Developing a machine learning integrated hybrid model for bioprocess kinetic modelling

Luca Riezzo[1], Harry Walter Kay[1], Dongda Zhang[1]*

[1]*Department of Chemical Engineering, The University of Manchester, Oxford Road, Manchester, M1 3AL, UK*

**dongda.zhang@manchester.ac.uk*

Highlights

- A hybrid model is constructed by integrating a data-driven model with a kinetic model.
- The model is used to simulate a complex bioreaction system for algal lutein production.
- The model can accurately predict process kinetics over a range of reaction conditions.

1. Introduction

Industrial biotechnology is at the spearhead of biochemical engineering, distributing commercially valued bioproducts for use in many different industries such as biorenewables, healthcare and medicine. To accelerate the industrialisation of a bioprocess, developing an accurate mathematical model capable of simulating and predicting the underlying biochemical reaction kinetics is a top priority. One the one hand, substantial effort has been made to develop kinetic models within the past decades[1]. However, due to the poor understanding of intricate bioreaction mechanisms, significant challenges are encountered when selecting the correct model structure to describe microbial cell behaviours and product formation dynamics. On the other hand, recent successes in machine learning have attracted great attention within the biochemical reaction engineering community[2]. However, a prerequisite to build an accurate data-driven model is the availability of a large amount of data, which is extremely time consuming to obtain for many fermentation and algal photo-production processes.

Therefore, to overcome the present challenges, this work proposes an advanced hybrid modelling technique that integrates kinetic modelling with machine learning. Specifically, a hybrid model consists of a kinetic model (i.e. prior knowledge) to describe the overall trajectory of the reaction system and a data-driven model to estimate the temporal evolution of key kinetic model parameters. Through this approach, it is possible to accurately quantify and predict the behaviours of complex biochemical processes. To demonstrate the performance of the proposed hybrid model, microalgal lutein production is used as a case study.

2. Methods

2.1 Introduction to kinetic model structure

First, a kinetic model must be developed based on prior knowledge as a reliable backbone of the hybrid model. The current work therefore simplified a sophisticate kinetic model constructed in a recent study as shown in Eq. 1-3[3]. In addition, the studied lutein synthesis system was carried out in a photobioreactor that contains a non-uniform light distribution throughout the reactor. To account the associated light attenuation effects in the reactor, Eq. 4 is the result of deriving the average light intensity used to account for light attenuation. Eq.1-4 therefore simulate biomass growth, nitrate consumption, and lutein production in a batch photobioreactor under different light intensities and nitrate concentrations,

$$\frac{dX}{dt} = \mu_m \cdot \frac{I}{I+k_s} \cdot \frac{N}{N+k_N} \cdot X - \mu_d \cdot X^2 \quad (1) \qquad \frac{dN}{dt} = -Y_N \cdot \mu_m \cdot \frac{I}{I+k_s} \cdot \frac{N}{N+k_N} \cdot X \quad (2)$$

$$\frac{dL}{dt} = Y_L \cdot \mu_m \cdot \frac{I}{I+k_s} \cdot \frac{N \cdot X}{N+k_{N,l}} - k_d \cdot X \cdot L \quad (3) \qquad \left(\frac{I}{I+k_s}\right)_{avg} = \ln\left(\frac{I+k_s}{I \cdot e^{-aXW}+k_s}\right) \cdot \frac{1}{aXW} \quad (4)$$

where μ_m is the biomass specific growth rate, μ_d is the biomass specific death rate, k_d is lutein consumption rate constant, Y_N is the nitrate yield coefficient, Y_L is the lutein yield coefficient, k_N is the nitrate half velocity constant for cell growth, $k_{N,L}$ is the nitrate half velocity constant for lutein synthesis, I is the incident light intensity, k_s and $k_{s,L}$ are the light saturation coefficient for cell growth and lutein synthesis, respectively, α is the algal light absorption coefficient, and W is the photobioreactor width.

2.2 Introduction to hybrid model construction

The hybrid model uses the same structure as the kinetic model. However, it further improves the model's accuracy by allowing parameters with the greatest uncertainty to vary over the batch process time course. This is achieved by constructing a machine learning model to estimate key parameters' values at each time step using the system's corresponding state[4]. Specifically, an artificial neural network was used in this work as the data-driven model. Initially, values of time-varying parameters at different time steps were estimated through the same parameter estimation framework used for kinetic model construction. Subsequently, an ANN was developed to correlate time-varying kinetic parameters with the three state variables (concentrations of biomass, nitrate, and lutein) and incident light intensity at each time step. ANN hyperparameter optimisation was also conducted to avoid overfitting. Finally, the ANN and the kinetic model were integrated to complete construction of the hybrid model. Figure 1 shows the concept of the hybrid model construction.

3. Results and discussion

The current study used *Desmodesmus sp*. as the algal species for lutein synthesis. When exploring the suitability of the current kinetic model structure, the proposed simple kinetic model can describe the overall bioprocess behaviours for most batches. However, clear model-process mismatch can be found and the values of some parameters differ greatly between each batch. As such, three kinetic parameters, μ_m, Y_N and Y_L were selected to build the hybrid model. Specifically, a 2-layer ANN was built to estimate values of these three parameters based on the four input variables. From Figure 2, it is observed that the hybrid model is capable of accurately simulating dynamics of the evolution of biomass, nitrate and lutein concentration with time. A deeper analysis also found that the model has good overall accuracy over a range of operating conditions, thus evidencing the performance of the hybrid model.

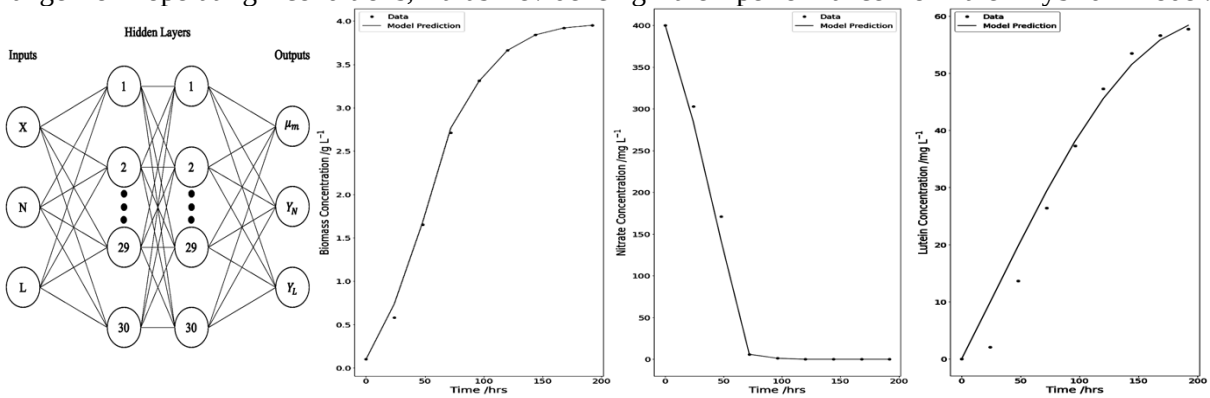

Figure 1: A schematic of the constructed ANN. Figure 2: Simulation result of the hybrid model.

4. Conclusions

Overall, combining the kinetic model and machine learning component, in this case an ANN, constructed a hybrid model that can successfully simulate a complex biochemical reaction system for different operating conditions. The introduction of prior mechanistic knowledge greatly reduces the amount of data required for a machine learning model construction, whilst the use of a data-driven model significantly simplifies the complexity of a kinetic model. Furthermore, the hybrid modelling framework proposed in this work is highly flexible and can be applied to general biochemical reaction processes. This work, therefore, paves a novel way to constructing accurate mathematical models for biochemical reaction system predictive modelling, optimisation, and upscaling.

References

[1] E. Lee, M. Jalalizadeh, and Q. Zhang, 'Growth kinetic models for microalgae cultivation: A review', *Algal Research*, vol. 12. Elsevier, pp. 497–512, Nov. 01, 2015.

[2] M. Mowbray *et al.*, 'Machine learning for biochemical engineering: A review', *Biochem Eng J*, vol. 172, Aug. 2021.

[3] E. del Rio-Chanona, N. r. Ahmed, D. Zhang, Y. Lu, and K. Jing, 'Kinetic modeling and process analysis for Desmodesmus sp. lutein photo-production', *AIChE Journal*, vol. 63, no. 7, pp. 2546–2554, Jul. 2017.

[4] A. W. Rogers, Z. Song, F. V. Ramon, K. Jing, and D. Zhang, 'Investigating "greyness" of hybrid model for bioprocess predictive modelling', *Biochem Eng J*, vol. 190, Jan. 2023.

Keywords: Bioreaction kinetics, machine learning, hybrid modelling, biorenewable production.