Systems & Control Transactions

Research Article - Peer Reviewed Conference Proceeding
ESCAPE 35 - European Symposium on Computer Aided Process Engineering
Ghent, Belgium. 6-9 July 2025
Jan F.M. Van Impe, Grégoire Léonard, Satyajeet S. Bhonsale,
Monika E. Polańska, Filip Logist (Eds.)

PSE PRESS

AI-Driven Automatic Mechanistic Model Transfer Learning for Accelerating Process Development

Alexander W. Rogers[a], Amanda Lane[b], Philip Martin[a], and Dongda Zhang[a]*

[a] The University of Manchester, Department of Chemical Engineering, Manchester, UK
[b] Unilever R&D Port Sunlight, Liverpool, UK
* Corresponding Author: dongda.zhang@manchester.ac.uk.

ABSTRACT

Accurate mechanistic models provide valuable physical insight and are crucial for efficient process scale-up and optimisation, but their identification requires lengthy experimental data collection, model construction, validation and discrimination. Traditional black-box machine learning transfer methods leverage prior knowledge but lack interpretability and physical insights. To address this, we propose a novel approach using artificial neural network feature attribution to automatically locate corrections and symbolic regression to make structural modifications to an inaccurate or low-fidelity mechanistic model. In a comprehensive in-silico case study, the framework adapted a kinetic model from one biochemical system to a different but related one, enhancing predictive accuracy. Integrated within an iterative model-based design of experiments routine, it minimised the number of new experiments required. The study also discusses the impact of the inductive bias trade-off and alternative ways of incorporating prior knowledge to improve performance, highlighting the framework's flexibility and potential for advancing automated knowledge discovery and high-fidelity predictive digital twin design for novel (bio)chemical processes.

Keywords: Knowledge Discovery, Dynamic Modelling, Genetic Algorithm, Artificial Intelligence, Biosystems, Model-Based Design of Experiments, Interpretable Machine Learning

1. INTRODUCTION

Developing accurate models to predict bioprocess dynamics is vital for commercial process design, scale-up, and control. Kinetic models, formulated as ordinary differential equations (ODEs), incorporate known reaction mechanisms and parameters with physical significance, thus making them interpretable. However, their construction is difficult due to incomplete and uncertain knowledge of biochemical reaction networks. Data-driven methods such as statistical and machine learning models (e.g., artificial neural networks, Gaussian processes, random forests) can capture complex dynamics without explicit mechanistic knowledge. Still, they lack enough interpretability and demand large, high-quality datasets that are experimentally resource-intensive.

Hybrid modelling has emerged as a pragmatic solution when systems are only partially understood and data is limited. Such models combine mass balances and simplified reaction terms with data-driven components to describe unknown kinetics as time-varying parameters. This corrects for structural mismatch and reflects biochemical systems' inherent transient or hysteretic nature.

Transfer learning accelerates model development by updating pre-trained models from related systems. It has proved effective in fields such as image classification, natural language processing, and, increasingly, chemical processes. In bioprocessing, knowledge of one microalgal strain is leveraged to reduce experimentation for mutant strains or the influence of plant hormones to accelerate screening during early bioprocess development while integrated transfer learning and hybrid modelling carry over mechanistic knowledge [1]. The assumption is that the new system's dynamics are sufficiently similar to the source. However, the degree of similarity is often uncertain, while these approaches remain non-interpretable: by only updating the data-driven and mechanistic parameters of an existing hybrid model, the mechanistic model structure remains unchanged, and updates to the embedded data-driven terms are opaque, providing little physical insight into the new system.

In recent years, the push for interpretable,

parsimonious models has led to algorithms that automatically discover governing equations. Sparse identification techniques have been used to elucidate stoichiometry and kinetics in bioprocesses [2]. However, constructing suitable libraries of candidate functions is not always feasible for complex systems. This has motivated symbolic regression (SR) [3], whereby closed-form expressions built from primitive operators (e.g., addition, multiplication) are generated by genetic programming, transformers, reinforcement learning or recurrent neural networks. SR has succeeded in discovering chemical and approximate biochemical mechanistic equations [4] but risks producing non-interpretable local approximations. SR is NP-hard, and when many features are possible, discovering accurate, parsimonious expressions is intractable.

Existing methodologies also demand sampling frequencies beyond typical experimental capabilities, while biomass, substrate, and product measurements may be noisy. This complicates gradient estimation for ODE discovery unless relatively computationally expensive numerical integration is embedded in the fitness function.

Researchers often utilise prior knowledge when exploring new systems. This work, therefore, proposes a framework integrating SR with artificial neural network (ANN) feature attribution to automatically adapt the structure of an existing mechanistic model from literature or a common concept. This will allow the transfer of physical knowledge between different but related systems to rapidly generate accurate predictive models with limited data – a common scenario in bioprocess development.

In Section 4.1, an in-silico case study will demonstrate the framework when prior knowledge is abundant but partially incorrect for the target domain. Moreover, the framework will be combined with a model-based design of experiments (MbDoE) for model discrimination to iteratively refine the model structure while minimising the number of experiments needed to discover the correct equations for the newly investigated system. Section 4.2 will then discuss how the amount of prior knowledge will impact the flexibility and efficiency of the framework.

2. METHODOLOGY

A kinetic model represents the rate of change, $\mathrm{d}\boldsymbol{y}/\mathrm{d}t$, in the concentrations of (bio)chemical species, $\boldsymbol{y}$, with respect to time, t, as a system of ODEs, $f(\boldsymbol{y}, \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ are the model parameters. The goal of the framework is to take a mechanistic model for a different but related system or a pre-existing low-fidelity mechanistic model and adapt or improve its structure to more accurately fit and make predictions on a new target system. The initial equations, or source model, are denoted by $f_s(\boldsymbol{y}, \boldsymbol{\theta}_s)$, and the adapted equations, or transfer model for the target domain, are denoted by $f_t(\boldsymbol{y}, \boldsymbol{\theta}_t)$.

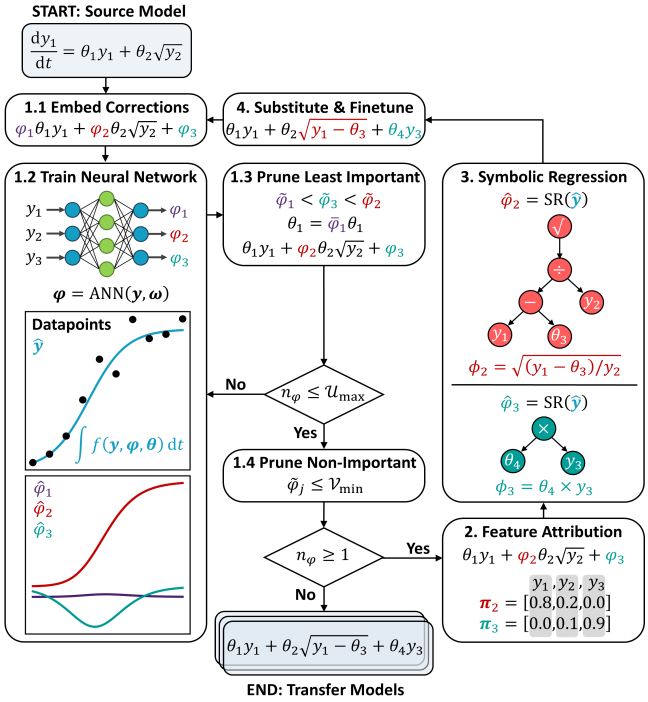

Figure 1: ANN attribution guided symbolic regression for dynamic systems. Takes equations for the source model and data for the target domain, and outputs candidate transfer models for model-based design of experiments.

An outline of the framework is depicted in Figure 1. Step 1 identifies which terms need altering by embedding initially unknown correction terms into the kinetic model (Step 1.1). A regularised ANN then learns the relationship between the correction terms and states (Step 1.2), then Steps 1.3-1.4 iteratively prune the corrections with the lowest variance. In Step 2, feature attribution determines which input variables matter most, providing weighting vectors, $\boldsymbol{\pi} = [\pi_1, \pi_2, \pi_3]$, for each feature, $\boldsymbol{y} = [y_1, y_2, y_3]$. Step 3 applies SR to build symbolic corrections guided by $\boldsymbol{\pi}$. Finally, in Step 4, these symbolic corrections are substituted back into the equations and parameters are fine-tuned, yielding $f_t(\boldsymbol{y}, \boldsymbol{\theta}_t)$. Variables $\boldsymbol{y}$ and t are normalised and the source model automatically nondimensionalised; afterwards $f_t(\boldsymbol{y}, \boldsymbol{\theta}_t)$ is automatically re-dimensionalised. Throughout this article, $\hat{\boldsymbol{y}}$ denotes fitted states, $\hat{\boldsymbol{\varphi}}$ fitted corrections, and $\boldsymbol{\phi}$ their symbolic forms or approximations generated by SR. This algorithm's output can be a single model or multiple candidates to be further discriminated by model-based design of experiments (MbDoE).

2.1. ANN-Based Algorithm for Locating Structural Corrections

Consider $g_t(\boldsymbol{y}, \boldsymbol{\theta}_t)$ in Eqn. 1a as the target-domain ground-truth equations and $f_s(\boldsymbol{y}, \boldsymbol{\theta}_s)$ in Eqn. 1b as the source equations, where $\boldsymbol{\theta}_t \neq \boldsymbol{\theta}_s$.

$$g_t(\boldsymbol{y}, \boldsymbol{\theta}_t) = \theta_1 \frac{y_1}{y_2 + \theta_2} \frac{y_2}{y_1 + \theta_3} - \theta_4 \frac{y_3}{y_4} \qquad (1a)$$

$$f_s(\boldsymbol{y}, \boldsymbol{\theta}_s) = \theta_1' \frac{y_1 y_2}{y_2 + \theta_2'} - \theta_4' \frac{y_3}{y_1} \tag{1b}$$

$$f_t(\boldsymbol{y}, \boldsymbol{\theta}_t) = \theta_1 \frac{y_1 y_2}{y_2 + \theta_2} \left[\frac{1}{y_1 + \theta_3}\right] - \theta_4 \frac{y_3}{y_1} \left[\frac{y_1}{y_4}\right] \tag{1c}$$

Comparing Eqn. 1a and Eqn. 1b reveals structural errors when using the source model to approximate the ground-truth for the target domain. By inserting $[1/(y_1 + \theta_3)]$ and $[y_1/y_4]$ into Eqn. 1b and re-optimising the parameters, Eqn. 1c and 1a match, illustrating the *simplest* symbolic corrections, $\boldsymbol{\phi}(\boldsymbol{y})$. While these corrections (e.g., $\phi_1(\boldsymbol{y}) = [1/(y_1 + \theta_3)]$, $\phi_2(\boldsymbol{y}) = [y_1/y_4]$) are deduced here by inspection; in practice, they are unknown. Thus, we use an ANN parameterised by a set of weights and biases, $\boldsymbol{\omega}$, as an approximator in Eqn. 2b, to learn $\boldsymbol{\varphi}(\boldsymbol{y})$, embedded into the model as in Eqn. 2c.

$$\min_{\boldsymbol{\theta}, \boldsymbol{\omega}} \frac{1}{n_i n_e} \sum_{i=1}^{n_i} \sum_{e=1}^{n_e} \left\| \hat{\boldsymbol{y}}_{i,e}(\boldsymbol{\theta}, \boldsymbol{\omega}) - \boldsymbol{y}_{i,e} \right\|_2 + n_\varphi P(\boldsymbol{y}, \boldsymbol{\omega}) \tag{2a}$$

s. t.

$$\hat{\boldsymbol{\varphi}}(\hat{\boldsymbol{y}}, \boldsymbol{\omega}) = \mathrm{ANN}(\hat{\boldsymbol{y}}, \boldsymbol{\omega}) \tag{2b}$$

$$f_t(\boldsymbol{y}, \boldsymbol{\theta}_t, \boldsymbol{\omega}) = \theta_1 \frac{y_1 y_2}{y_2 + \theta_2} \varphi_1(\boldsymbol{y}, \boldsymbol{\omega}) - \theta_4 \frac{y_3}{y_1} \varphi_2(\boldsymbol{y}, \boldsymbol{\omega}) \tag{2c}$$

$$\hat{\boldsymbol{y}}_{i,e}(\boldsymbol{\theta}, \boldsymbol{\omega}) = \boldsymbol{y}_0^e + \int_0^{t_i} f_t(\boldsymbol{y}, \boldsymbol{\theta}_t, \boldsymbol{\omega})\, \mathrm{d}t \tag{2d}$$

$$P(\boldsymbol{y}, \boldsymbol{\omega}) = \lambda_1 \left\| \frac{\partial \hat{\boldsymbol{\varphi}}}{\partial \hat{\boldsymbol{y}}} \right\|_2 + \lambda_2 \| \hat{\boldsymbol{\varphi}}(\boldsymbol{y}, \boldsymbol{\omega}) - \boldsymbol{b} \|_2 \tag{2e}$$

The transfer model, $f_t(\boldsymbol{y}, \boldsymbol{\theta}_t, \boldsymbol{\omega})$, incorporates ANN-based correction terms, $\boldsymbol{\varphi}(\boldsymbol{y}, \boldsymbol{\omega})$, into the ODEs. The ANN is trained by minimising the mean-square error $\mathcal{L}_{\mathrm{ANN}}(\boldsymbol{\theta}, \boldsymbol{\omega})$ in Eqn. 2a between measured concentrations, $\boldsymbol{y}_{i,e}$, and predictions, $\hat{\boldsymbol{y}}_{i,e}$, obtained by numerically integrating $f_t(\boldsymbol{y}, \boldsymbol{\theta}_t, \boldsymbol{\omega})$ from initial conditions $\boldsymbol{y}_0^e$ to time, t_i, where the subscripts i and e are the measurement time interval and experiment number. By automatic differentiation through numerical integration the kinetic parameters $\boldsymbol{\theta}$ and ANN parameters $\boldsymbol{\omega}$ were optimised simultaneously to avoid bias in the estimated corrections. Adaptive Moment Estimation (ADAM) was used to optimise the parameters, as implemented in the *Python* package *PyTorch*.

The λ-weighted penalty, $P(\boldsymbol{y}, \boldsymbol{\omega})$, in Eqn. 2e is crucial to prevent overfitting. Its first component, the sum of the squares of the elements in the Jacobian, $\partial \hat{\boldsymbol{\varphi}} / \partial \hat{\boldsymbol{y}}$, limits variation in the correction terms, $\hat{\boldsymbol{\varphi}}$, with respect to the fitted states, $\hat{\boldsymbol{y}}$, ensuring only truly influential state variables are retained. Its second component involving a binary vector $\boldsymbol{b}$, pushes additive corrections toward zero and multiplicative ones toward unity, helping to isolate genuine correction locations.

Because correction locations are initially unknown, multiple potential corrections are embedded. Figure 2 illustrates how $\mathcal{D}_{\max}$ sets the initial placement of potential corrections. These, represented by an ANN, are trained together. Unimportant corrections – those with the smallest variance, $\tilde{\varphi}$, over the input domain – are pruned iteratively and the ANN retrained, reducing the number of corrections, n_φ, until it falls below a pre-defined threshold, $\mathcal{U}_{\max} = 3$. Corrections with variance $\tilde{\varphi} < \mathcal{V}_{\min} = 0.1$ are then also finally removed. On an NVIDIA RTX A1000 GPU, ANN training takes ~20 seconds, so 10 pruning iterations require only a few minutes. Automatic symbolic manipulations using *SymPy* streamline expression modifications throughout.

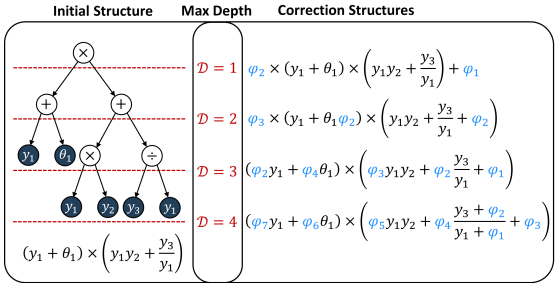

Figure 2: Illustration of correction term embedding at various depths ($\mathcal{D}_{\max}$). The initial equation structure (left) transforms into equations with embedded φ_j (right). For binary branching trees $n_\varphi = F(\mathcal{D}_{\max}) + \mathcal{D}_{\max}$ where, $F(n)$ is the nth entry in the Fibonacci sequence.

2.2. Attribution Guided Symbolic Regression to Construct Correction Equations

Not all features are required for every correction term. Integrated Gradients [5] quantifies each feature's, y_v, contribution to each fitted correction term, $\hat{\varphi}$. This weights which input features SR should use to construct corrections with the effect of either adding or removing them from the source equations.

$$\mathcal{L}_{\mathrm{SR}}(\phi, \boldsymbol{\theta}) = \frac{1}{n_i n_e} \sum_{i=1}^{n_i} \sum_{e=1}^{n_e} \left(\phi(\hat{\boldsymbol{y}}_{i,e}, \boldsymbol{\theta}) - \hat{\varphi}_{i,e}\right)^2 \tag{3}$$

Once Steps 1 and 2 identify the minimal corrections and key features, in Step 3 SR discovers the symbolic forms, $\phi(\boldsymbol{y}, \boldsymbol{\theta})$, for each correction $\hat{\varphi} = \mathrm{ANN}(\boldsymbol{y}, \boldsymbol{\omega})$. Tournament selection promotes and mutates candidates from a population of expressions using the *Julia* package *SymbolicRegression* [3]. The objective function for fitting corrections is shown in Eqn. 3 as the mean-square-error between the value evaluated by each symbolic expression, $\phi(\hat{\boldsymbol{y}}_{i,e}, \boldsymbol{\theta})$, and the values of the fitted corrections, $\hat{\varphi}_{i,e}$, averaged over each time point, i, and experiment, e.

After substituting the symbolic corrections into the source equations, yielding the new target equations, the target parameters, $\boldsymbol{\theta}_t$, are then fine-tuned and sparsity in $\boldsymbol{\theta}_t$ promoted by penalising the absolute magnitude, $\|\boldsymbol{\theta}_t\|_2$.

2.3. Integrating SR and MbDoE for Model Discrimination

Model-based design of experiments (MbDoE) for model structure discrimination maximises information gain by selecting experimental conditions that maximise differences in predicted responses among model candidates, reducing time and resource expenditure.

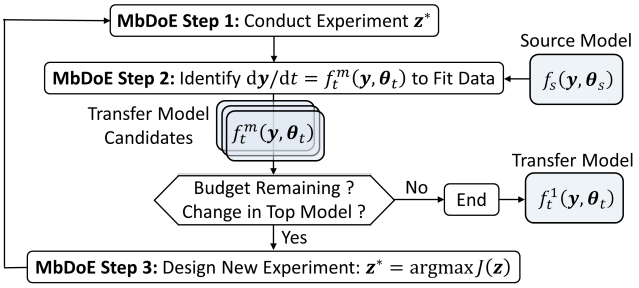

Figure 3: Transfer learning integrated with SR-MbDoE.

The SR-MbDoE framework [6] is adapted in Figure 3 to integrate model structure transfer learning. In MbDoE-Step 1, experiments parameterised by $\boldsymbol{z}$ are conducted. MbDoE-Step 2 uses model structure transfer learning to propose different candidate transfer models. In MbDoE-Step 3, so long as the experimental budget has not been exceeded or the equations from one iteration to the next are still being refined, a new experiment is designed by optimising to an objective function, $J(\boldsymbol{z})$.

In model structure transfer learning, SR generates a Pareto set of candidate expressions. These are scored by the derivative in the log-loss with respect to complexity. The top-scoring expressions represent a balance between parsimony and fitting accuracy. By substituting combinations of the top two scoring corrections into the source equations, various candidate transfer models, $f_t^m(\boldsymbol{y}, \boldsymbol{\theta}_t)$, indexed by m, are created. In this work, the design variables, $\boldsymbol{z}$, were the initial concentrations of each (bio)chemical species, $\boldsymbol{y}_0$, while the objective function was the Hunter-Reiner criterion.

3. BIOPROCESS CASE STUDY

This study considers an in-silico case study representing microalgae cultivation for producing a high-value nutraceutical via carbon substrate consumption. The *source* model, in Eqn. 4, describes the evolution of biomass (X), substrate (S), and product (P) concentration, with the parameters: $\boldsymbol{\theta}_s = (\mu_m, K_S, Y_S, Y_P, \beta)$, where $\mu_m = 0.1\ \mathrm{h}^{-1}$ is the maximum specific growth rate, $K_S = 10\ \mathrm{g\,l}^{-1}$ is the substrate half-saturation constant, $Y_S = 1\ \mathrm{g\,g}^{-1}$ is the growth-dependent substrate yield coefficient, while $Y_P = 0.5\ \mathrm{g\,g}^{-1}$ and $= 0.003\ \mathrm{h}^{-1}$ are the growth-dependent and independent product yield coefficients, respectively.

$$\frac{\mathrm{d}X}{\mathrm{d}t} = \mu_m \frac{S}{S + K_S} X \tag{4a}$$

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -Y_S\, \mu_m \frac{S}{S + K_S} X \tag{4b}$$

$$\frac{\mathrm{d}P}{\mathrm{d}t} = Y_P \mu_m \frac{S}{S + K_S} \cdot X + \beta X \tag{4c}$$

In-silico data for the *target* domain, representing a more productive strain, was generated using the ground-truth model in Eqn. 5. The target model parameters, $\boldsymbol{\theta}_t = (\mu_m, K_S, \mu_d, Y_S, Y_P, \beta, k_d)$, share the same meanings as in the source model but with different values: $K_S = 5\ \mathrm{g\,g}^{-1}$, $\beta = 0.05\ \mathrm{h}^{-1}$. New is a biomass decay constant, $\mu_d = 0.1\ \mathrm{h}^{-1}$, and product reversal rate under substrate-limiting conditions, $k_d = 0.07\ \mathrm{g}^{-1}\,\mathrm{l\,h}^{-1}$, while Eqn. 5 uses Contois, not Monod kinetics, altogether reflecting the physical differences between the source and target systems.

$$\frac{\mathrm{d}X}{\mathrm{d}t} = \mu_m \frac{S}{S + K_S X} \cdot X - \mu_d X \tag{5a}$$

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -Y_S\, \mu_m \frac{S}{S + K_S X} \cdot X \tag{5b}$$

$$\frac{\mathrm{d}P}{\mathrm{d}t} = Y_P \mu_m \frac{S}{S + K_S X} \cdot X + \beta X - k_d X^2 \tag{5c}$$

Batch simulations each run for 144 hours from initial conditions $\boldsymbol{y}_e^0 = (X_{0,e}, S_{0,e}, P_{0,e})$. States $\boldsymbol{y}_{i,e} = (X_{i,e}, S_{i,e}, P_{i,e})$ were recorded every 14 hours with 5% normally distributed noise. This represents practical conditions with measurement uncertainty, and offline data is not frequently collected. Each case study starts with three experiments from different initial conditions. To assess reproducibility and data sensitivity, the framework is re-run ten times, starting from different sets of three experiments generated from different random initial conditions.

4. RESULTS AND DISCUSSION

Section 4.1 will demonstrate the robustness of the proposed model structure transfer learning framework, given a detailed but imperfect source model. Section 4.2 will then discuss the impact of the quantity of prior knowledge on the efficiency of the framework.

4.1. Scenario with Moderate Inductive Bias

The source model to be modified is given by Eqn. 4. In Section 4.1.1, three initial experiments are used for each of the ten repetitions; Section 4.1.2 extends this with further experiments designed via MbDoE.

4.1.1. Single vs Multi-Pass Structure Modification

A single-pass approach first embeds multiple correction terms into the source model, which are then sparingly refined. First, in Step 1 in Figure 1, Eqn. 6 shows where twelve correction terms, φ, were automatically

embedded into the source model structure. Note that overall specific growth rate, μ, is a shared sub-function. Here ϕ_j^{GT} will denote the *simplest* symbolic corrections when comparing the corrections with the ground truth equations. Comparison with the target ground-truth in Eqn. 5 indicates three key corrections – one for adapting the Monod to the Contois kinetics ($\phi_2^{\mathrm{GT}} = X$), one for biomass decay ($\phi_7^{\mathrm{GT}} = -\mu_d X$) and another for product reversal ($\phi_{12}^{\mathrm{GT}} = -k_d X^2$), while other terms ideally vanish. Note others are possible; should $\phi_7^{\mathrm{GT}} = -\mu_d X$ and $\phi_{12}^{\mathrm{GT}} = -k_d X^2$ then $\phi_3^{\mathrm{GT}} = (S + K_S X - K_S)/S$ or $\phi_4^{\mathrm{GT}} = (S + K_S)/(S + K_S X)$ for example are equally valid, though are not the *simplest*.

$$\frac{\mathrm{d}X}{\mathrm{d}t} = \varphi_6 \mu X + \varphi_7 \tag{6a}$$

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\varphi_8 Y_S \mu X + \varphi_9 \tag{6b}$$

$$\frac{\mathrm{d}P}{\mathrm{d}t} = \varphi_{11} Y_P \mu X + \varphi_{10} \beta X + \varphi_{12} \tag{6c}$$

$$\mu = \varphi_4 \mu_m \frac{S + \varphi_1}{\varphi_3 S + \varphi_2 K_S} + \varphi_5 \tag{6d}$$

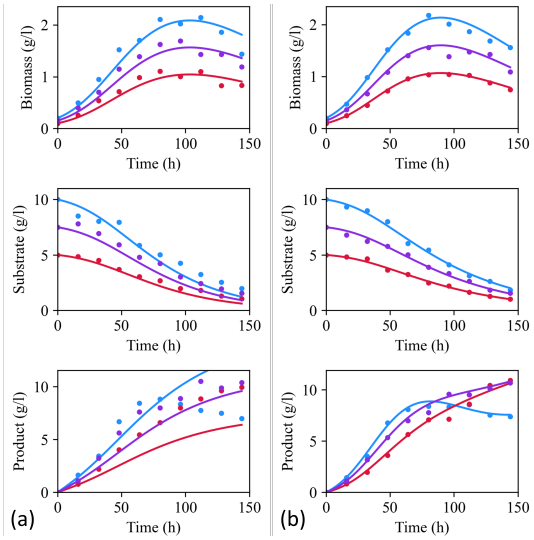

Figure 4: Comparison of a model fitting built from one of the random initial experiments after one pass (a) and after three passes (b) of model structure modification.

Across ten repetitions, the source structure was partially corrected in a single pass 10% of the time, where Eqn. 7 illustrates one such partially corrected model; this improves to 50% with sparsity promotion. However, certain terms were often incorrectly inferred: for example, falsely attributing product reversal under substrate-limiting conditions to a product accumulation term that simply diminished towards the end of the batch (i.e., $\phi_{12} = 1.88S$ rather than $\phi_{12}^{\mathrm{GT}} = -k_d X^2$). It could be said that it is prior knowledge that such decay terms should be negligible at the start of the batch; in fact, this information could be incorporated during ANN training but to demonstrate the generality of the method we assume we do not know this.

Now, the equations are returned for further passes to assess whether partially corrected equations can be fully adapted to the target domain. No further passes are required once there are no correction terms with a variance, $\tilde{\varphi}$, greater than the threshold, $\mathcal{V}_{\min}$. Allowing multiple passes of model structure modification raised the success rate from 50% to 80%, illustrating that iteratively refining partially corrected equations is advantageous. For one model, the mean absolute percentage error (MAPE) of fitting improved from 17% to 6.6% as shown in Figure 4a to 4b, once multiple passes discovered the ground truth. Meanwhile, the MAPE of prediction on the unseen experiments dropped from 23% to 10.7%.

4.1.2. Model-Based Design of Experiments

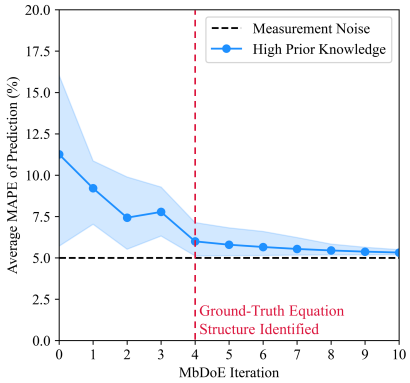

Figure 5: The mean absolute percentage error (MAPE) of prediction over each model-based design of experiment (MbDoE) iteration. The shaded region denotes three standard deviations in the MAPEs at each MbDoE iteration over the ten random initialisations, while the markers are the average. Here, multiple passes of structure modification are allowed.

When only a few noisy experiments are available, multiple candidate structures may explain the data equally well. MbDoE identifies experimental conditions for discriminating among those candidates. In Figure 5, given three initial experiments, the average MAPE was 11.2% with a standard deviation of 3.1% across the 10 repetitions. This decreased to 5.3% with a standard deviation of 0.1% after thirteen new MbDoE experiments. Eight out of ten runs achieved a MAPE of 6.0% by the fourth MbDoE iteration, at which point all the runs discovered the ground truth equation structures for the target domain. The product-reversal term ($\phi_{12}^{\mathrm{GT}} = -k_d X^2$) was typically discovered last when it was elucidated that $\phi_{12} = f(X)$ not $\phi_{12} = f(X, S)$ was the simplest correction, highlighting how separating the causal relationship can be difficult when state variables are strongly correlated – such as when one (bio)chemical species reacts to another, and the experiments are poorly designed.

4.2. Impact of Prior Knowledge Quantity

The framework's efficiency will be case-dependent. Eqn. 7b replicates the biomass growth model used here, while Eqn. 7a reflects the fundamental assumption that growth is proportional to the current cell population. Starting from a more minimal model will likely require more MbDoE experiments, as more kinetics must be uncovered. However, incorporating more physical knowledge – as in Eqn. 7c with a substrate inhibition term (K_I/S) and a product toxicity term $(P/(P+K_P))$ – increases the risk of introducing incorrect inductive bias, complicating model modification. For a completely new bioprocess, such as a novel strain, it is prudent to start with a simpler source model structure to identify key correction terms, then compare the generated equations with those from higher knowledge source models.

$$\frac{\mathrm{d}X}{\mathrm{d}t} = \mu X \tag{7a}$$

$$\frac{\mathrm{d}X}{\mathrm{d}t} = \mu_m \frac{S}{S+K_S} X \tag{7b}$$

$$\frac{\mathrm{d}X}{\mathrm{d}t} = \mu_m \frac{S}{S+K_S+\frac{K_I}{S}} \cdot X \; - \; \mu_d X \frac{P}{P+K_P} \tag{7c}$$

Prior knowledge can be incorporated beyond the source model structure. Previously, the correction $\phi_{12} = -k_d X^2$ was mistaken as $\phi'_{12} \propto S$ using only the first three experiments. When state variables are strongly correlated – such as when one (bio)chemical species reacts to form another (e.g., $A + 2B \rightarrow C$) – separating the causal relationship becomes more difficult. Fortunately, the framework is highly flexible. In these cases, extra constraints during ANN training are valuable. For example, bioprocess kinetics often follow Eqn. 8, with a growth-dependent term, $\varphi_{GD}(\boldsymbol{y})$, and a growth-independent term, $\varphi_{GI}(\boldsymbol{y})$, (e.g., biomass decay or substrate consumption for maintenance). By providing synthetic data to the ANN that pushes $\varphi_{GI}(\boldsymbol{y})$ to zero when biomass is zero, $\phi_{12} = -k_d X^2$ was reliably identified without MbDoE.

$$\frac{\mathrm{d}\boldsymbol{y}}{\mathrm{d}t} = \mu \cdot \varphi_{GD}(\boldsymbol{y}) + \varphi_{GI}(\boldsymbol{y}) \tag{8}$$

4. CONCLUSION

This study introduces a structural transfer learning framework combining symbolic regression (SR) with ANN feature attribution to adapt mechanistic models from literature or common concept for newly investigated systems. Unlike traditional black-box-based transfer learning approaches that only update model parameters, the proposed method modifies the structure of mechanistic models, enabling the discovery of interpretable differential equations. An in-silico bioprocess case study demonstrated the framework's robustness under low-sample frequency, noisy data conditions. Integration with model-based experiment design (MbDoE) minimised experimental effort by efficiently discriminating among candidate transfer models for the new system. Requiring several hyperparameters, the framework offers great flexibility for human-machine interaction, allowing experts to guide model complexity and the interpretability necessary in highly regulated industries. Overall, this general methodology advances automated model discovery for accelerating novel bioprocess development, and within other fast-paced (bio)chemical industries.

ACKNOWLEDGEMENTS

This research was supported by the Engineering and Physical Sciences Research Council (EPSRC) and Unilever. The authors gratefully acknowledge the crucial financial support and resources provided.

REFERENCES

1. S. Kay, H. Kay, A.W. Rogers, D. Zhang, Integrating hybrid modelling and transfer learning for new bioprocess predictive modelling, in: CACE, Elsevier, 2023: pp. 259–260. https://doi.org/10.1016/B978-0-443-15274-0.50412-1.
2. G.A. Pimentel, F.N. Santos-Navarro, L. Dewasme, A.V. Wouwer, Elucidation of Macroscopic Stoichiometry and Kinetics of Bioprocesses using Sparse Identification, IFAC 58 (2024) 422–427. https://doi.org/10.1016/j.ifacol.2024.08.373.
3. M. Cranmer, Interpretable Machine Learning for Science with SymbolicRegression.jl, (2023). https://doi.org/10.48550/ARXIV.2305.01582.
4. T. Forster, D. Vázquez, C. Müller, G. Guillén-Gosálbez, Machine learning uncovers analytical kinetic models of bioprocesses, CES (2024) 120606. https://doi.org/10.1016/j.ces.2024.120606.
5. M. Sundararajan, A. Taly, Q. Yan, Axiomatic Attribution for Deep Networks, (2017). http://arxiv.org/abs/1703.01365.
6. A.W. Rogers, A. Lane, C. Mendoza, S. Watson, A. Kowalski, P. Martin, D. Zhang, Integrating knowledge-guided symbolic regression and model-based design of experiments to automate process flow diagram development, CES 300 (2024) 120580. https://doi.org/10.1016/j.ces.2024.120580.

© 2025 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See https://creativecommons.org/licenses/by-sa/4.0/