This item was submitted to [Loughborough's Research Repository](#) by the author.
Items in Figshare are protected by copyright, with all rights reserved, unless otherwise indicated.

Comparison of physics-based and data-driven modelling techniques for dynamic optimisation of fed-batch bioprocesses

PLEASE CITE THE PUBLISHED VERSION

PUBLISHER

Wiley

VERSION

AM (Accepted Manuscript)

PUBLISHER STATEMENT

This is the peer reviewed version of the following article: Del Rio-Chanona EA, Ahmed NR, Wagner J, Lu Y, Zhang D, Jing K. Comparison of physics-based and data-driven modelling techniques for dynamic optimisation of fed-batch bioprocesses. *Biotechnology and Bioengineering*. 2019;116:2971–2982, which has been published in final form at <https://doi.org/10.1002/bit.27131>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.

LICENCE

CC BY-NC-ND 4.0

REPOSITORY RECORD

Del Rio-Chanona, Ehecatl Antonio, Nur Rashid Ahmed, Jonathan Wagner, Yinghua Lu, Dongda Zhang, and Keju Jing. 2019. "Comparison of Physics-based and Data-driven Modelling Techniques for Dynamic Optimisation of Fed-batch Bioprocesses". figshare. <https://hdl.handle.net/2134/9579449.v1>.

Comparison of physics-based and data-driven modelling techniques for dynamic optimisation of fed-batch bioprocesses

Ehecatl Antonio Del Rio-Chanona^{1,‡}, Nur Rashid Ahmed^{2,‡}, Jonathan Wagner³, Yinghua Lu², Dongda Zhang^{1,4,*}, Keju Jing^{2,*}

1: Centre for Process Systems Engineering, Imperial College London, South Kensington Campus, London, SW7 2AZ, UK.

2: Department of Chemical and Biochemical Engineering, College of Chemistry and Chemical Engineering, Xiamen University, Xiamen 361005, China.

3: Department of Chemical Engineering, Loughborough University, Loughborough Leicestershire, LE11 3TU, UK.

4: Centre for Process Integration, University of Manchester, The Mill, Sackville Street, Manchester, M1 3AL, UK.

‡: These authors contributed equally to this work.

*: Corresponding authors, email: dongda.zhang@manchester.ac.uk, tel: 44 (0)161 306 5153 (Dongda Zhang); jkj@xmu.edu.cn, tel: 86 592 2186038 (Keju Jing).

Abstract

The development of digital bioprocessing technologies is critical to operate modern industrial bioprocesses. This study conducted the first investigation on the efficiency of using physics-based and data-driven models for the dynamic optimisation of long-term bioprocess. More specifically, this research exploits a predictive kinetic model and a cutting-edge data-driven model to compute open-loop optimisation strategies for the production of microalgal lutein during a fed-batch operation. Light intensity and nitrate inflow rate are used as control variables given their key impact on biomass growth and lutein synthesis. By employing different optimisation algorithms, several optimal control sequences were computed. Due to the distinct model construction principles and sophisticated process mechanisms, the physics-based and the data-driven models yielded contradictory optimisation strategies. The experimental verification confirms that the data-driven model predicted a closer result to the experiments than the physics-based model. Both models succeeded in improving lutein intracellular content by over 40% compared to the highest previous record; however, the data-driven model outperformed the kinetic model when optimising total lutein production and achieved an increase of 40-50%. This indicates the possible advantages of using data-driven modelling for optimisation and prediction of complex dynamic bioprocesses, and its potential in industrial bio-manufacturing systems.

Keywords: kinetic modelling; machine learning; artificial neural network; dynamic optimisation; fed-batch operation

1. Introduction

Microbe-based biosynthetic systems are a central feature of most industrial bioprocesses. For instance, bacteria and yeasts have been utilised extensively to yield a large variety of commercial products ranging from fuels and platform chemicals to pharmaceuticals (Bankar, Dhumal, Bhotmange, Bhagwat, & Singhal, 2014; Jing et al., 2018). Recently, attention has increasingly shifted towards microalgae and cyanobacteria to produce renewable biofuels and sustainable high-value products (Harun et al., 2018; Dongda Zhang & Vassiliadis, 2015). These approaches provide a competitive alternative to traditional synthetic routes and have considerable demand in the global market. In contrast to chemical processes, where most systems are operated at steady-state conditions with long operational periods (up to a year), most bioprocesses are conducted in batch or fed-batch modes, with constantly changing conditions and short operational periods. Such dynamic systems pose particular challenges for the modelling, control, and optimisation of the process (Bernard, Dochain, Genovesi, Gouze, & Guay, 2008; Del Rio-Chanona, Cong, Bradford, Zhang, & Jing, 2018). As a result, it is essential to develop specialised dynamic simulation and optimisation strategies which effectively predict and maximise the biological production of these processes.

Enabling the mathematical treatment of bioprocess operations at any spatial or temporal scale requires the construction of accurate physical models (Vatcheva, de Jong, Bernard, & Mars, 2006; Dongda Zhang et al., 2015). Physics-based models, in particular kinetic models, are constructed using available biological and chemical engineering knowledge of the system. Model parameters are based on distinct physical mechanisms with the expectation that the models can predict the behaviours of unknown processes reliably (Adesanya, Davey, Scott, & Smith, 2014; del Rio-Chanona, Ahmed, Zhang, Lu, & Jing, 2017). During the last century, classic kinetic models, *e.g.* the Monod model and the Droop model, have been specifically developed to describe bioprocesses and they have since been adapted to both bacterial

fermentation and algal photo-production systems (Lemesle & Mailleret, 2008; Sommer, 1991). Despite their success and wide application, the construction of accurate kinetic models is highly reliant on a good understanding of the system. More importantly, emerging bioprocesses are increasingly complex, requiring sophisticated and nonlinear model structures, causing numerical issues, *e.g.* parameter estimation and stiffness, which impose severe obstacles on the accuracy of physics-based models (Bernard et al., 2008; Dongda Zhang et al., 2015) and make them incompetent for process prediction.

An alternative approach to bioprocess modelling is the adoption of machine learning techniques and design data-driven models. A variety of data-driven models have been developed based on several machine learning approaches including artificial neural networks (del Rio-Chanona, Manirafasha, Zhang, Yue, & Jing, 2016), Gaussian Processes (Bradford, Schweidtmann, Zhang, Jing, & del Rio-Chanona, 2018), and recurrent neural networks (Lim & Shin, 2013). A major advantage of data-driven models is that they do not require extensive knowledge of the physical process of the underlying system. Several articles comparing the performance of data-driven and mechanistic models for simulating and predicting bioprocess behaviour have been published in recent years (Bradford et al., 2018; Del Rio-Chanona et al., 2018). However, most of these works were restricted to evaluating the capability of these approaches for process simulation, with little exploration of their potentials for the dynamic optimisation of bioprocesses. As bioprocesses are mainly operated periodically (around one week per batch), great improvements of process profitability could be achieved by identifying optimal control sequences for key operational factors prior to the execution of new processes. Thus, industrially relevant models should be able to accurately predict the behaviour of new bioprocesses over long operational periods in order to estimate the entire optimal operating conditions for the process, known as open-loop control.

The current study aims to compare the efficiency of physics-based modelling and data-driven modelling for the open-loop control of microalgal lutein production. The main reason for choosing this system as the case study is that algal photo-production systems are usually more complex than microbial fermentation processes (Y.-P. Xie et al., 2014). Furthermore, lutein is a high-value product which has received great attention in several commercial sectors. Previous studies have shown that lutein production is primarily affected by nitrate concentration and light intensity. However, the effects of these two factors on lutein synthesis and biomass growth are yet to be determined and inconsistent observations regarding lutein accumulation and cell growth have been presented in several publications when nitrate concentration and light intensity are fixed at different levels (del Rio-Chanona, Ahmed, et al., 2017). Therefore, it is worthy using simulation techniques to further investigate this system.

2. Materials and Methods

2.1 Introduction and construction of the physics-based model

The first simulation approach to explore the effects of light intensity and nitrate concentration on lutein synthesis and biomass growth is to construct a high-fidelity physics-based model using the available biological knowledge. In our recent work, we presented a rigorous kinetic model capable of effectively simulating and predicting biomass growth and lutein production under different operating conditions (del Rio-Chanona, Ahmed, et al., 2017). This model is presented in Eqs. 1(a)-1(f), and the parameter values and units are listed in Table 1.

In this model, Eq. 1(a) represents biomass growth rate, where the first and second terms on the right denote cell growth and cell decay, respectively. It is assumed that biomass growth is dependent on both light intensity and nitrate concentration, and these effects are quantified by the Aiba model and the Monod model, respectively. In contrast to growth, cell death is assumed to depend only on biomass concentration. Eq. 1(b) calculates the change of the culture's nitrate concentration, in which the first term on the right estimates the uptake of nitrate by cell growth,

whereas the second term represents the dosing of concentrated nitrate solution into the reactor. In our previous fed-batch experiments, addition of concentrated nitrate solution was commenced after 60 hours until the end of the experiment to continuously replenish nitrogen consumed by the culture. As the resulting increase in culture volume was negligible it is reasonable to assume fixed-volume fed-batch operation. Eq. 1(c) represents lutein production, and the first term on the right denotes lutein accumulation. As lutein is a primary carotenoid, it can be consumed to synthesise other secondary metabolites required for cell growth. This degradation is calculated by the second term on the right of Eq. 1(c) and is assumed to follow first order reaction kinetics.

$$\frac{dc_X}{dt} = u_m \cdot \frac{I(z)}{I(z) + k_s + \frac{I(z)^2}{k_i}} \cdot \frac{c_N}{c_N + K_N} \cdot c_X - u_d \cdot c_X \quad (1a)$$

$$\frac{dc_N}{dt} = -Y_{N/X} \cdot u_m \cdot \frac{I(z)}{I(z) + k_s + \frac{I(z)^2}{k_i}} \cdot \frac{c_N}{c_N + K_N} \cdot c_X + F_{in} \cdot c_{N,in} \quad (1b)$$

$$\frac{dc_L}{dt} = k_m \cdot \frac{I(z)}{I(z) + k_{sL} + \frac{I(z)^2}{k_{iL}}} \cdot \frac{c_N}{c_N + K_{NL}} \cdot c_X - k_d \cdot c_L \cdot c_X \quad (1c)$$

$$I(z) = I_0 \cdot (e^{-(\tau \cdot X + K_a) \cdot z} + e^{-(\tau \cdot X + K_a) \cdot (L-z)}) \quad (1d)$$

$$\frac{I(z)}{I + k_s + \frac{I^2}{k_i}} = \frac{1}{20} \cdot \sum_{n=1}^9 \left(\frac{I_0}{I_0 + k_s + \frac{I_0^2}{k_i}} + 2 \cdot \frac{I_n}{I_n + k_s + \frac{I_n^2}{k_i}} + \frac{I_L}{I_L + k_s + \frac{I_L^2}{k_i}} \right) \quad (1e)$$

$$\frac{I(z)}{I + k_{sL} + \frac{I^2}{k_{iL}}} = \frac{1}{20} \cdot \sum_{n=1}^9 \left(\frac{I_0}{I_0 + k_{sL} + \frac{I_0^2}{k_{iL}}} + 2 \cdot \frac{I_n}{I_n + k_{sL} + \frac{I_n^2}{k_{iL}}} + \frac{I_L}{I_L + k_{sL} + \frac{I_L^2}{k_{iL}}} \right) \quad (1f)$$

where c_X , c_N , c_L : concentration of biomass, nitrate, and lutein, respectively. I_0 and I_n : incident and local light intensity (distance of $\frac{n \cdot L}{10}$ from the photobioreactor (PBR) front surface), respectively.

Apart from involving more sophisticated metabolic pathways, compared to traditional bacteria fermentations, the modelling of algal photo-production systems is impeded by the non-uniform distribution of local light intensity throughout the PBR (D. Zhang, Chanona, Vassiliadis, & Tamburic, 2015). This phenomenon, called light attenuation, is caused by bubble scattering and light absorption by the algal cells. Although light attenuation can be approximated using the modified Lambert-Beer's law (presented in Eq. 1(d)), the resulting model involves both spatial and temporal dimensions, requiring a set of partial differential equations, which may lead to convergence issues and numerical intractability. As a result, the 10-step Trapezoidal rule (Eqs. 1(e)-1(f)) is used to eliminate the spatial dimension and simplify the model complexity for parameter estimation and process optimisation. Although Eqs. 1(e)-1(f) are based on flat-plate rather than tubular PBR geometry, the additional mathematical complexity of using cylindrical coordinates would cause great difficulty in parameter estimation. As a result, the tubular PBR in this work has been approximated as a column with a square cross section, illuminated from two opposite sides of the reactor, allowing the application of the Trapezoidal rule. A more detailed explanation for model construction, simplification, and verification can be found in our previous work (del Rio-Chanona, Ahmed, et al., 2017).

2.2 Introduction and construction of the data-driven model

Unlike for physics-based models, the pre-processing of original experimental data and identification of the desirable model structure (*i.e.* optimisation of model hyperparameters) represents a critical step in the construction of data-driven models. So far, several machine learning techniques have been adopted to simulate chemical and biochemical processes. The current study uses an artificial neural network (ANN) designed in our recent work (del Rio-Chanona, Fiorelli, et al., 2017), and found to be of high accuracy for process modelling and prediction. A brief introduction to the construction of this ANN is given here.

Fig. 1 shows the ANN structure. The network inputs consist of five key operational factors: concentrations of three state variables (biomass, nitrate, and lutein), incident light intensity, and nitrate inflow rate. The network outputs are the concentration changes of the three state variables within one time interval (12 hours in this work). Five previous datasets (Experiment 1-5 in Table 2) covering a broad spectrum of incident light intensity and nitrate inflow rate were used to train the ANN with another three (Experiment 6-8 in Table 2) for model testing. These datasets were also used to construct the kinetic model. Before training the ANN, it is essential to normalise the original experimental datasets due to the large magnitude differences between the values of each input variable. Several scaling methods can be used, but their suitabilities are directly dependent on the use of activation functions in the network. Specific to our previous work, a hyperbolic tangent activation was used due to its high performance for regression purposes. A standardisation approach was implemented which centres the data around zero and defines the standard deviation to be one. Once the data is reformulated, it is necessary to identify the optimal structure of the ANN.

Hyperparameters include the number of hidden layers, the number of neurons in each layer, and the type of connection between different layers (*e.g.* feedforward or recurrent). Although recurrent neural networks (RNNs) have been specifically developed to simulate time-series (Lim & Shin, 2013), our previous work employed a feedforward neural network in order to simplify the model structure. To maximise the model accuracy, three strategies were adopted. The first is to augment the datasets artificially to consolidate the model regression performance. Given the limited number of experimental datasets (5 datasets), 50 replications of artificial datasets are generated by carefully embedding Gaussian random noise into each original dataset, so that the network receives a sufficient amount of training information; helping to mitigate overfitting. The efficiency of this data augmentation strategy has been well demonstrated in our previous studies (del Rio-Chanona, Fiorelli, et al., 2017; del Rio-Chanona,

Manirafasha, et al., 2016). The second is to use a hyperparameter selection framework to seek optimal combinations between the number of hidden layers and the number of neurons in each layer. This framework calculates the trade-off between *variance* and *bias* under different model configurations, allowing a reduction in computational cost and the risk of over-fitting. The third strategy is the *k*-fold method, where a random selection of *k-1* from *k* datasets is used to train an ANN and the remaining dataset is used to evaluate the model prediction error. Then, another *k-1* datasets are chosen to repeat this procedure until the model with the least prediction error is identified. A more detailed explanation regarding the construction of this ANN can be found in the recent work (del Rio-Chanona, Fiorelli, et al., 2017). The final ANN used in the present work is comprised of 2 hidden layers, each containing 15 neurons.

2.3 Open-loop dynamic optimisation via physics-based model

To maximise lutein production using the physics-based model, two 144-hour fed-batch processes were designed and optimised. The dynamic equations (Eq. 1(a) to 1(f)) were modified accordingly to model this mode of operation. Using fed-batch processes, it is possible to treat the nitrate inflow rate and light intensity as the control variables, allowing the formulation of an optimal control problem. To maintain the experimental feasibility of this scenario, controls were allowed to vary only once per day. The optimal control problem (OCP 1) is the following:

$$\max_{F_{in}(t), L(t)} C_L(t_f)$$

subject to:

$$\text{Eq. 1(a) – Eq. 1(f)}$$

$$F^{lb} \leq F_{in}(t) \leq F^{ub}$$

$$L^{lb} \leq L(t) \leq L^{ub}$$

where the super-indices *lb* and *ub* represent lower and upper bounds, respectively.

To avoid unnecessary changes in nitrate inflow rate and light intensity, a quadratic penalty for changing control values was added to the objective function. OCP 1 was discretised through orthogonal collocation over finite elements, and the collocation points were placed according to a fifth order Radau quadrature. This transformation therefore yields a nonlinear programming (NLP) problem and was solved by the interior point solver IPOPT (Wächter & Biegler, 2006) in a Python environment. OCP 1 was solved in a multi-start framework using a Latin hypercube sampling scheme in the controls to achieve a high lying optimum. Given the complex structure and nonlinearity of this kinetic model, different optimal solutions were found with different starting points. The two best optimal solutions were selected for further experimental verification.

2.4 Open-loop dynamic optimisation via data-driven model

Similar to Section 2.3 where a physics-based model is used to solve the dynamic optimisation in OCP 1, it is also possible to use the ANN to identify two optimal control strategies. However, it is well known that neural networks present very many saddle points or local minima (Goodfellow, Bengio, & Courville, 2016), and therefore it is ill advised to use gradient-based optimisation algorithms directly. As ANNs are very rapid at inference, the current study applied a hybrid evolutionary algorithm (del Rio-Chanona et al., 2018; Estrada-Wiese, del Río-Chanona, & del Río, 2018) to initially optimise the ANN process model. Once the neighbourhood of a solution was found, a gradient based optimisation method was used to find the optimal inflow rate and light intensity. Thus, a new optimal control problem (OCP 2) was formulated by replacing the constraints Eq. 1(a) to 1(f) in OCP 1 with the current ANN. This problem was also programmed and solved in a Python environment. Other stochastic based operational search method (del Rio-Chanona, Manirafasha, et al., 2016) can also be used to optimise an ANN model.

2.5 Microalgae strain and pre-culture

The thermo-tolerant microalga *Desmodesmus* sp. F51 used in this study was generously donated by professor Ching-Nen Nathan Chen from National Sun Yet-sen University, Taiwan. The medium used for pre-culture was modified Bristol's medium (Yeh & Chang, 2012), with the following compositions (g L^{-1}): NaNO_3 , 0.75; $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 0.075; $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$, 0.025; K_2HPO_4 , 0.075; KH_2PO_4 , 0.175; NaCl , 0.025; $\text{FeCl}_3 \cdot 6\text{H}_2\text{O}$, and 1mL trace element solution containing (g L^{-1}): H_3BO_3 , 0.061; $\text{MnSO}_4 \cdot 7\text{H}_2\text{O}$, 0.169; $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$, 0.287; $(\text{NH}_4)_6\text{Mo}_7\text{O}_{24} \cdot 0.004\text{H}_2\text{O}$, 0.00124; $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$, 0.0025. Algal cells were grown for 5 days at 28 °C with an agitation rate of 300 rpm and continuous supply of 2.5% CO_2 in air at an aeration rate of 0.2 vvm. The culture was illuminated with continuous light at 60 μE and the light intensity was controlled using a Li – 250 Light meter (Li – COR In., Lincoln, NE).

2.6 Experimental verification and analytical procedures

Experimental verification of four optimally designed experiments (two estimated by the kinetic model, and the other two by the ANN) were executed in a 1L glass tubular photobioreactor (15.5 cm in length and 9.5 cm in diameter) with an artificial external light source (14W TL 5 tungsten filament lamps, Philip Co., China) fixed on both sides of the reactor. Initial biomass concentration and initial nitrate concentration in all experiments were fixed at 0.27 g L^{-1} and 9 mM, respectively. Influent nitrate concentration and culture temperature were kept constant at 0.1 M and 35 °C, respectively, for all experiments. The culture was continuously aerated with 2.5% CO_2 (carbon source) in air at 0.2 vvm and the pH was controlled at 7.5 with 300 rpm agitation rate. Both nitrate feeding rate and the light intensity were changed daily from the beginning until the end of the process, except for one of the ANN derived optimal experiments, which was specifically designed to commence nitrate feeding only from Day 3. All the experiments were duplicate and the optimal control actions are listed in Table 3.

Measurement of biomass concentration and nitrate concentration: Biomass concentrations (DCW) were determined every 24 hours by measuring the optical density (OD_{685}) of the culture

using a UV/VIS spectrophotometer (UV-1780 Shimadzu, Japan) after proper dilution of samples using deionised water. The OD₆₈₅ values were converted to DCW using an calibration equation from (Y. Xie et al., 2013). Nitrate concentration was analysed by a spectrophotometer method (Ho, Kondo, Hasunuma, & Chang, 2013). The supernatant liquid sample after centrifugation was filtered using a 0.22 µm pore size filter and then diluted with deionised water for each sample. The optical density of each sample was measured using the UV-1780 spectrophotometer at a wavelength of 220 nm (OD_{220 nm}). The calibration between nitrate concentration and absorbance (OD_{220 nm}) was established using sodium nitrate as the baseline.

Analysis of lutein concentration: Lutein concentration was analysed using a previously modified protocol (Sánchez Mirón, Cerón García, García Camacho, Molina Grima, & Chisti, 2002). 10 mg of dry lyophilised biomass was added to 1 mL aqueous KOH solution (60% w/w) and disturbed for 7 minutes by a bead-beater. The mixture was kept at 40 °C for 40 minutes on a water bath. The carotenoids were repeatedly extracted from the mixture by adding 2 mL diethyl ether until the yellow colour of the supernatant had fully faded. The combined carotenoid solvent was dried with nitrogen gas and mixed again with 3 mL acetone for carotenoid estimation. The analysis of lutein content was carried out using a high performance liquid chromatography (Agilaten technology 1200, USA). A YMC carotenoid column with a flow rate of 1 mL min⁻¹ at 25 °C was operated to separate carotenoids. The mobile phase consisted of (A) methanol with 3% of distilled H₂O containing 0.05M ammonium acetate, (B) 100% tetra butyl methyl ether (TBME). The mobile phases (A&B) contain both 0.1% (w/v) butylated hydroxytoluene and 0.05% triethylamine. Lutein content was estimated by measuring absorbance at 450 nm wavelength.

3. Results and discussion

The four optimal control strategies identified in this study are summarised in Table 3. In our previous research, eight experiments were carried out to investigate effects and interactions of

nitrate concentration and light intensity on algal biomass growth and lutein production. These experiments were used as a benchmark to test the performance of the proposed optimal control strategies. A summary of these eight experiments is listed in Table 2. From the table, it is seen that Exp. 8 yields the highest lutein production (4.93 mg L^{-1}) and biomass concentration (2.69 g L^{-1}), and the lowest final nitrate concentration (1.75 g L^{-1} , highest nitrate conversion efficiency). In terms of lutein cellular content, which critically affects downstream separation costs and process profitability, Exp. 3 shows the best result (2.34 mg g^{-1}), 15.4% higher than for Exp. 8 (1.98 mg g^{-1}). Exp. 3 also has a similar lutein production compared to Exp. 8 (91.7%) but requires a lower light intensity. However, nitrate usage in this experiment is substantially higher than Exp. 8 (approx. 9 times higher), indicating its low nitrate utilisation efficiency and high raw material cost.

3.1 Results of physics-based model derived optimal control strategies

Fig. 2 shows the results of the two optimal control strategies estimated by the physical model and Fig. 3 presents the corresponding optimal control sequences. To facilitate comparison between the different control strategies, the current experimental results are listed in Table 4 together with the best previous experiment (Exp. 8). From Fig. 2, it can be seen that the kinetic model fails to accurately estimate nitrate accumulation in both experiments, as the actual increase in nitrate concentration towards the end of the experiments was not predicted. In contrast, it is observed that biomass growth and lutein production closely follow the the kinetic model predictions over the entire time course of the first experiment. Nonetheless, in the second experiment, the predictions significantly underestimated the real process during the initial operating stage (first 60 hours), but caught up with the actual process trajectory for the second half of the experiment (after 60 hours).

There are several factors that may have contributed to the discrepancy between model prediction and experimental verification. The kinetic model was fitted using data from previous

experiments in which nitrate supply was always sufficient (*i.e.* final nitrate concentrations are around 2.0 g L⁻¹). However, nitrate concentrations in the new experiments were much lower (0.2 to 0.6 g L⁻¹), and thus, the conditions for biomass growth may have switched from nitrate-sufficient to nitrate-limiting, changing the metabolic activities of the cells. From a mechanistic modelling perspective, this transition is accomplished by either changing the value of the nitrate yield coefficient ($Y_{N/X}$) or shifting the model structure from the Monod to the Droop model (Vatcheva et al., 2006). As the current model is exclusively built upon the Monod model, it may not be able to accurately estimate the trend of nitrate uptake under nitrate-limiting conditions, consistent with the results shown in Fig. 2. Nonetheless, due to the complex process mechanism, it is difficult to rule out other potential reasons such as the effect of frequent changes of light intensity on nitrate uptake at this early research stage. More experiments must be carried out to clarify this observation.

Another reason to explain the plant-model mismatch arises from the difficulty in simulating the microbial lag phase. At the beginning of an experiment, algal cells have to adapt into a new environment, causing a delay for cell growth and nutrient consumption. This period is known as the lag phase and is affected by many environmental factors which are difficult to quantify. The rapid biomass growth and nitrate uptake during the initial stage of the second experiment (Fig. 2(b) and 2(d)) compared to the simulation results suggest that the current model underestimates the speed of cell adaption (short lag phase) under this specific cultivation condition.

Comparing the two optimised experiments with the best previous experiment (Exp. 8, Table 4), shows that both optimised experiments yielded significantly higher lutein intracellular contents (by 38.4% and 53.5%, respectively) and reduced nitrate use by over 65%. As mentioned already, higher lutein cellular content are expected to result in reduced downstream separation costs, and therefore, from a process separation and material consumption point of

view, both experiments could result in significant increases in process profitability. However, only the first experiment resulted in an increase of lutein production (by 36.3%), whereas the second experienced a slight decrease (by 3.2%). This is attributed to the fact that the current model predicts the first experiment more accurately than the second, and hence the optimal control sequences implemented in the first experiment are more effective. Despite this fact, it is important to realise that the kinetic model estimated similar optimal control principles for both experiments, combining a high incident light intensity with a low nitrate concentration (*i.e.* slow nitrate replenishment, as shown in Fig. 3).

3.2 Results of data-driven model derived optimal control strategies

Fig. 4 shows the results of the two open-loop optimisation experiments designed by the ANN and the comparison between model prediction and experimental verification. The optimal control actions implemented in these experiments are presented in Fig. 5. From Fig. 4(a)-(f), it is seen that the trajectories predicted by the ANN model are relatively close to the actual experimental results for all the three state variables over the entire time course of operation. Compared to the kinetic model, this result confirms a high predictive power of the current data-driven model. It should also be noted that the final nitrate concentrations in both experiments are higher than for the best previous experiment (Exp. 8), indicating that nitrate concentrations are sufficient in both experiments and do not reach nitrate-limiting conditions.

The results of these two experiments are further summarised in Table 4 and compared to Exp.8., It can be seen that both experiments succeed in improving lutein production and cellular lutein content by approx. 50%, suggesting the advantage of applying data-driven modelling and evolutionary optimisation algorithms to bioprocess simulation and optimisation. When comparing the two experiments, it is found that the first experiment achieves a 10% higher lutein production, a 30% reduction in nitrate usage, but a 10% lower lutein cellular content than the second experiment. The main difference between the two experiments is that for experiment

1, light intensity and nitrate inflow rate were allowed to change from Day 1, whereas for the second, the addition of nitrate was only allowed to commence from Day 3. The difference of their results therefore reflects the importance of initial operating conditions and early stage system control on bioprocess optimisation and design. However, as a high lutein cellular content is critical to downstream separation, the trade-off between lutein production and lutein cellular content has to be balanced through more detailed economic and technical assessments.

3.3. Comparison of different model based optimal control strategies

From Table 4, it can be observed that compared to the physics-based model derived control strategies, the data-driven model based strategies clearly show a better result in terms of lutein production and lutein cellular content. Both of the ANN based strategies yield an increase of 40% to 50% in lutein production, whilst only one of the physical model based strategies leads to an increase in lutein production by 35%, with the other producing less lutein than the best previous, unoptimised study. Despite this, there is no obvious increase in biomass concentration for any of the four experiments and the increase in lutein production can be solely attributed to increases in the lutein cellular content by 40% to 50%. This suggests that all the optimal control strategies effectively enhance the conversion efficiency from nitrate to lutein, benefiting downstream separation and raw material utilisation. The current study also concludes that data-driven models can be effectively used to substitute physics-based models for process prediction and optimisation of complex systems, *e.g.* a general biological system, with poor understanding of the underlying process mechanisms.

Another interesting observation is that the current physics-based model and data-driven model identify opposite optimal control concepts. For both kinetic-model driven experiments, conditions with high light intensities (800 μE to 1000 μE) but low nitrate concentration (approx. 0.5 g L⁻¹) are preferred, whereas the ANN-driven experiments appear to favour high nitrate concentration (approx. 4.0 g L⁻¹) but low light intensities (400 μE to 600 μE). This discrepancy

implies that either the two models have completely distinct understandings of the same system due to their own predictive ability, or that the optimisation algorithms selected in this work are not effective enough to handle the model complexity. In fact, given the high nonlinearity of complex physics-based models, the existence of multiple local optima can impede the efficiency of deterministic optimisation solvers, returning low quality process control strategies. For instance, the two optimal control sequences estimated by the kinetic model were generated at different starting points, meaning that they are local optima. In contrast, although the evolutionary algorithms used for data-driven model optimisation cannot guarantee global optimality; they may provide good optimal solutions in practice. As a result, different combinations of model construction strategies and optimisation algorithms should be explored for the modelling, optimisation, and design of complex systems.

To further investigate whether the better results of ANN are attributed to its higher predictive capability or a more efficient optimisation method, the kinetic model was used to predict lutein productions under the two optimal control sequences estimated by the ANN. If the kinetic model can well predict these experimental results, it means that the optimisation method used by the kinetic model is not efficient and more competent algorithms should be adopted; otherwise if the kinetic model cannot predict accurately the process behaviours, it means that the kinetic model's predictive power is not as good as that of the ANN. Through this approach, it is found that the kinetic model calculates a lutein production around 4.2 mg L^{-1} in both cases, much lower than the experimental results (6.90 to 7.53 mg L^{-1}). Therefore, it is concluded that although this physics-based model shows good fitting and predictive capability within a certain operating region, it is not capable of predicting process dynamics over a broader range of operating conditions.

In addition, it is worth exploring the physical insight behind the optimal control actions discovered by the ANN and the physical model. Lutein is known as an intracellular nitrogen storage and a non-photochemical quenching agent (Sun, Li, Zhou, & Jiang, 2015). It is overproduced under high light intensities or high nitrate concentrations. However, it is also consumed by cells to synthesise other metabolites during cell growth. Thus, a rapid biomass growth may result in a quick consumption of cellular lutein concentration. The physical model based optimisation strategies aim to restrain cell growth by providing a nitrate-limiting environment; whilst the data-driven model derived optimal control actions try to balance this trade-off by lowering the light intensity. From experimental verification, the ANN based strategies are more effective. In terms of the significant difference of light intensity at the 24th to the 48th hours between the two ANN derived experiments, it is noted that nitrate was not replenished in the second experiment until the 48th hour, whereas for the first experiment nitrate was fed from the beginning. To compensate for this lack of nitrogen influx, the light intensity was significantly reduced in the second experiment to avoid the rapid growth of biomass (consuming more nitrogen storage *e.g.* lutein). On the contrary, since nitrate influent was added from the beginning of the first experiment, a higher light intensity was applied to facilitate biomass growth and achieve a higher total lutein production.

4. Conclusion

Overall, both the physics-based and machine learning based models successfully improved lutein cellular content compared to the best previous experiment. In addition, optimal control strategies identified by the mechanistic model greatly reduced the use of nitrate, whereas experiments controlled by the data-driven model yielded significant increases in the total lutein production. By comparing the results, it is concluded that data-driven models can serve as a competent alternative to traditional mechanistic models for process prediction, optimisation, and design, and more attention should be given to their applications in generic complex systems.

Furthermore, to guarantee success, it is vital to ensure that the open-loop optimisation is conducted within the operational region for which the system model is valid. Nonetheless, it is a challenging task to identify model validity prior to the execution of new experiments. Hence, in practice model based bioprocess dynamic optimisation may have to follow several iterations until the model's predictive capability is confirmed.

Specific to lutein production, this study yields the highest lutein production and intracellular concentration ever reported for the current microalgae strain. Nonetheless, there are other microalgae species showing a higher lutein cellular content up to 7.6 mg g⁻¹ compared to the value of 3.98 mg g⁻¹ observed in this work (Fernández-Sevilla, Acién Fernández, & Molina Grima, 2010). However, most of these works used a smaller PBR with a volume ~ 200 mL compared to the current 1 L reactor. As light attenuation aggravates exponentially with the size of PBR, it is difficult to judge if these higher yields can be reproduced in a larger PBR. Moreover, the strain selected in this work is a thermo-tolerant microalga *Desmodesmus* sp. F51 which can grow in tropical and subtropical regions where temperature can easily surpass 45°C during the daytime, showing an additional advantage for industrial application (Y. Xie et al., 2013).

It is also important to highlight that the current proposed modelling technologies can be directly transferred to other generic photo-production and fermentation processes for production of high-value biorenewables and sustainable biofuels. In particular, model based process design is considered as an essential procedure for future bioprocess manufacturing, control and scale-up. In order to identify a competent process modelling technique, this work carried out the first thorough study to compare performance of different modelling methods for long-term bioprocess prediction and optimisation. Moreover, to avoid potential risks in the lack of datasets for data-driven model construction and the issue of model-plant mismatch for kinetic model development, one approach to be investigated in the future is to adopt advanced online

optimisation frameworks. Both model predictive control (MPC) which has been widely used in chemical processes (Mayne, 2014) and reinforcement learning which has been recently applied to real-time bioprocess optimisation should be given priorities in future study (Petsagkourakis, Sandoval, Bradford, Zhang, & del Rio-Chanona, 2019). With an online optimisation mode, optimal control sequences can be synchronised during the ongoing process and model-plant mismatch can be kept minimum. Moreover, newly collected data can be directly used to enhance the data-driven model accuracy. In particular, economic model predictive control (del Rio-Chanona, Zhang, & Vassiliadis, 2016), which is specifically designed for dynamic systems, can be considered as an efficient strategy for bioprocess optimisation. It is envisaged that through the integration of state-of-the-art online optimal control methodologies, accuracy and reliability of both physical model and machine learning based dynamic process optimisation will be greatly improved.

Acknowledgement

This work was supported by the National Natural Science Foundation of China No. 21776232 and No. 21736009. It also received the funding from the EPSRC project (EP/P016650/1). The authors also sincerely appreciate the academic collaborations between Xiamen University, Imperial College London, University of Manchester, and Loughborough University.

References

- Adesanya, V. O., Davey, M. P., Scott, S. A., & Smith, A. G. (2014). Kinetic modelling of growth and storage molecule production in microalgae under mixotrophic and autotrophic conditions. *Bioresource Technology*, 157, 293–304. <https://doi.org/10.1016/j.biortech.2014.01.032>
- Bankar, S., Dhumal, V., Bhotmange, D., Bhagwat, S., & Singhal, R. (2014). Empirical predictive modelling of poly- ϵ -lysine biosynthesis in resting cells of *Streptomyces noursei*. *Food Science and Biotechnology*, 23(1), 201–207. <https://doi.org/10.1007/s10068-014->

- Bernard, O., Dochain, D., Genovesi, A., Gouze, J.-L., & Guay, M. (2008). *Bioprocess Control*. (D. Dochain, Ed.). London, UK: ISTE. <https://doi.org/10.1002/9780470611128>
- Bradford, E., Schweidtmann, A. M., Zhang, D., Jing, K., & del Rio-Chanona, E. A. (2018). Dynamic modeling and optimization of sustainable algal production with uncertainty using multivariate Gaussian processes. *Computers & Chemical Engineering*, 118, 143–158. <https://doi.org/10.1016/j.compchemeng.2018.07.015>
- del Rio-Chanona, E. A., Ahmed, N. rashid, Zhang, D., Lu, Y., & Jing, K. (2017). Kinetic modeling and process analysis for *Desmodesmus* sp. lutein photo-production. *AIChE Journal*, 63(7), 2546–2554. <https://doi.org/10.1002/aic.15667>
- Del Rio-Chanona, E. A., Cong, X., Bradford, E., Zhang, D., & Jing, K. (2018). Review of advanced physical and data-driven models for dynamic bioprocess simulation: Case study of algae–bacteria consortium wastewater treatment. *Biotechnology and Bioengineering*, bit.26881. <https://doi.org/10.1002/bit.26881>
- del Rio-Chanona, E. A., Fiorelli, F., Zhang, D., Ahmed, N. rashid, Jing, K., & Shah, N. (2017). An efficient model construction strategy to simulate microalgal lutein photo-production dynamic process. *Biotechnology and Bioengineering*, 114(11), 2518–2527. <https://doi.org/10.1002/bit.26373>
- del Rio-Chanona, E. A., Manirafasha, E., Zhang, D., Yue, Q., & Jing, K. (2016). Dynamic modeling and optimization of cyanobacterial C-phycocyanin production process by artificial neural network. *Algal Research*, 13, 7–15. <https://doi.org/10.1016/j.algal.2015.11.004>
- del Rio-Chanona, E. A., Wagner, J. L., Ali, H., Fiorelli, F., Zhang, D., & Hellgardt, K. (2018). Deep learning-Based surrogate modeling and optimization for microalgal biofuel production and photobioreactor design. *AIChE Journal*, aic.16473.

<https://doi.org/10.1002/aic.16473>

del Rio-Chanona, E. A., Zhang, D., & Vassiliadis, V. S. (2016). Model-based real-time optimisation of a fed-batch cyanobacterial hydrogen production process using economic model predictive control strategy. *Chemical Engineering Science*, 142, 289–298. <https://doi.org/10.1016/j.ces.2015.11.043>

Estrada-Wiese, D., del Río-Chanona, E. A., & del Río, J. A. (2018). Stochastic optimization of broadband reflecting photonic structures. *Scientific Reports*, 8(1), 1193. <https://doi.org/10.1038/s41598-018-19613-6>

Fernández-Sevilla, J. M., Acien Fernández, F. G., & Molina Grima, E. (2010). Biotechnological production of lutein and its applications. *Applied Microbiology and Biotechnology*, 86(1), 27–40. <https://doi.org/10.1007/s00253-009-2420-y>

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. Retrieved from <http://www.deeplearningbook.org>

Harun, I., Del Rio-Chanona, E. A., Wagner, J. L., Lauersen, K. J., Zhang, D., & Hellgardt, K. (2018). Photocatalytic Production of Bisabolene from Green Microalgae Mutant: Process Analysis and Kinetic Modeling. *Industrial & Engineering Chemistry Research*, 57(31), 10336–10344. <https://doi.org/10.1021/acs.iecr.8b02509>

Ho, S.-H., Kondo, A., Hasunuma, T., & Chang, J.-S. (2013). Engineering strategies for improving the CO₂ fixation and carbohydrate productivity of *Scenedesmus obliquus* CNW-N used for bioethanol fermentation. *Bioresource Technology*, 143, 163–171. <https://doi.org/10.1016/j.biortech.2013.05.043>

Jing, K., Tang, Y., Yao, C., del Rio-Chanona, E. A., Ling, X., & Zhang, D. (2018). Overproduction of L-tryptophan via simultaneous feed of glucose and anthranilic acid from recombinant *Escherichia coli* W3110: Kinetic modeling and process scale-up. *Biotechnology and Bioengineering*, 115(2), 371–381. <https://doi.org/10.1002/bit.26398>

531 Lemesle, V., & Mailleret, L. (2008). A mechanistic investigation of the algae growth “Droop”
 532 model. *Acta Biotheoretica*, 56(1–2), 87–102. <https://doi.org/10.1007/s10441-008-9031-3>

533 Lim, H. C., & Shin, H. S. (2013). *Fed-Batch Cultures*. Cambridge: Cambridge University Press.
 534 <https://doi.org/10.1017/CBO9781139018777>

535 Mayne, D. Q. (2014). Model predictive control: Recent developments and future promise.
 536 *Automatica*, 50(12), 2967–2986. <https://doi.org/10.1016/j.automatica.2014.10.128>

537 Petsagkourakis, P., Sandoval, I. O., Bradford, E., Zhang, D., & del Rio-Chanona, E. A. (2019).
 538 Reinforcement Learning for Batch Bioprocess Optimization. Retrieved from
 539 <http://arxiv.org/abs/1904.07292>

540 Sánchez Mirón, A., Cerón García, M.-C., García Camacho, F., Molina Grima, E., & Chisti, Y.
 541 (2002). Growth and biochemical characterization of microalgal biomass produced in
 542 bubble column and airlift photobioreactors: studies in fed-batch culture. *Enzyme and*
 543 *Microbial Technology*, 31(7), 1015–1023. [https://doi.org/10.1016/S0141-](https://doi.org/10.1016/S0141-0229(02)00229-6)
 544 [0229\(02\)00229-6](https://doi.org/10.1016/S0141-0229(02)00229-6)

545 Sommer, U. (1991). A Comparison of the Droop and the Monod Models of Nutrient Limited
 546 Growth Applied to Natural Populations of Phytoplankton. *Functional Ecology*, 5(4), 535.
 547 <https://doi.org/10.2307/2389636>

548 Sun, Z., Li, T., Zhou, Z., & Jiang, Y. (2015). Microalgae as a Source of Lutein: Chemistry,
 549 Biosynthesis, and Carotenogenesis. In *Advances in Biochemical Engineering* (Vol. 153,
 550 pp. 37–58).

551 Vatcheva, I., de Jong, H., Bernard, O., & Mars, N. J. I. (2006). Experiment selection for the
 552 discrimination of semi-quantitative models of dynamical systems. *Artificial Intelligence*,
 553 170(4–5), 472–506. <https://doi.org/10.1016/j.artint.2005.11.001>

554 Wächter, A., & Biegler, L. T. (2006). On the implementation of an interior-point filter line-
 555 search algorithm for large-scale nonlinear programming. *Mathematical Programming*,

106(1), 25–57. <https://doi.org/10.1007/s10107-004-0559-y>

Xie, Y.-P., Ho, S.-H., Chen, C.-Y., Chen, C.-N. N., Liu, C.-C., Ng, I.-S., ... Lu, Y.-H. (2014). Simultaneous enhancement of CO₂ fixation and lutein production with thermo-tolerant *Desmodesmus* sp. F51 using a repeated fed-batch cultivation strategy. *Biochemical Engineering Journal*, 86, 33–40. <https://doi.org/10.1016/j.bej.2014.02.015>

Xie, Y., Ho, S.-H., Chen, C.-N. N., Chen, C.-Y., Ng, I.-S., Jing, K.-J., ... Lu, Y. (2013). Phototrophic cultivation of a thermo-tolerant *Desmodesmus* sp. for lutein production: Effects of nitrate concentration, light intensity and fed-batch operation. *Bioresource Technology*, 144, 435–444. <https://doi.org/10.1016/j.biortech.2013.06.064>

Yeh, K.-L., & Chang, J.-S. (2012). Effects of cultivation conditions and media composition on cell growth and lipid productivity of indigenous microalga *Chlorella vulgaris* ESP-31. *Bioresource Technology*, 105, 120–127. <https://doi.org/10.1016/j.biortech.2011.11.103>

Zhang, D., Chanona, E. A. D.-R., Vassiliadis, V. S., & Tamburic, B. (2015). Analysis of green algal growth via dynamic model simulation and process optimization. *Biotechnology and Bioengineering*, 112(10). <https://doi.org/10.1002/bit.25610>

Zhang, D., Dechatiwongse, P., Del-Rio-Chanona, E. A., Hellgardt, K., Maitland, G. C., & Vassiliadis, V. S. (2015). Analysis of the cyanobacterial hydrogen photoproduction process via model identification and process simulation. *Chemical Engineering Science*, 128, 130–146. <https://doi.org/10.1016/j.ces.2015.01.059>

Zhang, D., & Vassiliadis, V. S. (2015). *Chlamydomonas reinhardtii* Metabolic Pathway Analysis for Biohydrogen Production under Non-Steady-State Operation. *Industrial & Engineering Chemistry Research*, 54(43), 10593–10605. <https://doi.org/10.1021/acs.iecr.5b02034>

581

Table 1: Parameters in the current kinetic model

Parameter	Value	Parameter
u_m [h^{-1}]	0.152	maximum cell specific growth rate
u_d [h^{-1}]	5.95×10^{-3}	cell specific decay rate
K_N [mg L^{-1}]	30.0	nitrate half-velocity constant
$Y_{N/X}$ [mg g^{-1}]	305.0	nitrate yield coefficient
k_m [$\text{mg g}^{-1} \text{h}^{-1}$]	0.350	maximum lutein synthesis rate constant
k_d [$\text{L mg}^{-1} \text{h}^{-1}$]	3.71×10^{-3}	lutein degradation rate constant
K_{NL} [mg L^{-1}]	10.0	nitrate half-velocity constant for lutein synthesis
k_s [μE]	142.8	light saturation term for cell growth
k_i [μE]	214.2	light inhibition term for cell growth
k_{sL} [μE]	320.6	light saturation term for lutein synthesis
k_{iL} [μE]	480.9	light inhibition term for lutein synthesis
τ [$\text{m}^2 \text{g}^{-1}$]	0.120	cell absorption coefficient
K_a [m^{-1}]	0.0	bubble scattering coefficient
F_{in} [L h^{-1}]	--	nitrate influent flow rate
$c_{N,in}$ [M]	0.1	nitrate influent concentration
L [m]	0.084	width of the photobioreactor

582

583

584

Table 2: Previous experimental results. Nitrate inflow rate was fixed at 3.0 mL h⁻¹ in all cases.

Lutein content is the average lutein cellular content in each experiment (mg lutein per g

biomass). All experiments last for 6 days (144 hours). Initial biomass concentration was kept

same in all experiments.

Experiment	1	2	3	4	5	6	7	8
Incident light intensity, μE	150	300	480	600	150	300	480	600
Nitrate inflow con., g L ⁻¹	42.5	42.5	8.5	42.5	8.5	8.5	42.5	8.5
Final nitrate con., g L ⁻¹	12.7	13.2	16.2	15.8	2.21	2.13	1.89	1.75
Final biomass con., g L ⁻¹	2.04	2.74	2.26	2.45	1.54	2.21	2.36	2.69
Lutein production, mg L ⁻¹	2.11	3.77	4.52	4.41	2.36	3.47	4.56	4.93
Lutein content, mg g ⁻¹	0.97	1.44	2.34	2.08	1.60	1.96	2.01	1.98

Table 3: Optimal control actions implemented in the current study for fed-batch process optimisation. “Kinetic 1” and “Kinetic 2” denote the first and second optimal experiment determined by the physics-based model, respectively. “ANN 1” and “ANN 2” denote the first and second optimal experiment determined by the data-driven model, respectively. All fed-batch processes are operated for 6 days, with the nitrate inflow rate and incident light intensity changed once per day.

Optimal nitrate inflow rate	Kinetic 1	Kinetic 2	ANN 1	ANN 2
Nitrate inflow rate on Day 1, mg h ⁻¹	1.70	1.82	7.25	0.0
Nitrate inflow rate on Day 2, mg h ⁻¹	2.25	2.55	4.44	0.0
Nitrate inflow rate on Day 3, mg h ⁻¹	4.40	4.86	25.8	101.2
Nitrate inflow rate on Day 4, mg h ⁻¹	8.22	6.01	30.3	76.9
Nitrate inflow rate on Day 5, mg h ⁻¹	9.21	6.25	25.5	33.5
Nitrate inflow rate on Day 6, mg h ⁻¹	8.64	5.79	25.5	30.4
Optimal incident light intensity	Kinetic 1	Kinetic 2	ANN 1	ANN 2
Incident light intensity on Day 1, μE	350	370	580	800
Incident light intensity on Day 2, μE	880	800	610	240
Incident light intensity on Day 3, μE	1010	880	500	610
Incident light intensity on Day 4, μE	960	850	550	610
Incident light intensity on Day 5, μE	930	820	610	450
Incident light intensity on Day 6, μE	960	800	610	500

Table 4: Comparison between the current optimal control experiments and the previous best experiment (Exp. 8). “Kinetic 1” and “Kinetic 2” denote the first and second optimal experiment identified by the physics-based model, respectively. “ANN 1” and “ANN 2” denote the first and second optimal experiment identified by the data-driven model, respectively.

Experimental result	Exp. 8	Kinetic 1	Kinetic 2	ANN 1	ANN 2
Final biomass con., g L ⁻¹	2.69	2.83	2.07	2.35	1.97
Final nitrate con., g L ⁻¹	1.75	0.58	0.54	3.67	4.91
Lutein production, mg L ⁻¹	4.93	6.72	4.77	7.53	6.90
Lutein cellular content, mg g ⁻¹	1.98	2.74	3.04	2.82	3.01
Maximum lutein content, mg g ⁻¹	2.16	4.27	3.98	3.27	3.50
Performance of the current optimal experiments compared to Exp. 8					
Increase of biomass con., %	--	5.2%	No	No	No
Reduction of nitrate usage, %	--	66.9%	69.1%	No	No
Increase of lutein production, %	--	36.3%	No	52.7%	40.0%
Increase of lutein content, %	--	38.4%	53.5%	42.4%	52.0%

Figure 1: Structure of the artificial neural network. Δc_i denotes the change of concentration of state variable i from the current time step to the next time step (12 hours).

Figure 2: Results of the two fed-batch optimal control experiments estimated by the kinetic model. (a), (c), (e): concentrations of biomass, nitrate, and lutein, respectively, in the first optimal experiment. (b), (d), (f): concentrations of biomass, nitrate, and lutein, respectively, in the second optimal experiment. Points: experimental measurements (experimental verification). Lines: physical model predictions.

Figure 3: Optimal control sequences estimated by the physics-based model. (a): nitrate inflow rate in the two optimal experiments. (b): incident light intensity in the two optimal experiments. Solid line: optimal control actions for the first experiment. Dashed line: optimal control actions for the second experiment. The control actions are changed once per day.

Figure 4: Results of the two fed-batch optimal control experiments designed by the data-driven model. (a), (c), (e): concentrations of biomass, nitrate, and lutein, respectively, in the first experiment. (b), (d), (f): concentrations of biomass, nitrate, and lutein, respectively, in the second experiment. Points: experimental measurements (experimental verification). Lines: data-driven model prediction result. ANN only predicts concentrations of state variables once per 12 hours. For visualisation purpose, these discrete points are jointed into a solid line.

Figure 5: Optimal control sequences identified by ANN. (a): nitrate inflow rate in the two experiments. (b): incident light intensity in the two experiments. Solid line: optimal control actions in the first experiment. Dashed line: optimal control actions in the second experiment. The control actions are allowed to change once per day.