to limitations in predictive accuracy. In this work, we propose a novel hybrid modeling approach that combines a conventional first-principles model with the strengths of a data-driven multi-layer perceptron (MLP) model and also a linear regression (LR) model to enhance the predictability of polymerization processes. Utilizing this hybrid approach significantly reduces the mean absolute error for predicting the concentrations of main reagents by 84% and 86%, respectively, in experiments with significantly deviant outcomes. Our results indicate that the model is capable of predicting the concentrations of both the main and side products with a maximum error margin of 3.5%.

1. Introduction

In order to effectively control, optimize, or monitor complex chemical processes, it is crucial to have an accurate model that captures their nonlinear dynamic behaviors. Achieving these objectives is feasible with a reliable process model, typically developed based on first-principles equations [1]. In the realm of chemical reaction engineering, where the underlying physical and chemical laws may be unknown or uncertain, traditional phenomenological models often prove insufficient or prone to inaccuracies due to necessary approximations [2]. Recently, artificial intelligence (AI) and machine learning (ML) techniques have emerged as promising alternatives, demonstrating substantial success in theoretical and computational chemistry [3]. A comprehensive review by Zhu et al. [4] discusses the growing application of ML in chemical reactors, particularly highlighting its potential in developing accurate predictive models for multiphase flow systems. Despite their success, these ML techniques, often referred to as "black box" models, exhibit significant limitations. Their primary drawback is a restricted ability to generalize beyond the scope of their training data, coupled with a lack of interpretability. This shortcoming hinders a deeper understanding of the chemical processes at play, representing a substantial challenge for their application in complex scenarios [3,5].

Hybrid modeling, combining the predictive power of black box models with the explanatory depth of white box models, offers a balanced approach that leverages the strengths of both. This methodology not only enhances the model's interpretability and ability to generalize across different conditions but also improves extrapolation capabilities by incorporating fundamental physical and chemical principles [6–8].

Reflecting on the theoretical underpinnings and advantages of hybrid modeling, empirical studies have been conducted to validate its application and efficacy in diverse scenarios. Mjalli and Ibrehem [9] use statistical methods in their analysis to identify key model parameters that significantly affect the outcome and to provide optimal estimates for these parameters. In another study, Di Caprio et al. [10] initially developed a first principles model and used experimental data to fit two different types of functions. They then utilize these fitted functions to dynamically estimate two different model parameters. Expanding on this, recent work by Bangi and Kwon [11] introduces a deep hybrid model-based predictive control (DHM-based MPC) framework that enhances predictive accuracy while ensuring system stability within the model's domain of applicability, particularly in continuous stirred tank reactors. Pahari et al. [12] highlight the use of hybrid models to estimate spatiotemporally varying parameters in complex systems like reaction–diffusion processes. They illustrate how such models can capture latent chemical mechanisms that are otherwise difficult to discern using traditional methods. Sitapure and Kwon [13] demonstrate the application of hybrid modeling with time-series transformers in chemical systems, underscoring the potential for improving process modeling and control through enhanced temporal predictions. Similarly, Shah

* Corresponding author.
E-mail address: seyedomid.sobhani@uantwerpen.be (O. Sobhani).

https://doi.org/10.1016/j.ceja.2024.100662
Received 2 July 2024; Received in revised form 30 September 2024; Accepted 12 October 2024
Available online 19 October 2024

et al. [14] developed a hybrid model for an industry-scale fermentation process, combining kinetic models with deep neural networks to predict time-varying dependencies among parameters, thus enhancing the model's robustness and accuracy beyond traditional approaches.

Despite the advancements in hybrid modeling for optimizing parameters and improving predictive accuracy in diverse scenarios [15], the complexity of polymeric reactions introduces additional challenges that demand further refinement of these models. Polymeric reactions play a crucial role in creating materials with specific properties and applications, from medical devices to structural components. However, these reactions often face challenges due to the gel effect, a phenomenon characterized by a rapid increase in reaction rate and medium viscosity at certain conversion levels, complicating process control and product quality [16]. Ensuring product quality in polymerization processes is significantly more challenging than in reactions with small molecules due to the profound impact of the polymer's morphological and molecular characteristics on its physical, chemical, thermal, rheological, and mechanical properties. These attributes critically determine the polymer's suitability for its intended applications, making precise control over the polymerization process essential for achieving the desired product quality [17].

Hybrid neural network models for polymerization processes offer a promising approach to modeling complex chemical reactions. By combining neural networks with simplified phenomenological models, researchers have developed methods that improve prediction accuracy and operational efficiency in polymer production. These models effectively capture the dynamics of polymerization reactions, including those influenced by gel and glass effects, demonstrating a significant advance over traditional modeling techniques. The integration of machine learning with established chemical engineering principles marks a pivotal contribution to the field, enabling more precise control and optimization of polymerization processes [18–20].

The studies previously mentioned developed an extensive first-principles model for a batch reactor operating at constant temperature and attempted to predict the specifications of the polymer under different reactor conditions. A significant limitation of this approach is its inapplicability to fed-batch or continuous stirred-tank reactors (CSTR), where the system's parameters change dynamically. Furthermore, integrating complex differential equations with neural networks to simulate the reactor proves to be time-consuming. This complexity restricts the practical use of these models as simulation environments for optimization.

Regarding industrial priorities, ensuring the quality of the end product and swiftly addressing process deviations to prevent gelation is paramount. Hence, optimization, while important, becomes a secondary consideration, provided that product quality remains uncompromised. This emphasis on quality assurance and the necessity for rapid troubleshooting underpin the need for more adaptable and efficient modeling approaches that can support real-time monitoring and optimization without sacrificing product integrity.

The main contribution of this work lies in the novel integration of a regression module with a simplified hybrid model for the computation of reaction kinetics and quality of the final product. This approach enables real-time monitoring of both main (*i.e.* the polymer with the desired chain length) and side product concentrations during the process. Given its simplified foundation, this approach demonstrates remarkable generalizability across different reactor types. For this purpose, we investigate a fed-batch polymerization reactor as a case study and utilize provided experimental dataset (detailed in Section 2.2) for training and validation.

The manuscript is structured as follows: Section 2 outlines the materials and methods, establishing the experimental framework. Section 3 describes the hybrid modeling approach, detailing its development and theoretical underpinnings. Section 4 presents the results and discussion, analyzing the data and interpreting the findings. The paper concludes with Section 5, summarizing the study's contributions and suggesting avenues for future research.

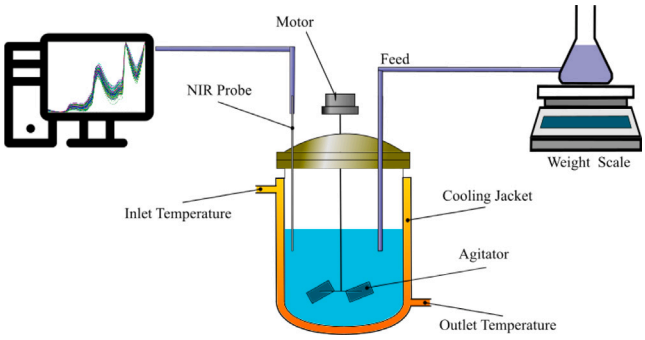

Fig. 1. Schematic representation of the experimental setup: Initially, the reactor contains a mixture of component A and a catalyst. Component B is stored in a separate tank and is fed into the reactor according to various strategies. The reactor's temperature is regulated by a surrounding jacket.

2. Materials and methods

In this section, we first describe the experimental setup and detail the datasets collected from these experiments. Following this, we introduce a simple first-principle model (also known as white-box model), comprised a set of ordinary differential equations. Subsequently, we present the multi-layer perceptron and linear regression model as a function estimators to estimate parameters of the model. Finally, we discuss the metrics used for training and evaluating our model.

2.1. Reaction system

The reactor studied here, depicted in Fig. 1 , is a pilot-scale polymerization fed-batch reactor located at the R&D center of allnex Resin Company in Drogenbos, Belgium. As depicted in this figure, the reactor is equipped with a stirrer to ensure thorough mixing of the reactants. The temperature of the reaction mixture is controlled by circulating oil at an appropriate temperature through the reactor's jacket. Additionally, the reactor is fitted with a near-infrared (NIR) probe to monitor changes in the mixture's composition. The process begins with a mixture of catalysts and one of the reagents pre-heated to a specified temperature, followed by the gradual addition of the second reagent according to a specific recipe. Once the addition is complete, the mixture is allowed to react in the reactor for a predetermined period to ensure the reaction reaches completion. The polymerization reactions involve chaining monomers, with the chain length highly dependent on reaction conditions, including temperature and the availability of free radicals. Furthermore, the reaction is exothermic, requiring careful temperature control within a specific range. The heat-bath setpoints are determined by human experts, and the temperature is controlled with a primary PID. Different feed flow rates can be applied by a peristaltic pump, and can be set by the expert. The molar equivalents of the reactants are estimated using NIR spectroscopy, measured online during the reaction. It is important to prevent the accumulation of side reactions during the process as they can lead to the formation of a highly viscous product, also known as gel, which is undesirable. Since the desired product and by-products share the same chemical bond, it is impossible to distinguish and monitor the production of the main products and the by-products using NIR spectroscopy. Therefore, indirect methods are necessary to model and estimate the concentrations of both the desired product and by-products.

2.2. Data description

In this study, 44 lab-scale experiments were conducted with two recipes under different initial conditions and strategies. The main difference between these two recipes is the feeding time. Among these 44

experiments, there are 4 golden batches representing the best execution based on human expert knowledge. These golden batch experiments are optimized to achieve the minimum concentration of side products at the end of the experiments. The remaining experiments are designed to cover a wide range of initial and operational conditions and recipes, aiming to produce a specific product with quality outcomes ranging from acceptable to significantly deviant. There are three sources of data for each experiment. The first source is related to the `Camile` software and includes measurements such as the initial weight of the loaded reactor, flow rate, stirrer speed, and the inlet and outlet jacket temperatures. The second source of data is the NIR probe, from which it is possible to estimate the concentration of the main reagents. These two data sets are recorded every minute during the experiments. The third source is HPLC analysis, providing information about the mixture's composition and is a one-time measurement that take place only at the end of the experiments.

Based on these measurements, and following the first principal model outlined in Section 2.3, we create an auxiliary dataset for each experiment. This auxiliary dataset includes the reaction rate and stoichiometry coefficients that best fit the concentrations of the main reagents at each time step for all experiments and will be used to train the MLP model.

For training purposes, we randomly divide the experiments into train and validation datasets, ensuring there is at least one golden batch experiment in the validation dataset. The remaining experiments are concatenated together to form a comprehensive training dataset, consisting more than 10^4 data points. This approach allows us to closely simulate real-world conditions, ensuring that our model is not only accurate but also robust enough to handle variations within the production environment.

It is important to note that due to the confidentiality agreement, all chemical names and variables have been masked, and the presented graphs are normalized versions specifically created for this study.

2.3. Simple first principal model

The stoichiometry of the investigated reaction is defined as follows:

$$A + 2B \rightarrow \alpha SP + \beta MP, \tag{1}$$

where A and B are identified as the main reagents, with SP and MP representing the side and main products, respectively. The variables α and β denote the rates of production for the side and main products, resulting from the consumption of one mole of component A and two moles of component B, respectively. Importantly, β is influenced by the properties of the main product (MP) and is determined by the reaction conditions. By first calculating β and then applying mass balance equations, we can accurately determine α.

The first principal model is developed based on the work of Merfeld et al. [21] as follows:

$$\frac{dC_A}{dt} = \frac{\dot{m}_A}{V_l M_A} - k_0 e^{-\frac{E}{RT}} C_A^m C_B^n C_{cat}^o - \frac{C_A}{V_l}\frac{dV_l}{dt} \tag{2a}$$

$$\frac{dC_B}{dt} = \frac{\dot{m}_B}{V_l M_B} - \gamma k_0 e^{-\frac{E}{RT}} C_A^m C_B^n C_{cat}^o - \frac{C_B}{V_l}\frac{dV_l}{dt} \tag{2b}$$

$$\frac{dC_{sp}}{dt} = \alpha k_0 e^{-\frac{E}{RT}} C_A^m C_B^n C_{cat}^o - \frac{C_{sp}}{V_l}\frac{dV_l}{dt} \tag{2c}$$

$$\frac{dC_{mp}}{dt} = \beta k_0 e^{-\frac{E}{RT}} C_A^m C_B^n C_{cat}^o - \frac{C_{mp}}{V_l}\frac{dV_l}{dt} \tag{2d}$$

$$\frac{dm^{total}}{dt} = \sum_i \dot{m}_i \implies \frac{dV_l}{dt} = \frac{1}{\rho_{avg}} \sum_i \dot{m}_i, \tag{2e}$$

where C_is denote the molar concentrations of components. M_i represent the molar weights of reagents. V_l and dV_l represent the volume and volume change of the mixture, respectively. $\dot{m}_i$ is the flow rate of the reagents, and $k_0 = 8.6 \times 10^6$, $E = 61.07$ kJ/mol, $R = 8.314$ J/(mol K), and T are the reaction constant, activation energy, gas constant, and reactor temperature, respectively. The terms $m = 1.0096$, $n = 0.3899$, and $o = 0.11$ represent the reaction orders with respect to concentration of species A, B, and the catalyst cat, respectively. Finally, ρ_{avg} is the average density of the mixture. Note that this set of equations is not independent and one can be derived from the others.

The coefficients in these equations are optimized to fit the reagent concentrations for golden-batch experiments at each time step. Fig. 5 compares the output of the white box model with experimental data for component B in three different experimental conditions. As demonstrated in this figure, for a golden batch experiment with minimal side reactions, the white box model's predictions of the concentration closely match the experimental values. However, the model is unable to accurately predict the values when side reactions are present in the system.

2.4. Multi-layer perceptron

Multi-layer perceptron (MLP) is a machine learning technique for processing information, and has gained popularity for its ability to predict both continuous and discrete variables across numerous fields [22, 23]. MLPs are composed of neurons and layers, with neurons in one layer receiving inputs from those in the previous layer, applying an activation function, and forwarding the output to the next layer [24]. This process flows from the input to the output layer, passing through any number of hidden layers, whose configuration – number of layers, neurons per layer, and activation functions – are critical hyper-parameters that need to be defined in advance. The estimated output $\hat{y}$ of an MLP with one hidden layer can be succinctly represented by the equation:

$$\hat{y} = f\left(\sum_{j=1}^{M} w_j^{(2)} \cdot g\left(\sum_{i=1}^{N} w_{ij}^{(1)} \cdot x_i + b_j^{(1)}\right) + b^{(2)}\right) \tag{3}$$

where $\hat{y}$ denotes the estimated output of the MLP. x_i represents the ith input to the model, with i ranging from 1 to N, indicating the total number of input features. $w_{ij}^{(1)}$ stands for the weight from the ith input unit to the jth neuron in the hidden layer. $b_j^{(1)}$ is the bias associated with the jth neuron in the hidden layer. $w_j^{(2)}$ refers to the weight from the jth neuron in the hidden layer to the output neuron. $b^{(2)}$ represents the bias of the output neuron. $g(\cdot)$ is the activation function applied to each neuron in the hidden layer, and $f(\cdot)$ is the activation function applied at the output layer, the choice of which depends on the specific task.

This formulation encapsulates the process by which an MLP with one hidden layer processes input data. The model initially applies a series of weighted sums and biases to the inputs, followed by a non-linear transformation through the activation function g in the hidden layer. The resulting activations are then linearly combined, adjusted by another bias, and passed through the output activation function f, producing the final estimate $\hat{y}$. This structure enables the MLP to capture complex relationships between the input features and the target output, making it a versatile tool for a wide range of predictive modeling tasks.

2.5. Linear regression

Linear regression (LR) is a statistical method used to model the relationship between a dependent variable and one or more independent variables by fitting a linear equation to the observed data. In this study, we employed linear regression to analyze the influence of the independent variables on the dependent variable. The linear regression model can be expressed as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \epsilon, \tag{4}$$

where β_0 is the intercept, $\beta_1, \beta_2, \ldots, \beta_n$ are the coefficients of the independent variables, and (ϵ) represents the error term.

In our analysis, we used ordinary least squares (OLS) estimation to determine the parameters $(\beta_0, \beta_1, \ldots, \beta_n)$. This method minimizes the sum of the squared differences between the observed values and the

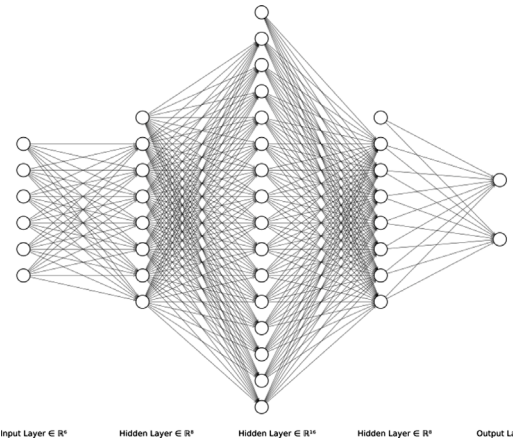

Fig. 2. Structure of Multi-Layer Perceptron (MLP) model. The model consists of 6 inputs, 3 hidden layers with [8, 16, 8] neurons, and 2 outputs. It uses ReLU activation for the hidden layers and a linear activation for the output layer.

values predicted by the model. We evaluated the model fit using R-squared, and conducted hypothesis tests on the regression coefficients to determine their statistical significance. The results of the linear regression analysis provided insights into the strength and direction of the relationships between the variables under investigation.

2.6. Metrics and sensitivity

For evaluating the proposed model, we utilized Mean Squared Error (MSE) and Mean Absolute Error (MAE). These metrics succinctly capture the model's accuracy and predictive reliability.

$$\mathrm{MSE} = \frac{1}{m}\sum_{i=1}^{m}(\hat{y}_i - y_i)^2 \tag{5}$$

$$\mathrm{MAE} = \frac{1}{m}\sum_{i=1}^{m}|\hat{y}_i - y_i| \tag{6}$$

where, $\hat{y}_i$ and y_i are the ith prediction and experimental values, respectively, while m is the number of data points that is equal to the length of each experiment.

To investigate the robustness of the proposed modeling approach, we augmented measurements in different parts of the model with varying levels of Gaussian noise. Subsequently, we compared the error metrics to experimental measurements of the reagents' concentrations.

3. Hybrid modeling approach

In this section, we motivate and present our hybrid modeling approach, which is composed of three parts: an MLP model, a LR model, and an ODE (*i.e.* white-box) model, as shown in Fig. 4. This structure enables us to more effectively generalize the model to a wider range of reactor conditions and accurately capture the production trends of both main and side products. Initially, in the first part (Section 3.1), we utilize an MLP to estimate the reaction rate and coefficients. Subsequently, in the second part of our hybrid model (Section 3.2), we introduce new features alongside a linear regression model to estimate the production rates of the main and side products. By employing the action strategies in the data set and the out put of the MLP and regression models, we are able to update the concentration of the components accurately. The comprehensive working model is discussed in detail at the conclusion of this section and the input and output of each part of the model is summarized in Table 1.

3.1. Reaction rate and coefficient

Hybrid modeling approaches combine principles from equations and experimental data to enhance the model's generalizability. There are several approaches to developing hybrid models, including transfer learning [25,26], a serial model refining structure [27], and internal parameter optimization [10,28]. The structure of the hybrid model and its specific application depend on the intervening variable to estimate and the available experimental data.

In Eq. (2b), the concentration of component B, the feeding reagent, is determined by several factors including the feeding flow ($\dot{m}_B$), reaction rate (υ_t), stoichiometry coefficient (γ_t), and volume changes (dV_l). From Eq. (2), the reaction rate can be defined as follows:

$$\upsilon_t = k_0 e^{-\frac{E}{RT}} C_A^m C_B^n C_{cat}^o \tag{7}$$

Therefore, reaction rate is influenced by the concentrations of both reagents (C_A and C_B) and the catalyst (C_{cat}), in addition to reaction conditions such as reactor temperature (T^r). These equations are adequate for predicting the concentrations of reagents in golden-batch (*i.e.*, ideal) experiments. However, in non-ideal experiments, characterized by numerous uncertainties and side reactions, the predicted values may differ.

In this work, we have obtained experimental concentrations for both main reagents and the reactor temperature. We also calculated the catalysis concentration at each time step. However, the predicted values for non-ideal experiments, derived from the first-principle model, show a noticeable discrepancy when compared with the experimental results. To address this, we hypothesized that the reaction rate follows a more complex form than that described by Eq. (2). Furthermore, we propose that additional experimental conditions, such as stirrer speed (S_t), can significantly affect the consumption rate of the reagents.

By substituting Eq. (7) in Eq. (2a) and (2b), we will have:

$$\frac{dC_A}{dt} = \frac{\dot{m}_A}{V_l M_A} - \upsilon_t - \frac{C_A}{V_l}\frac{dV_l}{dt} \tag{8a}$$

$$\frac{dC_B}{dt} = \frac{\dot{m}_B}{V_l M_B} - \gamma_t \upsilon_t - \frac{C_B}{V_l}\frac{dV_l}{dt} \tag{8b}$$

As mentioned in Section 2.2, the experimental data sets include the concentrations of the main reagents and the flow rate of the feeding reagent. The molar weights (M_i) of the components are known, allowing the calculation of volume and its changes. Therefore, the only elements required to determine the change in the concentration of the components are the reaction rate and the stoichiometric coefficient. By performing an extensive grid search for each time step and using Eq. (8), it is possible to calculate υ_t and γ_t for each time step in each specific experiment. These newly calculated variables, based on experimental data, are used as ground truth to train the MLP model.

As is shown in Fig. 2, in this study, we employed a multi-layer perceptron (MLP) neural network, configuring its hidden layers with 8, 16, and 8 neurons respectively. We incorporated `ReLU` activation functions in the initial layers to effectively capture non-linear relationships, while a linear activation function in the output layer facilitates the prediction of continuous kinetic parameters. The model inputs, at each time step, include concentrations of the main reagents and catalyst, reagent B's feed rate, the reactor's temperature, and the stirrer speed, chosen for their critical roles in reaction dynamics. Our neural network features two output 'heads,' each designed to estimate one of the two parameters: reaction rate and stoichiometry coefficient. Mean Absolute Error (MAE) is used as the loss function to be minimized during the training of the model. We implemented the MLP model using the `PyTorch` library [29]. To train the model, we utilized 60% randomly selected data points of the training data set and used the remaining 40% as test data.

3.2. Estimate final products quality

As discussed in Section 3.1, using the MLP model, enables to estimate the concentration of the main reagents (*i.e.* Eqs. (2a), (2b) and (2e)). The υ_t output of the MLP model is also used in Eqs. 2c and 2d as follows:

$$\frac{dC_{sp}}{dt} = \alpha \upsilon_t - \frac{C_{sp}}{V_l}\frac{dV_l}{dt} \tag{9a}$$

$$\frac{dC_{mp}}{dt} = \beta \upsilon_t - \frac{C_{mp}}{V_l}\frac{dV_l}{dt} \tag{9b}$$

As mentioned before, equation set 2 is not independent. Therefore, by calculating either 8a or 8b, the other one is also calculated. To solve these equations at each time step, υ_t is calculated by the MLP model. Moreover, C_{mp}, C_{sp}, and V_l are the values from previous time step and dV_l/dt is calculated using 2e. Therefore, to solve equation set 8, it is sufficient to estimate either α or β.

As mentioned previously, data on the mass fraction of components are available only at the end of experiments, measured through HPLC analysis. Consequently, the experimental production rates of the main and the side products (*i.e.* α and β) remains unknown during the process and it is not straightforward to estimate them.

Assuming that the α value calculated from HPLC represents an average during the process, we explored various features to model this coefficient. We identified the three most significant features, which are strongly correlated with the final product's quality.

- Temperature difference between inlet and outlet of jacket.

Based on the predefined recipe, the reactor temperature (T^r) should follow a certain trajectory (*i.e.*, dynamic set-points). The PID controller manipulates the jacket inlet temperature (T^j) to adhere to these set-points. As suggested by Zhang et al. [30], the difference in temperature between the inlet and outlet (here equal to T^r) of the jacket indicates the amount of heat produced by the reaction. By measuring this amount of heat, it is possible to estimate the quality of the final polymer.

- Reaction rate

The reaction rate (υ_t) or consumption rate of reagents is another feature that can indicate the quality of the final polymer. This rate depends on the concentrations of the components as well as the reactor temperature and should follow a certain trajectory to maintain the quality of the polymer. Different trajectories can lead to polymers with varying chain lengths.

- Availability of component B,

In polymerization reactions, the availability of initiators (in this case, reagent B) dictates the quantity of primary radicals that control the propagation of polymer production [31]. If this propagation halts prematurely, the resulting polymer's length will be shorter than anticipated, leading to a reduction in the product's viscosity. Conversely, extended propagation can result in longer polymers and potentially cause a gel effect in the product. The availability of the initiator can be represented by the molar equivalent of component B (meq^B) and can be determined using NIR spectroscopy during the process.

Based on golden-batch experiments, we calculate the desired trends of the aforementioned features and fit a piecewise polynomial to achieve smooth trends these features. Assuming these trends lead to the desired polymer, we hypothesize that any deviation during the experiments can decrease the quality of the product. These deviations are calculated as mean absolute error (MAE) and then converted using `log` function.

Fig. 3 depicts the scatter plots of the α coefficient versus the average deviation of introduced features from the desired trends for all experiments. To prevent overfitting, a multivariate linear regression (LR) was performed to estimate the effect of various features on the side production rate (*i.e.* α).

Table 1
A summary of input and output of different part of the model and their corresponding section.

Model	Input	Output	Described in
MLP	T_t^r, $\dot{m}_t^B$, m_t^{total}, C_t^A, C_t^B, S_t	υ_t, γ_t	Section 3.1
LR	$meq_{:t}^B$, $T_{:t}^r$, $T_{:t}^j$, υ_t	α_t , β_t	Section 3.2
ODE	υ_t, γ_t, α_t , β_t , C_t^i	C_{t+1}^i	Section 2.3
Full model	T_t^r, T_t^j, $\dot{m}_t^B$, m_0, S_t	C_{t+1}^i	Section 3.3

This regression model enables us to estimate the production of both main and side products during the process. It is important to note that it is not necessary to wait until the end of the process to calculate the α coefficient; it can be calculated at each time step as a function of the deviation from the beginning of the process up to that time step. Consequently, if any deviation occurs during the process, since it is logged, its effect will not immediately fade in the next time step and will remain for a longer period. This allows us to mimic the cumulative effect of producing side products.

3.3. Full simulation model

In this study, we model the polymerization reaction as irreversible, aligning with scenarios where monomers form strong covalent bonds to produce polymer chains, rendering depolymerization highly improbable under typical conditions. This approach simplifies the modeling process by eliminating the need to account for the reverse reaction, common in radical polymerizations and others where the energy barrier for bond breaking is significantly high. Unlike reversible polymerization, where equilibrium allows for both polymer formation and breakdown, our irreversible model focuses on unidirectional reactant conversion to products, mirroring practical observations in many polymerization processes [32]. This assumption streamlines the analysis, reflecting the predominance of product formation in these reactions.

Fig. 1 illustrates the comprehensive hybrid simulation model where two data driven model (*i.e.* MLP and LR) are combined with the first-principle (*i.e.* ODE) model. The model receives initial values of the reactor, including the initial weight and temperature of component A, as well as dynamic variables such as feed rate, jacket inlet and outlet temperatures, and total weights of the reactor at time step t. These inputs are processed through the MLP model to predict the reaction rates and stoichiometry coefficients. Then, the reaction rate, along with other variables, is inputted into the LR model to estimate the side production rate. The output from the LR model, combined with other measurements, is fed into an ordinary differential equation (ODE) solver which serves as the white box model. The white box model updates the concentrations of the reactants and products in the system for time step $t + 1$. The updated concentrations are then looped back as inputs into the MLP and the LR model for continuous real-time simulation, allowing for dynamic adjustments based on the process conditions.

This iterative process ensures a detailed and adaptive representation of the system, enabling the prediction of both main and side product concentrations throughout the chemical process.

4. Results and discussion

In this section, we first investigate the efficiency of the proposed MLP model in predicting the reagent's concentration. Afterwards, the results of the regression model are presented. We then examine the robustness of the hybrid model against noise in different parts of the model.

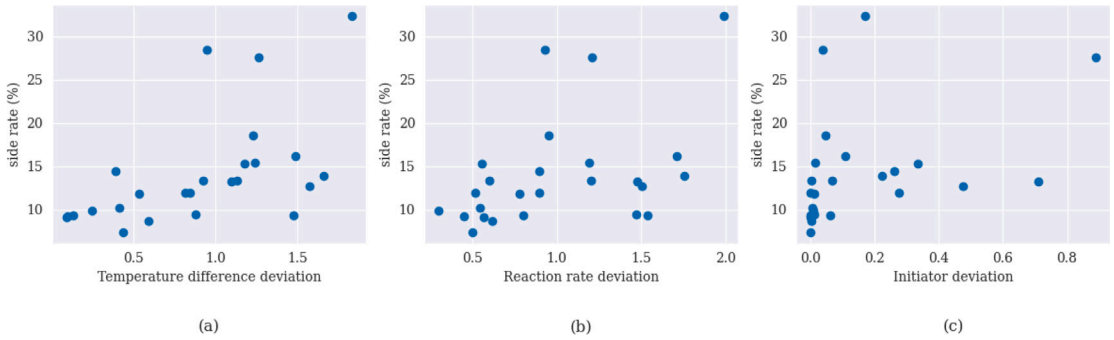

Fig. 3. Scatter plots illustrating the impact of various features on the average side reaction rate: (a) temperature difference deviation between jacket inlet and outlet, (b) reaction rate deviation, and (c) initiator availability deviation from reference trends.

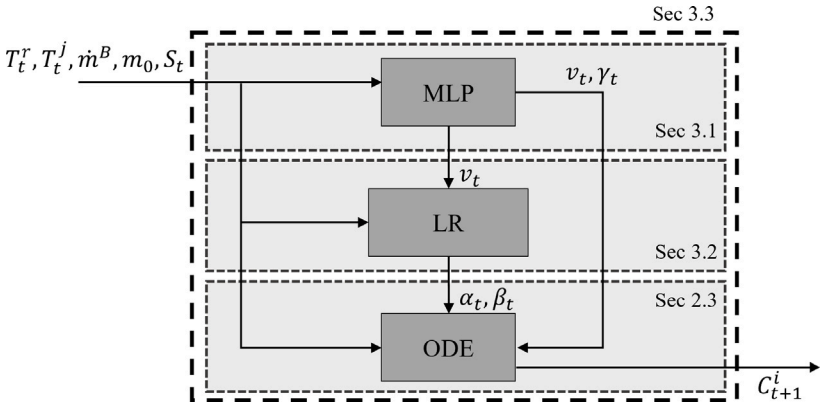

Fig. 4. Scheme of the connection of different parts of the hybrid model at time step t. The details of each part are described in the corresponding section.

Table 2
Comparison of Mean Squared Error (MSE) and Mean Absolute Error (MAE) for white-box (W) and proposed hybrid (H) models in a golden batch experiment, an experiment with acceptable outcome and a significantly deviant experiment.

		Golden batch		Acceptable outcome		Significantly deviant	
		MSE	MAE	MSE	MAE	MSE	MAE
C_A	W	**0.0023**	0.0425	0.0431	0.1931	0.4092	0.5808
	H	0.0025	**0.0366**	**0.0338**	**0.1643**	**0.0101**	**0.0877**
C_B	W	**0.0011**	**0.0297**	0.0407	0.1903	0.4293	0.5997
	H	0.0015	0.0398	**0.0359**	**0.1709**	**0.0084**	**0.0782**

4.1. Reaction rate and stoichiometry coefficient model

Table 2 compares the MSE and MAE of predicted values for C_A and C_B across ideal and non-ideal experiments. The table shows that, in the ideal experiment, the MSE for both C_A and C_B slightly increases when the hybrid model is applied. Conversely, in the non-ideal experiment, these values decrease by approximately 22% when applying the proposed hybrid model. A similar trend is observed for the MAE values.

Fig. 5 compares the predicted values for component A with experimental measurements under different experimental conditions. Fig. 5a illustrates the results from a golden batch experiment, where, as expected, the hybrid model maintains the accuracy of the white-box model. Fig. 5b and c present the same values for a non-ideal experiments. In this scenario, the predictions made by the hybrid model are more accurate and closer to the experimental values than those made by the white-box model. Fig. 6, illustrate the same comparison for the concentration of component B. As we expected, the hybrid model, significantly improve the accuracy of the predictions for non-ideal experiments while preserve the accuracy for ideal experiment.

Table 3
Comparison of MP and SP between Simulation and HPLC for a golden batch experiment, an experiment with acceptable outcome and a significantly deviant experiment.

	Golden batch		Acceptable outcome		Significantly deviant	
	MP %	SP %	MP %	SP %	MP %	SP%
Simulation	84.8	11.3	78.1	17.9	62.5	27.3
HPLC	82	10.3	77.1	19.2	65.9	28.3

4.2. Product quality estimation

Table 3 presents the results of a full model simulation and compares the percentages of the main product (MP) and the side product (SP) with the experimental values obtained via HPLC for both an ideal and a non-ideal experiment. As indicated in this table, the simulation results show deviations of 1% for the MP and 1.7% for the SP in the ideal experiment. These deviations fall within an acceptable range for non-ideal experiments, at 2% for the MP and 7.5% for the SP, respectively.

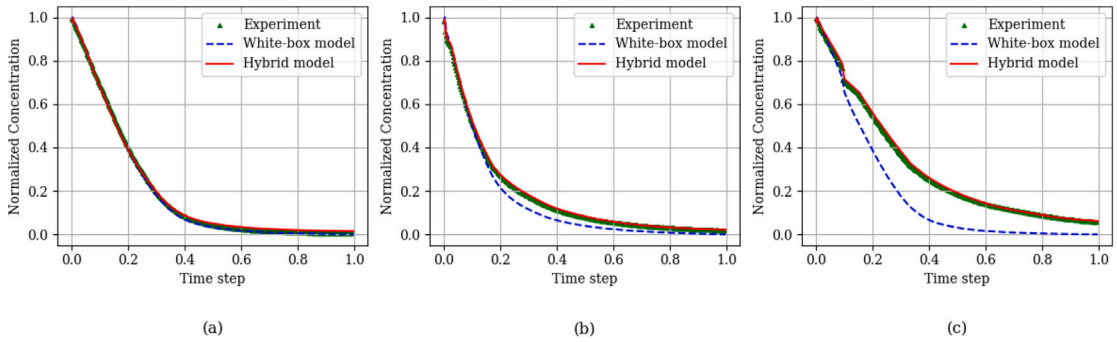

Fig. 5. Comparison of Predicted vs. Experimental Concentrations of Component *A* under Three Experimental Conditions: (a) Golden Batch, (b) Acceptable Outcome, and (c) Significantly Deviant Outcome.

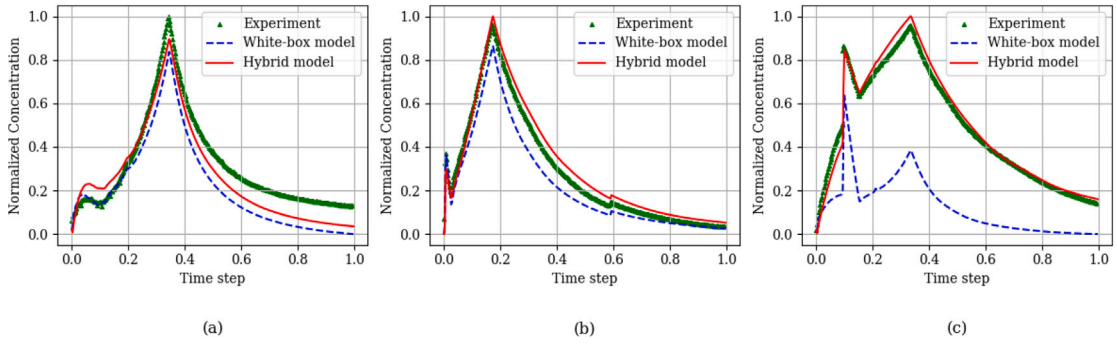

Fig. 6. Comparison of Predicted vs. Experimental Concentrations of Component *B* under Three Experimental Conditions: (a) Golden Batch, (b) Acceptable Outcome, and (c) Significantly Deviant Outcome.

4.3. Side reaction rate

Since the model directly estimates the production rates of both the main and side products, we can log and monitor this data throughout the process. Fig. 7 compares the side product rates from a golden batch experiment with those from two non-ideal experiments. As illustrated, the side product rate for the golden batch experiment is consistently below 10%, with an average of about 7%. These rates are significantly higher in non-ideal experiments.

It is worth noting that the variability observed in Fig. 7 reflects differences in the execution of the experimental procedures, despite similar initial and reactor conditions across different experiments. These variations in execution can influence the production rates of both the main and side products. While these specific values could not be directly validated due to limited data, the model's weighted average predictions remain within an acceptable range, indicating the physical plausibility of the results.

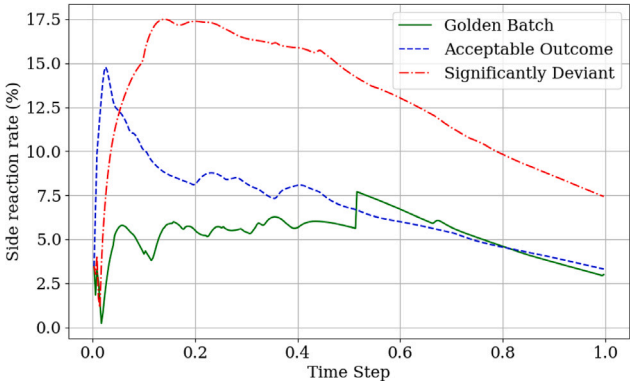

Fig. 7. Estimation of the rate of creating side product during the process for different experimental conditions.

4.4. Sensitivity analysis

In this section, we investigate the robustness of the proposed model against different levels of noise in various parts of the model. The sensitivity of the reagents to the reaction rate and reactor temperature is discussed in Section 4.4.1 and Section 4.4.2, respectively. The sensitivity of predicting the concentration of the final products is discussed in Section 4.4.3.

4.4.1. Sensitivity of reagent concentrations to reaction rate

First, we investigate the sensitivity of reagents concentration predictions to the reaction rate by introducing different levels of noise into the reaction rate estimates generated by the MLP model. Fig. 8 illustrates the average error in predicting the concentrations of the main reagents over 10 repetitions of the experiment, with the reaction rate subjected to various levels of noise. This figure demonstrates that the concentration of component *A* is more sensitive to changes in the reaction rate, compared to the concentration of component *B*. This figure suggests that predicting the concentration is robust against the noise in the reaction rate, up to a level of 20%. It also shows that while adding up to 30% noise does not significantly increase the average error, there is a noticeable rise in the standard error (shadow in Fig. 8).

4.4.2. Sensitivity of reagent concentrations to the temperature

Fig. 9 illustrates the average error in predicting the concentrations of the main reagents over 10 repetitions of the experiment when the temperature is subjected to various levels of noise. This figure shows that the model's ability to predict concentration is more sensitive to temperature noise than to noise in the reaction rate. While an up to 30% noise level in the reaction rate results in an acceptable average error – implying the model's predictions remain reliable – the model maintains robustness up to only 5% noise in temperature.

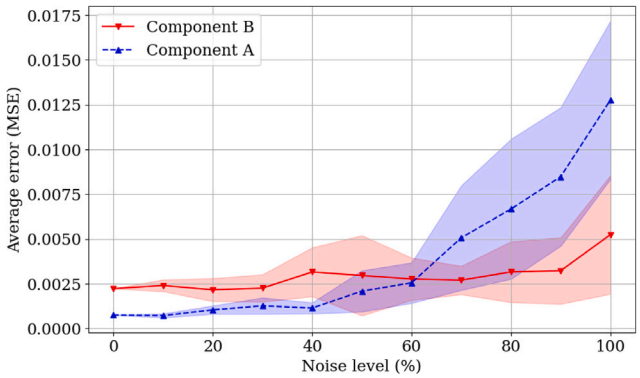

Fig. 8. Sensitivity of the concentration of the main reagents to the noise of reaction rate.

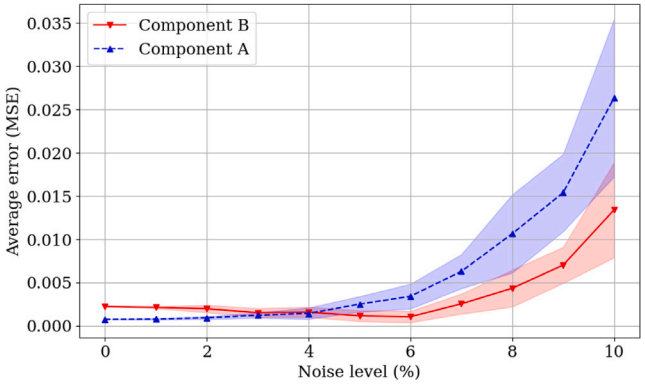

Fig. 9. Sensitivity of the concentration of the main reagents to the different level of noise in reactor temperature.

4.4.3. Sensitivity of final products to the temperature

To investigate the robustness of a regression model's performance in predicting the concentration of final products, we introduced different levels of noise to the reactor temperature. Fig. 10 illustrates the discrepancy between the measured and predicted concentrations of the main and side products. As shown in this figure, with an increase in noise (or deviation in reactor temperature), the predicted concentrations begin to diverge from their initial values. Up to a 4% noise level, this divergence increases linearly, and the standard error for 10 repetitions at the same noise level is relatively small. Beyond this noise level, the divergence accelerates rapidly, and the standard error also increases. After exceeding 8% noise, the standard error decreases, indicating that the model becomes more certain about predicting the creation of side products over main products.

5. Conclusion

In this study, we presented a novel combination of a hybrid and a regression model to predict the concentrations of reagents as well as the quality of the final product of a polymerization fed-batch reactor. As discussed, utilizing a neural network to model the reaction rate and stoichiometry coefficient can noticeably increase the model's performance. As an example the MAE is reduced by 84% and 86% when predicting the concentration of components A and B, respectively, in an experiment with significantly deviant outcome. The improved performance of the hybrid model over the white-box model is due to its ability to adapt to different conditions by utilizing additional input parameters, thereby capturing complexities that the white-box model cannot. Moreover, we introduced three features that are more informative for predicting the final product quality. These three features were combined through

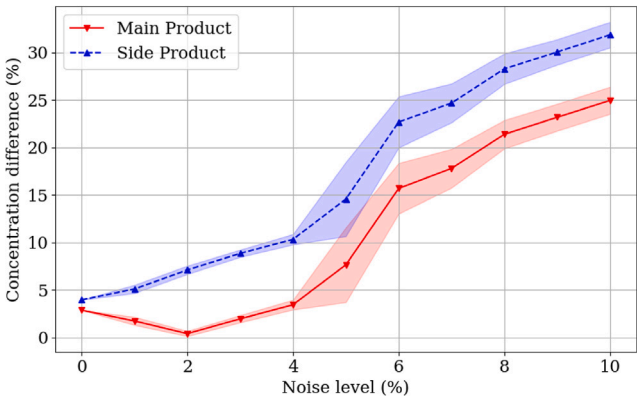

Fig. 10. Sensitivity of predicting the main and side product concentration to the different level of noise in reactor temperature.

a linear regression model to predict the rate of producing the main and side products. We investigated the effect of these features and the performance of the regression model in different experiments. The error for predicting the percentage of main and side products is maximum 3.5% which is an acceptable rate for these kind of reactions in the literature.

Additionally, the introduced model can run simultaneously with industrial settings. This capability offers an opportunity to monitor the reactor and detect malfunctions immediately, preventing detrimental phenomena in the process, such as gelation.

Although the proposed procedure has good performance in experiments with the same recipe, the model is not generalizable to different recipes. To be able to generalize to different recipes, the model need only a few golden-batch experiment datasets following the desired recipes.

Moreover, from Fig. 3, we can see that the effects of different features are not linear, and a more complex model could improve the prediction power, especially for experiments where these deviations are relatively high. For future work, we suggest to investigate more informative features and utilize polynomial regression to model the side product rate. To this end, more data from significantly deviant experiments are required.

The proposed approach was designed to be compatible with the simple first principle equations following ordinary differential equations (ODE). Another direction for further research is to generalize the applicability of this approach to systems dealing with differential algebraic equations (DAE) and partial differential equations (PDE).

Our future plan includes using our full model as an environment to train a reinforcement learning (RL) agent to optimize and control the process.

CRediT authorship contribution statement

Omid Sobhani: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Hamid Toliati:** Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Furkan Elmaz:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Shahab Pormoradi Gerdposhteh:** Writing – review & editing, Resources, Data curation, Conceptualization. **Benedict Carius:** Writing – review & editing, Resources, Data curation, Conceptualization. **Kevin Mets:** Writing – review & editing, Supervision. **Siegfried Mercelis:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used OpenAI ChatGPT in order to proofread. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of competing interest

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

Acknowledgments

This work was performed within the CHAI imec icon project. It was realized with the financial support of Flanders Innovation & Entrepreneurship, Belgium (VLAIO, project no. HBC.2021.0671).

Data availability

The data that has been used is confidential.

References

[1] S.R. Ponnuswamy, S.L. Shah, C.A. Kiparissides, Computer optimal control of batch polymerization reactors, Ind. Eng. Chem. Res. 26 (11) (1987) 2229–2236, http://dx.doi.org/10.1021/IE00071A010.

[2] K. Choudhary, B. DeCost, C. Chen, A. Jain, F. Tavazza, R. Cohn, C.W. Park, A. Choudhary, A. Agrawal, S.J.L. Billinge, E. Holm, S.P. Ong, C. Wolverton, Recent advances and applications of deep learning methods in materials science, npj Comput. Mater. 8 (1) (2022) 59, http://dx.doi.org/10.1038/s41524-022-00734-6.

[3] N.C. Frey, R. Soklaski, S. Axelrod, S. Samsi, R. Gómez-Bombarelli, C.W. Coley, V. Gadepally, Neural scaling of deep chemical models, Nat. Mach. Intell. 5 (11) (2023) 1297–1305, http://dx.doi.org/10.1038/s42256-023-00740-3.

[4] L.-T. Zhu, X.-Z. Chen, B. Ouyang, W.-C. Yan, H. Lei, Z. Chen, Z.-H. Luo, Review of machine learning for hydrodynamics, transport, and reactions in multiphase flows and reactors, Ind. Eng. Chem. Res. 61 (28) (2022) 9901–9949, http://dx.doi.org/10.1021/acs.iecr.2c01036.

[5] L. Peterson, J. Bremer, K. Sundmacher, Challenges in data-based reactor modeling: A critical analysis of purely data-driven and hybrid models for a CSTR case study, Comput. Chem. Eng. 184 (2024) 108643, http://dx.doi.org/10.1016/j.compchemeng.2024.108643.

[6] N. Sharma, Y.A. Liu, A hybrid science-guided machine learning approach for modeling chemical processes: A review, AIChE J. 68 (5) (2022) http://dx.doi.org/10.1002/AIC.17609.

[7] S. Zendehboudi, N. Rezaei, A. Lohi, Applications of hybrid models in chemical, petroleum, and energy systems: A systematic review, Appl. Energy 228 (2018) 2539–2566, http://dx.doi.org/10.1016/j.apenergy.2018.06.051.

[8] J. Sansana, M.N. Joswiak, I. Castillo, Z. Wang, R. Rendall, L.H. Chiang, M.S. Reis, Recent trends on hybrid modeling for Industry 4.0, Comput. Chem. Eng. 151 (2021) 107365, http://dx.doi.org/10.1016/j.compchemeng.2021.107365.

[9] F.S. Mjalli, A.S. Ibrehem, Optimal hybrid modeling approach for polymerization reactors using parameter estimation techniques, Chem. Eng. Res. Des. 89 (7) (2011) 1078–1087, http://dx.doi.org/10.1016/j.cherd.2010.11.018.

[10] U. Di Caprio, M. Wu, F. Elmaz, Y. Wouters, N. Vandervoort, A. Anwar, S. Mercelis, S. Waldherr, P. Hellinckx, M.E. Leblebici, Hybrid modelling of a batch separation process, Comput. Chem. Eng. 177 (2023) 108319, http://dx.doi.org/10.1016/J.COMPCHEMENG.2023.108319.

[11] M.S.F. Bangi, J.S.-I. Kwon, Deep hybrid model-based predictive control with guarantees on domain of applicability, AIChE J. 69 (5) (2023) e18012, http://dx.doi.org/10.1002/aic.18012.

[12] S. Pahari, P. Shah, J. Sang-Il Kwon, Unveiling latent chemical mechanisms: Hybrid modeling for estimating spatiotemporally varying parameters in moving boundary problems, Ind. Eng. Chem. Res. 63 (3) (2024) 1501–1514, http://dx.doi.org/10.1021/acs.iecr.3c03531.

[13] N. Sitapure, J. Sang-Il Kwon, Introducing hybrid modeling with time-series-transformers: A comparative study of series and parallel approach in batch crystallization, Ind. Eng. Chem. Res. 62 (49) (2023) 21278–21291, http://dx.doi.org/10.1021/acs.iecr.3c02624.

[14] P. Shah, M.Z. Sheriff, M.S.F. Bangi, C. Kravaris, J.S.-I. Kwon, C. Botre, J. Hirota, Deep neural network-based hybrid modeling and experimental validation for an industry-scale fermentation process: Identification of time-varying dependencies among parameters, Chem. Eng. J. 441 (2022) 135643, http://dx.doi.org/10.1016/j.cej.2022.135643.

[15] A.M. Schweidtmann, D. Zhang, M. von Stosch, A review and perspective on hybrid modeling methodologies, Digit. Chem. Eng. 10 (2024) 100136, http://dx.doi.org/10.1016/J.DCHE.2023.100136.

[16] S. Curteanu, Machine learning techniques applied to a complex polymerization process, in: Machine Learning in Chemistry: The Impact of Artificial Intelligence, The Royal Society of Chemistry, 2020, http://dx.doi.org/10.1039/9781839160233-00227.

[17] Y. Tian, J. Zhang, J. Morris, Modeling and optimal control of a batch polymerization reactor using a hybrid stacked recurrent neural network model, Ind. Eng. Chem. Res. 40 (21) (2001) 4525–4535, http://dx.doi.org/10.1021/ie0010565.

[18] J.-S. Chang, S.-C. Lu, Y.-L. Chiu, Dynamic modeling of batch polymerization reactors via the hybrid neural-network rate-function approach, Chem. Eng. J. 130 (2007) 19–28, http://dx.doi.org/10.1016/j.cej.2006.11.011.

[19] C.-G. Piuleac, S. Curteanu, G. Asachi, Different methods of neural network based modelling for polymerization process, Mater. Plast. 47 (3) (2010).

[20] S. Curteanu, F. Leon, Hybrid neural network models applied to a free radical polymerization process, Polym.-Plast. Technol. Eng. 45 (9) (2006) 1013–1023, http://dx.doi.org/10.1080/03602550600726285.

[21] G. Merfeld, C. Molaison, R. Koeniger, A.E. Acar, S. Mordhorst, J. Suriano, P. Irwin, R.S. Warner, K. Gray, M. Smith, K. Kovaleski, G. Garrett, S. Finley, D. Meredith, M. Spicer, T. Naguy, Acid/epoxy reaction catalyst screening for low temperature (120°C) powder coatings, Prog. Org. Coat. 52 (2) (2005) 98–109, http://dx.doi.org/10.1016/j.porgcoat.2004.09.004.

[22] A. Jain, J. Mao, K. Mohiuddin, Artificial neural networks: a tutorial, Computer 29 (3) (1996) 31–44, http://dx.doi.org/10.1109/2.485891.

[23] G.S. Fesaghandis, A. Pooya, M. Kazemi, Z.N. Azimi, Comparison of multilayer perceptron and radial basis function in predicting success of new product development, Eng. Technol. Appl. Sci. Res. 7 (2017).

[24] L. Noriega, Multilayer perceptron tutorial, Sch. Computing. Staffordshire Univ. 4 (5) (2005) 444.

[25] E. Shim, J.A. Kammeraad, Z. Xu, A. Tewari, T. Cernak, P.M. Zimmerman, Predicting reaction conditions from limited data through active transfer learning, Chem. Sci. 13 (2022) 6655–6668, http://dx.doi.org/10.1039/D1SC06932B.

[26] F. Elmaz, S. Ghane, T. Huybrechts, A. Anwar, S. Mercelis, P. Hellinckx, Transfer learning-based hybrid modeling approach for indoor temperature modeling, in: IECON 2022 – 48th Annual Conference of the IEEE Industrial Electronics Society, 2022, pp. 1–6, http://dx.doi.org/10.1109/IECON49645.2022.9968939.

[27] M. von Stosch, R. Oliveira, J. Peres, S. Feyo de Azevedo, Hybrid semi-parametric modeling in process systems engineering: Past, present and future, Comput. Chem. Eng. 60 (2014) 86–101, http://dx.doi.org/10.1016/j.compchemeng.2013.08.008.

[28] A. Afram, F. Janabi-Sharifi, Gray-box modeling and validation of residential HVAC system for control system design, Appl. Energy 137 (2015) 134–150, http://dx.doi.org/10.1016/j.apenergy.2014.10.026.

[29] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala, PyTorch: An imperative style, high-performance deep learning library, in: H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché Buc, E. Fox, R. Garnett (Eds.), Advances in Neural Information Processing Systems 32, Curran Associates, Inc., 2019, pp. 8024–8035.

[30] J. Zhang, E.B. Martin', A.J. Morris, C. Kiparissides, Inferential estimation of polymer quality using stacked neural networks, Comput. Chem. Eng. 21 (1997) 1025–1030.

[31] L. Ghiba, E.N. Drăgoi, S. Curteanu, Neural network-based hybrid models developed for free radical polymerization of styrene, Polym. Eng. Sci. 61 (3) (2021) 716–730, http://dx.doi.org/10.1002/pen.25611.

[32] G. Odian, Principles of Polymerization, John Wiley & Sons, 2004.