

Supporting Information

Dynamic Modeling of Microalgae Growth and Lipid Production under Transient Light and Nitrogen Conditions

Diyuan Wang, Yi-Chun Lai, Amanda L. Karam, Francis L. de los Reyes, III, Joel J. Ducoste*

Department of Civil, Construction, and Environmental Engineering, North Carolina State University, Raleigh, NC 27695, United States

Table of Contents

S-1 Light measurement and estimation.....	2
S-2 Model calibration methods	4
S-2.1 Use of data in model calibration and validation.....	4
S-2.2 Ranges of parameters	4
S-2.3 Identifiability analysis	5
S-2.4 Parameter estimation and uncertainty quantification	5
S-2.5 Sensitivity analysis.....	6
S-3 Model calibration results	7
S-3.1 Identifiability analysis results	7
S-3.2 Parameter estimation and uncertainty quantification results	8
S-3.3 Sensitivity analysis results.....	12
S-3.4 Light estimation performance	21
S-4 Model validation results.....	21
References	23

List of Figures

Figure S1. Schematic showing the top view of the photobioreactor setup.....	2
Figure S2. Regression curve obtained by fitting the background light measurements at different distances from the light source	3
Figure S3. Light measurements and light estimations at location $l = z_l$	3
Figure S4. Profile likelihood plots for parameter identifiability assessment	7
Figure S5. Demonstration of the asymptotic behavior of the profile likelihood with increasing Q_{nm}	8
Figure S6. Parameter chain plots.....	9
Figure S7. Parameter posterior density plots.....	10
Figure S8. Parameter pairwise correlation plots.....	11
Figure S9. Calibration results of the best fit.....	12

Figure S10. Sensitivity index plots under different conditions ($N_s = 2000$). 19

Figure S11. Sensitivity index plots under different conditions ($N_s = 3000$). 20

Figure S12. Comparison of I_{avg} estimations using the experimental data and the model-simulated data. 21

Figure S13. Comparison of model-predicted light profiles at location $l = z_l$ with both experimental estimation and light sensor measurements. 21

Figure S14. Model validation by predicting the new data. 22

Figure S15. Model validation of I_{avg} predictions. 23

Figure S16. Model validation of I_{z_l} predictions. 23

List of Tables

Table S1. Experimental data divided into two subsets for model calibration and validation. 4

Table S2. Ranges of parameters. 4

Table S3. Experimental data used in sensitivity analysis. 6

Table S4. Settings for the eFAST algorithm. 7

Table S5. Geweke’s p -value of each parameter calculated by MCMC chains. 10

Table S6. Table series (a-e) of parameter sensitivity (S_i) rankings for different state variables. . 14

Table S7. Table series (a-e) of parameter sensitivity (S_{Ti}) rankings for different state variables. 16

S-1 Light measurement and estimation

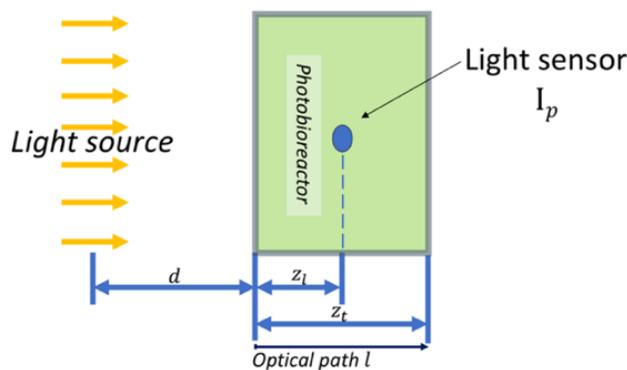


Figure S1. Schematic showing the top view of the photobioreactor setup ($z_l = 0.039624$ m, $z_t = 0.08255$ m).

To quantify the light attenuation due to background absorbing material, we measured the light intensities $I_{R,l}$ at different locations (l) inside the reactor at different light source positions (d). The reactor was filled with culture media and deionized water (DI water). Figure S2 shows the observed relationship between the local background light intensity inside the reactor and the corresponding distance from the light source. We can thus obtain Eqn (8) $I_{R,l} = 1.7324(l + d)^{-1.993}$. This relationship closely followed the inverse square law, suggesting that the algae media and DI water in our photobioreactor setup may have minor impacts on the light intensity.

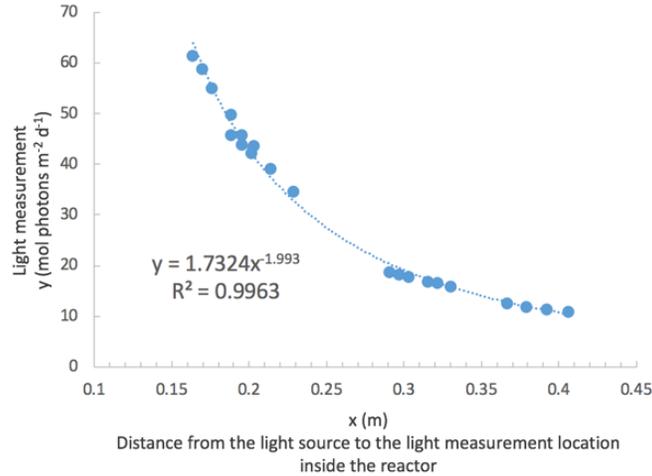


Figure S2. Regression curve obtained by fitting the background light measurements at different distances from the light source.

We then used the Eqn (7) $I_l = I_{R,l} \cdot \exp(-A \cdot l)$ to estimate the actual light intensity experienced by algae cells inside the reactor. We monitored the variation of light intensity during the entire algae cultivation period at location $l = z_l$ to help us decide the values of parameter a and b , which were used to quantify the light absorbance A due to Chl-a and non Chl-a biomass. The calibrated values of a and b were $80.5 \text{ m}^2 (\text{mol } X_{chl}\text{-C})^{-1}$ and $1.4718 \text{ m}^2 (\text{mol C})^{-1}$, respectively. Five experiments were used for the light calibration. We compared the estimated light intensities to the measured data in Figure S3.

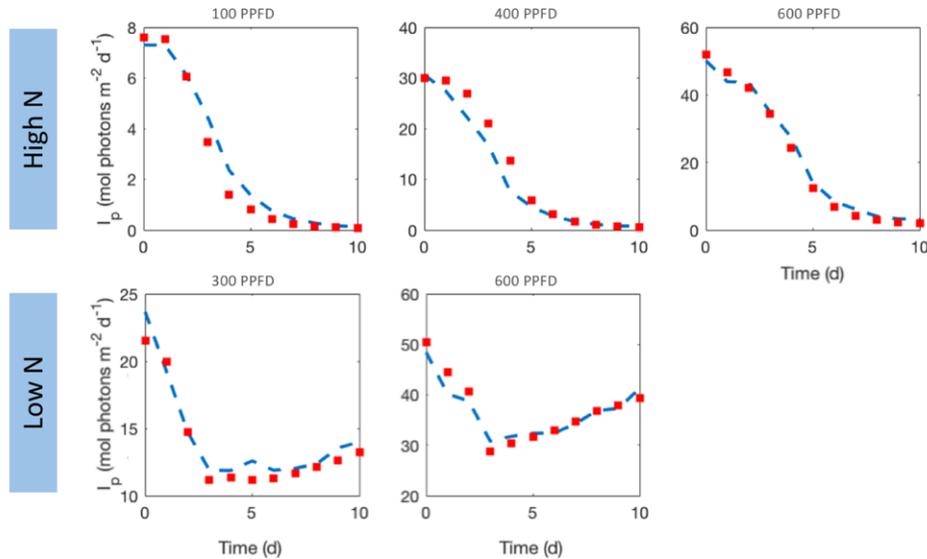


Figure S3. Light measurements and light estimations at location $l = z_l$. The red squares denote the light intensities measured by the light sensor located at $l = z_l$. Blue dashed lines represent light intensities calculated from the experimental data of the X_{chl} and X_C and the fitted values of a and b . The light and nitrogen conditions correspond to the experiments in the calibration group.

S-2 Model calibration methods

S-2.1 Use of data in model calibration and validation

Table S1. Experimental data divided into two subsets for model calibration and validation

Experiments	Light level	Light intensity (PPFD)	N level	NO ₃ ⁻ concentration (mM)
Calibration group (6 experiments)	Low	100	High N	4.95
	Moderate	400		4.89
	High	600		4.88
	Low	100	Low N	0.62
	Moderate	300		0.66
	High	600		0.71
Validation group (5 experiments)	Moderate	300	High N	4.90
		300		3.80
		300	Low N	0.67
		300		0.43
		400		0.43

S-2.2 Ranges of parameters

Table S2. Ranges of parameters

Parameter	Min	Max
P_m	0.1	10
Y_E	0.01	1
k_{ye}	0.001	0.1
q_n	0.001	0.1
ν_{nm}	0.05	2
m	0.001	10
β	0.01	0.5
k_{nl}	0.001	0.1
ϕ	0.001	0.5
m_{cN}	0.001	10
m_c	0.001	15
k_{mc}	0.0001	0.2
θ	0.01	1
k_{chl}	0.001	0.1
k_I	0.0001	0.1
k_N	0.01	20
Q_{nm}	0.16	100

The proposed model initially involved 17 parameters. We determined the parameter ranges based on the underlying biological mechanisms and the reported values from prior literature. The upper and lower bounds were further modified during the preliminary optimization process and the identifiability analysis to ensure a reasonably wide range for each parameter. Table S2 lists the final range of each parameter used during the calibration process.

S-2.3 Identifiability analysis

Based on the profile likelihood method,¹ we started with collecting the baseline values of model parameters from the optimization via the MATLAB MultiStart fmincon routine. The optimization was performed by minimizing the weighted sum of squared error (SSE) between the model predictions and the experimental data.

$$\text{weighted SSE} = \sum_{k=1}^{N_k} \sum_{j=1}^{N_j} \sum_{i=1}^{N_i} \left(\frac{f_{ijk}(\Theta) - y_{ijk}}{\sigma_{ijk}^{exp}} \right)^2 \quad (1)$$

N_k is the number of experiments used in calibration, $N_k = 6$;

N_j is the number of state variables, $N_j = 5$;

N_i is the number of time points, $N_i = 11$;

Θ is the parameter vector;

$f_{ijk}(\Theta)$ are the model responses of state variable j at time i for the experiment k ;

y_{ijk} are the experimental measurements associated with measurement error σ_{ijk}^{exp} .

We used the minimal value of weighted SSE to approximate the parameter likelihood $\chi^2(\Theta)$. We then estimated the likelihood profile of each parameter by reoptimizing the model with the parameter of interest Θ_p fixed at different values. These values were chosen by increasing or decreasing the baseline value of Θ_p until the increment of the likelihood met the threshold, which was determined by the 95% quantile of the χ^2 -distribution with the degree of freedom set to the number of parameters. The shape of the resulted likelihood profile could then indicate the identifiability of the parameter. Specifically, (i) a parameter is identifiable if its likelihood profile exceeds the threshold region in both the positive and negative directions, (ii) a parameter is structurally non-identifiable if its likelihood profile is flat in both directions, and (iii) a parameter is practically non-identifiable if its likelihood profile extends infinitely in either direction.

S-2.4 Parameter estimation and uncertainty quantification

According to the theory of Bayesian inference, the parameter posterior density is inferred from the prior knowledge of the parameter density $\pi(\Theta)$ and the likelihood $\pi(y|\Theta)$.² We used the uniform priors bounded by the parameter ranges (Table S2) due to no prior knowledge of the parameter distributions. The likelihood $\pi(y|\Theta)$ was established using the DRAM algorithm³ by specifying the statistical model for the observed data y . We employed the statistical model (Eqn. 2) to relate the experimental data y_{ijk} to the model responses $f_{ijk}(\Theta)$ by introducing the observation errors ϵ_{ijk} in an additive manner. The observation errors were assumed to be independent and identically distributed (IID) and follow $\epsilon_{ijk} \sim \mathcal{N}(0, \sigma_j^2)$, where σ_j are independent of the observational time and the experimental conditions. The error variances σ_j^2 were estimated by the DRAM algorithm and updated for each iteration.

$$y_{ijk} = f_{ijk}(\Theta) + \epsilon_{ijk} \quad (2)$$

To run the DRAM algorithm, we used the best fit from the MultiStart fmincon optimization generated during the identifiability analysis as the starting point. We also estimated the initial error variance and covariance matrix as the input to initialize the algorithm. Haario *et al.*³ suggested that the value of the non-adaptive period n_0 was a few thousand for models with parameter dimensions larger than 15, but they provided no exact value recommendation. Thus, we selected $n_0 = 1000$ and $n_0 = 5000$, running 250000 iterations each to evaluate the convergence. Geweke's method⁴ was used to check the convergence. The posterior density of each parameter including the mean and standard deviation was estimated from samples that reached a steady range. Parameter uncertainties and the observation errors could propagate to the model output, enabling the estimation of the credible and prediction intervals for the model predictions.² In this work, we randomly selected 2000 parameter sets from the posterior samples to generate the model outputs. The 2.5% and 97.5% quantiles of the model outputs were used to calculate the 95% credible intervals. Based on the statistical model defined previously (Eqn. 2), we determined the 95% prediction intervals by adding observation errors to the model predictions. The observation errors were randomly generated from normal distributions $\mathcal{N}(0, \sigma_j^2)$ based on the variance chain estimated by the DRAM algorithm. The credible intervals measured the limits of the model fit due to parameter uncertainties, and the prediction intervals estimated the limits of the model observations after considering the measurement noises^{2,5}.

S-2.5 Sensitivity analysis

eFAST⁶ estimates the model output variance by varying the input parameters in the form of sinusoidal functions with different set frequencies for different parameters. The output variance is then decomposed by Fourier analysis to quantify the variances due to the particular parameter of interest (Θ_i), the complementary set of parameters, or the higher-order interactions between the parameters of interest and other parameters. As a result, the first-order sensitivity index (S_i) and the total-order sensitivity index (S_{Ti}) of each parameter can be determined. Note in this approach the term S_{Ti} also includes the effects of higher-order interactions between parameters other than (Θ_i).

Table S3. Experimental data used in sensitivity analysis

		Light level	N Level
Stress conditions	No stress	Moderate	High
	Single light stress	High	High
	Single N stress	Moderate	Low
	Dual stress	High	Low
Time points	Day 2, Day 4, Day 10		
State variables	$X_{cd}, X_{carb}, X_{nl}, X_{chl}, S_{NO3}$		

We systematically performed the sensitivity analysis for each parameter under four conditions (Table S3) corresponding to the effects of different stress types and levels. For each condition, we evaluated five state variables at three representative time points (Table S3). Day 2, Day 4 and Day 10 were chosen to represent the entire time course. Day 2 and Day 4 also corresponded to the transient states of nitrogen depletion under low nitrogen conditions. Details of the eFAST settings are listed in Table S4. We used two sample sizes to check the accuracy of the sensitivity

results. Also note that after resolving the non-identifiability associated with a certain parameter, a total of 17 parameters including one dummy parameter were evaluated in the sensitivity analysis.

Table S4. Settings for the eFAST algorithm

eFAST Settings	Values
Number of samples per curve N_S	2000, 3000
Number of parameters N_p	17
Number of repeated times N_R	5

S-3 Model calibration results

S-3.1 Identifiability analysis results

Q_{nm} appeared to be the only non-identifiable parameter as indicated by the flatness of its profile likelihood extending to the upper bound of the search range (Figure S4). The ratio $\frac{Q_{nm}-Q_n}{Q_{nm}-q_n}$ would approach one, if the value of Q_{nm} becomes very large. As a result, $\frac{Q_{nm}-Q_n}{Q_{nm}-q_n}$ could be removed from the nitrate uptake rate equation. We further compared the changes in profile likelihood by reoptimizing the model via either increasing the value of Q_{nm} or eliminating the term $\frac{Q_{nm}-Q_n}{Q_{nm}-q_n}$ (Figure S5). As Q_{nm} increases, the profile likelihood decreases and approaches the yellow line, indicating that the removal of $\frac{Q_{nm}-Q_n}{Q_{nm}-q_n}$ improved the model performance by reducing the discrepancy between the model prediction and the experimental data.

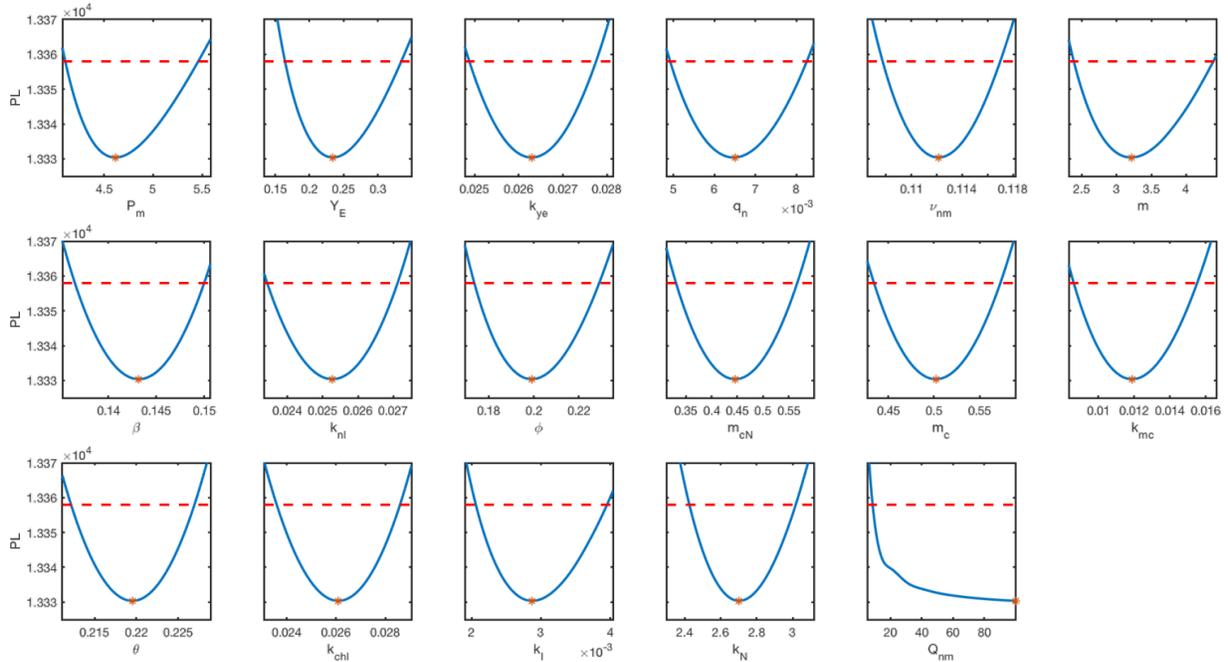


Figure S4. Profile likelihood plots for parameter identifiability assessment. Solid lines represent profile likelihood versus parameter value. The baseline parameter values optimized from the MultiStart fmincon are indicated by asterisks. The dashed lines denote the threshold of the likelihood confidence region.

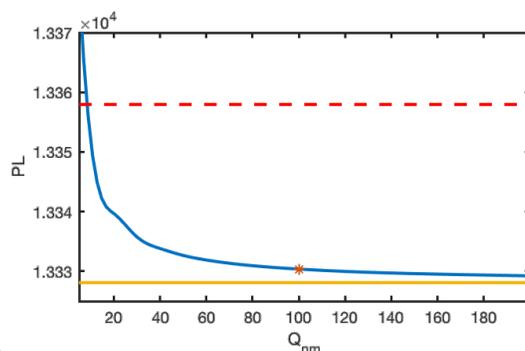


Figure S5. Demonstration of the asymptotic behavior of the profile likelihood with increasing Q_{nm} . The yellow solid line denotes the minimum weighted SSE value obtained by removing the term $\frac{Q_{nm}-Q_n}{Q_{nm}-q_n}$ from the nitrate uptake equation.

S-3.2 Parameter estimation and uncertainty quantification results

After removing Q_{nm} , our model had 16 parameters remaining to be calibrated. Parameters were estimated to simultaneously fit the model under different experimental conditions. We obtained two sample chains by running the DRAM algorithm with $n_0 = 1000$ and $n_0 = 5000$. The corresponding MCMC sample chains and the steady ranges are illustrated in Figure S6. After 250,000 iterations, parameter values in both chains converged to similar steady ranges. We then compared the p -values of these two chains using the Geweke's method.⁴ For most parameters, we noticed that the p -values calculated from the chain of $n_0 = 1000$ were closer to 1 than the p -values calculated from the chain of $n_0 = 5000$, indicating that the chain of $n_0 = 1000$ was more stable over the iterations (Table S5). We thus chose the chain of $n_0 = 1000$ to generate the posterior parameter densities (Figure S7) and the pairwise parameter correlations (Figure S8).

Figure S9 further compares the experimental data with the model predictions by using two sets of best-fit parameter values. Parameters with maximum likelihood function could have a slightly better fit to the lipid production, but we could not find a significant difference between the two best fits. This small difference could reflect certain arbitrariness in choosing a single best fit, especially for models with a high-dimensional parameter space. Moreover, defining a “best fit” could be difficult due to measurement noises. The MCMC method, by accounting for the observation errors and searching for all the potential good fits, allowed us to quantify the uncertainties of fitting parameters and model predictions. Thus, we could evaluate the model performance in a non-arbitrary way.

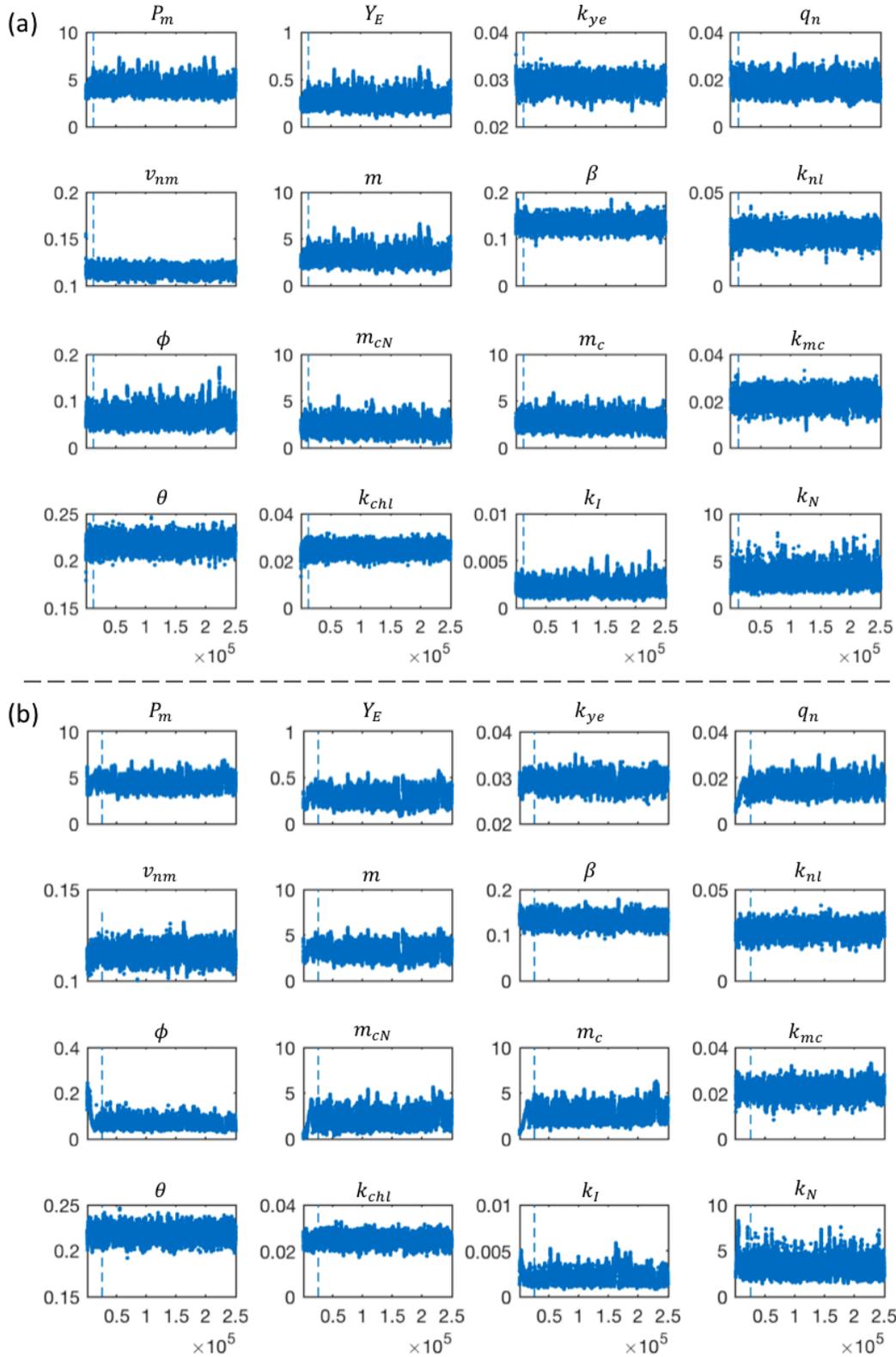


Figure S6. Parameter chain plots (a) MCMC chain with $n_0 = 1000$, (b) MCMC chain with $n_0 = 5000$. Dashed lines indicate the location at which the chain settled to the steady range.

Table S5. Geweke's p -value of each parameter calculated by MCMC chains with $n_0 = 1000$ and $n_0 = 5000$

Parameter	p -value ($n_0 = 1000$)	p -value ($n_0 = 5000$)
P_m	0.9906	0.9273
Y_E	0.9886	0.9020
k_{ye}	0.9999	0.9920
q_n	0.9605	0.8552
ν_{nm}	0.9998	0.9929
m	0.9913	0.8804
β	0.9981	0.9878
k_{nl}	0.9939	0.9667
ϕ	0.9061	0.9759
m_{cN}	0.9268	0.9617
m_c	0.9550	0.9930
k_{mc}	0.9819	0.9945
θ	0.9967	0.9909
k_{chl}	0.9923	0.9591
k_I	0.9402	0.7753
k_N	0.9911	0.9816

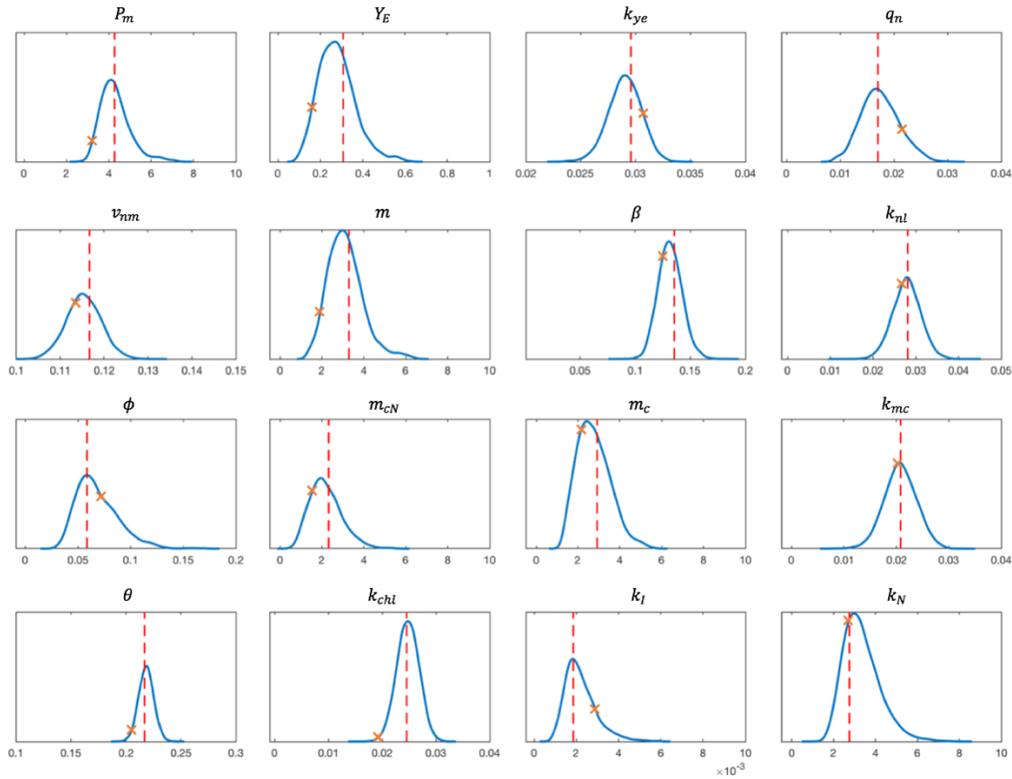


Figure S7. Parameter posterior density plots. Red dashed lines indicate the parameter values of the best fit with maximum likelihood $\pi(y|\theta)$. Cross signs denote the parameter values with minimum SSE.

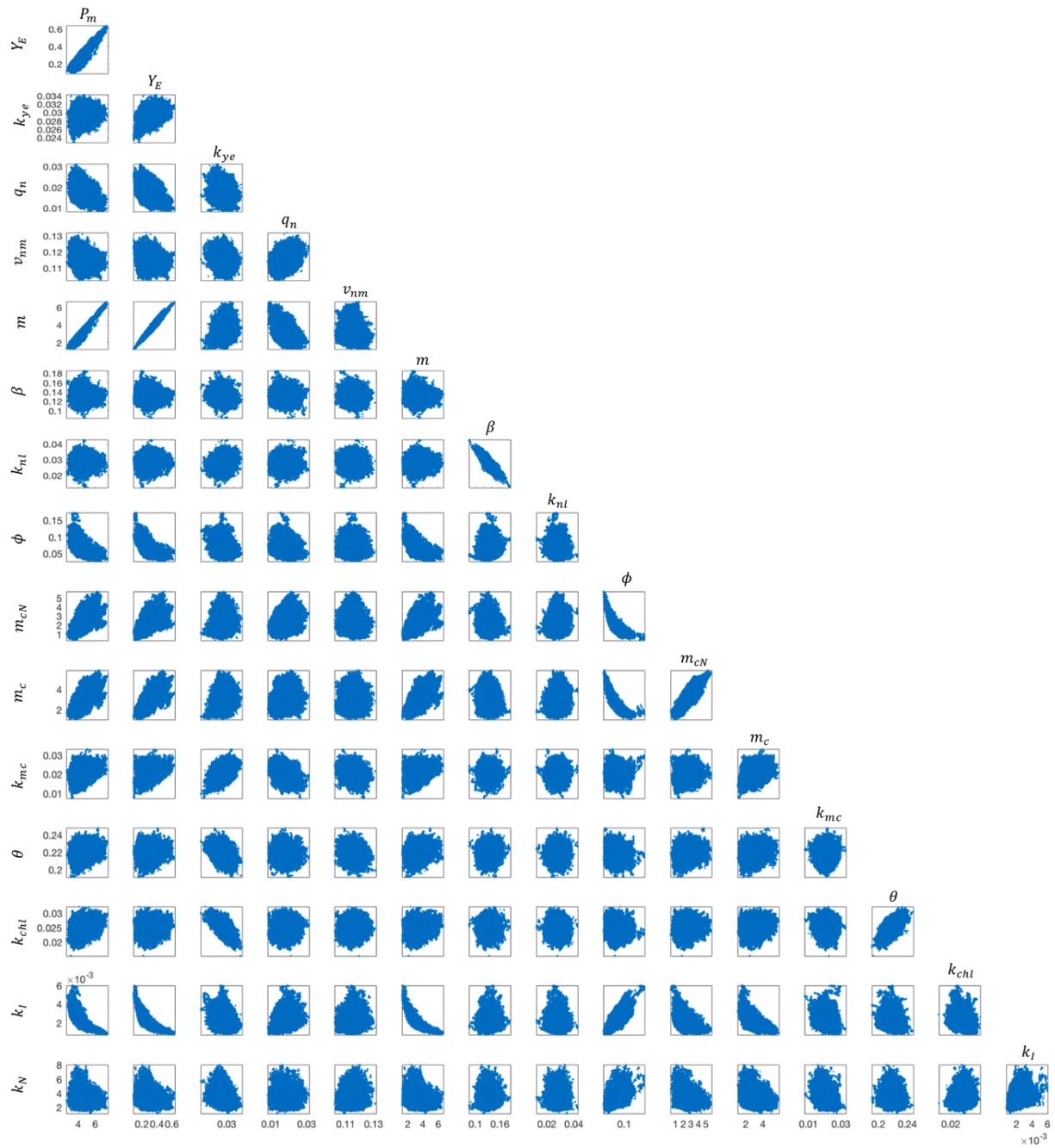


Figure S8. Parameter pairwise correlation plots

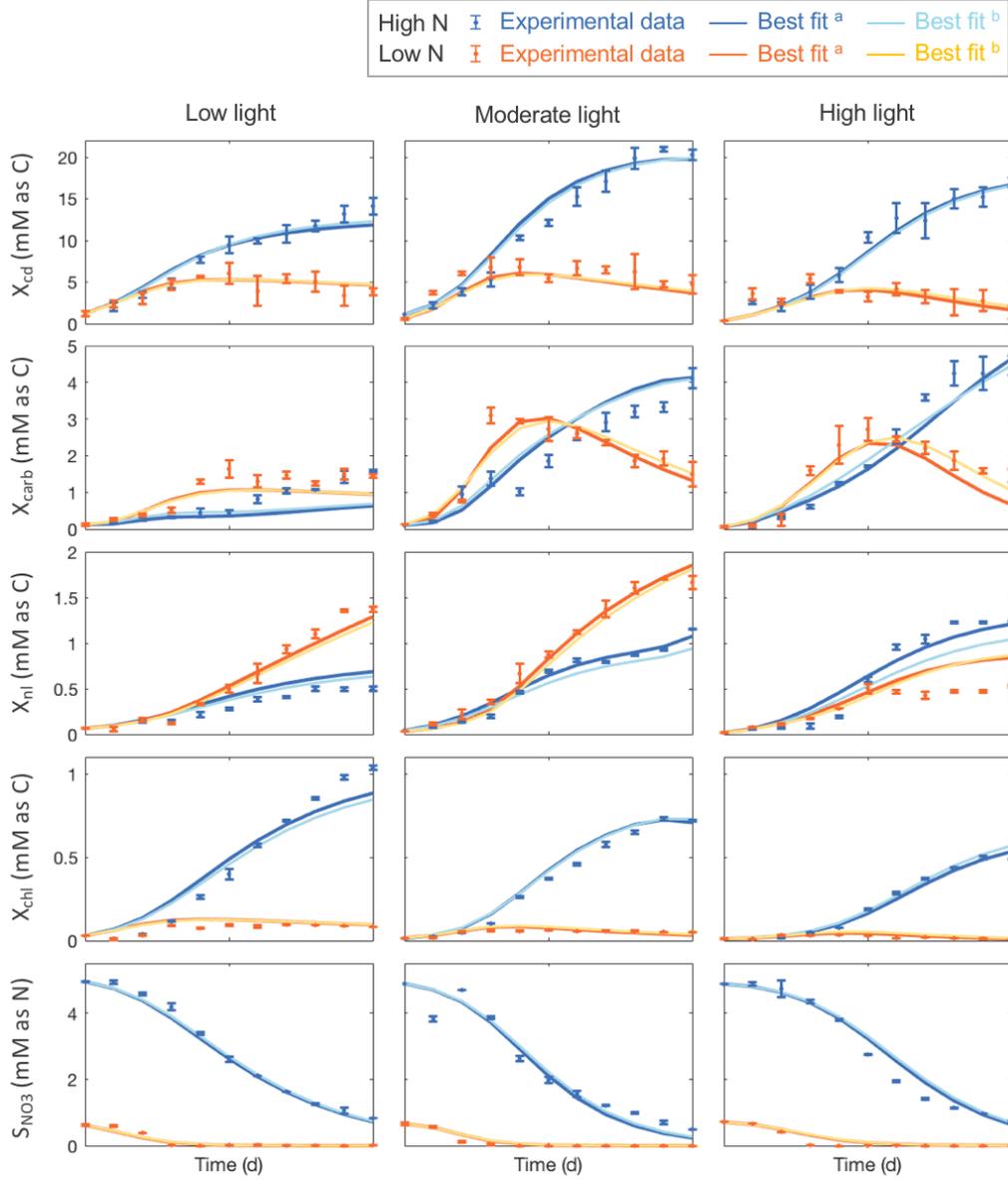


Figure S9. Calibration results of the best fit ^a based on maximum $\pi(y|\theta)$ and the best fit ^b based on minimum SSE. Solid lines denote the model predictions. Symbols with error bars represent experimental data and measurement noises.

S-3.3 Sensitivity analysis results

We first used 2000 samples per search curve to analyze the parameter sensitivities. Increasing to 3000 samples per search curve did not change the sensitivity ranks much but slightly increased the number of parameters whose sensitivity indices were significantly different from the dummy parameter. Therefore, we used the results from $N_S = 3000$ to report the parameter sensitivity ranks for both S_i (Table S6) and S_{T_i} (Table S7). The sensitivity plots for $N_S = 2000$ and $N_S = 3000$ are shown in Figure S10 and Figure S11, respectively.

Potential model reduction. Some parameters were mainly sensitive to certain variables or under certain conditions. For example, ϕ was mainly sensitive to the neutral lipid production; k_N and k_I became more sensitive under light and nitrogen stress; k_{mc} , m_{cN} , and m_c were primarily sensitive to carbohydrate production. ϕ could not be reduced as it represented the major carbon flux for lipid production under nitrogen depletion. As tuning parameters, k_N and k_I were used to adjust carbon redirection to lipid accumulation and to regulate Chl-a degradation when the low-nitrogen and high-light stresses would affect the corresponding processes. We could potentially set these two parameters at certain nominal values, such as the mean values of their posterior distributions, to improve the estimation of other parameters. Among the parameters related to carbohydrate maintenance, we noticed that m_c was less sensitive to carbohydrate, however, k_{mc} was more sensitive. Since m_c and k_{mc} were coupled in the process R_4 , further tests would be needed to see if the process R_4 could be reduced. Simplifying the carbohydrate maintenance equation may help reduce the parameter uncertainties associated with the carbohydrate prediction.

Maintenance and respiration. Different stress conditions may affect the maintenance and respiration process in complex ways. Comparing to the no-stress condition, single low-nitrogen stress increased the sensitivities of the parameter m for X_{cd} and the parameters m_c , k_{mc} , and m_{cN} for X_{carb} as shown in Table S6. Such increases may indicate that the functional biomass and carbohydrate were more prone to maintenance processes under nitrogen deficiency. In contrast, high-light stress seemed to decrease the sensitivities of these parameters (Table S6), potentially indicating a relatively lower activity of respiration caused by photoinhibition.⁷ It is also interesting to notice that the sensitivities of these parameters decreased under the dual-stress condition, which may suggest that the effect of high-light stress outweighed the low-nitrogen stress in our case. In addition, parameters such as m , m_c , and k_{mc} were formulated based on a first-order maintenance relationship and they had a slightly higher sensitivity on later time periods, indicating an accumulative effect over time. In contrast, the sensitivity of m_{cN} seemed to decrease with time, which was reasonable because m_{cN} was related to the nitrate uptake rate. Once the nitrate was depleted, m_{cN} would no longer affect the model output.

Neutral lipid accumulation. The sensitivity variations of parameters related to lipid production further verified the effects of light and nitrogen stresses on the activities of R_6 and R_7 . Specifically, the low-nitrogen stress decreased the sensitivities of R_6 -related parameters β and k_{nl} but increased the sensitivity of R_7 -related parameter ϕ . This variation was in line with our model formulation, as R_6 was active under nitrogen-available conditions whereas R_7 was active under nitrogen depletion. In addition, the sensitivity rank of ϕ was lower than β and k_{nl} under dual-stress conditions, potentially indicating the reduced activity of R_7 when photoinhibition affected the carbon fixation.

Table S6. Table series (a-e) of parameter sensitivity (S_i) rankings for different state variables ($N_S = 3000$). Parameters are listed from the most sensitive (rank 1) to least sensitive (rank 17). Related parameters are grouped by colors based on their functions: P_m , Y_E , k_{ye} , q_n , v_{nm} , and m are highlighted in yellow; k_{nl} and β in red; k_{chl} and θ in green; k_{mc} , m_{cN} , and m_c in purple; k_I and k_N in blue; ϕ in pink; and the dummy parameter in gray.

(a) Xcd			S_i	Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
No Stress	400HN	Day 2	qn	vnm	kye	Pm	YE	kmc	kchl	m	mcN	theta	mc	beta	kN	knI	kl	dum	phi		
		Day 4	Pm	qn	kye	vnm	YE	kmc	theta	m	kchl	mcN	beta	mc	knI	kl	kN	phi	dum		
		Day 10	Pm	qn	kye	vnm	m	YE	theta	kmc	mcN	beta	kchl	knI	kl	mc	kN	dum	phi		
Light Stress	600HN	Day 2	knI	kye	qn	Pm	vnm	beta	YE	kchl	mcN	kmc	m	theta	kN	phi	kl	mc	dum		
		Day 4	knI	kye	qn	Pm	beta	vnm	YE	kchl	theta	m	mcN	kmc	kN	phi	kl	mc	dum		
		Day 10	kye	knI	qn	Pm	beta	vnm	YE	m	theta	kchl	kmc	mcN	kN	dum	mc	phi	kl		
N Stress	300LN	Day 2	qn	Pm	vnm	YE	m	kye	mcN	kmc	theta	kchl	kN	beta	mc	knI	kl	phi	dum		
		Day 4	Pm	qn	vnm	m	YE	kye	mcN	kmc	theta	kN	kchl	kl	beta	mc	knI	phi	dum		
		Day 10	Pm	qn	vnm	m	YE	kye	mcN	kN	kmc	theta	kchl	kl	knI	beta	mc	phi	dum		
Dual Stress	600LN	Day 2	knI	qn	kye	Pm	YE	kchl	beta	theta	kmc	m	vnm	kN	mcN	mc	kl	phi	dum		
		Day 4	knI	qn	kye	Pm	YE	kchl	beta	m	kN	theta	vnm	mcN	kmc	mc	kl	phi	dum		
		Day 10	knI	qn	kye	Pm	YE	beta	kchl	kN	m	vnm	theta	kmc	mcN	mc	kl	phi	dum		

(b) Xcarb			S_i	Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
No Stress	400HN	Day 2	kye	vnm	Pm	qn	YE	mcN	kmc	m	theta	kchl	mc	beta	kN	knI	kl	phi	dum		
		Day 4	kye	Pm	qn	vnm	m	kmc	YE	theta	mcN	kchl	mc	beta	kl	kN	knI	phi	dum		
		Day 10	Pm	kye	qn	kmc	vnm	m	YE	theta	mcN	mc	beta	kchl	kl	kN	knI	dum	phi		
Light Stress	600HN	Day 2	kye	knI	qn	Pm	vnm	beta	kmc	YE	mcN	kchl	theta	m	phi	kN	kl	mc	dum		
		Day 4	kye	knI	qn	Pm	kmc	beta	YE	vnm	m	kchl	mcN	theta	kN	kl	phi	mc	dum		
		Day 10	kye	knI	qn	kmc	Pm	YE	beta	vnm	m	kchl	theta	mcN	kN	mc	dum	phi	kl		
N Stress	300LN	Day 2	Pm	qn	vnm	m	YE	kmc	kye	mcN	theta	mc	kN	kchl	kl	beta	phi	knI	dum		
		Day 4	Pm	vnm	qn	kmc	m	YE	kye	mcN	mc	theta	kN	kchl	kl	beta	phi	knI	dum		
		Day 10	vnm	Pm	qn	kmc	m	mc	YE	mcN	kye	kN	theta	kchl	kl	beta	phi	knI	dum		
Dual Stress	600LN	Day 2	kye	knI	qn	Pm	YE	kchl	kmc	beta	m	theta	mcN	vnm	kN	mc	phi	kl	dum		
		Day 4	kye	knI	qn	Pm	kmc	YE	kchl	beta	m	kN	vnm	theta	mcN	mc	kl	phi	dum		
		Day 10	knI	kye	qn	Pm	kmc	YE	kchl	beta	kN	m	vnm	mcN	theta	mc	kl	phi	dum		

(c) Xnl			S_i	Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
No Stress	400HN	Day 2	knI	m	beta	vnm	Pm	kye	kchl	phi	YE	mcN	qn	kN	theta	mc	kl	kmc	dum		
		Day 4	m	Pm	knI	vnm	beta	kye	phi	kchl	YE	qn	mcN	kl	kN	theta	mc	kmc	dum		
		Day 10	m	Pm	knI	phi	kye	kl	vnm	qn	beta	kchl	YE	mcN	theta	kN	mc	kmc	dum		
Light Stress	600HN	Day 2	knI	vnm	m	beta	kye	Pm	qn	YE	kchl	phi	kN	dum	theta	kl	kmc	mcN	mc		
		Day 4	m	knI	vnm	kye	Pm	beta	qn	phi	YE	kchl	kl	mcN	dum	kN	theta	kmc	mc		
		Day 10	m	knI	vnm	kl	kye	Pm	phi	beta	YE	qn	kchl	theta	mcN	kN	dum	mc	kmc		
N Stress	300LN	Day 2	m	Pm	phi	vnm	YE	knI	qn	kye	theta	kN	beta	kchl	kl	mcN	dum	mc	kmc		
		Day 4	m	Pm	phi	qn	vnm	YE	theta	kye	kl	kN	knI	kchl	mcN	beta	dum	mc	kmc		
		Day 10	m	Pm	phi	qn	kl	YE	theta	kye	vnm	kN	knI	mcN	kchl	beta	dum	mc	kmc		
Dual Stress	600LN	Day 2	knI	m	vnm	beta	Pm	kye	YE	phi	qn	kN	kchl	theta	kmc	kl	mcN	mc	dum		
		Day 4	m	knI	beta	vnm	Pm	kye	YE	phi	kl	kchl	qn	theta	kN	kmc	mcN	mc	dum		
		Day 10	m	knI	beta	kye	kl	phi	vnm	YE	kchl	Pm	theta	qn	kN	dum	kmc	mc	mcN		

(d) **Xchl**

		S_i	Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
No Stress	400HN	Day 2		theta	m	Pm	kchl	vnm	kye	YE	knl	beta	qn	kN	kl	mcN	mc	kmc	phi	dum
		Day 4		theta	m	kchl	kye	Pm	YE	vnm	kl	knl	kn	beta	qn	mcN	phi	mc	kmc	dum
		Day 10		theta	m	kchl	YE	kye	kl	kN	Pm	vnm	knl	beta	mcN	qn	dum	kmc	phi	mc
Light Stress	600HN	Day 2		kchl	m	theta	vnm	Pm	kye	qn	knl	YE	kN	beta	kl	mcN	kmc	phi	dum	mc
		Day 4		kchl	theta	m	kye	vnm	Pm	qn	knl	YE	kN	kl	beta	mcN	kmc	phi	dum	mc
		Day 10		kchl	theta	kye	YE	m	kN	vnm	knl	Pm	kl	qn	beta	dum	mcN	phi	kmc	mc
N Stress	300LN	Day 2		kchl	m	theta	kN	vnm	kl	YE	Pm	kye	qn	knl	mcN	beta	kmc	phi	dum	mc
		Day 4		kchl	theta	m	kN	kl	Pm	vnm	YE	kye	qn	beta	mcN	kmc	knl	phi	dum	mc
		Day 10		kchl	theta	kN	m	Pm	kl	YE	kye	vnm	qn	beta	knl	kmc	mcN	phi	dum	mc
Dual Stress	600LN	Day 2		kchl	theta	m	kN	vnm	kye	Pm	kl	knl	YE	qn	beta	mcN	kmc	mc	dum	phi
		Day 4		kchl	theta	m	kN	kl	kye	vnm	YE	qn	Pm	knl	kmc	mc	beta	dum	mcN	phi
		Day 10		kchl	theta	kN	m	qn	Pm	YE	vnm	kye	kl	knl	beta	mcN	kmc	mc	phi	dum

(e) **Sno3**

		S_i	Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
No Stress	400HN	Day 2		m	Pm	vnm	kye	knl	qn	YE	beta	kchl	theta	mcN	kmc	kN	mc	dum	kl	phi
		Day 4		Pm	m	vnm	kye	knl	kchl	beta	qn	YE	theta	mcN	kmc	kl	kN	mc	dum	phi
		Day 10		Pm	m	vnm	kye	knl	kchl	YE	beta	theta	qn	mcN	kmc	kN	mc	kl	dum	phi
Light Stress	600HN	Day 2		m	vnm	Pm	kye	qn	YE	knl	kchl	theta	beta	mcN	kmc	kN	phi	mc	kl	dum
		Day 4		m	vnm	Pm	kye	qn	knl	YE	kchl	theta	beta	mcN	kmc	kN	phi	dum	kl	mc
		Day 10		m	vnm	Pm	kye	qn	knl	YE	beta	kchl	theta	mcN	kN	kmc	mc	dum	kl	phi
N Stress	300LN	Day 2		m	vnm	Pm	kye	YE	qn	kchl	knl	theta	beta	kN	mcN	kmc	kl	mc	dum	phi
		Day 4		m	Pm	vnm	kye	YE	qn	knl	kchl	theta	beta	kN	kmc	mcN	mc	kl	dum	phi
		Day 10		m	Pm	vnm	kye	YE	qn	knl	beta	theta	kchl	kN	kmc	mcN	mc	phi	dum	kl
Dual Stress	600LN	Day 2		m	vnm	Pm	kye	qn	knl	YE	kN	beta	kchl	theta	mcN	kmc	mc	dum	kl	phi
		Day 4		m	vnm	Pm	kye	qn	knl	YE	kN	beta	kchl	mcN	theta	kmc	kl	dum	mc	phi
		Day 10		m	vnm	Pm	kye	qn	knl	kN	YE	beta	kchl	mcN	kmc	kl	theta	dum	mc	phi

Table S7. Table series (a-e) of parameter sensitivity (S_{Ti}) rankings for different state variables ($N_S = 3000$). Parameters are listed from the most sensitive (rank 1) to least sensitive (rank 17). Related parameters are grouped by colors based on their functions: P_m , Y_E , k_{ye} , q_n , v_{nm} , and m are highlighted in yellow; k_{nl} and β in red; k_{chl} and θ in green; k_{mc} , m_{cN} , and m_c in purple; k_I and k_N in blue; ϕ in pink; and the dummy parameter in gray.

(a) Xcd			S_{Ti}	Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
No Stress	400HN	Day 2	Pm	kye	qn	vnm	kchl	YE	m	theta	kmc	mc	mcN	beta	kN	kl	knI	phi	dum		
		Day 4	Pm	kye	qn	kchl	YE	m	theta	vnm	kmc	beta	mc	mcN	kN	kl	knI	phi	dum		
		Day 10	Pm	kye	qn	YE	kchl	m	theta	vnm	beta	mcN	kmc	mc	kl	kN	phi	knI	dum		
Light Stress	600HN	Day 2	kye	knI	qn	YE	Pm	vnm	beta	kchl	m	kN	theta	phi	mcN	kl	dum	kmc	mc		
		Day 4	kye	knI	qn	YE	Pm	vnm	beta	kchl	m	kN	theta	mcN	phi	kl	dum	kmc	mc		
		Day 10	kye	knI	qn	YE	Pm	vnm	beta	kchl	m	kN	theta	mcN	kl	phi	kmc	dum	mc		
N Stress	300LN	Day 2	Pm	vnm	qn	kye	YE	m	kchl	mcN	theta	kmc	beta	kN	mc	dum	knI	phi	kl		
		Day 4	Pm	vnm	qn	kye	YE	m	kchl	theta	mcN	kmc	kN	beta	mc	knI	phi	dum	kl		
		Day 10	Pm	vnm	qn	kye	YE	m	kchl	mcN	theta	kmc	kN	mc	knI	beta	phi	dum	kl		
Dual Stress	600LN	Day 2	kye	knI	Pm	qn	kchl	YE	beta	vnm	mcN	theta	kmc	m	phi	kN	kl	mc	dum		
		Day 4	kye	knI	qn	Pm	kchl	YE	beta	vnm	kN	m	kmc	mcN	theta	phi	mc	kl	dum		
		Day 10	kye	knI	qn	Pm	beta	kchl	YE	vnm	kN	mcN	m	kmc	theta	phi	kl	mc	dum		

(b) Xcarb			S_{Ti}	Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
No Stress	400HN	Day 2	Pm	kye	m	kchl	YE	vnm	qn	theta	kmc	beta	mc	mcN	kN	kl	knI	phi	dum		
		Day 4	Pm	kye	m	YE	qn	kchl	theta	vnm	kmc	beta	mc	kN	kl	mcN	knI	phi	dum		
		Day 10	Pm	kye	YE	vnm	qn	kmc	theta	m	kchl	beta	kl	mc	mcN	kN	knI	phi	dum		
Light Stress	600HN	Day 2	kye	knI	qn	YE	Pm	vnm	beta	kchl	kmc	m	phi	kN	theta	dum	mcN	kl	mc		
		Day 4	kye	knI	qn	YE	Pm	beta	vnm	kchl	kmc	m	kN	phi	theta	dum	mcN	kl	mc		
		Day 10	kye	knI	qn	YE	Pm	kmc	vnm	beta	kchl	m	kN	theta	phi	kl	mcN	dum	mc		
N Stress	300LN	Day 2	Pm	vnm	kye	YE	qn	m	kmc	kchl	mcN	theta	mc	kN	beta	knI	phi	dum	kl		
		Day 4	Pm	vnm	kye	YE	qn	kmc	m	kchl	mcN	theta	mc	kN	beta	knI	phi	dum	kl		
		Day 10	Pm	vnm	YE	kye	qn	kmc	mc	mcN	kchl	theta	m	kN	beta	phi	knI	dum	kl		
Dual Stress	600LN	Day 2	kye	knI	Pm	qn	YE	kchl	beta	vnm	kmc	m	mcN	theta	kN	phi	kl	mc	dum		
		Day 4	kye	knI	qn	Pm	YE	kchl	beta	kmc	vnm	m	kN	mcN	theta	kl	mc	phi	dum		
		Day 10	kye	knI	qn	Pm	YE	kchl	kmc	beta	vnm	kN	mcN	m	theta	kl	phi	mc	dum		

(c) Xnl			S_{Ti}	Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
No Stress	400HN	Day 2	m	Pm	kchl	kye	vnm	YE	theta	mcN	beta	qn	phi	knI	mc	kmc	dum	kN	kl		
		Day 4	m	kye	kchl	Pm	vnm	theta	YE	mcN	phi	qn	beta	knI	kN	mc	kmc	kl	dum		
		Day 10	m	kye	kchl	Pm	vnm	mc	YE	theta	mcN	qn	phi	beta	kN	knI	kmc	kl	dum		
Light Stress	600HN	Day 2	qn	knI	kye	m	vnm	beta	YE	kchl	Pm	theta	phi	dum	kmc	kl	kN	mcN	mc		
		Day 4	m	kye	qn	kchl	vnm	YE	Pm	knI	dum	beta	mcN	theta	phi	kl	kN	kmc	mc		
		Day 10	m	kchl	kye	theta	qn	YE	Pm	vnm	mcN	kl	kN	beta	dum	knI	phi	kmc	mc		
N Stress	300LN	Day 2	m	Pm	kye	YE	phi	qn	theta	mcN	kchl	vnm	kN	beta	knI	dum	kmc	mc	kl		
		Day 4	m	Pm	kye	YE	qn	kchl	phi	theta	vnm	mcN	kN	beta	knI	dum	kl	mc	kmc		
		Day 10	m	Pm	kye	YE	qn	kchl	phi	theta	vnm	kN	beta	mcN	dum	mc	knI	kl	kmc		
Dual Stress	600LN	Day 2	knI	kye	m	Pm	kmc	beta	YE	vnm	kchl	qn	phi	theta	kN	mcN	kl	mc	dum		
		Day 4	m	knI	kye	Pm	YE	kchl	kl	kmc	beta	theta	phi	vnm	qn	mcN	mc	kN	dum		
		Day 10	m	kye	knI	kl	Pm	YE	dum	kchl	beta	theta	phi	mc	kmc	qn	vnm	mcN	kN		

(d)		Xchl																		
		S_{Ti}	Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
No Stress	400HN	Day 2		theta	m	Pm	kye	vnm	kchl	YE	qn	knl	beta	mcN	kN	kl	kmc	mc	phi	dum
		Day 4		m	kye	theta	vnm	YE	Pm	kchl	qn	beta	knl	kN	kl	mcN	kmc	mc	phi	dum
		Day 10		m	vnm	kye	YE	theta	Pm	kchl	beta	qn	knl	kN	kl	mcN	kmc	mc	phi	dum
Light Stress	600HN	Day 2		kchl	kye	m	theta	vnm	Pm	YE	qn	knl	beta	kN	mcN	kl	mc	kmc	dum	phi
		Day 4		kchl	kye	theta	m	vnm	YE	Pm	qn	knl	beta	kN	mcN	kl	kmc	mc	dum	phi
		Day 10		kye	kchl	vnm	m	YE	theta	knl	Pm	beta	qn	kl	kN	mcN	phi	kmc	dum	mc
N Stress	300LN	Day 2		m	kchl	theta	vnm	Pm	YE	kye	kN	qn	kl	beta	mcN	kmc	knl	mc	dum	phi
		Day 4		m	kchl	theta	vnm	Pm	YE	kye	kN	qn	kl	beta	kmc	mcN	knl	mc	dum	phi
		Day 10		kchl	theta	m	Pm	YE	vnm	kye	kN	beta	kl	qn	kmc	knl	mcN	phi	mc	dum
Dual Stress	600LN	Day 2		kchl	theta	m	vnm	kye	YE	Pm	kN	qn	knl	kl	beta	mcN	dum	mc	phi	kmc
		Day 4		kchl	m	vnm	theta	kye	YE	Pm	qn	kN	knl	beta	dum	kl	kmc	mcN	mc	phi
		Day 10		kchl	m	kye	vnm	theta	qn	Pm	YE	kN	knl	beta	dum	mc	mcN	phi	kmc	kl

(e)		Sno3																		
		S_{Ti}	Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
No Stress	400HN	Day 2		Pm	m	vnm	kye	YE	kchl	theta	knl	qn	beta	mcN	kmc	mc	kN	phi	dum	kl
		Day 4		Pm	m	kye	vnm	YE	kchl	theta	knl	beta	qn	mcN	kmc	kN	mc	phi	kl	dum
		Day 10		Pm	kye	m	vnm	YE	kchl	knl	theta	beta	qn	kmc	mcN	kN	mc	kl	dum	phi
Light Stress	600HN	Day 2		kye	vnm	m	Pm	YE	knl	qn	kchl	beta	theta	mcN	kmc	kl	kN	phi	mc	dum
		Day 4		kye	m	vnm	Pm	knl	YE	kchl	qn	beta	theta	mcN	kl	kmc	phi	mc	dum	kN
		Day 10		kye	vnm	Pm	m	knl	YE	qn	kchl	beta	theta	mcN	kmc	kl	dum	kN	phi	mc
N Stress	300LN	Day 2		Pm	m	vnm	kye	YE	qn	theta	kchl	beta	knl	kN	mcN	kmc	dum	kl	mc	phi
		Day 4		Pm	m	vnm	kye	YE	qn	theta	kchl	beta	knl	kN	mcN	kmc	kl	dum	mc	phi
		Day 10		Pm	m	vnm	YE	kye	qn	kN	theta	knl	kchl	beta	mcN	kmc	kl	mc	dum	phi
Dual Stress	600LN	Day 2		m	vnm	kye	Pm	knl	YE	qn	kchl	beta	theta	kN	mcN	dum	kmc	mc	kl	phi
		Day 4		m	kye	vnm	knl	Pm	YE	qn	beta	kchl	kN	theta	mcN	kmc	dum	kl	mc	phi
		Day 10		kye	m	knl	vnm	Pm	YE	qn	beta	kchl	kN	theta	mcN	kmc	mc	dum	phi	kl

(d) $N_s = 2000$, Dual stress (600LN)

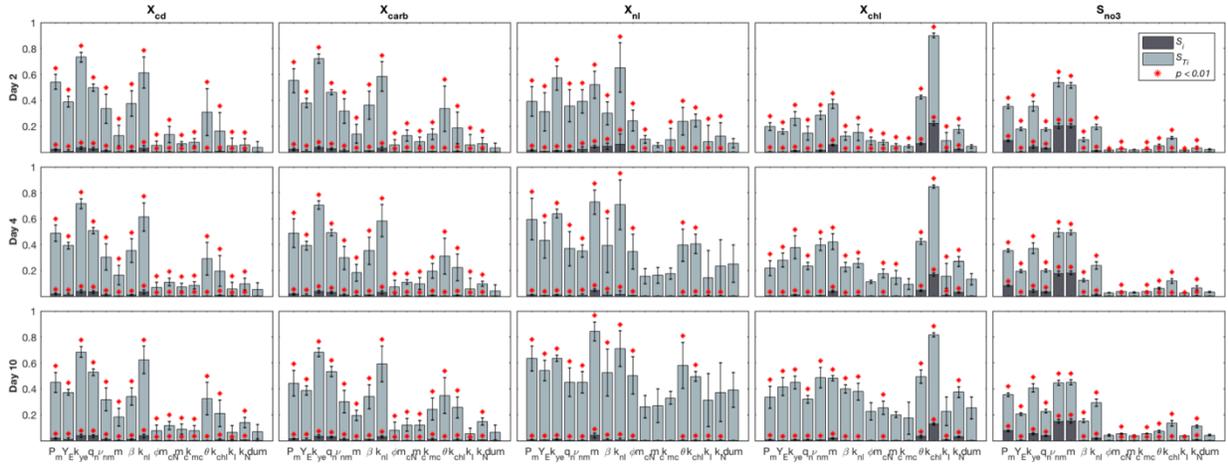
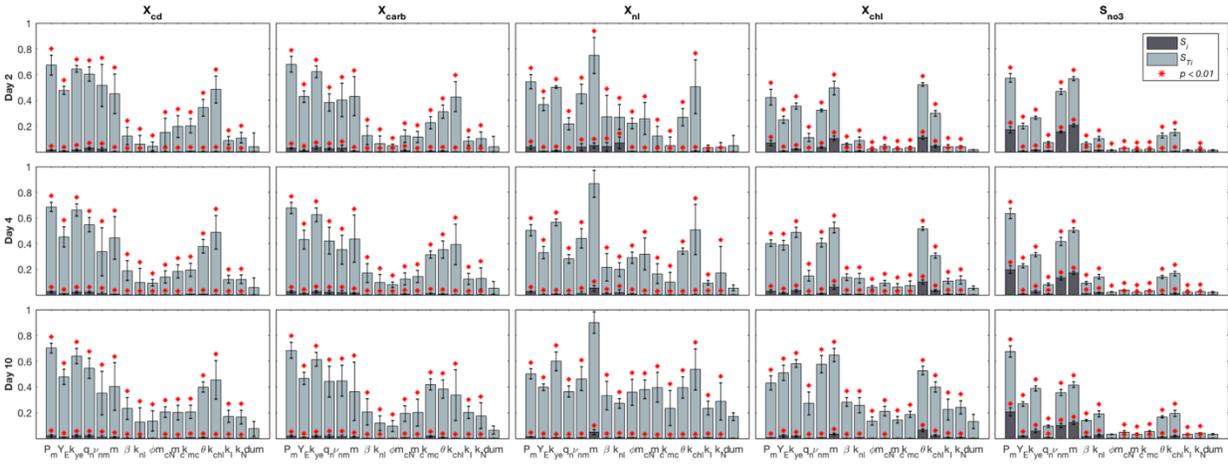
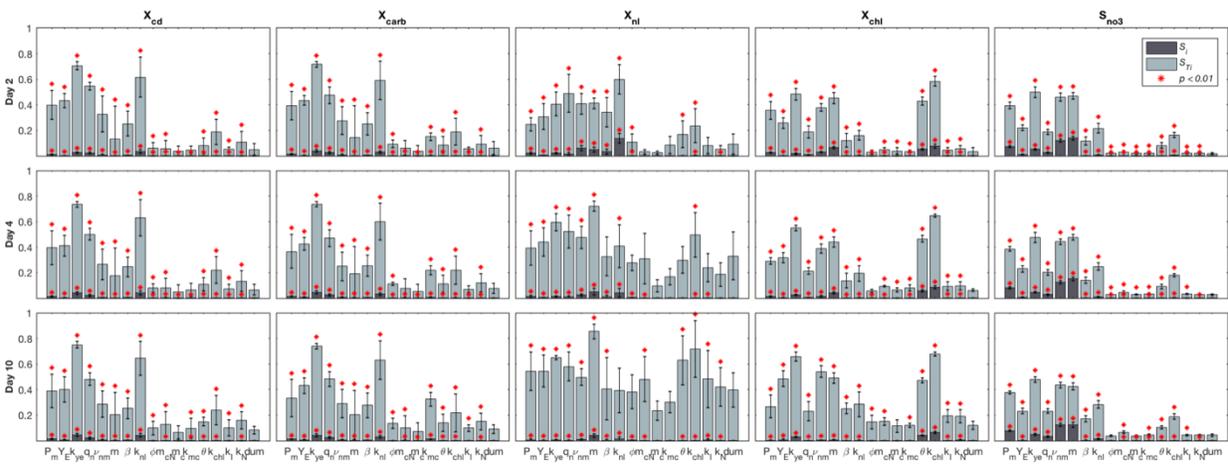


Figure S10. Sensitivity index plots under different conditions ($N_s = 2000$).

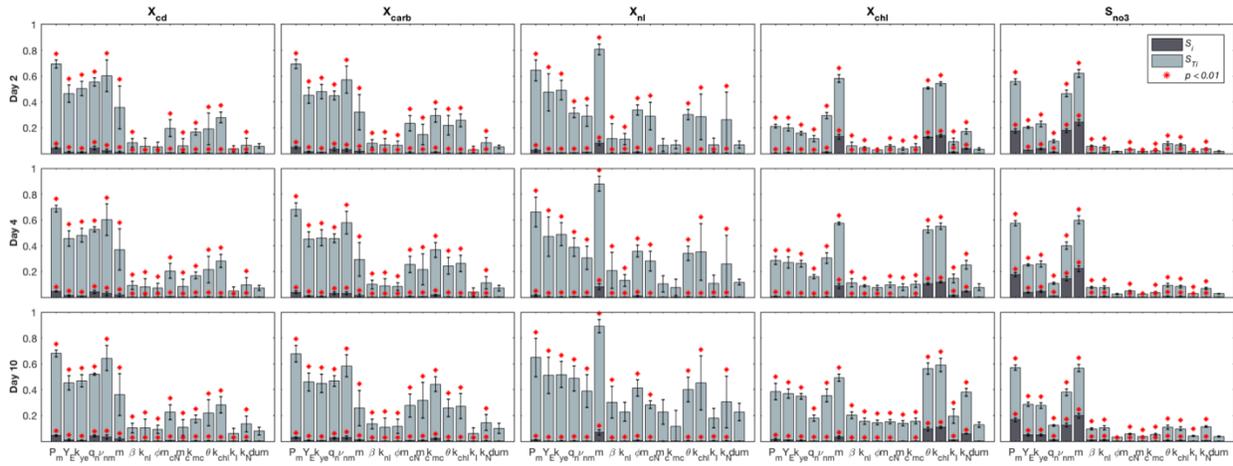
(a) $N_s = 3000$, No Stress (400HN)



(b) $N_s = 3000$, Single light stress (600HN)



(c) $N_s = 3000$, Single N stress (300LN)



(d) $N_s = 3000$, Dual stress (600LN)

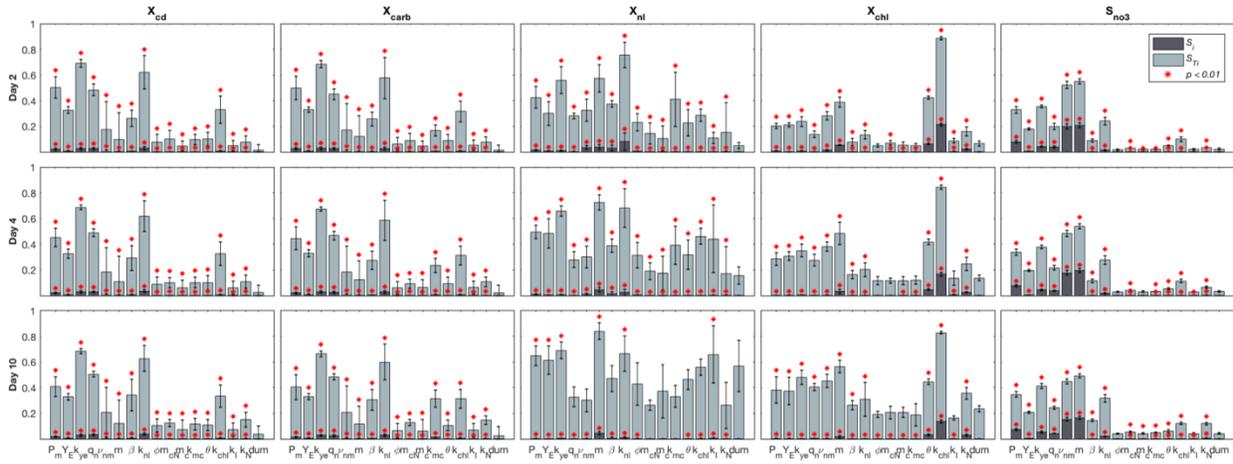


Figure S11. Sensitivity index plots under different conditions ($N_s = 3000$).

S-3.4 Light estimation performance

The light estimations of I_{avg} and I_{zl} were predicted by using the best-fit parameter set with the maximum likelihood $\pi(y|\theta)$.

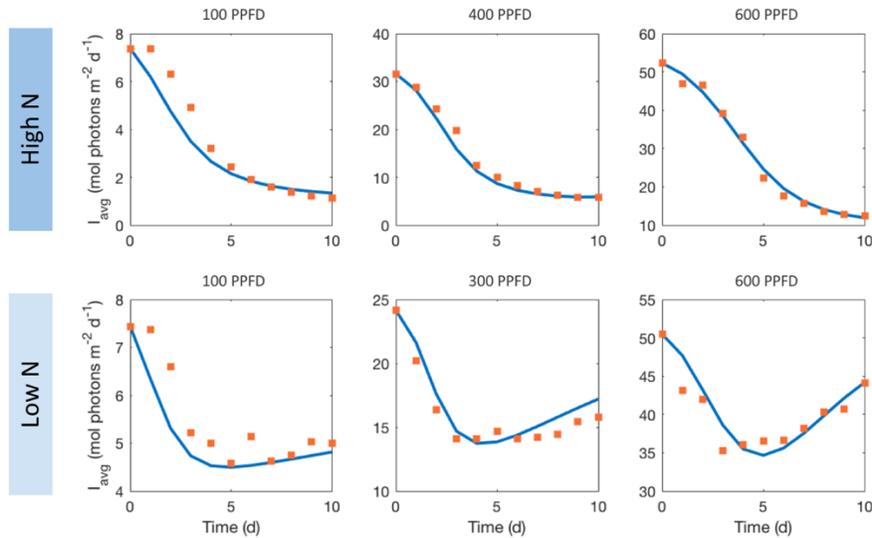


Figure S12. Comparison of I_{avg} estimations using the experimental data (squares) and the model-simulated data (solid lines) of X_{chl} and X_C .

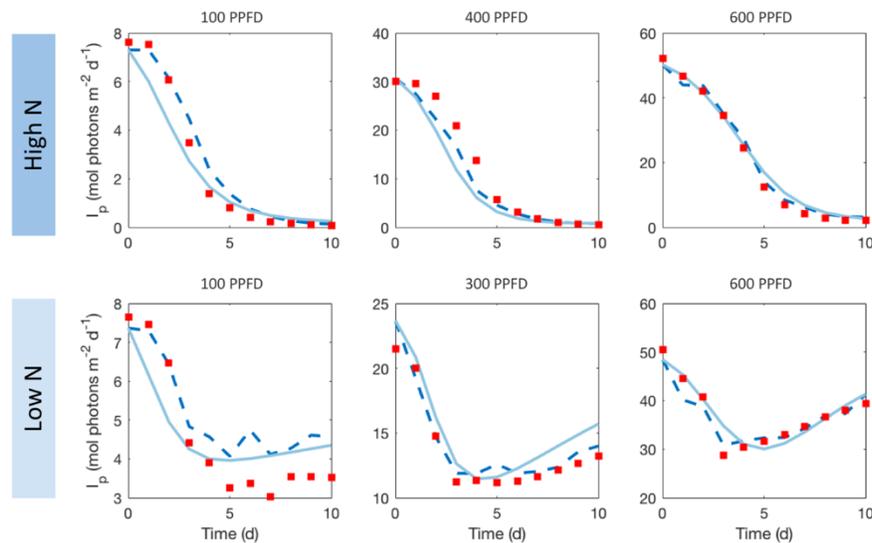


Figure S13. Comparison of model-predicted light profiles at location $l = zl$ (solid lines) with both experimental estimation (dashed lines) and light sensor measurements (squares).

S-4 Model validation results

We simulated five more experiments that were not used for calibration to validate the model. Similarly, we sampled 2000 parameter sets from the parameter posterior distributions and established the prediction intervals based on the corresponding error distributions. These experiments contained data for two light intensities (300 PPFD, 400 PPFD) and four nitrate

concentrations (4.90 N, 3.80 N, 0.67 N, 0.43 N) as summarized in Table S1. Figure S14 compares the model predictions with the experimental data. Figure S15 and Figure S16 further demonstrate the model performance of light estimation using the best-fit parameter set with the maximum likelihood $\pi(y|\theta)$.

Note that some experiments were not predicted from day zero due to either the lack of the initial data or the initial culture being too diluted. Thus, we tested the model with data starting from day one for the experiments without a complete set of the initial values. For the diluted culture with longer lag phase, the model had to start from the day when the culture biomass achieved a level comparable to the first-day biomass concentrations in other experiments. As such, the assumption that the initial intracellular nitrogen $X_{N_0} = 0.2X_{cd_0}$ may not hold true for these experiments due to the variations of the initial states, which can thus contribute to the bias in model predictions.

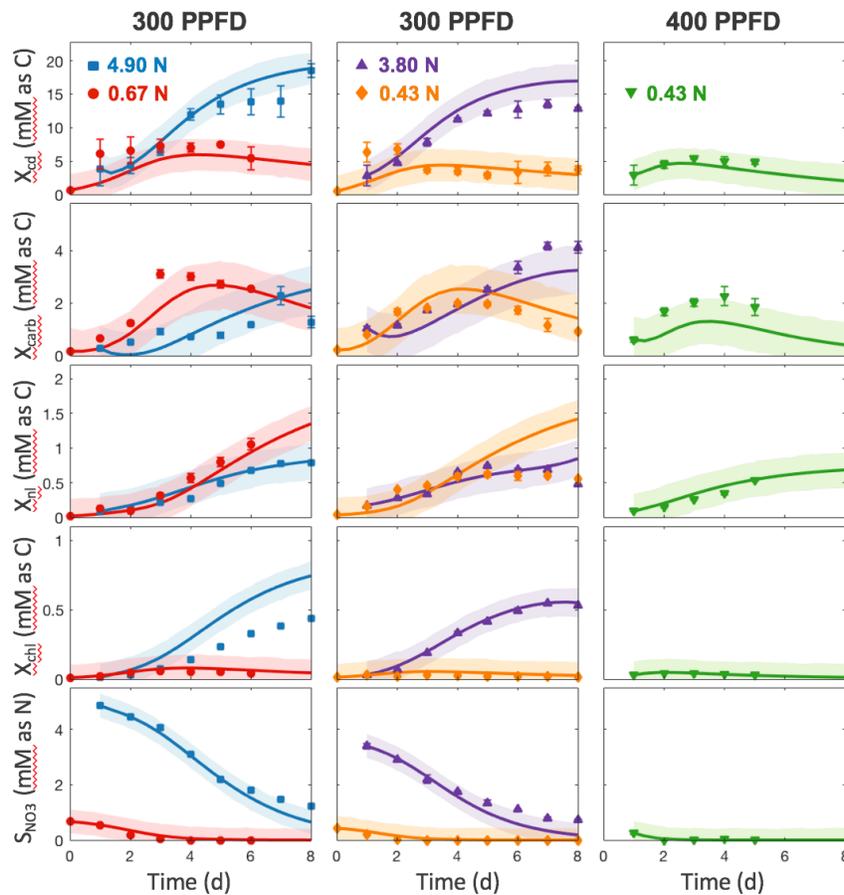


Figure S14. Model validation by predicting the new data. Solid lines correspond to the medians of model predictions bounded by 95% prediction intervals (shades). Symbols with error bars represent experimental data and measurement noises.

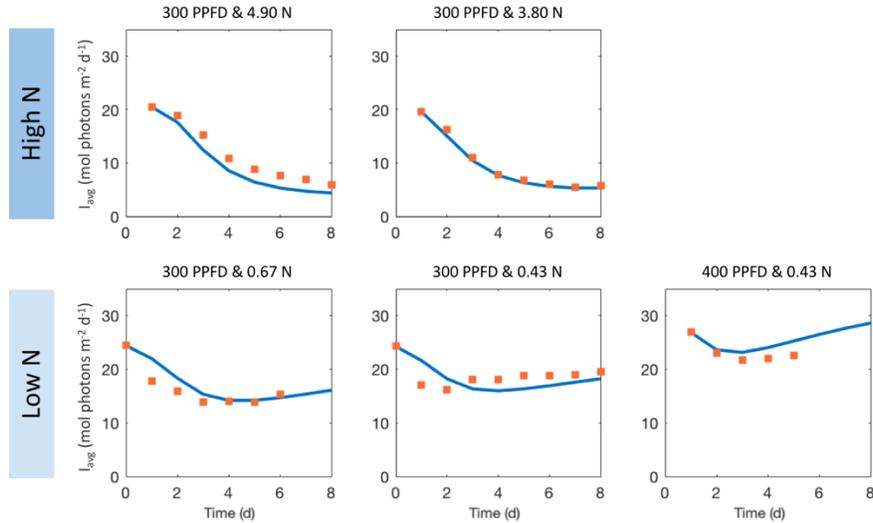


Figure S15. Model validation of I_{avg} predictions. Comparison of I_{avg} estimations using the experimental data (squares) and the model-simulated data (solid lines) of X_{chl} and X_C .

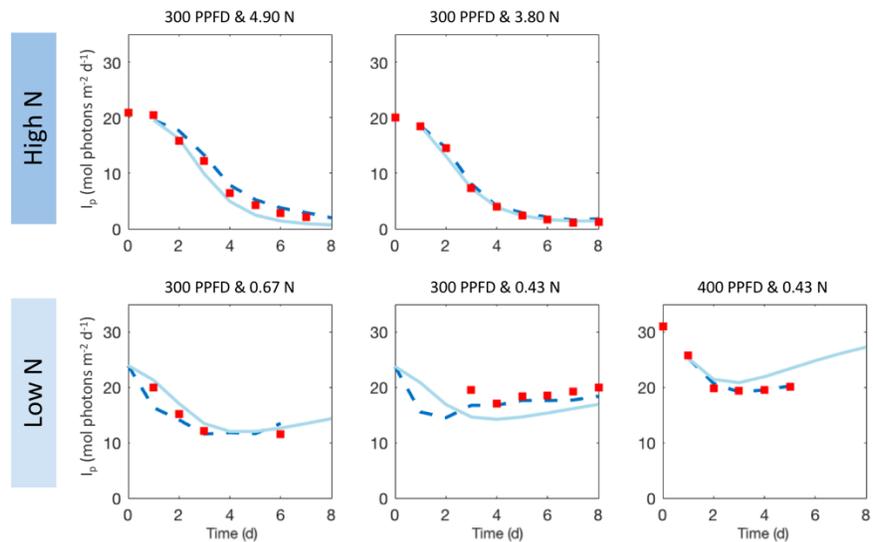


Figure S16. Model validation of I_{zl} predictions. Comparison of model-predicted light profiles at location $l = zl$ (solid lines) with both experimental estimation (dashed lines) and light sensor measurements (squares).

References

- (1) Raue, A.; Kreutz, C.; Maiwald, T.; Bachmann, J.; Schilling, M.; Klingmüller, U.; Timmer, J. Structural and Practical Identifiability Analysis of Partially Observed Dynamical Models by Exploiting the Profile Likelihood. *Bioinformatics* **2009**, *25* (15), 1923–1929.
- (2) Smith, R. C. *Uncertainty Quantification: Theory, Implementation, and Applications*; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 2013.
- (3) Haario, H.; Laine, M.; Mira, A.; Saksman, E. DRAM: Efficient Adaptive MCMC. *Stat. Comput.* **2006**, *16* (4), 339–354.
- (4) Geweke, J. Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of

- Posterior Moments. *Fed. Reserv. Bank Minneapolis, Res. Dep.* **1991**, 1–29.
- (5) Malve, O.; Laine, M.; Haario, H.; Kirkkala, T.; Sarvala, J. Bayesian Modelling of Algal Mass Occurrences—Using Adaptive MCMC Methods with a Lake Water Quality Model. *Environ. Model. Softw.* **2007**, *22* (7), 966–977.
 - (6) Marino, S.; Hogue, I. B.; Ray, C. J.; Kirschner, D. E. A Methodology for Performing Global Uncertainty and Sensitivity Analysis in Systems Biology. *J. Theor. Biol.* **2008**, *254* (1), 178–196.
 - (7) Ninnemann, H. Photoinhibition of Respiration in Bacteria and the Cyanophyceae *Vitreoscilla Stercoraria*. *FEBS Lett.* **1972**, *27* (1), 179–180.