



A Physics-Informed Neural Network (PINN) framework for generic bioreactor modelling

Monesh kumar Thirugnanasambandam ^{a,b} , José Pinto ^{a,b,*} , Ekaterina Moskovkina ^{a,b}, Rafael S. Costa ^{a,b} , Rui Oliveira ^{a,b,*}

^a Associate Laboratory i4HB - Institute for Health and Bioeconomy, NOVA School of Science and Technology, Universidade NOVA de Lisboa, Caparica 2829-516, Portugal

^b UCBIO - Applied Molecular Biosciences Unit, Department of Chemistry, NOVA School of Science and Technology, Universidade NOVA de Lisboa, Caparica 2829-516, Portugal

ARTICLE INFO

Keywords:

Bioprocess modelling
Physics-informed neural networks
Hybrid semiparametric modelling
Deep artificial neural networks
ADAM algorithm
Fed-batch bioreactor

ABSTRACT

Many previous studies have explored hybrid semiparametric models merging Artificial Neural Networks (ANNs) with mechanistic models for bioprocess applications. More recently, Physics-Informed Neural Networks (PINNs) have emerged as promising alternatives. Both approaches seek to incorporate prior knowledge in ANN models, thereby decreasing data dependency whilst improving model transparency and generalization capacity. In the case of hybrid semiparametric modelling, the mechanistic equations are hard coded directly into the model structure in interaction with the ANN. In the case of PINNs, the same mechanistic equations must be “learned” by the ANN structure during the training. This study evaluates a dual-ANN PINN structure for generic bioreactor problems that decouples state and reaction kinetics parameterization. Furthermore, the dual-ANN PINN is benchmarked against the general hybrid semiparametric bioreactor model under comparable prior knowledge scenarios across 2 case studies. Our findings show that the dual-ANN PINN can level the prediction accuracy of hybrid semiparametric models for simple problems. However, its performance degrades significantly when applied to extended temporal extrapolation or to complex problems involving high-dimensional process states subject to time-varying control inputs. The latter is primarily due to the more complex multi-objective training of the dual-ANN PINN structure and to physics-based extrapolation errors beyond the training domain.

1. Introduction

Due to the intricate mechanistic complexity of cell culture systems, developing predictive models for optimization and control at an acceptable cost remains a challenge in the bioprocessing industries (Hong and Braatz, 2021). Traditional modelling approaches involve mathematical equations derived from first principles, including mass and energy balances, reaction kinetics, and transport phenomena. While these models can elucidate numerous aspects of process behaviour, their development is time-consuming and requires extensive domain knowledge. Furthermore, these models are inherently limited in their capacity to represent the global characteristics of most biological systems in specific application scenarios, and the efficiency of their solution is significantly lower than theoretically possible (Glassey et al., 2011; Van

Impe et al., 2013).

In recent years, data-driven modelling has received significant attention leveraging on technical progress in data acquisition and process analytical technology. Particularly deep artificial neural networks (deep ANNs) are increasingly considered for bioprocess modelling due to their ability to infer intricate patterns from data with minimal domain knowledge requirements (Hellekes et al., 2023). However, their dependency on high-quantity and high-quality data, which are often scarce, and their limited generalization capability constitute major drawbacks (Kipf et al., 2018). To mitigate these limitations, hybrid modelling approaches have emerged that combine ANNs and mechanistic modelling in a common workflow. The most frequently reported approach is hybrid semiparametric modelling, wherein ANNs and mechanistic equations are interlinked in a common model structure,

Abbreviations: FFNN-S, State variables neural network; FFNN-R, Reaction kinetics neural network; Nw, Number of network weight; K, Number of collocation points; SILU, Sigmoid weighted linear unit; tanh, hyperbolic tangent; ADAM, Adaptive Moment Estimation; WMSE, Weighted mean squared error.

* Corresponding authors at: Department of Chemistry, NOVA School of Science and Technology, Universidade NOVA de Lisboa, Caparica, 2829-516, Portugal

E-mail addresses: jmm.pinto@campus.fct.unl.pt (J. Pinto), rmo@fct.unl.pt (R. Oliveira).

<https://doi.org/10.1016/j.compchemeng.2025.109354>

Received 10 May 2025; Received in revised form 23 July 2025; Accepted 14 August 2025

Available online 17 August 2025

0098-1354/© 2025 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

thereby enhancing the model's accuracy and transparency (Psichogios and Ungar 1992; Thompson and Kramer, 1994; Schubert et al., 1994; Oliveira, 2004; Teixeira et al., 2007; Kadlec et al., 2009; Von Stosch et al., 2014b). Numerous application examples have demonstrated the effectiveness of hybrid semiparametric modelling for bioreactor systems (e.g., Pinto et al., 2022; Narayanan et al., 2023; Bayer et al., 2023; Ramos et al., 2024). A common path involves using intensified Statistical Design of Experiments (iDoE) in combination with hybrid modeling to dynamically infer the impact of critical process parameters (CPP) on critical quality attributes (CQA) (Von Stosch et al., 2014a). These methods not only help optimize process conditions but also reduce experimental runs, thus saving both time and resources. As a result, process engineers can achieve more reliable, scalable, and precise outcomes thereby enhancing overall productivity in industrial applications (Smiatek et al., 2020).

A recent addition to bioprocess modelling is Physics-Informed Neural Networks (PINNs). PINNs incorporate a system's known governing physical law, usually described by partial differential equations (PDEs) or ordinary differential equations (ODEs), into the training process of a deep neural network (DNN) (Raissi et al., 2019). The training minimizes two loss functions simultaneously, a data loss function based on process observations, and a physics loss function derived from physics equations. Unlike conventional DNNs, PINNs can handle complex problems, where precise data collection is difficult due to real-world measurement limitations. PINNs and hybrid semiparametric models are thus related concepts with similar goals, i.e. the integration of prior process knowledge in ANN models thereby reducing data dependency. Whereas in semiparametric modelling the physics equations are hard coded directly in the model structure, in the case of PINNs the same physics equations are embedded in a physics loss function that is minimized during the training. The general principle of a PINN was already postulated in the early study by Thompson and Kramer (1994) who categorized hybrid modelling as design methods, where prior knowledge is hard coded in the model structure, and training methods where, prior knowledge is embedded in the training loss function. Since the pioneering study by Raissi et al. (2019), PINNs are taking the first steps in the bioprocess modelling field. Due to their similarities, hybrid semiparametric modelling and PINNs are sometimes interchangeably referred to in the literature (e.g., Bangi et al., 2022; Cui et al., 2024). Adebar et al. (2024) recently proposed a PINN approach for mammalian cultivations using truncated Taylor series expansions to approximate growth kinetics. Simplifications such as pseudo-first-order kinetics and decoupled individual process outcomes were introduced. Despite the simplifications, they successfully modelled with reasonable accuracy the dynamics of key variables such as viable cell density, glucose, lactate, and product. In a different study, Li et al. (2024) developed a complex PINN framework using the multi-stage Koopman method for microbial growth modelling. The Koopman method served to map the process dynamics into a high-dimensional linear space and to model each growth stage separately in the linear space. Yang et al. (2024) compared parallel hybrid modelling, serial hybrid modelling, and PINNs in a pilot fed-batch CHO culture. They applied the multiple-shooting method to divide the culture into 24-hour intervals with constant feed. Their results suggested that PINNs have superior predictive power than the widely adopted semiparametric approaches. Moayedi et al. (2024) compared a Recurrent Neural Network (RNN) with a recurrent PINN in a simulation case study with 2 continuous reactors connected in series. They used the Euler method to approximate the derivatives of state variables to evaluate the physics loss function. They concluded that the PINN outperformed the conventional RNN. Jul-Rasmussen et al. (2025) recently presented a comparative study of hybrid semiparametric modelling and physics-informed recurrent neural networks (PIRNNs) for a pilot-scale bubble column aeration process. The study compared each approach in terms of training ease, adherence to governing equations, prediction accuracy with less frequent measurements, and performance with limited data. Their study found that hybrid semiparametric modelling

outperformed the recurrent PINN approach for the pilot-scale bubble column aeration process.

PINNs have gained significant traction in scientific computing, but their advantages and development opportunities for bioprocess applications need more evidence. Typical problems arising with PINNs are the choice of the ANN structure, handling time-varying control inputs, enforcing initial condition (IC), and convergence of the multi-objective loss function, which combines physics, data, and IC constraints. One critical issue in PINN training is the trade-off between data loss and physics loss, which directly influences model convergence. In this study, we evaluate a dual-ANN PINN structure for bioreactor problems that decouples the parametrization of dynamic state variables and static reaction kinetics. The first ANN parameterizes dynamic state variables as function of time, initial conditions and control inputs. The second ANN parametrizes static reaction kinetics variables as function of state variables and control inputs. This dual-ANN PINN structure is generally applicable to stirred-tank bioreactor problems under time-varying and batch-varying control inputs. This structure is evaluated across 2 case studies and compared with the general bioreactor hybrid model (Oliveira, 2004; Pinto et al., 2022). The remainder of this paper is organized as follows: Section 2 details the methodological framework, including the dual-ANN PINN structure and training strategies. Section 3 introduces the case studies, encompassing both simple and complex bioreactor problems. Section 4 presents the modelling results and comparative performance analysis. Section 5 provides a discussion of the findings in the context of current literature. Finally, Section 6 presents the main conclusions.

2. Methods

2.1. Dual-ANN PINN structure for bioreactor systems

The state-space equation of a perfectly mixed stirred tank bioreactor takes the following general form,

$$\frac{dC}{dt} = Sr(C, u) - DC + DC_{in}, \quad C(0) = C_0, \quad (1a)$$

where C is a $(n \times 1)$ state vector of concentrations of relevant biochemical species, S is a $(n \times m)$ matrix of yield coefficients, $r(C, u)$ is a $(m \times 1)$ vector of biological reaction kinetics, $D = F/V$ is the dilution rate (for simplicity we consider the liquid volume, V , to be controlled by a single feed stream with flow rate, F), C_{in} is a $(n \times 1)$ vector of concentrations in the feed stream and t is the independent variable time. The overall material balance equation for constant liquid density and a single feed stream is defined as,

$$\frac{dV}{dt} = F, \quad V(0) = V_0. \quad (1b)$$

Based on the bioreactor state-space Eqs. (1a,b), a dual-ANN PINN structure is outlined under the assumption that the reaction kinetics term is unknown, i.e., there are no defining equations for the term $r(C, u)$ (Fig. 1A). This PINN structure is composed of a state parameterization feed-forward neural network (FFNN-S) and a reaction kinetics feed-forward neural network (FFNN-R). The rationale behind this dual-ANN structure is to decouple the parameterization of dynamic state variables over time and reaction kinetics as functions of state variables. This modular design reflects the underlying structure of the process and enhances the model's generalization capacity, as further elaborated in the results section.

Both network structures are multi-layered FFNNs with an input layer, one or more hidden layers (to capture the complex interactions and nonlinearities in the system), and an output layer:

$$y = \sigma(W_L \sigma(W_{L-1} \sigma(W_1 x + b_1) + b_{L-1}) + b_L), \quad (2)$$

where y is a vector of outputs, x is a vector of inputs, W_i is a matrix of

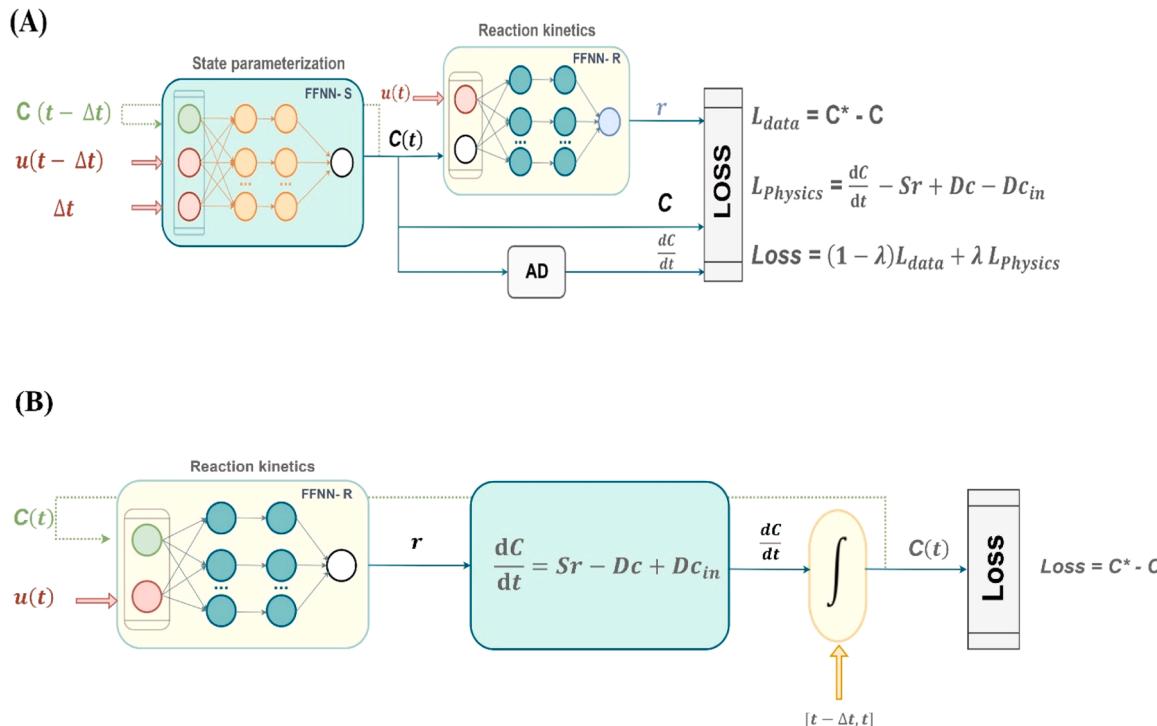


Fig. 1. Schematic diagram of hybrid model structures for a generic stirred-tank bioreactor system. **(A)** Dual-ANN PINN structure where physical equations are embedded in the training loss function, **(B)** Hybrid semiparametric structure where physical equations are hard-coded directly in the model structure.

weights of the i^{th} layer, b_i is a vector of bias parameters and σ is the activation function.

Following the structure of Eq. (2), the FFNN-S is implemented as a one-step predictor. It receives as inputs the state variables, $C(t - \Delta t)$, the control inputs $u_{[t-\Delta t, t]}$ (which are constant over time intervals $[t - \Delta t, t]$), and time step, Δt . The time step, Δt , might not be constant due to heterogeneous data sparsity. When constant, it may drop from the FFNN-S inputs. The predicted concentrations are computed from the neural network outputs as follows:

$$C(t) = C(t - \Delta t) + y(t)\Delta t. \quad (3)$$

Eq. (3) ensures that $C(t) = C(t - \Delta t)$ when $\Delta t = 0$, which eliminates the need for initial boundary conditions in the loss function used to train the PINN. Automatic differentiation (AD) is applied to compute dC/dt , which is required in the physics loss function. The second neural network, FFNN-R, computes the reaction kinetics at the current time, $r(t) = y(t)$, as a function of concentrations at the current time, $C(t)$, and of control inputs at the current time, $u(t)$. The outputs of the PINN are thus $C(t)$, $r(t)$ and dC/dt . These are used to calculate the data loss and physics loss terms. Of note, the prior knowledge of the material balance equation is not directly embedded in the model structure but is rather used to compute the physics loss term.

2.2. Hybrid semiparametric model for bioreactor systems

This study compares the dual-ANN PINN with the more traditional hybrid semiparametric approach. Here we follow the previously published general bioreactor hybrid model concept (Oliveira, 2004; Pinto et al., 2022) applied in many different studies. The hybrid semiparametric model Fig. (1B) contains a single neural network that models the reaction kinetics as a function of the concentrations of species and control inputs. This network parallels the FFNN-R of the dual-ANN PINN structure. The calculated reaction kinetics pass to a system of ODEs derived from the material balance equations. This is a fundamental difference to the dual-ANN PINN structure, as the prior knowledge ODEs are hard-coded directly in the model structure. The ODEs are

numerically integrated over time intervals $[t - \Delta t, t]$ from initial conditions, $C(t - \Delta t)$ to the current concentrations, $C(t)$. The integration was done numerically with a Runge-Kutta 4th order method (RK4). This is another fundamental difference to PINNs, as the hybrid semiparametric approach requires integration whereas the PINN approach requires differentiation. However, it is important to note that the discretization of the process to match sampling time steps can impact the derivative approximation and overall solution accuracy, as highlighted by Cruz-Bournazou et al. (2022). Comparing computed and measured concentrations results in a data loss that is minimized during the.

2.3. Training methods

To ensure comparability, the training, validation and testing methods were identical for both modelling approaches, except for the calculation of the loss function which is intrinsically different in both approaches. The standard Adaptive Moment Estimation Method (ADAM) was applied to minimize the total loss. Specifically, the standard ADAM method described in Kingma and Ba (2014) was applied. In the case of the PINN structure, a loss function with two terms is minimized. The data loss term is computed as the weighted mean squared error (WMSE) as follows,

$$\mathcal{L}_{data} = \frac{1}{N \times M} \sum_{i=1}^N \sum_{j=1}^M \left(\frac{C_{ij}^* - C_{ij}}{\sigma_{Cj}} \right)^2, \quad (4)$$

where N is the number of observations, M is the number of biochemical species, C_{ij}^* and C_{ij} are the measured and model predicted concentrations, respectively, of biochemical species j at time instant i , and σ_{Cj} is the standard deviation of the concentration of biochemical species j over the set of observations. The physics loss is computed as follows:

$$e = \frac{dc}{dt} - Sr + DC - DC_{in}, \quad (5a)$$

$$\mathcal{L}_{\text{physics}} = \frac{1}{K \times M} \sum_{i=1}^K \sum_{j=1}^M \left(\frac{\Delta t e_{ij}}{\sigma_{Cj}} \right)^2, \quad (5b)$$

where K represents the number of collocation points and e_{ij} is the physics residual of biochemical species j at collocation point i . The number of collocation points does not coincide with the number of observations. A random collocation method was employed to distribute these points across the domain space. Furthermore, it was ensured that the total number of residuals, $(N + K) M$, is always higher than the total number of weights of the dual-ANN PINN structure. Finally, the total loss is computed as a weighted sum of the two loss terms:

$$\mathcal{L}_{\text{total}} = (1 - \lambda) \mathcal{L}_{\text{data}} + \lambda \mathcal{L}_{\text{physics}}. \quad (6)$$

The parameter λ is set by the user between 0 and 1 to adjust the relative importance of the physics and data loss terms. For the hybrid semiparametric structures, the physics loss term drops, thus the total loss is simply given as $\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{data}}$.

The computation of gradients is a critical step for the success of both modelling methods. In the hybrid semiparametric case, sensitivity equations are often applied (Oliveira, 2004; Pinto et al., 2022). However, for comparability, the automatic differentiation method implemented in the python package PyTorch (Baydin et al., 2018) was applied in both modelling approaches.

The data was partitioned consistently for both the PINN and semiparametric structures to ensure comparability. Specifically, the dataset was divided into training, validation, and testing subsets, with the specific partitioning varying across case studies. To reduce overfitting, standard cross-validation was adopted in case study 1, while in case study 2 a mini-batch stochastic regularization scheme was employed following Pinto et al. (2022) (further details are provided in Section 3). When comparing different PINN models, the selection of the best model was based on the total test loss considering both the data and physics loss terms. When comparing different model types, the selection of the best model was based on the test data loss term only.

All methods reported in this study were implemented in Python 3.11.5. The results were obtained using Windows 11 Pro on a PC with an Intel Core i9 CPU with 32 GB RAM.

3. Case studies

3.1. Case study 1 - logistic biomass growth in a fed-batch bioreactor

Case study 1 is a very simple logistic growth process in a fed-batch bioreactor with only 2 state variables described by two ODEs, Eqs. (7a,b), and the logistic growth model, Eq. (7c):

$$\frac{dX}{dt} = \mu X - DX, \quad (7a)$$

$$\frac{dV}{dt} = F, \quad (7b)$$

$$\mu = \mu_{\max} \left(1 - \frac{X}{X_{\max}} \right), \quad (7c)$$

where X is the biomass concentration, V is the liquid volume, μ is the specific growth rate, F is the feed rate into the reactor, $D = F/V$ is the dilution rate, μ_{\max} is the maximum specific growth rate, and X_{\max} is the maximum biomass concentration. Synthetic experiments were simulated dynamically using a Runge-Kutta 4th/5th order ODE solver with $\mu_{\max} = 0.3 \text{ h}^{-1}$ and $X_{\max} = 47.3 \text{ g/L}$. Data points were sampled with 1-hour intervals and 5 % Gaussian noise (5 % of true value). A central composite design (CCD) with 3 factors (initial biomass concentration ranging from 0.5–2.5 g/L, initial volume ranging from 1.9–3.5 L, and feed rate, F , ranging from 0.05–0.5 L/h) yielded a data set comprising 15 experiments. Given the underlying simplicity, the main objective in this

case study was to evaluate whether training the models on data from a single fed-batch experiment could effectively capture the underlying process dynamics. As such, the data partition consisted of a training subset with a single reactor experiment (the CCD center point corresponding to 24 observations), cross-validation with a single reactor experiment (a repetition of the center point experiment corresponding to 24 observations) and validation and testing with the remaining 14 experiments (336 observations). Details of the data set and respective partition are provided in Supplementary File 1.

3.2. Case study 2 – Park & Ramirez fed-batch bioreactor

Case study 2 involves a highly nonlinear fed-batch bioreactor of yeast cells expressing a foreign protein. The process is characterized by five state variables and time-varying input feed rate (Park and Ramirez, 1988). The state variables are the biomass concentration (X), glucose concentration (S), secreted protein concentration (P_m), total protein concentration (P_t), and liquid volume in the reactor (V). The bioreactor is governed by the following set of material balance equations:

$$\frac{dX}{dt} = \mu X - DX, \quad (8a)$$

$$\frac{dS}{dt} = -Y\mu X + D(S_{in} - S), \quad (8b)$$

$$\frac{dP_m}{dt} = \theta(P_t - P_m) - DP_m, \quad (8c)$$

$$\frac{dP_t}{dt} = f_p X - DP_t, \quad (8d)$$

$$\frac{dV}{dt} = F \quad (8e)$$

This process has a single feed stream with flow rate, F , and substrate concentration, $S_{in} = 20 \text{ g/L}$. The specific reaction rates are defined by highly nonlinear kinetic equations:

$$\mu = \frac{21.87 S}{(S + 0.4)(S + 62.5)}, \quad (9a)$$

$$\theta = \frac{4.75\mu}{0.12 + \mu}, \quad (9b)$$

$$f_p = \frac{Se^{-5.0S}}{0.1 + S}, \quad (9c)$$

where μ is the specific growth rate, Y is the substrate/biomass yield coefficient, θ is the foreign protein secretion rate, and f_p is the foreign protein synthesis rate. A total of 16 experiments were simulated across a 0–15 h time window, always with the same initial condition ($X(0) = 1.0 \text{ g/L}$, $S(0) = 5.0 \text{ g/L}$, $P_t(0) = 0 \text{ g/L}$, $P_m(0) = 0 \text{ g/L}$, $V(0) = 1.0 \text{ L}$) but with varying feed rate, F , time profiles. Fifteen experiments were designed by a three-factor CCD. The factors were step changes in the feed rate over time, defined as follows: feed rate between 0 and 5 h (F_1), feed rate between 5 and 10 h (F_2) and feed rate between 10 and 15 h (F_3). Each factor was varied between 0 and 2 L/h. The 9 experiments corresponding to the CCD cube and center points were used for training with embedded stochastic regularization. The 6 experiments corresponding to the CCD star points were used for validation. A 16th experiment was simulated using the optimal control feed rate profile that delivers the maximum possible secreted protein in a 10 L bioreactor (32.4 g) (Park and Ramirez, 1988). This optimal control experiment was used as a test experiment. A Runge-Kutta 4th/5th order ODE solver was adopted for process simulation. Data points were sampled with 1-hour intervals and 5 % Gaussian noise (5 % of true value). Details of the data set and respective partition are given in Supplementary File 2.

4. Results

4.1. Development of a bioreactor PINN model for case study 1

The dual-ANN PINN structure was thoroughly investigated for both bioreactor case studies, starting with the simpler logistic growth fed-batch reactor problem. The key objective was to evaluate the PINN approach under conditions of high data sparsity by training it with data from a single experiment (24 observations) and validation and testing it with data from the remaining 14 experiments (336 observations). The data partition was kept the same across all trials. The prior knowledge used to compute the physics loss included the biomass and volume material balances, Eqs. (7a,b), but excluded the logistic growth rate law, Eq. (7c). The data loss included biomass but excluded volume measurements. This was intentional to evaluate whether the PINN could predict unmeasured volume dynamics given that the physics loss term includes Eq. (7b), which precisely defines volume dynamics.

To optimize the PINN structure, a systematic hyperparameter search was conducted using Optuna (Akiba et al., 2019), an automated framework employing Bayesian optimization and the Tree-structured Parzen Estimator (TPE) algorithm. The optimization targeted key hyperparameters, including the learning rate, activation functions in both networks (\tanh , SiLU, and gelu), number of layers (n_{layers}), nodes (n_{units}) in FFNN-S, and nodes in FFNN-R (model2_nodes ; a single hidden layer was kept given the simplicity of the unknown kinetic term defined by Eq. (7c)) over 25 trials. At this stage, the pre-optimized weighting parameter, $\lambda = 0.5$, and number of collocation points, K , equivalent to the size of FFNN-S ($K = NWS$) were excluded from the hyperparameter search (further discussed below). Fig. 2 depicts a parallel coordinate plot of Optuna's search, where lines represent trials and color indicates objective value. Optuna converged to an optimal dual-ANN PINN configuration with FFNN-S sized as $3 \times 12 \times 12 \times 12 \times 2$ with SiLU activation in the hidden layers, FFNN-R sized as $1 \times 3 \times 1$ with \tanh activation in the hidden layer, and a learning rate of 0.005. This configuration achieved a total training loss of 0.002 and a total validation loss of 0.007.

The automated search was further refined by a manual grid search of key hyperparameters. The SiLU and \tanh activation functions derived from Optuna were kept but the sizes and learning rate were further investigated in the manual grid search. Detailed performance metrics for each PINN variation are presented in Table S1 of Supplementary File 3. The manual refinement converged to an optimal PINN structure (FFNN-S: $3 \times 8 \times 8 \times 8 \times 2$ with SiLU, FFNN-R: $1 \times 5 \times 1$ with \tanh) and a learning rate of 0.0075. The slightly higher learning rate was shown to accelerate convergence without compromising performance. This configuration achieved the same training (0.002) and validation (0.007)

total loss of Optuna, but with a lower number of weights.

Fig. (3A–D) illustrates the predictive capability and adherence to physics constraints of the selected PINN structure for case study 1. Fig. (3A) and (3B) show the model's performance on the training and validation experiments, respectively. The coefficient of determination for both biomass (X) and volume (V) is high and comparable for the train and validation data subsets. Moreover, the final train data loss was 0.0046, practically coincident with the noise level (0.0049), showing negligible overfitting. The final validation data loss was 0.013, almost twofold the noise level (0.007), but even so very low.

Fig. (3C) depicts the model's adherence to the underlying physical laws at randomly generated collocation points. These points were sampled throughout the design space (biomass, volume, and feed rate) to enforce the physical constraints across the relevant operational range. At each training epoch, a different set of collocation points was randomly generated within the known bounds to ensure a comprehensive coverage of the design space. The FFNN-S output derivatives (dX/dt and dV/dt) were computed at the collocation points by automatic differentiation and then compared against the derivative values of the governing differential equations, Eqs. (7a,b). The strong linearity observed in the derivative parity plot of Fig. (3C) highlights the model's ability to accurately satisfy the governing differential equations. In addition, Fig. (3D) extends the physics adherence analysis to the validation data points showing that the physics constraints are also obeyed in the validation data space.

Overall, these results show that the dual-ANN PINN succeeded in describing the biomass and volume dynamics in all validation experiments using training data from a single experiment. Notably, the prediction of volume dynamics relied solely on the embedded physics law, Eq. (7b). This simple example demonstrates that when full and precise physics of a subset of state variables are known, observations of those state variables are in principle not needed to train the PINN. This highlights a key advantage of PINNs over conventional neural networks since the latter always need data for their training.

4.2. Development of a bioreactor PINN model for case study 2

The dual-ANN PINN framework was applied to the more complex Park and Ramirez (1988) fed-batch bioreactor benchmark, which has a higher dimensionality of state variables (biomass (X), substrate (S), total protein (P_t), secreted protein (P_m), and volume (V)), highly nonlinear kinetics and time-varying control inputs. The key objective was to evaluate whether a PINN trained on a sparse data set (9 out of the 15 CCD experiments) could accurately predict the optimal production scenario deduced by Park and Ramirez (1988) (32.4 g of secreted protein when the optimal feed profile is applied). The physics loss included

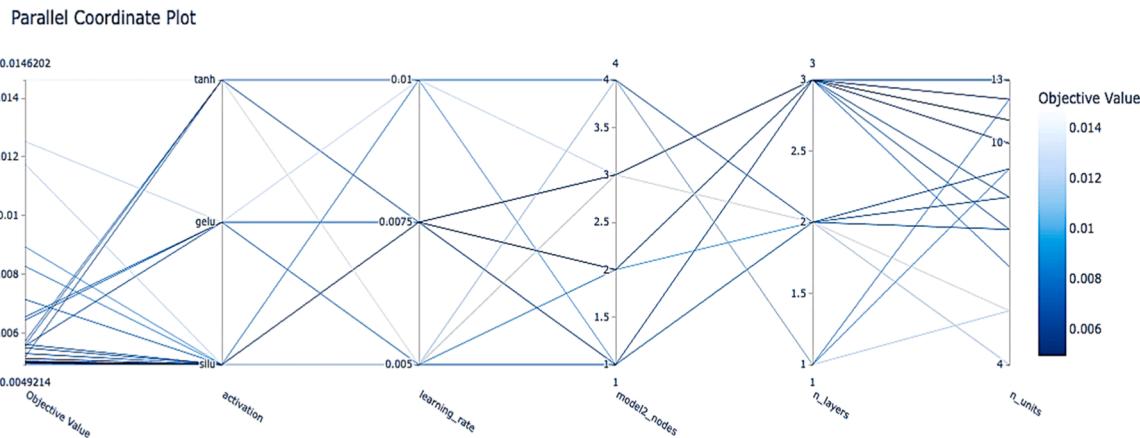


Fig. 2. Optuna hyperparameter search results: Parallel coordinate plot showing the relationship between hyperparameters and objective function values for case study 1 over 25 trials.

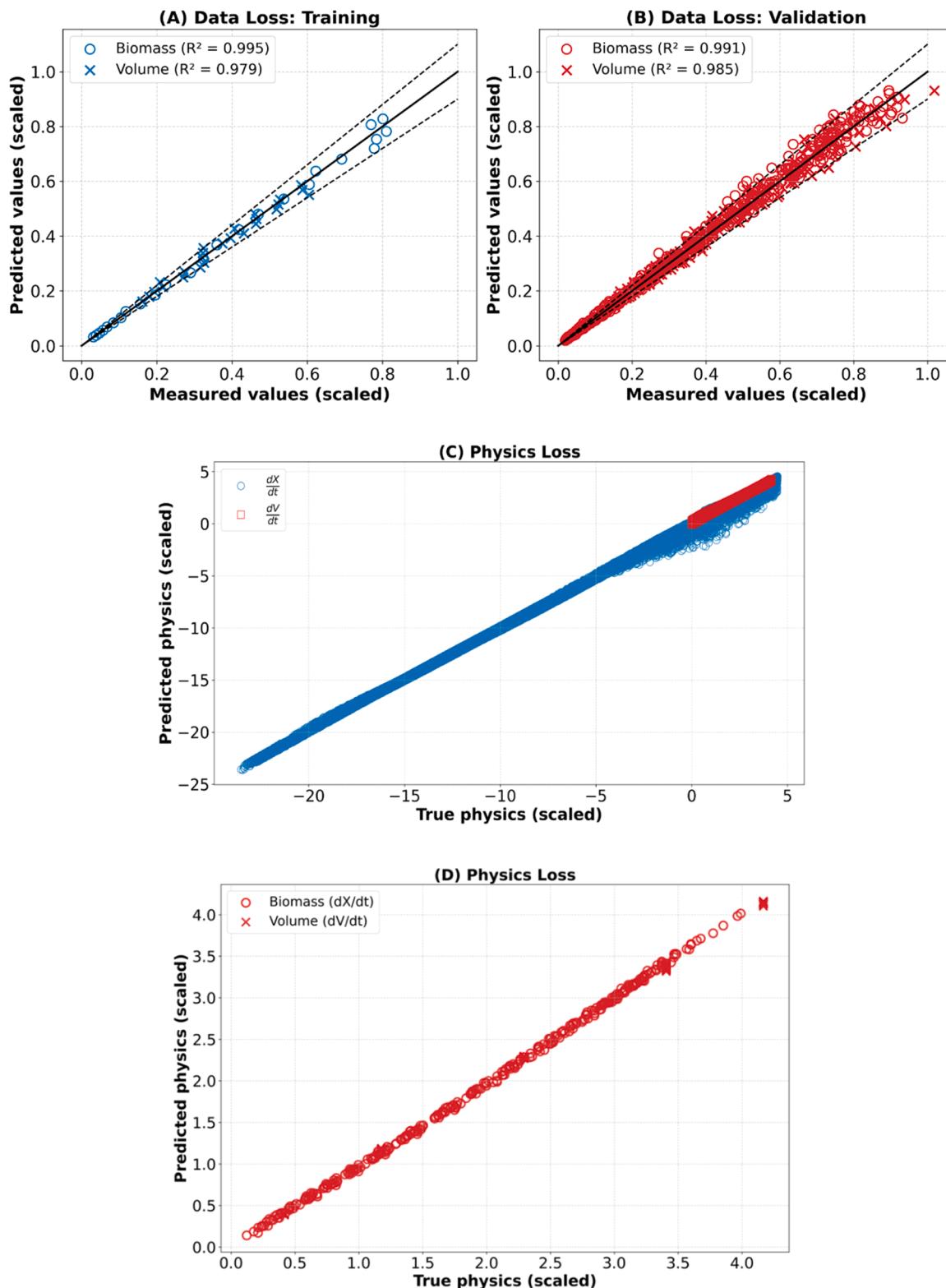


Fig. 3. Training results for case study 1 using a PINN structure (FFNN-S: $3 \times 8 \times 8 \times 8 \times 2$ with SiLU, FFNN-R: $1 \times 5 \times 1$ with tanh) trained with a learning rate of 0.0075, $\lambda = 0.5$ and $K = 194$ collocation points. (A) Predicted versus measured biomass and volume data for the training set, (B) Predicted versus measured biomass and volume data for the validation set, (C) Biomass and volume physics at the training collocation points, (D) Biomass and volume physics at the validation data points.

the material balance equations of the 5 state variables, Eqs. (8a-e), and excluded the reaction kinetics, Eqs. (9a-c). The same collocation points scheme of case study 1 was applied. The data loss included measurements of all state variables except volume (X, S, P_t , and P_m). The volume

training relied exclusively on the physics loss term (based on Eq. (8e)) as in case study 1. The CCD center and square experiments (9) were used for training whereas the CCD star experiments (6) were used for validation. Overfitting was minimized by stochastic regularization whereby

6 out of 9 training experiments were randomly selected to compute the training data loss at each epoch. Building on the optimal dual-ANN PINN structural features already identified for case study 1, a manual grid search of the optimal FFNN-S and FFNN-R sizes was conducted as detailed in Table S2 of Supplementary File 3. This procedure converged to the optimal structure (FFNN-S: $6 \times 17 \times 17 \times 5$ with SiLU, FFNN-R: $1 \times 6 \times 3$ with tanh). Fig. (4A–D) show the training, validation and testing performance of the selected dual-ANN PINN. A strong agreement between predicted and measured values across all five state variables is observed for the training data set (Fig. 4A). Notably, the coefficient of determination (R^2) exceeded 0.99 for all state variables showing excellent training performance. However, the model's generalization to the test experiment degraded significantly (Fig. 4B). The (R^2) values for X (0.95), S (0.90), and V (0.95) remained relatively high, but a significant performance degradation was observed for P_t (0.58) and P_m (0.68). The parity plot of Fig. (4C) illustrates the PINN's adherence to the underlying physical laws at randomly generated collocation points. The close alignment along the diagonal confirms a high level of adherence to the physics constraints. However, the generalization of physics equations to the test data points is severely compromised (Fig. 4D), particularly those associated with protein dynamics (P_t and P_m), which exhibit a significant scatter.

The Park and Ramirez (1988) problem clearly poses a greater challenge for the PINN's ability to generalize than in case study 1. Given the excellent training performance, the degradation of predictive power could be caused by insufficient information content in the training set. To test this hypothesis, the selected PINN structure was retrained using the full CCD data set (the center, square, and star points of the CCD; 15 experiments). Stochastic regularization was adjusted to randomly selected 10 out of 15 experiments for the calculation of the training data loss at each epoch. The overall results are shown in Fig. (4E, F). Indeed, augmenting the training set significantly improved the testing performance. As before, the PINN succeeded in describing the full training data set, with R^2 higher than 0.99. The augmented training set likely provided a more comprehensive data coverage enabling the PINN to more accurately learn the intricate nonlinearities. Despite the tenfold decrease in the test data loss (from 0.133 to 0.0128), some scatter persists with a few predictions laying outside the 5 % error bound, suggesting there is still room for improvement (further to this in the comparison with the hybrid semiparametric model below).

4.3. Training convergence

The training of PINNs is complicated by the need to simultaneously minimize the training and physics loss terms. This study followed the most common approach whereby a total loss function, defined as the weighted sum of physics and data loss terms, Eq. (6), is minimized. In this approach, it is critically important that both loss terms are correctly normalized. Fig. (5A, B) elucidates the training convergence behavior of the selected dual-ANN PINN structures for case studies 1 and 2, which include the evolution of the data loss, physics loss, and total loss across training epochs.

In both cases a stable although noisy convergence behavior is observed, characterized by a rapid initial decline in the total loss function, encompassing both data and physics losses, followed by a prolonged plateau at low loss values. In case study 1, stochasticity is introduced in the physics loss function via the random selection of collocation points, which propagates to the data loss function via the update of neural network parameters. In case study 2, stochasticity originates from both the physics loss (random collocation points) and data loss (random experiments to compute the data loss). The physics loss term tends to plateau always below the data loss term. This is explained by the random experimental error that is intrinsic to the data residuals but absent in the physics residuals. Importantly, the data loss terms plateau close to the WMSE noise level (0.0049 for case study 1

and 0.013 for case study 2) showing that in both cases overfitting was successfully mitigated.

4.4. Relative importance of data and physics loss

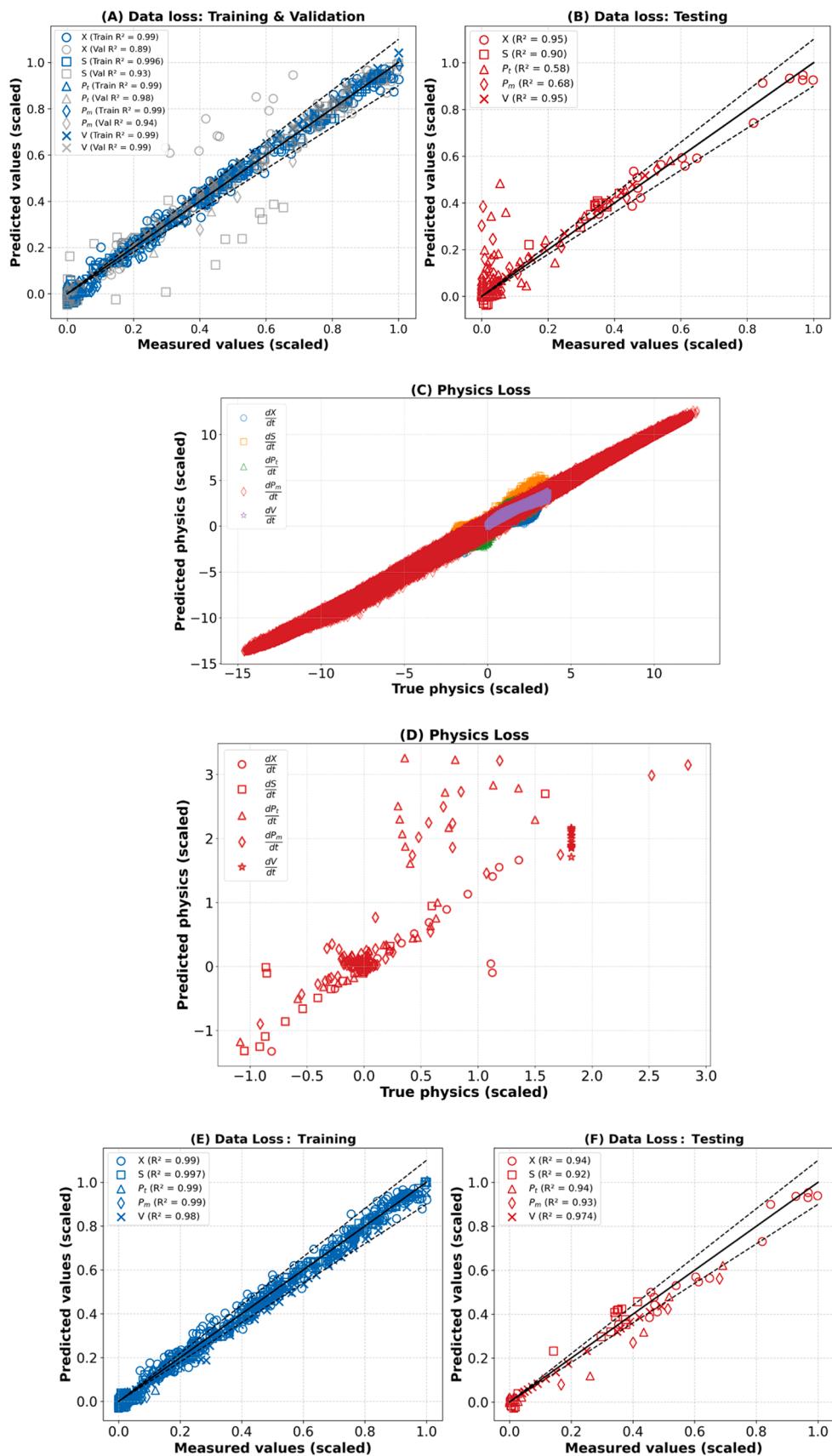
The weighting parameter, λ , sets the relative importance of the physics loss (λ) and data loss ($1 - \lambda$). The choice of λ may significantly influence the training convergence and overall model performance. The results above assumed equal importance of physics and data loss terms ($\lambda = 0.5$). Fig. (6A) shows the effect of λ on the final training and validation losses for the selected dual-ANN PINN structure for case study 1. At $\lambda = 0.0$, the physics loss is not minimized thus the PINN becomes analogous to a conventional neural network model. As expected, with $\lambda = 0.0$ the PINN succeeded in describing the training data (low final data loss). However, the validation loss was the highest among all runs performed. This highlights the importance of "physics learning" for the PINN's ability to extrapolate to unseen observations. As λ increases from 0.25 to 0.9, a clear decrease trend in the training and validation physics loss is observed. Notably, this is accompanied by a decreasing trend in the validation data loss denoting synergies between physics and data learning. Setting $\lambda = 0.5$ resulted in the lowest loss values, thus providing the optimal trade-off between data and physics loss terms. Fig. (6B) shows similar results for the selected dual-ANN PINN structure of case study 2. As in the previous case, a weighting factor of $\lambda = 0.5$ yielded the optimal balance between data and physics loss minimization. All in all, these results highlight the synergy between the data and physics learning components in the PINN framework. Learning physics is shown to be critically important for the PINN's extrapolation capacity, proving a clear advantage over conventional neural models, as further discussed below.

4.5. Effect of collocation points

The number of collocation points (K) directly influences the physics loss minimization and indirectly the training convergence and overall model metrics. In this study, the number of collocation points was set as multiples of the number of weights of the FFNN-S (NWS). Choosing $K/NWS \geq 1$ thus provides a sufficient number of physics residuals to train FFNN-S even when no data points are available. Increasing the K/NWS ratio could in theory positively influence the training and validation metrics. Fig. (7A, B) shows the effect of the ratio $K/NWS = 1, 2, 3$ on the training and validation loss terms for the selected dual-ANN PINN structures of case studies 1 and 2. As anticipated, increasing K/NWS generally reduces the training and validation physics loss term, with, however, a more marked improvement in case study 1 than in case study 2. The physics loss reduction is, nevertheless, not reflected in the data loss terms, which seem to be relatively insensitive to the K/NWS ratio. It was concluded that a $K/NWS = 1$ is sufficient for achieving robust model convergence. Importantly, the computational cost (CPU) escalates linearly with K/NWS . Consequently, a ratio of 1 presents an optimal balance between model accuracy and computational efficiency. For this reason, a value of $K/NWS = 1$ was used in all trials conducted in the present study.

4.6. Comparison between single- and dual-ANN PINN models

While most of previously published studies addressed single-ANN structures (e.g. Raissi et al., 2019; Wang et al., 2021, 2022; Yang et al., 2024; Jul-Rasmussen et al., 2025) the present study focuses on a dual-ANN PINN structure. To elucidate this aspect in a bioreactor modeling context, dual-ANN PINN structures were systematically compared with single-ANN PINN structures where all state variables and kinetic rates are computed by a single FFNN. To ensure a fair comparison, models with equivalent sizes (number of hidden layers and number of nodes) were subject to identical training, validation, and testing processes. Detailed performance metrics for every single-ANN PINN



(caption on next page)

Fig. 4. Training results for case study 2 using a PINN structure (FFNN-S: $6 \times 17 \times 17 \times 5$ with SiLU, FFNN-R: $1 \times 6 \times 3$ with tanh), trained with learning rate 0.0075, $\lambda = 0.5$ and $K = 515$ collocation points, when the model is trained over a partial data set with 9 batch experiments (center and square points of the CCD) (A-D) or over the full data set with 15 experiments (full CCD) (E-F). (A) Predicted versus measured X, S, P_t, P_m, V for the partial training set (9 experiments), (B) Predicted versus measured X, S, P_t, P_m, V for the testing set (optimal Park and Ramirez (1988) bioreactor with 32.4 g secreted protein production), (C) X, S, P_t, P_m, V physics at the training collocation points, (D) X, S, P_t, P_m, V physics at the test data points, (E) Predicted versus measured X, S, P_t, P_m, V when the PINN is trained on the full CCD (15 experiments), (F) Predicted versus measured X, S, P_t, P_m, V for the testing set (optimal Park and Ramirez (1988) bioreactor with 32.4 g secreted protein production).

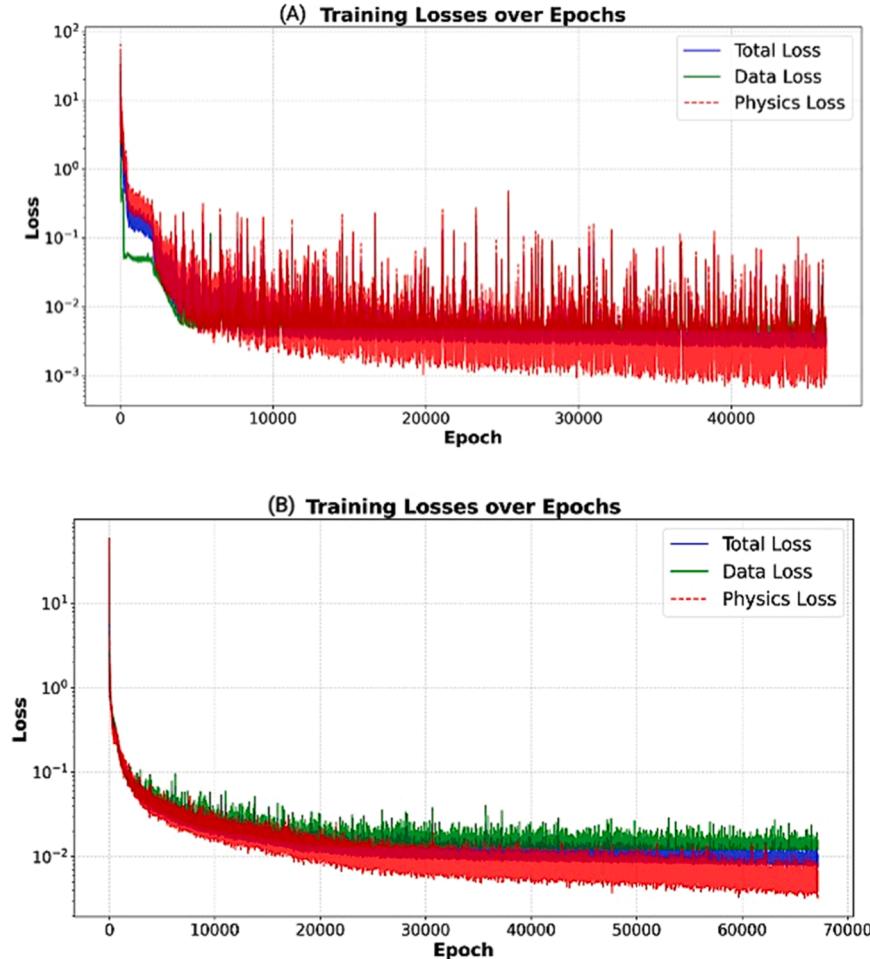


Fig. 5. Loss function evolution as function of training epoch with a learning rate of 0.0075, $\lambda = 0.5$ and $K/NWS = 1$. (A) PINN structure (FFNN-S: $3 \times 8 \times 8 \times 8 \times 2$, FFNN-R: $1 \times 5 \times 1$) for case study 1, (B) PINN structure (FFNN-S: $6 \times 17 \times 17 \times 5$, FFNN-R: $1 \times 6 \times 3$) for case study 2.

structure investigated are given in Tables S1 and S2 of Supplementary File 3 for case studies 1 and 2 respectively.

Fig. (8A, B) show the obtained results in terms of data, physics and total loss ratios (single-ANN PINN) / (dual-ANN PINN) for structures with equivalent sizes. In case study 1 (Fig. 8A), the training loss ratio medians are close to 1, indicating that single- and dual-ANN PINNs of comparable sizes achieve similar training performance. However, a notable difference is observed in the validation dataset, where the loss ratio is significantly greater than 1, especially for the data loss and total loss components. Fig. (8B) extends the loss ratio analysis to the more complex case study 2. The differences between the single- and dual-ANN PINNs are not so straightforward in this case. The training performance are better for single-ANN structures, suggesting an additional training burden of dual-ANN structures when complex physics are involved. However, the validation data loss was lower for the dual-ANN PINN.

Taking both case studies together, it may be concluded that the dual-ANN PINN more accurately captured the nonlinear process dynamics translating in higher predictive power. Since both structures underwent

the same training procedures, the dual-ANN PINN predictive advantage may be explained by a better structural alignment with the actual process. Specifically, the dual-ANN PINN assumes static reaction rates that depend solely on concentrations, whereas the single-ANN PINN dynamically parameterizes the kinetic rates, a formulation that does not reflect the true system behavior. In fact, the dual-ANN PINN follows a modular design, where neural network models are interconnected based on the topological and functional structure of the process under study (Thompson and Kramer, 1994). In other words, the dual-PINN structure is informed by prior knowledge in a classical hybrid modeling sense, providing an additional descriptive advantage in representing the underlying process.

4.7. Comparison with hybrid semiparametric modelling

PINNs have been shown to yield lower prediction errors than conventional neural networks in several modelling studies (e.g., Bangi et al., 2022; Moayedi et al., 2024; Velioglu et al., 2025). Only a few studies

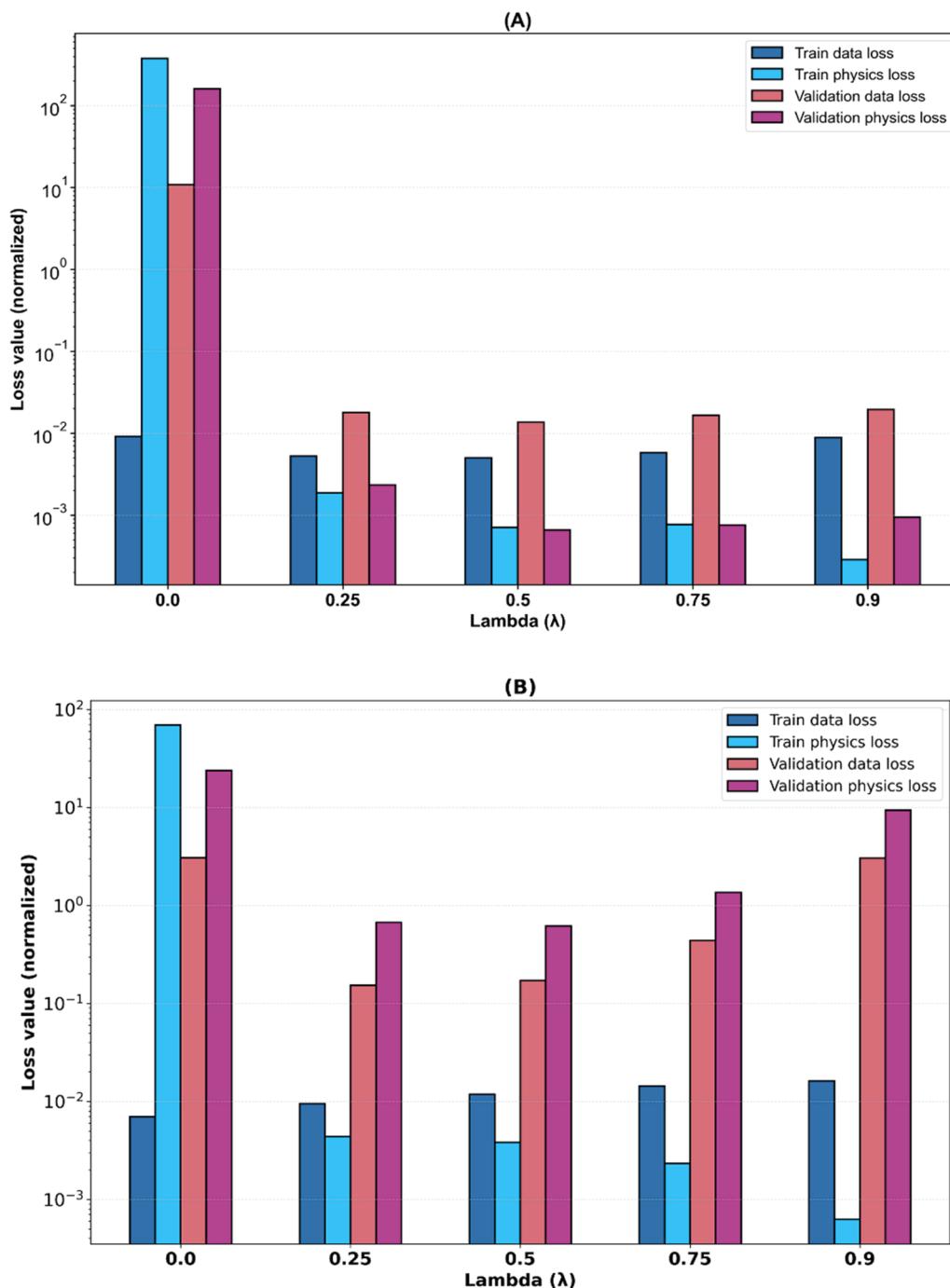


Fig. 6. Dependency of data and physics losses for different λ values. (A) Selected PINN structure for case-study 1 (FFNN-S: $3 \times 8 \times 8 \times 8 \times 2$, FFNN-R: $1 \times 5 \times 1$), (B) Selected PINN structure for case-study 2 (FFNN-S: $6 \times 17 \times 17 \times 5$, FFNN-R: $1 \times 6 \times 3$).

compared PINNs and hybrid semiparametric modelling (Yang et al., 2024; Jul-Rasmussen et al., 2025). It is particularly interesting to analyze if both approaches perform differently when exposed to the same data and prior knowledge. To assess this, selected dual-ANN PINN, hybrid semiparametric and conventional neural network structures (without the physics component) were compared for both case studies. The FFNN-R structure of the hybrid semiparametric model and the FFNN-S structure of the conventional FFNN model were systematically evaluated by manual grid search using the same training, validation and testing procedures applied to the dual-ANN PINN. Details of performance metrics are provided in Tables S1 and S2 for case studies 1 and 2 respectively (Supplementary File 3). The selection of the best model in

each category was based on the validation loss, whereas the comparison of selected best models was based on the test data loss. The case study 1 test experiment presents a particularly challenging scenario, as the time scale was extended to 166 h to assess the models' ability to perform temporal extrapolation. The case study 2 test experiment is equally challenging as the models are inputted with the optimal feed rate profile (unseen during the training) and tasked with predicting the process dynamics that maximizes total secreted protein concentration at the end of the cultivation. A summary of the results is presented in Table 1.

In case study 1, the dual-ANN PINN, hybrid semiparametric, and conventional FFNN models all achieved the same final training data loss (Table 1). However, the hybrid semiparametric model demonstrated

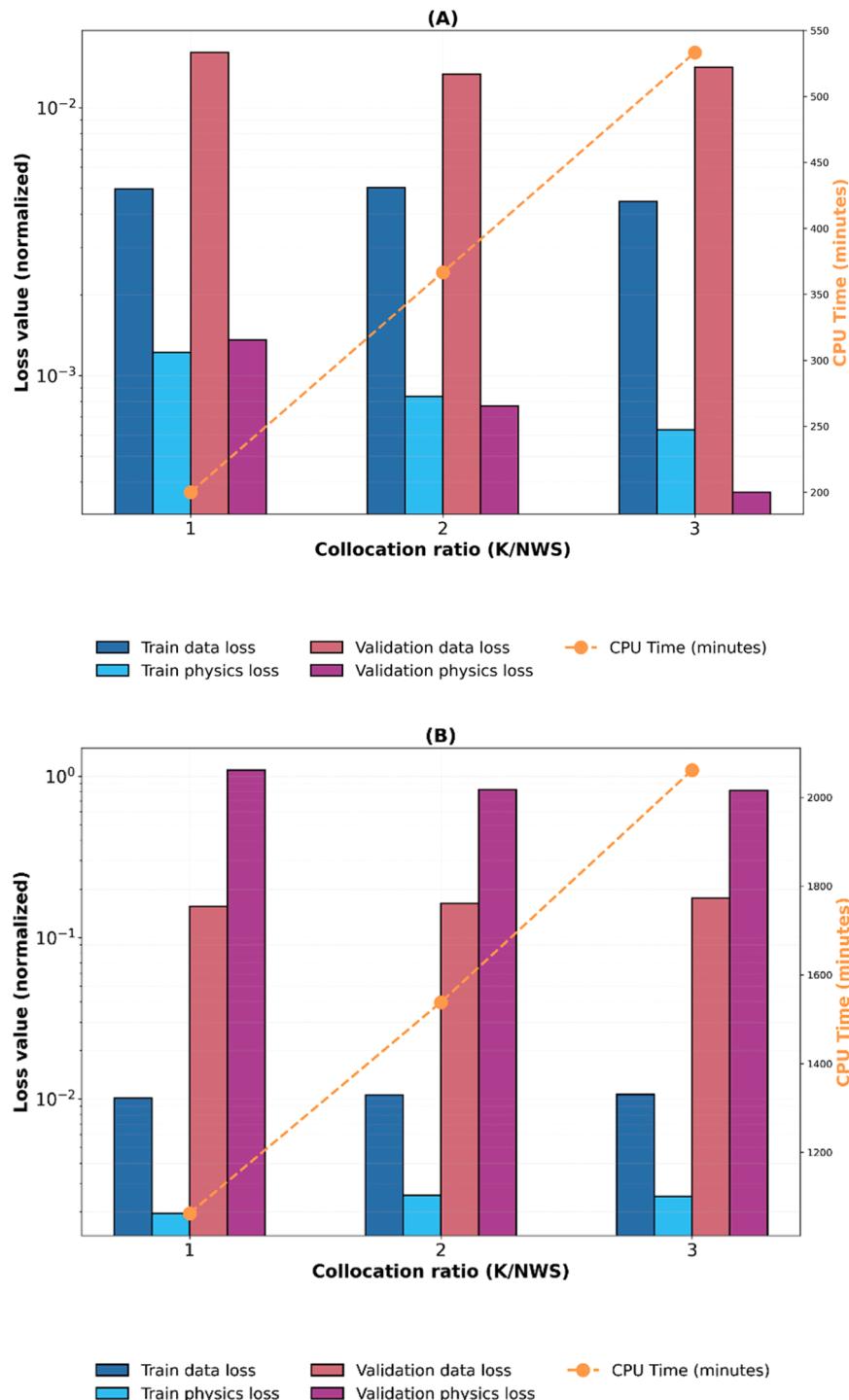


Fig. 7. Dependency of data and physics losses for different collocation point ratios $K/NWS = 1, 2, 3$. **(A)** Selected dual-ANN PINN structure for case-study 1 (FFNN-S: $3 \times 8 \times 8 \times 8 \times 2$, FFNN-R: $1 \times 5 \times 1$), **(B)** Selected dual-ANN PINN structure for case-study 2 (FFNN-S: $6 \times 17 \times 17 \times 5$, FFNN-R: $1 \times 6 \times 3$).

superior prediction accuracy to that of the PINN (~50 % lower test data loss). Fig. (9A, B) further detail the prediction profiles for the test experiment of case study 1. Both the dual-ANN PINN and semiparametric models effectively approximated the growth curve within the scatter of the experimental data. In contrast, the conventional FFNN severely failed to predict the true dynamic trajectory. When evaluated under extended temporal extrapolation conditions, the hybrid semiparametric model converged to a stable final biomass concentration with an 8 % underprediction off-set relative to the true value. The dual-ANN PINN displays, however, unstable behaviour, diverging from the

final biomass concentration over time. This apparent disadvantage of the dual-ANN PINN is not incidental as similar behaviour was consistently observed across multiple experiments. Furthermore, the hybrid semiparametric model demonstrated perfect volume prediction (Fig. 9B) because the full physics is embedded in the model, leaving numerical integration as the sole source of error. In contrast, the PINN shows an increasing error over time, likely due to limitations in the extrapolation of the learned $\frac{dy}{dt} = F$ equation. Of note, the conventional FFNN does not predict volume as volume data points were not available for training. Fig. (9C, D) presents the data loss for all CCD experiments over the

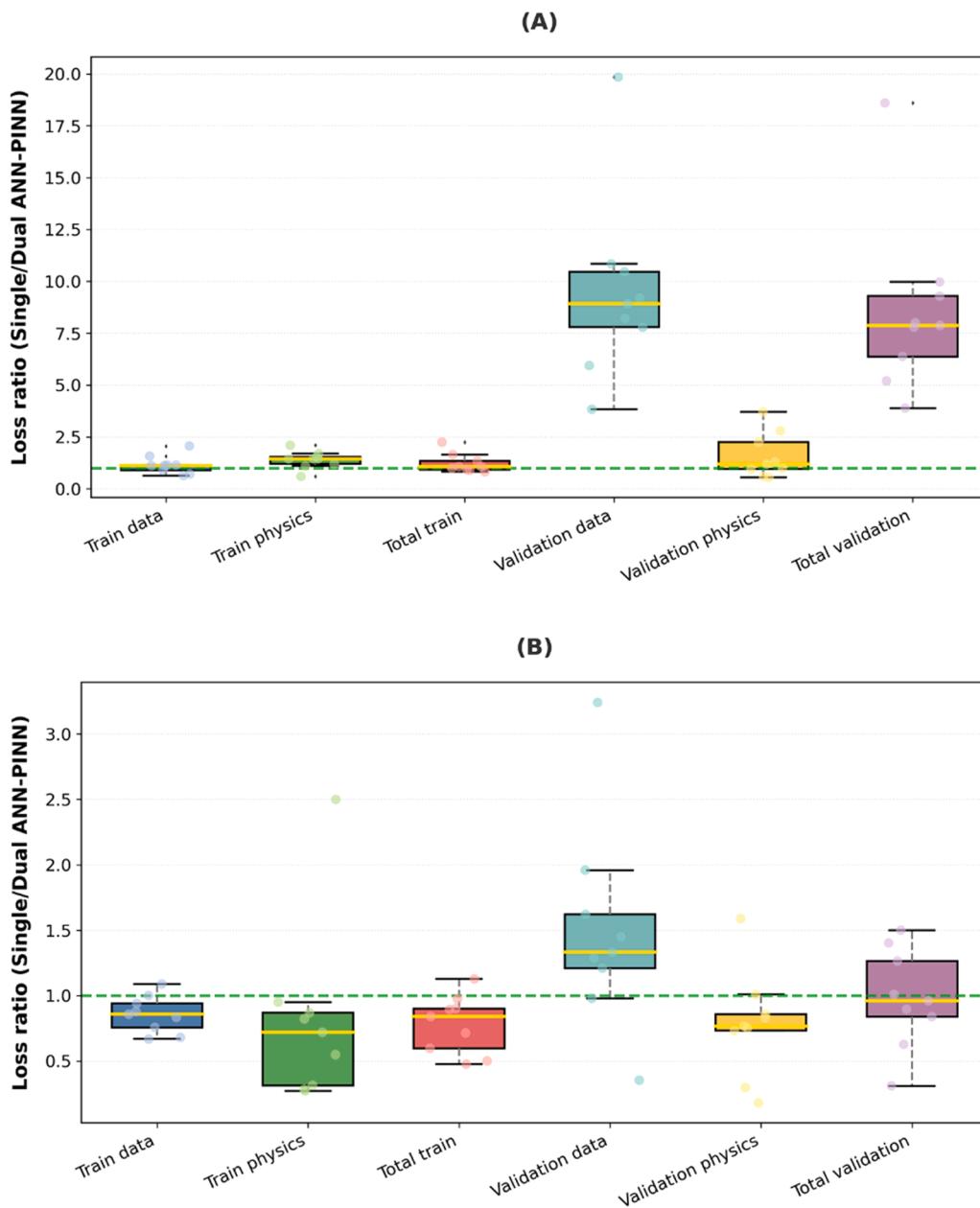


Fig. 8. Loss ratio (Single-ANN PINN) / (Dual-ANN PINN) for structures with equivalent size. **(A)** Case-study 1, **(B)** Case-study 2.

Table 1

. Comparison of training, validation and testing performance metrics of selected Dual-ANN PINN, conventional FFNN and Hybrid semiparametric models across case studies 1 and 2.

Case Study	Model	FFNN-S	FFNN-R	Train data loss	Cross-Validation data loss	Validation data loss	Test data loss	Nw	CPU (min)
1	Dual-ANN PINN	$3 \times 8 \times 8 \times 8 \times 2$	$1 \times 5 \times 1$	0.0046	0.0054	0.013	0.318	210	109.2
	Conventional FFNN [‡]	$3 \times 3 \times 1$	NA	0.0046	0.0070	168.8	1604.5	16	5.6
	Hybrid semiparametric *	NA	$1 \times 5 \times 1$	0.0046	0.0085	0.007	0.036	16	103.9
2	Dual-ANN PINN	$6 \times 17 \times 17 \times 5$	$1 \times 6 \times 3$	0.012	NA	0.133	0.021	548	732.3
	Conventional FFNN [‡]	$6 \times 5 \times 5 \times 4$	NA	0.011	NA	0.521	0.327	89	125.2
	Hybrid semiparametric*	NA	$1 \times 6 \times 3$	0.013	NA	0.039	0.012	33	702.0

(‡) The conventional neural network models consisted of a single FFNN-S structure trained with data of measured state variables without the physics loss.

(*) The hybrid semiparametric models followed the general structure of Fig. (1B). The material balance ODEs were inserted directly in the model structure in replacement of the FFNN-S, namely Eqs. (7a, b) for case study 1 and Eqs. (8a-e) for case study 2. The reaction kinetics were considered unknown thus described by a FFNN-R with the same topology of the selected dual-ANN PINN structure;

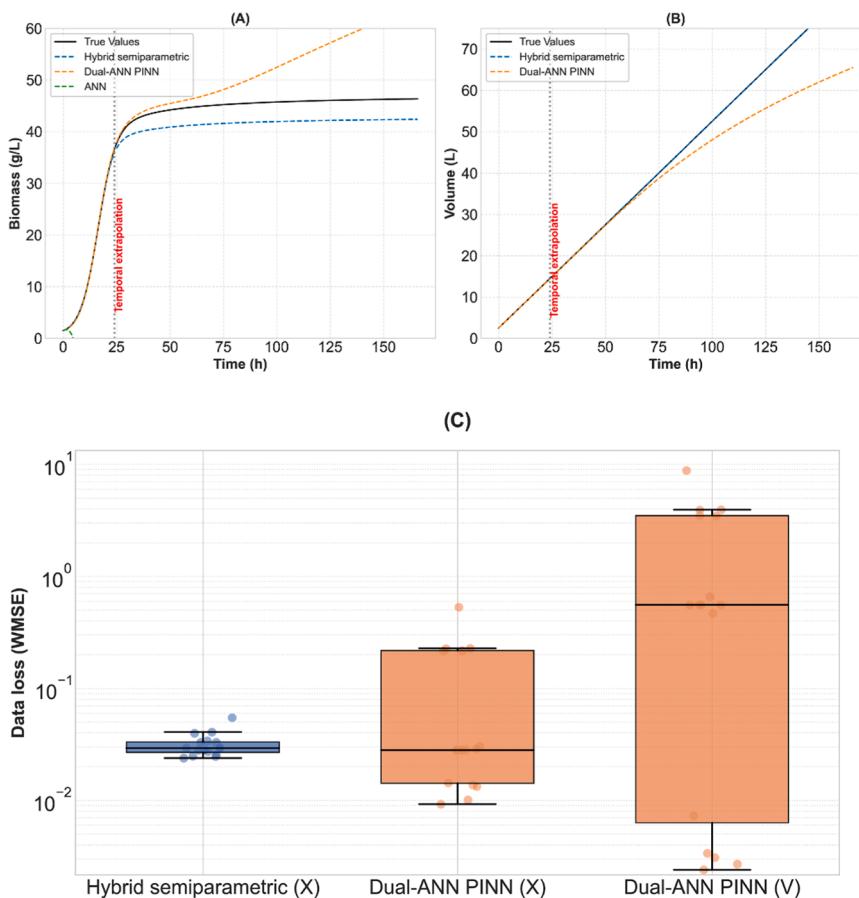


Fig. 9. Compares the dual-ANN PINN (FFNN-S: $3 \times 8 \times 8 \times 8 \times 2$, FFNN-R: $1 \times 5 \times 1$) and hybrid semiparametric (ODE(2); FFNN-R: $1 \times 5 \times 1$) extrapolation capacity for case study 1. (A) Biomass prediction for the test experiment over an extended timeframe 0–166 h, (B) Volume prediction for the test experiment over an extended timeframe 0–166 h, (C) Biomass and volume data loss distribution across all (training + validation + testing) batches under extended temporal extrapolation (0 - 166 h).

extended timeframe, confirming that the hybrid semiparametric model consistently outperformed the PINN under temporal extrapolation conditions. It is worth mentioning that training with a single experiment represents an extreme case of data sparsity. Increasing the number of training batches to two significantly reduced the hybrid semiparametric offset and significantly enhanced the PINN extrapolation performance (results not shown).

As for case study 1, the PINN, hybrid semiparametric, and conventional FFNN models achieved practically the same final training data loss in case study 2, denoting a comparable descriptive power of the training data. The prediction accuracy of the Park and Ramirez (1988) optimal experiment is however very divergent among the 3 models. The hybrid semiparametric model achieved the lowest test data loss of 0.039, followed by the PINN, with a 3.4-fold increase. As expected, the conventional FFNN model had the lowest prediction accuracy, showing a 13.1-fold increase in the test data loss in relation to the hybrid semiparametric model. Fig. (10A–F) depict the predicted profiles of state variables for the optimal Park and Ramirez (1988) feed rate profile. Fig. (10F) shows the piecewise constant optimal feed rate profile (1 h step size) that delivers optimal protein production ($P_m(15)V(15) = 32.4\text{ g}$ at 15 h cultivation time). The hybrid semiparametric model provides a more accurate prediction of the true process dynamics than the other approaches. The final prediction of secreted protein production was 32.48 g, which is virtually identical to the optimal value of 32.4 g. In contrast, the dual-ANN PINN shows significant underperformance compared to the hybrid semiparametric model, particularly in the predictions of S , P_t , P_m and V . The final prediction of secreted protein was only 28.16 g (13.1 % underprediction). The conventional FFNN showed

the highest deviations from the true process profiles (as for case study 1, the conventional FFNN model was not trained with volume data thus it is unable to predict volume). Importantly, the conventional FFNN model occasionally predicted negative concentration values of S , P_t and P_m . This problem was largely mitigated (except for substrate) in the case of the dual-ANN PINN due to the physics regularization during the training. Substrate poses a significant challenge because its concentration drops to nearly zero in the final four cultivation hours. The dual-ANN PINN could not cope with negative concentrations most likely to inaccurate material balance equations extrapolation (inaccuracies in the learned physics). The hybrid semiparametric model practically eliminated the negative concentrations as the material balance equations are numerically integrated.

5. Discussion

This study presents a comparison of a dual-ANN PINN structure and hybrid semiparametric modelling for bioprocesses, evaluated across two case studies with varying levels of mechanistic complexity. The comparative analysis elucidates the strengths and limitations of each modelling paradigm, particularly under data-scarce regimes and extrapolative tasks. In case study 1, which involved logistic microbial growth governed by relatively simple dynamics, both the dual-ANN PINN and hybrid semiparametric models achieved excellent predictive performance within the data domain. The dual-ANN PINN effectively inferred unobserved volume dynamics using only biomass measurements, showcasing the potential of embedding physical knowledge within the loss function to predict unobserved target variables.

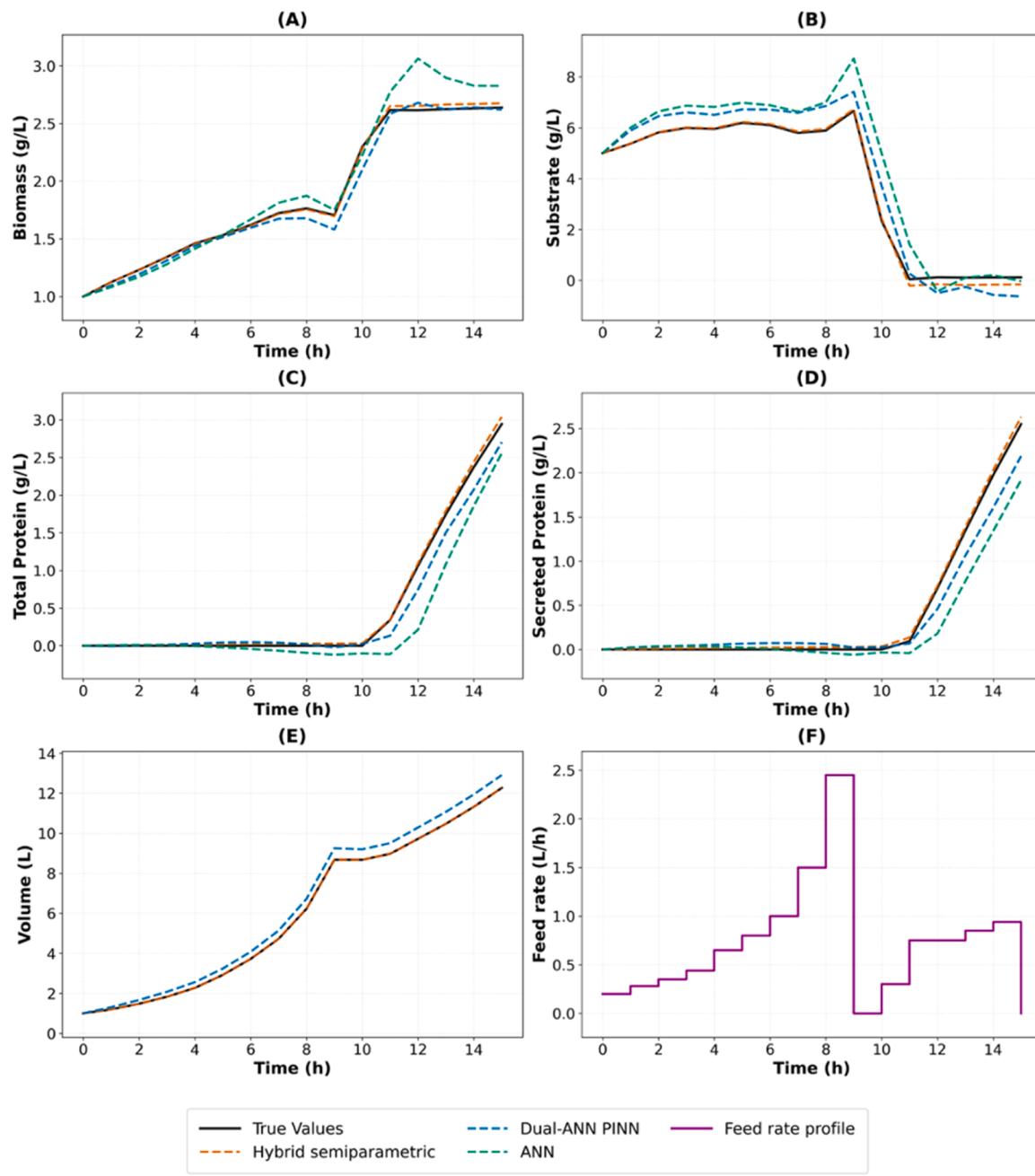


Fig. 10. Compares the selected best dual-ANN PINN, hybrid semiparametric, and conventional FFNN models (Table 1) for the Park and Ramirez (1988) optimal control experiment (test data set). (A) Biomass overtime, (B) Substrate overtime, (C) Total protein overtime, (D) Secreted protein overtime, (E) Volume overtime, (F) Pricewise constant optimal feed rate profile.

However, as shown in Fig. 9, the hybrid semiparametric model consistently outperformed the dual-ANN PINN in extended temporal extrapolation, benefiting from the direct integration of governing equations in the model structure.

Case study 2 posed a significantly more complex scenario, involving five interacting state variables, nonlinear reaction kinetics, and time-varying control inputs. Under these conditions, the performance of the dual-ANN PINN deteriorated notably, particularly in extrapolating the dynamics of secreted protein (P_m) and total protein (P_t). The hybrid semiparametric model again outperformed the dual-ANN PINN, offering lower training and testing loss values and more accurate predictions across all state variables. The dual-ANN PINN struggled to generalize effectively to unseen feed profiles, a result that is consistent with literature studies that reported PINN shortcomings in capturing nonlinear

dynamics due to gradient pathologies and the stiffness in the optimization process, where different parts of the solution space vary on different scales (Wang et al., 2021, 2022). Li and Feng (2022) emphasized the importance of balancing physics and data losses during training, a challenge that our study addressed through manual tuning of the weighting parameter λ , as discussed in Section 4.4. By contrast, the hybrid semiparametric model maintained stable and accurate performance across both case studies. Its architecture, that incorporates differential equations directly in the model structure, enabled transparent integration of mechanistic knowledge, reduced training complexity, and improved convergence. Similar results were reported by Jul-Rasmussen et al. (2025).

Unsurprisingly, the conventional FFNN consistently underperformed in both case studies due to its lack of mechanistic structure, displaying

high variance and poor generalization comparatively to the dual-ANN PINN and hybrid semiparametric models. These findings are supported by the recent work of [Velioglu et al. \(2025\)](#), who proposed a differential-algebraic equation-based heuristic to determine whether a PINN can estimate hidden states in partially known systems. Their results demonstrated the improved extrapolation and state reconstruction capabilities of PINNs compared to conventional neural networks, even in the absence of full constitutive equations. This aligns with the present study where the dual-ANN PINN has accurately extrapolated beyond the training domain for unobserved states such as the case of unmeasured volume. However, this extrapolative advantage is not guaranteed in more complex systems as highlighted in case study 2.

Recent methodological contributions further insight into PINN limitations. [Subramanian et al. \(2023\)](#) proposed adaptive collocation schemes that target high-error regions to improve physics learning and physics extrapolation. While the present study did not adopt adaptive resampling, a stochastic collocation strategy was employed that regenerated collocation points at each training epoch across the bounded state space. This approach proved effective for enforcing broad physical coverage and stabilizing the physics loss. To avoid overfitting, Gaussian Process (GP) smoothening strategies were suggested by [Bajaj et al. \(2023\)](#). This study followed conventional approaches, for instance, case study 1 employed cross-validation, while case study 2 used stochastic regularization via minibatch size data resampling. Furthermore, the present study relied on standard automatic differentiation, aligned with conventional PINN methods, but results from [Chiu et al. \(2022\)](#) suggested that hybrid differentiation strategies by coupled automatic-numerical differentiation method (Can-PINN) may enhance computational efficiency and accuracy in solving complex physics systems, an avenue worth exploring in future work.

Overall, this study highlights the nuanced capabilities of each modelling framework. Hybrid semiparametric models offer high accuracy and stability within well-defined process regimes. This is particularly the case when the backbone of the model is precisely defined by a mechanistic equation, such as Eqs. (1a,b) for bioreactor problems. A key advantage is that the extrapolation of physics is not endangered by training limitations. The requirement of numerical integration may, however, constitute a limitation, particularly for stiff ODE systems. PINNs also have high potential to learn from data and physics simultaneously and may be more flexible for problems with partially known mechanistic knowledge. They show a clear advantage over conventional ANNs, but their multi-objective training is far more complex. While PINNs are not affected by the stiffness issues associated with ODEs in hybrid semiparametric models, their main drawback lies in the extrapolation of prior physical equations. This extrapolation is highly sensitive to the training methodology used, which can significantly impact the reliability of PINN predictions. The choice between these paradigms should thus be dictated by the modelling objective whether high-fidelity interpolation or robust extrapolation is required, by the nature of available process knowledge and encountered numeric implementation challenges.

6. Conclusion

This study compared a dual-ANN PINN, hybrid semiparametric, and conventional ANN structures for generic bioreactor modelling across 2 case studies. A key aspect of the dual-ANN PINN was the decoupling of dynamic state variables and static reaction kinetics parameterization, which consistently led to enhanced generalization capacity. In line with previous studies, our results highlight that PINNs, through the incorporation of physics in the loss function, exhibit stronger extrapolation capabilities than conventional ANNs, particularly in scenarios of high data sparsity. Hybrid semiparametric models, which embed the physics equations directly into the model structure, generally offer better predictive accuracy and more stable convergence than the dual-ANN PINN approach, particularly for high-dimensional nonlinear dynamical

systems with time-varying control inputs. The findings highlight that, although PINN architectures offer a more flexible way to incorporate partial prior knowledge and data, their performance is quite sensitive to system complexity, hyperparameter selection, and potential training instabilities. Since they do not require the integration of ODEs they are well-suited for process control applications that require fast model simulations. Hybrid semiparametric models, in contrast, inherit all the numerical challenges associated with the mechanistic equations. This can result in infeasible or stiff systems that require significantly longer simulation times and slower training, ultimately limiting practical applicability. The dual-ANN PINN framework struggled with time-varying and batch-varying control inputs (feed rates) due to physics-based extrapolation errors beyond the training domain. In contrast, explicitly incorporating process control inputs into the mechanistic component of the hybrid semiparametric model can lead to more robust process control designs and represents a significant advantage of this approach. The choice between the two hybrid modeling methods ultimately depends on carefully weighing their respective advantages and disadvantages for the specific bioreactor problem at hand.

CRediT authorship contribution statement

Monesh kumar Thirugnanasambandam: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **José Pinto:** Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization. **Ekaterina Moskovkina:** Writing – review & editing, Investigation. **Rafael S. Costa:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Rui Oliveira:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

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 is financed by national funds from FCT - Fundação para a Ciência e a Tecnologia, I.P., in the scope of the project UIDP/04378/2020 (DOI: 10.54499/UIDP/04378/2020) and UIDB/04378/2020 (DOI: 10.54499/UIDB/04378/2020) of the Research Unit on Applied Molecular Biosciences - UCIBIO and the project LA/P/0140/2020 (DOI: 10.54499/LA/P/0140/2020) of the Associate Laboratory Institute for Health and Bioeconomy - i4HB. This work received funding from the European Union's Horizon 2020 research and innovation program under grant agreement no.101099487-BioLaMer-HORIZON-EIC-2022-PATHFINDEROPEN-01 (BioLaMer).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.compchemeng.2025.109354](https://doi.org/10.1016/j.compchemeng.2025.109354).

Data availability

Data will be made available on request.

References

- Adebar, N., Arnold, S., Herrera, L.M., Emenike, V.N., Wucherpfennig, T., Smiatek, J., 2024. Physics-informed neural networks for biopharmaceutical cultivation processes: consideration of varying process parameter settings. *Biotechnol. Bioeng.* <https://doi.org/10.1002/bit.28851>.

- Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M., 2019. Optuna: a next-generation hyperparameter optimization framework. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 2623–2631. <https://doi.org/10.1145/3292500.3330701>.
- Bajaj, C., McLennan, L., Andeen, T., Roy, A., 2023. Recipes for when physics fails: recovering robust learning of physics informed neural networks. *Mach. Learn. Sci. Technol.* 4 (1), 015013. <https://doi.org/10.1088/2632-2153/acb416>.
- Banki, M.S.F., Kao, K., Kwon, J.S., 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. <https://doi.org/10.1016/j.cherd.2022.01.041>.
- Baydin, A.G., Pearlmutter, B.A., Radul, A.A., Siskind, J.M., 2018. Automatic differentiation in machine learning: a survey. *J. Mach. Learn. Res.* 18 (153), 1–43. <http://jmlr.org/papers/v18/17-468.html>.
- Bayer, B., Duerkop, M., Pörtner, R., Möller, J., 2023. Comparison of mechanistic and hybrid modeling approaches for characterization of a CHO cultivation process: requirements, pitfalls and solution paths. *Biotechnol. J.* 18 (1), 2200381. <https://doi.org/10.1002/biot.202200381>.
- Chiu, P.H., Wong, J.C., Ooi, C., Dao, M.H., Ong, Y.S., 2022. CAN-PINN: a fast physics-informed neural network based on coupled-automatic-numerical differentiation method. *Comput. Methods Appl. Mech. Eng.* 395, 114909. <https://doi.org/10.1016/j.cma.2022.114909>.
- Cruz-Bourazou, M.N., Narayanan, H., Fagnani, A., Butté, A., 2022. Hybrid Gaussian process models for continuous time series in fed-batch cultures. *IFAC-Papers OnLine* 55 (7), 204–209. <https://linkinghub.elsevier.com/retrieve/pii/S2405896322008461>.
- Cui, T., Bertalan, T., Ndahiro, N., Khare, P., Betenbaugh, M., Maranas, C., Kevrekidis, I. G., 2024. Data-driven and physics-informed modeling of Chinese Hamster ovary cell bioreactors. *Comput. Chem. Eng.* 183, 108594. <https://doi.org/10.1016/j.compchemeng.2024.108594>.
- Glassey, J., Gerena, K.V., Clemens, C., 2011. Process analytical technology (PAT) for biopharmaceuticals. *Biotechnol. J.* 6 (4), 369–377. <https://doi.org/10.1002/biot.201000356>.
- Hellekes, L.M., Hemmerich, J., Wiechert, W., von Lieres, E., Grünberger, A., 2023. Machine learning in bioprocess development: from promise to practice. *Trends Biotechnol.* 41 (6), 817–835. <https://doi.org/10.1016/j.tibtech.2022.10.010>.
- Hong, M.S., Braatz, R.D., 2021. Mechanistic modeling and parameter-adaptive nonlinear model predictive control of a microbioreactor. *Comput. Chem. Eng.* 147, 107255. <https://doi.org/10.1016/j.compchemeng.2021.107255>.
- Jul-Rasmussen, P., Kumar, M., Pinto, J., Oliveira, R., Liang, X., Huusom, J.K., 2025. Incorporating first-principles information into hybrid modeling structures: comparing hybrid semi-parametric models with physics-informed Recurrent Neural Networks. *Comput. Chem. Eng.*, 109119 <https://doi.org/10.1016/j.compchemeng.2025.109119>.
- Kadlec, P., Gabrys, B., Strandt, S., 2009. Data-driven soft sensors in the process industry. *Comput. Chem. Eng.* 33 (4), 795–814. <https://doi.org/10.1016/j.compchemeng.2008.12.012>.
- Kingma, D.P. and Ba, J., 2014. Adam: a method for stochastic optimization. CoRR, abs/1412.6980.
- Kipf, T.N., Fetaya, E., Wang, K.C., Welling, M., Zemel, R., 2018. Neural relational inference for interacting systems. [10.48550/arXiv.1802.04687](https://arxiv.org/abs/1802.04687).
- Li, S., Feng, X., 2022. Dynamic weight strategy of physics-informed neural networks for the 2D navier-stokes equations. *Entropy* 24 (9), 1254. <https://doi.org/10.3390/e24091254>. PMID: 36141140; PMCID: PMC9497516.
- Li, Q., Zhang, J., Wan, H., Zhao, Z., Liu, F., 2024. Physics-informed neural networks for multi-stage Koopman modeling of microbial fermentation processes. *J. Process Control* 143, 103315. <https://doi.org/10.1016/j.jprocont.2024.103315>.
- Moayed, F., Chandrasekar, A., Rasmussen, S., Sarna, S., Corbett, B., Mhaskar, P., 2024. Physics-informed neural networks for process systems: handling plant-model mismatch. *Ind. Eng. Chem. Res.* 63 (31), 13650–13659. <https://doi.org/10.1021/acs.iecr.4c00690>.
- Narayanan, H., von Stosch, M., Feidl, F., Sokolov, M., Morbidelli, M., Butté, A., 2023. Hybrid modeling for pharmaceutical processes: advantages, opportunities, and implementation. *Front. Chem. Eng.* 5, 1157889. <https://doi.org/10.3389/fceng.2023.1157889>.
- Oliveira, R., 2004. Combining first principles modelling and artificial neural networks: a general framework. *Comput. Chem. Eng.* 28 (5), 755–766. <https://doi.org/10.1016/j.compchemeng.2004.02.014>.
- Park, S., Ramirez, W.F., 1988. Optimal production of secreted protein in fed-batch reactors. *AIChE J.* 34 (9), 1559–1568. <https://doi.org/10.1002/aic.690340917>.
- Pinto, J., Mestre, M., Ramos, J., Costa, R.S., Striedner, G., Oliveira, R., 2022. A general deep hybrid model for bioreactor systems: combining first principles with deep neural networks. *Comput. Chem. Eng.* 165, 107952. <https://doi.org/10.1016/j.compchemeng.2022.107952>.
- Psychogios, D.C., Ungar, L.H., 1992. A hybrid neural network-first principles approach to process modeling. *AIChE J.* 38 (10), 1499–1511. <https://doi.org/10.1002/ai.690381003>.
- Raiassi, M., Perdikaris, P., Karniadakis, G.E., 2019. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* 378, 686–707. <https://doi.org/10.1016/j.jcp.2018.10.045>.
- Ramos, J.R., Pinto, J., Poiares-Oliveira, G., Peeters, L., Dumas, P., Oliveira, R., 2024. Deep hybrid modeling of a HEK293 process: combining long short-term memory networks with first principles equations. *Biotechnol. Bioeng.* 121 (5), 1554–1568. <https://doi.org/10.1002/bit.28668>.
- Schubert, J., Simutis, R., Dors, M., Havlík, I., Lübbert, A., 1994. Hybrid modelling of yeast production processes—combination of a priori knowledge on different levels of sophistication. *Chem. Eng. Technol.* 17, 10–20. <https://doi.org/10.1002/ceat.270170103>.
- Smiatek, J., Jung, A., Bluhmki, E., 2020. Towards a digital bioprocess replica: computational approaches in biopharmaceutical development and manufacturing. *Trends Biotechnol.* 38 (10), 1141–1153. <https://doi.org/10.1016/j.tibtech.2020.05.008>.
- Subramanian, S., Kirby, R.M., Mahoney, M.W. and Gholami, A., 2023. Adaptive self-supervision algorithms for physics-informed neural networks. [10.3233/FAIA230521](https://arxiv.org/abs/2303.13233).
- Teixeira, Ana P., Carinhas, N., Dias, J.M.L., Cruz, P., Alves, P.M., Carrondo, M.J.T., Oliveira, R., 2007. Hybrid semi-parametric mathematical systems: bridging the gap between systems biology and process engineering. *J. Biotechnol.* 132 (4), 418–425. <https://doi.org/10.1016/j.biote.2007.08.020>.
- Thompson, M.L., Kramer, M.A., 1994. Modeling chemical processes using prior knowledge and neural networks. *AIChE J.* 40. <https://doi.org/10.1002/aic.690400806>.
- Van Impe, J.F., Vanderleyden, J., Versyck, K.J., et al., 2013. Quantitative analysis of microbial metabolism: probing and modeling as driving forces in biotechnology. *Biotechnol. Adv.* 31 (1), 1–19. <https://doi.org/10.1016/j.foodcont.2012.06.019>.
- Velioglu, M., Zhai, S., Rupprecht, S., Mitsos, A., Jupke, A., Dahmen, M., 2025. Physics-informed neural networks for dynamic process operations with limited physical knowledge and data. *Comput. Chem. Eng.* 192, 108899. <https://doi.org/10.1016/j.compchemeng.2024.108899>.
- Von Stosch, M., Davy, S., Francois, K., Galvanauskas, V., Hamelink, J.M., Luebbert, A., Mayer, M., Oliveira, R., O'Kennedy, R., Rice, P., Glassey, J., 2014a. Hybrid modeling for quality by design and PAT—benefits and challenges of applications in pharmaceutical industry. *Biotechnol. J.* 9 (6), 719–726. <https://doi.org/10.1002/biot.201300385>.
- Von Stosch, M., Oliveira, R., Peres, J., de Azevedo, S.F., 2014b. Hybrid semi-parametric modeling in process systems engineering: past, present and future. *Comput. Chem. Eng.* 60, 86–101. <https://doi.org/10.1016/j.compchemeng.2013.08.008>.
- Wang, S., Teng, Y., Perdikaris, P., 2021. Understanding and mitigating gradient flow pathologies in physics-informed neural networks. *SIAM J. Sci. Comput.* 43 (5), A3055–A3081. <https://doi.org/10.1137/20M1318043>.
- Wang, S., Yu, X., Perdikaris, P., 2022. When and why PINNs fail to train: a neural tangent kernel perspective. *J. Comput. Phys.* 449, 110768. <https://doi.org/10.1016/j.jcp.2021.110768>.
- Yang, S., Fahey, W., Truccollo, B., Browning, J., Kamyar, R., Cao, H., 2024. Hybrid modeling of fed-batch cell culture using physics-informed neural network. *Ind. Eng. Chem. Res.* <https://doi.org/10.1021/acs.iecr.4c01459>.