Antonis Kokossis, Michael C. Georgiadis, Efstratios N. Pistikopoulos (Eds.)
PROCEEDINGS OF THE 33rd European Symposium on Computer Aided Process Engineering (ESCAPE33), June 18-21, 2023, Athens, Greece
 http://dx.doi.org/10.1016/B978-0-443-15274-0.50014-7

Investigating physics-informed neural networks for bioprocess hybrid model construction

Alexander William Rogers,[a] Ilya Orson Sandoval Cardenas,[b] Ehecatl Antonio Del Rio-Chanona,[b] Dongda Zhang,[a*]

[a] *Department of Chemical Engineering, University of Manchester, Oxford Road, Manchester, M1 3AL, UK.*
[b] *Department of Chemical Engineering, Imperial College London, South Kensington Campus, London, SW7 2AZ, UK.*
[*] *Corresponding author email: dongda.zhang@manchester.ac.uk*

Abstract

Integrating physical knowledge and machine learning is a cost-efficient solution to modelling complex biochemical processes when the underlying mechanisms are not fully understood. However, hybrid model structure identification is still time-consuming for new processes, requiring iteration over different hypotheses to explain the observed process dynamics while minimizing over-parameterization. Unfortunately, conventional approaches to automatic model structure identification do not always converge for highly nonlinear models and cannot estimate time-varying model parameters. To address this and accelerate the design of new biochemical processes, a Reinforcement Learning (RL) based framework recently reformulated synchronous hybrid model structure-parameter identification into a process optimal control problem. To further investigate other possible solutions, in this study, a novel Physics Informed Neural Network (PINN) based framework was proposed for the first time to infer time-varying kinetic parameters. This framework first combines possible kinetic structures from phenomenological knowledge, then simultaneously identifies the most likely hybrid model structure and time-varying parameter trajectories. To demonstrate the performance of the PINN based framework, several in-silico case studies were conducted using a known ground truth bioprocess. We thoroughly examined the advantages and limitations of the framework, elucidating its potential for high-fidelity hybrid model construction in biochemical engineering research.

Keywords: automatic model structure identification, time-varying parameter estimation, physics-informed neural network, hybrid modelling, machine learning.

1. Introduction

Mathematical modelling is pivotal to understanding and designing biochemical processes. A thoroughly validated dynamic model can predict biomass growth and product synthesis under different operating conditions, reducing the number of experiments required to characterize and optimize novel biochemical processes. Kinetic and data-driven models have been proposed to describe biological processes. However, identifying a suitable kinetic model structure is time-consuming, and they are often overparameterized (i.e., multiple structures and many parameters) in an effort to capture the complex dynamics, leading to high parameter and propagated state uncertainties. On the other hand, data-driven models risk overfitting without enough experimental data – which is time-consuming to generate – and poor generalization to new operating conditions.

Integrating physical knowledge and machine learning is a cost-efficient solution to modelling complex biochemical processes when the underlying mechanisms are not fully understood. Hybrid models either use a data-driven model to correct the discrepancy between a kinetic model and the observed process dynamics or update selected time-varying kinetic model parameters. A properly validated hybrid model can effectively resolve the issue of incomplete physical knowledge and low-quality-quantity data (Zhang et al., 2020) while improving prediction accuracy and confidence compared with pure kinetic models (Vega-Ramon et al., 2021). However, hybrid models inherit the risk of over-parameterizing the kinetics and overfitting the data-driven model, which can lead to high uncertainty and poor generalization. Consequently, it is essential to identify the kinetic model structure that best represents the underlying mechanisms to reduce the burden on the data-driven model to compensate and risk overfitting. This is no trivial task since the current bioprocess kinetics depend on the present culture conditions and microenvironmental history due to stochastic effects on the controlling mechanisms or metabolic stores and systematic intracellular metabolic regulation mechanisms. As a result, the kinetic model parameters and structure will evolve with time and history.

Despite this challenge, there are few attempts outside case-specific studies to resolve automatic model structure identification for complex combinatorial, history-dependent or time-varying kinetics. Conventional approaches such as mixed integer nonlinear programming (MINLP) do not always converge for highly nonlinear models and cannot estimate time-varying model parameters. To address this, we recently proposed a novel Reinforcement Learning (RL) based framework to reformulate synchronous hybrid model structure-parameter identification into a process optimal control problem (Wu et al., 2022). The RL-based framework proved promising for recovering the correct kinetic model structure or time-varying kinetic parameters, but the combined case with more combinatorial options and time-varying kinetic parameters remains unexplored. In recent years, Physics-Informed Neural Networks (PINNs) (Raissi et al., 2019) have emerged as a novel approach to discovering underlying governing equations; however, they have not been applied before to infer time-varying physical parameters. Therefore, in this work we aim to propose a novel PINN-based framework for hybrid bioprocess model construction. Several in-silico case studies will thoroughly examine the advantages and limitations of this technique for synchronous hybrid model structure-parameter identification.

2. Problem Statement

The PINN-based framework first combines several possible kinetic structures from phenomenological knowledge, then simultaneously identifies the most likely hybrid model structure and time-varying parameter trajectories. To demonstrate the performance of this framework, several in-silico case studies were conducted using a known ground truth model, shown in Equation 1, that we developed in our previous work (Rogers et al., 2022). This high-fidelity hybrid model can predict the temperature-dependent biomass growth, glucose consumption and γ-linolenic acid (GLA) accumulation rates during fermentation of the fungus *Cunninghamella echinulata*. In Equation 1, X, F, S and P are the total biomass, fat-free biomass, glucose and GLA concentration, respectively, while T is temperature. The half-saturation K_S and maintenance k_S coefficients, the total k_X and fat-free k_F biomass decay rates and specific GLA decay rate k_P are all constants, while the total Y_X and fat-free Y_F biomass growth rates, the glucose-to-biomass yield coefficient Y_S and growth-dependent GLA yield coefficient Y_P are all time-varying parameters. In

our previous work, we estimated the constants and built the Gaussian Process (GP) model mapping the states and temperature to the time-varying parameters (Rogers et al., 2022).

$$\frac{\mathrm{d}X}{\mathrm{d}t} = Y_X(X,F,S,P,T) \cdot \frac{S}{S + K_S \cdot X} \cdot X - k_X \cdot X \tag{1a}$$

$$\frac{\mathrm{d}F}{\mathrm{d}t} = Y_F(X,F,S,P,T) \cdot \frac{S}{S + K_S \cdot X} \cdot X - k_F \cdot X \tag{1b}$$

$$\frac{\mathrm{d}S}{\mathrm{d}t} = Y_S(X,F,S,P,T) \cdot \frac{S}{S + K_S \cdot X} \cdot X - k_S \cdot X \cdot \frac{S}{S + 0.1} \tag{1c}$$

$$\frac{\mathrm{d}P}{\mathrm{d}t} = Y_P(X,F,S,P,T) \cdot \frac{S}{S + K_S \cdot X} \cdot X - k_P \cdot X \tag{1d}$$

Now the ground-truth model was used to generate six in-silico batches from low ($S_0 = 60$ g L^{-1}) and high ($S_0 = 100$ g L^{-1}) initial glucose concentrations under three different temperatures ($T = 14$℃, $T = 28$℃, $T = 37$℃). The four GP models updated the four time-varying parameters every 24 hours while the constants were fixed. Each batch was 'fermented' for 408 hours, and the four state variables 'measured' every 24 hours.

3. Methodology

3.1. Physics-Informed Neural Network Structure

The PINN aims to learn the solution to the system of ordinary differential equations (ODEs) presented in Equation 1 by reformulating the parameter estimation and numerical integration problem into a single nonlinear optimization problem. The ANN aims to learn the concentration profile and kinetic parameters that satisfy the measured states and the system of ODEs of the form: $d\boldsymbol{x}/dt = \mathcal{N}(\boldsymbol{x}, \boldsymbol{\phi}, \boldsymbol{\theta})$, where $\boldsymbol{x} = [X, F, S, P]^T$, while $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ are vectors of time-varying and time-constant parameters, respectively. The unknown solution to the concentration profile $\boldsymbol{u_x}(\boldsymbol{\lambda}, t, \boldsymbol{x}_0)$ and kinetic parameters $\boldsymbol{u_\phi}(\boldsymbol{\lambda}, t, \boldsymbol{x}_0)$ was represented by a single ANN as a function of batch time $t \in [0, 408\ \text{hours}]$ and the initial state $\boldsymbol{x}_0 = [X_0, F_0, S_0, P_0]^T$, where $\boldsymbol{\lambda}$ denotes the tunable weights and biases.

$$\boldsymbol{\lambda}^* = \min_{\boldsymbol{\lambda}} \mathcal{L}(\boldsymbol{\lambda}) = \min_{\boldsymbol{\lambda}} [\omega_0 \mathcal{L}_0(\boldsymbol{\lambda}) + \omega_d \mathcal{L}_d(\boldsymbol{\lambda}) + \omega_c \mathcal{L}_c(\boldsymbol{\lambda})] \tag{2a}$$

$$\mathcal{L}_0(\boldsymbol{\lambda}) = \frac{1}{N_0} \sum_{i=1}^{N_0} \left| \widehat{\boldsymbol{x}}_0^i - \boldsymbol{x}_0^i \right|^2 \tag{2b}$$

$$\mathcal{L}_d(\boldsymbol{\lambda}) = \frac{1}{N_d} \sum_{i=1}^{N_d} \left| \widehat{\boldsymbol{x}}_d^i - \boldsymbol{x}_d^i \right|^2 \tag{2c}$$

$$\mathcal{L}_c(\boldsymbol{\lambda}) = \frac{1}{N_c} \sum_{i=1}^{N_c} \left| \frac{\partial \widehat{\boldsymbol{x}}_c^i}{\partial t} - \mathcal{N}\left(\widehat{\boldsymbol{x}}_c^i, \widehat{\boldsymbol{\phi}}_d^i, \widehat{\boldsymbol{\theta}}\right) \right|^2 \tag{2d}$$

The PINN was trained by minimizing the composite loss function in Equation 2, where ω_0, ω_d and ω_c balance the interplay between the different loss terms during training. Here $\{\boldsymbol{x}_0^i\}_{i=1}^{N_0}$, $\{t_d^i, \boldsymbol{x}_d^i\}_{i=1}^{N_d}$ and $\{t_c^i, \boldsymbol{x}_c^i\}_{i=1}^{N_c}$ are the time-state value pairs for the initial conditions,

measured data points and collocation points, respectively. Each batch is divided into N intervals of $\Delta t = 24$ hours, where $t_d \in \{0,\ \Delta t,\ 2\Delta t,\ 3\Delta t \ldots N\Delta t\}$, while each interval is subdivided into $\Delta t/\delta t = 5$ collocation points, such that $t_c \in \{0,\ \delta t,\ 2\delta t,\ 3\delta t \ldots n\delta t\}$ is on the same timescale. The terms $\widehat{\boldsymbol{x}}_d = \boldsymbol{u}_{\boldsymbol{x}}(\boldsymbol{\lambda}, t_d, \boldsymbol{x}_0)$ and $\boldsymbol{x}_d$ denote the predicted and measured states, respectively, at the known data points, while $\widehat{\boldsymbol{x}}_c = \boldsymbol{u}_{\boldsymbol{x}}(\boldsymbol{\lambda}, t_c, \boldsymbol{x}_0)$ denotes the predicted state at the collocation points for which the states were not measured. Once $\widehat{\boldsymbol{x}}_c$ is evaluated, the gradients with respect to time $\partial\widehat{\boldsymbol{x}}_c/dt$ were computed by reverse automatic differentiation. The time-constant $\widehat{\boldsymbol{\theta}} = \boldsymbol{u}_{\boldsymbol{\phi}}(\boldsymbol{\lambda}, 0,0)$ and time-varying $\widehat{\boldsymbol{\phi}}_d = \boldsymbol{u}_{\boldsymbol{\phi}}(\boldsymbol{\lambda}, t_d, \boldsymbol{x}_0)$ kinetic parameters were both predicted by the ANN, implicitly coupling them to the same network parameters for more stable training. However, unlike the continuous state variables, $\widehat{\boldsymbol{\theta}}$ is a fixed constant, while $\widehat{\boldsymbol{\phi}}$ is a piecewise constant. To represent this behaviour, $\widehat{\boldsymbol{\theta}}$ was evaluated at a fixed input while $\widehat{\boldsymbol{\phi}}_d$ was evaluated at each data sampling point t_d and the result used over $t_d \le t_c < t_d + \Delta t$ when computing $\mathcal{N}(\widehat{\boldsymbol{x}}_c, \widehat{\boldsymbol{\phi}}_d, \widehat{\boldsymbol{\theta}})$. The network was built from a single hidden layer of 40 neurons with a hyperbolic tangent activation function in the hidden layer and a linear output layer. This combination mirrors the typical bioprocesses sigmoid-like growth and parameter profiles.

3.2. Physics-Informed Neural Network Training and Simulation

The PINN was fitted simultaneously to all six in-silico batches. The PINN was trained in two stages: first (i) the network was trained for 6000 epochs with $\{\omega_0 = 5, \omega_d = 1, \omega_c = 0\}$ and a learning rate of 5×10^{-2} until $\mathcal{L}_d(\boldsymbol{\lambda})$ converged, then (ii) the network was trained for 10000 epochs with $\{\omega_0 = 5, \omega_d = 1, \omega_c = 5 \times 10^3\}$ and a learning rate of 5×10^{-4} until $\mathcal{L}_c(\boldsymbol{\lambda})$ converged. Fitting the state profile before the kinetic parameters approximates the initial gradients $\partial\widehat{\boldsymbol{x}}_c/dt$ and was found empirically to be more robust to becoming trapped within low-quality local optima than single-stage training.

Once the PINN was constructed, the multistep-ahead state trajectory was re-simulated using Equation 3 from $\boldsymbol{x}_0 = [X_0, F_0, S_0, P_0]^T$ using the predicted time-constant parameters $\widehat{\boldsymbol{\theta}}^* = \boldsymbol{u}_{\boldsymbol{\phi}}(\boldsymbol{\lambda}^*, 0,0)$ and the predicted time-varying parameters $\widehat{\boldsymbol{\phi}}_d^* = \boldsymbol{u}_{\boldsymbol{\phi}}(\boldsymbol{\lambda}^*, t_d, \boldsymbol{x}_0)$ updated once every 24 hours.

$$\widehat{\boldsymbol{x}}'_{d+1} = \widehat{\boldsymbol{x}}'_d + \int_{t_d}^{t_{d+1}} \mathcal{N}(\boldsymbol{x}', \widehat{\boldsymbol{\phi}}_d^*, \widehat{\boldsymbol{\theta}}^*)\, dt \tag{3}$$

4. Results and Discussion

Once the PINN was fitted to the six in-silico batches, the fitted state $\widehat{\boldsymbol{x}}_d^* = \boldsymbol{u}_{\boldsymbol{x}}(\boldsymbol{\lambda}^*, t_d, \boldsymbol{x}_0)$ trajectory, the time-constant parameters $\widehat{\boldsymbol{\theta}}^* = \boldsymbol{u}_{\boldsymbol{\phi}}(\boldsymbol{\lambda}^*, 0,0)$ and the time-varying kinetic parameter $\widehat{\boldsymbol{\phi}}_d^* = \boldsymbol{u}_{\boldsymbol{\phi}}(\boldsymbol{\lambda}^*, t_d, \boldsymbol{x}_0)$ trajectory was retrieved. In addition, the fitted PINN was used to re-simulate the multistep-ahead state trajectory $\widehat{\boldsymbol{x}}'_d$ for the six in-silico batches from their different initial states and operating temperatures, as described in Section 3.2. Figure 1 compares the fitted (i.e., $\widehat{\boldsymbol{x}}_d^*$) and the re-simulated (i.e., $\widehat{\boldsymbol{x}}'_d$) state trajectories for one of the six in-silico batches, showing that the PINN could fit and re-simulate the state trajectory with very high accuracy, with a mean absolute percentage error (MAPE) of 1% and 5%, respectively. Since $\boldsymbol{x}_0$ is satisfied (MAPE of 1%) by strong regularization (i.e., $\omega_0 > \omega_d$) the mismatch between the fitted and re-simulated trajectories is due to the PINN not being strictly required to satisfy the system of ODEs during PINN training.

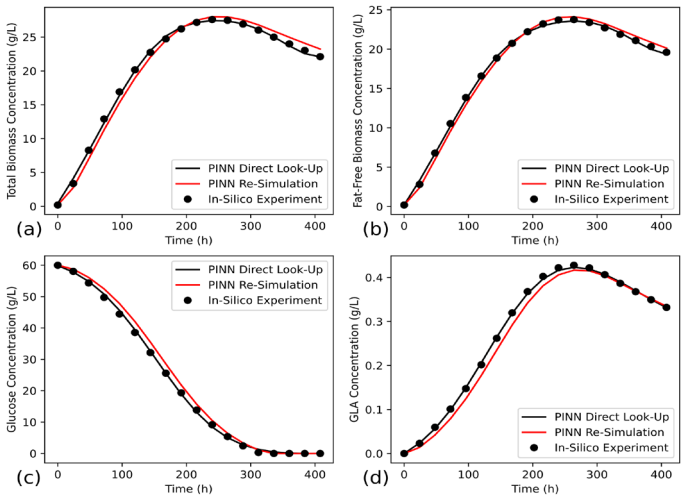

Figure 1: Fitted and re-simulated state trajectory for one of the six in-silico training batches for total biomass (a), fat-free biomass (b), glucose (c) and GLA (d) concentration.

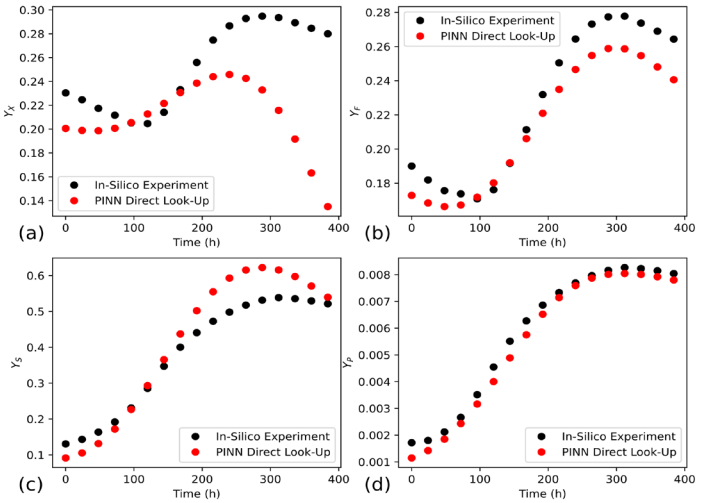

Figure 2: Ground truth and estimated time-varying parameters for one of the six in-silico batches for the specific growth rate Y_X (a) and Y_F (b), Y_S (c) and Y_P (d) yield coefficients.

Table 1 compares the PINN estimated and ground-truth constants, demonstrating that the PINN can accurately recover (MAPE of 8%) the correct constants when there is no measurement uncertainty. The residual between the ground truth and estimated values can likely be attributed to the practical identifiability of the parameters given only six in-silico batches. Figure 2 compares the PINN estimated and ground-truth time-varying parameter trajectories, demonstrating that the PINN can recover Y_F, Y_S and Y_P well (MAPE of 15%), particularly considering the time-varying parameters (i.e., Y_X, Y_F, Y_S and Y_P) become non-identifiable as $S \to 0$ at the end of the batch, as their influence becomes negligible compared to the specific decay constants (i.e., k_X, k_F, k_S and k_P). However, the PINN could not recover well Y_X (MAPE of 150%). Given the excellent fit in Figure 1,

there could be two reasons for this: either (i) the PINN cannot converge to the globally optimum ground-truth parameters, and the model is insensitive to the exact parameter values, or (ii) K_S is partially non-identifiable, and extra regularization is required.

Table 1: Comparison of ground-truth and estimated time-constant parameters.

Parameter	K_S	k_X	k_F	k_S	k_P
Ground-Truth	48.64	1.70E-03	1.35E-03	3.10E-03	3.23E-05
Estimated	49.60	1.50E-03	1.32E-03	2.57E-03	2.96E-05
Absolute Error (%)	2.0	11.8	2.2	17.1	8.4

5. Conclusion

The PINN-based approach has potential for kinetic parameter estimation and for providing a way to understand real bioprocess underlying system dynamics. At present, embedded hybrid model construction follows a two-step procedure: (i) simultaneous time-constant and time-varying parameter estimation, then (ii) data-driven model construction to correlate the time-varying parameters with the state and operating conditions for predictive simulation. This novel PINN-based framework has the potential to accelerate hybrid model construction by directly building the predictive data-driven model. However, there remains the challenge of identifying a high-quality solution to the PINN network parameters. Therefore, future work will explore more advanced techniques, such as adaptive time-sampling, self-attenuation, time-marching and causal training (Wang et al., 2022). In addition, the PINN-based framework is currently only employed to estimate the kinetic parameters for a pre-defined kinetic model structure. Therefore, future work will also explore synchronous parameter-structure identification and compare the performance of the PINN-based framework against an extended RL-based framework (Wu et al., 2022) when multiple possible structures are possible.

References

Raissi, 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. *Journal of Computational Physics*, *378*, 686–707.

Rogers, A. W., Song, Z., Vega-Ramon, F., Jing, K., & Zhang, D. (2022). Investigating 'greyness' of hybrid model for bioprocess predictive modelling. *Biochemical Engineering Journal*.

Vega-Ramon, F., Zhu, X., Savage, T. R., Petsagkourakis, P., Jing, K., & Zhang, D. (2021). Kinetic and hybrid modeling for yeast astaxanthin production under uncertainty. *Biotechnology and Bioengineering*, *118*(12), 4854–4866.

Wang, S., Sankaran, S., & Perdikaris, P. (2022). *Respecting causality is all you need for training physics-informed neural networks*.

Wu, C., Mowbray, M. R., Rogers, A. W., Rio-Chanona, E. A. Del, & Zhang, D. (2022). *A reinforcement learning based hybrid modelling framework for bioprocess kinetics identification*.

Zhang, D., Savage, T. R., & Cho, B. A. (2020). Combining model structure identification and hybrid modelling for photo-production process predictive simulation and optimisation. *Biotechnology and Bioengineering*, *117*(11), 3356–3367.