Biodiesel from Microalgae: Application of Dynamic Optimisation Techniques to a Microalgae Bioreactor

Samuel Dale, Seyoung Lee, Dominik Klotz and Luke Whittington

Department of Chemical Engineering, Imperial College London, U.K.

Abstract: Microalgae are a promising source of fatty acids for biodiesels, showing huge advantages over other production methods. In order to construct industrially feasible bioreactors for this purpose, it is vital to ensure that these systems are accurately modelled, accounting for their metabolically-governed reaction networks. Del-Rio Chanona et al. [1] formulated a dynamic model of cylindrical microalgae bioreactors, backing this up with experimental work. This project builds on this research, re-attempting the parameter estimation of the dynamic model presented in this paper, with improvements in parameter estimations over the source material demonstrated quantitatively. Sensitivity analysis using Sobol's method [2] demonstrated biomass growth rate and nitrogen consumption rate to be the most important parameters to control in the process. Dynamic optimisation is used to set process conditions, suggesting maximising light intensity and reducing batch duration to improve productivity. Finally, multiple methods are used to optimise experimental design for the most efficient gathering of information.

1 Introduction

In light of the global drive towards sustainability, the development of renewable, carbon-neutral fuels is of high priority. For compression-ignition internal combustion engines, biodiesel is an increasingly important alternative to fossil fuels. Fatty acid methyl esters (FAME), derived from organic oils via transesterification, may be used directly in unmodified diesel engines. Consequently, these are promising candidates for the partial replacement of petrodiesel, which they outperform in several important metrics such as a higher combustion number[3], biodegradability[4] and, lower NO_x emissions (a crucial concern for urban areas) [5–7].

A wide variety of potential oil feedstocks and their effect on the physical and chemical properties of the resulting biodiesel have been investigated [8]. In the US, rapeseed and soybean oil feedstocks account for more than half of current biodiesel production [9]. The low energetic efficiency of traditional feedstocks and their competition with foodstuffs for agricultural land have, however, put into focus more unconventional lipid sources – microalgae in particular. Owing to their highly efficient lipid anabolisms, specific algae are capable of producing tri-glycerides (TAG), the FAME precursor, at space productivities over 20 times higher than comparable rapeseed cultivations. Additionally, under suitable biological conditions, algae cells may achieve lipid contents in excess of 70wt.%, particularly in nitrate-deficient systems.

The underlying dynamics of biolipid formation are highly non-linear and come with numerous complex operational trade-offs. For example, in nitrate-deficient environments the synthesis of biolipids is favoured at the expense of decreased overall biomass generation. Photosynthetic activity, and hence both biolipid and biomass formation, strongly depends on the incident light radiation – high biomass concentration in turn lowers the effective axial light intensity due to increased scattering. Considering these complexities, it comes as no surprise that the first scale-up studies of microalgae-enabled biolipid production have performed significantly below lab-scale levels (60% decreases in lipid content achieved) and are, in their current state, not economically feasible. The successful operation of microalgae-based bioreactors at high FAME productivity will hence require implementation of sophisticated process control and monitoring techniques.

Del-Rio Chanona et al.'s [1] extensive research on the topic illustrated the strong basis for the application of real-time optimisation and model predictive control to the biolipid-bioreactor setting and has been used as the primary reference throughout this paper. This report outlines and implements the dynamic model of FAME biosynthesis proposed by Del-Rio Chanona et al. and aims to find an optimal dynamic control strategy for the microalgae culture, aiming to replicate their results. Additional model parameter estimation using experimental data will be conducted and, after appropriate sensitivity analysis and uncertainty quantification, used to obtain a second updated optimal control strategy under detailed discussion of the multi-objective trade-offs. Lastly, an optimal set of future experiments will be proposed using design of experiments (DoE) approaches, providing guidance for the best next steps in practical research enabling maximum data utilisation. *gPROMS* [10] was used extensively throughout this project.

2 Dynamic Model of FAME Bioproduction

2.1 Methodology

Del-Rio Chanona et al.outlined the key variables and relationships affecting biolipid production and, based on these, formulated an accurate dynamical model of the microalgae lipid synthesis process in a bioreactor. The model proposed by Del-Rio Chanona et al. was used in this report, and is now revisited briefly. The primary growth of the biomass concen-

Table 1: Overview of the 5 state variables

Equation	Units	Description
$\frac{\mathrm{d}}{\mathrm{d}t}X = (\mu_0 - \mu_d)X$	$\mathrm{g\,l^{-1}\,h^{-1}}$	Biomass Growth
$\frac{\mathrm{d}}{\mathrm{d}t}N = -\mu_N \frac{N}{N+k_N}X$	$\mathrm{mg\,l^{-1}\,h^{-1}}$	Nitrate Consumption
$\frac{\mathrm{d}}{\mathrm{d}t}q = \mu_N \frac{N}{N+k_N} - \mu_M(I)\left(1 - \frac{k_q}{q}\right)q$	$\mathrm{mg\,l^{-1}\,h^{-1}}$	Intracellular Nitrogen
$\frac{\mathrm{d}}{\mathrm{d}t}f = \left(1 - \frac{k_q}{q}\right)\left(\mu_M(i)(\theta q - \varepsilon f) - \gamma\mu_N \frac{N}{N+k_N} - q\right)$	$\mathrm{mg\,l^{-1}\,h^{-1}}$	FAME Synthesis
$Y(II) = \frac{\exp(\tau q)}{\exp(\tau q) + \delta} + \phi$	$\mathrm{mg\,l^{-1}\,h^{-1}}$	Photosynthetic Activity

tration X is governed by nitrate concentration N, incident light intensity I and, in some cases, some intrinsic biomass decay μ_d – a simple first order rate law expression with a decay coefficient μ_d and a growth coefficient μ_0, technically $\mu_0(I, N)$, is used. Nitrate consumption was described using a Monod expression with an uptake parameter μ_N and half-velocity coefficient K_N. Synthesis of lipids within the biomass is mediated by the intracellular nitrogen content (the nitrogen quota q) and the sufficiency of incident light intensity – nitrogen quota evolution was hence modelled using the balance of nitrate consumption to cellular nitrogen uptake. Three differential mass balances for biomass content X, nitrate concentration N and nitrogen quota q are hence used to describe the outlined dynamics; application of a modification capturing the transesterification kinetics to the lipid mass balance as proposed by Gnansounou & Raman [11] allows formulation of the fourth mass balance directly in terms of the amount of FAME produced (f). In order to include a real–time measurable variable in the model description for both future process feedback and current model tuning purposes, the photosynthetic activity, measured via the chlorophyll fluorescence $Y(II)$, is used as a further state variable defined via an algebraic relation – all equations of which are presented in table 1.

The effect of local light intensity on the biomass production was modelled using the well known Aiba model with two parameters k_i and k_S, with a modified Lambert-Beer law capturing the effects of adsorption (via the biomass adsorption coefficient α) and bubble scattering (via scattering coefficient β) in the radial reactor direction. To avoid the introduction of spatial model dependencies, a 10 step trapezoidal approximation, which has been previously proven to be highly accurate, was applied. The relation between the maximum biomass growth rate and the real biomass growth rate was described in terms of the ratio between current and minimum (k_q) nitrogen quota. These additional model equations may be found in table 2, and a systematic summary of the full dynamic model may be found in the appendix, section A.

2.2 Dynamic Modelling Result

This dynamic model was then implemented in *gPROMS*, using model parameter values adapted from Del-Rio Chanona et al.'s findings – with a few minor corrections regarding typographical errors in the paper. Simulations were then performed for the initial and operating conditions of experiment 1 and 2 respectively of Del-Rio Chanona et al.'s work, the results of which are shown in figure 1.

Generally, the results show satisfactory agreement with the experimental data points. The shape of the trends corresponds well to the experimental data, and the range of experimental measurement uncertainty encompasses the simulation results at the majority of

Table 2: Light intensity approximation & relation between maximum and real biomass growth

Equation	Units	Description
$I(z) = I_0 \exp\left(-z(\alpha X + \beta)\right)$	$\mu\mathrm{mol\,m^{-2}\,s^{-1}}$	Light intensity
$\mu_m(I) = \frac{\mu_{M0}}{20}\sum_{i=1}^{9}\left(\frac{I_{i=0}}{I_{i=0} + k_s + \frac{I_{i=0}^2}{k_i}} + \frac{I_{z=\frac{iL}{10}}}{I_{z=\frac{iL}{10}} + k_s + \frac{I_{z=\frac{iL}{10}}^2}{k_i}} + \frac{I_{z=10}}{I_{z=10} + k_s + \frac{I_{z=10}^2}{k_i}}\right)$	$\mathrm{h^{-1}}$	Trapezoidal approx.
$\mu_0 = \mu_M\left(1 - \frac{k_q}{q}\right)$	$\mathrm{h^{-1}}$	max to real growth

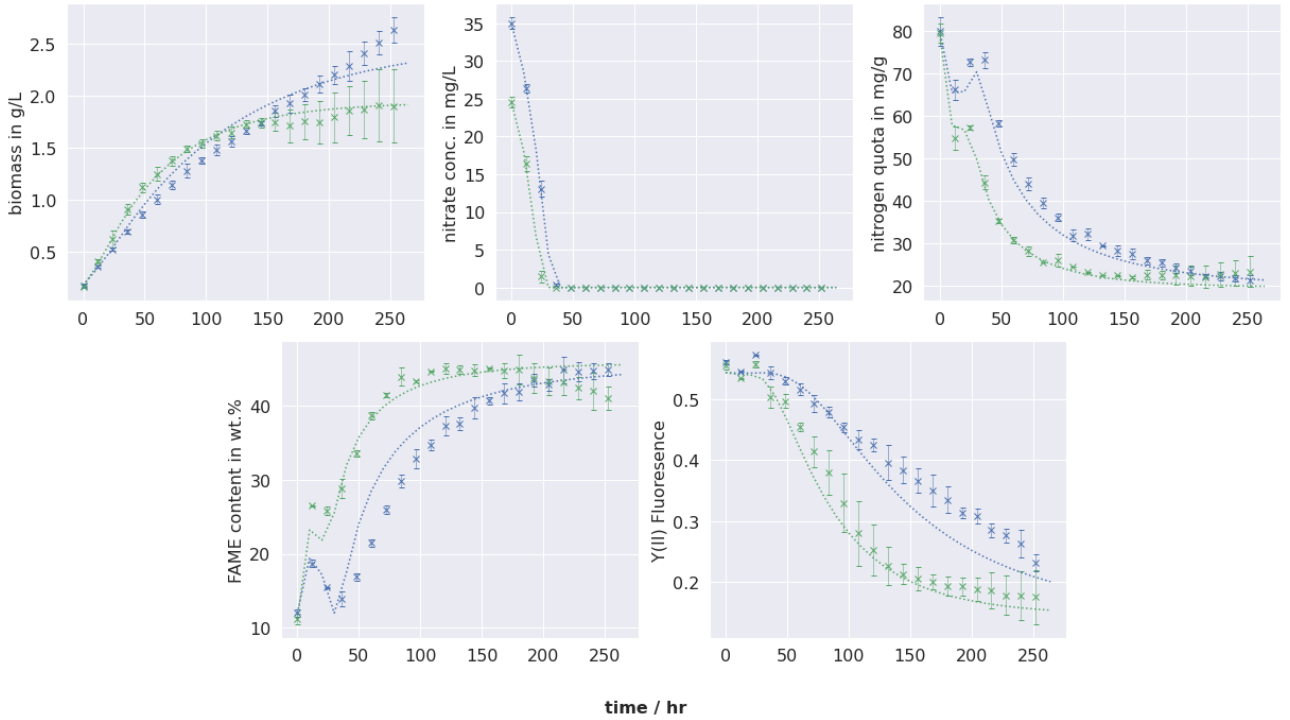


Figure 1: Simulation results vs the experimental data from Del-Rio Chanona et al. for the 5 state variables – experiment 1 in blue, experiment 2 in green

points. Simple visual inspection showed that a better fit to the data obtained from experiment 2. In order to quantitatively assess the fit of the predictions, the widely-used root relative mean square error was redefined as mean root relative square error, accounting for the large variations in the input variables:

$$\text{MRRSE} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{y_i}\sqrt{(y_i - \hat{y}_i)^2} \tag{1}$$

The nitrate concentration profile seemed to be most accurately described by the model as it shows perfect fit with the experimental data at a MRRSE value of 1.1% $\pm$ 1.3. This is due to the fact that both biomass concentration and nitrogen quota values mostly fit well in the range of standard deviation proposed by the experiment, as change in nitrate concentration is dependent on change in biomass and nitrogen quota. On the other hand, the result of FAME yield and fluorescence modelling deviated more from the experimental values, with MRRSE's of 8.1% and 8.9% respectively and a noticeable trend of error propagation with increasing time. Since the calculations for these two variables involve many parameters, it was deemed that the parameter estimation could be improved to fit our purpose of analysing experiment 1 and 2 only, while del-Rio Chanona, et al. involved more experiments to provide more holistic estimation towards parameters. Moreover, it is possible that since IPOPT was leveraged to estimate parameters by solving nonlinear programming (NLP) based on line-search filter barrier method, which only guarantees global convergence not global optimality [12], potential for further optimisation lies within.

3 Parameter Estimation

3.1 Methodology

From the opportunity spotted to ameliorate the accuracy of modelling via improving the accuracy of values of parameters, further parameter estimation was carried out in *gPROMS* based on the experimental dataset of del-Rio Chanona, et al. The main aim of parameter estimation is to find the set of parameters and operating conditions such that the prediction error, which is described by the log-likelihood estimator in *gPROMS*, is minimised.

Table 3: Equations of available Variance Models

Variance Model	Constitutive Relation
Constant Model	$\sigma = \omega$
Relative Model	$\sigma = \omega z$
Heteroscedastic Model	$\sigma = \omega \lvert z \rvert^{\gamma}$
Linear Model	$\lvert \alpha z + \beta \rvert$

To perform parameter estimation, an appropriate variance model, which is a function that describes the relationship between the value of the variable and its standard deviation, had to be estimated. There are four different variance models that *gPROMS* accepts: constant, relative, heteroscedastic and linear. Mathematical description of the models is summarised in Table 3. The mean squared error (MSE) between the prediction from each model and the actual value of standard deviation was calculated using *SciPy* library

in *Python 3.9* and the model minimising MSE was chosen, with few exceptions. The results of the parameter estimation were analysed both for their validity, associated uncertainty and sensitivity. The generalised improvement of the predictive capabilities was assessed via comparison of the pre-estimation and post-estimation relative root square error between experimental data and simulated values at all available timesteps and various statistical tests encompassing goodness-of-fit, bias and lack-of-fit tests. While this allowed evaluation of model performance on the known data, uncertainty quantification and sensitivity analysis needed to be conducted to ensure the model was not only able to reproduce the experimental results but also to provide an accurate overall description of the general physical system, capable of extrapolating from the experimental basis.

The inherent uncertainty in the estimated parameters and its impact on the model output was quantified using an ensemble simulation approach, making use of the `EnsembleProblem` functionality module of the `DifferentialEquations.jl` library for the Julia programming language [13, 14]. The parameters were assumed to be normally distributed around the mean $\tilde{x}_i$ and variance σ_i^2 found in the estimation procedure, and were hence independently sampled from distributions $x \sim \mathcal{N}(\tilde{x}_i, \sigma_i^2)$. Parameters for which no variance information was obtained (primarily those whose mean converged to a given bound) were simply set as having a variance of zero, due to lack of information. From each parameter sample, an ODE problem was defined and then numerically integrated, resulting in 500 individual trajectories of the underlying system. At each point in time the (0.05, 0.95) quantile bounds were then computed and taken as a suitable confidence interval around the mean of the system's state variables' evolution.

Lastly, in addition to the results of the ensemble simulation, further detailed sensitivity analysis was conducted. Variance-based sensitivity analysis following Sobol's method [2] was deemed a more suitable approach than simple one-at-a-time (OAT) sensitivity analysis for the complexities of the dynamic system considered. Variance-based sensitivity analysis aims to find a decomposition of the total output variance Var($\boldsymbol{Y}$) in terms of individual variance sensitivities to each input parameter ("first order interactions") and cross–parameter variation effects ("higher order interactions"):

$$\mathrm{Var}(\boldsymbol{Y}) = \sum_{i=1}^{d} V_i + \sum_{i<j}^{d} V_{ij} + ... \tag{2}$$

The ratio between a an input variance parameter V_i and the corresponding total output variance Var($\boldsymbol{Y}$) is referred to as the sensitivity index S_i of the input with respect to the output – or, in case of higher order indices, the sensitivity of the output to input parameter interactions. These sensitivity indices were computed via Julia's global sensitivity analysis capabilities, using a Sobol-sampling method based on a quasi-Monte Carlo generated sampling design matrix with size N=40.000, and parameter ranges corresponding to $\tilde{x}_i \pm \sigma_i$.

3.2 Variance Model Result

The functional parameters for each proposed model that provide the lowest value of MSE were calculated and the results are summarised in Table 4. Overall, both heteroscedastic and linear models showed superior performances over other models. This was expected from the visual examination of error bars in Figure 1, as the standard error showed tendency to amplify with the progression of the experiment. The relative model was not chosen for any of the variables, since it was constantly outperformed by the linear model which is similar in essence but has an adjustment factor of β.

For the variable model of biomass concentration (X), although both the heteroscedastic and linear models showed practically the same levels of MSE, the heteroscedastic model was chosen because it provided lower value of log likelihood function of 74.06 compared to the 81.11 of the linear model, signalling more accurate prediction of parameter estimation. This is suspected to be a consequence of the negative value of β, which has a tendency to cancel out the ωz term at certain range of values, providing impractically small estimation of variance. Furthermore, for the variance model of fluorescence ($Y(II)$), the constant model was chosen despite the fact that its MSE is higher than those of heteroscedastic and linear models. This is because the negative value of γ in the

Table 4: Variance Modelling Results (Highlighted are the chosen model)

	Constant Model		**Relative Model**		**Heteroscedastic Model**			**Linear Model**		
	ω	**MSE**	ω	**MSE**	ω	γ	**MSE**	α	β	**MSE**
X	8.95e-2	1.48e-2	6.12e-2	1.13e-2	**5.48e-2**	**1.18**	**1.12e-2**	7.32e-2	-2.11e-2	1.12e-2
N	1.12e-1	1.55e-1	2.99e-2	5.96e-2	**5.19e-1**	**1.35e-1**	**1.65e-2**	2.85e-2	3.57e-2	5.74e-2
Y(II)	**1.85e-2**	**3.31e-4**	3.82e-2	5.81e-4	7.20e-3	-8.24e-1	2.43e-4	-5.63e-2	3.89e-2	2.18e-4
q	1.32	1.73	3.34e-2	1.82	2.74e-1	4.49e-1	1.60	**1.57e-2**	**7.71e-1**	**1.58**
f	9.43	6.63e+1	2.57e-2	5.74e+1	1.08e-1	7.61e-1	5.70e+1	**2.06e-2**	**1.97**	**5.68e+1**

Table 5: Parameter Estimation Results - full significant figure values found in Appendix B. Note: † indicating parameter values having converged to lower, § upper bound

Final	α	β	δ	ε	γ	k_i	k_N	k_q	k_s	μ_d	μ_M	μ_N	ϕ	τ	θ
$\tilde{x}_i$	300§	0†	12.86	2e-4†	7.29	250	2§	20.37	200§	0†	0.78	2.64	-0.46	0.15	6.51
σ	-	-	2.62	-	0.17	80.30	-	0.26	-	-	1.6e-2	2.7e-2	5.9e-3	9.1e-3	0.10
Initial	196.4	0	9.90	1e-3	7.53	100	0.80	19.60	91.20	0	0.36	2.69	-0.46	0.14	6.69

heteroscedastic model is unacceptable in *gPROMS*, due to its tendency to amplify the variance exponentially at small values. In addition, the negative value of α also causes unreasonably small variance at certain values like the case of negative β. This is confirmed by smaller log likelihood objective functions, with the constant model's 74.06 compared to linear model's 90.24.

3.3 Parameter Estimation Result

The updated parameter estimates and their respective initial values under the outlined parameter estimation scheme are presented in table 5.

As can be seen, several parameter values have changed relatively significantly and are still associated with high uncertainty, manifesting in a high standard deviation. The parameters α, k_N, k_s have converged to their upper bounds and hence require further experimental data and repetition of the estimation procedure on a larger parameter range. On the contrary, the two parameters describing the biomass decay μ_d and the bubble scattering β were estimated to be zero, and the underlying effects can hence be deemed physically irrelevant over the considered time and condition range.

In order to further confirm the accuracy of the parameters estimated, goodness of fit test (chi-squared test) was conducted and the result is summarised in table 6. In general, the model with the new parameter estimation showed better performance in experiment 2 than in experiment 1. For experiment 2, all χ^2 values except the one for the FAME yield (f) were smaller than their critical values, suggesting the deviation of predicted values from the experimental values is within the tolerance level of the confidence interval, therefore a good fit. For experiment 1, although nitrogen quota (q) is the only variable that passed the test, other variables except for f provided χ^2 value close to the critical value. This is clearly seen in the χ^2 values of both experiments combined. The test proved good fit for all variables except f. The potential reason for inaccuracy in FAME yield estimation was suspected to be derived from the fact that the equation for f involves the most number of the parameters, hence the combined effect of inaccuracies was amplified.

Similar results obtained from the lack of fit test further validate the result of goodness of fit test result, which is shown in Table 7. The F value in the lack of fit test was greater for the case of N and f in both experiments and X in experiment 1, indicating that null hypothesis of proposed model fitting well is rejected. Furthermore, from the observation of t-value given in Table 12 in Appendix B, one reason behind the underperformance of the model could be inaccurate prediction of k_i, the light inhibition term, because its t-value is smaller than the reference t-value, signalling that the data used for modelling may be insufficient.

To complement the information gained from statistical tests with data both possible to visualise and physically grasp, the simulation of experiment 1 from Del-Rio Chanona et al. was then repeated using the obtained updated parameter estimates. The previously discussed RRSE was now first considered over the entire evolution of the simulation instead of averaging. Figure 2 illustrates the RRSE for experiment 1 us-

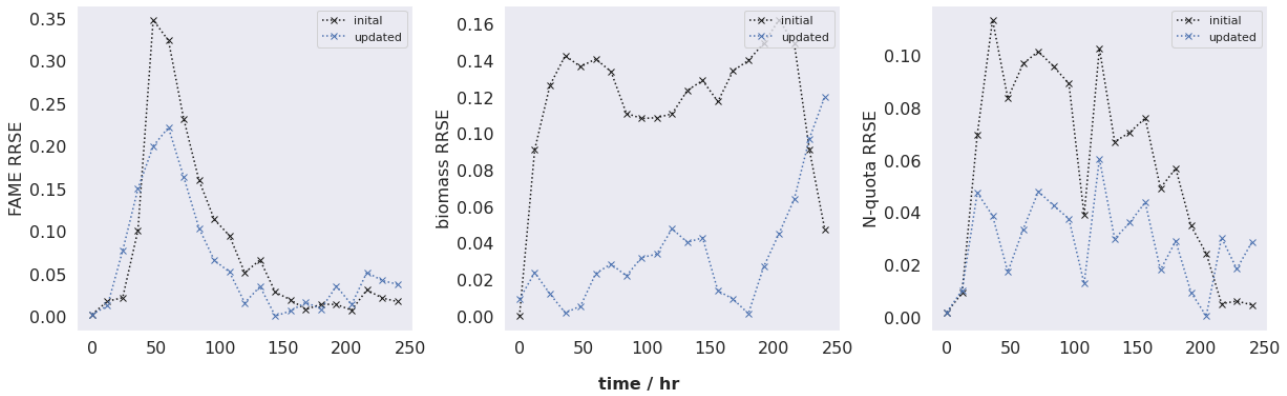


Figure 2: RRSE for Del-Rio Chanona et al. Experiment 1 - FAME, biomass and N-quota: initial parameters in black, and updated parameters in blue

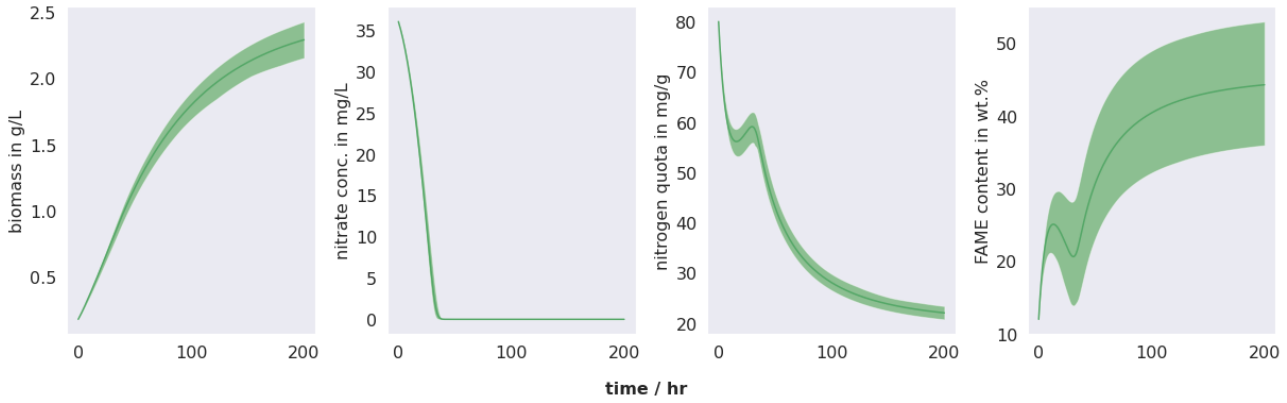


Figure 3: mean and (0.05, 0.95) quantiles for ensemble simulation of N=500 trajectories and $\mathcal{N}$ distributed model parameters

Table 6: Goodness of Fit Test Result

Variable	Exp 1		Exp 2		Exp 1 & 2	
	χ^2	χ^2_{crit}	χ^2	χ^2_{crit}	χ^2	χ^2_{crit}
N	15.6	14.1	6.7	14.1	22.3	42.6
X	21.4	14.1	4.6	14.1	26.0	42.6
$Y(II)$	22.2	14.1	12.0	14.1	34.2	42.6
f	132	14.1	67.2	14.1	199	42.6
q	11.5	14.1	27.4	14.1	39.0	42.6
Total	203	119	118	119	321	239

Table 7: Lack of Fit Test Result

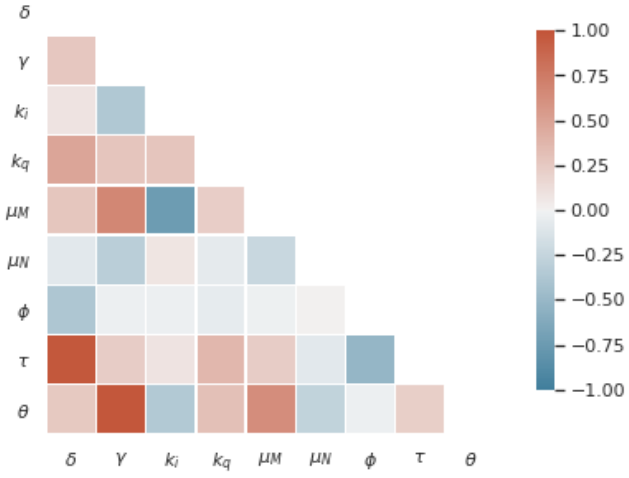
Variable	Exp 1		Exp 2		Exp 1 & 2	
	F	F_{crit}	F	F_{crit}	F	F_{crit}
N	8.73	2.46	3.05	2.46	2.54	1.73
X	4.70	2.46	-0.46	2.46	0.67	1.73
$Y(II)$	2.17	2.46	0.72	2.46	0.18	1.73
f	10.90	2.46	8.27	2.46	4.03	1.73
q	1.05	2.46	2.70	2.46	0.35	1.73
Total	3.30	1.38	2.38	1.38	2.51	1.25

ing the initial parameter estimates in black and as obtained from the updated parameters in blue. Qualitative assessment immediately shows the improvement achieved by the updated model; in quantitative terms, the average RRSE of all three curves decreased by 48% – implying a significant increase in the models predictive capabilities.

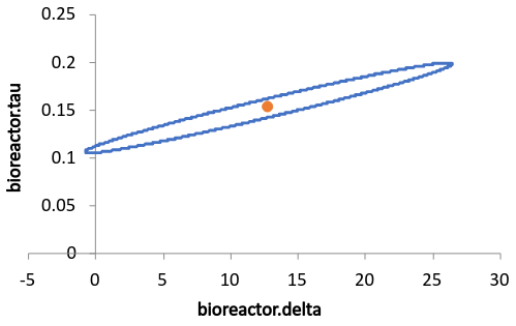
The resulting confidence intervals around the mean of the ensemble simulation is presented in figure 3. Clearly, the output variation in the predicted biomass and nitrate concentrations as well as the nitrogen quota are remarkably consistent over the considered one-σ parameter range. The FAME cell weight content is the the noisiest of the key variables, which comes at no surprise when comparing the complexity of the underlying equations and the number of parameters directly appearing.

3.4 Sensitivity Analysis

While Sobol's method may be formulated for a multivariate output, consideration of the two aggregated KPIs, mean and maximum FAME productivity, allowed good analysis of the sensitivity indices with applicability to the entire problem without having to implement the much more complex multivariate analysis. First-order indices, the sensitivity indices wrt. to individual parameter variations, are given in figure 5a. Sobol's

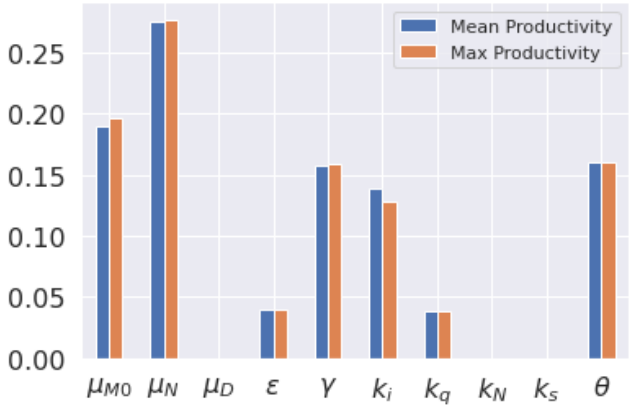


(a) Correlation Matrix of the estimated parameters

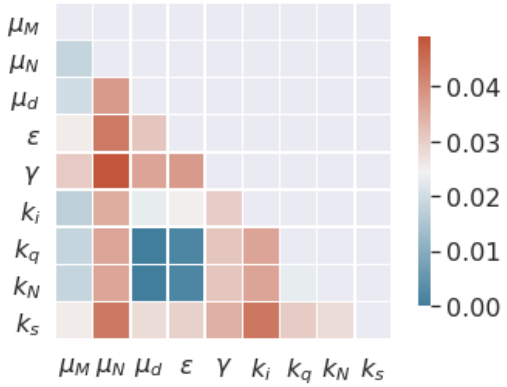


(b) Illustrated elongated confidence ellipsoid of Y(II) parameter pair δ and τ

Figure 4: Parameter estimation plots



(a) First-order sensitivity coefficients



(b) Second-order sensitivity coefficient matrix for the mean productivity

Figure 5: Sensitivity analysis plots

analysis puts forth the biomass growth rate μ_M and nitrogen consumption rate μ_N as the parameters determining most of the output variance. The pair of γ and θ discussed previously are, likely due to their high correlation, deemed to be near-equally contributing to the overall output variance. The parameters representing specific biological properties such as the minimum nitrogen quota k_q and the half-velocity coefficient of N-uptake showed relatively low sensitivity, accounting for 4% and near 0% of the output variance. A sensitive parameter however is again the light saturation parameter k_i – the parameter estimation struggled to lower the standard deviation of k_i, which is reflected in the large variation range leading to a higher first-order Sobol index.

The second-order coefficients, a matrix representation of which is given in figure 5b, showed some relations that were to be expected from the underlying mathematics (f.e. the high γ-μ_N index, as these parameters are directly multiplied in the FAME balance) but also some surprising ones: for example, the comparatively high sensitivity index of the k_s–μ_N pair is not explicable via a direct mathematical interaction and therefore of note. It should also be noted that μ_N is in general a sensitive parameter for higher order interactions – this might likely be related to the fact that μ_N is one of the two parameters of the model coupled to all four differential equations and its higher estimation uncertainty than the other such parameter, the biomass growth rate μ_M.

4 Model Optimisation

4.1 Methodology

Building on the previously-formulated dynamic model, optimisation of experimental operating conditions was carried out. In order to construct an objective function, some efficiency measures, the productivities of FAME and microalgal biomass, were formulated. Productivity here is defined as the amounts produced per unit of batch time and the mathematical definition can be found in eq. 3 and eq. 4. The biomass concentration was multiplied by the reactor volume (V_r) to convert it from concentration to mass. Furthermore, 6 hours was added to the total time of operation as changeover between cycles takes a fixed time of 6 hours.

$$p_{FAME} = \frac{V_r(f(t)X(t) - f(0)X(0))}{t+6} \tag{3}$$

$$p_{BIOMASS} = \frac{V_r(X(t) - X(0))}{t+6} \tag{4}$$

In order to optimise the productivity measures, the effects of initial nitrate concentration (N_0), total operating time (t_{op}) and incident light intensities (I_0) were investigated. Some other important initial conditions, such as biomass concentration, FAME yield and nitrogen quota, were not investigated as they were determined by the pre-culture, therefore were assumed constant. Additionally, different control strategies of incident light intensity, that include varying number of control intervals and switching between piecewise constant and piecewise linear model, were also investigated to provide best control strategy.

Furthermore, multi-objective optimisation was performed using ϵ-constraint method by adding an additional inequality constraint of minimum biomass productivity, while maximising the FAME productivity.

4.2 Single Objective Optimisation Result

FAME productivity and biomass productivity were optimised using parameters obtained from both del-Rio Chanona, et al. and Section 3.3, and the results can be seen in Table 8 and Table 9, respectively. The result indicate that FAME productivity of 42.6 - 48.8 $mg\,h^{-1}$ can be obtained through optimisation, which is 2.29 - 2.63 times the end FAME productivity obtained from initial experiment 1. Similarly, biomass productivity of 11.73 - 12.84 $mg\,h^{-1}$ can be obtained through optimisation, which is 2.96 - 3.24 times the end biomass productivty of initial experiment 1.

In general, for both productivities, variation in control strategy of incident light intensities did not seem to have noticeable influence to the productivity values. This is mainly due to the fact that the optimal controlling strategy calculated by *gPROMS* turned out to be

Table 8: FAME Productivity ($mg\,h^{-1}$) Optimisation Result

Method	Horizon Division	del-Rio Chanona, et al.			From section 3.3		
		p_{FAME}	N_0	t_{op}	p_{FAME}	N_0	t_{op}
Piecewise-Constant	1 (constant)	4.25	10.6	34.0	4.88	0.0	12.8
	10	4.26	10.4	33.7	4.88	0.0	12.8
	20	4.26	10.4	33.6	4.88	0.0	12.8
	50	4.26	10.4	33.7	4.88	0.0	12.8
Piecewise-Linear	5	4.26	10.4	33.6	4.88	0.0	12.8
	10	4.26	10.4	33.6	4.88	0.0	12.8
	20	4.26	10.4	33.6	4.88	0.0	12.8
	50	4.26	10.4	33.7	4.88	0.0	12.8

Table 9: Biomass Productivity ($mg\,h^{-1}$) Optimisation Result

Method	Horizon Division	del-Rio Chanona, et al.			From section 3.3		
		$p_{BIOMASS}$	N_0	t_{op}	$p_{BIOMASS}$	N_0	t_{op}
Piecewise-Constant	1 (constant)	11.42	66.1	59.3	12.41	62.0	52.3
	10	11.74	80.0	72.4	12.84	80.0	65.5
	20	11.74	80.0	72.4	12.84	80.0	65.4
	50	11.74	80.0	72.4	12.84	80.0	65.4
Piecewise-Linear	5	11.74	80.0	72.4	12.84	80.0	65.9
	10	11.74	80.0	72.4	12.84	80.0	65.4
	20	11.74	80.0	72.4	12.84	80.0	65.4
	50	11.74	80.0	72.4	12.84	80.0	65.4

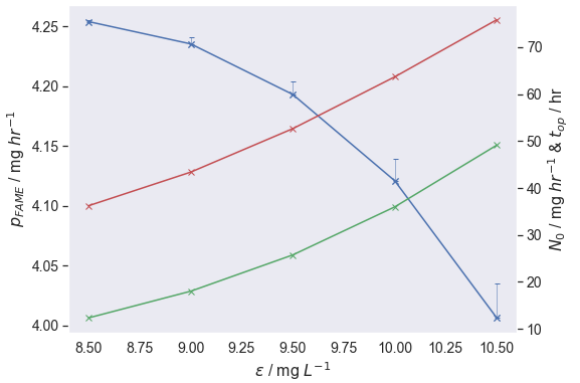
keeping the light intensity level at the upper bound of 300 $\mu mol\,m^{-2}\,s^{-1}$ for most of the time. Nevertheless, opportunity for maximising energy efficiency was observed in time-varying control, as same level of productivity can be achieved by initially providing lower light intensity and gradually increasing to upper bound. This result can be seen in Figure 7 in Appendix.

Noticeable trade between achieving optimal FAME productivity and optimal biomass productivity was observed. In order to achieve high FAME productivity, low concentration of nitrate of 0.0 - 10.6 $mg\,L^{-1}$ in the medium and short batch duration of 12.8 - 34.0 hours is recommended, whereas for biomass productivity very high nitrate concentration near the upper bound of 80 $mg\,L^{-1}$ and longer duration of 52.3 - 72.4 hours is recommended. Overall, the batch duration is recommended to be decreased from the initial duration of 264 hours to obtain higher productivity, because the media runs out of nitrate after around 40 hours time and nitrogen quota starts to drastically decrease, thereby not able to sustain the high nitrogen consumption from high light intensities. Moreover, careful decision over initial nitrate concentration is required as high nitrate concentration can promote the growth of biomass but it might not be efficient in terms of FAME production.

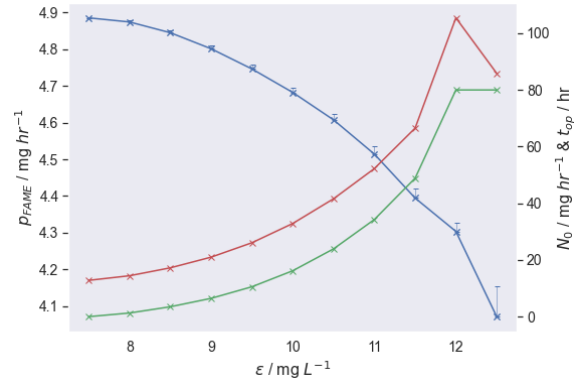
Significant difference between the result of using different sets of parameters was also observed. With the new parameter values, the initial nitrate concentration (N_0) was recommended to be lower even kept at the lower bound of 0 $mg\,h^{-1}$ for FAME productivity. One possible reason behind it could be the increase of half-velocity coefficient (k_N) that increased by 2.5 times after parameter estimation. Since $\frac{N}{N+k_N}$ term was included in many differential equations, effect of increased k_N, would have caused the presence of nitrate concentration to be faded. The total operating time was also recommended to be kept lower than the prediction of the initial result. This is because since the nitrate gets depleted earlier in the media due to lower initial concentration, the nitrogen quota starts decreasing earlier as well. Therefore, there is less benefit of running longer batch times according to the new parameters.

4.3 Multi-Objective Optimisation Result

Multi-objective optimisation was conducted and the effect of setting different ϵ-constraint was investigated over the range of feasibility. Increase in both N_0 and t_{op} was observed with tightening of biomass concentration constraint, but the FAME productivity tend to decrease with this change. Moreover, the Lagrange multiplier, which indicates the change in objective function upon relaxing the constraint, increases with ϵ suggesting that the FAME productivity becomes exponentially inefficient with higher biomass productivity. This can be seen in the error bars of Figure 6 and they indicate the increase in p_{FAME} from relaxation of ϵ by 0.1. The analysis suggests that in order to operate robustly with potential disturbances during batch process, it is recommended to operate around the middle value of $X = 9.5 mg\,l^{-1}$.



(a) With initial parameters



(b) With revised parameters

Figure 6: Multi-Objective Optimisation plots - N_0 in green, t_{op} in red and p_{FAME} in blue, the upper error bar represents potential increase in p_{FAME} upon relaxation of ϵ-constraint

5 Design of Experiments (DoE)

5.1 Methodology

gPROMS' inbuilt optimal experiment design functionality was used to design a further experiment, improving the statistical quality of future parameter estimations. Taking the variance model results calculated in 3.2 and the optimal parameter estimates calculated in 3.3, the initial value of nitrate concentration N_0, incident light intensity I_0 and experiment duration D were optimised according to several criteria. The initial values of X, f and q are treated as properties of the algal culture set before experiments are run and are thus treated as fixed. For the parameters optimised, optimality criteria were based on the covariance matrix of their estimators $C(\theta)$. Three optimality criteria were investigated, based on the results of experiments 1 and 2:

1. **A-optimality**: minimisation of the average variance of the estimators
2. **D-optimality**: minimisation of the determinant of $C(0)$ (minimising confidence ellipsoid area)
3. **E-optimality**: minimisation of the maximum eigenvalue of $C(0)$ (minimising longer axis of confidence ellipsoid), treated as the base case here

5.2 Discussion & Results for DoE

Initial guesses and final values of time duration of the experiment D, light intensity I_0, and the initial nitrate concentration N_0 for E-optimal, A-optimal, and D-optimal experiment design can be found in table 10. E-optimality is considered here as the "base case".

The initial guess for D was close to the maximum duration of the experiment as the longer the duration, the more information available. The initial guess for I_0 was also set close to its upper bound as it would stimulate the responses for f and q, thus providing more information than is already available. For N_0, the initial guess was also set close to its upper limit, as lower concentrations had already been experimented with for experiment 1 and 2. Other initial guesses were used, but none provided better solutions. In all cases tested, optimal D and I_0 were both found to be at their upper limits, with the same rationale as setting initial guesses near these points explaining this. However, optimal N_0 values varied in each test case and were found to be significantly displaced from the upper bound of 80 mg l^{-1}. This is due to the interaction between the governing differential equations, resulting in a trade-off scenario where no one governing equation dominates. Furthermore, approach of the upper or lower bounds of N_0 prevents estimation of multiple parameters – in the case of evaluating nitrate consumption, high N_0 results in removal of dependence on k_n, whereas low N_0 allows estimation of only the ratio $\frac{\mu_N}{k_N}$ as opposed to their individual values.

In the base case, E-optimality, an optimal N_0 of 57.50 mg l^{-1} is produced, with a design criterion (objective function) value of 4512. This base case uses data from both of Del-Rio Chanona et al.'s experiments, and comparison is made between this and inclusion of only experiment 1 or 2 with the E-optimality criteria. If only experiment 1 is considered, the results are quite similar to the base case, with the value of the design criterion increasing somewhat, indicating the lesser quantity of information encompassed. There is a tiny decrease in N_0 compared to the base case. However, if only experiment 2 is considered, the value of the design criterion is much higher than for the base case while the nitrate concentration is significantly lower. It is unclear whether if these different dynamics correspond to a global minimum or if the solver may have gotten stuck at less desirable local minimum. Otherwise, it be a result of the experimental design's time-variant incident light intensity, unlike in the first experiment, and the effects of this on the quality of parameter estimation, given this intensity was considered as time-invariant (although an optimal I_0 at the upper bound was still produced).

In comparison between the results of different optimality conditions, A and E-optimality showed little difference, with optimal N_0 being within within 0.5% of each other. In contrast, D-optimality presented somewhat different results, with optimal N_0 of 74.99 mg l^{-1}, close to the initial value. D-optimality minimises confi-

Table 10: DoE results

Optimality condition	Experimental data	D (h) Initial	**Final**	N_0 (mg l^{-1}) Initial	**Final**	I_0 (µmol m^{-2} s^{-1}) Initial	**Final**	Design criterion value
E (base)	1 & 2	250	264	70	57.500	270	300	4511.95
E	1	250	264	70	58.4721	270	300	5121.86
E	2	250	264	70	44.5814	270	300	18,556.6
A	1 & 2	250	264	70	57.2456	270	300	301.862
D	1 & 2	250	264	70	74.9942	270	300	0.001743

dence ellipsoid areas – this can potentially lead to the prioritisation of uncertainty reduction for well-defined parameters over less well-determined parameters, as these ellipsoids will approach zero area if they have close to zero length in the short axis. This presents a substantially different approach to A and E-optimality conditions, which tend to favour general uncertainty reduction more. There is little to differentiate between the suitability of one over the other here, but either A or E-optimality likely present the more sensible choice of condition in this case.

6 Conclusion

Del-Rio Chanona et al.'s dynamic model was again demonstrated to be effective through its implementation here, capturing overall trends well, although not uniformly predicting exact numerical values of experimental data, due to its complex, nonlinear behaivour.

The parameter estimations performed showed reduced average RSSE values and so higher accuracy compared to del-Rio Chanona et al.'s source material. However, χ^2 tests indicated that some predictions, especially for FAME yield, were not satisfactory. FAME yield is dependent on the greatest number of parameters, amplifying uncertainties. The model's sensitivity to some parameters is high and so these effects are observed here. Furthermore, some parameters converged to their bounds, showing the experimental condition ranges were not varied enough to show the effects of these variables' variation.

Model optimisation demonstrated that, although varying light intensity control strategies showed near-negligible difference on FAME and biomass productivity compared to keeping light intensity I_0 at its maximum of 300 µmol m^{-2} s^{-1}, small opportunities for energy saving were identified in initially setting lower I_0 and ramping up to the maximum. A clear trade-off over initial nitrate concentration was found between optimising FAME and biomass productivity. In industrial bioreactors, reducing energy consumption could be of key economic concern, so formulating optimisations with objective functions balancing minimisation of energy consumption with maximisation of FAME production could be more important than simply maximising the production alone. The balance of these two variables would effectively be an economic potential (EP) function when including energy and FAME pricing, which is generally the key element to maximise in an industrial process. Further variables may also be included in more complex EP functions. However, the cost of control implementation for light intensity may outweigh the small energy savings produced by simply leaving intensity at its maximum.

For optimal experiment design, application of E and A-optimality conditions resulted in virtually the same recommendation, maximising light intensity and experimental duration to provide the most information, whilst satisfying the trade-off between the differential equation system by producing an intermediate initial nitrate concentration N_0. D-optimality resulted in significantly different results, with increased N_0 in comparison. D-optimality prioritises the greatest uncertainty reduction in well-defined parameters over reduction of all individual uncertainties – as multiple parameters were poorly-defined here, E or A-optimality are therefore suggested as the superior methods of experimental design. Inclusion of data from the redefined experiment would lead to decreased uncertainty in these parameters overall and should be carried out as a next stage in the parameter estimation process.

Overall, the application of further, more rigorous optimisation techniques could be utilised in the refinement of modelling of this system. *gPROMS* is limited to gradient-based optimisation techniques which do not give guaranteed convergence to global optima. Spatial branch-and-bound algorithms could be ideal for ensuring that global optima are produced. Although computing times would be increased, the potential impacts on increasing process productivity would likely outweigh this. Through this project, the intricacies of producing a completely-specified model of this complex process have been demonstrated.

References

1. Del Rio-Chanona, E. A. *et al.* Dynamic modeling of green algae cultivation in a photobioreactor for sustainable biodiesel production. *Biotechnology and Bioengineering* **115,** 359–370. ISSN: 10970290 (2018).
2. Sobol, I. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation* **55.** The Second IMACS Seminar on Monte Carlo Methods, 271–280. ISSN: 0378-4754. https://www.sciencedirect.com/science/article/pii/S0378475400002706 (2001).
3. Demirbas, A. Fuel properties and calculation of higher heating values of vegetable oils. English. *Fuel* **77,** 1117–1120 (1998).
4. Speidel, H. K. *et al.* Biodegradability of new engineered fuels compared to conventional petroleum fuels and alternative fuels in current use. eng. *Applied Biochemistry and Biotechnology* **84-86,** 879–897 (2000).
5. World Health Organisation. *Air pollution* www.who.int/sustainable-development/cities/health-risks/air-pollution/en/ (2019).
6. López, J. M. *et al.* Comparison of GHG emissions from diesel, biodiesel and natural gas refuse trucks of the City of Madrid. English. *Applied Energy* **5,** 610–615. https://www.infona.pl//resource/bwmeta1.element.elsevier-68e41cdd-3340-3241-8ba5-620cc3f760c6 (2009).
7. USEPA. A comprehensive analysis of biodiesel impacts on exhaust emissions. eng. *Draft Technical Report* (2002).
8. Singh, S. P. & Singh, D. Biodiesel production through the use of different sources and characterization of oils and their esters as the substitute of diesel: A review. en. *Renewable and Sustainable Energy Reviews* **14,** 200–216. https://ideas.repec.org/a/eee/rensus/v14y2010i1p200-216.html (2010).
9. Administration, U. E. I. *Monthly Biodiesel Production Report* tech. rep. (U.S. Department of Energy, 2019), 1–10. https://www.eia.gov/biofuels/biodiesel/production/biodiesel.pdf.
10. Process Systems Enterprise. *gPROMS* 2020. http://www.psenterprise.com/products/gproms.
11. Gnansounou, E. & Raman, J. K. Life cycle assessment of algae biodiesel and its co-products. *Applied energy* **161,** 300–308 (2016).
12. Wächter, A. & Biegler, L. On the Implementation of an Interior-Point Filter Line-Search Algorithm for Large-Scale Nonlinear Programming. *Mathematical programming* **106,** 25–57 (2006).
13. Rackauckas, C. & Nie, Q. Differentialequations. jl–a performant and feature-rich ecosystem for solving differential equations in julia. *Journal of Open Research Software* **5** (2017).
14. Rackauckas, C. *Parallel Ensemble Simulations* (GitHub repository). %5Curl%7Bhttps://github.com/SciML/DiffEqDocs.jl/blob/master/docs/src/features/ensemble.md%7D (2021).

A Systematic Dynamic Model Summary

Table 11: Systematic listing of all model equations

Equation	# of eqs	# of vars	Description
$\frac{\mathrm{d}}{\mathrm{dt}}X = (\mu_0 - \mu_d)X$	1	2	Biomass Growth
$\frac{\mathrm{d}}{\mathrm{dt}}N = -\mu_N \frac{N}{N + k_N}X$	1	1	Nitrate Consumption
$\frac{\mathrm{d}}{\mathrm{dt}}q = \mu_N \frac{N}{N + k_N} - \mu_M(I)\left(1 - \frac{k_q}{q}\right)q$	1	2	Intracellular Nitrogen
$\frac{\mathrm{d}}{\mathrm{dt}}f = \left(1 - \frac{k_q}{q}\right)\left(\mu_M(i)(\theta q - \varepsilon f) - \gamma\mu_N \frac{N}{N + k_N} - q\right)$	1	1	FAME Synthesis
$Y(II) = \frac{\exp(\tau q)}{\exp(\tau q) + \delta} + \phi$	1	1	Photosynthetic Activity
$I(z) = I_0 \exp\left(-z(\alpha X + \beta)\right)$	1	1	Scattering + Adsorption
$\mu_m(I) = \frac{\mu_{M0}}{20}\sum_{i=1}^{9}\left(\frac{I_{i=0}}{I_{i=0} + k_s + \frac{I_{i=0}^2}{k_i}} + \frac{I_{z=\frac{iL}{10}}}{I_{z=\frac{iL}{10}} + k_s + \frac{I_{z=\frac{iL}{10}}^2}{k_i}} + \frac{I_{z=10}}{I_{z=10} + k_s + \frac{I_{z=10}^2}{k_i}}\right)$	1	0	Trapezoidal approximation
$\mu_0 = \mu_M\left(1 - \frac{k_q}{q}\right)$	1	0	max to real growth
TOTAL	8	8	

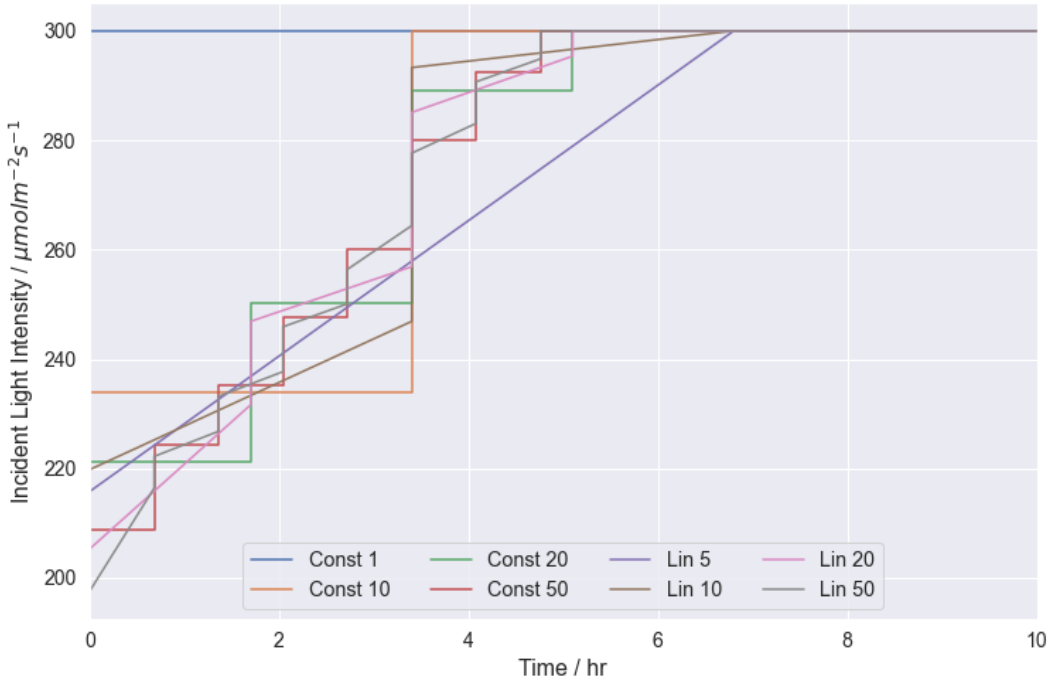
B Parameter Estimation

Table 12: Parameter Estimation Results - full significant figure values
Note: † indicating parameter values having converged to lower, § upper bound

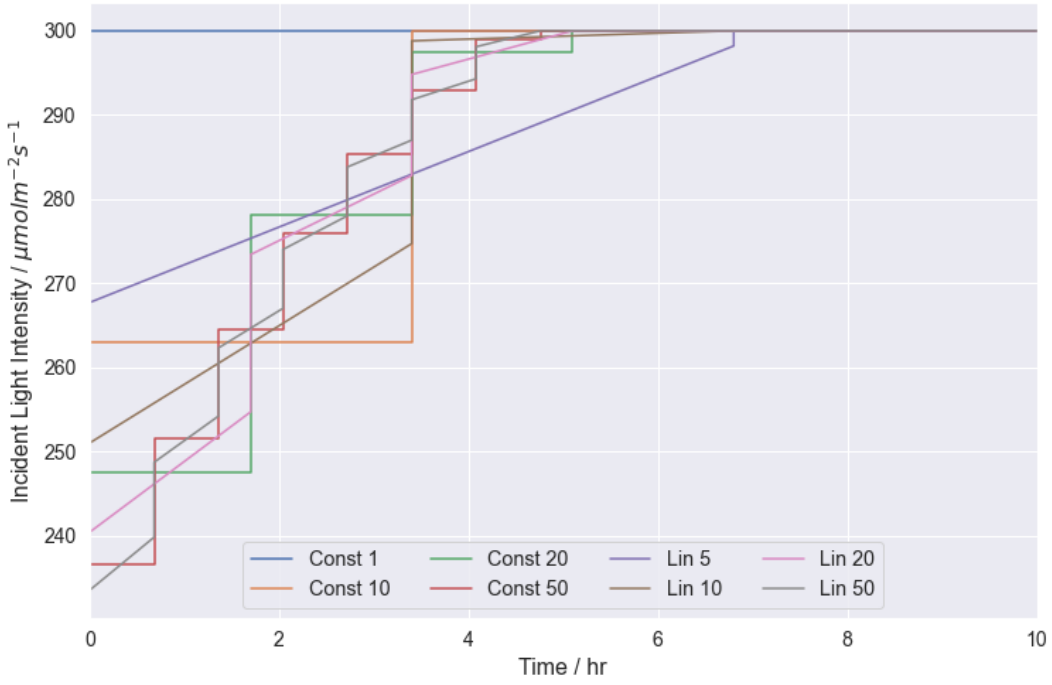
Name	Unit	Final	Initial	Lower Bound	Upper Bound	Confidence Interval 90%	95%	99%	95% t-value	STD
α	$m^2\,kg^{-1}$	300§	196.4	50	300					
β	m^{-1}	0†	0	0	50					
δ	-	12.8	9.9	5	20	4.33	5.17	6.82	2.486	2.624
ϵ	-	0.0002†	0.001	0.0002	0.004					
γ	-	7.28	7.53	1	10	0.279	0.333	0.439	21.85	0.169
k_i	$\mu mol\,m^{-2}\,s^{-1}$	250	100	80	250	132.6	158.3	208.7	1.58	80.3
K_N	$mg\,l^{-1}$	2§	0.8	0.2	2					
k_q	$mg\,g^{-1}$	20.4	19.6	10	100	0.431	0.514	0.678	39.64	0.261
k_s	$\mu mol\,m^{-2}\,s^{-1}$	200§	91.2	50	200					
μ_d	h^{-1}	0†	0	0	0.5					
μ_M	h^{-1}	0.782	0.36	0.3	1	0.027	0.032	0.042	24.55	0.016
μ_N	$mg\,g^{-1}\,h^{-1}$	2.643	2.69	1	10	0.045	0.053	0.070	49.64	0.027
ϕ	-	-0.46	-0.456	-2	2	0.0098	0.012	0.015	39.59	0.0059
τ	-	0.152	0.138	0.05	0.2	0.015	0.018	0.024	8.48	0.0091
θ	-	6.508	6.69	4	15	0.167	0.199	0.262	32.74	0.1008

Reference t-value (95%): 1.652

C Time-varying Control Decisions



(a) Optimal Control Strategy for FAME Productivity



(b) Optimal Control Strategy for Biomass Productivity

Figure 7: Optimal Control Strategy of Incident Light Intensity - Const refers to piecewise constant control and Lin refers to piecewise linear control; numbers represent number of horizon divisions; time after 10 hours was omitted because it was kept at 300µmol m^{-2} s^{-1} for all control regimes