Developing process models for γ-linolenic acid production by *Cunninghamella echinulata*

A thesis submitted to

The University of Manchester

for the degree of

Master of Philosophy

in the Faculty of Science & Engineering

2023

Candidate's name: Ziqi Song

Department of Chemical Engineering and Analytical Science

School of Engineering

Table of contents

List of Tables

List of Figures

List of symbols

Symbols	Description
A_i	Specific parameters for each temperature-dependent parameter
B_i	Specific parameters for each temperature-dependent parameter
$B_C(\cdot)$	Data-driven terms underlying the parameters in the exponent of the Arrhenius expressions for glucose to biomass yield coefficient
$B_{X_B}(\cdot)$	Data-driven terms underlying the parameters in the exponent of the Arrhenius expressions for ratio of total to fat-free biomass
$B_{X_G}(\cdot)$	Data-driven terms underlying the parameters in the exponent of the Arrhenius expressions for the growth-dependent GLA yield coefficient
$B_{\mu_m}(\cdot)$	Data-driven terms underlying the parameters in the exponent of the Arrhenius expressions for maximum specific growth rate
C	Glucose concentration $(g\ L^{-1})$
$C_{e_{max}}$	Maximum measured glucose concentration
C_i	Specific parameters for each temperature-dependent parameter
K_{co}	Saturation growth constant $(g\ L^{-1})$
k_d	Specific rate of lipid turnover $(g\ g^{-1}h^{-1})$
k_m	Luedeking–Piret growth-associated constant $(g\ g^{-1})$
k_n	Luedeking–Piret non-growth-associated constant $(g\ g^{-1}h^{-1})$
k_0	Ratio of coefficient for fat free biomass on total biomass $(g\ g^{-1})$
k_p	Saturation product constant $(g\ L^{-1})$

m	Specific maintenance coefficient $(g\ g^{-1}h^{-1})$
P	Penalty term
X_B	Fat free biomass concentration $(g\ L^{-1})$
$X_{B_{e_{max}}}$	Maximum measured fat-free biomass concentration
X_G	GLA product concentration $(g\ L^{-1})$
$X_{G_{e_{max}}}$	Maximum measured GLA concentration
X_T	Total biomass concentration $(g\ L^{-1})$
$X_{T_{e_{max}}}$	Maximum measured total biomass concentration
Y_{C0}	Yield of coefficient for glucose on biomass $(g\ g^{-1})$
$Y_{C/X_T}(\cdot)$	Proportionality parameters for glucose consumption
$Y'_{C/X_T}(\cdot)$	Data-driven terms underlying glucose to biomass yield coefficient
$Y_{X_B/X_T}(\cdot)$	Proportionality parameters for fat-free biomass growth
$Y'_{X_B/X_T}(\cdot)$	Data-driven terms underlying ratio of total to fat-free biomass
$Y_{X_G/x_i}(\cdot)$	Proportionality parameters for GLA accumulation
$Y'_{X_G/x_i}(\cdot)$	Data-driven terms underlying growth-dependent GLA yield coefficient

Greek letters

δ_{ij}	Kronecker delta function
λ_i	Penalty weight
$\mu(\cdot)$	Proportionality parameters for total biomass growth
$\mu'(\cdot)$	Data-driven terms underlying maximum specific growth rate

μ_d	Specific cell death rate (h^{-1})
μ_m	Maximum specific growth rate (h^{-1})

Word count: 18,198

Abstract

Modelling plays an increasingly important role in chemical and bioprocesses nowadays and is widely used for process simulation, optimisation and real-time control. Especially for metabolic reactions with complex underlying reaction mechanisms, modelling for process analysis, prediction and control is a very cost-effective technique. In this MPhil project, a temperature-dependent kinetic model to simulate biomass growth, substrate consumption and the production of GLA by *Cunninghamella echinulata* was first proposed. The model was verified to be of high accuracy using data from a 1L bioreactor. Model aided upscaling to a 5L bioreactor with a two-stage temperature-shift strategy showed a 69.6% increment of GLA production and was verified experimentally. Then, hybrid modelling which is a state-of-the-art modelling technique and combine machine learning techniques and traditional kinetic models, was used to simulate and predict the performance of the GLA fermentation experiment by *Cunninghamella echinulata*. In addition, the hybrid models incorporated different amounts of kinetic information from a pre-existing complex kinetic model, representing different level of hybrid model 'greyness' was investigated for bioprocess predictive modelling. The results show that incorporating more specific kinetic information increased the risk of incorporating incorrect inductive bias that hindered rather than enhanced hybrid model performance. Nonetheless, the hybrid models demonstrated much improved predictive confidence with similar predictive accuracy to the original kinetic model.

Declaration

I declare that this research is the result of my own work except as cited in the references and no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright

Acknowledgements

First and foremost, I would like to thank my supervisor, Dr Dongda Zhang, for his patience, guidance and support. I have benefited greatly from his wealth of knowledge and detailed revisions. I am very grateful to him for taking me on as a student and for having faith in me over the years. I would like to express my deepest respect and heartfelt gratitude to him.

I would like to thank all my friends and colleagues for their help and support. I am grateful for all the enjoyable times I have spent with them during my MPhil.

Thanks for the financial support received for the National Natural Science Foundation of China (No. 21776232 and No. 21978244).

Thanks for my parents. They always give me the greatest encouragement and endless support.

Chapter 1 Background and thesis structure

1.1 Background

With the development of computer technology in recent decades, modelling is playing an increasingly important role in the fields of process engineering, simulation and optimisation. For bioprocesses, even within a single cell, there are thousands of enzymatically catalysed metabolic reactions, which can lead to very complex growth and product kinetics. Thus, modelling is a very important tool to qualitatively and quantitatively simplify the complex metabolic reactions within cells (Rohner & Meyer, 1995). This provides a significant contribution in safety, optimal plant design, monitoring and analysis (Maria, 2004).

Kinetic models are capable of representing complex intracellular metabolic reactions in a more complete way than most other types of models. They are able to assist in the rational design of the properties of cell factories or the production processes that they are utilized (Almquist et al., 2014), which is a useful tool for understanding bioprocesses and the impact of key operational variables due to their interpretable model structure and physical parameters (Maria, 2004). However, identifying a suitable kinetic model and appropriate parameters to accurately predict a bioprocess is challenging and time-consuming. In addition, unstructured models simplify the complex metabolic reactions and interactions within the cell, and therefore the use of unstructured models to describe and predict the dynamic of bioprocesses can lead to large model uncertainties.

Hybrid models as an alternative modelling approach combine the characteristics of data-driven models such as Artificial Neural Networks (ANNs) and Gaussian Processes (GPs) (Safarian et

al., 2021; Sheng et al., 2020) and kinetic models. Since the hybrid model balances the characteristics of a purely data-driven model and a dynamics model, it does not exhibit the same high dependence on training data as a purely data-driven model, which is prone to over-fitting. In addition, hybrid models rely on less professional knowledge and consume less time compared with dynamics models. In a hybrid model, the data-driven part is responsible for learning complex functions with concentrated parameters from the process data, and the kinetic part is responsible for reducing the non-linearity of the data-driven model. As a result, hybrid models have successfully applied in several biochemical process modelling studies (Cabaneros Lopez et al., 2021; Willis & von Stosch, 2017).

1.2 Thesis structure

The main aims of this MPhil project are to (i) investigate the effect of temperature on the *C. echinulata* biomass growth and GLA production through the construction of develop a rigorous kinetic model and (ii) develop three hybrid models of the three degrees of mechanistic complexity for case: γ-linolenic acid (C18:3n-6, GLA) fermentation by the oily fungus *Cunningham echinulata* (*C. echinulata*). This MPhil thesis is presented in the journal format, following the thesis submission guidelines approved by The University of Manchester[1]. The content of the thesis is prepared according to the following structure.

Chapter 2 summarise a dynamic model which can simulate biomass growth, substrate consumption, and GLA biosynthesis of *Cunninghamella echinulata* for a wide temperature

[1] https://documents.manchester.ac.uk/display.aspx?DocID=7420

range. Then, this model was validated using data from a 1L bioreactor, and it was found the optimal temperatures for biomass growth and GLA production. In addition, the new upscaling experiment with 5L bioreactor and a two-stage temperature-shift strategy was carried based on this dynamic model.

Hybrid modelling provides a cost-effective solution to modelling complex biochemical reaction kinetics when the underlying mechanisms are not fully understood. However, a bi-dimensional bias-variance trade-off is a challenge for hybrid modelling. Therefore, in Chapter 3, three hybrid models incorporated different amounts of kinetic information from a pre-existing complex kinetic model were presented to show the fitting and uncertainty performance for the fermentation of the fungus *Cunninghamella echinulata*. In addition, the validation experiment with the temperature-shift dynamics for the upscaled 5 L bioreactor was conducted to investigate predictive performance of hybrid models.

Chapter 2 Kinetic modelling of γ-linolenic acid production by *Cunninghamella echinulate*

This chapter sheds light on building a dynamic model capable of simulating biomass growth, substrate consumption, and GLA biosynthesis of *Cunninghamella echinulata* for a wide temperature range was proposed for the first time. In addition, the model was verified to be of high accuracy using data from a 1L bioreactor, and the optimal temperatures for biomass growth and GLA production was found. Model aided upscaling to a 5L bioreactor with a two-stage temperature-shift strategy of GLA production and was verified experimentally. This presents important advances for the upscaling of GLA production biotechnology from laboratory-scale to pilot-scale.

2.1 Introduction

2.1.1 Introduction of GLA and temperature-shift method

The polyunsaturated fatty acid (PUFA), γ-linolenic acid (C18:3n-6, GLA), is widely utilised within the pharmaceutical and nutraceutical industries due to its outstanding value for treating diseases like rheumatoid arthritis (Jäntti et al., 1989), multiple sclerosis (Barber, A. J., 1988), schizophrenia (D.F Horrobin, 1979) and atopic eczema (Scott, J., 1989). As an essential precursor for the biosynthesis of several prostaglandins, GLA is an essential fatty acid with proven anti-inflammatory and anti-cancer effects (Wan, 2009). However, with the body unable to manufacture its own (Somashekar et al., 2003), GLA must be assimilated through the consumption of poultry, beef, pork, chicken or egg yolk, or else as a dietary supplement.

However, its natural concentration in such foodstuffs is small. For instance, 100 g of raw beef lean contains 2.2 mg of GLA (Horrobin, 1992).

Plant seeds such as borage, black currant, evening primrose, and hemp have been used as commercial sources of GLA oil (Tanticharoen et al., 1994). However, cultivating these plants requires large swathes of arable land owing to their seeds low intracellular GLA content, which also hinders the economic viability of the seed-based extractive methods. In contrast, a promising alternative source: the oleaginous fungus *Cunninghamella echinulata* (*C. echinulata*), possesses a much higher GLA content of 33.5 mg g^{-1} (Kavadia et al., 2001), a higher growth rate of 0.04 $g\ L^{-1}\ h^{-1}$ (Chatzifragkou et al., 2010), and is simpler to cultivate (i.e., traditional fermentation). Taken together, this marks the fungus as an economically competitive source of GLA for large-scale production.

In response, there have been many attempts to develop a commercially viable *C. echinulata* GLA production bioprocess (Čertík et al., 2006; Fakas et al., 2008). Primarily, a multitude of cultivation conditions have been explored to optimise GLA yield and adjust the PUFAs composition, including different medium compositions (e.g., glucose, fructose, whey concentrate (Chatzifragkou et al., 2010; Fakas et al., 2008)), temperatures (e.g., 15 ℃ to 35 ℃ (Fakas et al., 2008)) and pH (e.g., 7 to 10 (Dyal et al., 2005)). However, the cost of feedstock (i.e., low-cost carbon and nitrogen sources) is also widely studied to lower the GLA production cost (Chen & Chang, 1996; Gema et al., 2002). Despite these efforts, the operating temperature remains a critical factor in the control and optimisation of the bioprocess. However, only a

handful of studies (Al-Hawash et al., 2018; Li et al., 2019) have investigated the effect of temperature on lipid accumulation and the composition of PUFAs within *C. echinulata*.

Temperature-shift: an approach where the operating temperature is switched during cultivation to increase the accumulation of the targeted metabolite has been reported to influence the intracellular PUFAs content of fungal species (Jang et al., 2005; Lindberg & Molin, 1993). Specifically, a two-stage temperature-shift strategy employed during the cultivation of the arachidonic acid (ARA)-producing fungus *Mortierella alpine* (*M. alpine*) increased ARA production and total fatty acid concentration by 26.1% and 20% by switching the temperature from 25 °C to 20 °C at the later stage of fermentation (Peng et al., 2010).

However, whilst both *C. echinulata* and *M. alpine* are oleaginous fungi, the optimal temperature for biomass growth, lipid production, and product accumulation might differ significantly between strains. Hence, it is valuable to investigate the optimal temperature-shift strategy for industrial GLA production with *C. echinulata*. Furthermore, previous work has overlooked kinetic model-driven optimisation and control of such temperature-shift strategies (Peng et al., 2010; Sivagurunathan et al., 2014; M. L. Sun et al., 2017), despite the insight kinetic models capture about the dynamic behaviour of the underlying biological system. Moreover, model-based design of experiments is considered an effective tool to accomplish bioprocess scale-up from laboratory to pilot and industrial scale (Zhang et al., 2015).

2.1.2 The aims of this work

This work aims to: (i) investigate the effect of temperature on the *C. echinulata* biomass growth and GLA production through the construction of a rigorous kinetic model; (ii) evaluate the

performance of the temperature-shift strategy when up-scaling the fermentation processes from a 1L to 5L bioreactor; (iii) evaluate the predictive accuracy and sensitivity of the kinetic model over different bioreactor scales. These objectives underpin this work and are organised into two sections: the model construction procedure and the discussion of the results and implications of the research.

2.2 Literature review

As mentioned in section 2.1, the oleaginous fungus *Cunninghamella echinulata* (*C. echinulata*) as a promising alternative source can be used for large-scale GLA production. The cultivation can be affected by a multitude of cultivation conditions and the operating temperature remains a critical factor in the control and optimisation of the bioprocess. This review will focus on three major themes throughout the literature review. These themes are: (i) the construction of a rigorous kinetic model of *C. echinulata* fermentation, (ii) the effect of temperature on the *C. echinulata* biomass growth and GLA production, (iii) the predictive accuracy and sensitivity of the kinetic model.

2.2.1 The construction of kinetic model

The kinetic model constructed will simulate *C. echinulata* total and fat-free biomass growth, glucose consumption, and GLA accumulation. The Monod model, as one of the most notable unstructured models, was first introduced by Jaccques Monod in 1942 to describe the relationship between specific growth rate and substrate utilisation rate in a bioreactor (Monod, 1949). The Monod model can be defined by Eq.2.1.

$$\mu = \frac{\mu_{max} S}{K_s + S} \tag{2.1}$$

Where, μ specific growth rate, μ_{max} maximum specific growth rate, S substrate concentration, X biomass concentration, K_s half saturation constant. The Monod model assumes that there is only one growth limiting substrate in the bioreactor (Monod, 1942). However, the Monod model has several limitations e.g. the maximum specific growth rate is independent of the substrate concentration at high substrate concentratiom; growth is dependent on substrate concentration at low substrate concentration; the model cannot describe lag and death phease during the growth stage.

The Contois model was first introducted by Contois in 1959 (Contois, 1959). Muloiwa, Stephen Nyende-Byakika and Megersa Dinka (2020) summaried the difference between Monod model and Contios model. They indicated that the Contios model is an extension of the Monod model and it is an unstructured model based on the both substrate and biomass concentration. The model structure of Contois model is shown below:

$$\mu = \frac{\mu_{max} S}{K_s X + S} \tag{2.2}$$

Where, μ specific growth rate, μ_{max} maximum specific growth rate, S substrate concentration, X biomass concentration, K_s half saturation constant. In the Contios model, the assumption is that biomass concentration is inversely proportional to specific growth rate. Mrwebi (2004) indicated that the Contios model has not been widely used to describe specific growth rate, and when used, it produced satisfactory fit on the experimental data. Abdullah et al., (2006) reported a satisfactory fit on *Aspergillus oryzae* in industrial natural rubber effluent serum. The Contois model produced R^2 of 80.2%.

Moreover, the Logistic model as another unstructure model which is also widely used. The Logistic model was first introduced by Pierre Verhulst in 1838 (Verhulst, 1838). This model is based on an assumption that the growth rate of cells is proportional to the current population, and the unutillised resources in a closed habitat (Horowitz et al., 2010). The model is a substrate independent model, it describe kinetic growth based on biomass cconcentration only (Ali et al., 2017). The Logistic model is defined below:

$$\mu = \mu_{max}\left[1 - \frac{X}{X_m}\right] \tag{2.3}$$

Where, μ specific growth rate, μ_{max} maximum specific growth rate, X biomass concentration, X_m is the maximum biomass concentration.

The Haldane model as another extension of the Monod model was introduced by Haldane in 1930. This model introduces an inhibition term K_i which deals with specific growth rate inhibition at very high or low substrate concentration. K_i is the inhibition constant which is equal to the highest substrate concentration at which the specific growth rate is equal to one half maximum growth rate in the absence of an inhibition (Muloiwa et al., 2020). The Logistic model is defined by Eq.2.4:

$$\mu = \frac{\mu_{max} S}{K_s + S + \frac{S^2}{K_i}} \tag{2.4}$$

Where, μ specific growth rate, μ_{max} maximum specific growth rate, S substrate concentration, X biomass concentration, K_s half saturation constant and K_i inhibition constant. The advantage of the Haldane model is that it is capable of describing all growth phases: lag, exponential, stationary, and death phase (Dutta et al., 2015). In addition, it is able to describe growth rate at low and high substrate concentration.

In this work, there is no inhibition of substrate so the Haldane model is not considered. On the other hand, the Contois model was adopted to describe biomass growth because of its accurate fit on the experimental data and successful capture on the inverse proportionality between biomass growth and biomass concentration.

Moreover, several literatures developed the kinetic model of lipid phenotypes of fungal strains, enabling to distinguish three lipid-producing stages, including low lipid-producing, lipid accumulation and lipid turnover stages. Wannawilai et al., (2020) developed the mathematical model which can be used to describe the lipid-producing stages of both the wild type and morphologically engineered strain of *Aspergillus oryzae.* The kinetic model structure is shown below:

$$\frac{dC_X}{dt} = \mu_{max} C_X \left(1 - \frac{C_X}{C_{X_m}}\right) \tag{2.5}$$

$$\frac{dC_S}{dt} = -\left[\left(\frac{1}{Y_{\frac{X}{S}}}\right)\left(\frac{dC_X}{dt}\right) + m_S C_X\right] \tag{2.6}$$

$$\frac{dC_P}{dt} = (\alpha\mu + \beta)C_X - k_d \left(1 - \frac{C_S}{C_{S_0}}\right)^n C_X \tag{2.7}$$

The kinetic model structure is constructed by the Logistic model. For lipid-producing stage, authors used $(\alpha\mu + \beta)C_X$ this term to sucessfully describe the lipid-producing and lipid accumulation. Moreover, $k_d \left(1 - \frac{C_S}{C_{S_0}}\right)^n C_X$ this term accurately fit the lipid-producing stages, including low lipid-producing and lipid turnover stages. The models fitted well with the experimental data ($R^2 \geq 0.96$), indicating that lipid production profiles could be reasonably modeled.

Moreover, Antimanon et al., (2020) also developed the mathematical models to describe that the lipid and Dihomo-γ-linolenic acid (DGLA) were growth-associated metabolites corresponding to the relevant kinetic parameters of fermentations. The kinetic model they constructed is as same as literature by Wannawilai et al., (2020).

$$\frac{dX}{dt} = \mu_{max} X \left(1 - \frac{X}{X_m}\right) \tag{2.8}$$

$$\frac{dS}{dt} = -\left[\left(\frac{1}{Y_{\frac{X}{S}}}\right)\left(\frac{dX}{dt}\right) + m_S X\right] \tag{2.9}$$

$$\frac{dP}{dt} = (\alpha\mu + \beta)X - k_d \left(1 - \frac{S}{S_0}\right)^n X \tag{2.10}$$

The formation mode of each product was defined according to the fungal growth. If α = 0 and β = 0, the product formation is the growth-related production. On the contrary, if α = 0 and β = 0, the product formation is unrelated to fungal growth. In case if the product formation is mixed with the growth production, the value α = 0 and β = 0. $k_d \left(1 - \frac{S}{S_0}\right)^n$ this term can be used to describe the lipid turnover phase which can cause the decay of lipid concentration.

This literature provides an informative perspective in the *n*-6 fatty acid production through physiological manipulation, thus leading to a prospect in viable production of the DGLA-enriched oil by the engineered strain. Therefore, GLA accumulation was simulated by adapting the Luedeking-Piret equation (Luedeking & Piret, 1959) to include a novel GLA consumption term which can be used to demonstrate lipid production rate when substrate was exhausted in this work.

2.2.2 The effect of temperature

As shown in Figure 2.1, the inverse correlation between biomass growth and GLA accumulation with increasing operating temperature. Therefore, an accurate kinetic model

capable of simulating GLA production must incorporate the effect of operating temperature. Phisalaphong et al. (2006) investigated temperature effect on kinetic parameters of ethonal fermentation using mathematical models. The relationship between temperature and parameters in their literature is shown below:

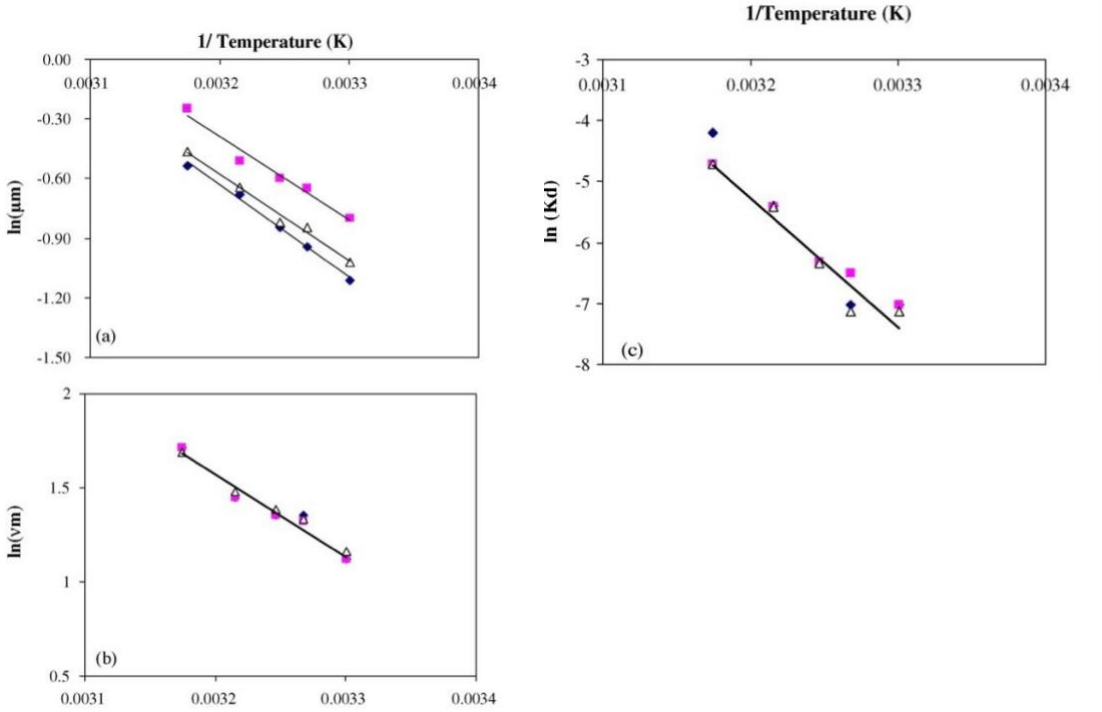


Figure 2.1 The result of the effect of temperature on parameters μ_m, K_d and v_m illustrated by Phisalaphong et al. (2006)

These parameters increased exponentially as the temperature increased. Expressed by the Arrhenius relationship, the temperature dependency of the reaction rate is fitted very well with the experimental data. They investigated and quantified the influence of temperature and initial sugar concentration on cell activites. Arrhenius relationships between operating temperature and the maxmium specific growth rate, speciific production rate, specific death rate were then established. The performance of a mathematical model to describe the dynamic

behaviour of the ethanol fermentation process were assessed.Similarly, Rivera et al. (2007) proposed an equation to describe the influence of temperature and fit it to the optimised values obtained for each temperature. The temperature-dependent parameter equations they used are shown below:

$$\theta_i = A_i \cdot \exp\left(\frac{B_i}{T}\right) \tag{2.11}$$

$$\theta_i = A_i \cdot \exp\left(\frac{B_i}{T}\right) + C_i \cdot \exp\left(\frac{D_i}{T}\right) \tag{2.12}$$

In these equations, A, B and C are constants and T is temperature in ℃. The fitting results illustrated in their literature indicate that the temperature-dependent kinetic parameters described accurately the experimental data.

Moreover, (Laidler, 1984) summarised the development of the Arrhenius equations. The author illustrated the most common used Arrhenius equation which introduced by Van't Hoff in 1884. The expression is shown below:

$$k = Ae^{-B/T} \tag{2.13}$$

In additions, other temperature-dependence equations were introduced. e.g. an equation in which the pre-exponential factor also has a temperature dependence was first proposed in 1893 by Kooi, whose equation expressed by:

$$k = AT^c e^{-B/T} \tag{2.14}$$

Also Van't Hoff pointed out that most of the previously presented equations with special cases of the equations in 1898:

$$k = AT^c e^{-(B-DT^2)/T} \tag{2.15}$$

Although these empirical equations are widely different, all the equations can give a reasonably good fit to the experimental data. These temperature-dependent equations have a great enlightenment to find the suitable empirical equations in my work.

2.2.3 Sensitivity analysis of the kinetic model

The sensitivity analysis is evaluated for their usefulness as part of the model-building within process analysis technology applications. Considering the complexity of typical mechanical models of microbial processes, sensitivity analysis is needed to minimise the risk of errors in the process of model construction. Sin et al. (2009) introduced the sensitivity analysis techniques. In their work, the subjective input uncertainty was defined after an expert review process. Expert review involves asking the opinion of process experts (and/or consulting the relevant literature resources) about the uncertainty of the parameters. In order to structure the expert review process, it is assumed that all the model parameters have a uniform probability distribution. In this way, the minmum and maximum values of the uniform distribution can be calculated as follows:

$$\theta_{min} = (1 - \%Variation) \times \theta_{mean} \quad (2.16)$$

$$\theta_{max} = (1 + \%Variation) \times \theta_{mean} \quad (2.17)$$

Then the well-known Latin Hypercube Sampling method was used for probabilistic sampling of the input space and the sampled input matrix $\theta_{N\times M}$was propagated through the dynamic model by performing *N* dynamic simulations. Overall, the sensitivity analysis is believed to help establish a reliable mechanical model and correctly interpret the model outputs, which helps to improve process understanding, process optimisation and control purposes.

2.3 Methodology

The experiment setup work was completed by our co-operators of Xiamen University. The experiment setup is detailly displayed in Appendix A.

2.3.1 Kinetic model construction

The kinetic model constructed will simulate *C. echinulata* total and fat-free biomass growth, glucose consumption, and GLA accumulation over a range of operating temperatures for the experiments visualised in Figure 2.2.

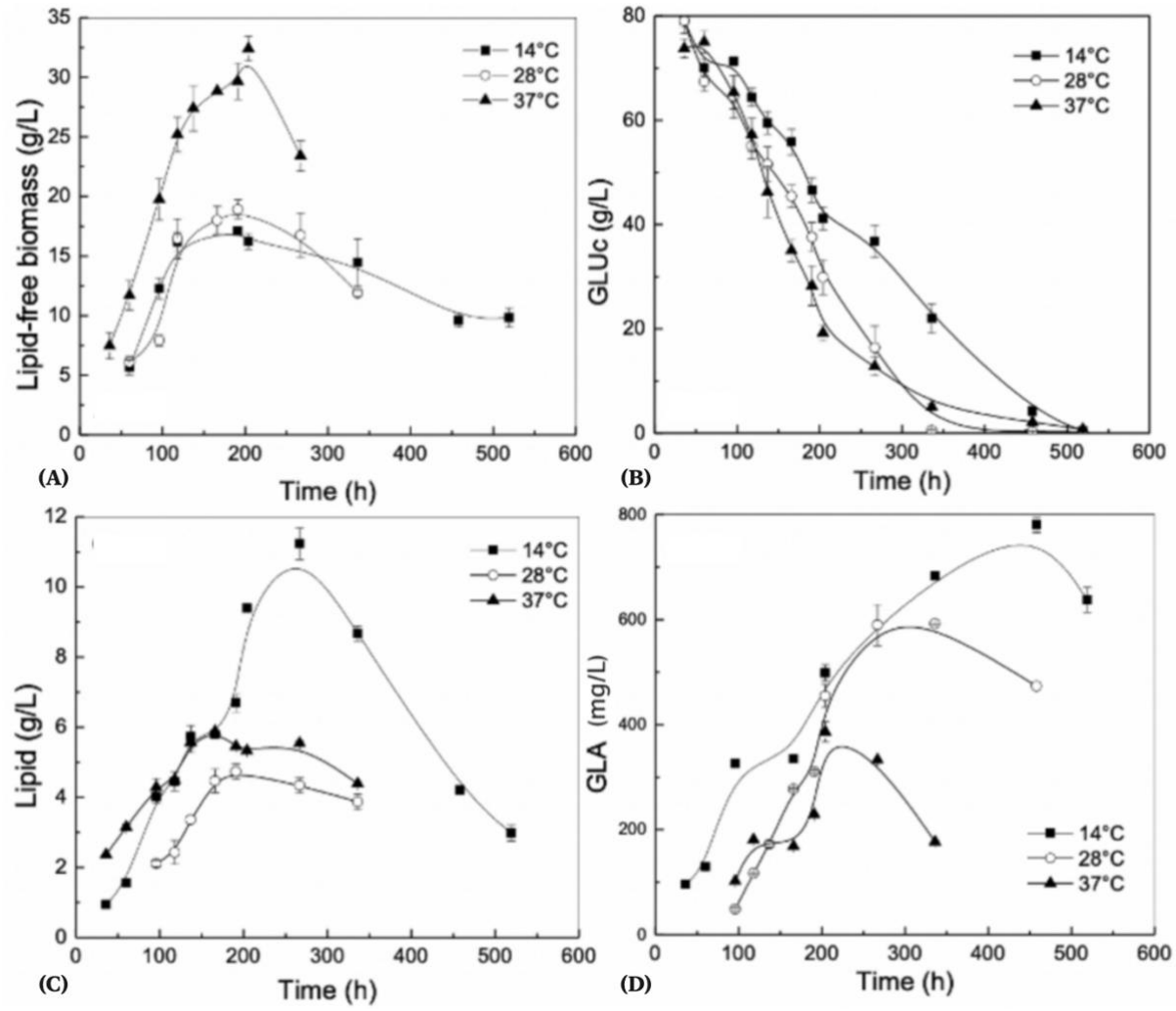


Figure 2.2 Experiment result obtained by Xiamen University on the effect of temperature on lipid-free biomass production (A), residual sugar concentration (B), lipid accumulation (C) and

GLA synthesis (D) during batch growth of C. echinulata. The symbols used were 14 ℃ (square), 28 ℃ (empty circle), 37 ℃ (triangle). Error bars indicated on the mean of triplicate experimental runs.

Two key features characterise the batch experiments shown in Figure 2.2: (i) the soft transition over 250 to 300 hours from glucose-replete to glucose-limiting conditions; (ii) the inverse correlation between biomass growth and GLA accumulation with increasing operating temperature. Therefore, an accurate kinetic model capable of simulating GLA production must incorporate the effects of culture glucose concentration and operating temperature. This will be the focus of the proceeding two sections.

2.3.1.1 Model structure identification

The soft transition from glucose-replete to glucose-limiting conditions between 250 and 300 hours impacts biomass growth earlier than it does GLA accumulation. From Figure 2.2(A), Figure 2.2(B) and Figure 2.2(C), the total biomass concentration is observed to peak then decline beyond 200 hours irrespective of temperature despite glucose concentration remaining relatively high (i.e., > 30 g L^{-1}). Whereas, Figure 2.2(B) and Figure 2.2(D) show that GLA concentration continues to accumulate, only decaying later once glucose drops below about 15 g L^{-1}. Taken together, this suggests that GLA is a secondary metabolite (Shamloo et al., 2017) with an accumulation tendency independent from biomass growth, where GLA is consumed to support cell maintenance once the carbon source becomes limiting. Leveraging this insight, the kinetic model structure could now be motived.

The Contois model accounts for the effect of cell flocculation and diffusional barriers that arise in high-density cell cultures. Unlike the Monod model (Zambrano & Carlsson, 2014), the Contois model (Abdullah et al., 2016; Muloiwa et al., 2020; S. L. Sun et al., 2009) successfully captures the inverse proportionality between biomass growth and biomass concentration that leads to the decline at 200 hours despite glucose concentration remaining high. Thus, the Contois model was adopted in this study to describe biomass growth. Equation (2.18) simulates the total biomass growth rate, the first term on the right-hand side represents total biomass growth whilst the second term represents endogenous cell decay. Building upon this, Equation (2.19) simulates fat-free biomass growth rate, assuming that the fat-free biomass growth rate is proportional to total biomass growth rate, as apparent in Figure 2.2(A) and Figure 2.2(C).

$$\frac{dX_T}{dt} = \mu_m \cdot \frac{C}{K_{co} \cdot X_T + C} \cdot X_T - \mu_d \cdot X_T \tag{2.18}$$

$$\frac{\mathrm{d}X_B}{\mathrm{d}t} = k_0 \cdot \frac{\mathrm{d}X_T}{\mathrm{d}t} = k_0 \cdot (\mu_m \cdot \frac{C}{C + X_T \cdot K_{co}} \cdot X_T - \mu_d \cdot X_T) \tag{2.19}$$

Where X_T is total biomass concentration (g L^{-1}), C is glucose concentration (g L^{-1}), μ_m is maximum specific growth rate (h^{-1}), μ_d is specific cell death rate (h^{-1}), K_{co} is half-saturation constant and k_0 is the ratio of fat-free biomass to total biomass (g g^{-1}).

Equation (2.20) simulates glucose consumption rate. The first term on the right-hand side is the glucose consumption rate for cell growth, whilst the second is glucose uptake for maintenance activities.

$$\frac{dC}{dt} = -Y_{C0} \cdot \left(\mu_m \cdot \frac{C}{K_{co} \cdot X_T + C} \cdot X_T - \mu_d \cdot X_T\right) - m \cdot X_T \tag{2.20}$$

Where Y_{C0} is yield coefficient from glucose to biomass (g g^{-1}) and m is biomass specific maintenance coefficient (g $g^{-1}h^{-1}$).

Finally, GLA accumulation was simulated by adapting the Luedeking-Piret equation (Luedeking & Piret, 1959) to include a novel GLA consumption term. In Equation (2.21), the first two terms reflect the growth-dependent and growth-independent synthesis rates, proportional to biomass growth rate and biomass concentration, respectively. Whereas the third term represents the rate of GLA consumption for maintenance under glucose limiting conditions. Examining the term more closely, when glucose is replete, $\frac{1}{C+K_p} \to 0$, glucose is consumed as usual; however, when glucose becomes limiting, $\frac{1}{C+K_p} \to \frac{1}{K_p}$, starved of any extracellular carbon source for cell maintenance, GLA is consumed instead. Thus:

$$\frac{\mathrm{d}X_G}{\mathrm{d}t} = k_m \cdot \left(\mu_m \cdot \frac{C}{C + K_{co} \cdot X_T} \cdot X_T - \mu_d \cdot X_T\right) + k_n \cdot X_T - k_d \cdot \frac{1}{C + K_p} \cdot X_T \qquad (2.21)$$

Where k_m is the growth-dependent synthesis constant (g g^{-1}), k_n is the growth-independent synthesis constant (g $g^{-1}h^{-1}$), k_d is the specific GLA decay rate (g $g^{-1}h^{-1}$) and K_p is the saturation product constant (g L^{-1}) for GLA decay.

2.3.1.2 Simulating temperature effects

Figures 2.2(A), 2.2(C), and 2.2(D) illustrate the inverse correlation between biomass growth and GLA accumulation with increasing operating temperature: whilst higher temperatures promote total and fat-free biomass growth, GLA accumulation is reduced. Therefore, some kinetic parameters will correlate positively with temperature within the proposed model structure, whilst others will correlate negatively.

In order to investigate the temperature dependence of each kinetic parameter, seven were modelled as a function of temperature (μ_m, K_{co}, Y_{C0}, K_0, k_m, k_d and K_p), whilst the remaining three parameters (m, μ_d, k_n) where considered temperature independent. The

positive or negative temperature dependence was captured by either Equation (2.22), or Equation (2.23) (Laidler, 1984), respectively, where the former is simply the standard Arrhenius equation.

$$\theta_i = A_i \cdot \exp\left(\frac{B_i}{T}\right) \tag{2.22}$$

$$\theta_i = C_i - A_i \cdot \exp\left(\frac{B_i}{T}\right) \tag{2.23}$$

Where A_i, B_i and C_i in the above equations are specific parameters to be fitted for each kinetic parameter θ_i and T is the temperature in Kelvins K. Thus, two or three constants describe each of the seven temperature-dependent parameters.

2.3.2 Parameter estimation method

The 24 parameters required by the model were estimated by formulating the nonlinear least-squares optimisation problem defined by Equations 2.24 (a) to (j), identifying parameter vector parameters θ by minimising the objective function (Del Rio-Chanona et al., 2015):

$$\min_{\theta} E(\theta) = \sum_{n=1}^{np} \left[\frac{\left(X_{T_n} - X_{T_{e_n}}\right)^2}{X_{T_{e_{max}}}^2} + \frac{\left(C_n - C_{e_n}\right)^2}{C_{e_{max}}^2} + \frac{\left(X_{B_n} - X_{B_{e_n}}\right)^2}{X_{B_{e_{max}}}^2} + \frac{\left(X_{G_n} - X_{G_{e_n}}\right)^2}{X_{G_{e_{max}}}^2} \right] \tag{2.24a}$$

Subject to:

$$\frac{dX_T}{dt} = f(X_T(t), \theta), \qquad t \in [t_0, t_f] \tag{2.24b}$$

$$\frac{dC}{dt} = f(C(t), \theta), \qquad t \in [t_0, t_f] \tag{2.24c}$$

$$\frac{dX_B}{dt} = f(X_B(t), \theta), \qquad t \in [t_0, t_f] \tag{2.24d}$$

$$\frac{dX_G}{dt} = f(X_G(t), \theta), \qquad t \in [t_0, t_f] \tag{2.24e}$$

$$\theta_{lb} \leq \theta \leq \theta_{ub} \tag{2.24f}$$

$$X_T(t_0) = X_{T_0} \tag{2.24g}$$

$$C(t_0) = C_0 \tag{2.24h}$$

$$X_B(t_0) = X_{B_0} \tag{2.24i}$$

$$X_G(t_0) = X_{G_0} \tag{2.24j}$$

Where $X_{T_{e_n}}$, C_{e_n}, $X_{B_{e_n}}$ and $X_{G_{e_n}}$ are the measured concentrations of biomass, substrate, fat-free biomass and GLA respectively, at each sampling time n. Whilst, X_{T_n}, C_n, X_{B_n} and X_{G_n} are the respective concentrations computed by the model at each sampling time n, and $X_{T_{e_{max}}}$, $C_{e_{max}}$, $X_{B_{e_{max}}}$ and $X_{G_{e_{max}}}$ are the maximum measured concentrations and np is the number of sampling points.

This derivative-based nonlinear programming problem (NLP) was solved by adopting the parameter estimation framework widely employed by (del Rio-Chanona et al., 2017). Given the high nonlinearity and stiffness of the system (i.e., components of different orders of magnitude and time scales), the differential system of equations was discretised by direct transcription by orthogonal collocation (Biegler, 1984) into a series of nonlinear algebraic equations. The NLP was then solved using the interior point nonlinear optimisation solver IPOPT (Wächter & Lorenz T. Biegler, 2006) implemented in the open-source Python optimisation environment: *Pyomo* (Hart et al., 2017).

2.3.3 Sensitivity analysis

Although the solution to a parameter estimation problem may provide point estimates that agree with the data instantaneously, questions about model stability remain unaddressed. Therefore, model sensitivity to parameter uncertainty was investigated by resampling each parameter in turn with all others fixed, simulating the state variable trajectories each time (Franceschini &

Macchietto, 2008). These trajectories were then aggregated into the propagated uncertainty bounds shown later. All the model parameters were assumed to be distributed uniformly between a lower bound (P_{min}) and an upper bound (P_{max}) centred on the mean (P_{mean}), defined as follows (Sin et al., 2009):

$$P_{min} = (1 - \%Variation) * P_{mean} \quad (2.8)$$

$$P_{max} = (1 + \%Variation) * P_{mean} \quad (2.9)$$

In total, Latin Hypercube Sampling (LHS) of the input space selected 200 parameter samples. All code was executed in Python version 3.7 using the *SMT 1.0.0*, *SciPy* libraries and *NumPy*.

2.4 Results and discussion

2.4.1 Results of model construction

The values of the parameter estimates are shown in Table 2.1.

Table 2.4.11 Parameters values with different operation temperatures

T (℃)	μ_m (h^{-1})	K_{co} (g L^{-1})	Y_{C0} (g g^{-1})	K_0 (g g^{-1})	k_m (g g^{-1})
14	0.115	41.345	1.420	0.700	0.011600
28	0.164	49.586	1.307	0.780	0.003085
37	0.218	55.000	0.591	0.901	0.001173
	k_d(g $g^{-1}h^{-1}$)	K_p (g L^{-1})	μ_d (h^{-1})	m (g $g^{-1}h^{-1}$)	k_n (g $g^{-1}h^{-1}$)
14	0.00352	25.013	0.0017	0.00498	0.00013814
28	0.00480	34.277	0.0017	0.00498	0.00013814
37	0.00727	37.888	0.0017	0.00498	0.00013814

In addition, the average relative error between model fitting performance and experimental results are shown in Table 2.2.

Table 2.4.12 The relative percentage error between model fitting and experimental results

Output variable	Relative percentage error (%)		
	T = 14 ℃	T = 28 ℃	T = 37 ℃
$\boldsymbol{X_T}$	6.56	2.55	8.06
$\boldsymbol{C}$	12.01	20.70	16.28
$\boldsymbol{X_B}$	8.63	2.95	6.98
$\boldsymbol{X_G}$	12.10	20.81	18.67

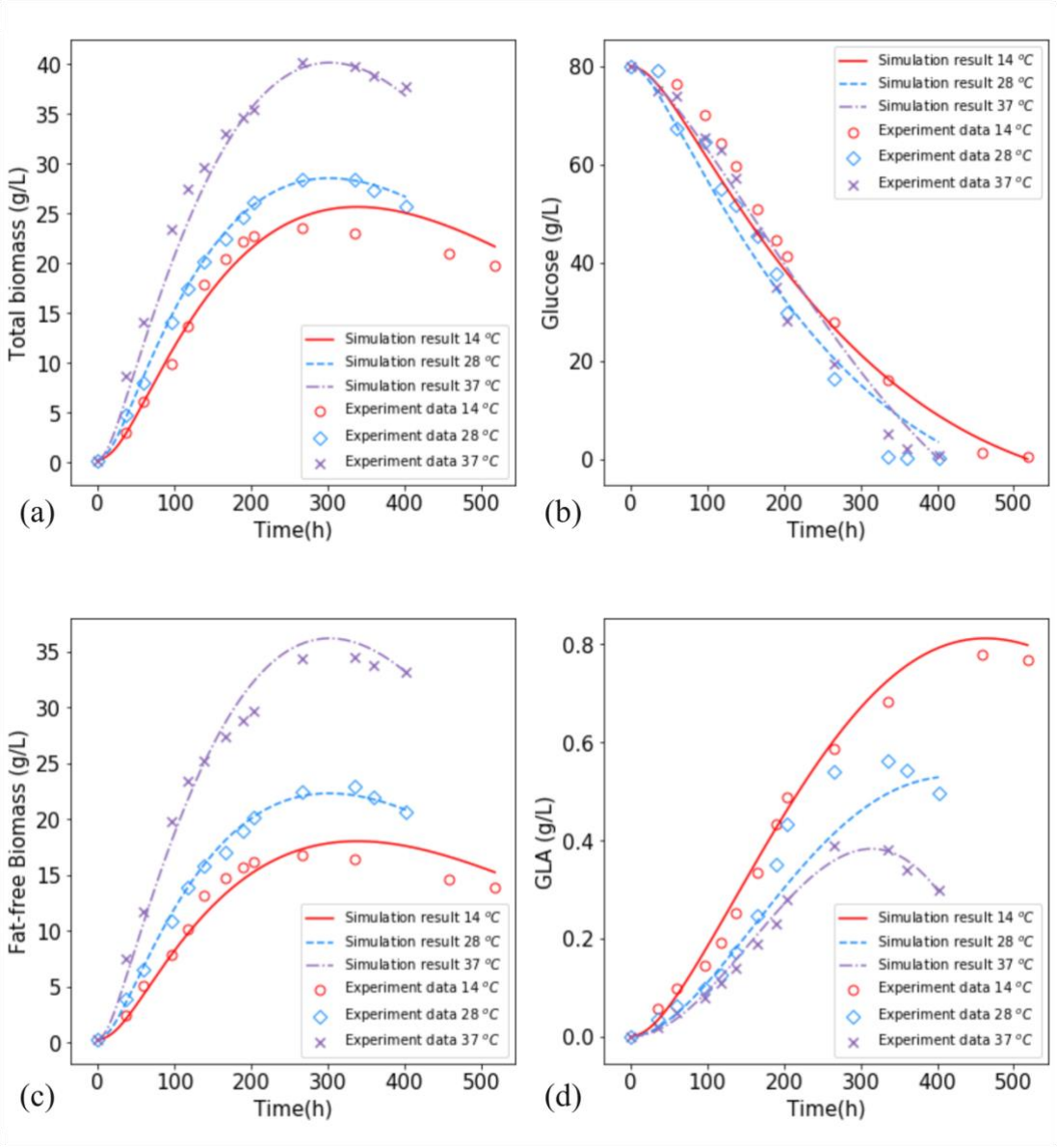


Figure 2.3 Experimental results and simulation fitting results for (a) total biomass, (b) glucose, (c) fat-free biomass and (d) GLA.

From Figure 2.3, it is evident that the model accurately captures the bioprocess dynamics over the different operating temperatures, with the state variables' mean relative percentage error falling within 20%, as shown in Table 2.2. Furthermore, features observed experimentally are successfully described by the proposed model structure. For instance, Figure 2.3 illustrates how the experimental and simulated total and fat-free biomass concentrations pass through a

maximum at around 270 hours; GLA concentration peaks similarly, albeit later on around 300 hours. The peaks at 270 and 300 hours mark the transition from glucose-replete to glucose-limiting conditions for cell growth and bioproduct synthesis, respectively.

Examining GLA accumulation more closely, GLA accumulation is characterised by an initial lag phase that proceeds exponential growth. This suggests that *C. echinulata* only biosynthesises GLA once sufficient primary metabolites have accumulated during cell growth under nutrient replete conditions, an observation consistent with the findings by (Chen & Chang, 1996), where GLA was reported to form as a secondary metabolite. Thus, physically motivating the first two terms on the right-hand side of Equation (21). However, beyond the exponential growth phase, when glucose becomes limiting, the third term on the right-hand side of Equation (21) dominates, and GLA concentration drops sharply. Thus, the kinetic model can accurately capture the GLA decay dynamics at 14 °C and 37 °C illustrated in Figure 2.2(D). This demonstrates the high fidelity of the modified Luedeking-Piret equation proposed herein for simulating the temperature-dependent biokinetics of *C. echinulata*. However, whilst the model struggles to fit the GLA decay stage at 28 °C, the error remains relatively small. Therefore, any marginal improvement made by introducing more parameters (e.g. adding the degree of lipid depletion to $1 / (C + K_p)$ which can reduce error by 5%), is outweighed by the risk of overfitting and poor predictive accuracy on unseen process states.

2.4.2 Model sensitivity analysis

The sensitivity of the simulated trajectories with respect to the kinetic model parameters was investigated as described in Section 2.4. That is, each parameter, in turn, was resampled 200

times between P_{min} and P_{max} with a variation of 5%, whilst the other parameters were held constant, simulating the state variable trajectories each time. This approach of uncertainty input characterisation is similar to the method which was mentioned by (Sin et al., 2009). Figure 2.4 displays the propagated uncertainty for total biomass, fat-free biomass, glucose, and GLA concentration resulting from a 5% variation in parameter m at an operating temperature of 14 °C.

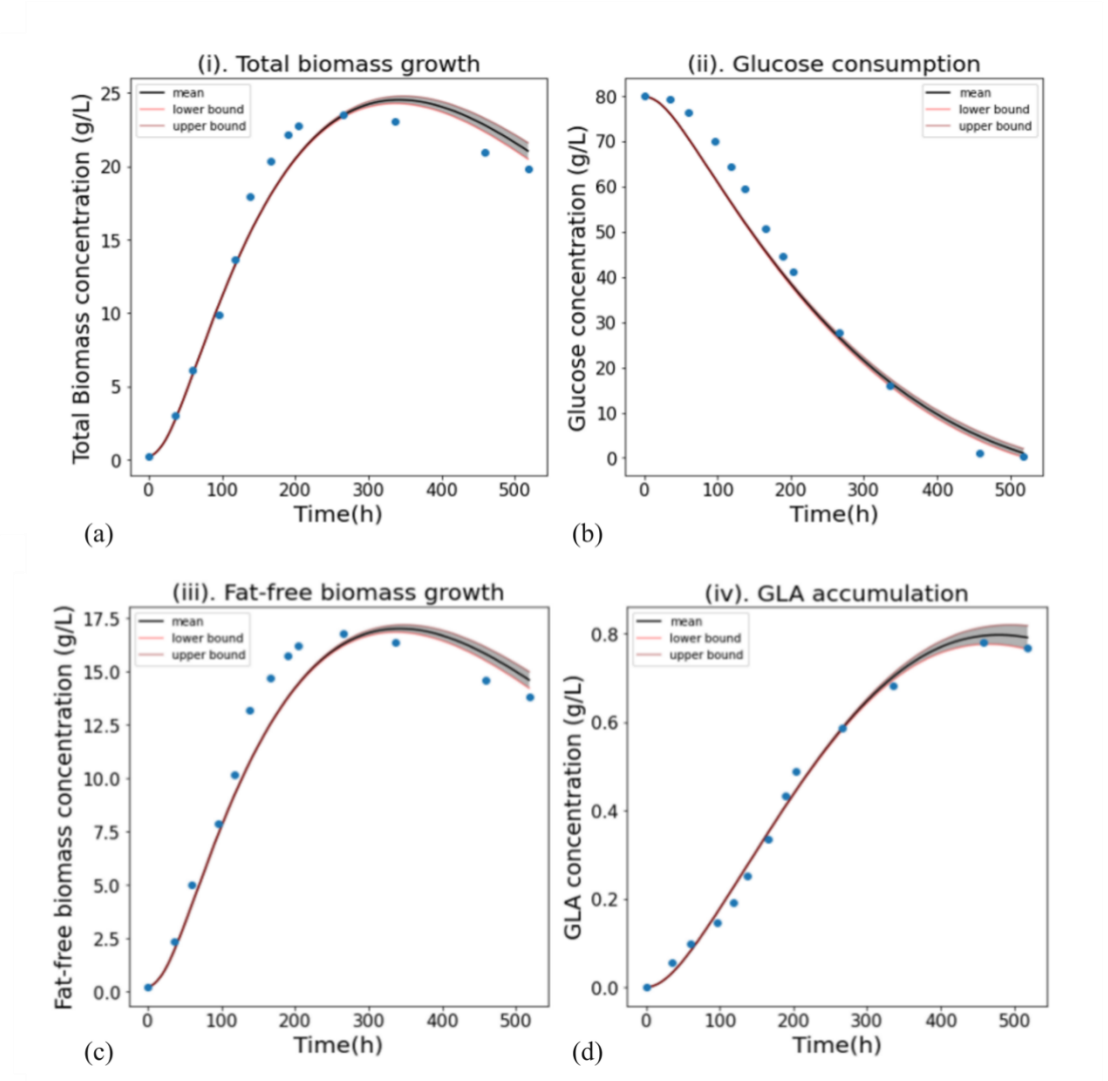


Figure 2.4 Uncertainty with 5% variation for parameter m in the model predictions for (a) total biomass, (b) glucose, (c) fat-free biomass and (d) GLA.

In Figure 2.4, the spread of the uncertainty bands reflects the sensitivity of the model's prediction to uncertainty in parameter m, where the propagated uncertainty increases with time. Supplementary uncertainty projections for the remaining model parameters and temperatures can be found in Appendix B.

For a more comprehensive comparison of the model's sensitivity to each parameter, Table B.1 (see Appendix B) tabulates the mean relative percentage deviation (MRPD) between the lower and upper bound for each state variable and operating temperature.

Upon inspecting Table A.1, it is evident that the model is more sensitive to B_i than either A_i or the remaining temperature-independent parameters, an expected result given that B_i modifies the exponent of the temperature-dependent parameters. Of these, the MRPD is largest for B_{μ_m}, $B_{k_{co}}$ and $B_{Y_{co}}$ in decreasing order, and therefore contribute most to model uncertainty. Moreover, the temperature-dependent parameters μ_m, k_{co} and Y_{co}, are critical to simulating all four state variables, unlike μ_o or k_m, k_d and K_p for instance, which are specific to fat-free biomass growth and GLA synthesis, respectively.

Therefore, when employing this model for robust online process control and optimisation, A_i and B_i for μ_m, k_{co} and Y_{co} should be estimated very accurately and fixed before operation. m, μ_d and k_n should be treated similarly because whilst the model is not as sensitive to these parameters, they impact all four state variables. Whereas A_i and B_i for μ_o, k_m, k_d and K_p are ideal candidates to be updated during online process control and optimisation because they are not "critical" to model accuracy as a whole but significantly impact state variables of interest.

2.4.3 Effects of temperature on bioprocess kinetics

The effect of temperature on *C. echinulate* fermentation was studied by inspecting the temperature correlations captured by fitting Equations (22) and (23). Table 2.3 tabulates A_i, B_i and C_i for each temperature-dependent parameter and displays the coefficient of determination (R^2) between the values predicted as functions of temperature and the optimal value fitted on an experiment-by-experiment basis.

Table 2.4.3 The coefficient of determination (R^2) of the optimal parameters fitting as functions of temperature and the obtained constant (A, B and C).

Parameter	R^2	A	B	C
μ_m	0.99	590.58	-2457.29	N/A
K_{C0}	0.98	1239.43	-966.81	N/A
K_0	0.96	18.41	-942.62	N/A
k_m	0.99	5.58E-16	8813.01	N/A
k_d	0.93	62.24	-2808.45	N/A
K_p	0.98	7670.33	-1640.48	N/A
Y_{C0}	0.99	5.60E+10	-7655.47	1.67

From Table 2.3, it can be seen that the majority of the kinetic parameters follow the standard Arrhenius relationship, with μ_m, K_{co}, k_0, k_d and K_p increasing exponentially from 14 ℃ to 37 ℃. This is expected given that these specific model parameters are closely related to the biosynthesis of biomass and lipid turnover, whereby their metabolic and enzymatic activities increase with temperature for as long as the temperature is sub-optimal. Similar findings were

reported in the literature by (Peng et al., 2010), who observed an increase in biomass productivity from 18 ℃ to 25 ℃, corroborating the observation herein. In contrast, parameters k_m and Y_{C0} decreased with temperature (See Table 2.1). However, whilst k_m and Y_{C0} exhibit the same trend overall, k_m decreases exponentially from the offset, unlike Y_{C0} which remains relatively static from 14 ℃ to 28 ℃ before plummeting over 28 ℃ to 37 ℃. Since Y_{C0} is associated with the efficiency with which glucose is utilised to synthesis biomass constitutes, the sudden drop in Y_{C0} indicates that high temperatures deactivated the overflow metabolism that previously produced by-product. However, higher temperatures are merely concomitant to an increased biomass biosynthesis rate and rapid glucose consumption.

2.4.4 Design of a temperature-shift strategy

A two-stage temperature-shift strategy promises to maximise GLA yield, given the different optimum temperatures for *C. echinulata* biomass growth and GLA accumulation. From Figure 2.2, it is clear that biomass productivity is maximised at a temperature of 37 ℃, whilst peak GLA accumulation is highest at 14 ℃. Therefore, the bioreactors were operated at 37 ℃ to maximise biomass density in the first stage before dropping to 14 ℃ to maximise GLA accumulation in the second stage. The remaining degree of freedom: when to shift the temperature from 37 ℃ to 14 ℃, was addressed by rigorous optimisation of the peak GLA concentration as predicted by the kinetic model. To validate the kinetic model's fidelity for bioprocess scale-up and process optimisation, two further experiments were conducted to verify the temperature-shift strategy. The two experiments were all carried out in a 5 L bioreactor, ultimately switching the temperature at 168 hours and 96 hours, respectively.

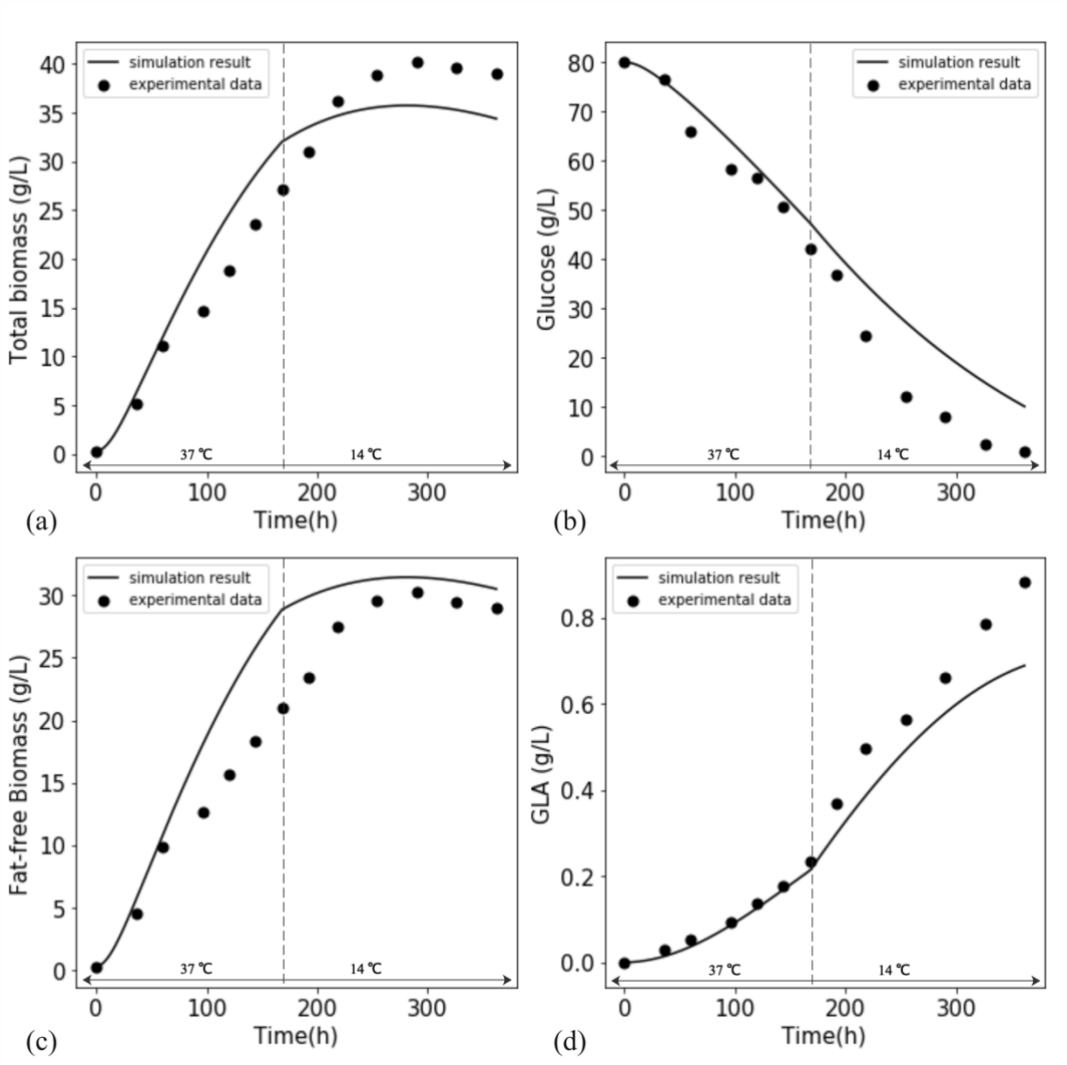


Figure 2.5 The experimental results and model prediction results for (a) total biomass, (b) glucose, (c) fat-free biomass and (d) GLA of temperature-shift strategy at 168hr in 5L fermentor.

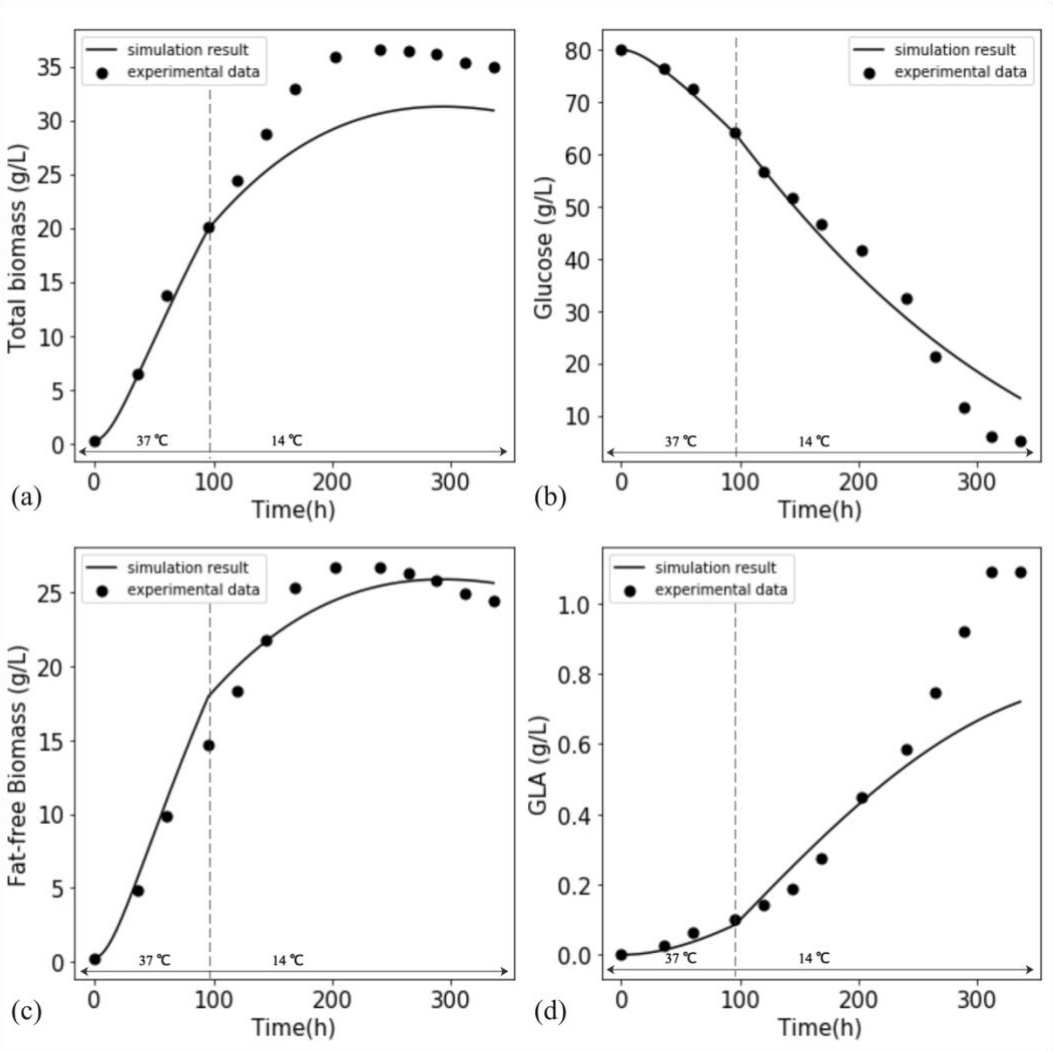


Figure 2.6 The experimental results and model prediction results for (a) total biomass, (b) glucose, (c) fat-free biomass and (d) GLA of temperature-shift strategy at 96hr in 5L fermentor.

Figure 2.5 and Figure 2.6 compare the model predictions and corresponding experimental results when employing the chosen temperature-shift policy in the 5L bioreactor. The state variables were predicted accurately within 30% error in both cases, as quantified in Table 2.4.

Table 2.4.4 Model prediction accuracy for the temperature-shift processes

Output variable	Temperature-shift at 96hr	Temperature-shift at 168hr
	Prediction error (%)	Prediction error (%)
X_T	29.86	13.00
C	36.07	28.77
X_B	16.86	14.89
X_G	27.30	22.73

Of particular success was the upscaled temperature-shift experiment, which despite being a first for *C. echinulata*, achieved a GLA concentration of 1323 mg L^{-1}, a 69.6% increase over 780 mg L^{-1} attained by the fixed temperature culture at 14 ℃ in Figure 2.2. Similarly, employing a two-stage temperature shift policy (Peng et al., 2010) increased ARA production by 26.1%; however, the greater increase here highlights the strategy's effectiveness for promoting *C. echinulate* GLA biosynthesis in particular.

The model's sensitivity to parameter uncertainty was re-evaluated with the new temperature-shift policy (See Appendix C). However, the conclusions drawn from here mirror those Section 3.2, so will not be elaborated upon, namely: that propagated uncertainty increases with time, high sensitivity to parameters B_{μ_m}, $B_{k_{co}}$ and $B_{Y_{co}}$; but low sensitivity to parameters μ_d, m and k_n. Therefore, leading to the same control policy recommendations when subscribing to the two-stage temperature shift strategy.

2.5 Conclusion

In this work, a temperature-dependent biokinetic model capable of simulating the fermentative biomass growth and GLA biosynthesis of *C. echinulata* was proposed for the first time. Using experimental data from a 1L bioreactor, the biokinetic parameters were identified and the prediction accuracy verified over a wide temperature range from 14 ℃ to 37 ℃. Higher cultivation temperatures around 37 ℃ were found to benefit cell biomass growth, whilst GLA accumulation favoured lower temperatures around 14 ℃. Thus, a two-stage temperature-shift strategy was designed and tested by optimising biomass growth and GLA biosynthesis of *C. echinulata* for the first time. Compared to fixed temperature cultivation at 14 ℃, the optimised two-stage temperature shift strategy increased GLA production by 69.6% when verified experimentally. The proposed biokinetic model's high predictive accuracy when up-scaling the bioreactor from 1L to 5L demonstrates the model's reliability for continued scale-up of the biotechnology. However, further studies on the impact of scale-dependent transport phenomena such as mixing induce shear rate, aeration and eddy size is recommended to improve the upscaling predictions. In addition, dissolved oxygen was found to run out in the later stage of the temperature-shift experiments, meaning that oxygen concentration may become a limiting factor for biomass growth. As a result, this parameter should be included in the model for future process control and optimisation.

Chapter 3 Investigating ‘greyness’ of hybrid model for bioprocess predictive modelling

This chapter presented three hybrid models for predicting the temperature-dependent rates of biomass growth, glucose consumption and γ-linolenic acid accumulation during fermentation of the fungus *Cunninghamella echinulata*. Each hybrid model incorporated different amounts of kinetic information from a pre-existing complex kinetic model, representing three levels of hybrid model ‘greyness’, then embedded a Gaussian Process (GP) to simulate the unknown kinetics inferred from experimental measurements. This observation also held when using the hybrid models to predict the temperature-shift dynamics for the upscaled 5 L bioreactor. The hybrid models demonstrated much improved predictive confidence with similar predictive accuracy to the original kinetic model, demonstrating the proficiency of hybrid modelling for accelerating the construction of confident bioprocess models for robust process optimisation and real-time monitoring.

3.1 Introduction

With the exponential increase in computing power in recent decades, modelling now plays a critical role in process engineering and is widely applied in simulation, optimisation and control (Marchetti et al., 2016; Voll & Marquardt, 2012; Zhang et al., 2019). In essence, process modelling aims to translate knowledge about a process into an abstract mathematical representation (von Stosch et al., 2014). For bioprocesses, even a single cell contains thousands of enzymes catalysing thousands of interacting metabolic reactions, leading to complex growth and product formation dynamics. Modelling is a valuable tool for qualitatively and

quantitatively simplifying the complex metabolic reactions within a fermenter (Rohner & Meyer, 1995), paving the way toward significant improvements in safety, optimal plant design, monitoring and analysis (Maria, 2004).

High fidelity kinetic models are effective tools for understanding biological processes and the influence of key operating variables due to their interpretable model structure and physical parameters (Maria, 2004). However, identifying the correct model structure and parameters to describe and predict a given dynamic bioprocess accurately is challenging and time-consuming. This is acutely challenging for biochemical systems that are not fully understood. Hence, to simplify, the cell culture is often assumed homogeneous, and the evolution of the process is described in terms of macroscopic variables (e.g. biomass, substrate and product concentration). Unstructured kinetic models achieve this by lumping the effect of countless metabolic reactions into a handful of kinetic constants. However, the kinetic constants in the model are macroscopic representations of the numerous metabolic reactions within the cell, leading to issues during parameter estimations and process optimisation (Nicoletti et al., 2009). Moreover, since, in reality, the lumped kinetic constants represent complex functions of extracellular nutrient and metabolite concentrations in the culture and environmental conditions (González-Figueredo et al., 2019), pure kinetic models cannot satisfy accurate, low-uncertainty long-term prediction performance of dynamic bioprocesses.

Alternatively, the step-change in the macroscopic state variables can be captured directly from process data as a function of the current state and operating conditions using data-driven models such as Artificial Neural Networks (ANNs) and Gaussian Processes (GPs) (Safarian et

al., 2021; Sheng et al., 2020). However, while data-driven methods streamline model construction by automatically capturing the complex process dynamics without a priori physical knowledge, they cannot be used to extrapolate outside the operational range they were trained on and are prone to overfitting (Dochain, 2008; Nicoletti et al., 2009). Moreover, data-driven methods require large amounts of information-rich data to capture the highly nonlinear bioprocess dynamics (Sansana et al., 2021) – a luxury that is seldom available without extensive, well-designed experiments. Nonetheless, fermentation processes can have long experimental cycles, in many cases requiring several weeks or months to complete an experiment and offline sample analysis. Hence, collecting sufficient bioprocess data for data-driven model construction is time and resource-consuming.

Hybrid models offer a solution and middle ground by embedding the data-driven model into the kinetic model structure to simulate the dynamic nature of the lumped parameters. This way, the data-driven model is responsible for learning the complex functions underling the lumped parameters from process data, while the structure of the kinetic model reduces the nonlinearity of the data-driven modelling problem compared with the original bioprocess dynamics. As a result, hybrid models have experienced success in several recent fermentation process modelling (Saraceno et al., 2010; Vega-Ramon et al., 2021; Wang et al., 2010) and biochemical process modelling studies (Cabaneros Lopez et al., 2021; Willis & von Stosch, 2017), where embedding a data-driven model has proven capable of enhancing the prediction accuracy of an unstructured kinetic model (Carinhas et al., 2011; von Stosch et al., 2014).

Hybrid models exist on a spectrum of 'greyness' between pure data-driven 'black-box' models and pure mechanistic-based 'white-box' models. Towards the 'black-box' end of the spectrum, the data-driven component is more responsible for capturing the process dynamics, while towards the 'white-box' end, the mechanistic backbone is more responsible for capturing the process dynamics. Hence, a bi-dimensional bias-variance trade-off exists for any given hybrid modelling problem (Sansana et al., 2021). Specifically, when building a hybrid bioprocess model, there exists a question of how much kinetic information to incorporate before offloading the remaining model complexity onto the data-driven component. This decision is often made based on experience given the fact that different balances of bioprocess data and mechanistic understanding can be available for each system.

Despite this, an approach to systematically determine the most suitable 'greyness' in the bidimensional hybrid modelling plane remains unaddressed. As a result, the aim in this chapter is to investigate this decision for dynamic bioprocess simulation by exploring the question for a real case study: γ-linolenic acid (C18:3n-6, GLA) fermentation by the oily fungus *Cunningham echinulata* (*C. echinulata*) for which a complex unstructured kinetic model was constructed in our previous work (Song, 2021). In spite of the high accuracy of the kinetic model, it was time consuming to identify the suitable model structure. Moreover, the kinetic model also suffers from large uncertainty, restricting its applicability and reliability for bioprocess control and optimisation. Specifically, this chapter aims to investigate the effect of 'greyness' by incorporating different amounts of kinetic information from the complex kinetic model structure to build three hybrid models of three degrees of mechanistic complexity. Each

hybrid model was used to capture the temperature-dependent kinetics before being used to predict the temperature-shift kinetics and the predictive performance alongside model uncertainty compared with the original complex unstructured kinetic model.

3.2 Literature review

As mentioned previously, hybrid models are the effective tools which have successfully applied in fermentation process modelling. However, hybrid models exist on a spectrum of 'greyness' between 'black-box' models and 'white-box' models. Hence, how to balance a bi-dimensional bias-variance for any given hybrid modelling in still a key problem. This review will focus on three major themes throughout the literature review. These themes are: (i) investigate the 'greyness' of γ-linolenic acid (C18:3n-6, GLA) dynamic bioprocess simulation by the *Cunningham echinulata* (*C. echinulata*), (ii) the predictive performance for temperature-shift fermentation with model uncertainty compared with the pure complex unstructured kinetic model.

3.3 Methodology

3.3.1 Introduction to case study

γ-linolenic acid (GLA) fermentation via *C. echinulata* X-15, a screened high-yield strain, was carried out in 1 L bioreactors with pH fixed during the cultivation. Three batch experiments were conducted under different temperatures: 14, 28, and 37 ℃. Cultivation time of the experiments ranged from 300 to 480 h. A total of 39 experimental data points were collected across the three experiments (13 experimental measurements per experiment) at different temperatures. Then, to enhance GLA production, two additional temperature-shift experiments

were designed in which GLA fermentation was carried out in a 5 L bioreactor. Temperature was switched from 37 to 14 ℃ at the 168th hour and the 96th hour, respectively, in the two experiments. Figure 3.1 depicts *C. echinulata* total and fat-free biomass growth, glucose consumption, and GLA accumulation over a range of operating temperatures for the constant-temperature experiments.

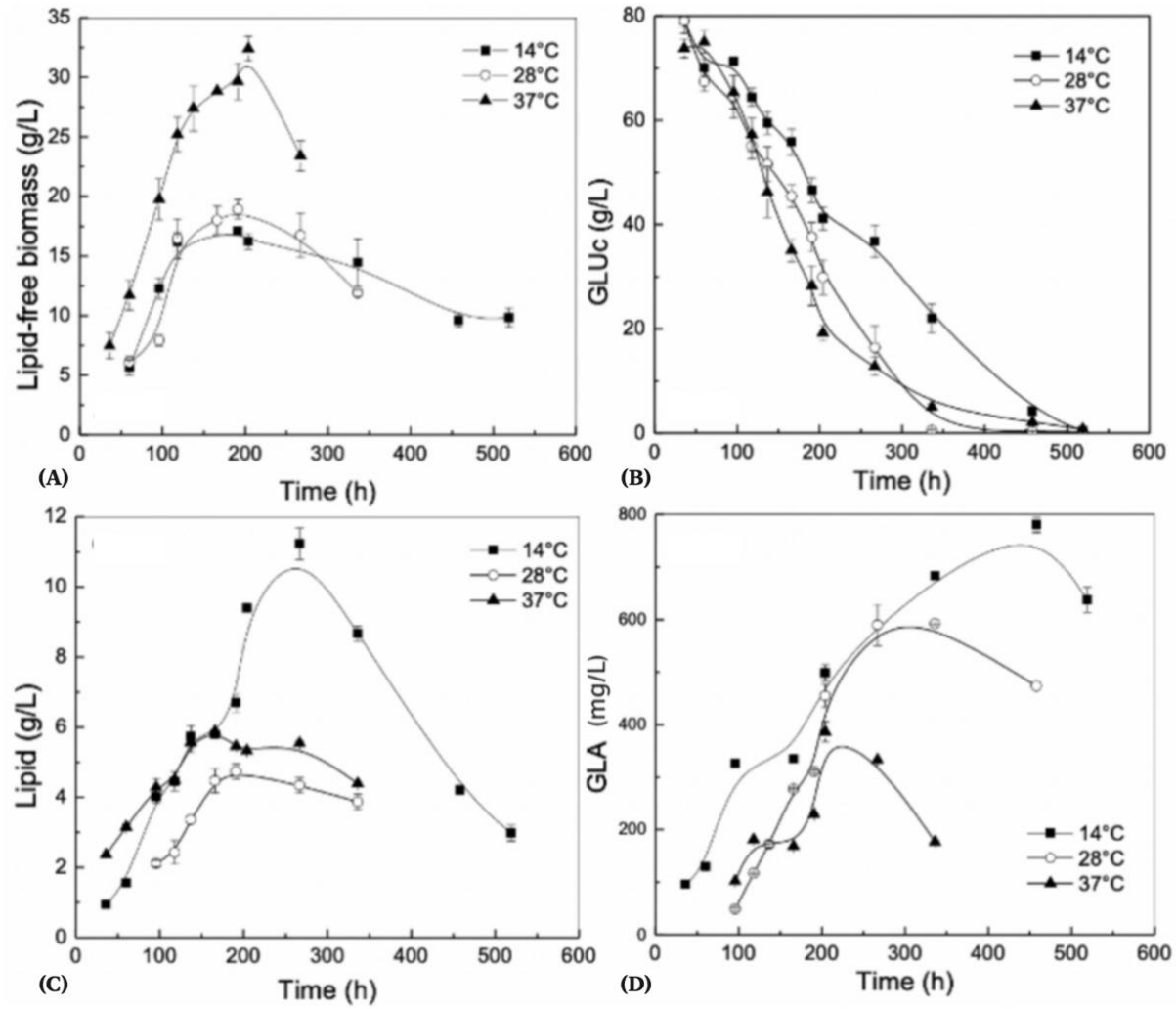


Figure 3.1 Experiment result on the effect of temperature on lipid-free biomass production (A), residual sugar concentration (B), lipid accumulation (C) and GLA synthesis (D) during batch growth of C. echinulata. The symbols used were 14 ℃ (square), 28 ℃ (empty circle), 37 ℃ (triangle). Error bars indicated on the mean of triplicate experimental runs.

3.3.2 Introduction to the kinetic model

A complex kinetic model was proposed in our previous study to simulate the effect of the temperature on total and fat-free biomass growth, glucose consumption and GLA production and is reproduced in Equations (3.1-3.4) (Song, 2021).

$$\frac{dX_T}{dt} = \mu_m \cdot \frac{C}{K_{co} \cdot X_T + C} \cdot X_T - \mu_d \cdot X_T \tag{3.1}$$

$$\frac{\mathrm{d}X_B}{\mathrm{d}t} = k_0 \cdot \frac{\mathrm{d}X_T}{\mathrm{d}t} = k_0 \cdot \left(\mu_m \cdot \frac{C}{C + X_T \cdot K_{co}} \cdot X_T - \mu_d \cdot X_T\right) \tag{3.2}$$

$$\frac{dC}{dt} = -Y_{C0} \cdot \left(\mu_m \cdot \frac{C}{K_{co} \cdot X_T + C} \cdot X_T - \mu_d \cdot X_T\right) - m \cdot X_T \tag{3.3}$$

$$\frac{\mathrm{d}X_G}{\mathrm{d}t} = k_m \cdot \left(\mu_m \cdot \frac{C}{C + K_{co} \cdot X_T} \cdot X_T - \mu_d \cdot X_T\right) + k_n \cdot X_T - k_d \cdot \frac{1}{C + K_p} \cdot X_T \tag{3.4}$$

In Equation 1, X_T is total biomass concentration (g L^{-1}), C is glucose concentration (g L^{-1}), μ_m is maximum specific growth rate (h^{-1}), μ_d is specific cell death rate (h^{-1}) and K_{co} is the half-saturation constant. In Equation 2, k_0 is the ratio of fat-free biomass to total biomass (g g^{-1}). In Equation 3.3, Y_{C0} is yield coefficient from glucose to biomass (g g^{-1}) and m is biomass-specific maintenance coefficient (g $g^{-1}h^{-1}$). Finally, in Equation 3.4, k_m is the growth-dependent synthesis constant (g g^{-1}), k_n is the growth-independent synthesis constant (g $g^{-1}h^{-1}$), k_d is the specific GLA decay rate (g $g^{-1}h^{-1}$) and K_p is the saturation product constant (g L^{-1}) for GLA decay. The parameter values and their uncertainties estimated in our previous work (Song, 2021) are tabulated in Appendix E Table S1.

Our previous work showed that this model could accurately simulate the effect of temperature on total and fat-free biomass growth, glucose consumption and GLA production over a wide temperature range. However, the complex and highly nonlinear model structure was not only

time-consuming to identify, but also suffered from large uncertainty (e.g. standard deviation of GLA production over 100%) which prevents its further application in process control and optimisation. Other researchers have also emphasised this issue in their works (Sadino-Riquelme et al., 2020; Vega-Ramon et al., 2021). Therefore, to improve the kinetic model, its structure was adapted to build three hybrid models, each with different amounts of mechanistic complexity, as described in Section 2.3. Then once built, the predictive performance of three hybrid models was benchmarked against the original kinetic model in Section 3.5.

3.3.3 Hybrid model construction

3.3.3.1 Hybrid model structures

The complex bioprocess dynamics arise from underlying biological mechanisms; some are known and are encoded in the kinetic model, others not and remain untapped within process data. Hybrid models leverage both sources of process knowledge by embedding a data-driven model inside a kinetic model, such that the kinetic model describes known mechanisms while the data-driven component fills in the gaps with knowledge extracted from process data. The structure of the hybrid model will depend on how confidently certain mechanisms are known to occur. For example, Equation (5), one of the embedded hybrid structures used to describe total biomass growth in this work, explicitly accounts for substrate saturation (i.e., the first term on the right) and endogenous biomass decay (i.e., the second term on the right). However, the complex function of the dynamic states: total biomass concentration X_T, glucose concentration C, fat-free biomass concentration X_B, GLA concentration X_G and temperature T that

underlies the maximum specific growth rate is learnt from process data by a data-driven sub-model $\mu'(\cdot)$.

$$\frac{\mathrm{d}X_T}{\mathrm{d}t} = \mu'(X_T, C, X_B, X_G, T) \cdot \frac{C}{C + X_T \cdot K_{C0}} \cdot X_T - \mu_d \cdot X_T \tag{3.5}$$

Compared with a purely mechanistic kinetic model, the mechanistic backbone of a hybrid model does not need to be as complex. Specifically, where a purely mechanistic kinetic model would strive to include all aspects of the kinetics, the mechanistic backbone of a hybrid model only needs to capture an approximation by drawing upon the standard kinetic theories available in the literature (e.g., Monod, Contois or Haldane models). By only tasking the data-driven component with simulating the unknown elements, the nonlinearity of the data-driven modelling problem is reduced compared with the original bioprocess dynamics. However, there remains the question of how much kinetic information to incorporate from the original complex kinetic model outlined in Section 2.2 while avoiding overfitting the data-driven model. Therefore, in this study, three hybrid models were built to simulate the fermentation process by incorporating different amounts of kinetic information from the original kinetic model, as shown in Equations (3.6) to (3.8), representing three levels of 'greyness'.

Hybrid Model 1, presented in Equations (3.6a) to (3.6d), only incorporated the most fundamental but confident assumption that biomass growth, glucose consumption, and GLA production are proportional to the current cell population. The proportionality parameters for total biomass growth $\mu(\cdot)$, glucose consumption $Y_{C/X_T}(\cdot)$, fat-free biomass growth $Y_{X_B/X_T}(\cdot)$ and GLA accumulation $Y_{X_G/x_i}(\cdot)$ were each represented by unique a data-driven function of the state concentrations and temperature. Hybrid Model 2, presented in Equations (3.7a) to

(3.7d), then incorporated substrate saturation, substrate maintenance and first-order endogenous biomass decay, while the data-driven terms account for the complex functions underlying maximum specific growth rate $\mu'(\cdot)$, the glucose to biomass yield coefficient $Y'_{C/X_T}(\cdot)$, the ratio of total to fat-free biomass $Y'_{X_B/X_T}(\cdot)$ and the growth-dependent GLA yield coefficient $Y'_{X_G/x_i}(\cdot)$. Finally, Hybrid Model 3, presented in Equations (3.8a) to (3.8d), incorporated glucose limiting GLA consumption and the temperature-dependent Arrhenius expression. Here the data-driven terms account for the complex functions underlying the parameters in the exponent of the Arrhenius expressions for maximum specific growth rate $B_{\mu_m}(\cdot)$, the glucose to biomass yield coefficient $B_C(\cdot)$, the ratio of total to fat-free biomass $B_{X_B}(\cdot)$ and the growth-dependent GLA yield coefficient $B_{X_G}(\cdot)$.

Hybrid Model 1:

$$\frac{\mathrm{d}X_T}{\mathrm{d}t} = \mu(X_T, C, X_B, X_G, T) \cdot X_T \tag{3.6a}$$

$$\frac{\mathrm{d}C}{\mathrm{d}t} = -Y_{\frac{C}{X_T}}(X_T, C, X_B, X_G, T) \cdot X_T \tag{3.6b}$$

$$\frac{\mathrm{d}X_B}{\mathrm{d}t} = Y_{X_B/X_T}(X_T, C, X_B, X_G, T) \cdot X_T \tag{3.6c}$$

$$\frac{\mathrm{d}X_G}{\mathrm{d}t} = Y_{X_G/x_i}(X_T, C, X_B, X_G, T) \cdot X_T \tag{3.6d}$$

Hybrid Model 2:

$$\frac{\mathrm{d}X_T}{\mathrm{d}t} = \mu'(X_T, C, X_B, X_G, T) \cdot \frac{C}{C + X_T \cdot K_{C0}} \cdot X_T - \mu_d \cdot X_T \tag{3.7a}$$

$$\frac{\mathrm{d}C}{\mathrm{d}t} = -Y'_{C/X_T}(X_T, C, X_B, X_G, T) \cdot \frac{C}{C + X_T \cdot K_{C0}} \cdot X_T - m \cdot X_T \tag{3.7b}$$

$$\frac{\mathrm{d}X_B}{\mathrm{d}t} = Y'_{X_B/X_T}(X_T, C, X_B, X_G, T) \cdot \frac{C}{C + X_T \cdot K_{C0}} \cdot X_T - k_0 \cdot X_T \tag{3.7c}$$

$$\frac{\mathrm{d}X_G}{\mathrm{d}t} = Y'_{X_G/x_i}(X_T, C, X_B, X_G, T) \cdot \frac{C}{C + K_{C0} \cdot X_T} \cdot X_T + k_g \cdot X_T \tag{3.7d}$$

Hybrid Model 3:

$$\frac{\mathrm{d}X_T}{\mathrm{d}t} = A_{\mu_m} \cdot e^{\frac{B_{\mu_m}(X_T,C,X_B,X_G,T)}{T}} \cdot \frac{C}{C + X_T \cdot K_{co}} \cdot X_T - \mu_d \cdot X_T \quad (3.8a)$$

$$\frac{\mathrm{d}C}{\mathrm{d}t} = -A_C \cdot e^{\frac{B_C(X_T,C,X_B,X_G,T)}{T}} \cdot \frac{C}{C + X_T \cdot K_{C0}} \cdot X_T - m \cdot X_T \quad (3.8b)$$

$$\frac{\mathrm{d}X_B}{\mathrm{d}t} = A_{X_B} \cdot e^{\frac{B_{X_B}(X_T,C,X_B,X_G,T)}{T}} \cdot \frac{C}{C + X_T \cdot K_{C0}} \cdot X_T - k_0 \cdot X_T \quad (3.8c)$$

$$\frac{\mathrm{d}X_G}{\mathrm{d}t} = A_{X_G} \cdot e^{\frac{B_{X_G}(X_T,C,X_B,X_G,T)}{T}} \cdot \frac{C}{C + X_T \cdot K_{C0}} \cdot X_T + (k_n - \frac{k_d}{C + k_p}) \cdot X_T \quad (3.8d)$$

For each hybrid model, only the minimum number of kinetic constants were replaced with data-driven sub-models to avoid over-parameterisation. However, using a set of lumped constants to represent dynamic properties results in large uncertainty and low accuracy, particularly if the kinetic model structure is non-identifiable. Hence only the parameters directly related to biomass growth, substrate consumption and production were replaced with data-driven sub-models. These parameters commonly suffer from the greatest uncertainty; at the same time, these parameters also have the greatest influence on the predicted process trajectory. The aim of replacing the most sensitive parameters was to reduce kinetic model complexity and prediction uncertainty.

3.3.3.2 Parameter estimation and mitigating overfitting

Since the data-driven parameters are functions of dynamic state variables, they are time-varying parameters $\boldsymbol{\varphi}$, while the remaining parameters $\boldsymbol{\theta} \in \left[K_{co},\ \mu_d, m, k_0, k_g, k_n, k_d, k_p \right]^T$ are constant kinetic parameters. To simulate the bioprocess, the data-driven sub-models updated the time-varying parameters as a function of the current state concentrations and temperature. However, before correlating the time-varying parameters with the state

concentrations and temperature, the time-varying and constant parameters were estimated simultaneously from experimental data.

Parameter estimation was formulated as the nonlinear least-squares regression problem shown in Equations (3.9a) to (3.9h), where the set of time-varying parameters $\boldsymbol{\varphi}$ and constant parameters $\boldsymbol{\theta}$ were found by minimising the difference between the experimentally measured $\hat{\boldsymbol{y}}_n$ and simulated $\boldsymbol{y}_n$ values of the state variables $\boldsymbol{y} \in [X_T, C, X_B, X_G]^T$ subject to the system of ordinary differential equations (ODEs) $f(\boldsymbol{y}(t), \boldsymbol{\theta}, \boldsymbol{\varphi})$ describing the rates of change of the state variables. Where in Equation (3.9b), $\boldsymbol{\Lambda}$ is a weighting matrix, and n and i index the batch number and time interval, respectively. Finally, $\boldsymbol{y}_{lb}$ and $\boldsymbol{y}_{ub}$ are the lower and upper bounds of the state variables, respectively, while $\boldsymbol{\theta}_{lb}$ and $\boldsymbol{\theta}_{ub}$ and $\boldsymbol{\varphi}_{lb}$ and $\boldsymbol{\varphi}_{ub}$ are the lower and upper bounds of the constant and time-varying parameters, respectively.

$$\min_{\boldsymbol{\theta},\boldsymbol{\varphi}} E(\boldsymbol{\theta}, \boldsymbol{\varphi}) = f(\boldsymbol{\theta}, \boldsymbol{\varphi}) + \boldsymbol{\lambda}^T \cdot P(\boldsymbol{\varphi}) \tag{3.9a}$$

$$f(\boldsymbol{\theta}, \boldsymbol{\varphi}) = \sum_{n=1}^{N} \sum_{i=1}^{I} \left(\hat{\boldsymbol{y}}_n^i - \boldsymbol{y}_n^i(\boldsymbol{\theta}, \boldsymbol{\varphi})\right)^T \boldsymbol{\Lambda} \left(\hat{\boldsymbol{y}}_n^i - \boldsymbol{y}_n^i(\boldsymbol{\theta}, \boldsymbol{\varphi})\right) \tag{3.9b}$$

$$P(\varphi_i) = \sum_{t=0}^{t_f} (\varphi_i^{t+1} - \varphi_i^t)^2 \tag{3.9c}$$

Subject to:

$$\frac{d\boldsymbol{y}}{dt} = f(\boldsymbol{y}(t), \boldsymbol{\theta}, \boldsymbol{\varphi}) \tag{3.9d}$$

$$\boldsymbol{y}_{lb} \leq \boldsymbol{y} \leq \boldsymbol{y}_{ub} \tag{3.9e}$$

$$\boldsymbol{\theta}_{lb} \leq \boldsymbol{\theta} \leq \boldsymbol{\theta}_{ub} \tag{3.9f}$$

$$\boldsymbol{\varphi}_{lb} \leq \boldsymbol{\varphi} \leq \boldsymbol{\varphi}_{ub} \tag{3.9g}$$

$$\boldsymbol{y}(t_0) = \boldsymbol{y}_0 \tag{3.9h}$$

This work follows the dynamic parameter estimation framework widely employed by (del Rio-Chanona et al., 2017). To estimate the model parameters, the system of ODEs was first discretised and transformed into a nonlinear programming (NLP) problem. To guarantee solution accuracy, the system of ODEs was discretised using fourth-order orthogonal collocation over finite elements in time into a system of nonlinear algebraic equations (Biegler, 1984). The interior-point optimisation solver IPOPT (Wächter & Biegler, 2006) was then used to solve the resulting NLP problem, as implemented in the open-source Python optimisation environment: *Pyomo* (Hart et al., 2017).

Given that it would be physically inconsistent for the data-driven sub-models to be a function of time span, the batch time was divided into equally spaced 24-hour intervals, each assigned a unique value for each time-varying parameter. Ideally, the number of intervals would equal the number of data points. However, experimental measurements were taken at nonuniform intervals (i.e., a mixture of 12, 24, 36, 48, 60, 72 or 120 hours between measurements), rendering the time-varying parameters strung between sparsely spaced measurements non-identifiable. Without a uniquely identifiable value associated with each interval, any correlation between the time-varying parameters and the state variables would be prone to overfitting. To rectify this problem, the time-varying parameters were penalised from changing too rapidly from one interval to the next, given that underlying metabolism should not change too drastically in this period. This was achieved by introducing the penalty term $P(\boldsymbol{\varphi})$ defined by Equation (3.9c) into the objective function. In Equation (3.9a), the penalty weight λ_i controls the 'rigidity' of the time-varying parameter φ_i with time – larger values will smoothen out

the influence of measurement noise and prevent overfitting. The value of λ_i for each φ_i was selected to balance hybrid model fitting accuracy and uncertainty on the training datasets.

3.3.3.3 Introduction to Gaussian processes

This work selected a Gaussian Process (GP) as the data-driven sub-model. GPs are a probabilistic machine learning technique capable of returning an estimate of the uncertainty associated with a prediction (Mowbray et al., 2021). Within each hybrid model, four independent multi-input single-output GPs correlated the estimated four time-varying parameter values over each 24-hour interval with the state concentrations and temperature. The posterior distribution of possible functions was inferred by specifying a prior distribution and then conditioning it with training data. The prior distribution was specified using a kernel function based on the squared-exponential function shown in Equation (3.10). For a fair comparison between each hybrid model, several kernels were screened to achieve the best performance for each hybrid model:

$$k(\boldsymbol{x}, \boldsymbol{x}^*) = \sigma^2 \exp\left(-\frac{(\boldsymbol{x} - \boldsymbol{x}^*)^T(\boldsymbol{x} - \boldsymbol{x}^*)}{2l^2}\right) + \delta_{ij}\sigma_{noise}^2 \tag{3.10}$$

where $\boldsymbol{x}$ and $\boldsymbol{x}^*$ are two different input locations, $k(\boldsymbol{x}, \boldsymbol{x}^*)$ is the covariance between them, σ, l and σ_{noise} are hyperparameters that control the properties of the underlying distribution over functions, and $\delta_{ij}$is the Kronecker delta function (Williams, 2006). The hyperparameters were optimised by maximum likelihood using the L-BFGS algorithm. The inputs $\boldsymbol{x} \in \mathbb{R}^{m \times d}$ to GP can be understood as sampling from a multivariate Gaussian distribution conditioned on training dataset $\boldsymbol{x}^* \in \mathbb{R}^{n \times d}$ with corresponding a mean function of zero and covariance kernel function $k(\boldsymbol{x}, \boldsymbol{x}^*)$ as shown in Equation (3.11) (Mowbray et al., 2021).

$$f(\boldsymbol{x}) \sim \mathrm{GP}\big(0, k(\boldsymbol{x}, \boldsymbol{x}^*)\big) \tag{3.11}$$

This work uses the open-source GP regression package for Python *GPy* (GPy, 2014), but for more detailed information on the theory, we guide the reader towa/rds (Rasmussen, 2004).

3.3.3.4 Multistep-ahead simulation and propagated uncertainty estimation

Only the system's initial state $\boldsymbol{x}_0$ was provided to the hybrid models, requiring the future state trajectory to be simulated in a closed-loop multistep-ahead manner. At each time step, as a function of the current state concentration $\boldsymbol{y}_i$ and temperature T_i, the GPs predicted and updated the vector of time-varying parameters $\boldsymbol{\varphi}_i \sim \mathrm{GP}(\boldsymbol{y}_i, T_i)$ which were then held constant from time t_i until the next interval at time t_{i+1} where it was updated again using the new state.

The confidence intervals for the constant parameters ($\boldsymbol{\theta}$) were estimated by taking the trace of an approximated covariance matrix obtained by taking the inverse of the Hessian matrix at the optimal parameter solution (Del Rio-Chanona et al., 2015; Franceschini & Macchietto, 2008). For all three hybrid models, it was verified that the contribution to the overall uncertainty by the constant parameters ($\boldsymbol{\theta}$) at the optimal solution was negligible compared to the time-varying parameters ($\boldsymbol{\varphi}$). Therefore, propagated state uncertainty was estimated by Monte Carlo simulation whereby a random value of $\boldsymbol{\varphi}_i$ was sampled from the 68.2% confidence interval of the normal distribution returned by the GP at each time step: $\mathcal{N}_{\boldsymbol{\varphi}}(\boldsymbol{\mu}_i, \boldsymbol{\sigma}_i) \sim \mathrm{GP}(\boldsymbol{y}_i, T_i)$, while $\boldsymbol{\theta}$ was held constant. This was repeated for each time step to produce one possible process trajectory, and then this was repeated 200 times from the same initial state $\boldsymbol{y}_0$ to generate 200 possible process trajectories. All code was executed in Python version 3.7 using *SciPy* version

1.8.0, and *NumPy* version 1.23.0. Numerical integration was performed using *LSODA* from *SciPy's ODEINT* library.

3.4 Results and discussion

3.4.1 Result of hybrid model construction

Measurements of the state concentrations taken during fermentation at 14 ℃, 28 ℃ and 37 ℃ were used together to build three dynamic hybrid models. The structures of the three hybrid models were those described in Section 2.3.1, where Hybrid Model 1 incorporated the least mechanistic information from the original kinetic model, Hybrid Model 2 incorporated an intermediate amount and Hybrid Model 3 incorporated the most mechanistic information. Table 3.1 presents the estimated constant parameters for the three hybrid models.

Table 3.4.11 Estimated values of fixed parameters in different hybrid models

Parameters	Hybrid Model 1	Hybrid Model 2	Hybrid Model 3
K_{co} (g L^{-1})	-	48.64	48.64
μ_d (h^{-1})	-	1.70×10^{-3}	1.70×10^{-3}
m (g $g^{-1}h^{-1}$)	-	3.10×10^{-3}	3.10×10^{-3}
k_0 (g $g^{-1}h^{-1}$)	-	1.35×10^{-3}	1.35×10^{-3}
k_g (g $g^{-1}h^{-1}$)	-	-3.23×10^{-5}	-
k_n (g $g^{-1}h^{-1}$)	-	-	1.20×10^{-4}
k_d (g $L^{-1}h^{-1}$)	-	5.20×10^{-3}	5.20×10^{-3}
k_p (g L^{-1})	-	32.39	32.39
A_{μ_m}	-	-	590.58
A'_C	-	-	1.27×10^{-4}
A_{X_B}	-	-	10872.57
A_{X_G}	-	-	3.29×10^{-13}

Once built using the three fermentation experiments, the hybrid models were used to simulate the state trajectories for the same three experiments to check the fitting result. Figures 3.2 to 3.4 display the fitting results and 68.2% confidence intervals when simulating the 14 ℃ fermentation experiment using Hybrid Models 1 to 3, respectively.

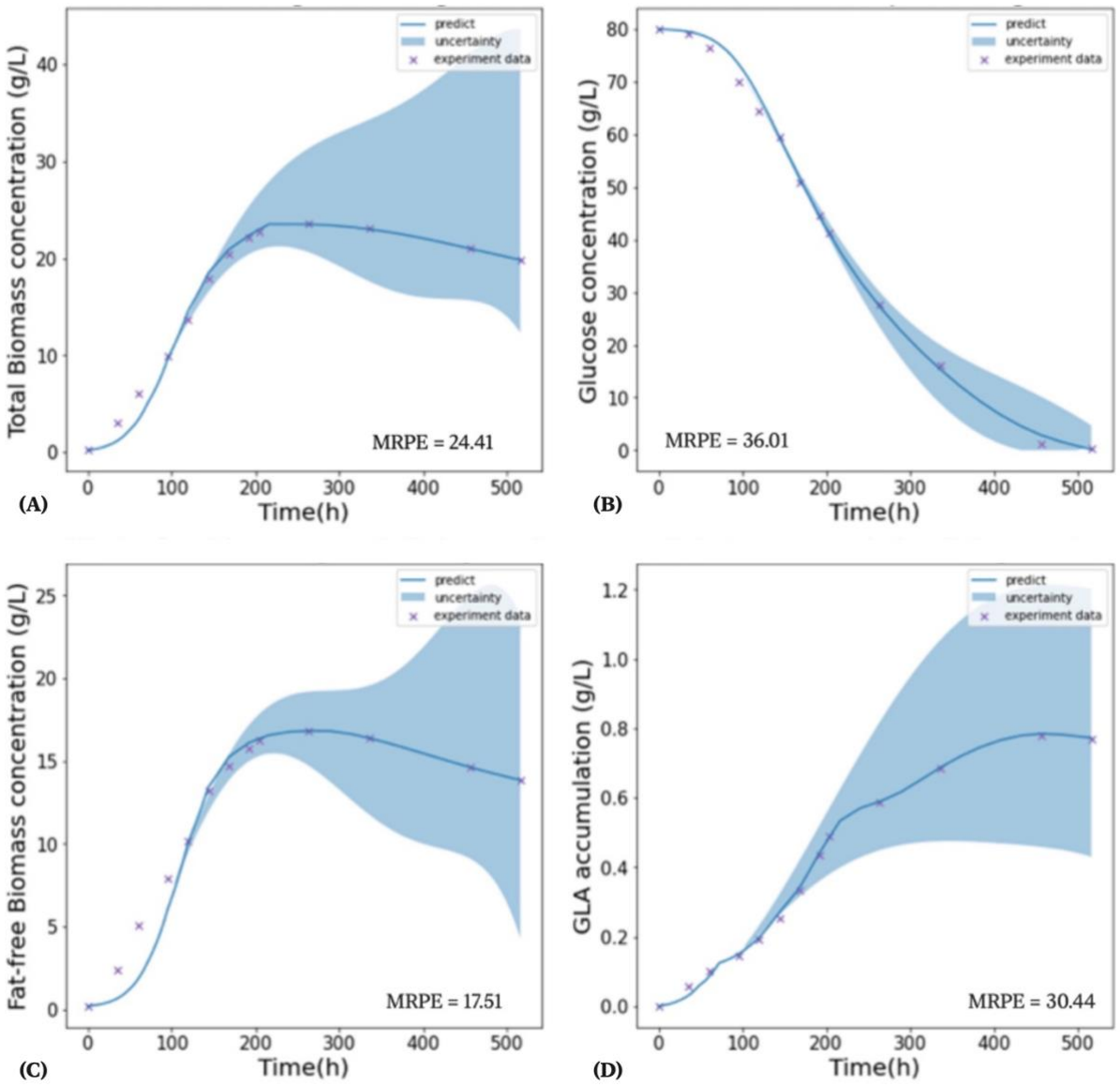


Figure 3.2 Simulation result of Hybrid Model 1 for total biomass concentration (A), glucose consumption (B), fat-free biomass concentration (C) and GLA accumulation (D) at 14℃.

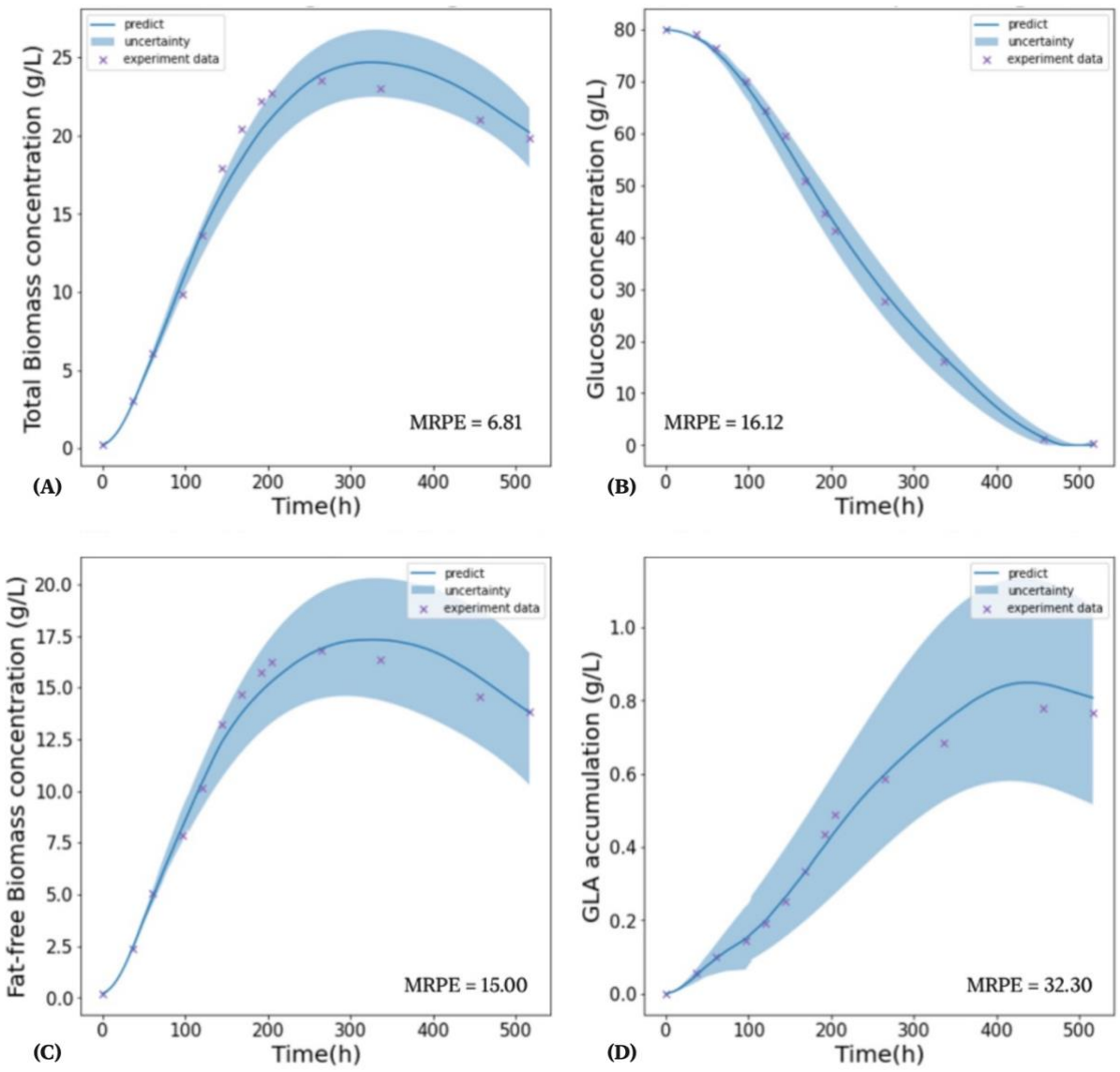


Figure 3.3 Simulation result of Hybrid Model 2 for total biomass concentration (A), glucose consumption (B), fat-free biomass concentration (C) and GLA accumulation (D) at 14℃.

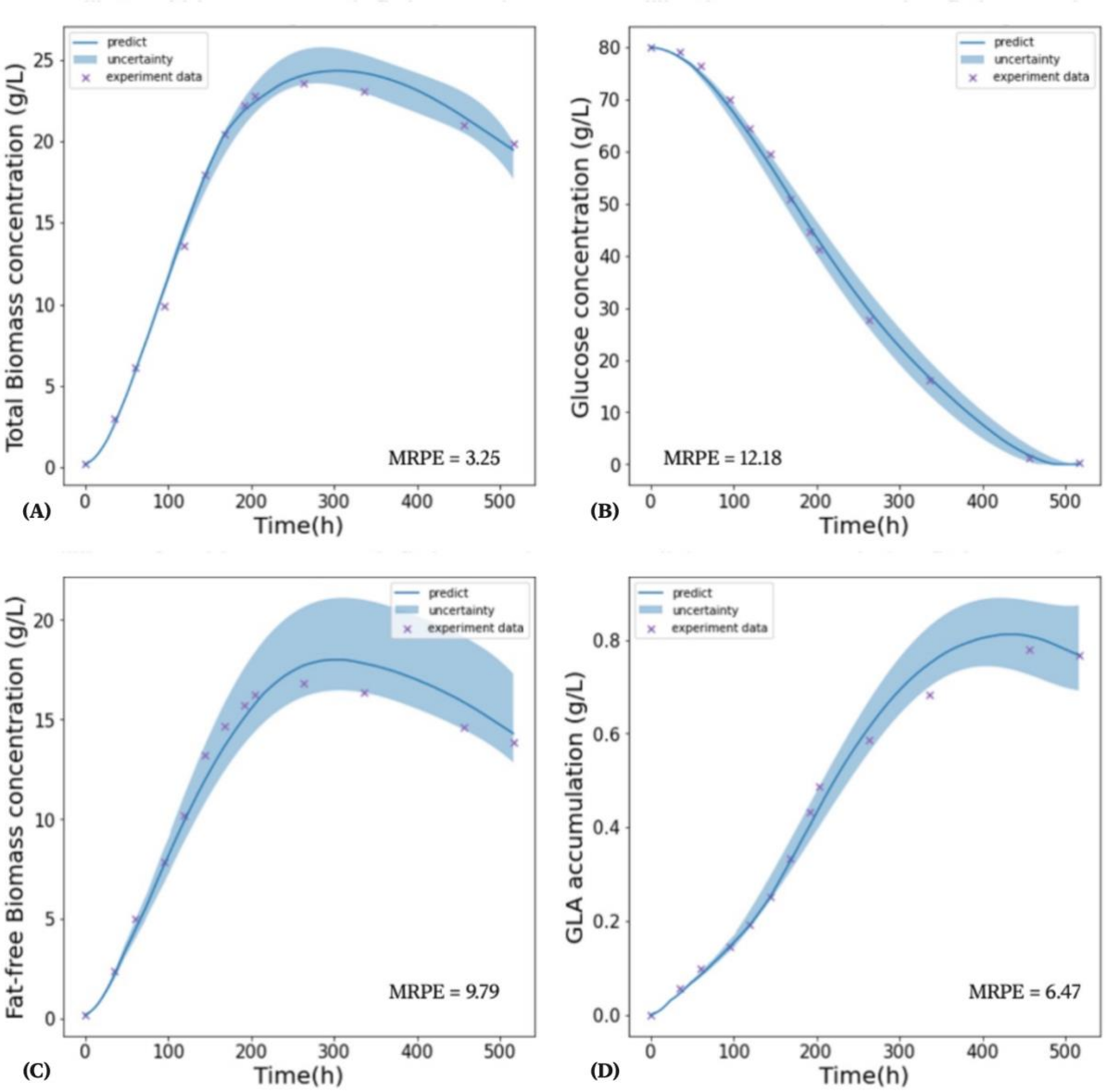


Figure 3.4 Simulation result of Hybrid Model 3 for total biomass concentration (A), glucose consumption (B), fat-free biomass concentration (C) and GLA accumulation (D) at 14℃.

The remaining fitting results for the 28 ℃ and 37 ℃ fermentation experiments are displayed in Appendix E. Together, these results show that all three hybrid models could fit relatively well, with an average mean relative percentage error (MRPE) between the measured and predicted states of 6.22 %, 18.3 %, 7.54 % and 6.65 % for total biomass, glucose, fat-free biomass and GLA concentration, respectively.

In terms of the value of the penalty term coefficient (i.e., $\boldsymbol{\lambda}$) for each fermentation experiment for each hybrid model was selected to balance fitting accuracy and uncertainty on the training dataset. For example, when fitting total biomass concentration for the 14 ℃ fermentation experiment, if λ_{X_T} is chosen as 0.1, the fitting error is 1.69% and while the prediction's uncertainty is 3.67%. However, if λ_{X_T} is chosen as 10, the fitting error increases to 3.94% while the prediction's uncertainty decreases by 3.26%. In addition, comparing the magnitude of $\boldsymbol{\lambda}^T \cdot P(\boldsymbol{\varphi})$ reveals that Hybrid Model 1 required a penalty five orders of magnitude larger than the other two hybrid models, whereas Hybrid Models 2 and 3 had a $\boldsymbol{\lambda}^T \cdot P(\boldsymbol{\varphi})$ of similar magnitude. From this simple observation, it can be concluded that Hybrid Model 1 was more prone to overfitting, necessitating a stronger penalty to prevent overfitting and reduce uncertainty. This is expected as Hybrid Model 1 has the least mechanistic information and its accuracy heavily relies on the construction of data-driven models.

3.4.2 Influence of hybrid model greyness on fitting performance

Section 3.1 demonstrated that the three hybrid models fit well under different cultivation temperatures. However, upon comparing the fitting results of the three hybrid models in Figures 3.2 to 3.4, a few key differences can be seen in their fitting accuracy and uncertainty. To begin with, Hybrid Model 1 suffered from greater uncertainty than Hybrid Models 2 or 3 (i.e., MRPU of 27.09% as opposed to 17.55% or 7.92%, respectively). This result stems from the fact that it is easier to overfit a limited number of observations when only incorporating the most fundamental assumptions about the underlying dynamics. Specifically, Hybrid Model 1 assumed that biomass growth, glucose consumption and GLA production were directly

proportional to the current cell population. As a result, the time-varying proportionality parameters $\mu(\cdot)$, $Y_{C/X_T}(\cdot)$, $Y_{X_B/X_T}(\cdot)$ and $Y_{X_G/x_i}(\cdot)$ had to account for more of the kinetics, requiring them to vary more drastically to capture the process dynamics than the time-varying parameters in Hybrid Models 2 or 3. Hence the data-driven modelling problem was more nonlinear for Hybrid Model 1, increasing uncertainty for a given number of observations. This trend continued: propagated uncertainty decreased from Hybrid Models 1 to 3 (i.e., MRPU of 27.09%, 17.55% and 7.92%, respectively) as more kinetic information was incorporated, and it became harder to overfit the highly nonlinear temperature-dependent bioprocess dynamics.

To reduce uncertainty to some extent, it was necessary to more strongly penalise variation in the time-varying parameters from one interval to the next during parameter estimation by increasing $\boldsymbol{\lambda}^T \cdot P(\boldsymbol{\varphi})$ for Hybrid Model 1 compared with Hybrid Models 2 or 3 – as discussed in Section 3.2. However, improved confidence came at the cost of worse fitting accuracy, where Figure 3.2 shows that Hybrid Model 1 could not fit the exponential growth phase well before 100 hours. The poor fit stems from the combination of three factors: (i) the low biomass concentration before 100 hours, (ii) the relatively large observed biomass growth, glucose consumption and productivity rates before 100 hours, and that (iii) Hybrid Model 1 assumed the rates of change of state were directly proportional to the current cell population. As a result, the time-varying proportionality parameters $\mu(\cdot)$, $Y_{C/X_T}(\cdot)$, $Y_{X_B/X_T}(\cdot)$ and $Y_{X_G/x_i}(\cdot)$ would have had to be large initially but then decrease rapidly as biomass concentration increased. Unfortunately, the rigidity imposed by a stronger penalty penalised the time-varying parameters from dropping rapidly enough to simultaneously fit the exponential growth phase

without overshooting the stationary phase. Accordingly, Figures 3.3 and 3.4 show that Hybrid Models 2 and 3 fit better (i.e., MRPE of 6.05% and 6.06%, respectively, as opposed to 11.2% for Hybrid Model 1) but also capture the exponential growth phase – a result of incorporating substrate inhibitory effects into the mechanistic backbone.

For a more comprehensive comparison of fitting accuracy and uncertainty, Table 3.3 presents the MRPE and MRPU, respectively, achieved by the three hybrid models for each of the three different temperature fermentation experiments. These results show that the three hybrid models fitted well overall (i.e., overall average MRPE of 9.67%). These results also reaffirm the observations drawn from Figures 3.2 to 3.4 above: fitting uncertainty decreased from Hybrid Models 1 to 3 (i.e., average MRPU of 29.93%, 21.36% and 10.78%, respectively) and fitting accuracy improved from Hybrid Models 1 to 2 (i.e., average MRPE of 13.1% and 7.35%, respectively) as more kinetic information was incorporated into the hybrid model structure.

However, Table 3.2 reveals an interesting exception: while Hybrid Model 2 fitted more accurately than Hybrid Model 1, Hybrid Model 3 fitted worse than Hybrid Model 2 despite incorporating more kinetic information.

Table 3.4.2 Mean relative percentage error (MRPE, %) and mean relative percentage uncertainty (MRPU, %) of different hybrid models over different temperatures. X_T, C, X_B and X_G are total biomass, glucose, fat-free biomass and GLA, respectively.

	Hybrid model 1		Hybrid model 2		Hybrid model 3	
14 ℃						
	MRPE	MRPU	MRPE	MRPU	MRPE	MRPU
X_T	9.39	24.41	5.05	6.81	3.85	3.25
C	16.78	36.01	10.02	16.12	9.63	12.18
X_B	12.78	17.51	3.78	15.00	5.05	9.79
X_G	6.04	30.44	5.34	32.30	5.69	6.47
28 ℃						
	MRPE	MRPU	MRPE	MRPU	MRPE	MRPU
X_T	13.07	15.38	2.13	5.57	3.05	4.88
C	28.64	39.88	26.18	26.05	25.51	23.06
X_B	16.11	8.55	2.43	10.01	4.12	9.95
X_G	6.18	31.97	5.62	45.66	6.91	8.67
37 ℃						
	MRPE	MRPU	MRPE	MRPU	MRPE	MRPU
X_T	13.11	25.58	2.03	4.88	4.32	4.91
C	15.62	53.29	13.93	27.33	18.07	24.37
X_B	14.23	10.73	3.94	7.23	5.45	13.59
X_G	4.62	65.37	7.79	59.46	11.62	8.23

In other words, fitting accuracy varied monotonically from Hybrid Models 1 to 3 (i.e., average MRPE of 13.1%, 7.35% and 8.61%, respectively) as more kinetic information was incorporated. While Hybrid Model 1 had to sacrifice some fitting accuracy to combat fitting uncertainty, the worse fit for Hybrid Model 3 is likely due to inductive bias caused by inadvertently incorporating incorrect kinetic information. The original kinetic model was built by hypothesising the underlying dynamics by drawing on empirical knowledge that explained the available experimental observations. However, unstructured kinetic models could be also slightly over-parameterised when aiming to improve its fitting accuracy, so more complex structures risk incorporating incorrect assumptions. Therefore, incorporating incorrect kinetic information into a hybrid will likely induce an inductive bias that negatively impacts bias-variance capital, either pushing nonlinearity back onto the data-driven component or causing the model to misrepresent the bioprocess dynamics.

3.4.3 Hybrid model temperature-shift prediction performance comparison

In order to maximise GLA yield, a two-stage temperature-shift strategy was employed based on the different optimum temperatures for *C. echinulata* biomass growth and specific GLA production. Our previous work identified that biomass growth is maximised at 37 ℃, while specific GLA production is maximised at 14 ℃. It was hypothesised that operating the bioreactor at 37 ℃ to maximise biomass concentration before dropping to 14 ℃ to maximise specific GLA production would maximise overall GLA yield. The original kinetic model was then used to predict and optimise the batch time at which the bioreactor temperature would be shifted during fermentation from 37 ℃ to 14 ℃ to maximise the final GLA concentration.

Finally, two validation experiments were conducted in an upscaled 5 L bioreactor, switching the temperature from 37 ℃ to 14 ℃ at either 168 hours or 96 hours, respectively. Therefore, in this work, to validate and compare the fidelity of the three hybrid models for bioprocess scale-up and optimisation, the three hybrid models were used to predict the upscaled temperature-shift bioprocess trajectories. Note that the original kinetic and three hybrid models were built using the same three fixed temperature (i.e., 14 ℃, 28 ℃ or 37 ℃) fermentation datasets.

Figures 3.5 and 3.6 show the total biomass, glucose, fat-free biomass and GLA concentration trajectories predicted by the three hybrid models for the upscaled 5 L bioreactor when shifting the temperature at 168 hours or 96 hours, respectively.

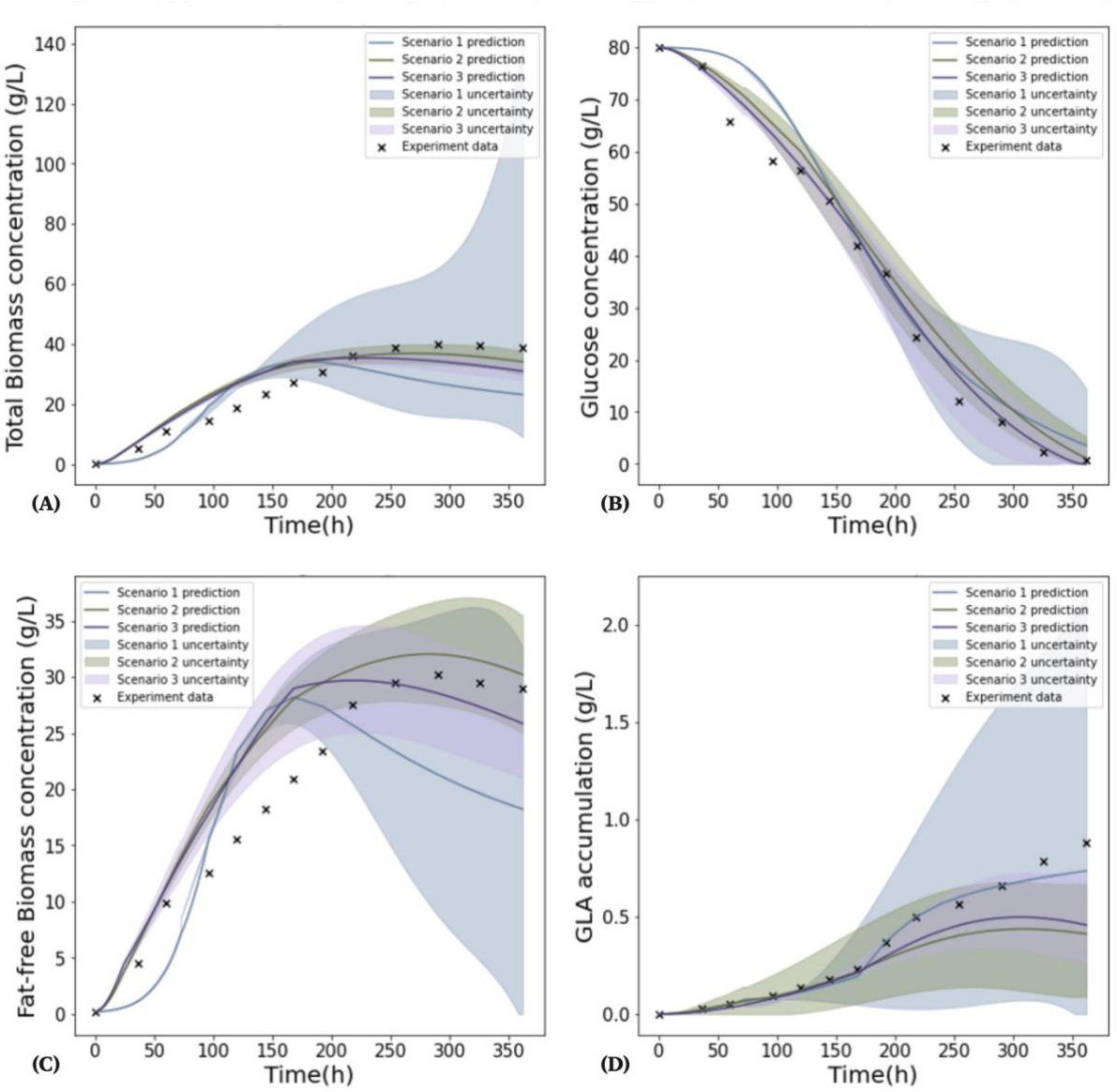


Figure 3.5 The experimental result and model prediction results for total biomass concentration (A), glucose consumption (B), fat-free biomass concentration (C) and GLA accumulation (D) of temperature-shift strategy at 168th hr in the 5 L fermenter.

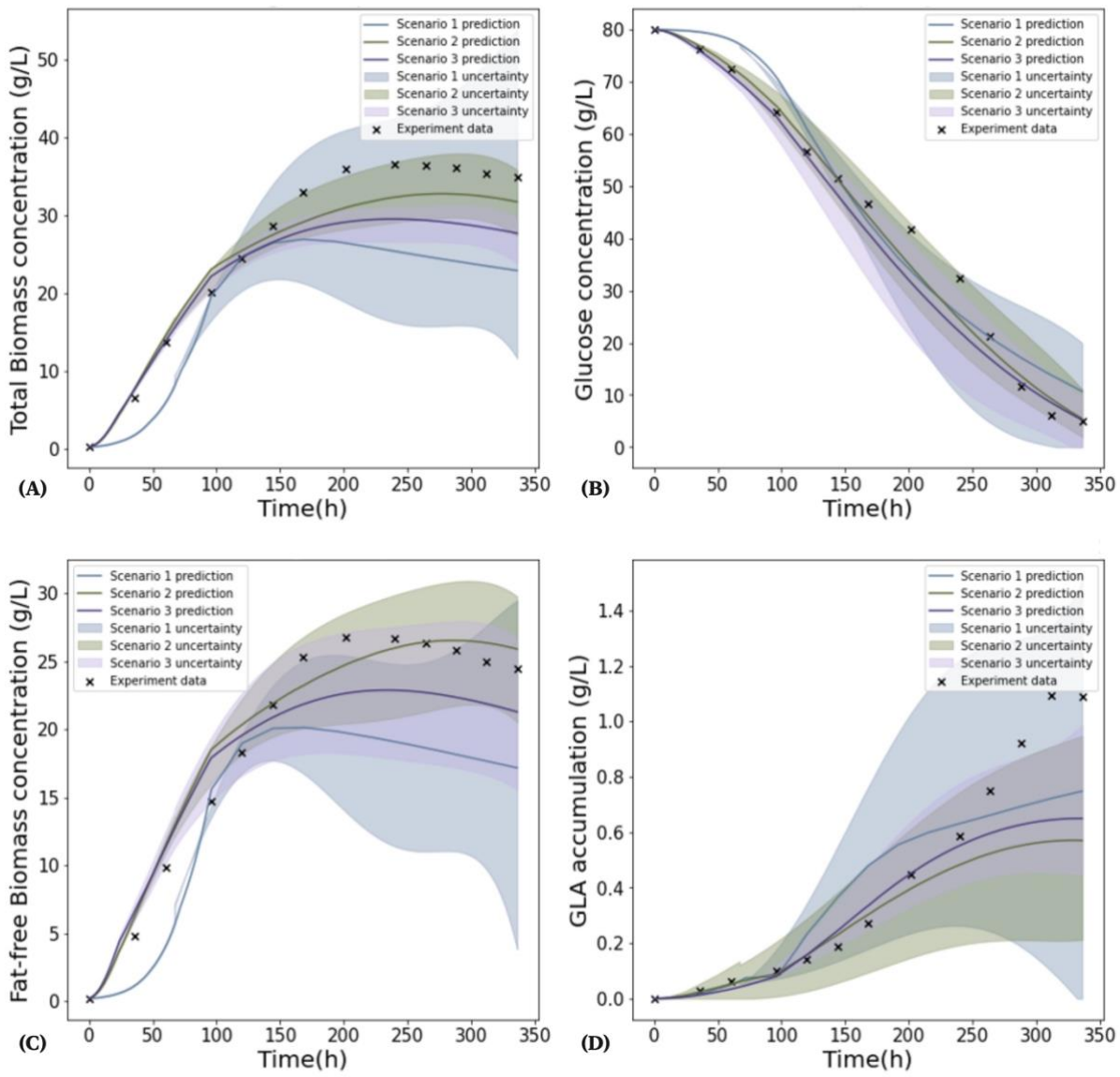


Figure 3.6 The experimental result and model prediction results for total biomass concentration (A), glucose consumption (B), fat-free biomass concentration (C) and GLA accumulation (D) of temperature-shift strategy at 96th hr in the 5 L fermenter.

Tables 3.3 compare the MRPE for the kinetic and three hybrid models when shifting the temperature at 168 hours or 96 hours, respectively, while Tables 3.3 show the MRPU when shifting the temperature at 168 hours or 96 hours, respectively.

Table 3.4.3 Model prediction accuracy (MRPE, %) and uncertainty (MRPU, %) for the temperature-shift processes with different temperature-shift strategies at 168th hr and 96th hr.

Temperature-shift strategy at 168th hr								
	Kinetic model		Hybrid model 1		Hybrid model 2		Hybrid model 3	
State	MRPE	MRPU	MRPE	MRPU	MRPE	MRPU	MRPE	MRPU
X_T	13.00	27.45	31.73	33.36	22.24	5.16	23.85	3.79
C	28.77	86.97	58.19	46.51	28.64	23.49	17.14	29.22
X_B	14.89	28.21	33.40	21.09	21.88	7.57	26.55	11.43
X_G	22.73	73.59	11.38	55.45	24.54	64.08	25.12	18.12
Temperature-shift strategy at 96th hr								
	Kinetic model		Hybrid model 1		Hybrid model 2		Hybrid model 3	
State	MRPE	MRPU	MRPE	MRPU	MRPE	MRPU	MRPE	MRPU
X_T	29.86	30.45	26.86	29.37	9.43	6.87	13.09	5.06
C	36.07	70.05	28.04	29.76	10.90	16.34	11.51	20.59
X_B	16.86	45.62	26.55	20.38	8.89	10.84	13.87	11.87
X_G	27.30	70.06	34.68	51.11	21.05	60.26	24.43	21.72

Taken together, it is evident that Hybrid Model 1 suffered from low prediction accuracy and large uncertainty (i.e., average MRPE of 31.4% and MRPU of 35.87%) compared with Hybrid Models 2 and 3 due to Hybrid Model 1 possessing a simpler mechanistic backbone that rendered it prone to overfitting – in line with observations made in Section 3.3. In contrast, Hybrid Model 2 (i.e., average MRPE of 18.4% and MRPU of 24.33%) and Hybrid Model 3 (i.e., average MRPE of 19.4% and MRPU of 15.23%) were more accurate and confident. Of these two hybrid models, Hybrid Model 2 was the most accurate, particularly when the temperature was switched at 96 hours, where the MRPE fell within 20% but within 10% for total biomass, glucose and fat-free biomass concentration. These results demonstrate the potential of Hybrid Model 2 for model-based bioreactor optimisation, scale-up and control. However, it is interesting that while Hybrid Model 3 was the most confident, it again had slightly worse predictive accuracy than Hybrid Model 2 – in line with observations made in Section 3.3 about hybrid model fitting. Therefore, incorporating more specific kinetic information about a system that is not fully understood is not without the risk of incorporating incorrect biases that might hinder rather than enhance hybrid model performance.

3.4.4 Comparison between kinetic and hybrid model predictive performance

Unstructured kinetic models are lumped macroscopic descriptions of many intracellular metabolic reactions so often suffer from large model-process mismatch. To mitigate this, kinetic models are often slightly over-parameterised, trading some accuracy for increased uncertainty. For this reason, while accurate, the kinetic model proposed in our previous work suffered from large uncertainty (192%). In contrast, Hybrid Models 2 and 3 achieve similar

predictive accuracy but much lower uncertainty, as illustrated in Figures 3.3 and 3.4. Specifically, the kinetic and two hybrid models had an average MRPE of 19.5% and 21.3%, respectively, but an average MPRU of 45.90% and 19.77%, respectively.

Examining the 96-hour temperature shift experiment, Hybrid Models 2 and 3 were, in fact, more accurate than the original kinetic model (i.e., average MRPE of 14.1% rather than 27.5%, respectively), particularly for predicting total-biomass, glucose and fat-free biomass concentration (i.e., average MRPE of 11.3% rather than 27.6%, respectively). However, while Hybrid Models 2 and 3 were more accurate at predicting the early accumulation of GLA compared to the kinetic model (i.e., average MRPE of 22.7% rather than 27.3%, respectively), Figures 3.5 and 3.6 show that the two hybrid models were slightly worse at predicting the accumulation of GLA towards the end of the batch. This is probably due to the fact that by incorporating less kinetic information than the original kinetic model, the hybrid models may not contain the necessary kinetic knowledge about GLA accumulation for accurate prediction. Without any doubt, this problem could be remedied by using a data-driven sub-model to update k_n as a function of the current state and temperature. However, selecting the best combination of time-varying parameters is challenging; Section 1.3.2 described how the time-varying parameters represented growth, consumption and production while all other parameters were non-growth associated constants to which the model had low sensitivity. Therefore, replacing k_n with a data-driven sub-model would also risk increasing hybrid model uncertainty when Hybrid Models 2 and 3 already have similar accuracy and smaller uncertainty than the original kinetic model.

3.5 Conclusion

This work proposed three hybrid models for simulating γ-linolenic acid (GLA) biosynthesis during fermentation of *C. echinulata*, each built with different levels of 'greyness' by incorporating different amounts of kinetic information from a pre-existing complex kinetic model. By embedding a data-driven Gaussian Process (GP) to simulate the dynamic nature of the otherwise constant lumped kinetic parameters, it was possible to accurately capture the bioprocess dynamics over a wide range of operating temperatures (i.e., 14 ℃ to 37 ℃). It was found that introducing too little kinetic information rendered the hybrid model prone to overfitting, resulting in high fitting uncertainty and low fitting accuracy. However, while incorporating more kinetic information always reduced hybrid model uncertainty, best fitting accuracy was achieved by incorporating only a moderate amount of kinetic information. Therefore, when incorporating more specific kinetic information about a system that is not fully understood, there is always a risk of incorporating incorrect kinetic information that creates an inductive bias that hinders rather than enhances hybrid model performance. Therefore, due to the unknowability of the ground truth, when building a hybrid model, either by hypothesising the nature of the underlying kinetics or from a pre-existing complex kinetic model, it may be necessary for the complexity of the mechanistic backbone to be built up incrementally on a case-by-case basis until a drop in hybrid model performance is observed.

Once built, the hybrid models were used to predict the upscaled (i.e., 1 L to 5 L bioreactor) bioprocess dynamics when shifting the temperature during fermentation from 37 ℃ to 14 ℃ at either 96 hours or 168 hours. Again, incorporating more kinetic information reduced

predictive uncertainty, while incorporating only a moderate amount yielded the hybrid model with the best predictive accuracy. The hybrid models achieved similar accuracy to the original kinetic model at much lower uncertainty and time cost due to the flexible data-driven terms more accurately representing the complex underlying kinetics. However, the mechanistic backbone only needs to capture an approximation of the process dynamics, while the specifics can be inferred from process data, greatly accelerating model construction. Therefore, hybrid modelling offers a cost-efficient technique for combining different sources of process knowledge for long-term robust process optimisation and real-time monitoring.

Chapter 4 Summary and recommendations for future work

4.1 Summary

Both the kinetic learning model and the hybrid model proposed in this MPhil project show accurate fitting and predictive performance in *C. echinulat* fermentation experiment. Furthermore, both models were applied in scale-up experiments from 1L to 5L bioreactor with temperature-shift strategy and showed the high prediction accuracy.

In Chapter 2, a first-time proposed temperature-dependent biokinetic model is able to simulate the fermentative biomass growth and GLA biosynthesis of *C. echinulata*. The kinetic parameters were identified, and the prediction accuracy was verified over a wide temperature range from 14 ℃ to 37 ℃ by using experimental data from a 1L bioreactor. Moreover, the kinetic model was applied to scale up the bioreactor from 1L to 5L and showed the high predictive accuracy. This demonstrates the model's reliability for continued scale-up of the biotechnology. Based on the obtained results, the further biokinetic model design with the impact of scale-dependent transport phenomena such as mixing induce shear rate, aeration and eddy size can be developed. In addition, the oxygen concentration may be a limiting factor for biomass growth because of the depletion of dissolved oxygen in the later stage of the temperature-shift experiments.

Three hybrid models for simulating fermentation of *C. echinulata* were presented in Chapter 3. Each model built with different levels of 'greyness' by incorporating different amounts of kinetic information. It was found that introducing too little kinetic information made the hybrid model prone to overfitting, resulting in high uncertainty and low fitting accuracy. Oppositely,

incorporating more kinetic information reduced hybrid model uncertainty. According to the simulation result, only a hybrid model with moderate amount of kinetic information can achieve the best fitting accuracy. Moreover, the hybrid models were used to predict the upscaling experiment from 1L to 5L bioreactor with temperature-shift strategy. The hybrid models achieved similar accuracy to the original kinetic model at much lower uncertainty and time cost due to the flexible data-driven terms more accurately representing the complex underlying kinetics. This promising result offers a cost-efficient technique for combining different sources of process knowledge for long-term robust process optimisation and real-time monitoring.

4.2 Future work

In this MPhil research, kinetic models have been designed and applied to simulate fermentation experiment of the fungus *Cunninghamella echinulata*. However, the effects of scale-dependent transport phenomena such as mixing induce shear rate, aeration and eddy size was deficiently investigated. These factors need to be taken into account in future studies in order to improve the upscaling predictions. Furthermore, as dissolved oxygen was depleted in the final stage of the temperature-shift experiment, this means that the effect of oxygen concentration is also an important factor in the accuracy of the model predictions, and it should be taken into account in future studies for further process control and optimisation.

In addition, the fermentation experiment of *Cunninghamella echinulate* also simulated by three hybrid model with different model structures. The hybrid models can achieve the similar accuracy to the original kinetic model where only an approximation of the process dynamics

is required to be captured by the mechanistic backbone. However, similar with kinetic models, the effects related to scale-up was lacked to investigate. The further validation is required of predicted performance for long-term robust process optimization, process control and real-time monitoring.

References

Abdullah, N. A. H., Nayan, N. A., Kamaludin, N. H. I., Idris, Z. M., & Tompang, M. F. (2016). Cell growth kinetics of Aspergillus Oryzae in industrial natural rubber effluent serum. *ARPN Journal of Engineering and Applied Sciences*, *11*(4), 2687–2692.

Al-Hawash, A. B., Li, S., Zhang, X., Zhang, X., & Ma, F. (2018). Productivity of γ-Linoleic acid by oleaginous fungus Cunninghamella echinulata using a pulsed high magnetic field. *Food Bioscience*, *21*(September 2016), 1–7. https://doi.org/10.1016/j.fbio.2017.10.007

Almquist, J., Cvijovic, M., Hatzimanikatis, V., Nielsen, J., & Jirstrand, M. (2014). Kinetic models in industrial biotechnology - Improving cell factory performance. *Metabolic Engineering*, *24*, 38–60. https://doi.org/10.1016/j.ymben.2014.03.007

Antimanon, S., Anantayanon, J., Wannawilai, S., Khongto, B., & Laoteng, K. (2020). Physiological Traits of Dihomo-γ-Linolenic Acid Production of the Engineered Aspergillus oryzae by Comparing Mathematical Models. *Frontiers in Microbiology*, *11*(November), 1–12. https://doi.org/10.3389/fmicb.2020.546230

Anye Cho, B., Ross, B. S., du Toit, J. P., Pott, R. W. M. C., del Río Chanona, E. A., & Zhang, D. (2021). Dynamic modelling of Rhodopseudomonas palustris biohydrogen production: Perturbation analysis and photobioreactor upscaling. *International Journal of Hydrogen Energy*, *46*(74), 36696–36708. https://doi.org/10.1016/j.ijhydene.2021.08.162

Biegler, L. T. (1984). Solution of dynamic optimization problems by successive quadratic programming and orthogonal collocation. *Computers and Chemical Engineering*, *8*(3–4), 243–247. https://doi.org/10.1016/0098-1354(84)87012-X

Bligh, E.G. and Dyer, W. J. (1959). Canadian Journal of Biochemistry and Physiology. *Canadian Journal of Biochemistry and Physiology*, *37*(8).

Cabaneros Lopez, P., Udugama, I. A., Thomsen, S. T., Roslander, C., Junicke, H., Iglesias, M. M., & Gernaey, K. V. (2021). Transforming data to information: A parallel hybrid model for real-time state estimation in lignocellulosic ethanol fermentation. *Biotechnology and Bioengineering*, *118*(2), 579–591. https://doi.org/10.1002/bit.27586

Carinhas, N., Bernal, V., Teixeira, A. P., Carrondo, M. J. T., Alves, P. M., & Oliveira, R. (2011). Hybrid metabolic flux analysis: Combining stoichiometric and statistical constraints to model the formation of complex recombinant products. *BMC Systems Biology*, *5*. https://doi.org/10.1186/1752-0509-5-34

Čertík, M., Sláviková, L., Masrnová, S., & Šajbidor, J. (2006). Enhancement of nutritional value of cereals with γ-linolenic acid by fungal solid-state fermentations. *Food Technology and Biotechnology*, *44*(1), 75–82.

Chaisutyakorn, P., Praiboon, J., & Kaewsuralikhit, C. (2018). The effect of temperature on growth and lipid and fatty acid composition on marine microalgae used for biodiesel production. *Journal of Applied Phycology*, *30*(1), 37–45. https://doi.org/10.1007/s10811-017-1186-3

Chang, L., Chen, H., Tang, X., Zhao, J., Zhang, H., Chen, Y. Q., & Chen, W. (2021). Advances in improving the biotechnological application of oleaginous fungus Mortierella alpina. *Applied Microbiology and Biotechnology*, *105*(16–17), 6275–6289. https://doi.org/10.1007/s00253-021-11480-y

Chatzifragkou, A., Fakas, S., Galiotou-Panayotou, M., Komaitis, M., Aggelis, G., & Papanikolaou, S. (2010). Commercial sugars as substrates for lipid accumulation in Cunninghamella echinulata and Mortierella isabellina fungi. *European Journal of Lipid Science and Technology*, *112*(9), 1048–1057. https://doi.org/10.1002/ejlt.201000027

Chen, H. C., & Chang, C. C. (1996). Production of γ-linolenic acid by the fungus Cunninghamella echinulata CCRC 31840. *Biotechnology Progress*, *12*(3), 338–341. https://doi.org/10.1021/bp960009y

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., Dechatiwongse, P., Zhang, D., Maitland, G. C., Hellgardt, K., Arellano-Garcia, H., & Vassiliadis, V. S. (2015). Optimal Operation Strategy for Biohydrogen Production. *Industrial and Engineering Chemistry Research*, *54*(24), 6334–6343. https://doi.org/10.1021/acs.iecr.5b00612

Dochain, D. (2008). Bioprocess Control. In *Bioprocess Control*. https://doi.org/10.1002/9780470611128

Dulf, F. V., Vodnar, D. C., Toşa, M. I., & Dulf, E. H. (2020). Simultaneous enrichment of grape pomace with γ-linolenic acid and carotenoids by solid-state fermentation with Zygomycetes fungi and antioxidant potential of the bioprocessed substrates. *Food Chemistry*, *310*(November 2019). https://doi.org/10.1016/j.foodchem.2019.125927

Dyal, S. D., Bouzidi, L., & Narine, S. S. (2005). Maximizing the production of γ-linolenic acid in Mortierella ramanniana var. ramanniana as a function of pH, temperature and carbon source, nitrogen source, metal ions and oil supplementation. *Food Research International*, *38*(7), 815–829. https://doi.org/10.1016/j.foodres.2005.04.002

Fabrikov, D., Guil-Guerrero, J. L., González-Fernández, M. J., Rodríguez-García, I., Gómez-Mercado, F., Urrestarazu, M., Lao, M. T., Rincón-Cervera, M. Á., Álvaro, J. E., & Lyashenko, S. (2019). Borage oil: Tocopherols, sterols and squalene in farmed and endemic-wild Borago species. *Journal of Food Composition and Analysis*, *83*(June 2018), 103299. https://doi.org/10.1016/j.jfca.2019.103299

Fakas, S., Čertik, M., Papanikolaou, S., Aggelis, G., Komaitis, M., & Galiotou-Panayotou, M. (2008). γ-Linolenic acid production by Cunninghamella echinulata growing on complex organic nitrogen sources. *Bioresource Technology*, *99*(13), 5986–5990. https://doi.org/10.1016/j.biortech.2007.10.016

Franceschini, G., & Macchietto, S. (2008). Model-based design of experiments for parameter precision: State of the art. *Chemical Engineering Science*, *63*(19), 4846–4872. https://doi.org/10.1016/j.ces.2007.11.034

Gema, H., Kavadia, A., Dimou, D., Tsagou, V., Komaitis, M., & Aggelis, G. (2002). Production of γ-linolenic acid by Cunninghamella echinulata cultivated on glucose and orange peel. *Applied Microbiology and Biotechnology*, *58*(3), 303–307. https://doi.org/10.1007/s00253-001-0910-7

González-Figueredo, C., Alejandro Flores-Estrella, R., & A. Rojas-Rejón, O. (2019). Fermentation: Metabolism, Kinetic Models, and Bioprocessing. In *Current Topics in Biochemical Engineering* (pp. 1–17). IntechOpen. https://doi.org/10.5772/intechopen.82195

GPy. (2014). *GPy: A gaussian process framework in python*. GPy: A Gaussian Process Framework in Phython. http://github.com/SheffieldML/GPy

Guilherme, A. A., Silveira, M. S., Fontes, C. P. M. L., Rodrigues, S., & Fernandes, F. A. N. (2012). Modeling and Optimization of Lactic Acid Production using Cashew Apple Juice as Substrate. *Food and Bioprocess Technology*, *5*(8), 3151–3158. https://doi.org/10.1007/s11947-011-0670-z

Hart, W. E., Laird, C., Watson, J.-P., & Woodruff, D. L. (2017). Pyomo – Optimization Modeling in Python. In *Advances in Modeling Agricultural Systems* (Vol. 67). http://link.springer.com/10.1007/978-1-4614-3226-5

Horrobin, D. F. (1992). NUTRITIONAL GAMMA-LINOLENIC IMPORTANCE ACID. *Progress in Lipid Research*, *31*(2), 163–194.

HORROBIN, D. F. (1979). SCHIZOPHRENIA: RECONCILIATION OF THE DOPAMINE, PROSTAGLANDIN, AND OPIOID CONCEPTS AND THE ROLE OF THE PINEAL. *The Lancet*, *313*(8115), 529–531.

Horrobin D.F. (1979). Multiple sclerosis: The rational basis for treatment with colchicine and evening primrose oil. *Medical Hypotheses*, *5*(3), 365–378.

Jang, H.-D., Lin, Y.-Y., & Yang, S.-S. (2005). Effect of culture media and conditions on polyunsaturated fatty acids production by Mortierella alpina. *Bioresource Technology*, *96*(15), 1633–1644. https://doi.org/10.1016/j.biortech.2004.12.027

Jäntti, J., Seppälä, E., Vapaatalo, H., & Isomäki, H. (1989). Evening primrose oil and olive oil in treatment of rheumatoid arthritis. *Clinical Rheumatology*, *8*(2), 238–244. https://doi.org/10.1007/BF02030080

Kang, H. W., Kim, Y., Kim, S. W., & Choi, G. W. (2012). Cellulosic ethanol production on temperature-shift simultaneous saccharification and fermentation using the thermostable yeast Kluyveromyces marxianus CHY1612. *Bioprocess and Biosystems Engineering*, *35*(1–2), 115–122. https://doi.org/10.1007/s00449-011-0621-0

Kavadia, A., Komaitis, M., Chevalot, I., Blanchard, F., Marc, I., & Aggelis, G. (2001). Lipid and γ-linolenic acid accumulation in strains of zygomycetes growing on glucose. *JAOCS, Journal of the American Oil Chemists' Society*, *78*(4), 341–346. https://doi.org/10.1007/s11746-001-0266-3

Korting, M.J., K. H. C. (1992). Treatment of atopic eczema with evening primrose oil: rationale and clinical results. *Clinical Pharmacology*, *70*, 167–171.

Laidler, K. J. (1984). The development of the arrhenius equation. *Journal of Chemical Education*, *61*(6), 494–498. https://doi.org/10.1021/ed061p494

Li, S., Yu, H., Liu, Y., Zhang, X., & Ma, F. (2019). The lipid strategies in Cunninghamella echinulata for an allostatic response to temperature changes. *Process Biochemistry*, *76*(November 2018), 85–94. https://doi.org/10.1016/j.procbio.2018.11.005

Luedeking, R., & Piret, E. L. (1959). Kinetic study of the lactic acid fermentation. Batch process at controlled pH. *Biotechnology and Bioengineering*, *1*(4), 393–412. https://doi.org/10.1002/(SICI)1097-0290(20000320)67:6<636::AID-BIT3>3.0.CO;2-U

Ma, L., Li, C., Yang, Z., Jia, W., Zhang, D., & Chen, S. (2013). Kinetic studies on batch cultivation of Trichoderma reesei and application to enhance cellulase production by fed-batch fermentation. *Journal of Biotechnology*, *166*(4), 192–197. https://doi.org/10.1016/j.jbiotec.2013.04.023

Marchetti, A. G., François, G., Faulwasser, T., & Bonvin, D. (2016). Modifier adaptation for real-time optimization - Methods and applications. *Processes*, *4*(4), 1–35. https://doi.org/10.3390/pr4040055

Maria, G. (2004). A review of algorithms and trends in kinetic model identification for chemical and biochemical systems. *Chemical and Biochemical Engineering Quarterly*, *18*(3), 195–222.

Mowbray, M., Savage, T., Wu, C., Song, Z., Cho, B. A., Del Rio-Chanona, E. A., & Zhang, D. (2021). Machine learning for biochemical engineering: A review. *Biochemical Engineering Journal*, *172*(April), 108054. https://doi.org/10.1016/j.bej.2021.108054

Muloiwa, M., Nyende-Byakika, S., & Dinka, M. (2020). Comparison of unstructured kinetic bacterial growth models. *South African Journal of Chemical Engineering*, *33*(July), 141–150. https://doi.org/10.1016/j.sajce.2020.07.006

Nicoletti, M. C., Jain, L. C., & Giordano, R. C. (2009). Computational intelligence techniques as tools for bioprocess modelling, optimization, supervision and control. In M. do Carmo

Nicoletti & L. C. Jain (Eds.), *Studies in Computational Intelligence* (Vol. 218, pp. 1–23). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-01888-6_1

Peng, C., Huang, H., Ji, X., Liu, X., You, J., Lu, J., Cong, L., Xu, X., & Ouyang, P. (2010). A temperature-shift strategy for efficient arachidonic acid fermentation by Mortierella alpina in batch culture. *Biochemical Engineering Journal*, *53*(1), 92–96. https://doi.org/10.1016/j.bej.2010.09.014

Qiao, L., Yang, X., Xie, R., Du, C., Chi, Y., Zhang, J., & Wang, P. (2020). Efficient production of ulvan lyase from Ulva prolifera by Catenovulum sp. LP based on stage-controlled fermentation strategy. *Algal Research*, *46*(July 2019), 101812. https://doi.org/10.1016/j.algal.2020.101812

Rasmussen, C. E. (2004). *Gaussian Processes in Machine Learning* (O. Bousquet, U. von Luxburg, & G. Rätsch (eds.); pp. 63–71). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-28650-9_4

Rivera, E. C., Costa, A. C., Atala, D. I. P., Maugeri, F., Maciel, M. R. W., & Filho, R. M. (2006). Evaluation of optimization techniques for parameter estimation: Application to ethanol fermentation considering the effect of temperature. *Process Biochemistry*, *41*(7), 1682–1687. https://doi.org/10.1016/j.procbio.2006.02.009

Rohner, M., & Meyer, H. P. (1995). Applications of modelling for bioprocess design and control in industrial production. *Bioprocess Engineering*, *13*(2), 69–78. https://doi.org/10.1007/BF00420432

Sadino-Riquelme, M. C., Rivas, J., Jeison, D., Hayes, R. E., & Donoso-Bravo, A. (2020). Making sense of parameter estimation and model simulation in bioprocesses. *Biotechnology and Bioengineering*, *117*(5), 1357–1366. https://doi.org/10.1002/bit.27294

Safarian, S., Saryazdi, S. M. E., Unnthorsson, R., & Richter, C. (2021). Artificial Neural Network Modeling of Bioethanol Production Via Syngas Fermentation. *Biophysical Economics and Sustainability*, *6*(1), 1–13. https://doi.org/10.1007/s41247-020-00083-2

Sansana, J., Joswiak, M. N., Castillo, I., Wang, Z., Rendall, R., Chiang, L. H., & Reis, M. S. (2021). Recent trends on hybrid modeling for Industry 4.0. *Computers and Chemical Engineering*, *151*, 107365. https://doi.org/10.1016/j.compchemeng.2021.107365

Saraceno, A., Curcio, S., Calabrò, V., & Iorio, G. (2010). A hybrid neural approach to model batch fermentation of "ricotta cheese whey" to ethanol. *Computers and Chemical Engineering*, *34*(10), 1590–1596. https://doi.org/10.1016/j.compchemeng.2009.11.010

Shamloo, M., Babawale, E. A., Furtado, A., Henry, R. J., Eck, P. K., & Jones, P. J. H. (2017). Effects of genotype and temperature on accumulation of plant secondary metabolites in Canadian and Australian wheat grown under controlled environments. *Scientific Reports*, *7*(1), 1–13. https://doi.org/10.1038/s41598-017-09681-5

Sheng, X., Ma, J., & Xiong, W. (2020). Smart Soft Sensor Design with Hierarchical Sampling Strategy of Ensemble Gaussian Process Regression for Fermentation Processes. *Sensors*, *20*(7), 1957.

Sin, G., Gernaey, K. V., & Lantz, A. E. (2009). Good modeling practice for PAT applications: Propagation of input uncertainty and sensitivity analysis. *Biotechnology Progress*, *25*(4), 1043–1053. https://doi.org/10.1002/btpr.166

Sivagurunathan, P., Sen, B., & Lin, C. Y. (2014). Overcoming propionic acid inhibition of hydrogen fermentation by temperature shift strategy. *International Journal of Hydrogen Energy*, *39*(33), 19232–19241. https://doi.org/10.1016/j.ijhydene.2014.03.260

Somashekar, D., Venkateshwaran, G., Sambaiah, K., & Lokesh, B. R. (2003). Effect of culture conditions on lipid and gamma-linolenic acid production by mucoraceous fungi. *Process Biochemistry*, *38*(12), 1719–1724. https://doi.org/10.1016/S0032-9592(02)00258-3

Song, Z. (2021). *Kinetic modelling of γ -linolenic acid production by Cunninghamella echinulata by*. The University of Manchester.

Souza, A. C. M., Mousaviraad, M., Mapoka, K. O. M., & Rosentrater, K. A. (2018). Kinetic modeling of corn fermentation with S. cerevisiae using a variable temperature strategy. *Bioengineering*, *5*(2). https://doi.org/10.3390/bioengineering5020034

Sun, M. L., Madzak, C., Liu, H. H., Song, P., Ren, L. J., Huang, H., & Ji, X. J. (2017). Engineering Yarrowia lipolytica for efficient γ-linolenic acid production. *Biochemical Engineering Journal*, *117*, 172–180. https://doi.org/10.1016/j.bej.2016.10.014

Sun, S. L., Wu, B., Zhao, D. Y., Zhang, X. X., Zhang, Y., Li, W. X., & Cheng, S. P. (2009). *Optimization of Xhhh strain biodegradation with metal ions for pharmaceautical wastewater treatment*. *30*(September), 877–882.

Tanticharoen, M., Reungjitchachawali, M., Boonag, B., Vonktaveesuk, P., Vonshak, A., & Cohen, Z. (1994). Optimization of γ-linolenic acid (GLA) production in Spirulina platensis. *Journal of Applied Phycology*, *6*(3), 295–300. https://doi.org/10.1007/BF02181942

Vega-Ramon, F., Zhu, X., Savage, T. R., Petsagkourakis, P., Jing, K., & Zhang, D. (2021). Kinetic and hybrid modeling for yeast astaxanthin production under uncertainty. *Biotechnology and Bioengineering*, *118*(12), 4854–4866. https://doi.org/10.1002/bit.27950

Voll, A., & Marquardt, W. (2012). Reaction Network Flux Analysis: Optimization-Based Evaluation of Reaction Pathways for Biorenewables Processing. *AIChE Journal*, *58*(6), 1788–1801. https://doi.org/10.1002/aic

von Stosch, M., Oliveira, R., Peres, J., & Feyo de Azevedo, S. (2014). Hybrid semi-parametric modeling in process systems engineering: Past, present and future. *Computers and Chemical Engineering*, *60*, 86–101. https://doi.org/10.1016/j.compchemeng.2013.08.008

Wächter, A., & Lorenz T. Biegler. (2006). Catalogue of an exhibition of the works of Dante Alighieri March to October 1909. In *Math. Program.* (Vol. 57).

Wang, X., Chen, J., Liu, C., & Pan, F. (2010). Hybrid modeling of penicillin fermentation process based on least square support vector machine. *Chemical Engineering Research and Design*, *88*(4), 415–420. https://doi.org/10.1016/j.cherd.2009.08.010

Wang, Y., Shi, J., & Gong, L. (2020). Gamma linolenic acid suppresses hypoxia-induced proliferation and invasion of non-small cell lung cancer cells by inhibition of HIF1α. *Genes and Genomics*, *42*(8), 927–935. https://doi.org/10.1007/s13258-020-00961-5

Wannawilai, S., Jeennor, S., Khongto, B., & Laoteng, K. (2020). Exploring differential traits of lipid-producing stages of the wild type and morphologically engineered strain of Aspergillus oryzae by comparative kinetic modeling. *World Journal of Microbiology and Biotechnology*, *36*(12), 1–16. https://doi.org/10.1007/s11274-020-02959-3

Williams, C. (2006). Gaussian Processes for Machine Learning (adaptive computation and machine learning). *Institute for Adaptive and Neural Computation*, *August*.

Willis, M. J., & von Stosch, M. (2017). Simultaneous parameter identification and discrimination of the nonparametric structure of hybrid semi-parametric models. *Computers and Chemical Engineering*, *104*, 366–376. https://doi.org/10.1016/j.compchemeng.2017.05.005

Zambrano, J., & Carlsson, B. (2014). *Optimizing zone volumes in bioreactors described by Monod and Contois growth kinetics*. *September*, 1–5. https://doi.org/10.13140/RG.2.1.4757.4640

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), 2025–2039. https://doi.org/10.1002/bit.25610

Zhang, D., Del Rio-Chanona, E. A., Petsagkourakis, P., & Wagner, J. (2019). Hybrid physics-based and data-driven modeling for bioprocess online simulation and optimization.

Biotechnology and Bioengineering, *116*(11), 2919–2930. https://doi.org/10.1002/bit.27120

Zhang, P., Chen, Y., Zhou, Q., Zheng, X., Zhu, X., & Zhao, Y. (2010). Understanding short-chain fatty acids accumulation enhanced in waste activated sludge alkaline fermentation: Kinetics and microbiology. *Environmental Science and Technology*, *44*(24), 9343–9348. https://doi.org/10.1021/es102878m

Appendix A:

A.1 Experimental setup

A.1.1 Microorganism and culture conditions

C. echinulata X-15 was a screened high-yield strain in our lab, which was maintained on potato dextrose agar (PDA) plates at 4 °C and transferred every 3 weeks to PDA plates. The culture was grown at 28 °C for 2 days, and then stored at 4 °C until fermentations were conducted. The seed culture medium consisted of (per liter): 80 g glucose, 6 g urea, 5 g $(NH_4)_2SO_4$, 1 g yeast extract, 5 g KH_2PO_4, 1 g $MgSO_4$, 0.06 mg Biotin. The fermentation medium consisted of (per liter): 100 g glucose, 6 g tryptone, 5 g yeast extract, 2 g $NaNO_3$, 3 g KH_2PO_4, 0.5 g $MgSO_4$. The culture was grown at 28 °C.

A.1.2 Inoculum preparation and fermentation

The 250 ml baffled flasks containing 50 ml seed medium was inoculated with a loop of mycelium of *C. echinulata* and cultivated at 28 °C, 150 rpm for 2 days. GLA fermentation was carried out in 1L bioreactors (Infors-2015 Bioprocess controller, Netherland) containing 0.7 L medium and cultivated at different temperatures (14, 28, and 37 °C) with 10% (v/v) of the seed culture without pH controlled. The cultivation time in the experiments ranged from 300 to 480 h as indicated. Samples were taken periodically for analysis. In the two temperature-shift experiments, the GLA fermentation was carried out in a 5L bioreactor (Infors-2015 Bioprocess controller, Netherland) containing 3.5L medium, ultimately switching the temperature from 37 to 14 °C at 168 hours and 96 hours, respectively. All the control conditions were same as those in the single temperature experiments. All cultivation experiments were performed in triplicate

and analysed individually.

A.1.3 Analytical methods

Mycelia after cultivation were harvested by vacuum filtration, washed with distilled water, and then dried at 70 °C until constant weight. Total lipids were extracted with chloroform/methanol (2:1, v/v) according to the methods of Bligh and Dyer (Bligh & Dyer, 1959). For fatty-acid methyl ester (FAME) analysis, lipids were saponated by 2 mL 0.5 M KOH-CH_3OH for 15 min at 65 °C, then esterified by 2 mL BF_3-CH_3OH for 30 min at 70 °C. Fatty-acid methyl ester (FAME) were extracted with hexane and analysed by gas chromatography. Gas chromatography was performed with an Agilent GC-7890A gas chromatograph (Agilent, USA) equipped with a flame ionisation detector (FID) and a split injector with a split-to-splitless ratio of 100:1. Glucose concentration was determined by biosensor with glucose oxidase electrode (SBA-40C, Shandong).

Appendix B:

Table B.1 tabulates the mean relative percentage deviation (MRPD) between the lower and upper bound for each state variable and operating temperature.

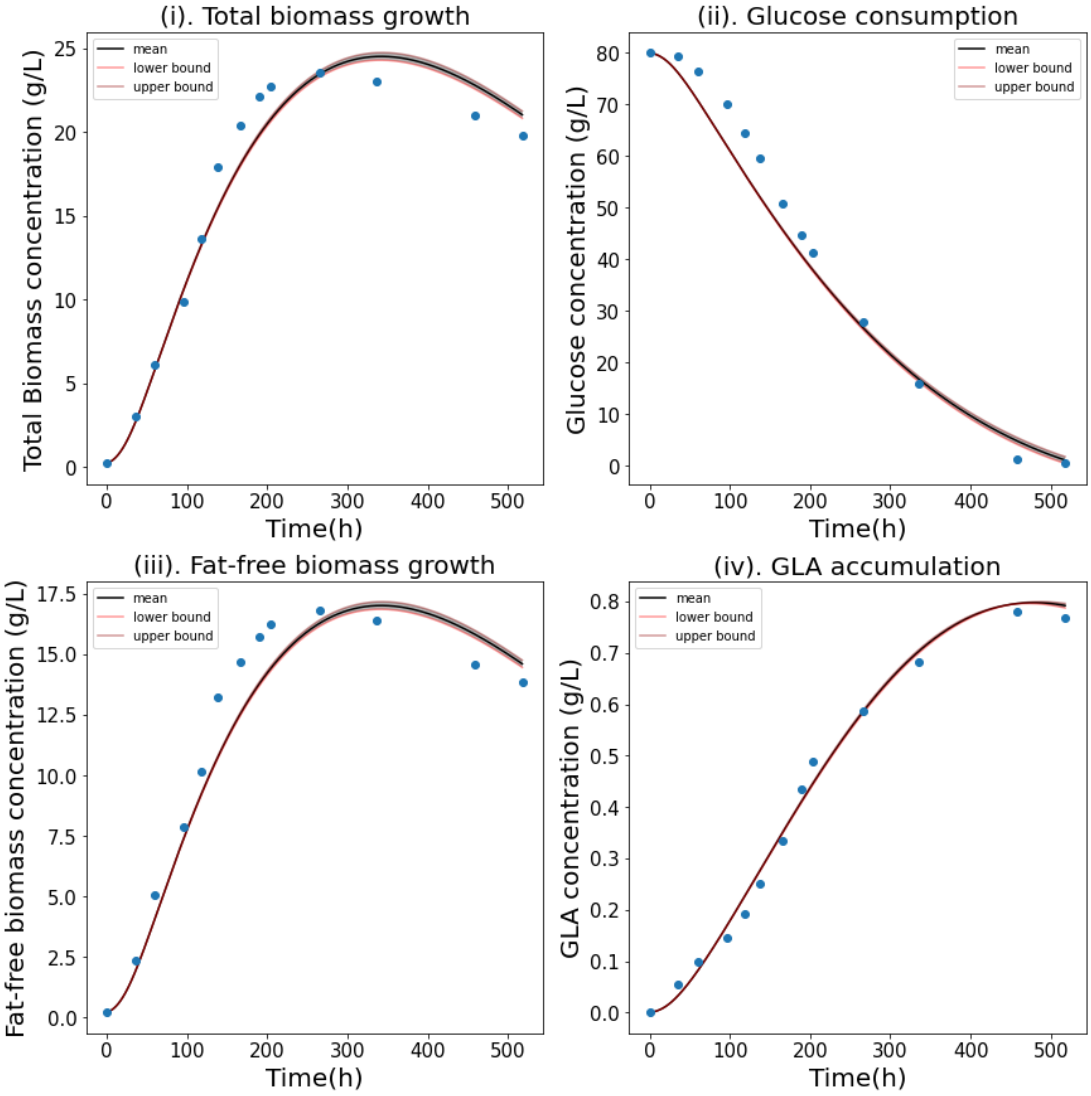
Table B.1. Model relative deviation (%) w.r.t. 5% variations of parameter permutation.

T/℃	Var.	m	μ_d	k_n	A_{μ_m}	B_{μ_m}	$A_{k_{co}}$	$B_{k_{co}}$	A_{k_0}	B_{k_0}
14	X_T	1.63	1.40	0	6.62	56.87	4.50	15.23	0	0
	C	15.68	10.60	0	24.57	182.13	20.76	66.37	0	0
	X_B	1.62	1.39	0	6.51	55.89	4.45	14.98	9.53	31.41
	X_G	1.88	0.73	8.28	6.99	59.96	4.18	14.07	0	0
28	X_T	1.25	1.22	0	7.46	60.11	5.41	17.33	0	0
	C	8.04	4.29	0	15.44	114.14	13.19	42.09	0	0
	X_B	1.24	1.21	0	7.41	59.59	5.39	17.25	9.59	30.16
	X_G	1.86	0.56	13.98	7.13	58.76	4.44	14.27	0	0
37	X_T	1.62	1.31	0	7.72	61.12	5.98	18.60	0	0
	C	16.62	6.61	0	23.38	145.10	20.67	57.95	0	0
	X_B	1.62	1.31	0	7.69	60.92	5.97	18.56	9.68	29.55
	X_G	3.62	0.85	21.25	8.07	64.47	5.67	17.67	0	0
		A_{k_m}	B_{k_m}	A_{k_d}	B_{k_d}	A_{K_p}	B_{K_p}	$A_{Y_{co}}$	$B_{Y_{co}}$	$C_{Y_{co}}$
14	X_T	0	0	0	0	0	0	0.28	10.57	3.17
	C	0	0	0	0	0	0	0.57	23.74	6.46
	X_B	0	0	0	0	0	0	0.28	10.53	3.16

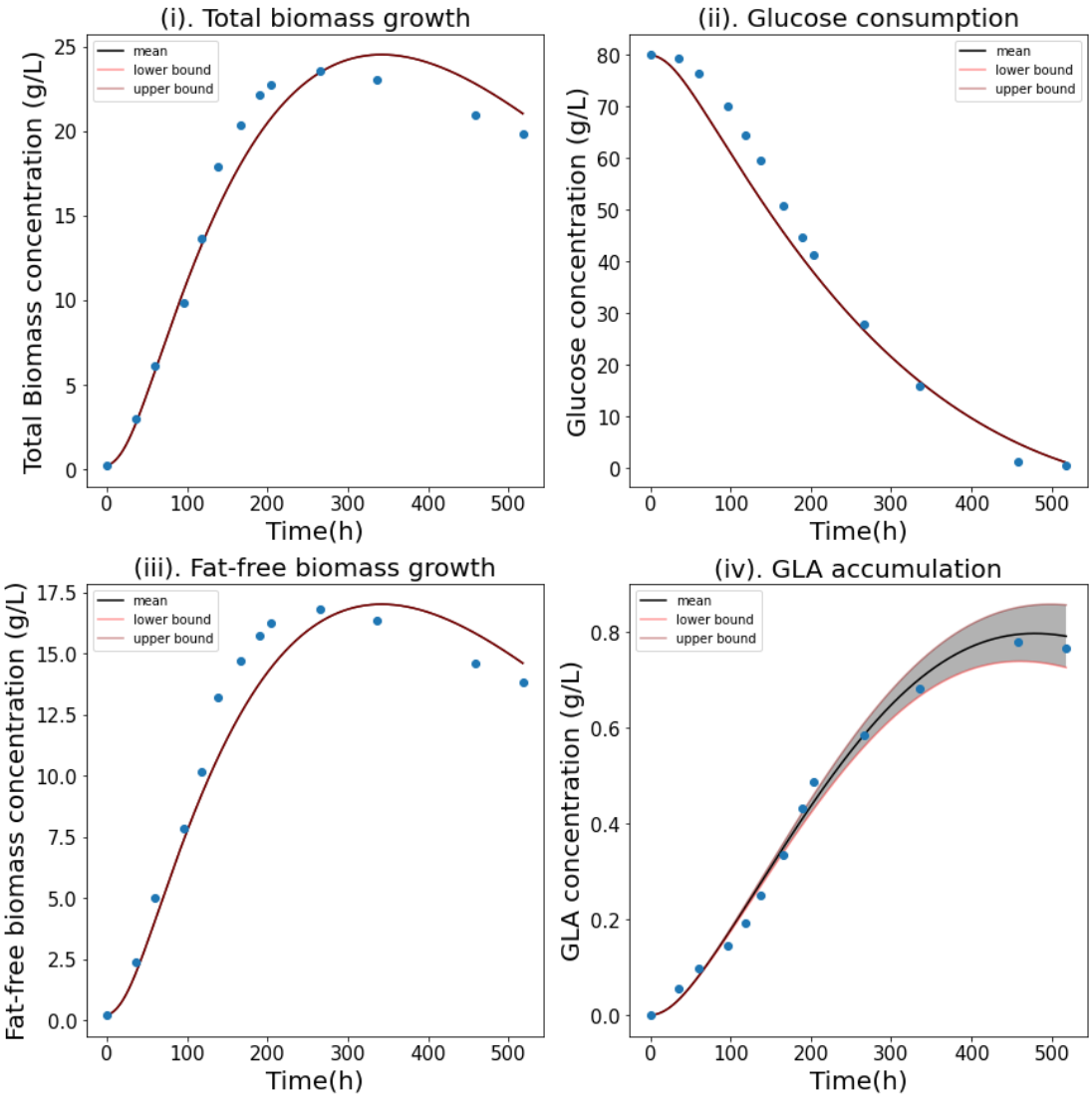
	X_G	5.31	234.27	3.65	37.13	1.79	10.23	0.32	12.07	3.62
28	X_T	0	0	0	0	0	0	1.05	44.19	3.46
	C	0	0	0	0	0	0	2.47	136.4	8.11
	X_B	0	0	0	0	0	0	1.05	44.11	3.46
	X_G	3.35	136.73	7.39	71.38	3.58	19.42	1.48	61.48	4.83
37	X_T	0	0	0	0	0	0	2.91	203.6	4.56
	C	0	0	0	0	0	0	8.48	1998	13.29
	X_B	0	0	0	0	0	0	2.91	203.5	4.55
	X_G	2.30	89.40	13.61	114.05	6.89	36.31	5.44	345.8	8.51

Appendix C:

Figure C.1. displays the propagated uncertainty for total biomass, fat-free biomass, glucose, and GLA concentration resulting from a 5% variation in part parameters at different operating temperatures.

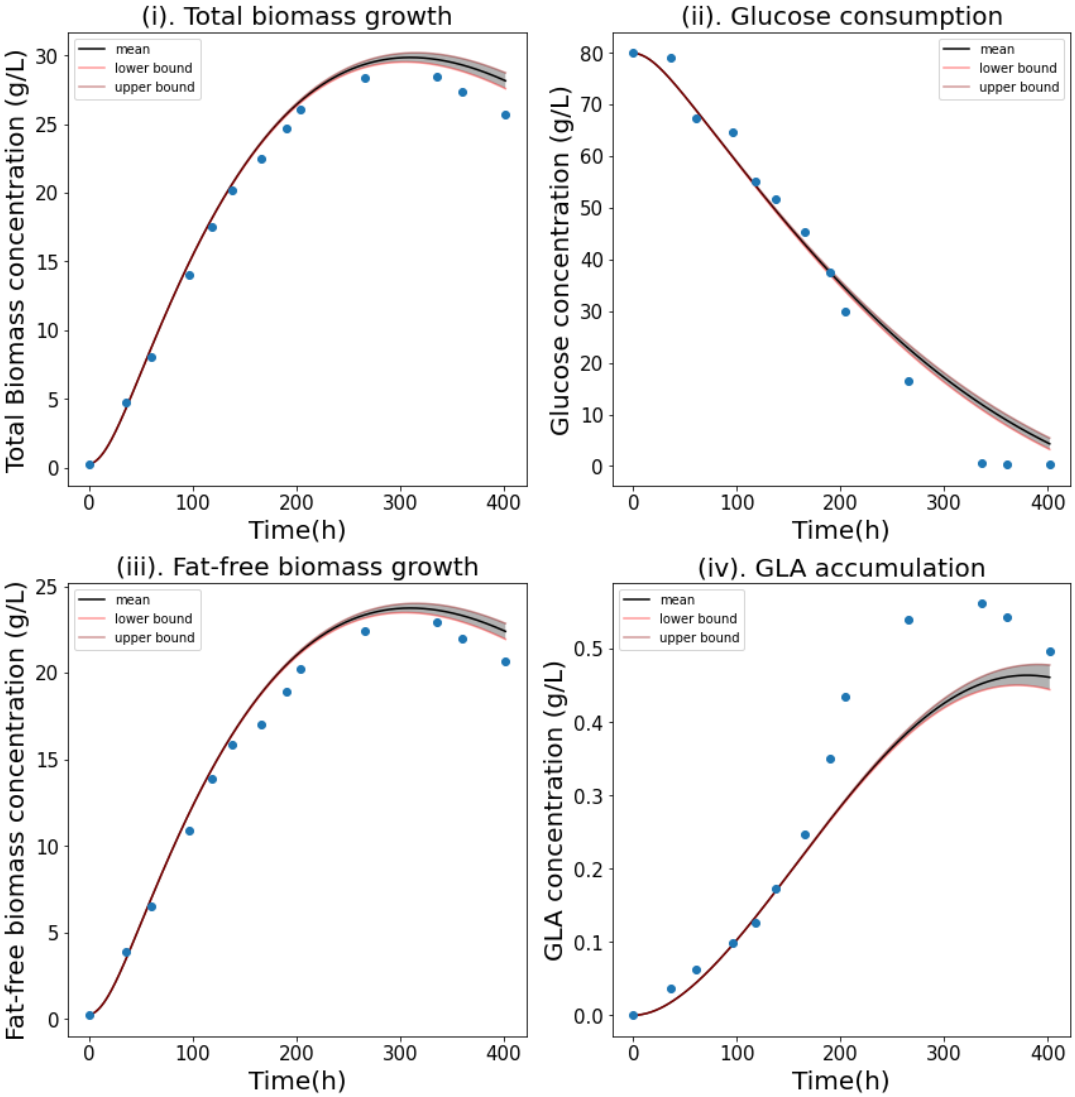


μ_d

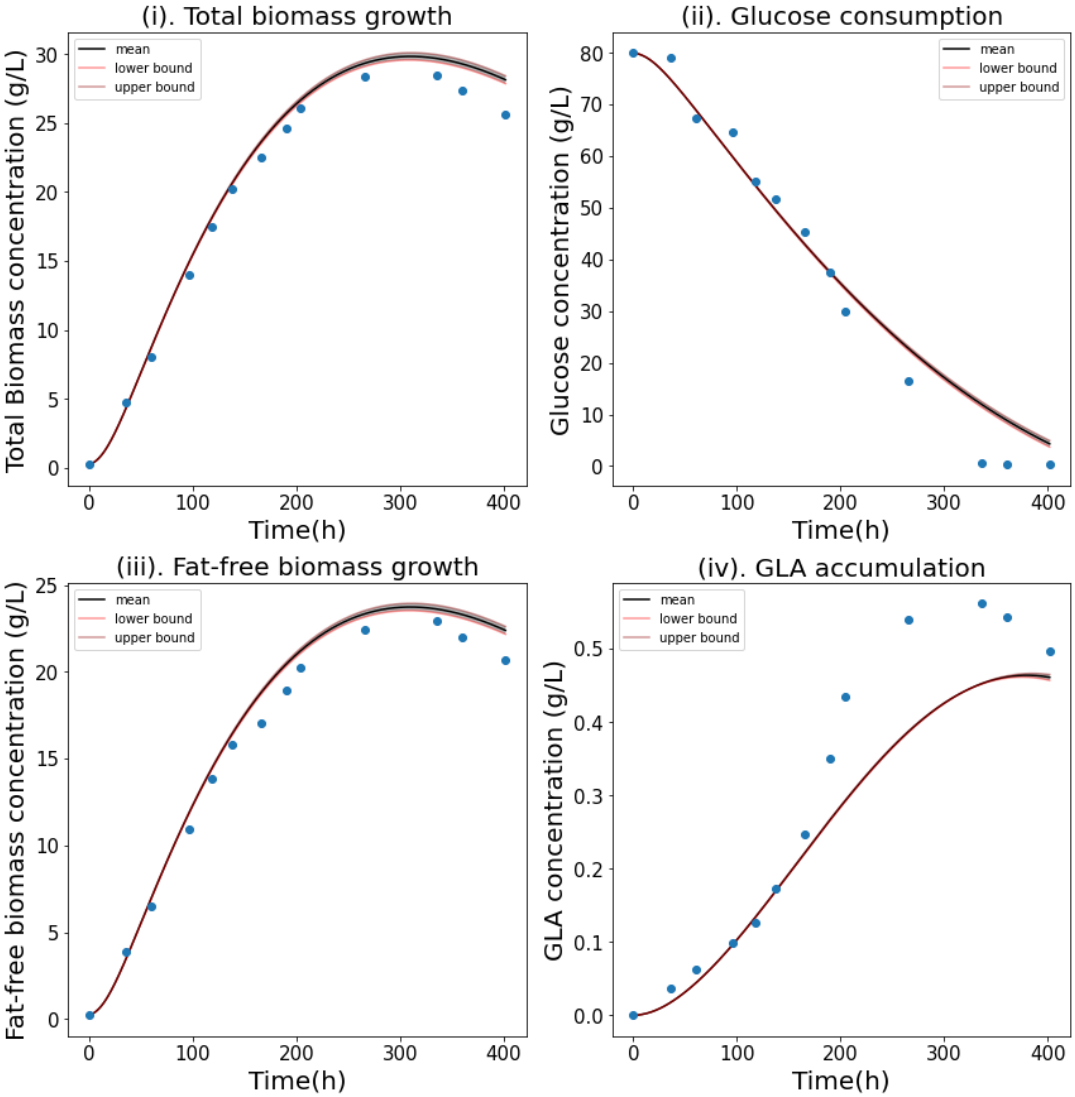

(i). Total biomass growth
Total Biomass concentration (g/L)
Time(h)
(ii). Glucose consumption
Glucose concentration (g/L)
Time(h)
(iii). Fat-free biomass growth
Fat-free biomass concentration (g/L)
Time(h)
(iv). GLA accumulation
GLA concentration (g/L)
Time(h)
mean
lower bound
upper bound

k_n

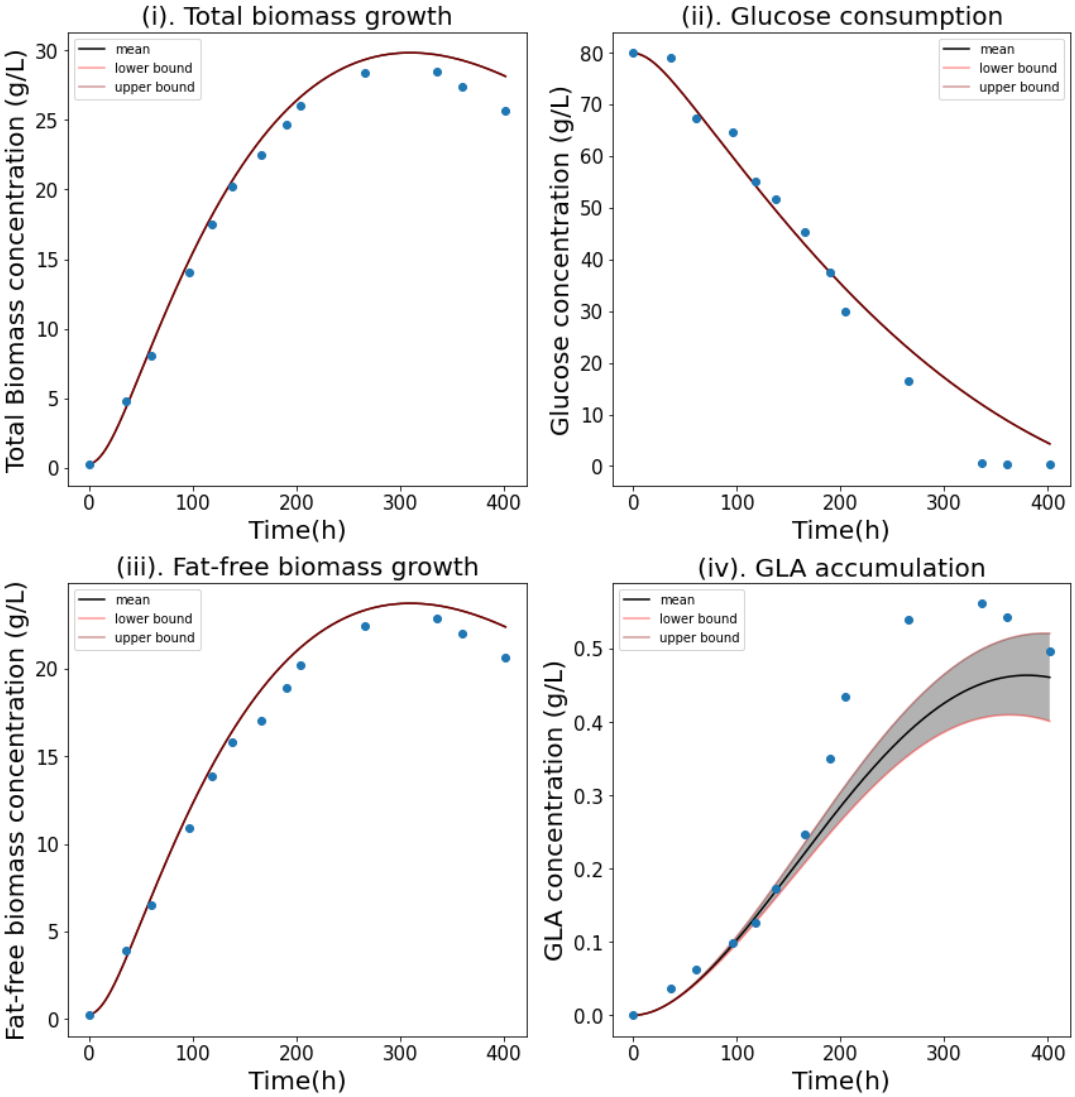
(A)


(i). Total biomass growth
Total Biomass concentration (g/L)
Time(h)
mean
lower bound
upper bound
(ii). Glucose consumption
Glucose concentration (g/L)
Time(h)
mean
lower bound
upper bound
(iii). Fat-free biomass growth
Fat-free biomass concentration (g/L)
Time(h)
mean
lower bound
upper bound
(iv). GLA accumulation
GLA concentration (g/L)
Time(h)
mean
lower bound
upper bound

m

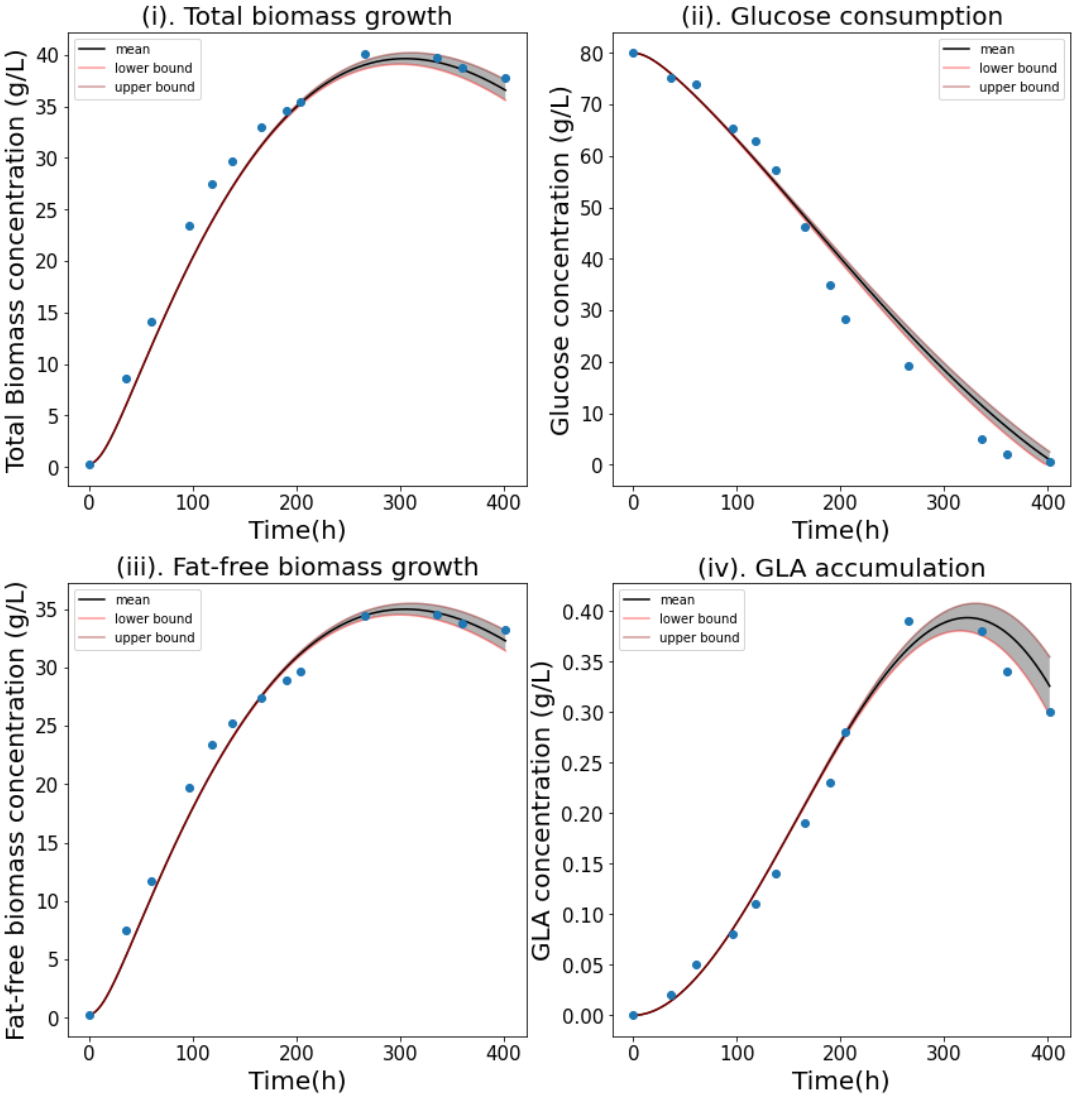

(i). Total biomass growth
Total Biomass concentration (g/L)
Time(h)
mean
lower bound
upper bound
(ii). Glucose consumption
Glucose concentration (g/L)
Time(h)
mean
lower bound
upper bound
(iii). Fat-free biomass growth
Fat-free biomass concentration (g/L)
Time(h)
mean
lower bound
upper bound
(iv). GLA accumulation
GLA concentration (g/L)
Time(h)
mean
lower bound
upper bound

μ_d

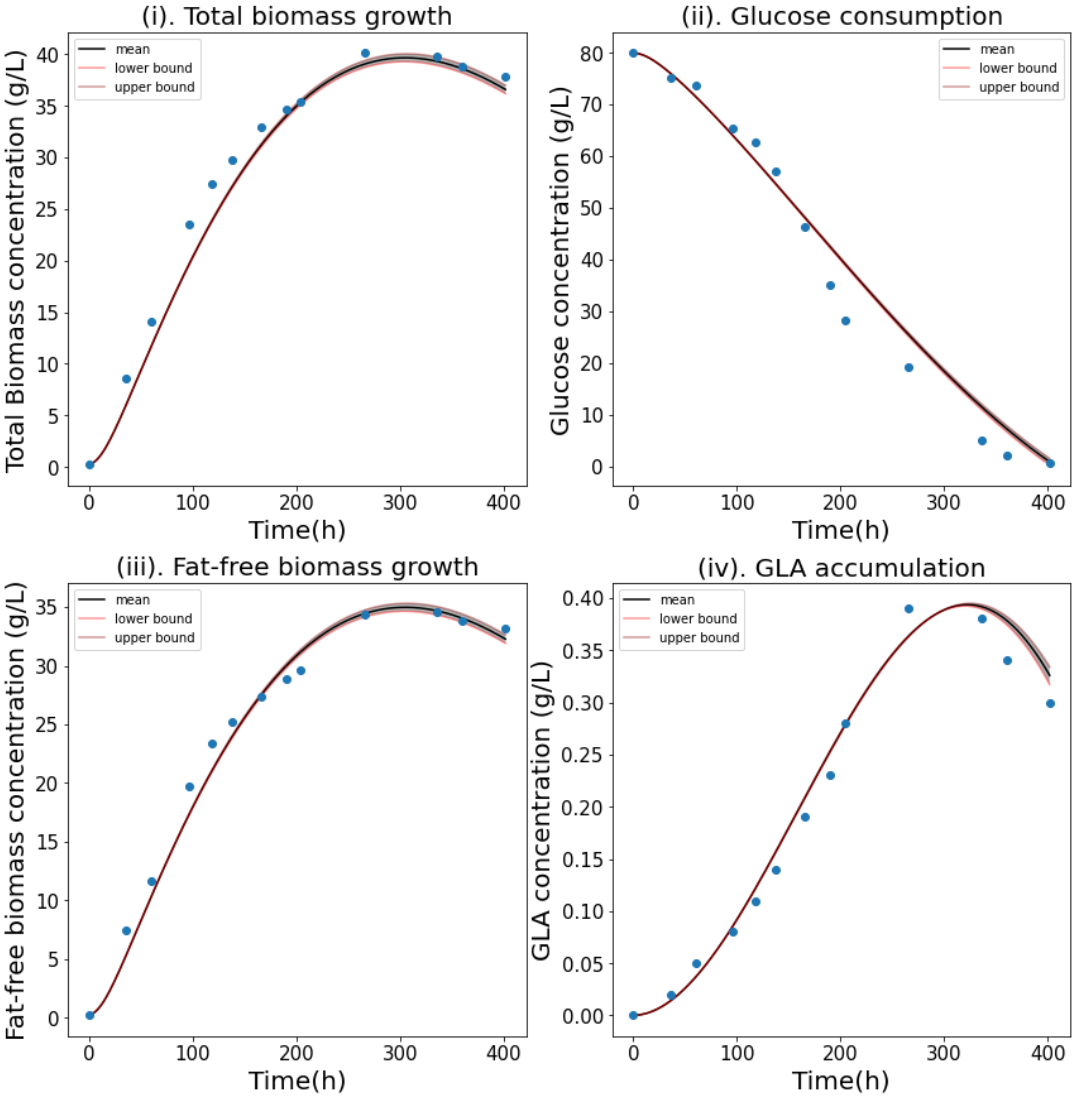

(i). Total biomass growth
Total Biomass concentration (g/L)
Time(h)
mean
lower bound
upper bound
(ii). Glucose consumption
Glucose concentration (g/L)
Time(h)
mean
lower bound
upper bound
(iii). Fat-free biomass growth
Fat-free biomass concentration (g/L)
Time(h)
mean
lower bound
upper bound
(iv). GLA accumulation
GLA concentration (g/L)
Time(h)
mean
lower bound
upper bound

k_n

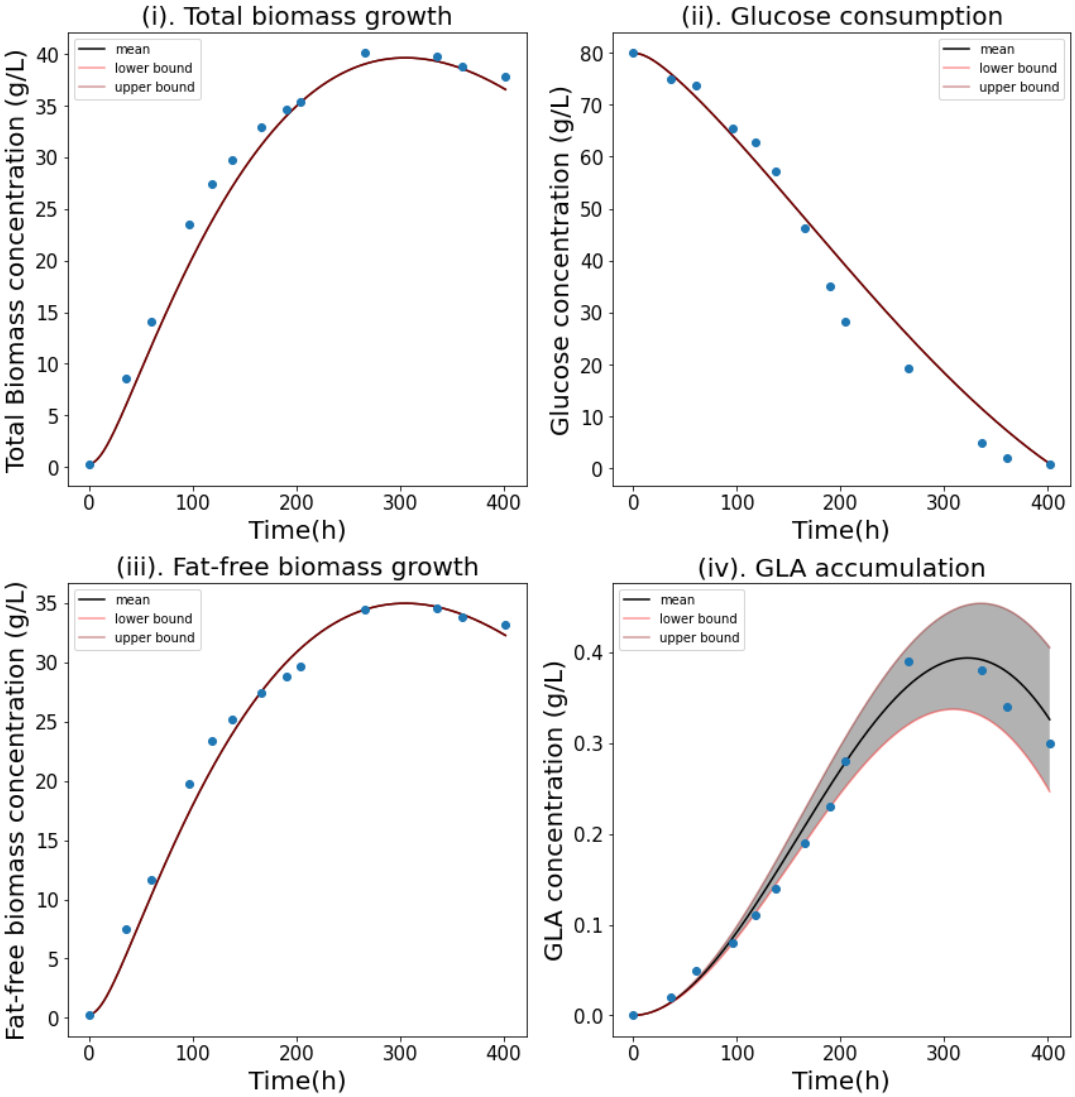
(B)


(i). Total biomass growth
Total Biomass concentration (g/L)
Time(h)
mean
lower bound
upper bound
(ii). Glucose consumption
Glucose concentration (g/L)
Time(h)
mean
lower bound
upper bound
(iii). Fat-free biomass growth
Fat-free biomass concentration (g/L)
Time(h)
mean
lower bound
upper bound
(iv). GLA accumulation
GLA concentration (g/L)
Time(h)
mean
lower bound
upper bound

m


(i). Total biomass growth
Total Biomass concentration (g/L)
Time(h)
mean
lower bound
upper bound
(ii). Glucose consumption
Glucose concentration (g/L)
Time(h)
mean
lower bound
upper bound
(iii). Fat-free biomass growth
Fat-free biomass concentration (g/L)
Time(h)
mean
lower bound
upper bound
(iv). GLA accumulation
GLA concentration (g/L)
Time(h)
mean
lower bound
upper bound

μ_d



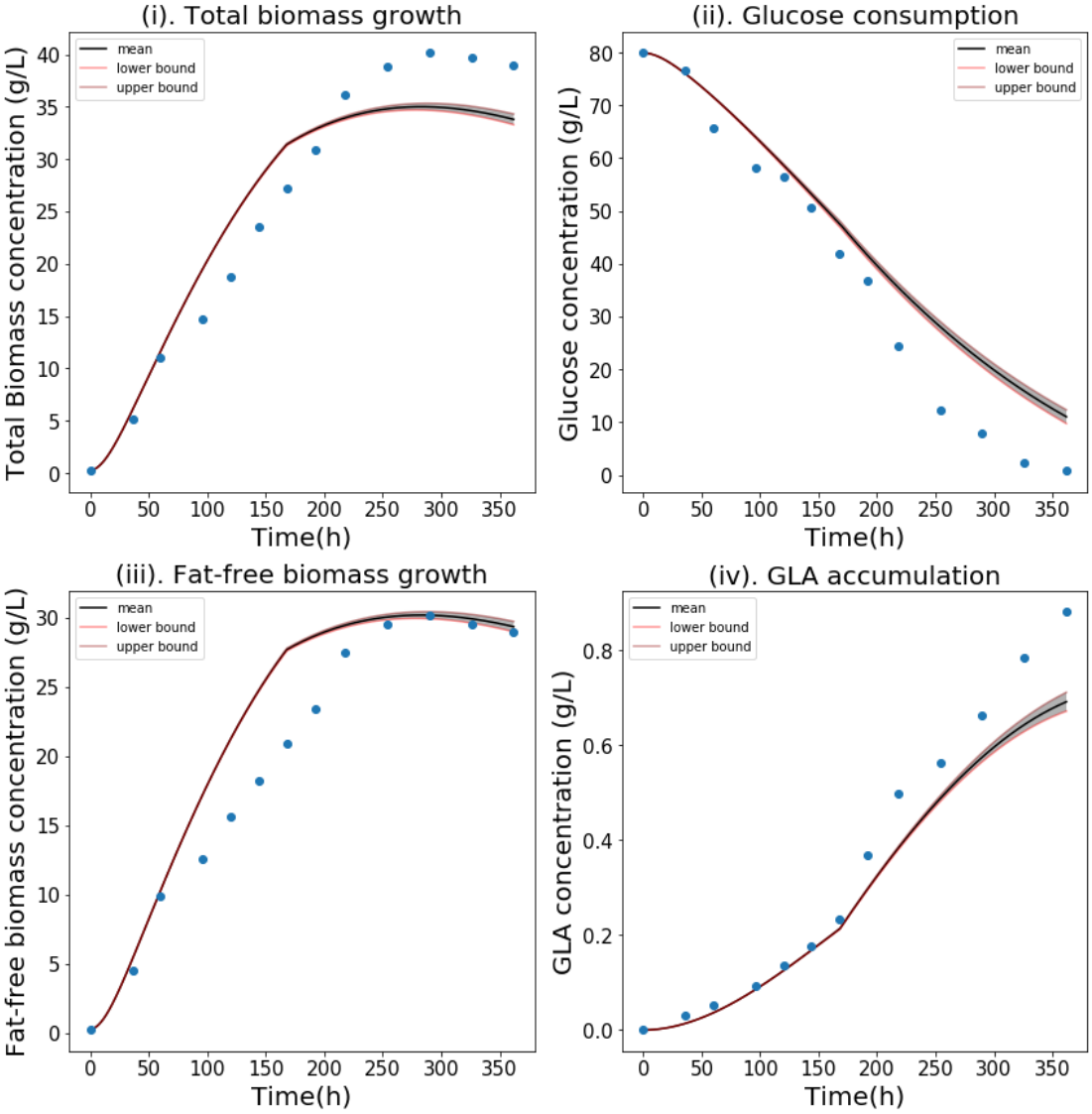
k_n

(C)

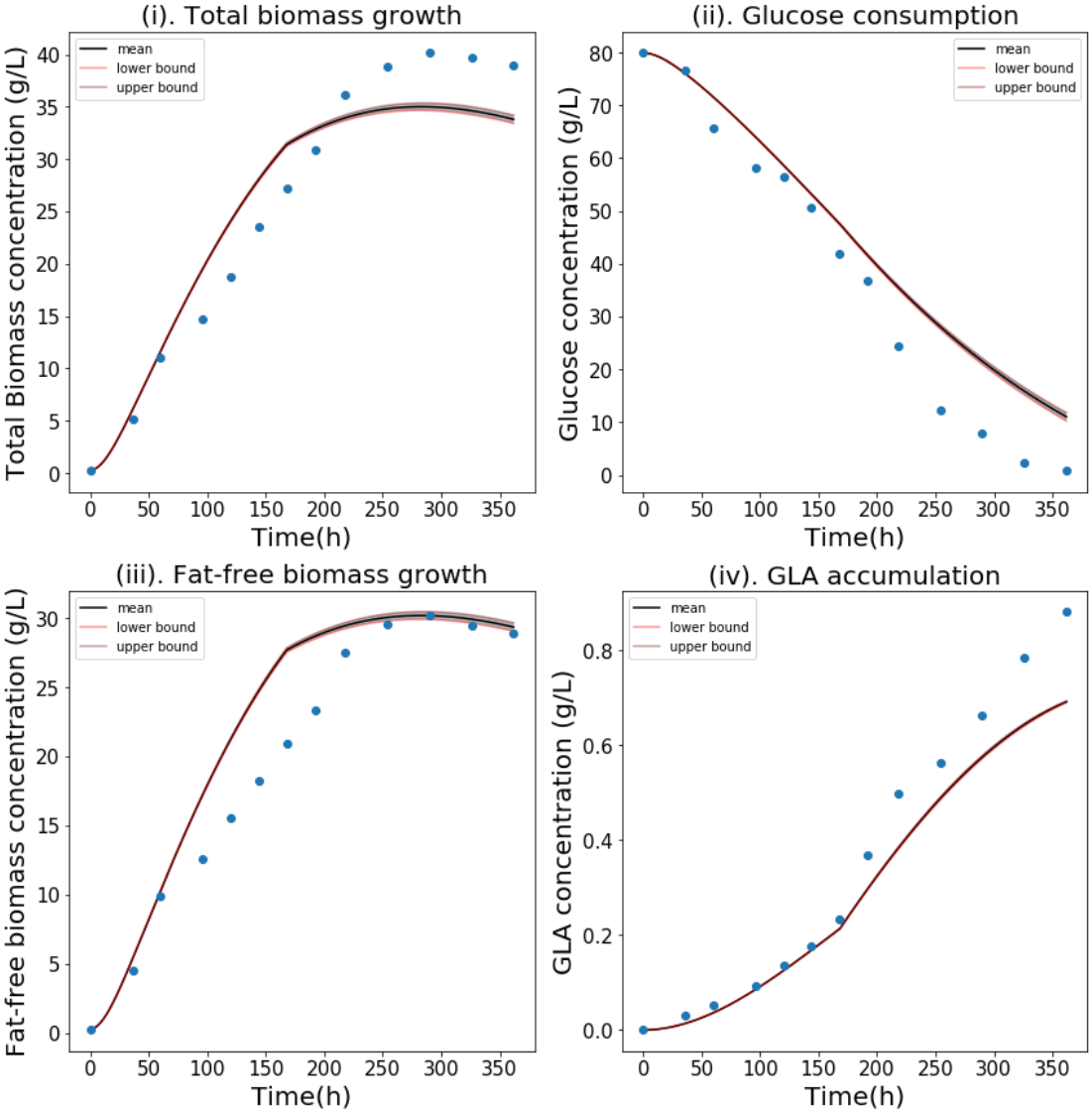
Figure C.1. Uncertainty with 5% variation for parameters in the model predictions for total biomass, glucose, fat-free biomass and GLA at 14 ℃ (A), 28 ℃ (B) and 37 ℃ (C).

Appendix D:

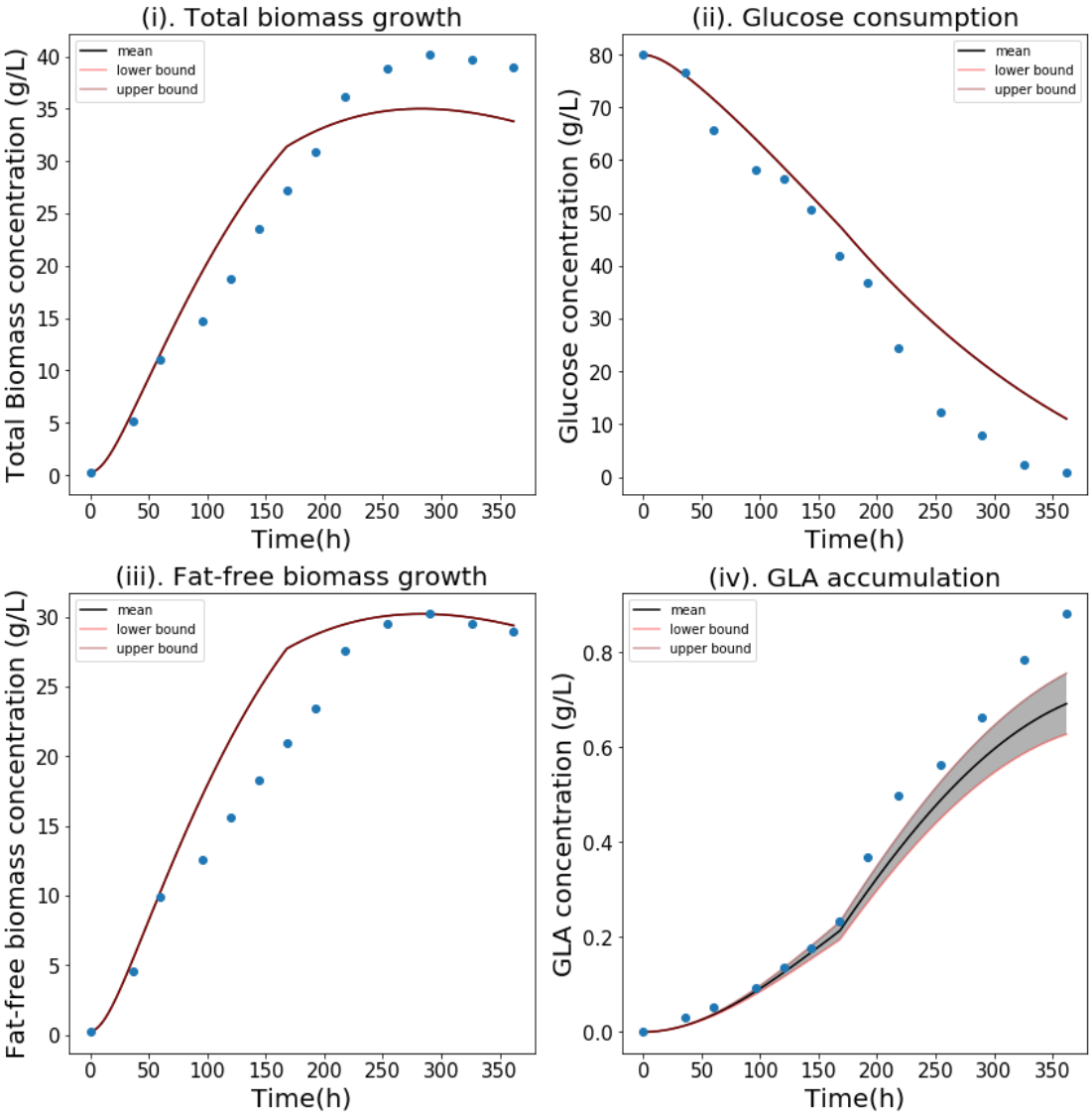
Figure D.1 show the uncertainty trajectories of part parameters generated by the Latin Hypercube Sampling (LHS) for 200 samples in temperature-shift at 168 hr and 96 hr.



m

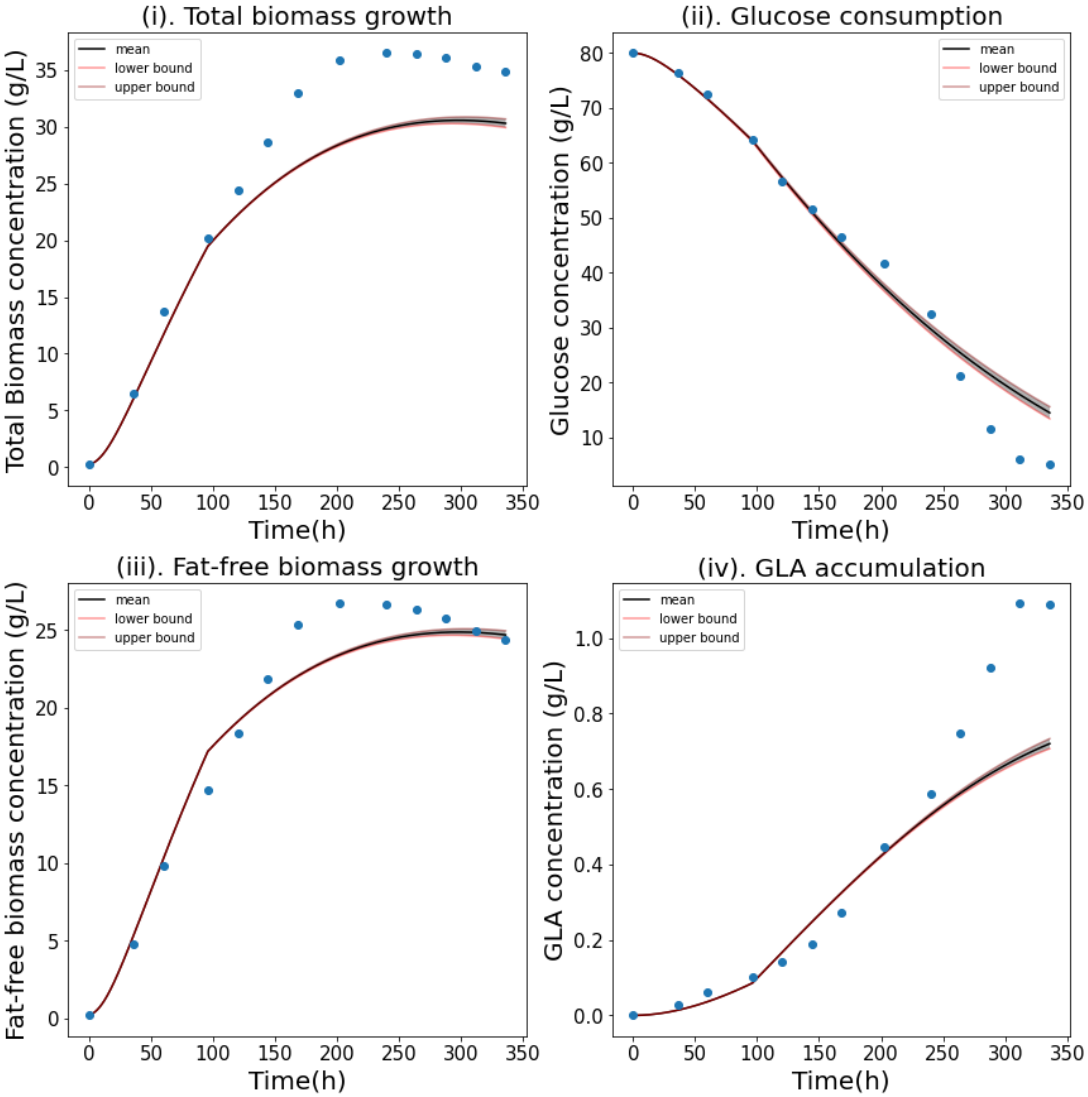

(i). Total biomass growth
Total Biomass concentration (g/L)
Time(h)
mean
lower bound
upper bound
(ii). Glucose consumption
Glucose concentration (g/L)
Time(h)
mean
lower bound
upper bound
(iii). Fat-free biomass growth
Fat-free biomass concentration (g/L)
Time(h)
mean
lower bound
upper bound
(iv). GLA accumulation
GLA concentration (g/L)
Time(h)
mean
lower bound
upper bound

μ_d

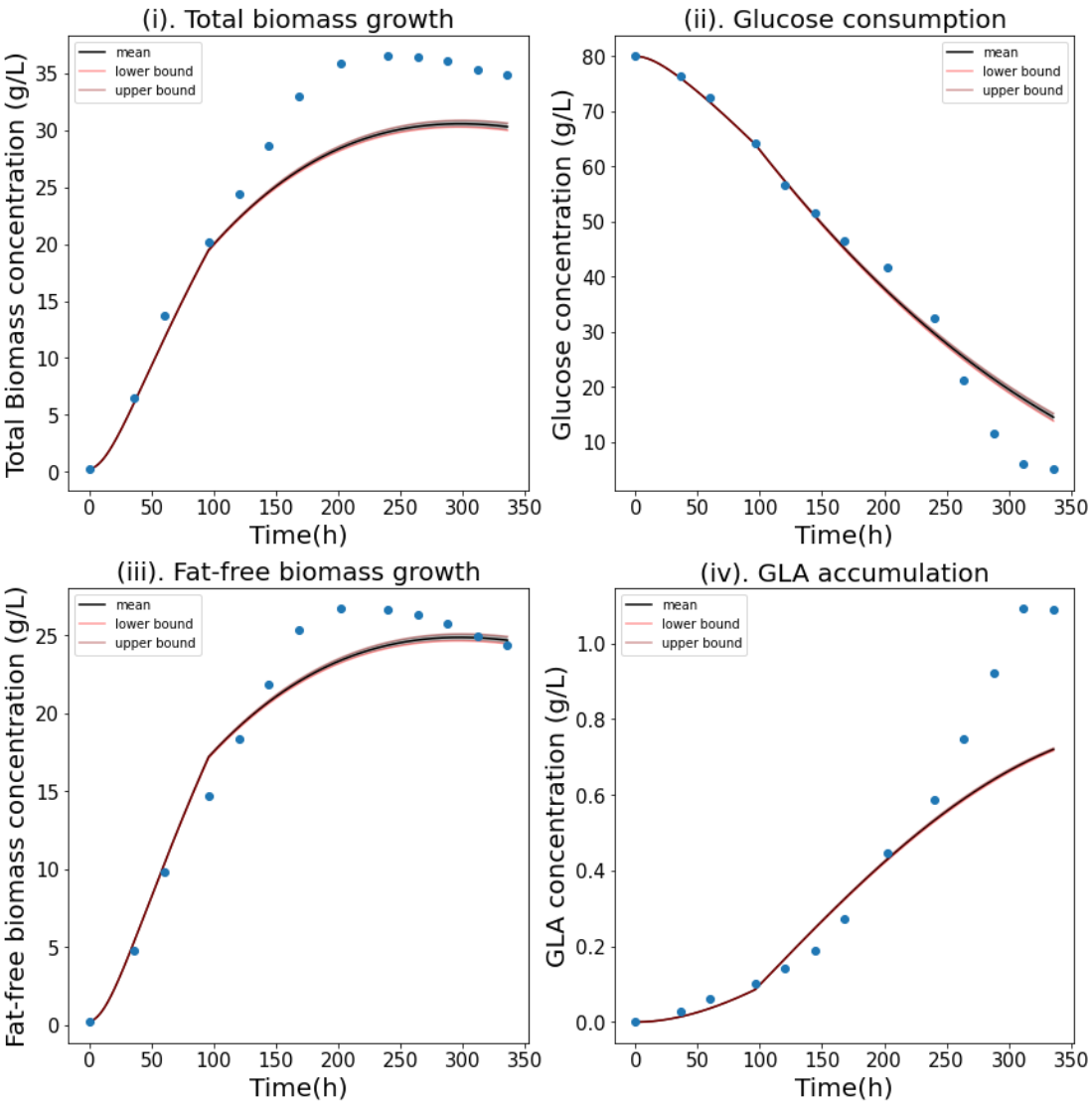

(i). Total biomass growth
Total Biomass concentration (g/L)
Time(h)
(ii). Glucose consumption
Glucose concentration (g/L)
Time(h)
(iii). Fat-free biomass growth
Fat-free biomass concentration (g/L)
Time(h)
(iv). GLA accumulation
GLA concentration (g/L)
Time(h)
mean
lower bound
upper bound

k_n

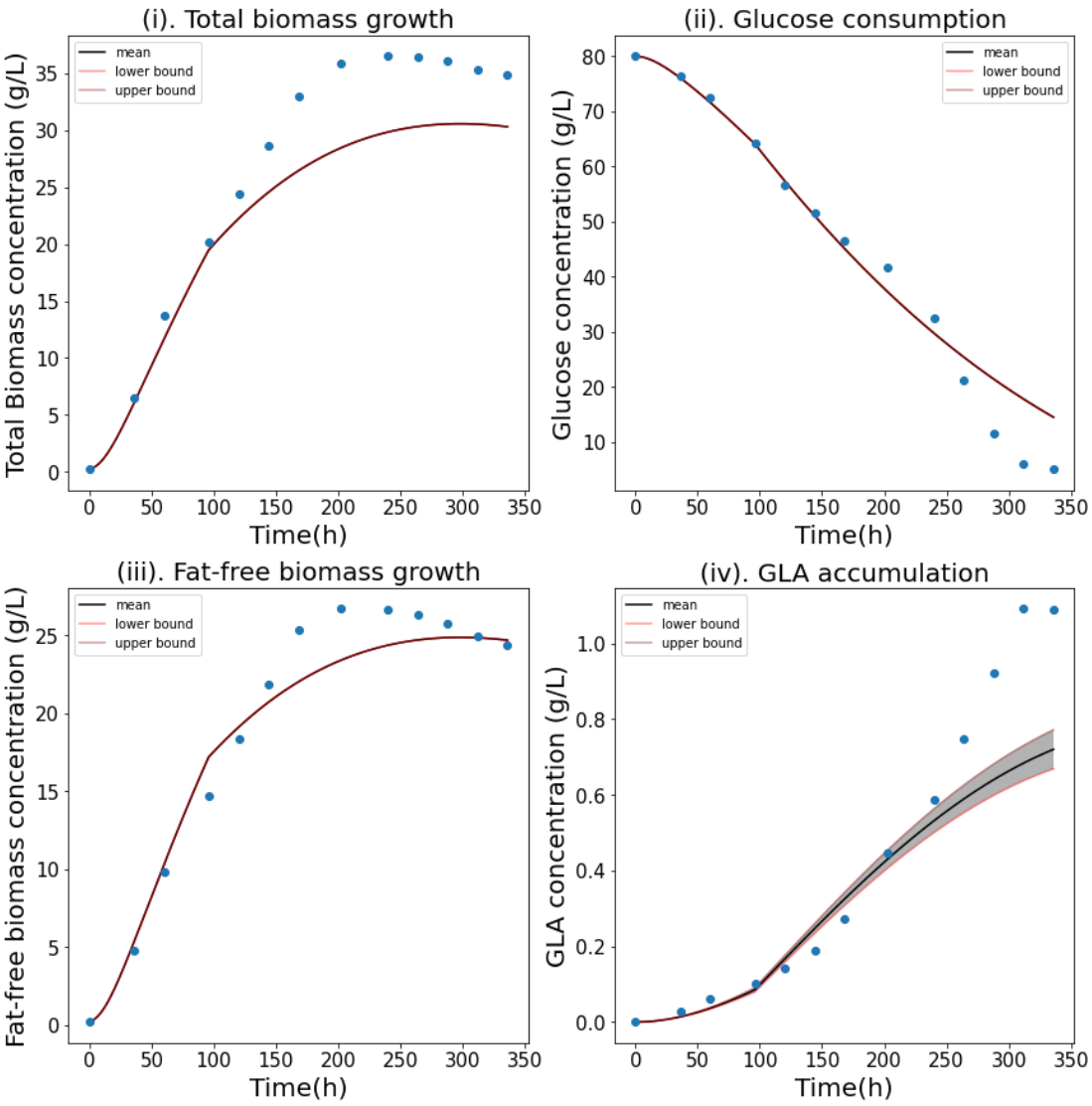
(A)


(i). Total biomass growth
Total Biomass concentration (g/L)
Time(h)
mean
lower bound
upper bound
(ii). Glucose consumption
Glucose concentration (g/L)
Time(h)
mean
lower bound
upper bound
(iii). Fat-free biomass growth
Fat-free biomass concentration (g/L)
Time(h)
mean
lower bound
upper bound
(iv). GLA accumulation
GLA concentration (g/L)
Time(h)
mean
lower bound
upper bound

m


(i). Total biomass growth
Total Biomass concentration (g/L)
Time(h)
mean
lower bound
upper bound
(ii). Glucose consumption
Glucose concentration (g/L)
Time(h)
mean
lower bound
upper bound
(iii). Fat-free biomass growth
Fat-free biomass concentration (g/L)
Time(h)
mean
lower bound
upper bound
(iv). GLA accumulation
GLA concentration (g/L)
Time(h)
mean
lower bound
upper bound

μ_d



k_n

(B)

Figure D.1. The uncertainty trajectories of models for different parameters with 5% input variation in temperature-shift at (A) 168 hr, (B) 96 hr.

Appendix E:

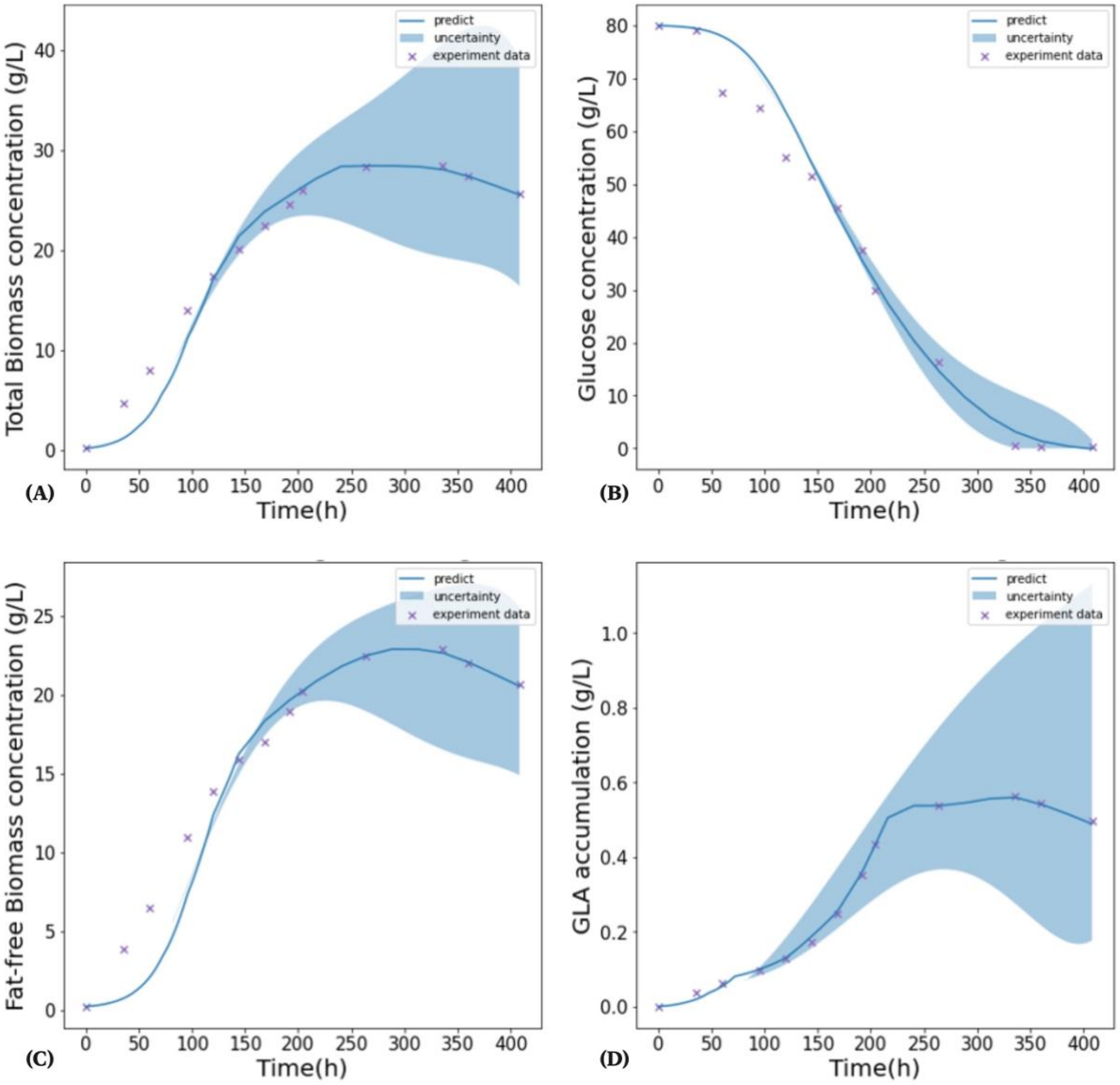


Figure E.1: Simulation result of Hybrid Model 1 for total biomass concentration (A), glucose consumption (B), fat-free biomass concentration (C) and GLA accumulation (D) at 28℃.

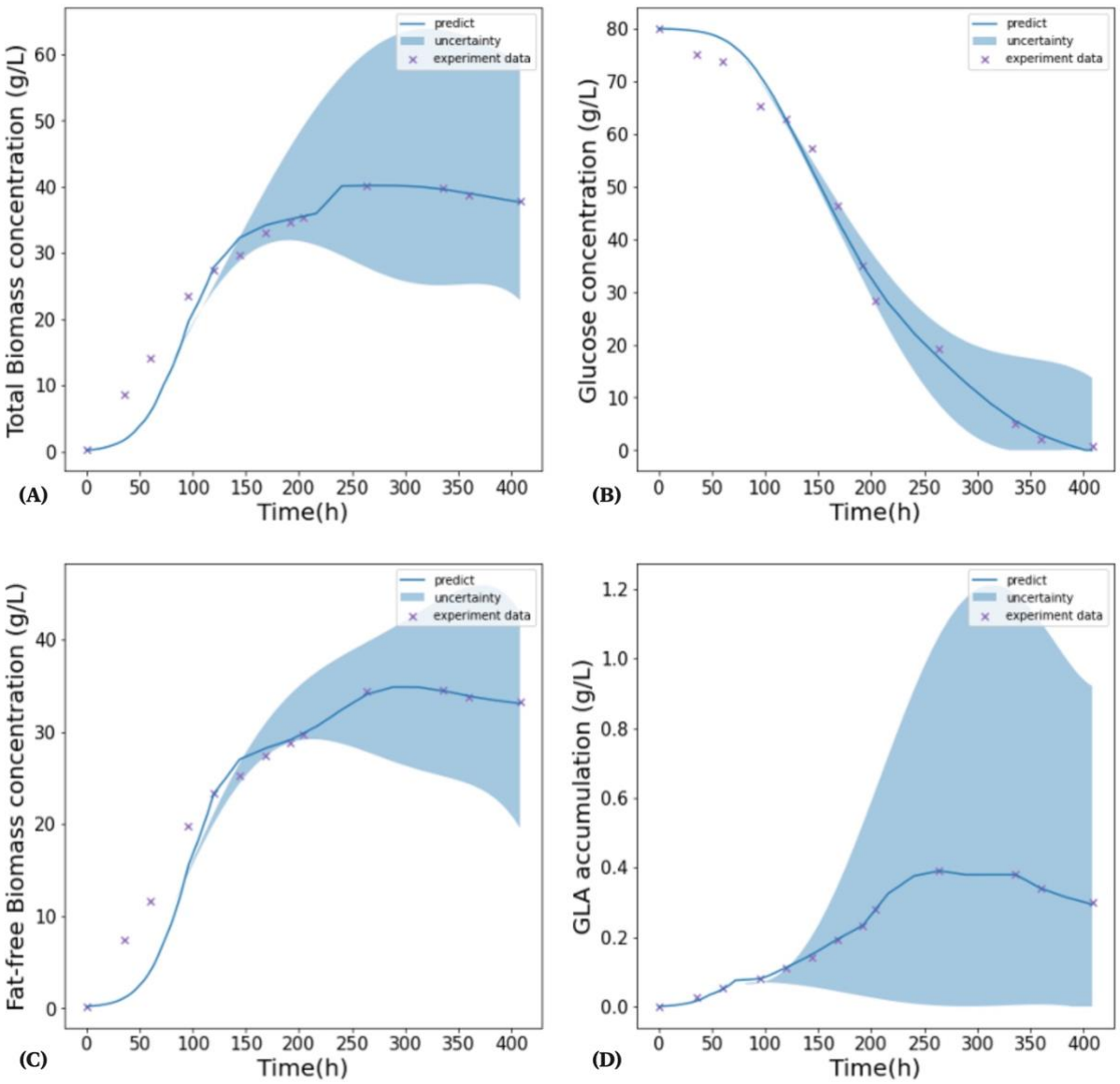


Figure E.2: Simulation result of Hybrid Model 1 for total biomass concentration (A), glucose consumption (B), fat-free biomass concentration (C) and GLA accumulation (D) at 37℃.

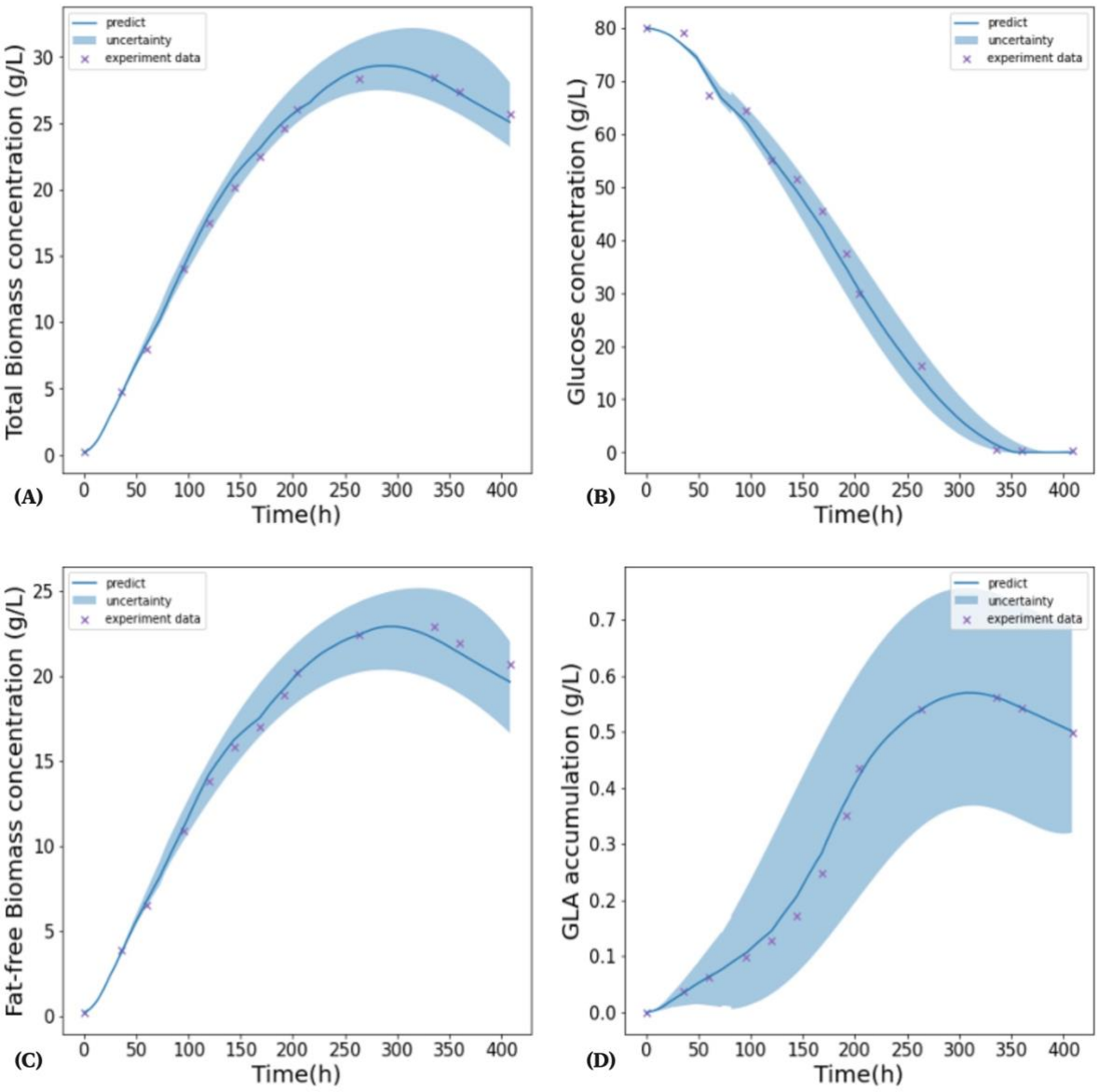


Figure E.3: Simulation result of Hybrid Model 2 for total biomass concentration (A), glucose consumption (B), fat-free biomass concentration (C) and GLA accumulation (D) at 28℃.

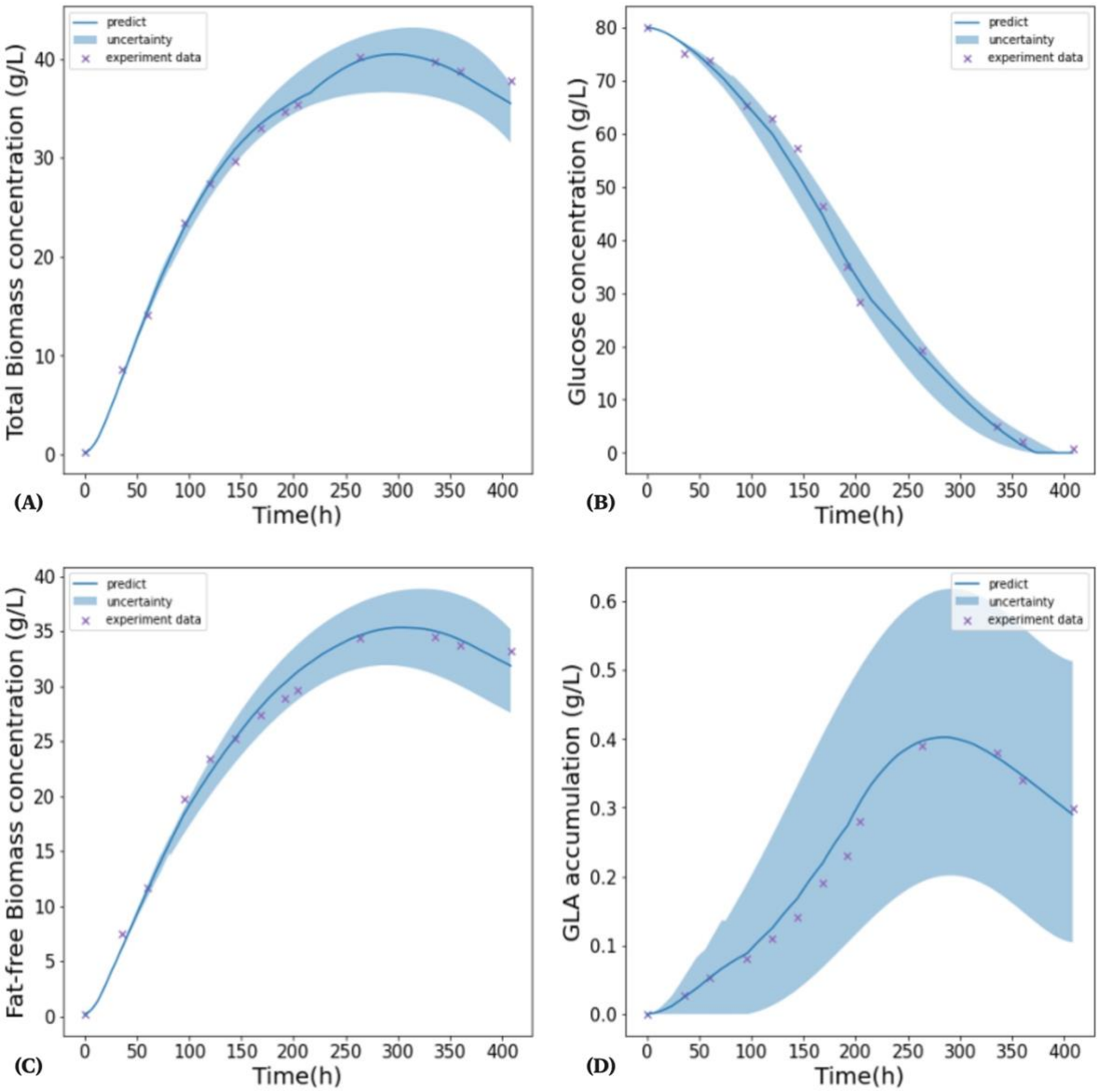


Figure E.4: Simulation result of Hybrid Model 2 for total biomass concentration (A), glucose consumption (B), fat-free biomass concentration (C) and GLA accumulation (D) at 37℃.

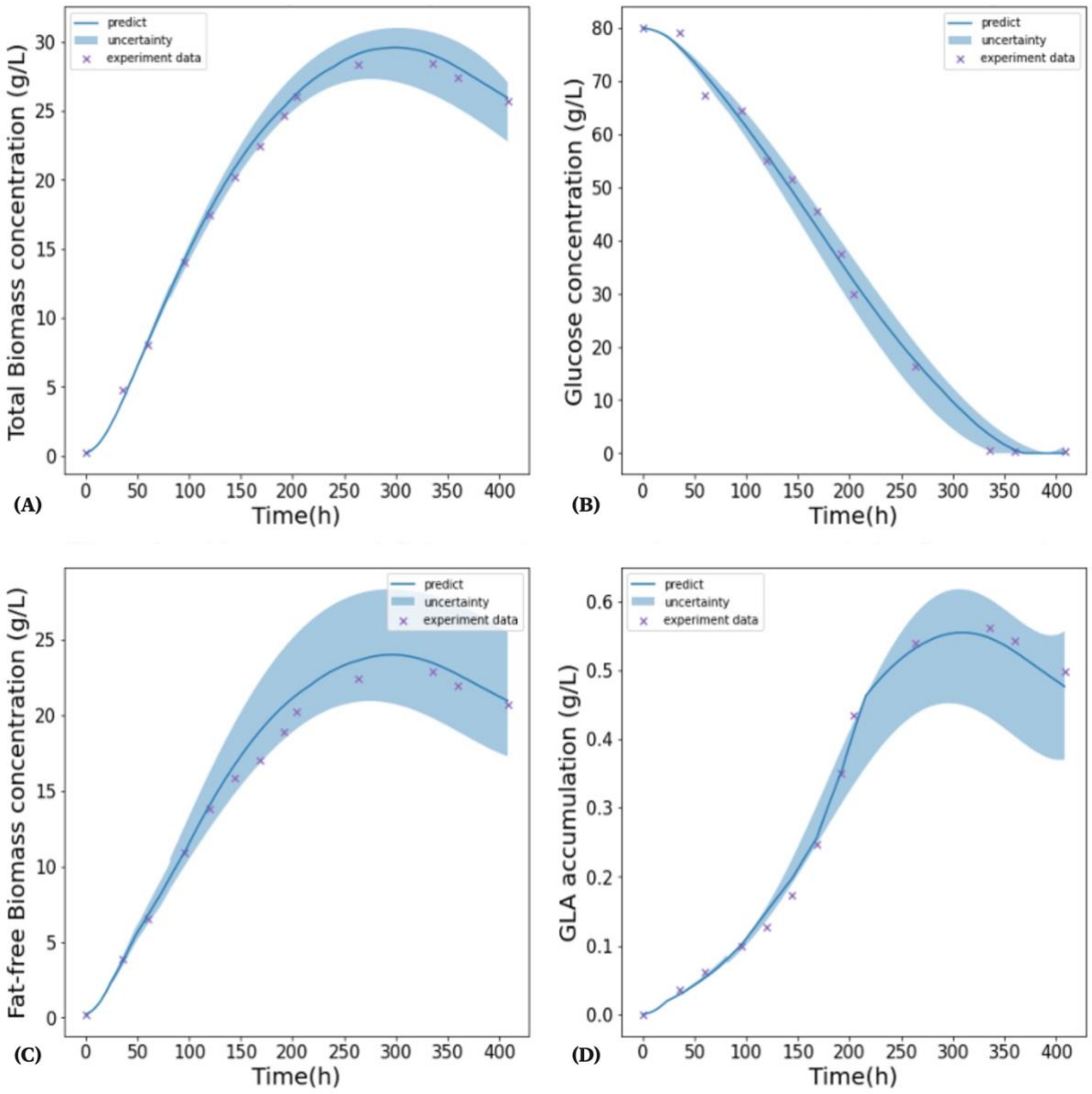


Figure E.5: Simulation result of Hybrid Model 3 for total biomass concentration (A), glucose consumption (B), fat-free biomass concentration (C) and GLA accumulation (D) at 28℃.

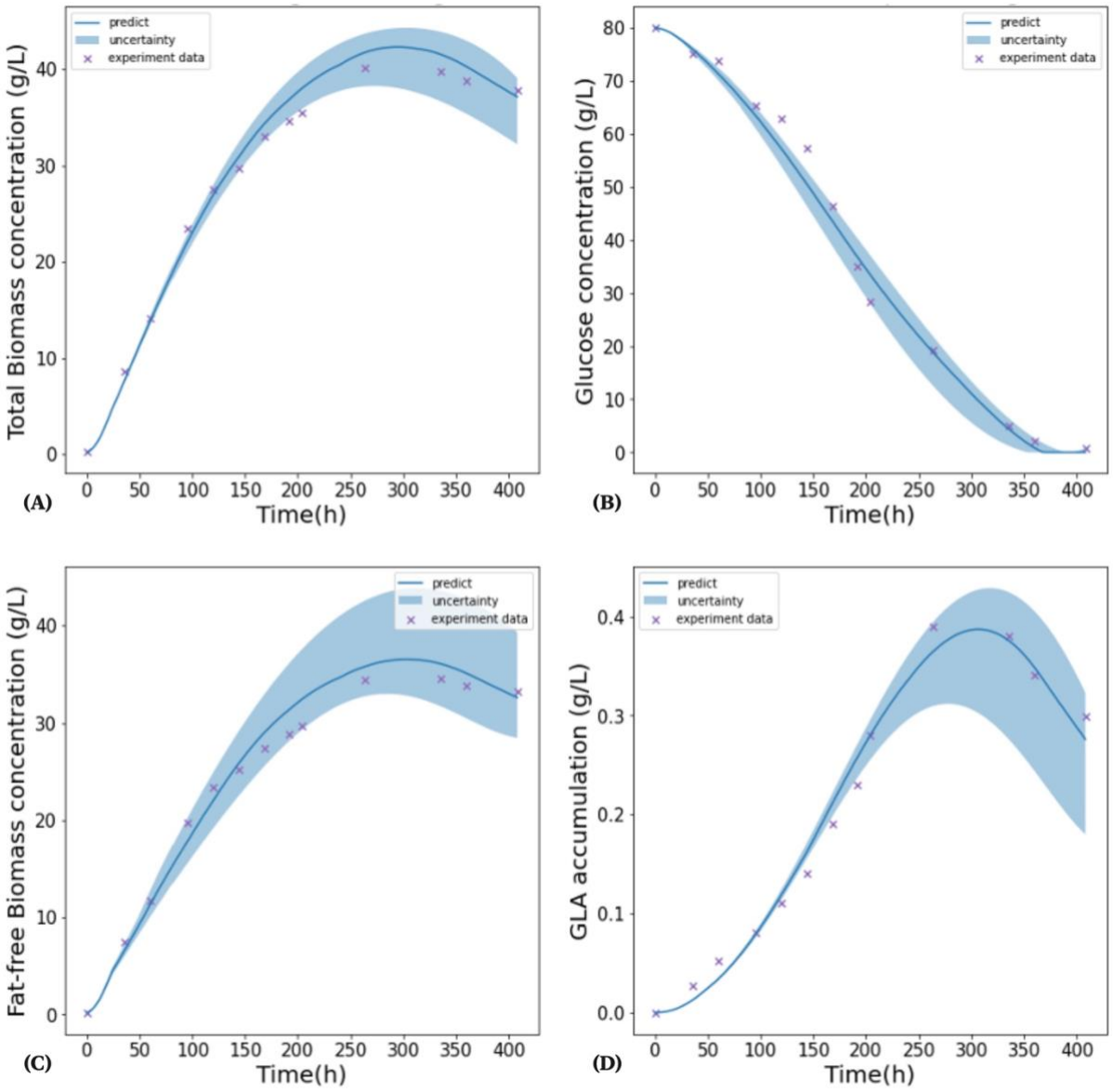


Figure E.6: Simulation result of Hybrid Model 3 for total biomass concentration (A), glucose consumption (B), fat-free biomass concentration (C) and GLA accumulation (D) at 37℃.

Table S1: Estimated kinetic model parameters values with 1 standard deviation at different temperatures for the unstructured kinetic model.

T (℃)	14	28	37
μ_m (h^{-1})	0.12 ± 0.0018	0.16 ± 0.32	0.22 ± 0.092
K_{co} ($g\ L^{-1}$)	41.35 ± 0.047	49.59 ± 129.35	55.00 ± 41.73
Y_{C0} ($g\ g^{-1}$)	1.42 ± 0.038	1.31 ± 1.64	0.59 ± 1.15
K_0 ($g\ g^{-1}$)	0.70 ± 0.19	0.78 ± 0.28	0.90 ± 0.38
k_m ($g\ g^{-1}$)	$1.16 \times 10^{-2} \pm 0.015$	$3.09 \times 10^{-3} \pm 0.012$	$1.17 \times 10^{-3} \pm 0.38$
k_d($g\ g^{-1}h^{-1}$)	3.52×10^{-3}	4.80×10^{-3}	7.27×10^{-3}
K_p ($g\ L^{-1}$)	25.01	34.28	37.89
μ_d (h^{-1})	$1.70 \times 10^{-3} \pm 0.00026$	$1.70 \times 10^{-3} \pm 0.0036$	$1.70 \times 10^{-3} \pm 0.0055$
m ($g\ g^{-1}h^{-1}$)	$4.98 \times 10^{-3} \pm 0.00041$	$4.98 \times 10^{-3} \pm 0.0080$	$4.98 \times 10^{-3} \pm 0.0060$
k_n ($g\ g^{-1}h^{-1}$)	$1.38 \times 10^{-4} \pm 6.41 \times 10^{-5}$	$1.38 \times 10^{-4} \pm 7.23 \times 10^{-5}$	$1.38 \times 10^{-4} \pm 5.84 \times 10^{-5}$