Chemical Engineering Research and Design 200 (2023) 538–549

Contents lists available at ScienceDirect

Chemical Engineering Research and Design

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

Improved modeling of crystallization processes by Universal Differential Equations

Fernando Arrais R.D. Lima [a,*], Carine M. Rebello [b], Erbet A. Costa [b], Vinícius V. Santana [b], Marcellus G.F. de Moares [c,d], Amaro G. Barreto Jr. [a], Argimiro R. Secchi [a,c], Maurício B. de Souza Jr. [a,c], Idelfonso B.R. Nogueira [b]

[a] Escola de Química, EPQB, Universidade Federal do Rio de Janeiro, P.O. Box 68542, Rio de Janeiro, RJ, 21941-909, Brazil
[b] Chemical Engineering Department, Norwegian University of Science and Technology, Trondheim, 793101, Norway
[c] Programa de Engenharia Química, PEQ/COPPE, Universidade Federal do Rio de Janeiro, PO Box 68502, Rio de Janeiro, RJ, 21941-972, Brazil
[d] Instituto de Química, Rio de Janeiro State University (UERJ), Rua São Francisco Xavier, 524, Maracanã, Rio de Janeiro, RJ, 20550-900, Brazil

ARTICLE INFO

Keywords:
UDE
Machine learning
Neural network
Batch process
Potassium sulfate

ABSTRACT

Crystallization is a crucial process in the pharmaceutical industry, usually modeled by Population Balance Models (PBM). This study introduces a novel approach, combining PBM with machine learning techniques, specifically the Universal Differential Equation (UDE), to describe a batch cooling crystallization process. Unlike conventional methods, UDE eliminates the need for defining a supersaturation term and was applied to model the nucleation, growth, and dissolution of potassium sulfate in water. In this study, three UDE models for supersaturated condition and one for undersaturated condition were developed, aiming to predict particle count, length, surface area, volume, and concentration. Initially, these models were trained using all available experimental data from a previous study for the potassium sulfate batch crystallization. They were validated with simulated data generated by the PBM. In a second scenario, the models were trained with a subset of the experimental data and tested with the remainder. The UDE models performed efficiently, exhibiting similar mean squared error and mean absolute error compared to conventional PBMs. Notably, the UDE approach demonstrated an advantage in requiring a smaller training dataset. This innovative coupling of UDE and PBM offers a less complex yet effective model to capture essential aspects of crystallization kinetics and variables' dynamics. Particularly, UDE models excel in predicting nucleation kinetics, making them a valuable alternative to conventional representations. Furthermore, this approach can be easily adapted to various crystallization processes involving different phenomena, enhancing its versatility and potential impact.

1. Introduction

Crystallization is a separation and purification process, in which a solid crystalline product can be obtained from a solution (Zhang et al., 2020). This process has been applied in different parts of the industry, such as in the pharmaceutical area for medicines fabrication and the production of food and agrochemicals (Ahn et al., 2022; Moraes et al., 2023a). A common way to run a crystallization process involves a batch cooling strategy (Su and Ward, 2019).

In a crystallization process, it is crucial to produce crystals with the desired size and morphology (Nagy and Braatz, 2012). Therefore, crystallization models are developed to predict this crystal's characteristics, estimate kinetic parameters, and optimize the operation conditions (Lewis et al., 2015). The most applied phenomenological approach for modeling crystallization processes uses population balance models (Braatz and Hasebe, 2002). Phenomenological models, rooted in conservation laws, integrate some heuristic elements to depict specific systems. These elements, symbolized by model parameters and the mass balance's sink/source term, often rely on empirical models. In crystallization, constitutive equations for nucleation, growth, agglomeration, and dissolution rates are coupled to population balance models (PBM). Regarding the use of these models, estimating kinetic parameters related to the different phenomena involved is inevitable, mainly due to the complexity involved in these processes. For this purpose, many experiments and their design are usually required, in which the determined parameters are often limited to the operational conditions and the system explicitly used.

* Corresponding author.
E-mail address: farrais@eq.ufrj.br (F.A.R.D. Lima).

https://doi.org/10.1016/j.cherd.2023.11.032
Received 30 September 2023; Received in revised form 31 October 2023; Accepted 15 November 2023
Available online 18 November 2023

Nomenclature

β	Nucleation exponent parameter
ΔC	Absolute supersaturation [g/L]
γ	Growth constant [cm^{-1}]
$\hat{B}$	Specific nucleation rate [$cm^{-3}min^{-1}$]
μ_i	ith order moment [cm^i/cm^3]
ψ	Volume fraction
ρ_c	Specific weight of the crystal [g/cm^3]
B	Nucleation rate [min^{-1}]
C	Solute concentration in solution [g/cm^3]
c_b	Nucleation exponent parameter
c_d	Dissolution exponent parameter
c_g	Growth exponent parameter
C_s	Solute equilibrium concentration [g/L]
D	Dissolution rate [cm/min]
E_{Ab}	Nucleation activation energy [J/mol]
E_{Ag}	Growth activation energy [J/mol]
G	Growth rate [cm/min]
G_0	Temperature-and-supersaturation-dependent factor in growth rate [cm/min]
k_b	Nucleation constant $[(\text{min cm}^3)^{-1}(\text{L/g})^{c_b}(\text{cm}^3/\mu\text{m}^3)^{\beta}]$
k_d	Dissolution rate constant $[(\text{cm/min})(\text{L/g})^{c_d}]$
k_g	Growth rate constant $[(\text{cm/min})(\text{L/g})^{c_g}]$
k_v	Volume shape factor
R	Universal gas constant [J/(mol K)]
r_i	Ratio between μ_i and μ_0 [cm^i]
S	Relative supersaturation
T	Temperature [K]

Many of these developments in crystallization modeling over the last few decades have focused on pharmaceutical crystallization. We can highlight the studies with paracetamol, which has become a benchmark substance for modeling investigations. Worlitschek and Mazzotti (2004) developed a PBM for the batch-cooling crystallization of paracetamol in an ethanol solution, considering crystal growth, dissolution, and nucleation. Then, they used the model to find the temperature trajectory to optimize the particle size distribution. Nagy et al. (2008) proposed an approach to estimate the kinetic parameters for the paracetamol crystallization in water and compared it to strategies presented in the literature. Ó'Ciardhá et al. (2012) developed a PBM for the paracetamol crystallization from methanol and water, considering nucleation, agglomeration, and growth. The model was developed using the quadrature method of moments and was applied to optimize the crystal size distribution and batch time. Recently, Kim et al. (2023) also developed a PBM for paracetamol crystallization from ethanol, accounting for the kinetics of primary and secondary nucleation, growth, and dissolution of crystals.

Still using PBM and crystallization kinetics estimation, McDonald et al. (2019) modeled the continuous reactive crystallization of β-lactam antibiotics and studied ways to optimize amoxicillin production. Liu et al. (2020) developed a PBM for the carbamazepine batch crystallization, evaluating different approaches to describe the nucleation and growth rates. Moraes et al. (2021) proposed a bivariate PBM for the batch crystallization of potassium dihydrogen phosphate (KDP) under supersaturated and undersaturated conditions. They also proposed an optimal control strategy to calculate temperature set points based on the optimal supersaturation policies. Trampuž et al. (2019) used the population balance approach for modeling the nucleation, dissolution, growth, and agglomeration in the fesoterodine fumarate crystallization in 2-butanone. Bosetti and Mazzotti (2020) modeled the secondary nucleation considering the potassium alum crystallization. Halfwerk et al. (2023) used a PBM for the lactose crystallization at sub-zero temperatures, including the effects of mutarotation and nucleation. Also, some works in the literature apply PBMs for the crystallization of potassium sulfate, the compound studied in the present work (Randolph and Rajagopal, 1970; Randolph and Cise, 1972; Jones et al., 1986; Budz et al., 1987).

The last decades have witnessed a growth in the commercial development of in situ real-time sensor technologies, providing more online data and computer hardware (Nagy and Braatz, 2012). Therefore, data-driven approaches appear as an efficient option for modeling and controlling chemical processes because of their ability to capture intrinsic nonlinearities and they usually demand lower computational cost compared to phenomenological models (McBride and Sundmacher, 2019; Schweidtmann et al., 2021). Despite their relevance, investigations on crystallization processes with these techniques are emerging (Xiouras et al., 2022).

Some studies present strategies for modeling crystallization processes based on machine-learning approaches. Griffin et al. (2016) proposed a machine-learning model to replace PBM for crystallization processes. They used experimental data to train the model and used it to obtain optimal control policies. Vasanth Kumar et al. (2008) developed neural networks (NNs) that efficiently predicted the crystal growth rate of sucrose. They compared the NN performance to the conventional nonlinear regression approaches, obtaining a better performance for the data-driven strategy. Recently, some studies applied recurrent neural networks (RNNs) for modeling batch cooling crystallization processes (Lima et al., 2022b,a; Moraes et al., 2023c; Zheng et al., 2022b,a).

Ma et al. (2020) proposed three semiempirical models and NNs to predict the metastable zone widths for the lithium carbonate reactive crystallization. The models were developed using only experimental data and the NNs could efficiently predict this characteristic accurately. Lima et al. (2023) developed NN models to predict the solute concentration and the first four moments for the potassium sulfate batch cooling crystallization. They used one NN to predict each of the four variables for the undersaturated and supersaturated conditions. Also, they tested the performance of the models for large prediction horizons and used them as internal models of an NMPC.

Yang and Wei (2006) developed NNs to model the growth rate and predict simultaneously the nucleation rate and the agglomeration coefficient. They considered experimental data of an antisolvent crystallization system to train the networks. Wu et al. (2023) proposed a physics-informed recurrent neural network (PIRNN) model for the aspirin crystallization process. They used only simulated data and considered the nucleation and crystal growth mechanisms to train this model. They showed that the PIRNN could efficiently model this process using less training data than a purely data-driven model. Highlighting this, Psichogios and Ungar introduced a hybrid structure in 1992, combining mechanistic models with Artificial Neural Networks (ANN) for a fed-batch reactor. The ANN was weaving with ordinary differential equations (ODE), symbolizing the mechanistic model. The resolution of ODEs had an impact on the learning process of ANNs, interweaving their numerical complexities. For two decades, this framework stood as the pinnacle of hybrid modeling due to its intricate nature and the constraints posed by computational capacity. Braake et al. (1998) delved deeper into hybridizing ODE models for chemical processes. Feyo de Azevedo et al. (1997) pinpointed the primary challenge of correlating mechanistic model solutions with ANN learning. They offered insights to tackle this, but computing limits restricted definitive solutions. Despite this, hybrid models' potential shone, elevating empirical models' reach using mechanistic knowledge. Recent computational advancements have reshaped hybrid modeling perspectives. For example, Narayanan et al.'s (2021) studies exploited a hybrid technique within Partial Differential Equations (PDE) to articulate single-component adsorption in a fixed bed unit. This methodology

excelled in modeling, sidestepping the challenge of defining appropriate mechanistic assumptions for sink/source phenomena and achieving better system descriptions with fewer experiments. Moreover, Boareto et al. (2007) proposed a hybrid model with NNs to predict the aqueous and intracellular lipase activity for lipase production under fed-batch operation with a constant mass feed rate. The proposed hybrid strategy presented a better performance than the phenomenological model.

Considering hybrid models for crystallization processes, Lauret et al. (2001) developed a hybrid model for sucrose batch evaporative crystallization. They used NNs for modeling the growth rate and adopted a loss function considering the difference between concentration values measured experimentally and the concentration obtained using the hybrid model. Georgieva et al. (2003) modeled an industrial batch evaporative crystallization process in sugarcane refining, using NNs to determine agglomeration, nucleation, and growth rates. The proposed hybrid strategy predicted values with a better agreement with the experimental data than the phenomenological model. Galvanauskas et al. (2006) applied the hybrid model for the industrial sugar crystallization process to solve an optimization problem. An objective function was minimized to define the values of the feeding rate of liquor/syrup and the steam supply rate that achieves the reference values of the crystal content and the crystal size distribution at the end of the batch cycle.

Oliveira et al. (2009) proposed a hybrid modeling strategy to calculate reaction rates using NNs and applied it to the sugar crystallization growth rate. Meng et al. (2019) developed a hybrid model for sugarcane crystallization, using NNs to calculate the nucleation, agglomeration, and growth rates. They used experimental measurements of crystal content to test the proposed approach. Nielsen et al. (2020) proposed a hybrid modeling strategy using NNs to calculate the kinetic rates trained with experimental measurements. Then, the NNs were combined with a PBM to calculate the particles' count, length, surface area, and volume. This methodology was applied for modeling a laboratory-scale lactose crystallization.

In the current work, a novel hybrid model for the crystallization process using the Universal Differential Equation (UDE) approach is proposed. UDE is a hybrid model denoted by Rackauckas et al. (2020), where a universal approximation replaces one or more terms in a differential equation. Considering applications for the chemical industry, UDE was used by Nogueira et al. (2022) for modeling a multicomponent separation by adsorption in a fixed bed column and was applied by Bangi et al. (2022) for modeling β–carotene fermentation using *Saccharomyces cerevisiae*. Although the hybrid models for crystallization processes by Georgieva et al. (2003), Galvanauskas et al. (2006), and Oliveira et al. (2009) present structure close to UDEs, the proposed training method has the advantage of accounting for more information from the phenomenological model. To train their hybrid models and obtain the parameters of the NNs, they minimize the difference between the predicted and the experimental values of the produced mass of crystals. The proposed training method has the advantage of considering the minimization of the error for the concentration, zero-order moment, and moment ratios to develop the NN to predict the rates related to crystallization. Therefore, this proposal used the flexibility of NNs considering the physical laws in the training. We applied this approach to investigate the modeling of potassium sulfate crystallization in water, using the experimental data reported by Moraes et al. (2023c) for nucleation and growth and by Moraes et al. (2023b) for dissolution. The hybrid model used NNs to predict the nucleation, dissolution, and growth kinetic rates, and the population balance equations to calculate the moments of the number density distribution and the concentration. Different structures of hybrid models were tested and compared to the performance of a single NN and to the PBM developed by Moraes et al. (2023c,b). The proposed approach has the advantage of not needing to define a proper supersaturation expression for modeling the crystallization process, which can be strongly dependent on the system and set of experiments performed when using rigorous PBM. Also, using the UDE models, a large dataset for the training was not needed, leading to a more straightforward way to represent the complex phenomena involved in crystallization.

2. Methodology

2.1. Phenomenological model of the process

From the number density distribution of crystals n $[(\mu m\ cm^3)^{-1}]$ obtained via image analysis, the first four moments $\mu_i (i = 0, 1, 2, 3)$ of the number density n were computed for the entire sampling time in each experiment, as described in Moraes et al. (2023c) for nucleation and growth experiments and in Moraes et al. (2023b) for dissolution experiments. The moments $\mu_i[(\mu m)^i/cm^3]$ are defined as

$$\mu_i(t) = \int_0^\infty n(L,t) L^i dL \tag{1}$$

The variables measured throughout each experiment were the moments μ_i and the concentration $C\,[g/cm^3]$ of K_2SO_4 in solution. Each moment of a specific order has a well-defined physical meaning: μ_0 is the total number of particles per volume of suspension; $r_1 = \mu_1/\mu_0$, $r_2 = \mu_2/\mu_0$, and $r_3 = \mu_3/\mu_0$ are the average characteristic size, surface size, and volume of the particles, respectively. The relative supersaturation S is defined as

$$S = C/C_s \tag{2}$$

The population balance model was established by applying the method of moments. For supersaturated conditions ($S \geq 1$), the phenomenological model results in a system of ordinary differential equations containing the moment equations, and solute mass balance (Eqs. (3a)–(3e)), where the specific nucleation ($\hat{B}$) and growth (G_0) rates are given in Eqs. (3f) and (3g). The PBM considers only secondary nucleation because it was observed that primary nucleation was negligible compared to secondary nucleation. Also, γ is the size-dependent growth constant and is equal to 2.390×10^{-3} μm^{-1}. k_v is the volume shape factor and is equal to 1.349, and ψ is the volume fraction of the total crystal population.

$$\frac{d\mu_0}{dt} = \hat{B} \tag{3a}$$

$$\frac{d\mu_1}{dt} = G_0\left(\mu_0 + \gamma\mu_1\right) \tag{3b}$$

$$\frac{d\mu_2}{dt} = 2G_0\left(\mu_1 + \gamma\mu_2\right) \tag{3c}$$

$$\frac{d\mu_3}{dt} = 3G_0\left(\mu_2 + \gamma\mu_3\right) \tag{3d}$$

$$\frac{dC}{dt} = \frac{-3\rho_c k_v G_0}{1-\psi}\left(\mu_2 + \gamma\mu_3\right) \tag{3e}$$

where:

$$\hat{B} = k_b \exp\left(\frac{-E_{Ab}}{RT}\right)\Delta C^{c_b}\mu_3^{\beta} \tag{3f}$$

$$G_0 = k_g \exp\left(\frac{-E_{Ag}}{RT}\right)\Delta C^{c_g} \tag{3g}$$

$$\psi = k_v\mu_3 \times 10^{-12} \tag{3h}$$

$$\Delta C = C - C_s \tag{3i}$$

$$C_s = -686.3 + 3.579T - 2.929 \times 10^{-3}T^2 \tag{3j}$$

Different from the model proposed by Moraes et al. (2023c) using the moments as state variables, the model was rearranged to use the moment ratios as state variables, as described by Eq. (4), where $B = \hat{B}/\mu_0$ is the nucleation rate. This change was done in order to the nucleation and the growth rates have similar orders of magnitude, which was necessary for the UDE training. Moreover, the nucleation rate was described only by one equation as in the proposal of Moraes et al. (2023c) for the specific nucleation rate. During the UDE development, it would not be possible to train a neural network the primary and the secondary nucleation rates independently because the predictions of the neural networks would be summed in the differential equations.

Consequently, the terms summed would not have a physical meaning individually, but only the result of this sum would have it.

$$\frac{d\mu_0}{dt} = B\mu_0 \tag{4a}$$
$$\frac{dr_1}{dt} = G_0\left(1+\gamma r_1\right) - Br_1 \tag{4b}$$
$$\frac{dr_2}{dt} = 2G_0\left(r_1+\gamma r_2\right) - Br_2 \tag{4c}$$
$$\frac{dr_3}{dt} = 3G_0\left(r_2+\gamma r_3\right) - Br_3 \tag{4d}$$
$$\frac{dC}{dt} = \frac{-3\rho_c k_v G_0}{1-\psi}\left(r_2+\gamma r_3\right)\mu_0 \tag{4e}$$

where:

$$B = k_b \exp\left(\frac{-E_{Ab}}{RT}\right)\Delta C^{c_b} r_3^{\beta}\mu_0^{\beta-1} \tag{4f}$$
$$G_0 = k_g \exp\left(\frac{-E_{Ag}}{RT}\right)\Delta C^{c_g} \tag{4g}$$
$$\psi = k_v r_3 \mu_0 \times 10^{-12} \tag{4h}$$
$$\Delta C = C - C_s \tag{4i}$$

Likewise, the phenomenological model for undersaturated conditions ($S < 1$) is described by Eq. (5), in which a dissolution rate (D) was established.

$$\frac{d\mu_0}{dt} = 0 \tag{5a}$$
$$\frac{d\mu_1}{dt} = D\mu_0 \tag{5b}$$
$$\frac{d\mu_2}{dt} = 2D\mu_1 \tag{5c}$$
$$\frac{d\mu_3}{dt} = 3D\mu_2 \tag{5d}$$
$$\frac{dC}{dt} = \frac{-3\rho_c k_v}{1-\psi} D\mu_2 \tag{5e}$$

where:

$$D = -k_d(1-S)^{c_d} \tag{5f}$$
$$\psi = k_v \mu_3 \times 10^{-12} \tag{5g}$$
$$S = C/C_s \tag{5h}$$

Eq. (5) was also modified to Eq. (6) in order to use the moment ratios as state variables instead of the moments.

$$\frac{d\mu_0}{dt} = 0 \tag{6a}$$
$$\frac{dr_1}{dt} = D \tag{6b}$$
$$\frac{dr_2}{dt} = 2Dr_1 \tag{6c}$$
$$\frac{dr_3}{dt} = 3Dr_2 \tag{6d}$$
$$\frac{dC}{dt} = \frac{-3\rho_c k_v G_0}{1-\psi} Dr_2\mu_0 \tag{6e}$$

where:

$$D = -k_d(1-S)^{c_d} \tag{6f}$$
$$\psi = k_v r_3 \mu_0 \times 10^{-12} \tag{6g}$$
$$S = C/C_s \tag{6h}$$

Similar to the model proposed by Moraes et al. (2023c), the model developed in this work did not consider breakage and aggregation phenomena. This is justified by the fact that these phenomena were not observed in the experimental batches.

2.2. Universal differential equations

The concept of Universal Differential Equations (UDEs) has its roots in the theory of universal approximation. This theory proposes that specific functions possess the capacity to replicate or depict any smooth function within a defined degree of accuracy as long as they operate within the constraints of predefined parameters, as proposed by Hornik et al. (1989).

UDEs stand as differential equations enhanced by integrating one or more universal approximators. This combination promotes a mathematical tool that incorporates the inflexibility and structure of differential equations with the flexibility and adaptability of universal approximators. This approach aims to get the advantages of differential equations and universal approximators, providing complex modeling systems with increased precision and flexibility.

Many functions and series have been recognized as universal approximators due to their adaptability and versatility. ANNs emerge as potent instruments within higher dimensions to be used as a universal approximator. This stems from their layered architecture and the application of nonlinear activation functions, demonstrating their efficacy in approximating functions within multidimensional spaces. The ability of ANNs to accommodate change, learn from data, and generalize their characteristics is the optimal choice for diverse high-dimensional approximation tasks, as this study illustrates.

Fig. 1 shows the essence of a UDE using ANNs as its universal approximation for the potassium sulfate batch crystallization, in which $\boldsymbol{r} = \{r_1, r_2, r_3\}$. This approach comprises two parts: one based on established physics, signifying existing structural knowledge, and another symbolizing the unknown physics, appropriately replaced by an ANN. Consequently, the ANN's output becomes an internal tool for calculating terms related to the Ordinary Differential Equation (ODE). Finally, the last prediction materializes through a numerical approach that resolves this ODE. Parameters of the ANN can be estimated through data sources, including experimental data, coupled with a loss function and a suitable optimization method. This approach highlights the connection between phenomenological modeling and the capabilities of ANNs, demonstrating their use in complex areas and linking theory with data-driven insights.

In Fig. 1, $NN(\mu_0, \boldsymbol{r}, C, T, \theta)$ represents the contribution from a neural network model with parameters θ, which aims to account for the unknown physics in the system. The arrow leads to an equation that breaks down the neural network's approximation into composite functions:

$$NN(\mu_0, \boldsymbol{r}, C, T, \theta) = (f_1 o f_2)(\mu_0, \boldsymbol{r}, C, T, \theta) \tag{7}$$

This suggests that the unknown physics of the neural network models may be represented as a composition of functions f_1 and f_2. The optimization equation $\theta = argmin_\theta \mathcal{L}\left(\theta, \left|\mu_0 - \mu_0^*\right|^2, ||\boldsymbol{r} - \boldsymbol{r}^*||^2, |C - C^*|^2\right)$ suggests that the parameters θ are determined by minimizing a cost function $\mathcal{L}$ based on the differences between the predicted values from the model $(\mu_0, \boldsymbol{r}, C)$ and the experimental or desired values $(\mu_0^*, \boldsymbol{r}^*, C^*)$.

2.2.1. Model identification

The methodology applied to develop the UDE is described in Fig. 2. The process starts with an input u, which feeds into a hybrid model, combining a neural network and an ordinary differential equation. The hybrid model outputs value, represented by α, to the ODE, then makes a prediction labeled as y^P. A decision is made based on whether training is needed:

- if "yes", the prediction y^p goes into an optimizer, which calculates the loss or error between the prediction y^p and the true value y^e using the given squared error formula. The optimizer then learns the function and presumably updates the parameters.
- if "no", the prediction y^p is directly taken as the output.

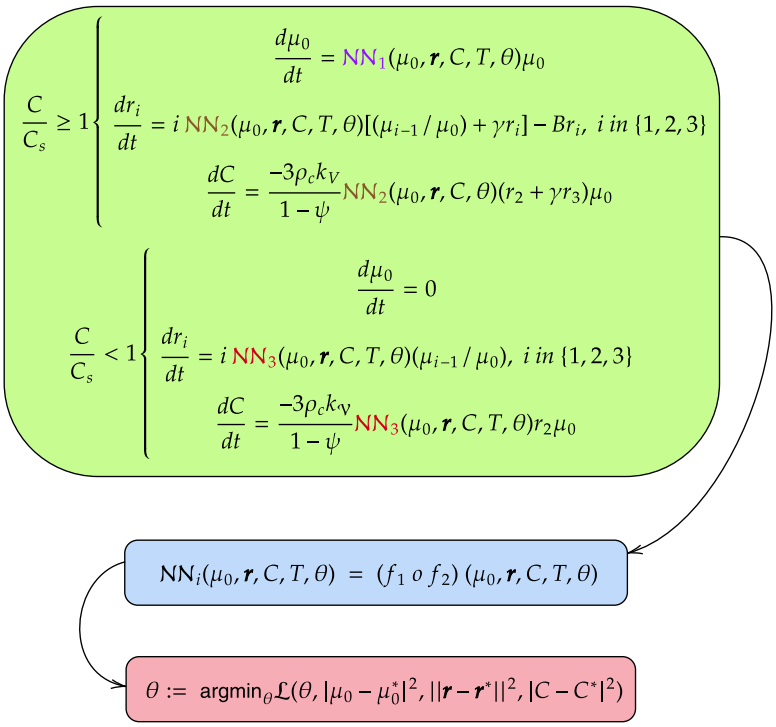

Fig. 1. Illustration of UDE approach for the potassium sulfate batch crystallization.

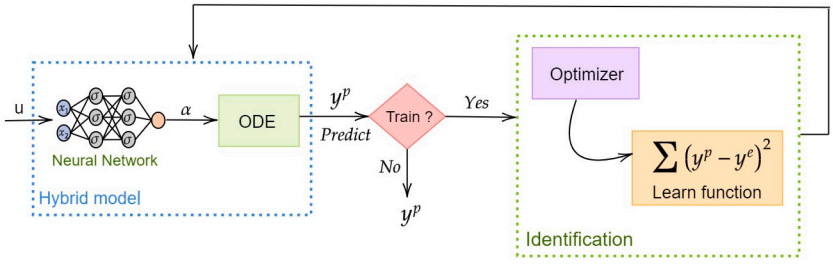

Fig. 2. Methodology for UDE employing ANN.

In this work, the UDE approach was used to model the potassium sulfate batch crystallization, considering nucleation, crystal growth, and dissolution. The structure of the UDE consists of predicting the nucleation, growth, and dissolution rates using the NNs, while the concentration (C), the zero-order moment (μ_0), and the moment ratios (r_1, r_2, and r_3) are predicted by the ODE system with the information provided by the NNs. Therefore, the optimization problem consisted of finding the parameters of the NNs that minimize the error between the predictions of the ODE (C_i^p, $\mu_{0_j}^p$, and $r_{i_j}^p$) system and the experimental values (C_i^e, $\mu_{0_j}^e$, and $r_{i_j}^e$) as described in Eq. (8). In this equation, N_e is the number of experimental values, and W and B are the vectors of weights and biases from the NN, respectively.

$$\min_{B,W} \sum_{j=1}^{N_e} \left(C_i^p - C_i^e\right)^2 + \sum_{j=1}^{N_e} \left(\mu_{0_j}^p - \mu_{0_j}^e\right)^2 + \sum_{i=1}^{3} \sum_{j=1}^{N_e} \left(r_{i_j}^p - r_{i_j}^e\right)^2 \tag{8}$$

For the supersaturated condition, three UDE models were developed, considering different strategies. First, a UDE with an NN was elaborated to predict only the nucleation rate (UDE-B); second, a UDE was built to predict only the growth rate (UDE-G); and third, a UDE constituted of two NNs to predict both rates (UDE-B-G). The inputs of the NNs to predict the growth rate were the temperature and the concentration, while the inputs of the NNs to predict the nucleation rate were the zero-order moment, the volume of the particles, the temperature, and the concentration. These inputs were chosen based on the nucleation and growth rates shown in Eq. (4). Considering the undersaturated condition, a UDE model was developed composed of an NN to predict the dissolution rate (UDE-D). The inputs of the NN to predict the dissolution rate were the temperature and the concentration. The inputs of this NN were chosen according to the dissolution rate in Eq. (6). Moreover, the hyperparameters used for the NNs in the UDEs were chosen by a sensitivity analysis for selecting the number of neurons in the hidden layer and the activation function in this layer.

Instead of using the SciML from Rackauckas et al. (2020) to develop the UDE, the hybrid models were developed using an algorithm developed in Python programming language. This decision was made because SciML was developed in Julia programming language, and we have been developing our material in Python. The optimization problem was solved using the sequential least squares programming (SLSQP) algorithm from the SciPy library (Virtanen et al., 2020) with an absolute tolerance value equal to 10^{-8}. The SLSQP presented an efficient performance to solve the optimization problem after comparing it with other approaches, such as Powell's method. The variables were normalized in the range of 0 to 1 in the objective function. To solve the ODE system, the integrator *odeint* from the SciPy library (Virtanen et al., 2020) was applied, which used the *lsoda* from the FORTRAN library odepack. In the integrator, the absolute and the relative tolerances were defined as 10^{-8}.

The performance of the UDE models was compared to a single Multilayer Perceptron NN (MLP) to predict the zero-order moment,

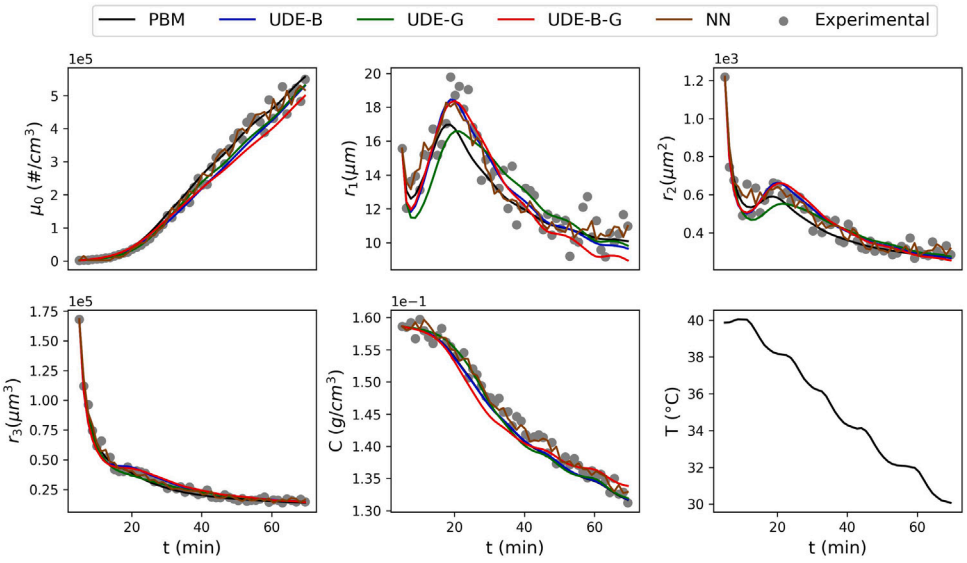

Fig. 3. Comparison between the NNs, the phenomenological model, and the three UDE structures trained with all experiments to predict the state variables under the supersaturated condition, considering experimental data for one of the batches.

the moment ratios, and the concentration, and to the PBMs developed by Moraes et al. (2023c,b). The NN was developed using the Keras library from Python (Chollet et al., 2015). Two NNs were developed: one for the supersaturated condition and another for the undersaturated condition. The inputs of these NNs were the temperature, the total number of particles per volume of suspension, the average characteristic length, the surface area, the volume of the particles, and the concentration. These NNs were trained using the Adam algorithm, defining the mean squared error as the loss function.

Initially, the machine learning models were trained with all experimental data available. Therefore, the models under the supersaturated condition were trained with ten experimental batches, and the dissolution models were trained with five experimental batches. The models were tested with simulated batches. Twenty batches were simulated; ten were conducted under the supersaturated condition, and the rest were under the undersaturated condition. Different initial conditions were defined in these simulated batches based on the experimental conditions. Also, some batches were simulated with temperature variation, and others were developed considering a constant temperature.

The UDE approach has the advantage of needing less data to develop an efficient model than the regular NN. The machine learning models were retrained using less experimental data for the training to demonstrate this fact. Therefore, the models under the supersaturated condition were trained with seven batches, and the dissolution models were trained with four batches. The remaining three batches under the supersaturated condition and the single batch under the undersaturated condition were used to validate these models. The codes and data used to develop the UDEs are available in the directory developed by Lima (2023).

3. Results and discussion

In this section, the performances of the UDE using a single NN to predict only the nucleation rate (UDE-B), the UDE using a single NN to predict only the growth rate (UDE-G), and the UDE using two NNs to predict both rates (UDE-B-G) are demonstrated. Their predictions for the zero-order moment, the moment ratios, and the concentration are compared to a single NN and the PBM developed by Moraes et al. (2023c). The performance of the UDE using one network to predict the dissolution rate (UDE-D) is also shown and compared to a single NN and the PBM developed by Moraes et al. (2023b). Even the PBM assuming the simplification of not considering the activity coefficients in the supersaturation equations, the model is based on the most used approach for modeling crystallization processes in the literature (Quilló et al., 2021). Therefore, the performance of the UDE is compared to a modeling strategy commonly used in the literature. Initially, the training was done using all the experimental batches available, and the models were tested with simulated data. Then, the machine learning models were trained with part of the experimental batches, using the remaining batches for testing.

3.1. Training with all experimental batches

The UDE using the NNs to predict the nucleation rate was trained (UDE-B) changing the number of neurons and activation functions. The hyperparameters chosen for the machine learning model were a single hidden layer with 20 neurons and the hyperbolic tangent as the activation function, which was shown to be enough to address the universal approximation theorem. The same hyperparameters were used for the other two UDE cases, UDE-G and UDE-B-G. For the NN trained for predicting the zero-order moment, the moment ratios, and the concentration, one hidden layer with 40 neurons and ReLU as the activation function were employed. Also, ReLU was used as the activation function of the output layer, and a batch size equal to 60 was adopted.

Fig. 3 presents the crystallization models' performance for nucleation and crystal growth for one experimental batch. All models could efficiently predict the zero-order moment, the concentration, and the moment ratios. However, the model based on a single NN showed an oscillatory response, which indicates that this model was trying to model the experimental noise, resulting in this overfitting problem. This behavior of the NN is explained by the small dataset used for training, and this approach needs a big dataset to be developed. Considering the UDE cases, they presented a similar performance to the phenomenological model. This confirms an advantage of UDE compared to a full empirical model as the regular NN, which is the need for less data for the development of an efficient model.

Table 1 presents the mean squared error (MSE) and mean absolute error (MAE) for the training and test datasets considering normalized values of the concentration, zero-order moment, and moment ratios. Considering the training dataset, all machine learning approaches presented MSE and MAE values close to the phenomenological model. The NN presented the best performance, reaching the lowest MSE and MAE values for μ_0, C, and r_3 for this case. However, the NN performance

Table 1
MSE and MAE values for the normalized training and test datasets of the models for nucleation and crystal growth trained with ten experimental batches.

MSE for the training dataset					
Case	μ_0	r_1	r_2	r_3	C
UDE-B	1.038×10^{-3}	5.204×10^{-3}	3.385×10^{-3}	9.402×10^{-4}	6.144×10^{-4}
UDE-G	8.177×10^{-4}	6.425×10^{-3}	4.275×10^{-3}	8.296×10^{-4}	1.127×10^{-3}
UDE-B-G	2.055×10^{-3}	6.387×10^{-3}	4.216×10^{-3}	1.114×10^{-3}	1.123×10^{-3}
NN	5.706×10^{-4}	4.035×10^{-3}	2.136×10^{-3}	4.456×10^{-4}	3.858×10^{-4}
PBM	1.130×10^{-3}	4.638×10^{-3}	2.708×10^{-3}	5.114×10^{-4}	6.729×10^{-4}
MSE for the test dataset					
Case	μ_0	r_1	r_2	r_3	C
UDE-B	5.916×10^{-3}	5.474×10^{-3}	4.464×10^{-3}	1.811×10^{-3}	8.226×10^{-5}
UDE-G	1.066×10^{-3}	3.143×10^{-3}	2.485×10^{-3}	1.368×10^{-3}	1.198×10^{-3}
UDE-B-G	1.255×10^{-2}	5.395×10^{-3}	4.663×10^{-3}	1.917×10^{-3}	2.799×10^{-3}
NN	6.688×10^{-4}	1.594×10^{-3}	8.546×10^{-4}	4.427×10^{-4}	6.647×10^{-3}
MAE for the training dataset					
Case	μ_0	r_1	r_2	r_3	C
UDE-B	2.169×10^{-2}	5.602×10^{-2}	4.331×10^{-2}	2.117×10^{-2}	2.020×10^{-2}
UDE-G	1.983×10^{-2}	6.140×10^{-2}	4.676×10^{-2}	1.865×10^{-2}	2.563×10^{-2}
UDE-B-G	3.042×10^{-2}	6.342×10^{-2}	4.976×10^{-2}	2.371×10^{-2}	2.767×10^{-2}
NN	1.551×10^{-2}	4.898×10^{-2}	3.497×10^{-2}	1.488×10^{-2}	1.558×10^{-2}
PBM	2.118×10^{-2}	5.171×10^{-2}	3.816×10^{-2}	1.518×10^{-2}	2.099×10^{-2}
MAE for the test dataset					
Case	μ_0	r_1	r_2	r_3	C
UDE-B	5.144×10^{-2}	5.366×10^{-2}	5.022×10^{-2}	3.277×10^{-2}	7.627×10^{-3}
UDE-G	2.381×10^{-2}	4.078×10^{-2}	3.640×10^{-2}	2.531×10^{-2}	2.202×10^{-2}
UDE-B-G	7.771×10^{-2}	5.555×10^{-2}	5.285×10^{-2}	3.419×10^{-2}	4.259×10^{-2}
NN	1.861×10^{-2}	2.828×10^{-2}	2.032×10^{-2}	1.457×10^{-2}	4.940×10^{-2}

is explained by its problem of overfitting, mimicking the experimental noise. Thus, MSE and MAE were not good metrics for this specific case. Comparing the UDE models, using two NNs leads to MSE and MAE values a bit higher than the UDE models using one network for the training. Even the UDE-B-G presented the highest values of MSE and MAE in most cases, these values were very close to the PBM. This fact confirms the efficiency of the approach, which shows the adequacy of the UDE models since the PBM was shown as a digital twin and experimentally validated (Moraes et al., 2023c). It is worth noting that to obtain the complete phenomenological model, great effort involving the design of experiments and the proper structure of the expressions for the supersaturation level and the kinetic rates of nucleation and growth are critical steps for the proper development of the PBM. Specifically for G, K_2SO_4 has a size-dependent growth rate ($G = G(L)$). The UDE models show a clear advantage in this context since they do not require the steps mentioned.

For the test datasets, the NN presented again the best results, with the lowest values of MSE and MAE, according to Table 1. The results obtained for the UDE applying only one NN were once again a bit better than those achieved by the UDE using two NNs. However, all approaches achieved similar MSE and MAE values, effectively predicting a new dataset. These results showed the efficiency of the proposed UDE strategies for being used in crystallization processes for modeling nucleation and crystal growth.

For developing the UDE for the dissolution model presented in Eq. (6), an NN with a single hidden layer, 20 neurons, and hyperbolic tangent as the activation function was used. The selection of these hyperparameters was done based on the previous case. Moreover, an NN to predict the state variables under the undersaturated condition was trained using the same hyperparameters as the previous case.

Fig. 4 presents a comparison between the PBM, UDE, and NN to predict one experimental dissolution batch. The NN presented a noisy answer, trying again to model the experimental error. The small dataset used to train the NN caused this negative point of the NN. However, all approaches presented a similar behavior and were effective in predicting r_1, r_2, r_3, and the concentration. This case also confirms the advantage of the UDE, which needs a small dataset to develop an efficient model.

Table 2
MSE and MAE values for the normalized training and test datasets of the models for dissolution, considering machine learning models trained with five experimental batches.

MSE for the training dataset				
Case	r_1	r_2	r_3	C
UDE-D	7.897×10^{-4}	8.393×10^{-4}	1.134×10^{-3}	3.094×10^{-4}
NN	4.312×10^{-4}	3.992×10^{-4}	4.819×10^{-4}	1.774×10^{-4}
PBM	7.849×10^{-4}	1.057×10^{-3}	1.294×10^{-3}	1.660×10^{-4}
MSE for the test dataset				
Case	r_1	r_2	r_3	C
UDE-D	1.956×10^{-3}	2.163×10^{-3}	1.876×10^{-3}	9.004×10^{-4}
NN	1.017×10^{-1}	9.086×10^{-2}	1.156×10^{-1}	1.810×10^{-2}
MAE for the training dataset				
Case	r_1	r_2	r_3	C
UDE-D	2.250×10^{-2}	2.287×10^{-2}	2.667×10^{-2}	1.363×10^{-2}
NN	1.518×10^{-2}	1.475×10^{-2}	1.686×10^{-2}	9.506×10^{-3}
PBM	2.271×10^{-2}	2.361×10^{-2}	2.851×10^{-2}	9.896×10^{-3}
MAE for the test dataset				
Case	r_1	r_2	r_3	C
UDE-D	2.411×10^{-2}	2.488×10^{-2}	2.265×10^{-2}	1.566×10^{-2}
NN	2.545×10^{-1}	2.353×10^{-1}	2.788×10^{-1}	1.053×10^{-1}

Table 2 presents the MSE and the MAE values for the training and test datasets considering normalized values of r_1, r_2, r_3, and C. Considering the training dataset, the NN presents the lowest MSE and MAE values in most cases, explained by this attempt at modeling the noise. All approaches presented small and close values of MSE and MAE for the training set, demonstrating their efficiency in modeling the process. The MSE and MAE values of the UDE approach were very similar to the PBM for this scenario, proving the efficacy of this data-driven strategy.

For the test dataset, the UDE and NN approaches presented small values of MSE and MAE, showing a performance similar to the phenomenological model for predicting new data. However, the UDE presented smaller MSE and MAE values than the NN, showing the better performance of the hybrid strategy. The worse performance of the NN is explained by the small dataset used in the training. The UDE does not need a big dataset to develop an efficient model.

One of the benefits of the UDE is the ability of the finalized neural network to make standalone predictions for specific variables, in this case, the growth, dissolution, and nucleation rates. This capability is not just a peripheral outcome; it is an intrinsic feature of the UDE framework. In addition to forecasting the model's state variables, the UDE methodology offers this supplementary benefit. This outcome, an inherent byproduct of the UDE, holds significant value for the case under consideration, facilitating direct predictions of these variables. Therefore, the performance of these NNs was compared to the phenomenological model in Fig. 5 for the training and test datasets. The nucleation, growth, and dissolution rates shown in the figure are normalized. For the growth and nucleation rates, the UDE approach showed an efficient performance, being very similar to the phenomenological model. The NNs made predictions of the nucleation and growth rates very similar to the phenomenological model, which is confirmed by the dots being close to the solid black line in all cases.

Considering the dissolution rate, most experimental points are close to the solid black line in both cases. This behavior proves that the

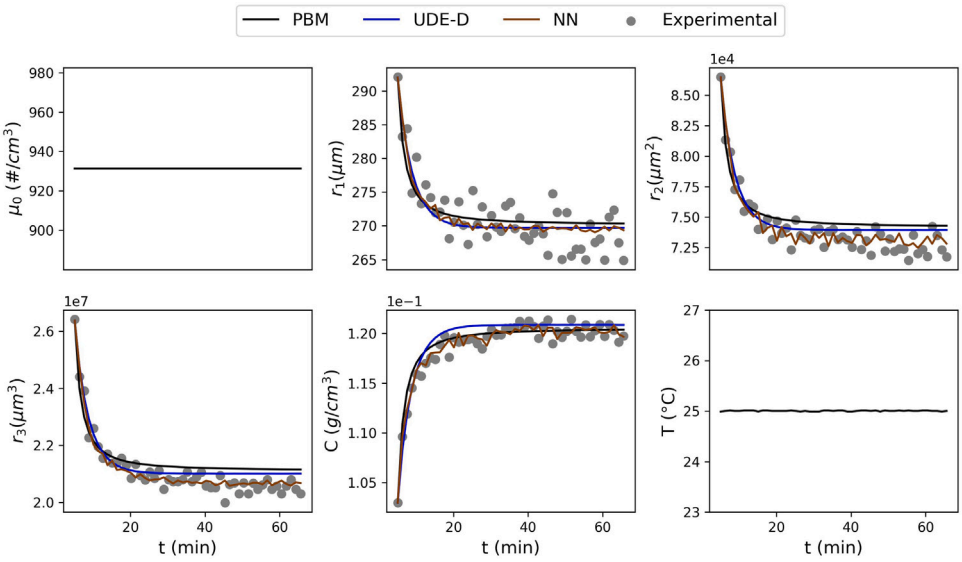

Fig. 4. Comparison between the NNs, the phenomenological model, and the UDE approach trained with all experiments to predict the state variables under the undersaturated condition, considering data for one of the batches.

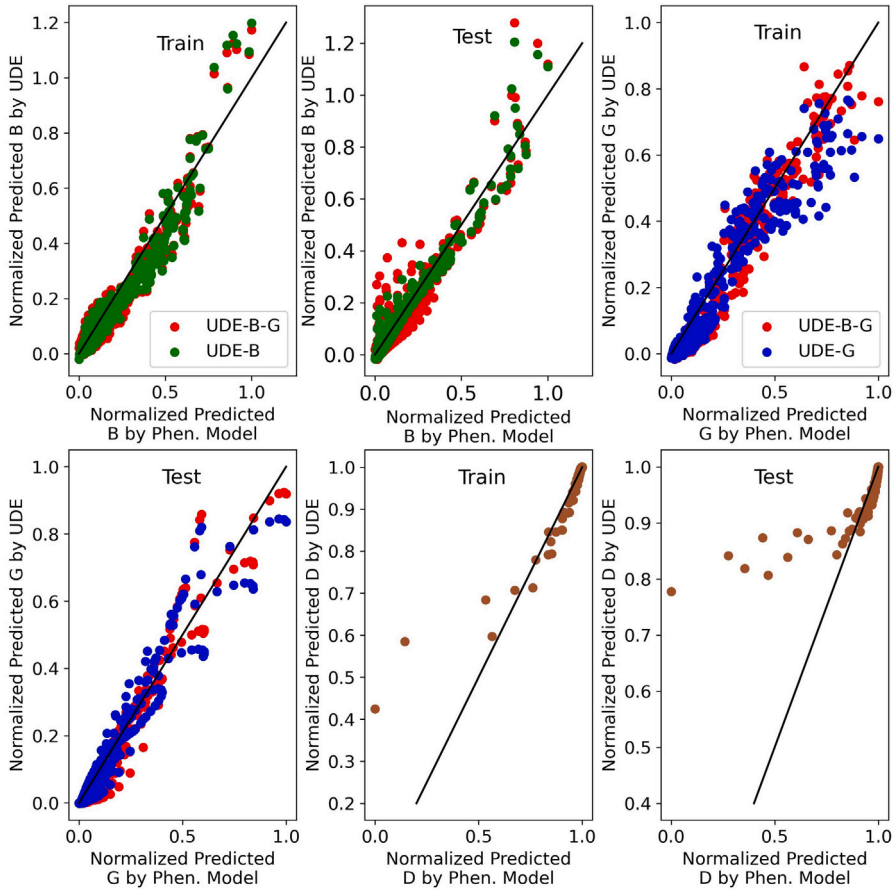

Fig. 5. Performance of the UDE models trained with ten experimental batches for predicting the nucleation, growth, and dissolution rates for the training and test datasets.

dissolution rate predicted by the UDE is very similar to the phenomenological model. For the undersaturated condition, most experimental points are at a plateau region due to the faster dissolution kinetics when compared to growth, reaching the solubility equilibrium at the end of the experiments.

3.2. Training with part of the experimental batches

In a second analysis, the machine learning models were trained and tested only with experimental data. Therefore, seven experimental batches were used for the training, and the other three were used to test

Table 3
MSE and MAE values for the normalized training and test datasets of the models for nucleation and crystal growth trained with seven experimental batches.

MSE for the training dataset					
Case	μ_0	r_1	r_2	r_3	C
UDE-B	1.031×10^{-3}	4.456×10^{-3}	3.082×10^{-3}	8.945×10^{-4}	5.311×10^{-4}
UDE-G	6.765×10^{-4}	5.692×10^{-3}	3.630×10^{-3}	5.924×10^{-4}	1.067×10^{-3}
UDE-B-G	1.965×10^{-3}	5.639×10^{-3}	3.791×10^{-3}	9.211×10^{-4}	1.211×10^{-3}
NN	6.197×10^{-4}	3.508×10^{-3}	1.826×10^{-3}	3.495×10^{-4}	4.060×10^{-4}
PBM	9.033×10^{-4}	3.805×10^{-3}	2.154×10^{-3}	3.588×10^{-4}	5.754×10^{-4}
MSE for the test dataset					
Case	μ_0	r_1	r_2	r_3	C
UDE-B	3.291×10^{-3}	8.270×10^{-3}	5.767×10^{-3}	1.304×10^{-3}	2.368×10^{-3}
UDE-G	3.179×10^{-3}	9.537×10^{-3}	8.089×10^{-3}	1.627×10^{-3}	2.368×10^{-3}
UDE-B-G	6.013×10^{-3}	9.005×10^{-3}	6.555×10^{-3}	1.513×10^{-3}	1.885×10^{-3}
NN	1.623×10^{-3}	7.624×10^{-3}	5.377×10^{-3}	9.546×10^{-4}	7.150×10^{-4}
PBM	3.427×10^{-3}	8.153×10^{-3}	5.663×10^{-3}	1.003×10^{-3}	1.302×10^{-3}
MAE for the training dataset					
Case	μ_0	r_1	r_2	r_3	C
UDE-B	2.079×10^{-2}	5.305×10^{-2}	4.143×10^{-2}	2.067×10^{-2}	1.868×10^{-2}
UDE-G	1.765×10^{-2}	5.875×10^{-2}	4.297×10^{-2}	1.709×10^{-2}	2.518×10^{-2}
UDE-B-G	2.960×10^{-2}	5.947×10^{-2}	4.712×10^{-2}	2.229×10^{-2}	2.799×10^{-2}
NN	1.783×10^{-2}	4.563×10^{-2}	3.260×10^{-2}	1.354×10^{-2}	1.567×10^{-2}
PBM	1.786×10^{-2}	4.825×10^{-2}	3.549×10^{-2}	1.330×10^{-2}	1.934×10^{-2}
MAE for the test dataset					
Case	μ_0	r_1	r_2	r_3	C
UDE-B	4.008×10^{-2}	6.801×10^{-2}	5.566×10^{-2}	2.521×10^{-2}	2.850×10^{-2}
UDE-G	4.017×10^{-2}	7.259×10^{-2}	6.472×10^{-2}	2.523×10^{-2}	3.832×10^{-2}
UDE-B-G	5.581×10^{-2}	6.980×10^{-2}	5.965×10^{-2}	2.727×10^{-2}	3.453×10^{-2}
NN	2.707×10^{-2}	6.647×10^{-2}	5.266×10^{-2}	2.099×10^{-2}	2.133×10^{-2}
PBM	4.151×10^{-2}	6.654×10^{-2}	5.284×10^{-2}	2.103×10^{-2}	3.001×10^{-2}

the models, considering the modeling of the supersaturated condition. The selection of the batches to compose the test and training datasets was made according to the work of Moraes et al. (2023c) for calibration and validation of the model. The same UDE and NN approaches were used, considering the models under the supersaturated condition. Also, the same hyperparameters from the previous case were used.

Table 3 presents the MSE and MAE values of the training and test datasets, considering the normalized values of the state variables for the PBM, NN, and UDE models under the supersaturated condition. Looking at these values, the NN presented the smallest values in all cases. However, all approaches presented similar values for all state variables. Also, even with the lowest values of the NN, all approaches presented similar values for the test dataset.

Fig. 6 shows a comparison between the PBM, NN, and UDE approaches for predicting one of the experimental batches from the test dataset. This experiment was also used for a test (called calibration) of the PBM in Moraes et al. (2023c). The NN still presents the disadvantage of a noisy prediction. On the other hand, the models based on UDE presented predictions close to the experimental data and very similar to the PBM. Fig. 6 also shows the improvement in μ_0 prediction by UDE models. The prediction of μ_0 by PBM shows a higher error, being the nucleation overestimated and with increasing prediction error throughout the experiment. Determination of phenomenological model parameters for nucleation (B) is challenging when compared to G and D rates, with the nucleation usually not well estimated using classical approaches in phenomenological models (Bosetti and Mazzotti, 2020; Erdemir et al., 2009; Gunawan et al., 2002). This fact is intrinsically related to the stochasticity of the nucleation phenomenon, as well as the experimental difficulty of detecting and determining the size of the nuclei or even when precisely they are formed.

Table 4
MSE and MAE values for the normalized training and test datasets of the models for dissolution, considering machine learning models trained with four experimental batches.

MSE for the training dataset				
Case	r_1	r_2	r_3	C
UDE-D	8.496×10^{-4}	1.221×10^{-3}	1.160×10^{-3}	3.665×10^{-4}
NN	4.879×10^{-4}	4.973×10^{-4}	5.247×10^{-4}	2.018×10^{-4}
PBM	7.149×10^{-4}	1.513×10^{-3}	1.302×10^{-3}	1.984×10^{-4}
MSE for the test dataset				
Case	r_1	r_2	r_3	C
UDE-D	1.094×10^{-2}	3.788×10^{-3}	9.803×10^{-3}	2.561×10^{-3}
NN	8.946×10^{-2}	8.447×10^{-1}	9.693×10^{-2}	5.451×10^{-2}
PBM	1.290×10^{-2}	3.582×10^{-3}	1.221×10^{-2}	1.625×10^{-3}
MAE for the training dataset				
Case	r_1	r_2	r_3	C
UDE-D	2.345×10^{-2}	2.852×10^{-2}	2.718×10^{-2}	1.546×10^{-2}
NN	1.675×10^{-2}	1.679×10^{-2}	1.809×10^{-2}	1.087×10^{-2}
PBM	2.069×10^{-2}	2.874×10^{-2}	2.716×10^{-2}	1.082×10^{-2}
MAE for the test dataset				
Case	r_1	r_2	r_3	C
UDE-D	8.909×10^{-2}	5.020×10^{-2}	8.326×10^{-2}	4.110×10^{-2}
NN	2.931×10^{-1}	9.086×10^{-1}	3.056×10^{-1}	2.251×10^{-1}
PBM	1.072×10^{-1}	5.022×10^{-2}	1.053×10^{-1}	3.304×10^{-2}

The NN and the UDE were also trained to predict the ratios and the concentration under the undersaturated condition. In this case, four batches were used for the training and one was used for the test. The batch chosen for the test dataset was randomly selected. The hyperparameters used in the machine learning models were the same as the previous models for the undersaturated condition. The MSE and the MAE values are presented in Table 4 for the training and test datasets, considering normalized values of the r_1, r_2, r_3, and the concentration. These values demonstrated a similar performance for all approaches, considering the training dataset. For the test, the NN presented the highest MSE and MAE values in all cases. This confirms the necessity of a huge dataset for training NNs. On the other hand, the UDE presented the smallest values in most cases, which shows its advantage of not needing a huge dataset to be efficiently developed.

Fig. 7 presents a comparison between the PBM, NN, and UDE for predicting the batch used in the test dataset. The NN predictions were very different from the experimental points, reaching incorrectly the values of the state variables in the equilibrium. However, the UDE performed very similarly to the PBM and the experimental behavior. This confirms once again the fact that the UDE needs much less experimental data for training an efficient model than the NN.

The intrinsic feature of the UDE framework allowed once again the development of neural networks to predict nucleation, dissolution and growth rates. The performance of these NNs was compared to the rate equations developed by Moraes et al. (2023c,b). This comparison can be seen in Fig. 8 for the training and test samples. For the nucleation and growth rates, the dots were close to the solid black line, demonstrating a close performance between the PBM and the UDE models. Even with a smaller training dataset, the performance of UDE models for modeling the crystal growth and nucleation is still very similar to the previous case and to the PBM. The performance of UDE-B-G is comparable to UDE-B and UDE-G, evidencing the ability of this model to predict both kinetics together. These conclusions present once again the advantage of the UDE of being able to predict the state variables and the rates efficiently. For the undersaturated condition, most dots are close to the black solid line, which confirms a similar behavior from both approaches. Therefore, the UDE could also efficiently predict the dissolution rate even with a small training dataset.

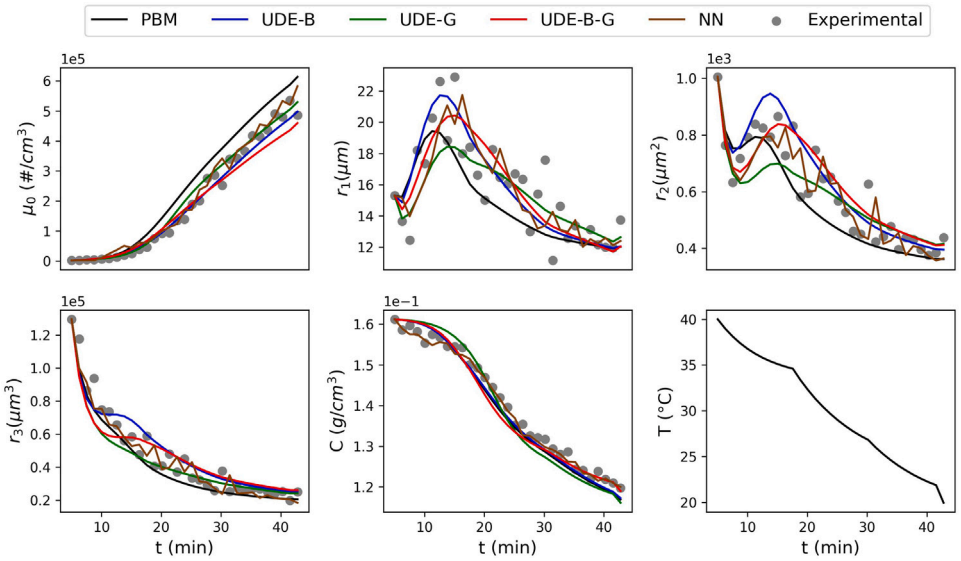

Fig. 6. Comparison between the NNs, the phenomenological model, and the three UDE structures trained with seven experimental batches to predict the state variables under the supersaturated condition. The models were compared to one experimental batch used for the test.

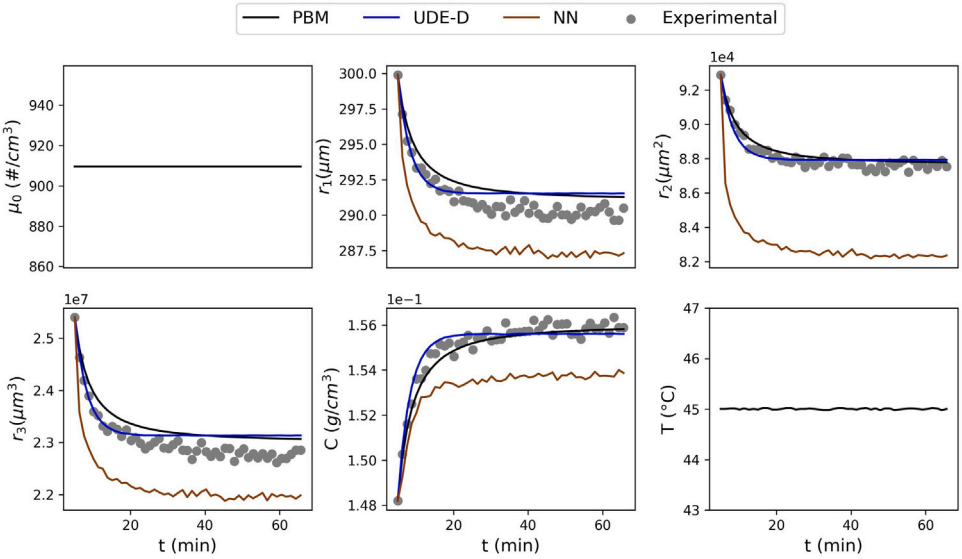

Fig. 7. Comparison between the NNs, the phenomenological model, and the three UDE structures trained with four experimental batches to predict the state variables under the undersaturated condition. The models were compared to one experimental batch used for the test.

Even with the advantages of the UDE models, previous knowledge about the phenomenological laws of the process is necessary for modeling it with the use of this hybrid approach. Considering modeling the potassium sulfate crystallization process, prior knowledge of the volume shape factor and the size-dependent growth constant was needed to train the UDEs.

4. Conclusions

This work aimed to synergize phenomenological insights with the power of machine learning techniques, such as the Universal Differential Equation (UDE), in the context of crystallization processes. In this study, we introduced an innovative approach by harnessing the capabilities of the Universal Differential Equation, a machine learning tool, to model the batch cooling crystallization of potassium sulfate in the presence of nucleation, growth, and dissolution phenomena. Notably, the distinctive feature of this approach lies in its ability to circumvent the necessity of defining a supersaturation term, as well as identifying representations and dependencies of the kinetic rates associated with the phenomena.

Comparing the performance of UDE models to traditional methodologies, such as the Population Balance Model (PBM) and individual NNs, demonstrated their efficacy. This held even when these models were trained on a narrower dataset. The developed UDE models presented prediction errors comparable to the PBM previously validated for the process. The validation process, which involved assessing the models against experimental data not utilized during training, underscored their robust generalization capabilities and adaptability.

The UDE models proved more suitable to represent the nucleation kinetics for validation experiments. The prediction errors associated with the determination of nucleation using purely phenomenological models are due to its strong dependence on disturbances and the intrinsic stochasticity of this mechanism. Thus, the UDE models prove to be a good hybrid model by showing an alternative to the classically used kinetic rates but guaranteeing the resolution of the population and

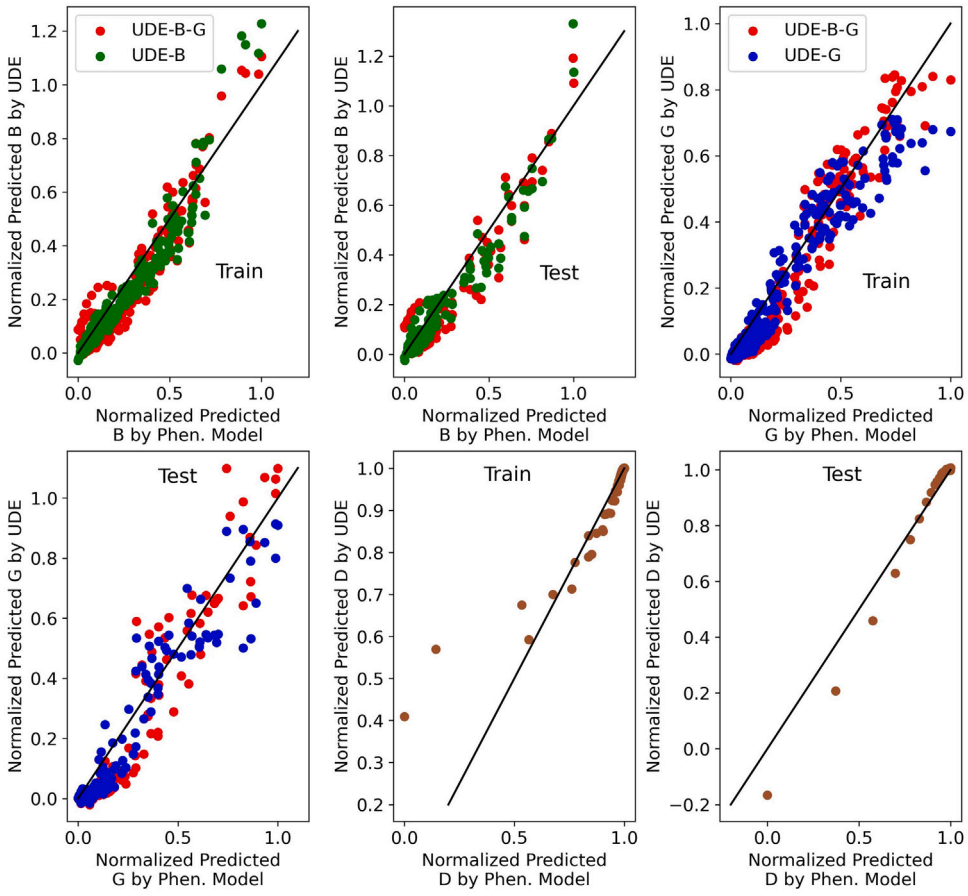

Fig. 8. Performance of the UDE models for predicting the nucleation, growth, and dissolution rates for the training and test datasets, considering machine learning models trained with four experimental batches.

mass balance equations. This predictive capability establishes UDE's potential as a robust tool for crystallization processes.

Although the immediate focus of this study was the crystallization of potassium sulfate, the novel application of the UDE approach hints at its promising prospect for extension to other crystallization systems. This approach opens up opportunities for advances in optimizing industrial processes and precision engineering materials with custom attributes, presenting a significant advance in process optimization and material design.

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

The authors acknowledge funding from the Brazilian Coordination for the Improvement of Higher Education Personnel (CAPES, Finance Code 001), the National Council for Scientific and Technological Development, Brazil (CNPq, Grants 303587/2020-2, 311153/2021-6, and 313337/2021-7), and the Research Support Foundation of the State of Rio de Janeiro, Brazil (FAPERJ, Grants E 26/111.671/2013 and E-26/201.148/2022).

References

Ahn, B., Bosetti, L., Mazzotti, M., 2022. Secondary nucleation by interparticle energies. II. Kinetics. Cryst. Growth Des. 22 (1), 74–86. http://dx.doi.org/10.1021/acs.cgd.1c00928.

Bangi, M.S.F., Kao, K., Kwon, J.S.I., 2022. Physics-informed neural networks for hybrid modeling of lab-scale batch fermentation for β-carotene production using saccharomyces cerevisiae. Chem. Eng. Res. Des. 179, 415–423. http://dx.doi.org/10.1016/j.cherd.2022.01.041.

Boareto, A.J.M., Souza, Jr., M.B., Valero, F., Valdman, B., 2007. A hybrid neural model (HNM) for the on-line monitoring of lipase production by Candida rugosa. J. Chem. Technol. Biotechnol. 82 (3), 319–327. http://dx.doi.org/10.1002/jctb.1678.

Bosetti, L., Mazzotti, M., 2020. Population balance modeling of growth and secondary nucleation by attrition and ripening. Cryst. Growth Des. 20 (1), 307–319. http://dx.doi.org/10.1021/acs.cgd.9b01240.

Braake, H., van Can, H., Verbruggen, H., 1998. Semi-mechanistic modeling of chemical processes with neural networks. Eng. Appl. Artif. Intell. 11 (4), 507–515. http://dx.doi.org/10.1016/S0952-1976(98)00011-6.

Braatz, R.D., Hasebe, S., 2002. Particle size and shape control in crystallization processes. In: AIChE Symposium Series. American Institute of Chemical Engineers, New York, pp. 307–327, URL: https://api.semanticscholar.org/CorpusID:14210359.

Budz, J., Jones, A.G., Mullin, J.W., 1987. On the shape-size dependence of potassium sulfate crystals. Ind. Eng. Chem. Res. 26 (4), 820–824. http://dx.doi.org/10.1021/ie00064a034.

Chollet, F., et al., 2015. Keras. https://keras.io.

Erdemir, D., Lee, A.Y., Myerson, A.S., 2009. Nucleation of crystals from solution: Classical and two-step models. Acc. Chem. Res. 42 (5), 621–629. http://dx.doi.org/10.1021/ar800217x.

Feyo de Azevedo, S., Dahm, B., Oliveira, F., 1997. Hybrid modelling of biochemical processes: A comparison with the conventional approach. Comput. Chem. Eng. 21, S751–S756. http://dx.doi.org/10.1016/S0098-1354(97)87593-X.

Galvanauskas, V., Georgieva, P., de Azevedo, S., 2006. Dynamic optimisation of industrial sugar crystallization process based on a hybrid (mechanistic+ANN) model. In: The 2006 IEEE International Joint Conference on Neural Network Proceedings. pp. 2728–2735. http://dx.doi.org/10.1109/IJCNN.2006.247177.

Georgieva, P., Meireles, M., Feyo de Azevedo, S., 2003. Knowledge-based hybrid modelling of a batch crystallisation when accounting for nucleation, growth and agglomeration phenomena. Chem. Eng. Sci. 58 (16), 3699–3713. http://dx.doi.org/10.1016/S0009-2509(03)00260-4.

Griffin, D.J., Grover, M.A., Kawajiri, Y., Rousseau, R.W., 2016. Data-driven modeling and dynamic programming applied to batch cooling crystallization. Ind. Eng. Chem. Res. 55 (5), 1361–1372. http://dx.doi.org/10.1021/acs.iecr.5b03635.

Gunawan, R., Ma, D.L., Fujiwara, M., Braatz, R.D., 2002. Identification of kinetic parameters in multidimensional crystallization pocesses. Internat. J. Modern Phys. B 16 (01n02), 367–374. http://dx.doi.org/10.1142/S0217979202009883.

Halfwerk, R., Yntema, D., Van Spronsen, J., Keesman, K., Van der Padt, A., 2023. Crystallization kinetics of lactose recovered at sub-zero temperatures: A population balance model combining mutarotation, nucleation and crystal growth. J. Food Eng. 345, 111412. http://dx.doi.org/10.1016/j.jfoodeng.2023.111412.

Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. Neural Netw. http://dx.doi.org/10.1016/0893-6080(89)90020-8.

Jones, A., Budz, J., Mullin, J., 1986. Crystallization kinetics of potassium sulfate in an MSMPR agitated vessel. AIChE J. 32 (12), 2002–2009. http://dx.doi.org/10.1002/aic.690321210.

Kim, Y., Kawajiri, Y., Rousseau, R.W., Grover, M.A., 2023. Modeling of nucleation, growth, and dissolution of paracetamol in ethanol solution for unseeded batch cooling crystallization with temperature-cycling strategy. Ind. Eng. Chem. Res. 62 (6), 2866–2881. http://dx.doi.org/10.1021/acs.iecr.2c03465.

Lauret, P., Boyer, H., Gatina, J.C., 2001. Hybrid modelling of the sucrose crystal growth rate. Int. J. Modelling Simul. 21 (1), 23–29. http://dx.doi.org/10.1080/02286203.2001.11442183.

Lewis, A., Seckler, M., Kramer, H., Van Rosmalen, G., 2015. Industrial Crystallization: Fundamentals and Applications. Cambridge University Press, http://dx.doi.org/10.1017/CBO9781107280427.

Lima, F.A.R.D.L., 2023. Universal differential equations (UDE) in Python. URL: https://github.com/fearrais96/Modelling-Crystallization-Process-with-Universal-Differential-Equations-UDE-.git.

Lima, F.A.R., de Miranda, G.F., Moraes, M.G., Capron, B.D., Souza, M.B., 2022a. A recurrent neural networks-based approach for modeling and control of a crystallization process. In: Montastruc, L., Negny, S. (Eds.), 32nd European Symposium on Computer Aided Process Engineering. In: Computer Aided Chemical Engineering, vol. 51, Elsevier, pp. 1423–1428. http://dx.doi.org/10.1016/B978-0-323-95879-0.50238-1.

Lima, F.A.R., Moraes, M.G., Barreto, A.G., Secchi, A.R., Souza, Jr., M.B., 2023. An NN-NMPC for controlling a crystallization process in the saturated and undersaturated zones. In: Kokossis, A.C., Georgiadis, M.C., Pistikopoulos, E. (Eds.), 33rd European Symposium on Computer Aided Process Engineering. In: Computer Aided Chemical Engineering, vol. 52, Elsevier, pp. 1835–1840. http://dx.doi.org/10.1016/B978-0-443-15274-0.50291-2.

Lima, F.A.R.D., Moraes, M.G.F., Secchi, A.R., Souza, Jr., M.B., 2022b. Development of a recurrent neural networks-based NMPC for controlling the concentration of a crystallization process. Digit. Chem. Eng. 5, 100052. http://dx.doi.org/10.1016/j.dche.2022.100052.

Liu, Y.C., Acevedo, D., Yang, X., Naimi, S., Wu, W.L., Pavurala, N., Nagy, Z.K., O'Connor, T.F., 2020. Population balance model development verification and validation of cooling crystallization of carbamazepine. Cryst. Growth Des. 20 (8), 5235–5250. http://dx.doi.org/10.1021/acs.cgd.0c00434.

Ma, S., Li, C., Gao, J., Yang, H., Tang, W., Gong, J., Zhou, F., Gao, Z., 2020. Artificial neural network prediction of metastable zone widths in reactive crystallization of lithium carbonate. Ind. Eng. Chem. Res. 59 (16), 7765–7776. http://dx.doi.org/10.1021/acs.iecr.9b06074.

McBride, K., Sundmacher, K., 2019. Overview of surrogate modeling in chemical process engineering. Chem. Ing. Tech. 91 (3), 228–239. http://dx.doi.org/10.1002/cite.201800091.

McDonald, M.A., Bommarius, A.S., Rousseau, R.W., Grover, M.A., 2019. Continuous reactive crystallization of β-lactam antibiotics catalyzed by penicillin G acylase. Part I: Model development. Comput. Chem. Eng. 123, 331–343. http://dx.doi.org/10.1016/j.compchemeng.2018.12.029.

Meng, Y., Yu, S., Zhang, J., Qin, J., Dong, Z., Lu, G., Pang, H., 2019. Hybrid modeling based on mechanistic and data-driven approaches for cane sugar crystallization. J. Food Eng. 257, 44–55. http://dx.doi.org/10.1016/j.jfoodeng.2019.03.026.

Moraes, M.G.F., Barreto, Jr., A.G., Secchi, A.R., Souza, Jr., M.B., Lage, P.L.d.C., Myerson, A.S., 2023a. Polymorphism of praziquantel: Role of cooling crystallization in access to solid forms and discovery of new polymorphs. Cryst. Growth Des. 23 (2), 1247–1258. http://dx.doi.org/10.1021/acs.cgd.2c01381.

Moraes, M.G., Grover, M.A., Souza, Jr., M.B., Lage, P.L., Secchi, A.R., 2021. Optimal control of crystal size and shape in batch crystallization using a bivariate population balance modeling. IFAC-PapersOnLine 54 (3), 653–660. http://dx.doi.org/10.1016/j.ifacol.2021.08.316.

Moraes, M., Lima, F., Barreto, Jr., A., Lage, P., Souza, Jr., M., Secchi, A.R., 2023b. Determination and model evaluation of pure dissolution kinetics of potassium sulfate using in-situ dynamic image analysis. In: WCCE11 - 11th World Congress of Chemical Engineering.

Moraes, M.G.F., Lima, F.A.R.D., Lage, P.L.d.C., Souza, Jr., M.B., Barreto, Jr., A.G., Secchi, A.R., 2023c. Modeling and predictive control of cooling crystallization of potassium sulfate by dynamic image analysis: Exploring phenomenological and machine learning approaches. Ind. Eng. Chem. Res. http://dx.doi.org/10.1021/acs.iecr.3c00739.

Nagy, Z.K., Braatz, R.D., 2012. Advances and new directions in crystallization control. Annu. Rev. Chem. Biomol. Eng. 3, 55–75. http://dx.doi.org/10.1146/annurev-chembioeng-062011-081043.

Nagy, Z.K., Fujiwara, M., Woo, X.Y., Braatz, R.D., 2008. Determination of the kinetic parameters for the crystallization of paracetamol from water using metastable zone width experiments. Ind. Eng. Chem. Res. 47 (4), 1245–1252. http://dx.doi.org/10.1021/ie060637c.

Narayanan, H., Dingfelder, F., Condado Morales, I., Patel, B., Heding, K.E., Bjelke, J.R., Egebjerg, T., Butté, A., Sokolov, M., Lorenzen, N., Arosio, P., 2021. Design of biopharmaceutical formulations accelerated by machine learning. Mol. Pharm. 18 (10), 3843–3853. http://dx.doi.org/10.1021/acs.molpharmaceut.1c00469.

Nielsen, R.F., Nazemzadeh, N., Sillesen, L.W., Andersson, M.P., Gernaey, K.V., Mansouri, S.S., 2020. Hybrid machine learning assisted modelling framework for particle processes. Comput. Chem. Eng. 140, 106916. http://dx.doi.org/10.1016/j.compchemeng.2020.106916.

Nogueira, I.B.R., Santana, V.V., Ribeiro, A.M., Rodrigues, A.E., 2022. Using scientific machine learning to develop universal differential equation for multicomponent adsorption separation systems. Can. J. Chem. Eng. 100 (9), 2279–2290. http://dx.doi.org/10.1002/cjce.24495.

Ó'Ciardhá, C.T., Hutton, K.W., Mitchell, N.A., Frawley, P.J., 2012. Simultaneous parameter estimation and optimization of a seeded antisolvent crystallization. Cryst. Growth Des. 12 (11), 5247–5261. http://dx.doi.org/10.1021/cg3006822.

Oliveira, C., Georgieva, P., Rocha, F., Feyo de Azevedo, S., 2009. Artificial neural networks for modeling in reaction process systems. Neural Comput. Appl. 18, 15–24. http://dx.doi.org/10.1007/s00521-008-0200-8.

Psichogios, D.C., Ungar, L.H., 1992. A hybrid neural network-first principles approach to process modeling. AIChE J. 38 (10), 1499–1511. http://dx.doi.org/10.1002/aic.690381003.

Quilló, G.L., Bhonsale, S., Gielen, B., Van Impe, J.F., Collas, A., Xiouras, C., 2021. Crystal growth kinetics of an industrial active pharmaceutical ingredient: Implications of different representations of supersaturation and simultaneous growth mechanisms. Cryst. Growth Des. 21 (9), 5403–5420. http://dx.doi.org/10.1021/acs.cgd.1c00677.

Rackauckas, C., Ma, Y., Martensen, J., Warner, C., Zubov, K., Supekar, R., Skinner, D., Ramadhan, A., Edelman, A., 2020. Universal differential equations for scientific machine learning. pp. 1–55. http://dx.doi.org/10.48550/arXiv.2001.04385.

Randolph, A.D., Cise, M.D., 1972. Nucleation kinetics of the potassium sulfate-water system. AIChE J. 18 (4), 798–807. http://dx.doi.org/10.1002/aic.690180423.

Randolph, A.D., Rajagopal, K., 1970. Direct measurement of crystal nucleation and growth rate kinetics in backmixed crystal slurry. Study of the K2SO4system. Ind. Eng. Chem. Fundam. 9 (1), 165–171. http://dx.doi.org/10.1021/i160033a027.

Schweidtmann, A.M., Esche, E., Fischer, A., Kloft, M., Repke, J.U., Sager, S., Mitsos, A., 2021. Machine learning in chemical engineering: A perspective. Chem. Ing. Tech. 93 (12), 2029–2039. http://dx.doi.org/10.1002/cite.202100083.

Su, P.C., Ward, J.D., 2019. Modeling of membrane-assisted seeded batch crystallization. Ind. Eng. Chem. Res. 58 (36), 16787–16797. http://dx.doi.org/10.1021/acs.iecr.9b02935.

Trampuž, M., Teslić, D., Likozar, B., 2019. Crystallization of fesoterodine fumarate active pharmaceutical ingredient: Modelling of thermodynamic equilibrium, nucleation, growth, agglomeration and dissolution kinetics and temperature cycling. Chem. Eng. Sci. 201, 97–111. http://dx.doi.org/10.1016/j.ces.2019.02.019.

Vasanth Kumar, K., Martins, P., Rocha, F., 2008. Modelling of the batch sucrose crystallization kinetics using artificial neural networks: Comparison with conventional regression analysis. Ind. Eng. Chem. Res. 47 (14), 4917–4923. http://dx.doi.org/10.1021/ie701706v.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., SciPy 1.0 Contributors, 2020. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods 17, 261–272. http://dx.doi.org/10.1038/s41592-019-0686-2.

Worlitschek, J., Mazzotti, M., 2004. Model-based optimization of particle size distribution in batch-cooling crystallization of paracetamol. Cryst. Growth Des. 4 (5), 891–903. http://dx.doi.org/10.1021/cg034179b.

Wu, G., Yion, W.T.G., Dang, K.L.N.Q., Wu, Z., 2023. Physics-informed machine learning for MPC: Application to a batch crystallization process. Chem. Eng. Res. Des. 192, 556–569. http://dx.doi.org/10.1016/j.cherd.2023.02.048.

Xiouras, C., Cameli, F., Quilló, G.L., Kavousanakis, M.E., Vlachos, D.G., Stefanidis, G.D., 2022. Applications of artificial intelligence and machine learning algorithms to crystallization. Chem. Rev. 122 (15), 13006–13042. http://dx.doi.org/10.1021/acs.chemrev.2c00141.

Yang, M., Wei, H., 2006. Application of a neural network for the prediction of crystallization kinetics. Ind. Eng. Chem. Res. 45 (1), 70–75. http://dx.doi.org/10.1021/ie0487944.

Zhang, J., Meng, Y., Wu, J., Qin, J., Hui wang, Yao, T., Yu, S., 2020. Monitoring sugar crystallization with deep neural networks. J. Food Eng. 280, 109965. http://dx.doi.org/10.1016/j.jfoodeng.2020.109965.

Zheng, Y., Wang, X., Wu, Z., 2022a. Machine learning modeling and predictive control of the batch crystallization process. Ind. Eng. Chem. Res. 61 (16), 5578–5592. http://dx.doi.org/10.1021/acs.iecr.2c00026.

Zheng, Y., Zhao, T., Wang, X., Wu, Z., 2022b. Online learning-based predictive control of crystallization processes under batch-to-batch parametric drift. AIChE J. 68 (11), e17815. http://dx.doi.org/10.1002/aic.17815.