Combining hybrid modelling and transfer learning to simulate fed-batch bioprocess under uncertainty

Oliver Pennington*, Youping Xie, Keju Jing***, Dongda Zhang*.**

** Department of Chemical Engineering and Analytical Science, University of Manchester, M1 3BU, UK (e-mail: oliver.pennington@manchester.ac.uk, dongda.zhang@manchester.ac.uk).*
*** Marine Biological Manufacturing Center of Fuzhou Institute of Oceanography, Fuzhou University, Fuzhou 350108, China (email: ypxie@fzu.edu.cn)*
**** Department of Chemical and Biochemical Engineering, College of Chemistry and Chemical Engineering, Xiamen University, (e-mail: jkj@xmu.edu.cn)*

Abstract: Hybrid modelling utilizes advantageous aspects of both mechanistic (white box) and data-driven (black box) modelling. Combining the physical interpretability of kinetic modelling with the power of a data-driven Artificial Neural Network (ANN) yields a hybrid (grey box) model with superior accuracy when compared to a traditional mechanistic model, while requiring less data than a purely data-driven model. This study aims to construct a hybrid model for the predictive modelling of a high-cell-density microalgal fermentation process for lutein production under uncertainty. In addition, transfer learning is combined with the hybrid model to simulate new fed-batches utilizing alternative substrates operated under a different reactor scale. By comparing with experimental data, the hybrid transfer model was found to be able to simulate the new fed-batch processes that achieve heightened cell densities and higher product quantities. Overall, this work presents a novel digital model construction strategy that can be easily adapted to general bioprocesses for model predictive control and process optimization under uncertainty.

Keywords: Machine learning, Transfer Learning, Hybrid modelling, Biosystem, Fed-batch processes.

1. INTRODUCTION

1.1 Motivation

Bioprocesses connect a multitude of global research interests, including the generation of renewable plastics, fuels, and other valuable bioproducts. Economically speaking, the UK bioeconomy alone was worth around £220 billion in 2018, and is set to double by 2030, making the field lucrative for research (Harrington, 2018). Supporting the development of bioprocesses requires overcoming several challenges, including substantial batch-to-batch variation, introducing challenges associated with quality control, deficient metabolic and secretory phenotypes for the production of proteins, accumulation of by-products, leading to heightened separation costs and loss of product, and finally lower yields in the scale-up of reactors. Tackling these challenges commences in several stages, including identifying suitable microbial strains and their optimal operating conditions, followed by the development of optimal operating and control strategies. Understanding how different strains interact with their environment is an essential step for uncovering these optimal operating conditions to overcome the aforementioned challenges that large-scale fermentation faces.

With the evolution of the fourth industrial revolution, the transition towards digitalization is becoming increasingly prevalent. Utilizing digital twins, in which a process model is required, has huge potential in the application of process optimization and control, and design of experiments. The concept of machine learning is being continuously applied in new ways to unveil more economical, safer, efficient, and sustainable approaches to chemical processing. Due to the continuous growth of interest within the fields of both bioprocess engineering and artificial intelligence, it is an excellent period to harness and fuse the advantages of each. However, accounting for uncertainty is an equally important step the development of a model, as it provides information surrounding the achievability of a found control strategy or optimum.

An example of where the combining of different modelling methodologies may be required is high-cell-density cultures (HCDCs). HCDCs can be capable of producing larger quantities of desired products, an example of which is lutein, a carotenoid pigment commonly used in the food industry. However, HCDCs can demonstrate self-inhibitory effects, where growth and/or substrate uptake is inhibited by the high cell density or high accumulation of products (Riesenberg et al., 1999). Gaining a better understanding of these inhibitory effects is pertinent for optimizing processes utilizing HCDCs.

1.2 Aims

This study aims to develop a predictive model to simulate a HCDC under various fed batch conditions. The dynamics of the HCDC are tackled by incorporating an Artificial Neural Network (ANN) where more in-depth physically derived biosystem kinetic models are challenging to identify. Once a suitable model is developed, uncertainty analysis and transfer

learning can be conducted for the utilization of a different substrate, in this case the replacement of nitrate with urea.

Ultimately, this model aspires to require minimal data and computational expense, while maximizing simulation accuracy for process optimization and control applications. This will make the modelling methodology described in this work applicable to other bioprocesses involving HCDCs, such as *Akkermansia muciniphila* which has received recent attention for its probiotic capabilities in the intestinal tract (Wu et al., 2024). With HCDCs being well suited to fed-batch and perfusion operation, they play a key role in moving towards the continuous operation of bioprocesses; such progression can be accelerated by the modelling and optimization strategy discussed in the recent study (Zinnecker et al., 2024).

In this work, the term *hybrid modelling* relates to the combination of a physically derivable (white-box) model, with a data-driven (black-box) model to yield an overall hybrid (grey-box) model. The term *transfer learning* refers to the transfer of information from one model to another in the form of penalizing deviations in parameters or simulated dynamics.

1.3 Case study

The case study at hand looks at the production of lutein from *Chlorella sorokiniana* under fed-batch conditions using sodium nitrate or urea as the source of nitrogen, depending on the case, and glucose as the carbon source. The dataset available also includes a preliminary batch using sodium nitrate as the nitrogen source. Experiments were run for 144 hours with measurements of biomass density and lutein content taken every 12 hours. For fed-batch operation, substrates were consumed until depletion, which was indicated by rising dissolved oxygen content in the medium. Further detail of the experiments can be found in a previous study (Xie et al., 2022).

2. METHODOLOGY

2.1 Overall approach

Initially, a single batch process in a 250 mL flask (with 100 mL of working volume) was conducted using sodium nitrate as the nitrogen source, to which a preliminary kinetic model was fitted. This model was then applied to a fed-batch process operating from the same initial conditions but in a 5 L bioreactor. Next, transfer learning was applied to simulate the fed-batch process with urea as an alternative nitrogen source to sodium nitrate. Uncertainty analysis was then conducted using bootstrapping to quantify the uncertainty in the model stemming from the experimental data.

2.2 Macro-scale kinetic modelling

For the initial batch experiment, Monod-inspired kinetics were utilized. The extracellular medium concentrations of the substrates glucose (G) and nitrate (N), as well as cell density (X) and lutein content (L_C), are described by the system of 4 Ordinary Differential Equations (ODEs) as in

$$\frac{dX}{dt} = \mu \cdot X - d_X \cdot X, \tag{1}$$

$$\frac{dG}{dt} = -v_{\max_G} \cdot \frac{G}{K_G + G} \cdot X, \tag{2}$$

$$\frac{dN}{dt} = -v_{\max_N} \cdot \frac{N}{K_N + N} \cdot X, \tag{3}$$

$$\frac{dL_c}{dt} = Y_{LX} - d_L \cdot L_C - \mu \cdot L_C, \tag{4}$$

$$\mu = \mu_{\max} \cdot \frac{G}{K_G + G} \cdot \frac{N}{K_N + N}, \tag{5}$$

where $\mu_{\max}$ refers to the maximum specific growth rate of biomass, with $v_{\max_i}$ being the maximum specific uptake rate for a given substrate i, K_i being the affinity constant for a given substrate i, Y_{LX} being the specific production rate of lutein from biomass, and d_i representing a decay term for a given product i. It should be noted that L_C represents an intracellular concentration measured in mg g^{-1} which is why dilution must be accounted for in Equation (4). To consider the overall lutein concentration in the reactor, product rule can be used with equations (1) and (4) to give the result as in

$$\frac{dL}{dt} = Y_{LX} \cdot X - d_X \cdot L - d_L \cdot L. \tag{6}$$

It is essential for parameters in the Monod model to remain positive in value to retain physical feasibility, as in

$$\boldsymbol{\beta} \geq \mathbf{0}, \tag{7}$$

where $\boldsymbol{\beta}$ is the vector of Monod model parameters. Parameters were identified using a stochastic optimization algorithm called Particle Swarm Optimization (PSO) (Wang et al., 2017) that minimizes a mean squared error objective function, as in

$$\min \frac{1}{n} \cdot \sum_{t_{\text{meas}}} \sum_{i} \left(\frac{C_{i_t} - C_{i_{\text{meas}_t}}}{\sigma_i} \right)^2 \tag{8}$$

where n is the number of datapoints. The objective function in Equation (8) minimizes the difference between the simulated extracellular concentration profiles and the measured averages at each timepoint t; each term i is weighted by its experimental measurement standard deviation; σ_i.

2.3 Hybrid modelling

Despite good fitting of the initial batch process, the parameters identified could not capture the entire fed-batch process. The most noticeable process-model mismatches can be seen later in the process during high cell densities, particularly once the culture reaches a cell density over 10 times that seen in the preliminary batch experiment. This indicates that substrate uptake and growth are being inhibited by high cell density, which can be due to poorer mass transfer of substrates to the cells due to their close proximity. The close proximity, or even contact, of cells can also inhibit growth and substrate uptake (Ruhe et al., 2013). Therefore, three time-varying parameters were introduced: two capture the respective substrate uptake

inhibition of glucose and nitrate ($\theta_G(\cdot)$ and $\theta_N(\cdot)$, respectively); and the other captures biomass growth inhibition ($\theta_X(\cdot)$). These time-varying parameters are functions of state variables, in the form of an ANN. Equations (1), (2), and (3) are therefore rewritten as in

$$\frac{dX}{dt} = \theta_X(X,G,N)\cdot\mu\cdot X - d_X\cdot X, \tag{9}$$

$$\frac{dG}{dt} = -\theta_G(X,G,N)\cdot v_{\max_G}\cdot\frac{G}{K_G+G}\cdot X, \tag{10}$$

$$\frac{dN}{dt} = -\theta_N(X,G,N)\cdot v_{\max_N}\cdot\frac{N}{K_N+N}\cdot X, \tag{11}$$

where the remaining parameters in Equations (4)-(5) and (9)-(11) take the same value as those previously identified. In order to train the ANN, time-varying parameter profiles are initially identified as functions of time, during which their deviations are penalised to try to follow the kinetic model dynamics as closely as possible. Penalisation is done using a penalty term, that is added to the objective function shown in Equation (8), as in

$$\min\frac{1}{n}\cdot\sum_{t_{\text{meas}}}\sum_{i}\left(\frac{C_{i_t}-C_{i_{\text{meas}_t}}}{\sigma_i}\right)^2 + \rho\cdot\sum_{t_{\text{TVP}}}\sum_{i}\left(\theta_i(t)-\theta_i(t-1)\right)^2 \tag{12}$$

where ρ is a hyperparameter to be tuned, t_{meas} illustrates the timepoints at which measurements are taken, and t_{TVP} illustrates the timepoints at which time-varying parameters are discretized. Discretization of time-varying parameters is conducted, at most, as frequently as measurements are taken. The time-varying parameters are then regressed as a function of state variables in the form of an ANN. The ANN is trained using the profiles of state variables and time-varying parameters generated by the objective function in Equation (12). The ANN is embedded within the system of ODEs as illustrated in Fig. 1, which shows a series hybrid model, as described in a previous study (Mowbray et al., 2021). State variables update the ANN which updates the time-varying parameters, which can be used in the system of ODEs to update the state variables, thus generating state variable profiles. Comparing the performance of several ANN architectures allowed the identification of the most appropriate ANN structure. The Akaike Information Criterion with correction for small sample sizes ($AICc$) was employed to quantify the performance of each ANN structure, as in

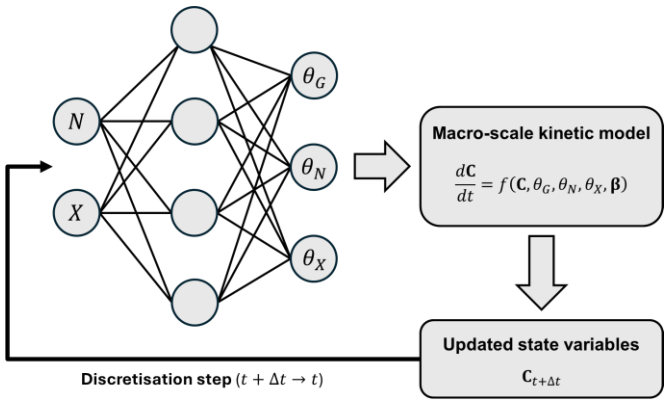

Fig. 1. Hybrid model structure: Simulation of time-varying parameters based on system components using an ANN.

$$AICc = n\ln Z + 2\,k + \frac{2\,k^2+2\,k}{n-k-1}, \tag{13}$$

where k is the number of parameters with a correction for the small sample size n. The ANN architecture with the lowest $AICc$ score was chosen as the best trade-off between minimal overfitting and accuracy.

2.4 Transfer learning

In order to simulate fed-batch operation using a different nitrogen source, in this case urea, transfer learning was employed. The same kinetic model parameters were used as those found in the preliminary kinetic model fitting; only the time-varying parameters were updated. For the fitting of the urea-fed fed-batch process, a penalty term was introduced to penalise deviations between the time-varying parameters in the different systems, alongside Equation (12), as in

$$\min\frac{1}{n}\cdot\sum_{t_{\text{meas}}}\sum_{i}\left(\frac{C_{i_t}-C_{i_{\text{meas}_t}}}{\sigma_i}\right)^2 + \rho\cdot\sum_{t_{\text{TVP}}}\sum_{i}\left(\theta_{i_{urea}}(t)-\theta_{i_{urea}}(t-1)\right)^2 + \lambda\cdot\sum_{t_{\text{TVP}}}\sum_{i}\left(\theta_{i_{NO_3}}(X,G,N)-\theta_{i_{urea}}(t)\right)^2 \tag{14}$$

where $\theta_{i_{NO_3}}(X,G,N)$ and $\theta_{i_{urea}}(t)$ are the time-varying parameters associated with the inhibition of phenomena i for the fed-batch operation utilizing nitrate and urea, respectively, and λ is a hyperparameter referred to, in this study, as the transfer learning factor. Increasing the value of λ increases the transfer of information from the nitrate-fed system to the urea-fed system. The hyperparameters (or penalty weights) were tuned iteratively to take the maximum possible value that maintained acceptable model fitting. The form of transfer learning represented in the objective function in Equation (14) is referred to as regularization-based transfer learning, due to the presence of a penalty term (Weiss et al., 2016). For the transfer learning model, $\theta_{i_{urea}}(\cdot)$ is preliminarily a function of time to be used to train an ANN to reformulate the parameter as a function of state variables, as has already been done with $\theta_{i_{NO_3}}(\cdot)$ in the original hybrid model. Although a higher value of lambda can mean less data is required to re-train the ANN, using a value that is too large can reduce the accuracy of the transfer learning model. An ANN of the same architecture was then re-trained using the updated time-varying parameters, thus generating a new hybrid model through transfer learning.

2.5 Uncertainty Analysis

Uncertainty analysis was conducted using bootstrapping which involved taking samples of the training dataset and

training the model on each sample separately, generating a distribution of model outputs, as illustrated in Fig. 2. A 95% confidence interval (z_{CI}) of each component z was generated by sampling from the resulting output distribution, as in

$$z_{CI} = \bar{z} \pm c \cdot \frac{\sigma_z}{\sqrt{\frac{n_z - 1}{n_z}}}, \quad (15)$$

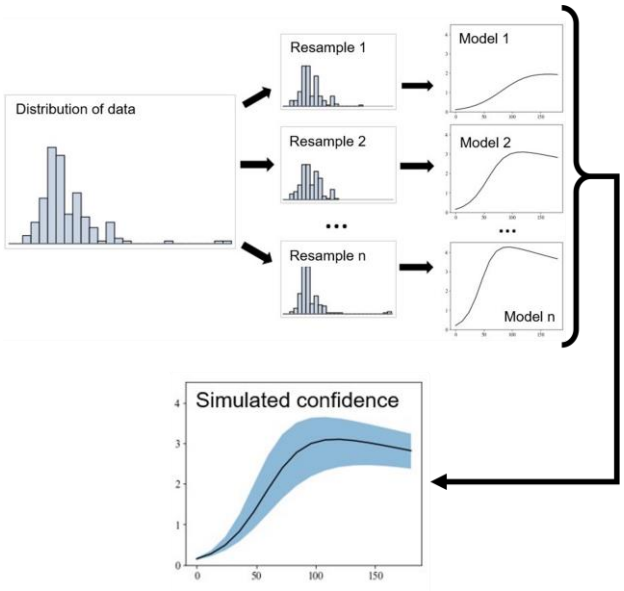

Fig. 2. Bootstrapping: Schematic.

where $\bar{z}$ is the mean of the component z, c is the confidence level, n_z is the number of samples of the component z, and σ_z is the standard deviation of the component z. In this study, bootstrapping was conducted by removing 48-hour segments of data, a third of the fed-batch time horizon.

3. RESULTS AND DISCUSSION

3.1 Results of hybrid modelling

The purely mechanistic macro-scale kinetic model was first fit with Equations (1)-(5) with no time-varying parameters. The fitting of the preliminary kinetic model has a mean R^2 value of 0.991 and is shown in Fig. 3. Parameter estimation converged after 50 iterations using a derivative-free Particle Swarm Optimization (PSO) algorithm. The selected algorithm has strong exploratory capabilities for parameter estimation in the stiff fed-batch system simulated in this study. The problem was run 5 times, each with 100 particles, to confirm the same optima was being found, increasing the confidence in it being a global optimal solution.

The application of this kinetic model to a 5 L fed-batch process, where the cell density reaches levels in excess of 10 times higher than the batch process, illustrated significant process-model mismatch, particularly in the latter stages at the highest cell densities, which is evident in the resulting near-zero R^2 value of 0.072 for biomass and lutein. This motivated the introduction of time-varying parameters to better fit the more challenging dynamics seen in the latter stages of the fed-batch process. Time-varying parameters were first introduced to a fed-batch operation that reinstated the same initial substrate concentrations as the preliminary batch process after each injection of concentrated medium. Time-varying parameters were initially simulated as a discretized function of time. Parameter estimation converged after 200 iterations using the same derivative-free PSO algorithm. The updated

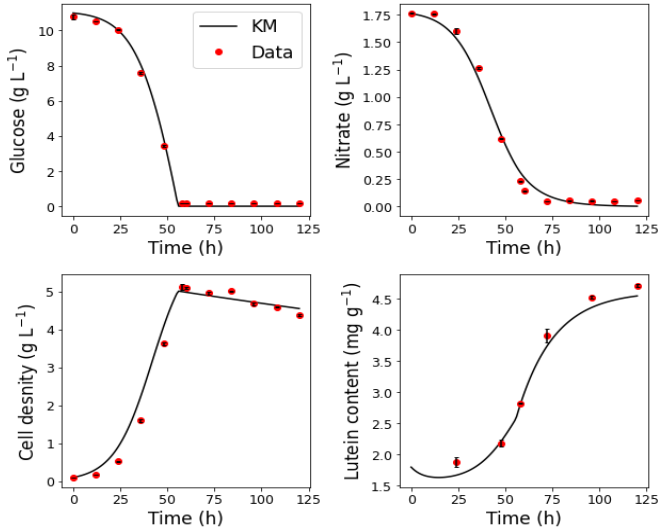

Fig. 3. Kinetic model (KM): Plots of substrates glucose and nitrate, biomass and lutein content fitted to the original nitrate-fed batch.

problem was also run 5 times, each with 100 particles.

The resulting time-varying parameter profiles were then used to train the feed-forward ANN. The most appropriate ANN architecture was that with the nitrogen source concentration and cell density as the inputs, with three nodes in a single hidden layer and sigmoid activation functions. The ANN was trained over 1700 epochs using data augmentation, following the method described in a previous study (Rogers et al., 2020). The fitting of the hybrid model is shown in Fig. 4, with the profiles of the corresponding time-varying parameters shown in Fig. 5. The introduction of time-varying parameters increased the mean R^2 value for cell density and lutein content from 0.072 to 0.961. The decreasing trend of all three time-varying parameters simulated by the ANN highlights the inhibitory effects that the high-cell-density has on substrate uptake and growth. It is observable that all parameters decrease over time, with negligible growth and nitrate uptake in the latter stages of the process, and only some glucose uptake.

3.2 Results of transfer learning

The production of lutein from *C. sorokiniana* can utilize an alternative nitrogen source in the form of urea. Since the same cell line and medium content is utilized (with the exception of the nitrogen source), transfer learning can be employed to maximize the passage of knowledge from the source domain to the target domain, thus minimizing the data requirement to accurately capture the dynamics of the urea-fed system through the re-training of the ANN. It should be noted that urea

was added so that the resulting nitrogen concentration was the same as that within the nitrate-fed process.

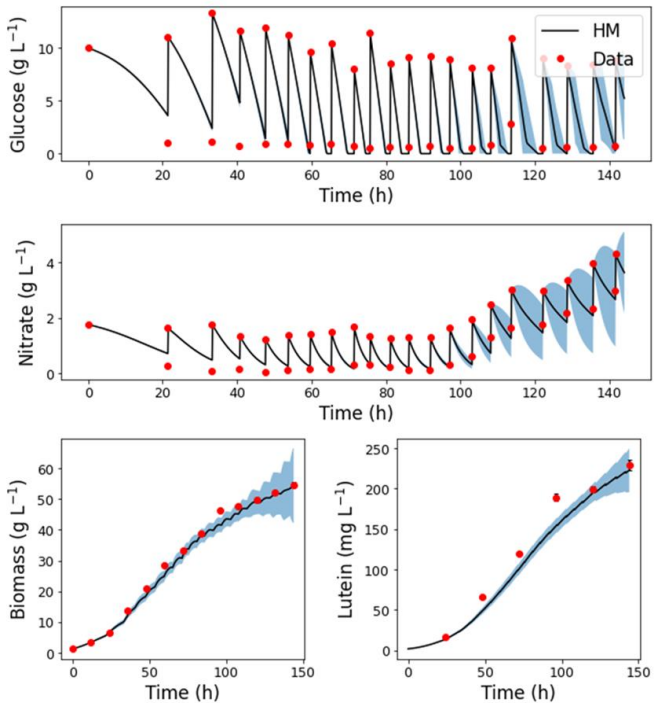

Fig. 4. Hybrid model (HM): Plots of substrates glucose and nitrate, biomass and lutein content for fed-batch simulation.

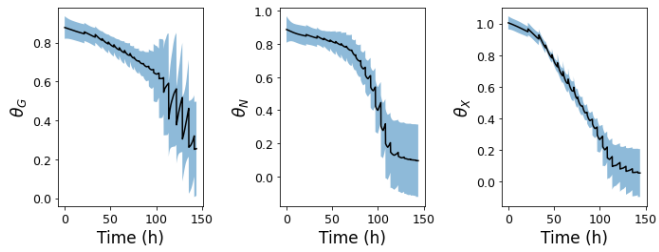

Fig. 5. Hybrid model: Time-varying parameters.

The comparison between the original hybrid model (trained on the nitrate system) and a subsequent transfer learning model for the urea-fed system is shown Fig. 6. It should be noted that 'Nitrate eq.' refers to the nitrate-equivalent concentration of urea. The R^2 value in the fitting of cell density was improved from 0.704 to 0.967 due to the implementation of transfer learning, thus demonstrating the power of transfer learning in HCDC applications utilizing different nitrogen sources.

Additionally, the fact that the production of lutein did not require the implementation of an additional time-varying parameter (to the time-varying parameter for specific growth rate) for either model indicates that the kinetic model adequately describes the dynamics of lutein content within the cells under batch and fed-batch conditions for both nitrate-fed and urea-fed systems. The structure of the hybrid model implies that, for the system at hand, lutein production is dependent solely on the growth of biomass, and not the nitrogen substrate used, which is evidenced by the model simulation result.

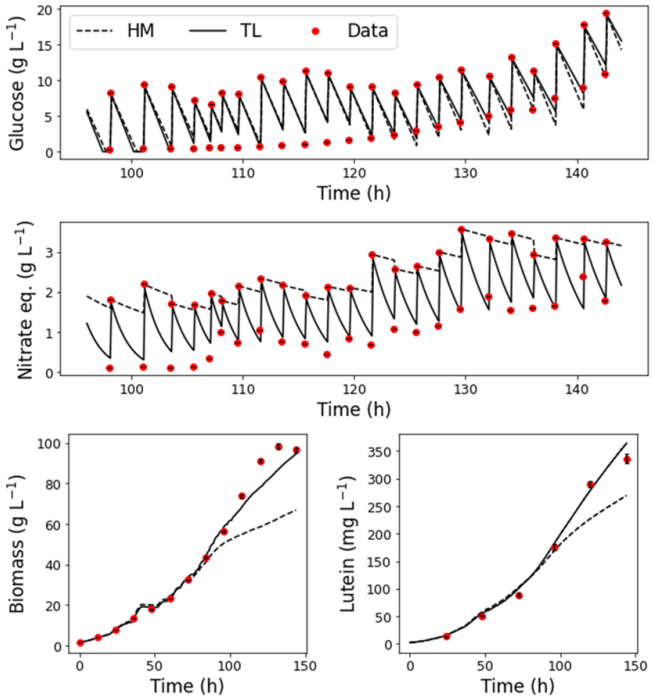

Fig. 6. Comparison between the original hybrid model (HM) the transfer learning model (TL).

3.3 Uncertainty Analysis

Uncertainty analysis was successfully applied through bootstrapping to both the original hybrid model and the transfer learning updated hybrid model. The original hybrid model uncertainty is illustrated in Fig. 4 and Fig. 5, and the transfer-learning generated model uncertainty is illustrated in Fig. 7 and Fig. 8 The narrow confidence intervals in Fig. 4 suggest that the data is not a significant source of uncertainty. Low model uncertainty from the experimental data is likely due to the high frequency and precision at which measurements were taken. The only inflations in uncertainty are observed in the latter stages of the process, beyond the 110-hour mark, which is likely due to the increased sensitivity of substrate uptake to the heightened cell density seen in the latter stages of the process, resulting in more complex inhibition mechanisms.

Despite the removal of 48-hour intervals of data to generate different ANNs through bootstrapping, the confidence intervals for both models remain narrow (as seen in Fig. 4 and Fig. 7), highlighting the capability of hybrid modelling to maintain low model uncertainty.

4. CONCLUSIONS

In this study, two hybrid models were constructed to simulate two fed-batch HCDCs for lutein production. Transfer learning was employed to construct a hybrid model using an alternative nitrogen source, in the form of urea, to compensate for the reduced availability of data. The introduction of time-varying parameters facilitated the capture of increasingly inhibited system dynamics during the latter stages of fed-batch operation. The varying system dynamics were accurately captured, thus maximizing the model's applicability to

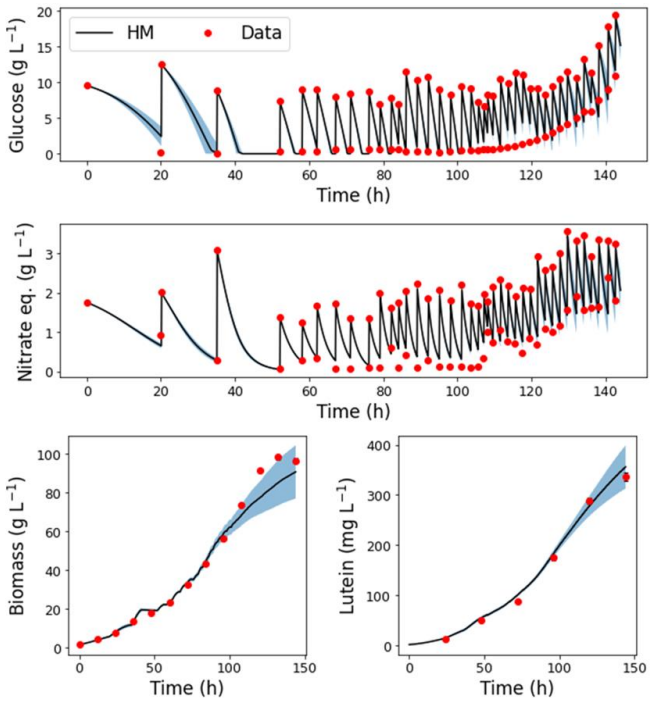

Fig. 7. Transfer learning: Uncertainty analysis.

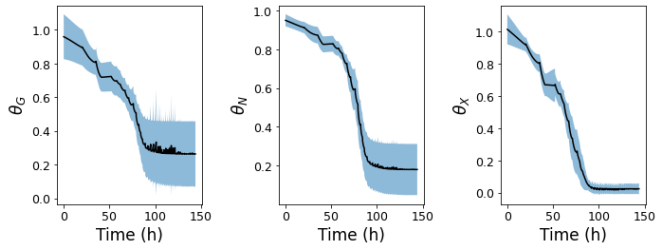

Fig. 8. Transfer learning: Time-varying parameters.

simulate fed-batch processes under alternate operating conditions. Time-varying parameter values were also penalized, which assisted in reducing model overfitting by avoiding dramatic changes between subsequent time-varying parameter values. The kinetic backbone used for the hybrid model construction incorporated as much physical information as possible, which meant that minimal additional data was required to train the ANN for both the original hybrid model and the transfer learning updated hybrid model. The ANN allowed the hybrid model to remain a function of observable state variables while being able to capture nonlinear dynamics that cannot be considered by the kinetic backbone. The use of transfer learning allowed for the optimal transfer of information to minimize the data requirement in training a hybrid model for the application to an alternative domain. Hyperparameters were fine-tuned iteratively to minimalize the overfitting of data and optimize the transfer of knowledge between the two domains. Uncertainty analysis was successfully conducted, facilitating future optimization and control strategies that ensure output quality and feasibility.

Overall, the transfer learning-based hybrid model showed good fitting capabilities, despite a low abundance of experimental data. The success of this modelling strategy paves the way towards the development of an effective digital twin for the modelling of fed-batch systems utilizing different substrates to identify an optimal control under uncertainty strategy for maximizing lutein production and yield for future fed-batch HCDC operation. The ability of this framework to incorporate a data-driven model and transfer learning with minimal data gives it substantial potential in applications to novel bioprocesses, where both experimental data and mechanistic understanding are limited.

REFERENCES

Harrington, R. (2018). Growing the bioeconomy: a national strategy to 2030, *HM Government*.

Mowbray, M., Savage, T., Chufan, W., Ziqi, S., Bovinille, A.C., Rio-Chanona, E.A.D., and Zhang, D. (2021). Machine learning for biochemical engineering: A review. *Biochemical Energy Journal*, volume 172, 108054.

Riesenberg, D., and Guthke, R. (1999). High-cell-density cultivation of microorganisms. *Applied Microbiology and Biotechnology*, volume 51, pages 422-430.

Rogers, A.W., Song, Z., Vega-Ramon, F., Jing, K., and Zhang, D. (2020). Investigating 'greyness' of hybrid model for bioprocess predictive modelling. *Biochemical Engineering journal*, volume 190.

Ruhe, Z.C., Low, D.A., and Hayes, C.S. (2013). Bacterial contact-dependent growth inhibition. *Trends in Microbiology*, volume 21, issue 5, pages 230-237.

Wang, D., Dapei, T., and Lei, L. (2017). Particle swarm optimization: an overview. *Soft Computing*, volume 22, pages 387-408.

Weiss, K., Taghi, M.K., and DingDing, W. (2016). A survey of transfer learning. *Journal of Big Data*, volume 3, article 9.

Wu, H., Shuhua, Q., Ruixiong, Y., Qihua, P., Yinghua, L., Chuanyi, Y., Ning, H., Song, H., and Xueping, L. (2024). Strategies for high cell density cultivation of Akkermansia muciniphila and its potential metabolism. *Journal of Clinical Microbiology*, volume 12, issue 1, e02386-23.

Xie, Y., Zhang, Z., Ma, Y., Liu, X., Miao, M., Ho, S.H., Chen, J., Leong, Y.K., and Chang, J.S. (2022). High-cell-density heterotrophic cultivation of microalga Chlorella sorokiniana FZU60 for achieving ultra-high lutein production efficiency. *Bioresource Technology*, volume 365, 128130.

Zinnecker, T., Reichl, U., and Genzel, Y. (2024). Innovations in cell culture-based influenza vaccine manufacturing – from static cultures to high cell density cultivations. *Human Vaccines & Immunotherapeutics*, volume 20, article 2373521.