Dynamic Modelling and Optimisation of Cyanobacterial C-phycocyanin Production Process by Artificial Neural Network

Ehecatl Antonio del Rio-Chanona[1,‡], Emmanuel Manirafasha[2,‡], Dongda Zhang[1,*], Qian Yue[2], Keju Jing[2,3,*]

1: Department of Chemical Engineering and Biotechnology, University of Cambridge, Pembroke Street, Cambridge CB2 3RA, UK.

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

3: The Key Lab for Synthetic Biotechnology of Xiamen City, Xiamen University, Xiamen 361005, China

‡: These authors contributed equally to this work.

*: corresponding authors, email: dz268@cam.ac.uk (Dongda Zhang), tel: 44 (0)1223 330132; email: jkj@xmu.edu.cn (Keju Jing), tel: 86 592 2186038.

Abstract

Artificial neural networks have been widely applied in bioprocess simulation and control due to their advantageous properties. However, their feasibility in long-term photo-fermentation process modelling and prediction as well as their efficiency on process optimisation have not been well studied so far. In the current study, an artificial neural network was constructed to simulate a 15-day fed-batch process for cyanobacterial C-phycocyanin production, which to the best of our knowledge has never been conducted. To guarantee the accuracy of artificial neural network, two strategies were implemented. The first strategy is to generate artificial data sets by adding random noise to the original data set, and the second is to choose the change of state variables as training data output. In addition, the first strategy showed the distinctive advantage of reducing the experimental effort in generating training data. By comparing with current experimental results, it is concluded that both strategies give the network great modelling and predictive power to estimate the entire fed-batch process performance, even when few original experimental data are supplied. Furthermore, by optimising the operating conditions of a 12-day fed-batch process, a significant increase of 85.6% on C-phycocyanin production was achieved compared to previous work, which suggests the high efficiency of artificial neural network on process optimisation.

Keywords: artificial neural network; dynamic simulation; C-phycocyanin production; process optimisation; bioprocess design; fed-batch operation.

1. Introduction

Sustainable production of renewable energy and high-value products from microalgae have been extensively investigated due to their outstanding advantages [1]. For example, the carbon and energy sources for bioproducts synthesis and microorganism growth are CO_2 and solar energy, respectively, which are low investment cost and always plentiful [2]. Bioenergy including biodiesel, bioethanol and biohydrogen can be used as an environmentally friendly resource to replace conversional diesel, gasoline and transport fuel [1], [3]. Meanwhile, high-value bioproducts such as C-phycocyanin and astaxanthin have been widely applied in the food, pharmaceutical, and cosmetic industries due to their unique antioxidant and anti-inflammatory properties [4], [5]. Furthermore, microalgae and cyanobacteria have also been utilised as healthy food in China and Mexico since antiquity [6], and are currently being cultivated as nutrient products in the United States and Thailand [7].

To facilitate the industrialisation of both high-value bioproducts and bioenergy production as well as high density biomass cultivation, a variety of long-term bioprocesses have been designed in recent studies. Different operation modes such as batch, fed-batch and continuous operations have been implemented for both high density biomass cultivation and biofuel as well as high-value bioproduct production [8]–[12]. However, recent studies have proved that the low productivity of both biomass and bioproducts in long-term operations significantly prevents the further scale-up of these processes. For example, in industrial processes final biomass concentration can barely go up to 1.0 g L^{-1} [14]. High-value bioproduct content such as C-phycocyanin and astaxanthin in long-term processes is even

61 less than half of their maximum content observed in short-term experiments. In order to solve
62 this challenge, process optimisation has to be executed to maximise bioprocess productivity
63 and efficiency.
64
65 Two methodologies have been predominantly used for bioprocess optimisation [19], namely
66 response surface methodology (RSM) and dynamic simulation. RSM is a statistical technique
67 which estimates the relationship between decision variables (*e.g.* temperature) and response
68 variables (*e.g.* biomass concentration) [20]. Experimental data are in general utilised to fit a
69 quadratic expression used for further process optimisation [21]. Because RSM does not
70 require biochemical kinetics, it greatly simplifies the procedure of process optimisation.
71 However, the main weakness of this methodology is that it needs substantial amount of data
72 sets to guarantee the accuracy of the quadratic function [22]. In addition, without the support
73 of biochemical kinetics, its optimisation result cannot be applied to other processes whose
74 operation mode and duration are different from the design experiments [19].
75
76 On the contrary, dynamic models are constructed based on biochemical mechanisms and do
77 not require a large amount of experiments [23]. The latter character of dynamic simulation
78 significantly promotes its application in photo-fermentation processes which are in general
79 time-consuming since each experiment usually has to undergo two to three weeks [8], [9],
80 [12]. For instance, it has been applied to optimise the operating conditions for the production
81 of biohydrogen [19], [24], biolipid [25], [26], C-phycocyanin [16], [27] and biomass [24],
82 [28]. However, because of the complexity of microorganism metabolic mechanisms,

constructing a dynamic model is always a difficult task especially if the bioproduct synthesis mechanisms have not been well identified. As a result, dynamic simulation has not been well applied in many cases such as bioproduct synthesis in biofilm and algae-bacteria consortium wastewater treatment.

To complement the deficiency of dynamic model, artificial neural networks (ANNs) have been recently applied to simulate and control dynamic bioprocesses [29]–[31]. ANN is a mathematical approximation of biological neural network [32]. It is capable of estimating the complex relation between system input and output without detailed kinetic knowledge of the system [31]. A typical ANN contains one or two hidden layers where neurons are allocated to formulate the relation between system input and output [30], [33], [34]. The most commonly used functions inside of neurons are sigmoid and linear functions [29], [30]. These functions take in the system inputs to the neurons and produce the outputs. The inputs are changed by multiplication with weights and addition of biases in neurons, both of which are tuned when the model is trained by experimental data (training data) [35].

Compared to the dynamic simulation methodology, ANNs have the advantage of treating the bioprocess as a black box so that biochemical kinetics is not necessary for model construction. Meanwhile, it is able to depict the dynamic performance of a bioprocess, which is an outstanding improvement compared to RSM. As a result, it shows great potential in bioprocess simulation and optimisation. Despite of its advantages, it also requires a great amount of training data for model construction, and efficient optimisation algorithm for

ANNs based process optimisation has not been developed [36]. As a result, little research has focused on using ANNs to simulate and design long-term bioprocesses. Therefore, the current work aims to explore the feasibility of ANNs on long-term photo-fermentation process simulation and prediction when only limited training data are available. The efficiency of ANNs on process optimisation will also be explored in the current study.

In particular, C-phycocyanin production from cyanobacterium *Arthrospira platensis* was selected as the photo-fermentation process in the present study. C-phycocyanin is a blue antenna pigment used to enhance the photosynthetic efficiency of red algae and cyanobacteria [12], [37]. It is currently considered as a promising high-value bioproduct because of its wide applications in different industries [4]. For example, it is currently used as a natural alternative to conventional toxic synthetic pigments in cosmetic and food production [38]. In addition, it has a great potential to be applied in pharmaceutical industry due to its outstanding anti-oxidant, anti-inflammatory and neuroprotective properties [16]. Therefore, implementing process optimisation in this bioprocess becomes an indispensable step to complete the transition of C-phycocyanin production from laboratory scale to industrial scale.

2. Theory of methodology

2.1 Experiment setup

2.1.1 Computational experiment setup

In the current study, computational experiments (CEs) were executed to develop the ANN modelling and optimisation strategies as they are more time-efficient compared to real

experiments. A dynamic model capable of accurately simulating both biomass growth and C-phycocyanin accumulation has been proposed in our previous study [16] and was shown in Equations 1 to 6. The model is valid for the simulation of both batch process and fed-batch process in which dense nutrient culture is pulsed into the reactor. As a result, the present computational experiment results will be generated by this model. The detailed dynamic model construction can be found in the previous research [16], and model parameter values are listed in Table 1. However, it is notable that computational experiments are purely selected in the current study due to their high time-efficiency, and can be completely replaced by real experiments in future work. Thus, this dynamic model is not necessary for ANN model construction in the current study.

$$\frac{dX}{dt} = u_0 \cdot \frac{N}{N + K_N} \cdot X - u_d \cdot X \quad (1)$$

$$\frac{dN}{dt} = -Y_{N/X} \cdot u_0 \cdot \frac{N}{N + K_N} \cdot X \quad (2)$$

$$\frac{d(q_c X)}{dt} = k \cdot X - \frac{k_d \cdot (q_c X)}{N + K_{Np}} \quad (3)$$

$$I = I_0 \cdot \left(e^{-(\tau \cdot X + K_a) \cdot z} + e^{-(\tau \cdot X + K_a) \cdot (L - z)}\right) \quad (4)$$

$$u_0 = \frac{u_m}{20} \cdot \sum_{n=1}^{9} \left(\frac{I_{i=0}}{I_{i=0} + k_s + \frac{I_{i=0}^2}{k_i}} + 2 \cdot \frac{I_{i=\frac{n \cdot L}{10}}}{I_{i=\frac{n \cdot L}{10}} + k_s + \frac{I_{i=\frac{n \cdot L}{10}}^2}{k_i}} + \frac{I_{i=L}}{I_{i=L} + k_s + \frac{I_{i=L}^2}{k_i}} \right) \quad (5)$$

$$k = \frac{k_m}{20} \cdot \sum_{n=1}^{9} \left(\frac{I_{i=0}}{I_{i=0} + k_{sp} + \frac{I_{i=0}^2}{k_{ip}}} + \frac{2 \cdot I_{i=\frac{n \cdot L}{10}}}{I_{i=\frac{n \cdot L}{10}} + k_{sp} + \frac{I_{i=\frac{n \cdot L}{10}}^2}{k_{ip}}} + \frac{I_{i=L}}{I_{i=L} + k_{sp} + \frac{I_{i=L}^2}{k_{ip}}} \right) \quad (6)$$

where X is biomass concentration, u_0 is cell specific growth rate, N is nitrate concentration, K_N is nitrate half-velocity coefficient, K_{Np} is nitrate half-velocity coefficient

for phycocyanin consumption, u_d is cell specific decay rate, $Y_{N/X}$ is nitrate yield coefficient, q_c is phycocyanin content in cells, k is phycocyanin production constant and k_d is phycocyanin consumption constant. u_m is the maximum specific growth rate, I is light intensity, k_s is light saturation term and k_i is light inhibition term. k_m is the maximum phycocyanin accumulation constant, k_{sp} is light saturation term for phycocyanin synthesis and k_{ip} is light inhibition term for phycocyanin synthesis. I_0 represents incident light intensity, τ is cell absorption coefficient, K_a is bubble reflection coefficient, z is the distance from light source and L is the width of the PBR.

Table 1: Parameters in the Monod model. Parameters in this model were fitted by the experimental data measured in our previous research [16].

Parameter	Value	Parameter	Value
u_m [h^{-1}]	0.0923	k_s [μmol m^{-2} s^{-1}]	178.85
u_d [h^{-1}]	0.0	k_i [μmol m^{-2} s^{-1}]	447.12
K_N [mg L^{-1}]	393.10	k_{sp} [μmol m^{-2} s^{-1}]	23.51
$Y_{N/X}$ [mg g^{-1}]	504.49	k_{ip} [μmol m^{-2} s^{-1}]	800.0
k_m [mg g^{-1} h^{-1}]	2.544	τ [m^2 g^{-1}]	0.0520
k_d [h^{-1}]	0.281	K_a [m^{-1}]	0.0
K_{Np} [mg L^{-1}]	16.89		

The dynamic model in this work was solved numerically using the Python programming language along with *numpy* [39] *and scipy* [40], [41] packages.

2.1.2 Real experiment setup

In order to verify the predictability of current ANN model, two real fed-batch experiments (RE1 and RE2) which were conducted in our recent study [12] were used in the current study. The experimental fed-batch processes were implemented such that dense nitrate solution was fed into the reactor to maintain the culture nitrate concentration at 5 mM and 10 mM, respectively. The detailed experimental fed-batch process setup can be found in [12]. In addition, another real experiment (RE3) was carried out in the current study to validate the simulation results of the current ANN regarding to the optimised 12-day fed-batch process which will be introduced in Section 2.4. All of the three real experiments were replicated three times to guarantee the accuracy of the experimental results. The photobioreactor used in the current study is a 1 L (15.5 cm in length and 9.5 cm in diameter) tank reactor where illumination is provided from both sides of the reactor [16]. 2.5% CO_2 with 0.2 vvm was constantly pumped into the reactor to provide carbon source and an agitation rate of 400 rpm was chosen to guarantee the well mix of the current culture [12].

2.2 Principle of training experiment design

In general, a large amount of training data is necessary to train ANNs. However specific to a bioprocess where the process duration is two to three weeks, obtaining enough data sets is a very time-consuming task since the data sets must cover a wide range of values for each input variable. In order to construct a highly accurate model and save experimental time effort, an experiment design strategy particular for ANN construction becomes vital in the present

research.

To achieve the aforementioned goal, a framework was proposed where only three experiments are designed to generate training data and reduce experimental time cost. To guarantee the large domain of training data sets, all of the computational training experiments (CE1-CE3) were constructed to be a 15-day fed-batch process in which dense nitrate feed can be intermittently added into the bioreactor to change the culture nitrate concentration. These fed-batch processes were operated in such a way that the nitrate culture concentration for most duration of the experiments was kept low (100~500 mg L^{-1}), medium (500~1000 mg L^{-1}) and high (1000~2000 mg L^{-1}), respectively, so that biomass concentration, nitrate concentration and phycocyanin production between each training experiment are sufficiently different.

The reason why conducting three experiments with different nitrate concentrations is due to the complicated effects of nitrate concentration on phycocyanin accumulation. Although it is known that nitrogen is an essential element for phycocyanin synthesis and cells will consume phycocyanin in a nitrate-limiting culture, recent study has found that maximum cellular phycocyanin content can be reduced in a culture with high nitrate concentration [12]. As a result, to guarantee the high accuracy of current ANN for further process optimisation, it is necessary to ensure that the model can well simulate cell growth and pigment production in both nitrate-limiting and nitrate-sufficient conditions. Therefore, the aforementioned computational experiments were carried out.

Meanwhile, it was assumed that biomass concentration, nitrate concentration and phycocyanin production can be measured 3 times per day (once every 8 hours), which is feasible for future real experiment implementations. The current training experiment design principle can then be directly applied in real experimental research, as all the computational training experiments are similar with our previous real experiment setups [12]. The detailed operating conditions for the training experiments are shown in Table 2.

Table 2: Operating conditions of training experiments. Initial biomass concentration and initial phycocyanin content are set the same as in our previous study [12], initial nitrate concentration are different in each training experiment. At the beginning of each pulse day, dense nitrate feed is pumped into the reactor. The replenished culture nitrate concentration is listed as pulse concentration.

	CE1	CE2	CE3
Ini. biomass [g L^{-1}]	0.113	0.113	0.113
Ini. nitrate [mg L^{-1}]	100.0	300.0	500.0
Ini. phycocyanin [mg g^{-1}]	1.120	1.120	1.120
First pulse day	3	4	3
First pulse con. [mg L^{-1}]	200.0	500.0	1500.0
Second pulse day	6	7	7
Second pulse con. [mg L^{-1}]	400.0	800.0	2000.0
Third pulse day	8	10	11

Third pulse con. [mg L^{-1}]	600.0	1000.0	2000.0
Fourth pulse day	12	14	14
Fourth pulse con. [mg L^{-1}]	500.0	1200.0	2500.0
Fifth pulse day	14	--	--
Fifth pulse con. [mg L^{-1}]	600.0	--	--

211

212 To further test the accuracy of current ANN, two additional computational fed-batch
213 processes (CE4 and CE5) were carried out. The operating conditions of both test experiments
214 are different from those in the training experiments, and are listed in Table 3. Finally, Table 4
215 summarises all the experiments included in the current study.

216

217 Table 3: Operating conditions of test experiments.

	CE4	CE5
Ini. biomass [g L^{-1}]	0.192	0.111
Ini. nitrate [mg L^{-1}]	300.0	500.0
Ini. phycocyanin [[mg g^{-1}]	3.081	1.102
First pulse day	3	3
First pulse con. [mg L^{-1}]	500.0	1500.0
Second pulse day	7	6
Second pulse con. [mg L^{-1}]	800.0	2100.0
Third pulse day	11	11
Third pulse con. [mg L^{-1}]	1700.0	1900.0

Fourth pulse day	--	14
Fourth pulse con. [mg L^{-1}]	--	2300.0

218

219 Table 4: Summary of current experiments.

	Classification	Aim
CE1	Computational experiment	Training experiment for ANN model construction
CE2	Computational experiment	Training experiment for ANN model construction
CE3	Computational experiment	Training experiment for ANN model construction
CE4	Computational experiment	Verify ANN model accuracy
CE5	Computational experiment	Verify ANN model accuracy
RE1	Real experiment	Verify ANN model predictability
RE2	Real experiment	Verify ANN model predictability
RE3	Real experiment	Validate ANN model process optimisation result

220

221 **2.3 Artificial neural network model construction**

222 In most recent dynamic applications [42], [43], ANNs are designed such that by feeding it
223 past states it can predict future ones. As previously mentioned in the current work, three
224 computational 15-day experimental runs have been carried out. One for which nitrate

concentration was in general high, another where it was low and the third one where it was in between (medium). As the states (biomass concentration, nitrate concentration and phycocyanin production) were measured once every eight hours, an original data set containing 135 points from all three experimental runs was generated.

However, to train an accurate ANN, a significantly larger amount of data should be provided [36]. This can be a major issue since it is very time consuming to obtain enough data sets in the underlying bioprocess. To overcome this challenge, two strategies were employed in the present work. The first one is to replicate the available experimental data, by adding a 3% normally distributed random noise to generate 100 additional artificial sets of data. This would enable the ANN to be trained with more information which would be infeasible to obtain in practice. The second strategy in the current study is to predict the rate of change for each state by given their past information, rather than directly predicting future states based on past ones. These combined strategies, as will be further discussed in Section 3.2, give the network great modelling and predictive power, even when very few original experimental data are actually supplied.

In the current study, the inputs of ANN include biomass concentration, nitrate concentration and phycocyanin production, and the outputs are change of biomass and nitrate concentrations as well as change of phycocyanin production. Different backpropagation structures for the ANNs were explored in this work, all of them using a sigmoid function for each neuron, and the results were presented in Section 3.1. When predicting the dynamic

247 performance of current test and real fed-batch experiments, only the initial conditions and the
248 times and magnitudes of nitrate pulses were supplied to the ANN. The neural network then
249 predicts the state profiles solely based on these inputs. The ANN construction was
250 implemented through *PyBrain* [44], a package in Python. All computational experiments
251 were run in an Intel Core i5, 4 GB RAM 2.53 GHz computer.
252
253 **2.4 Operation search for process development**
254 To explore the feasibility and efficiency of current ANN on process optimisation, a 12-day
255 fed-batch was then designed in the present study to find the optimal strategy for
256 C-phycocyanin production. Four different nitrate pulses were added along the time course of
257 operation, the first input is initially assumed to take place on day 3, the second on day 6, the
258 third on day 9 and the fourth on day 12. The culture nitrate concentration after replenishment
259 at each time can vary from 500-2500 mg L^{-1}. To find a high quality solution for the present
260 optimisation problem, the current study takes 100 step intervals for the added nitrate
261 concentration and allows the possibility of inputs to be implemented one day earlier or later
262 than that specified above. Thus, in total 2.5×10^5 runs of computational experiments were
263 carried out to search the best operating conditions for phycocyanin production. This
264 procedure enables the current work to search for the optimal operating conditions given 3
265 inputs in 12 days with a 50 mg L^{-1} tolerance in the refreshed culture nitrate concentration.
266
267 **3. Results and discussion**
268 **3.1 Result of model construction and validity**

A feedforward ANN (3,6,6,3) including a three neuron input layer, two six neuron hidden layers, and three neuron output layer is eventually selected due to its higher accuracy than other ANNs constructed in the current study. The resulting ANN is shown in Fig. 1(a), and in Fig. 1(b) a neuron with n inputs is also represented.

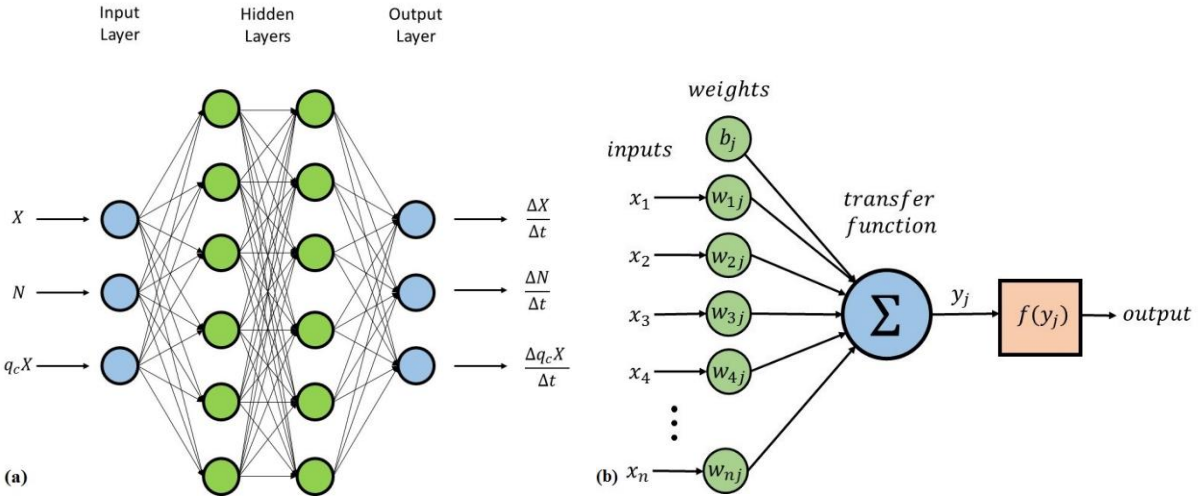


Figure 1: (a) A schematic representation of the feedforward three neuron input layer, two six neuron hidden layers, and three neuron output layer artificial neural network designed in this work. (b): A schematic representation of the neurons with n inputs, where the function $f(x)$ corresponds to a sigmoid function for the current work.

Fig. 2 shows the ANN model simulation result and the training data (CE1-CE3). From Fig. 2(a) to 2(c), it is found that the current model can well represent the entire dynamic performance of all the three training experiments where significant changes on the values of biomass density and nitrate concentrations are included, which indicates the high accuracy of the training procedure and current model.

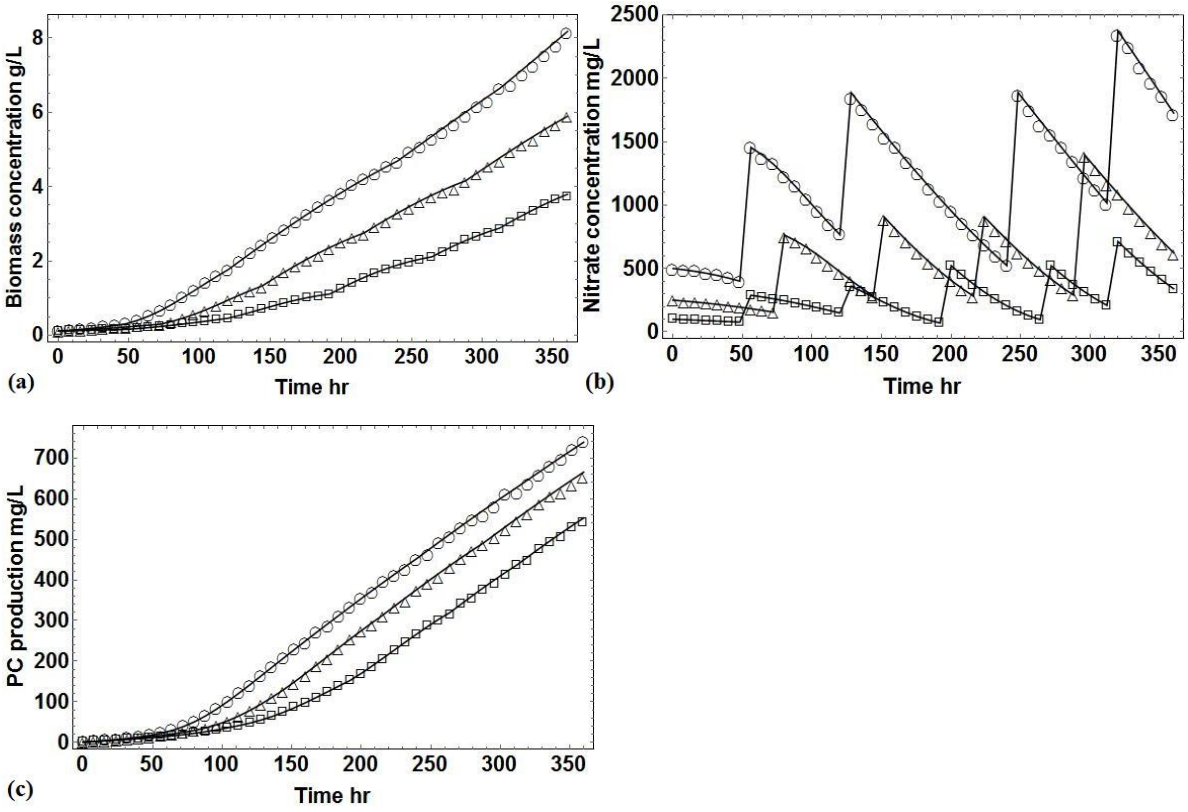


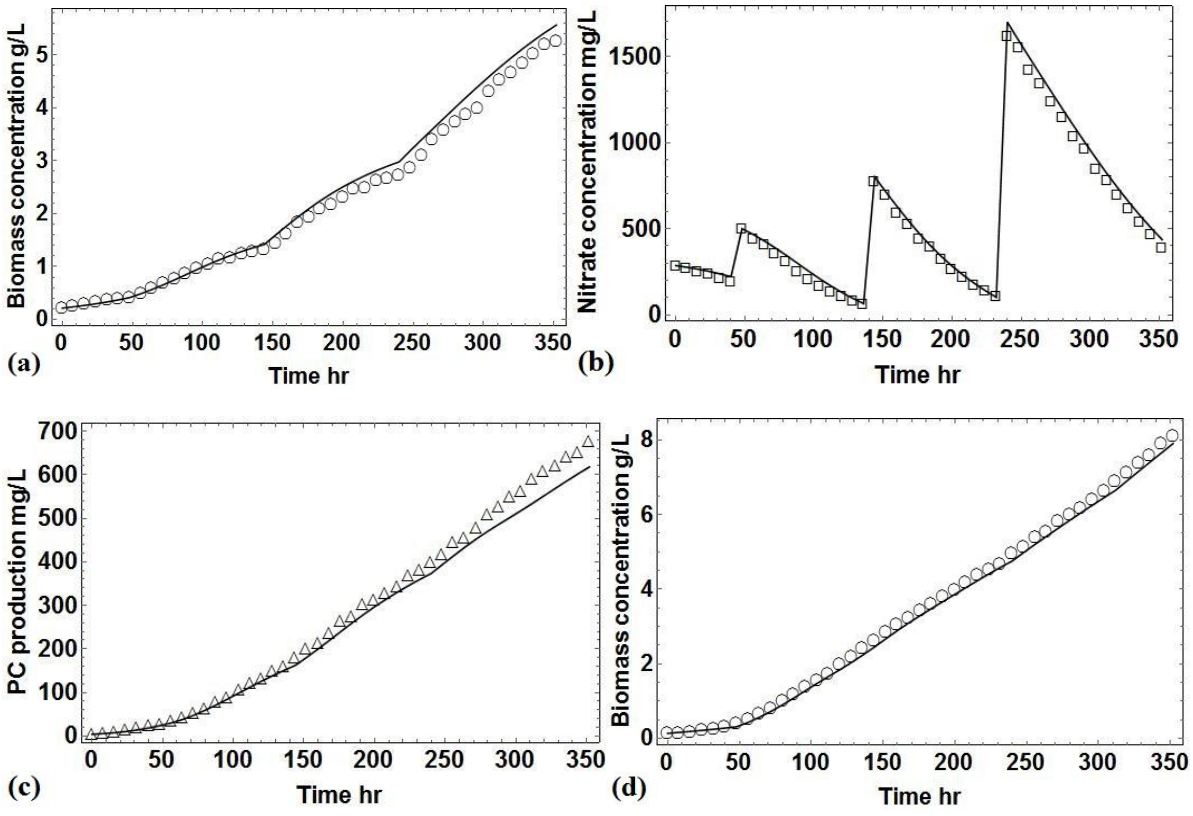
Figure 2: Comparison of ANN simulation result with training data. Open squares: simulation results of CE1, open triangles: simulation results of CE2, open circles: simulation results of CE3, lines: experimental results. (a): biomass concentration; (b): nitrate concentration; (c): C-phycocyanin production. The sharp change of nitrate concentration in Fig. 2(b) is due to the addition of nitrate feed.

3.2 Effect of data error on model accuracy

In the current work, artificial experimental data sets were generated by adding errors to the original data set. One hundred artificial data sets are computed by adding a 3% normally distributed error to the original data. This, as can be seen from the computational results in Section 3.1, is an essential strategy for ANN training. In order to further explore the effect of the amount of noise added on the accuracy of current ANN, two other ANN training schemes

were implemented, one where the normal distributed error has a 5% magnitude, whilst another ANN is trained without data replication. Their performance is next analysed.

Fig. 3 shows the modelling results of the two test fed-batch processes (CE4 and CE5) based on the model with a 3% data error. From Fig. 3, it can be seen that this model succeeds in accurately predicting both experiment performances where biomass densities increase rapidly and nitrate concentrations change dramatically, which suggests that adding an error of 3% for the augmentation of training data is highly feasible in practice and does not compromise the model efficiency.



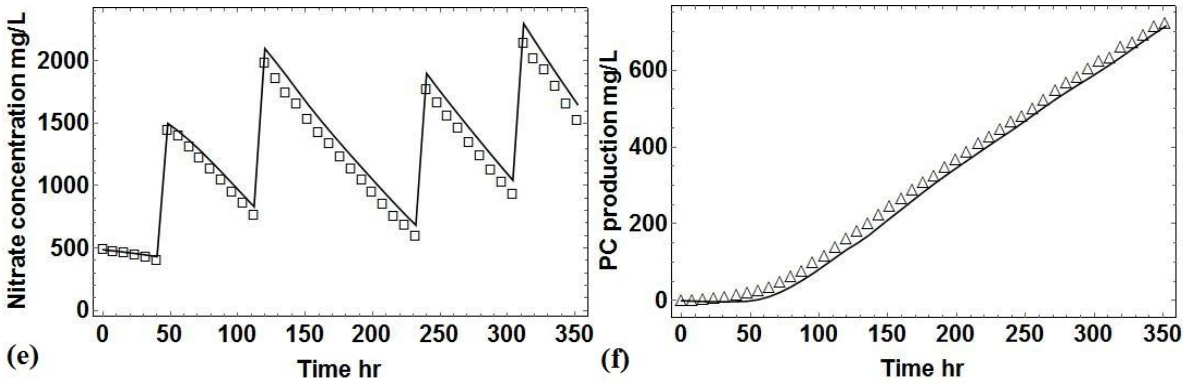
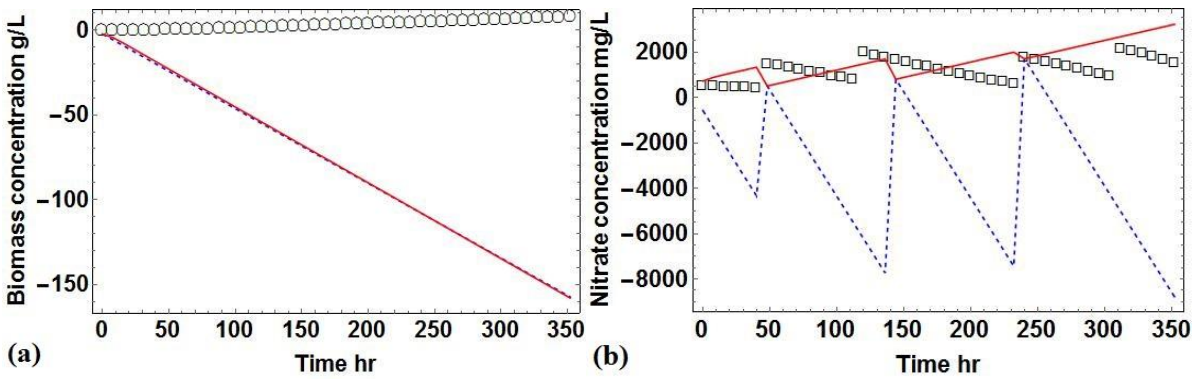


Figure 3: Comparison of ANN simulation results and test fed-batch processes experimental data. (a)-(c): biomass concentration, nitrate concentration and C-phycocyanin (PC) production in CE4; (d)-(e): biomass concentration, nitrate concentration and C-phycocyanin (PC) production in CE5. Solid lines: experimental results; points: simulation results. The sharp change of nitrate concentration in Fig. 3(b) and 3(e) is due to the addition of nitrate feed.

However, large errors are computed when the training informatio without data replication and with a 5% error are used to construct an ANN. For example, Fig. 4 compares the simulation results of both models with the second test experiment data. From Fig. 4, it can be seen that both of the models fail to predict the test experiment performance.



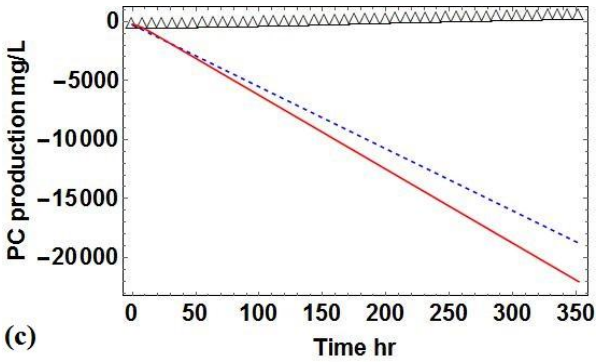


Figure 4: Comparison of the second test experiment data (CE5) with the simulation results from different models. Solid line: simulation results of the model with a 5% error; dashed line: simulation results of the model without data replication; points: second test experiment data. (a): biomass concentration; (b): nitrate concentration; (c): C-phycocyanin production. The sharp change of nitrate concentration in Fig. 3(b) is due to the addition of nitrate feed.

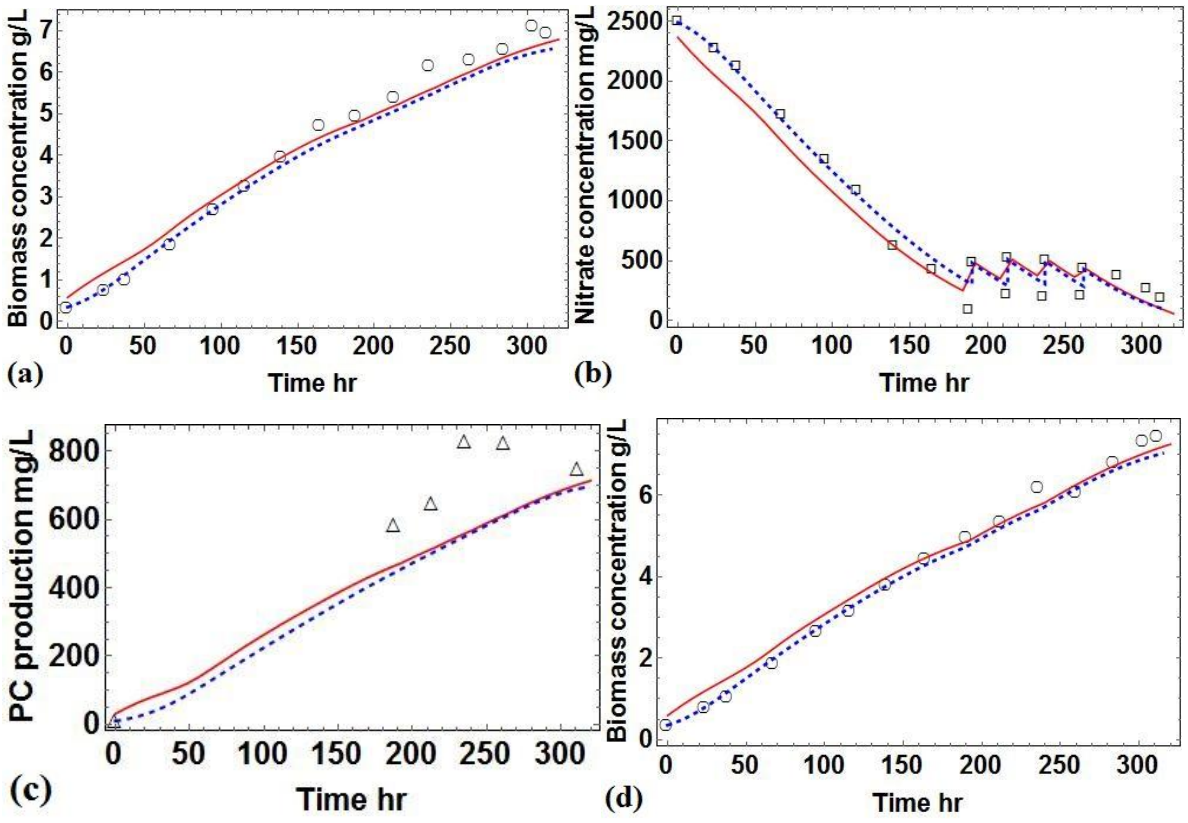
It can be then concluded that in order to construct a reliable ANN, a sufficient amount of training data is indispensable, and the model without data replication clearly does not satisfy this prerequisite. It is also demonstrated that selecting the magnitude of error is a crucial issue for ANN construction, as large error can directly mislead the model predictions. However, if the error is too small (*e.g.* 1%), the present study also found that the ANN is not able to well represent the dynamic performance of the current bioprocess, since the artificially generated data sets are very close to the original data set and the ANN recognises them as the same during its training procedure. On the contrary, if the errors induced are too large (*e.g.* 5%) the ANN is misguided and the actual dynamics of the process are hidden in the randomness induced. The threshold for the induced randomness in the error should be particular to each system. As a result, the model with an error of 3% in training data is selected in the current study.

3.3 Model predictability

The ANN model is then used to simulate the performance of two real experimental results (RE1 and RE2) conducted in our previous study mentioned in Section 2.2. In these two experiments, both the initial operating conditions (initial biomass concentration, initial nitrate concentration etc.) and the nitrate pulse locations are different from our computational experiments CE1-CE5. Fig. 5 then compares the results of real experiments, previous dynamic model and current ANN model. From Fig. 5, it is found that the current ANN can well predict the dynamic performance of both real experiments, and its results are also highly similar with those based on the dynamic model.

It is notable that during the model training procedure, as the time interval between training data input and output is 8 hours, in principle the current ANN can only well predict the process performance within 8 hours, and beyond that the accuracy of prediction results cannot be guaranteed. Furthermore, when the current ANN is used to predict the two test fed-batch processes and real experiments, only the initial operating conditions and nitrate pulses information are supplied to the ANN, and the model has to predict the entire time course of these processes (312 hours) without additional information. As a result, the prediction results shown in Fig. 3 and 5 strongly suggest that based on the current model construction strategies the proposed ANN is characterised by a high predictability for long-term dynamic bioprocess simulation, which to the best of our knowledge has not been reported before.

The approximate 10%-15% underestimation of phycocyanin production in the ANN is led by the limit of the previous dynamic model which is used to generate training data for current ANN construction. As a result, this underestimation may be eliminated in the future if real experimental data are directly used for ANN training, and the ANN model has the potential to show higher accuracy compared to the dynamic model due to its high agreement between the ANN simulation results and current computational experimental data. Therefore, the current comparison strongly indicates that ANN can be considered as a reliable replacement of dynamic model for practical applications in future work.



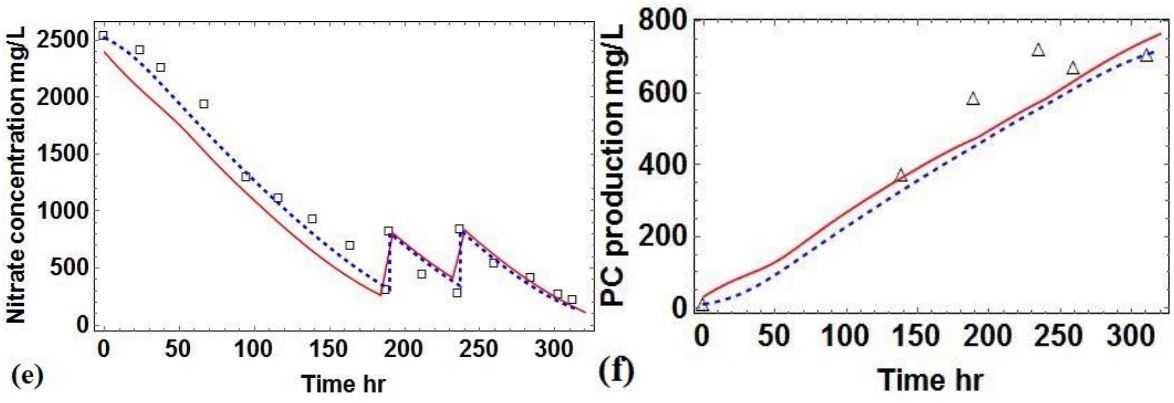


Figure 5: Comparison of ANN and dynamic model simulation results with real experiment data. Solid line: ANN results; dashed line: dynamic model results; point: experimental data. (a)-(c): biomass concentration, nitrate concentration and phycocyanin production in the 5 mM fed-batch process (RE1); (d)-(e): biomass concentration, nitrate concentration and phycocyanin production in the 10 mM fed-batch process (RE2).

3.4 Fed-batch process optimisation

Given the predictive power of the ANN designed and the high computational speed for which these structures are known for, it is possible to carry out an optimisation procedure to explore the control space (replenished culture nitrate concentration) to an acceptable tolerance.

The current ANN is used to optimise a 12-day fed-batch process with the same incident light intensity (300 μmol m^{-2} s^{-1}), temperature (28 °C), initial biomass concentration (0.40 g L^{-1}) and phycocyanin content (10.5%) in previous real experimental fed-batch processes [12]. The aim is to identify the feasibility and efficiency of current model on process optimisation.

Based on the current model and operation search methodology, the optimal operating

conditions found in the present study and the real experiment results of the 12-day fed-batch process are listed in Table 5. Compared to the previous experimental work, the optimised process achieves a higher biomass concentration than that in the 5 mM fed-batch process (6.96 g L^{-1}) and the 10 mM fed-batch process (7.46 g L^{-1}). Phycocyanin production in the current process remarkably increases by 74.9% and 85.6% compared to the 5 mM fed-batch process and 10 mM fed-batch process, respectively. Final phycocyanin intracellular content in the current optimised process is also significantly improved to be 17.2%, even higher than the highest phycocyanin content (16.1%) reported in previous research [12]. As a result, it is concluded that in future work it is highly feasible and efficient to utilise ANN for dynamic bioprocess optimisation.

Table 5: Optimal operating conditions and real experiment (RE3) results of the 12-day fed-batch process. Pulse day means that the pulse is added at the beginning of that day.

Initial nitrate con.	1300 [mg L^{-1}]	Third pulse day	day 11
First pulse day	day 5	Third addition con.	1100 [mg L^{-1}]
First addition con.	2100 [mg L^{-1}]	Final biomass con.	7.62 [g L^{-1}]
Second pulse day	day 7	Final PC production	1310 [mg L^{-1}]
Second addition con.	1700 [mg L^{-1}]	Final PC content	17.2%

Fig. 6 shows the predictability of the current ANN regarding to the optimised fed-batch process. From the figure, it is clearly concluded that the current model can well predict the dynamic performance of the investigated process, especially for biomass growth and nitrate

consumption. As mentioned before, the underestimation of phycocyanin production in the current ANN is led by the limitation of recent dynamic model which has been reported to underestimate phycocyanin production in a fed-batch process where nitrate feed is frequently pumped into the reactor [16]. This underestimation, however, can be totally eliminated by directly using the data from future real experiments to construct ANN models.

In the current optimised fed-batch process, it is found that biomass concentration does not increase rapidly (Fig. 6(a)) within the first 4 days as there is no nitrate pulse added (Fig. 6(b)). Previous research declared that a medium biomass density (around 1.5 g L^{-1}) can facilitate C-phycocyanin synthesis [16]. This is because local illumination in the reactor is moderately reduced due to cyanobacterial light absorption [16], and a lower light intensity is preferable for phycocyanin accumulation as it is used to help cells to trap solar energy [12]. However, high biomass concentration (higher than 4.0 g L^{-1}) can remarkably aggravate the light attenuation in the photobioreactor, which leads to the termination of C-phycocyanin accumulation and biomass growth [16]. Therefore, it is concluded that C-phycocyanin content in cells mainly increases in the beginning period of the experiment (Fig. 5(d)) when local illumination is not significantly reduced by biomass absorption.

As C-phycocyanin production is also dependent on biomass concentration, dense nitrate feed is intermittently added to facilitate cell growth. Despite that high nitrate concentration can suppress the consumption of phycocyanin [12], it also leads to a high biomass concentration which significantly reduces the light intensity in the reactor and accelerates the consumption

of phycocyanin. As a result, it can be found that the amount of nitrate added in the culture (Fig. 6(b)) decreases from the first pulse to the third pulse so that the conflicting effects of nitrate concentration on phycocyanin accumulation can be well balanced. Therefore, total phycocyanin production can keep increasing (Fig. 6(c)) without compromising the high phycocyanin content in cells (Fig. 6(d)).

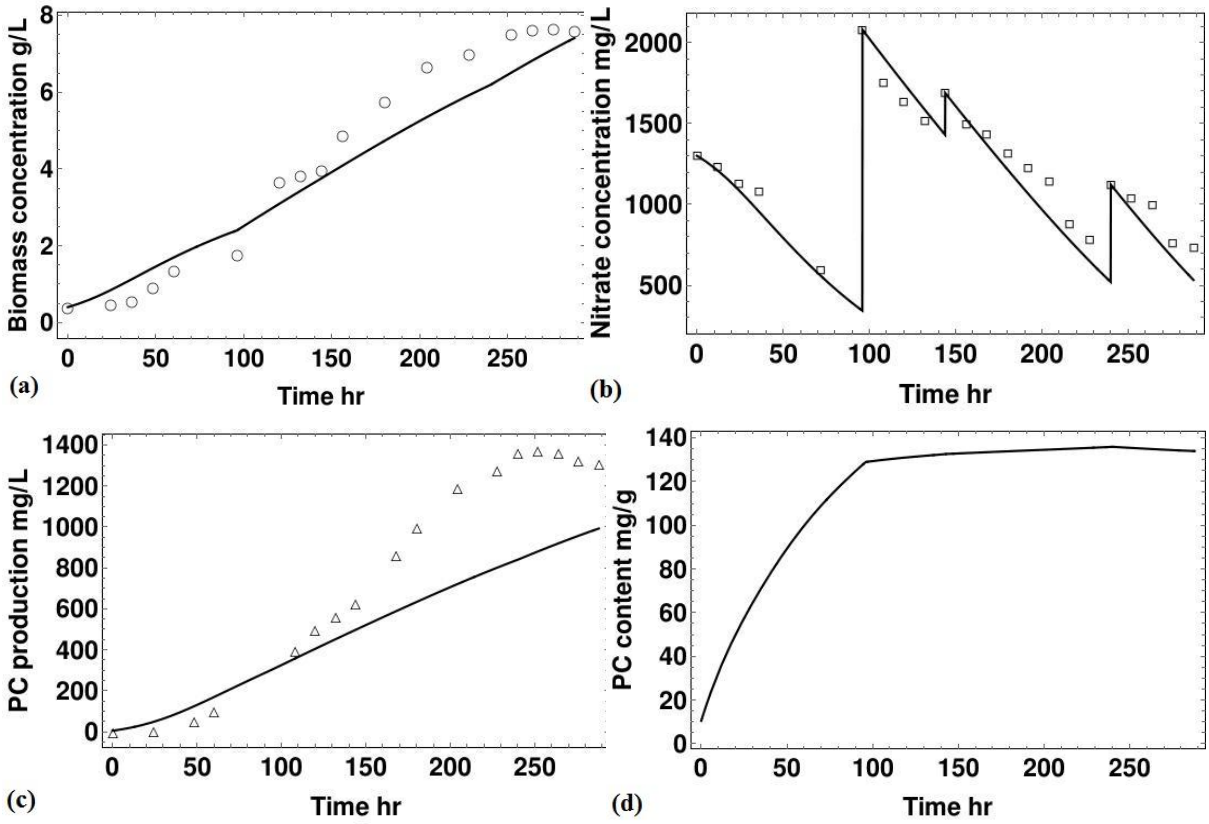


Figure 6: Performance of the optimised 12-day fed-batch process. (a) biomass concentration; (b) nitrate concentration; (c) phycocyanin production in the reactor; (d) phycocyanin content in cells. Lines are simulation results, and points are real experimental results (RE3).

Conclusion

In the current study, an ANN was constructed to simulate a 15-day fed-batch process for C-phycocyanin production. Computational experiments are carried out to develop the procedure for long-term bioprocess simulation and optimisation by ANN. An experiment design framework particularly for ANN construction is proposed to save experiment time effort. To guarantee the high accuracy of current model, two strategies are implemented during the model construction. It is found that adding random noise to augment the available training data is highly feasible in practice and does not compromise the model efficiency, whilst the magnitude of noise has to be cautiously determined.

By comparing with the result of test processes and real experiments, it is concluded that the current ANN is highly accurate and characterised by great predictable power. It can be then considered as an alternative to dynamic model for future process simulation and design. By searching the optimal operating conditions of a 12-day fed-batch process, it is concluded that the current model shows high efficiency on process optimisation, since a significant increase on phycocyanin production is achieved compared to previous research.

In terms of future work, due to the unique advantages of ANN which does not need the full knowledge of the investigated process kinetics and high efficiency on process optimisation, the current proposed ANN optimisation strategy should be applied in both fed-batch and repeated batch long-term operations for different high-value bioproducts and biofuels production. The difficulty of obtaining accurate kinetic models for bioprocess design can be therefore solved. Meanwhile, the amount of experimental data can be significantly reduced

by using the current strategy compared to that of RSM. Moreover, to further develop the current optimisation strategy, a stochastic optimisation algorithm particular for this application should be developed to improve the efficiency of ANN process optimisation.

Acknowledgement

Author E. A. del Rio-Chanona is funded by CONACyT scholarship No. 522530 from the Secretariat of Public Education and the Mexican government. Author D. Zhang gratefully acknowledges the support from his family. This work was also supported by the National High Technology Research and Development Program 863, China (No. 2014AA021701) and the National Marine Commonwealth Research Program, China (No. 201205020-2).

References

[1] T. M. Mata, A. a. Martins, and N. S. Caetano, "Microalgae for biodiesel production and other applications: A review," *Renew. Sustain. Energy Rev.*, vol. 14, no. 1, pp. 217–232, Jan. 2010.

[2] D. Zhang, P. Dechatiwongse, E. A. Del-Rio-Chanona, K. Hellgardt, G. C. Maitland, and V. S. Vassiliadis, "Analysis of the cyanobacterial hydrogen photoproduction process via model identification and process simulation," *Chem. Eng. Sci.*, vol. 128, pp. 130–146, 2015.

[3] R. Chaubey, S. Sahu, O. O. James, and S. Maity, "A review on development of industrial processes and emerging techniques for production of hydrogen from renewable and sustainable sources," *Renew. Sustain. Energy Rev.*, vol. 23, no. 0, pp. 443–462, Jul. 2013.

[4] M. Kuddus, P. Singh, G. Thomas, and A. Al-Hazimi, "Recent developments in production and biotechnological applications of C-phycocyanin.," *Biomed Res. Int.*, vol. 2013, p. 742859, Jan. 2013.

[5] E. Del Río, F. G. Acién, M. C. García-Malea, J. Rivas, E. Molina-Grima, and M. G. Guerrero, "Efficient one-step production of astaxanthin by the microalga Haematococcus pluvialis in continuous culture.," *Biotechnol. Bioeng.*, vol. 91, no. 7, pp. 808–15, Sep. 2005.

[6] D. Zhang, P. Dechatiwongse, E. a. del Rio-Chanona, G. C. Maitland, K. Hellgardt, and V. S. Vassiliadis, "Modelling of light and temperature influences on cyanobacterial growth and biohydrogen production," *Algal Res.*, vol. 9, pp. 263–274, May 2015.

[7] W. Chu, "Biotechnological applications of microalgae," *Int. e-Journal Sci. Med. Educ.*, vol. 6, no. 126, pp. 24–37, 2012.

[8] P. Dechatiwongse, G. Maitland, and K. Hellgardt, "Demonstration of a two-stage aerobic/anaerobic chemostat for the enhanced production of hydrogen and biomass from unicellular nitrogen-fixing cyanobacterium," *Algal Res.*, vol. 10, pp. 189–201, Jul. 2015.

[9] J. P. Kim, K.-R. Kim, S. P. Choi, S. J. Han, M. S. Kim, and S. J. Sim, "Repeated production of hydrogen by sulfate re-addition in sulfur deprived culture of Chlamydomonas reinhardtii," *Int. J. Hydrogen Energy*, vol. 35, no. 24, pp. 13387–13391, Dec. 2010.

[10] S.-K. Wang, Y.-R. Hu, F. Wang, A. R. Stiles, and C.-Z. Liu, "Scale-up cultivation of Chlorella ellipsoidea from indoor to outdoor in bubble column bioreactors.," *Bioresour. Technol.*, vol. 156, pp. 117–22, Mar. 2014.

[11] Z. Cheng-wu, O. Zmora, and R. Kopel, "An industrial-size flat plate glass reactor for mass production of Nannochloropsis sp . (Eustigmatophyceae)," *Aquaculture*, vol. 195, no. 0, pp. 35–49, 2001.

[12] Y. Xie, Y. Jin, X. Zeng, J. Chen, Y. Lu, and K. Jing, "Fed-batch strategy for enhancing cell growth and C-phycocyanin production of Arthrospira (Spirulina) platensis under phototrophic cultivation.," *Bioresour. Technol.*, vol. 180, pp. 281–7, Mar. 2015.

[13] J. Fábregas, A. Otero, A. Maseda, and A. Domínguez, "Two-stage cultures for the production of Astaxanthin from Haematococcus pluvialis," *J. Biotechnol.*, vol. 89, no. 1, pp. 65–71, Jul. 2001.

[14] D. Zhang, P. Dechatiwongse, E. Antonio Del Rio-Chanona, G. C. Maitland, K. Hellgardt, and V. S. Vassiliadis, "Dynamic modelling of high biomass density cultivation and biohydrogen production in different scales of flat plate photobioreactors.," *Biotechnol. Bioeng.*, Jun. 2015.

[15] P. Dechatiwongse, D. Zhang, E. A. del Rio-Chanona, V. S. Vassiliadis, G. C. Maitland, and K. Hellgardt, "Effects of Light and Temperature Regimes on Hydrogen Produciton of Nitrogen-fixing Cyanobacterium Cyanothece sp. ATCC 51142.," *Int. J. Hydrog. Energy (submited to journal)*, 2015.

[16] E. A. del Rio-Chanona, D. Zhang, Y. Xie, M. Emmanue, K. Jing, and V. S. Vassiliadis, "Dynamic Simulation and Optimisation for Arthrospira platensis Growth and C-phycocyanin Production: Part I. Model Construction and Process Analysis," *Biotechnol. Bioeng. (submitted to journal)*, 2015.

[17] M. Wan, D. Hou, Y. Li, J. Fan, J. Huang, S. Liang, W. Wang, R. Pan, J. Wang, and S. Li, "The effective photoinduction of Haematococcus pluvialis for accumulating astaxanthin with attached cultivation.," *Bioresour. Technol.*, vol. 163, pp. 26–32, Jul. 2014.

[18] M. Wan, J. Zhang, D. Hou, J. Fan, Y. Li, J. Huang, and J. Wang, "The effect of temperature on cell growth and astaxanthin accumulation of Haematococcus pluvialis during a light-dark cyclic cultivation.," *Bioresour. Technol.*, vol. 167, pp. 276–83, Sep. 2014.

[19] D. Zhang, N. Xiao, K. T. Mahbubani, E. a. del Rio-Chanona, N. K. H. Slater, and V. S. Vassiliadis, "Bioprocess modelling of biohydrogen production by Rhodopseudomonas palustris: Model development and effects of operating conditions on hydrogen yield and glycerol conversion efficiency," *Chem. Eng. Sci.*, vol. 130, pp. 68–78, Jul. 2015.

[20] M. G. Moghaddam, “Comparison of Response Surface Methodology and Artificial Neural Network in Predicting the Microwave-Assisted Extraction Procedure to Determine Zinc in Fish Muscles,” *Food Nutr. Sci.*, vol. 02, no. 08, pp. 803–808, 2011.

[21] M. A. Bezerra, R. E. Santelli, E. P. Oliveira, L. S. Villar, and L. A. Escaleira, “Response surface methodology (RSM) as a tool for optimization in analytical chemistry.,” *Talanta*, vol. 76, no. 5, pp. 965–977, Sep. 2008.

[22] G. E. P. Box and K. B. Wilson, “On the Experimental Attainment of Optimum Conditions,” *J. R. Stat. Soc.*, vol. 13, no. 1, pp. 1–45, 1951.

[23] I. Vatcheva, H. de Jong, O. Bernard, and N. J. I. Mars, “Experiment selection for the discrimination of semi-quantitative models of dynamical systems,” *Artif. Intell.*, vol. 170, no. 4–5, pp. 472–506, Apr. 2006.

[24] D. Zhang, E. A. Del-Rio Chanona, V. S. Vassiliadis, and B. Tamburic, “Analysis of green algal growth via dynamic model simulation and process optimisation.,” *Biotechnol. Bioeng.*, pp. 1–55, Apr. 2015.

[25] V. O. Adesanya, M. P. Davey, S. A. Scott, and A. G. Smith, “Kinetic modelling of growth and storage molecule production in microalgae under mixotrophic and autotrophic conditions.,” *Bioresour. Technol.*, vol. 157, pp. 293–304, Apr. 2014.

[26] S. J. Yoo, J. H. Kim, and J. M. Lee, “Dynamic modelling of mixotrophic microalgal photobioreactor systems with time-varying yield coefficient for the lipid consumption,” *Bioresour. Technol.*, vol. 162, pp. 228–235, 2014.

[27] H.-B. Chen, J.-Y. Wu, C.-F. Wang, C.-C. Fu, C.-J. Shieh, C.-I. Chen, C.-Y. Wang, and Y.-C. Liu, “Modeling on chlorophyll a and phycocyanin production by Spirulina platensis under various light-emitting diodes,” *Biochem. Eng. J.*, vol. 53, no. 1, pp. 52–56, Dec. 2010.

[28] S. Alagesan, S. B. Gaudana, S. Krishnakumar, and P. P. Wangikar, “Model based optimization of high cell density cultivation of nitrogen-fixing cyanobacteria.,” *Bioresour. Technol.*, vol. 148, pp. 228–33, Nov. 2013.

[29] J. R. Shie and Y. K. Yang, “Optimizations of a photoresist coating process for photolithography in wafer manufacture via a radial basis neural network: A case study,” *Microelectron. Eng.*, vol. 85, no. 7, pp. 1664–1670, 2008.

[30] M. A. Hosen, M. A. Hussain, and F. S. Mjalli, “Control of polystyrene batch reactors using neural network based model predictive control (NNMPC): An experimental investigation,” *Control Eng. Pract.*, vol. 19, no. 5, pp. 454–467, 2011.

[31] A. Arpornwichanop and N. Shomchoam, “Control of fed-batch bioreactors by a hybrid on-line optimal control strategy and neural network estimator,” *Neurocomputing*, vol. 72, no. 10–12, pp. 2297–2302, 2009.

[32] J. Wang and W. Wan, “Optimization of fermentative hydrogen production process using genetic algorithm based on neural network and response surface methodology,” *Int. J. Hydrogen Energy*, vol. 34, no. 1, pp. 255–261, 2009.

[33] S. Vlassides, J. G. Ferrier, and D. E. Block, “Using historical data for bioprocess optimization: Modeling wine characteristics using artificial neural networks and archived process information,” *Biotechnol. Bioeng.*, vol. 73, no. 1, pp. 55–68, 2001.

[34] Z. Xiong and J. Zhang, “Modelling and optimal control of fed-batch processes using a novel control affine feedforward neural network,” *Neurocomputing*, vol. 61, no. 1–4, pp. 317–337, 2004.

[35] M. Ronen, Y. Shabtai, and H. Guterman, “Hybrid model building methodology using unsupervised fuzzy clustering and supervised neural networks,” *Biotechnol. Bioeng.*, vol. 77, no. 4, pp. 420–429, 2002.

[36] A. Witek-Krowiak, K. Chojnacka, D. Podstawczyk, A. Dawiec, and K. Pokomeda, “Application of response surface methodology and artificial neural network methods in modelling and optimization of biosorption process,” *Bioresour. Technol.*, vol. 160, pp. 150–160, 2014.

[37] N. T. Eriksen, “Production of phycocyanin--a pigment with applications in biology, biotechnology, foods and medicine.,” *Appl. Microbiol. Biotechnol.*, vol. 80, no. 1, pp. 1–14, Aug. 2008.

[38] C.-Y. Chen, P.-C. Kao, C.-J. Tsai, D.-J. Lee, and J.-S. Chang, “Engineering strategies for simultaneous enhancement of C-phycocyanin production and CO2 fixation with Spirulina platensis,” *Bioresour. Technol.*, vol. 145, pp. 307–312, Oct. 2013.

[39] S. Van Der Walt, S. C. Colbert, and G. Varoquaux, “The NumPy array: A structure for efficient numerical computation,” *Comput. Sci. Eng.*, vol. 13, no. 2, pp. 22–30, 2011.

[40] T. E. Oliphant, “Python for scientific computing,” *Comput. Sci. Eng.*, vol. 9, no. 3, pp. 10–20, 2007.

[41] K. J. Millman and M. Aivazis, “Python for scientists and engineers,” *Computing in Science and Engineering*, vol. 13, no. 2. pp. 9–12, 2011.

[42] J. Sjoberg, “Neural networks for modelling and control of dynamic systems, M. Norgaard, O. Ravn, N. K. Poulsen and L. K. Hansen, Springer, London, 2000, xiv+246pp.,” *Int. J. Robust Nonlinear Control*, vol. 11, no. 9, pp. 881–882, 2001.

[43] Z. K. Nagy, “Model based control of a yeast fermentation bioreactor using optimally designed artificial neural networks,” *Chem. Eng. J.*, vol. 127, no. 1–3, pp. 95–109, 2007.

[44] T. Schaul, J. Bayer, D. Wierstra, Y. Sun, M. Felder, F. Sehnke, T. Ruckstieß, and J. Schmidhuber, “PyBrain,” *J. Mach. Learn. Res.*, vol. 11, pp. 743–746, 2010.