

# Contents

<b>4 A data assimilating state-space model for algal growth under controlled conditions within a photo-bioreactor</b>	<b>2</b>
4.1 Introduction . . . . .	2
4.2 Methods . . . . .	4
4.2.1 Data Model: Photo-bioreactor setup, experimental design and data collection methods . . . . .	4
4.2.2 Data model: Data treatment, distributions and measurement error . . . . .	6
4.2.3 Process model: Carbon chemistry . . . . .	7
4.2.4 Process model: Gas transfer equilibrium concentrations for $O_2$ and $CO_2$ . . . . .	12
4.2.5 Process model: Flux into cells by photosynthesis and respiration . . . . .	13
4.2.6 Process model: Dilution . . . . .	14
4.2.7 Process model summary and parameter model . . . . .	15
4.2.8 Design and setup of data assimilation model with both simulated and experimental data . . . . .	16
4.3 Results and Discussion . . . . .	17
4.3.1 Posterior results with simulated data where photosynthesis and respiration are constant through time . . . . .	18

4.3.2	Posterior results with simulated data where photosynthesis and respiration are changing through time . . . . .	22
4.3.3	Posterior results with experimental data where photosynthesis, respiration and respiratory quotients are changing through time . . . . .	27
4.3.4	[To update] Posterior results with experimental data where photosynthesis, respiration and respiratory quotients are changing through time and an offset on $O_2$ is introduced . . .	33
4.3.5	[To update] Posterior results with experimental data where photosynthesis, respiration and respiratory quotients are changing through time, $O_2$ has an offset and the $O_2$ observations were thinned further . . . . .	39
4.3.6	[To update] Posterior results with experimental data where respiratory quotients are changing through time and bounded to biologically feasible values . . . . .	44
4.4	Discussion . . . . .	50
<b>A LiBbi model code</b>		<b>52</b>
<b>B Posteriors with experimental data (photosynthesis, respiration and respiratory quotients are random walks, and an offset, and estimating obs error)</b>		<b>66</b>
<b>Bibliography</b>		<b>73</b>

## Chapter 4

# A data assimilating state-space model for algal growth under controlled conditions within a photo-bioreactor

### 4.1 Introduction

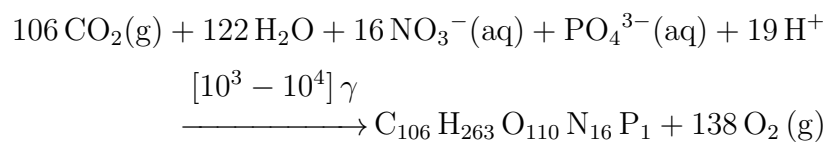
Microalgae has long been viewed as a potential platform for bioengineering. During the 1970's and 2000's much of the research focussed on biofuel production, while recently the potential for pharmaceuticals and other high-value chemicals has been explored. Whether the desired product is a primary or secondary metabolite it is important to be able to measure and the control the consumption and production.

- 'Fuse' high resolution measurements with more direct measure
- Incorporate prior information and constraints

The key goal of biofuels production is the optimisation of biomass productivity in large-scale microalgal culturing systems such as open ponds or closed photobioreactors. Primary goal is maximise the production raw biomass.

At the broadest scale / biggest picture the growth of microalgae can be expressed as the following chemical reaction in which C is taken from the atmosphere as  $\text{CO}_2$ , along with aqueous nutrients N and P, to produce biomass

species	conc
N	$880\mu M$
P	$36\mu M$
TA	$2300\mu M$



This equation captures a number of features of interest. The total biomass is determined by the concentrations of N and P. A typical growth media for microalgae is Guillard's Marine Enriched Seawater (F/2), which has

- **Biomass:** Taking the limiting nutrient to be Nitrogen, the maximum biomass of algae is  $200\text{mgL}^{-1}$ , of which about  $70\text{mgL}^{-1}$  is Carbon. This is equivalent 0.12L of pure CO<sub>2</sub> At Standard Temperature and Pressure, which around to 325L of air for each litre of media.
- **Change in alkalinity:** Alkalinity is central concept in seawater chemistry that allows one to calculate how much Carbon can be dissolved in solution, and what form it takes (Dissolved Inorganic Carbon exists as CO<sub>2</sub>, HCO<sub>3</sub><sup>-</sup> and CO<sub>3</sub><sup>2-</sup>). Alkalinity measures the charge imbalance between strongly and weakly dissociating ions in solution. As the charged nutrients are removed from the solution during growth the alkainty increases and more Carbon is able to be dissolved in the media.

## *Carbon in seawater*

Carbon and light availability are two of the most common limiting factors of biomass productivity [9].

## 4.2 Methods

### 4.2.1 Data Model: Photo-bioreactor setup, experimental design and data collection methods

All data collection methods for this chapter were part of a series of experiments examining microalgal responses to photobioreactor treatments (Peter Wood 2019 UTS PhD).

Microalgal culture *Nannochloropsis oceanica* (Droop) Green (strain CS-179) obtained from the Australian National Algae Culture Collection was cultured in 200 mL conical flasks; maintained in an incubator (Labec Pty Ltd) at 20°C, under an irradiance of  $50 \mu\text{mol m}^{-2} \text{s}^{-1}$  of cool-white fluorescent light at a 12 hour light/12 hour dark cycle. Stock cultures were grown in f/2 saltwater medium [6] and diluted 5 days prior to the start of experiments to ensure that *N. oceanica* was in the exponential growth phase and not nutrient deprived. f/2 was sparged prior to stock culture dilutions to maximise carbon and oxygen content.

*N. oceanica* was cultured in four, 500 mL environmental photo-bioreactors (eP-BRs, Phenometrics Inc) with a 10% v/v inoculation of stock culture. Top-side illumination over a path length of 25 cm was provided by a cool-white light LED, whilst temperature was maintained at 27°C using a Peltier heater-cooler connected to a water jacket. In-built thermocouples, calibrated against external temperature sensors attached to the Firesting module (TeX4; PyroScience GmbH), measured every 5 minutes were used to control the Peltier heater-cooler jacket through a feedback loop to an accuracy of  $\pm 0.2^\circ\text{C}$ . pH was also measured in 5 min intervals by in-built

pH electrodes (Van London Inc); controlled by periodic CO<sub>2</sub> (5%) injections using valves in the ePBRs. pH was 3-point calibrated using pH buffer solutions at pH  $4.00 \pm 0.02$ , pH  $7.00 \pm 0.02$  and pH  $10.00 \pm 0.02$ . PBR mixing was controlled by magnetic stirring bars at 110 rpm. All four ePBRs were aerated with filtered/humidified air through a 1.2 mm needle valve (Terumo Co).

A period of 2 days was allowed for *N. occulata* to acclimate to the ePBRs at an irradiance of  $500 \mu\text{mol m}^{-2} \text{s}^{-1}$  and a temperature of 27°C. Following this acclimation period, the ePBR was set to the experimental condition of  $2,000 \mu\text{mol photons m}^{-2} \text{s}^{-1}$  for another 2 days and a 12 hour light/12 hour dark cycle with a temperature of 27°C. ePBRs were maintained at an optical density (OD) of 0.4 using manual dilutions, creating a semi-batch culturing system. Dilutions occurred once per day (one hour before the light cycle), using aerated f/2 media. The experiment was conducted over a period of 4 days, samples were extracted post and prior dilution, as well as half way through the light cycle. 50 mL was extracted to examine total alkalinity and dissolved inorganic carbon. Dissolved oxygen (DO) was measured using a 3 mm robust optical probe (OXROB10-OI; PyroScience GmbH) attached to a FireStingO2 logger (PyroScience GmbH). DO measurements were taken every 60 seconds and temperature-corrected using a temperature extension module (TeX4; PyroScience GmbH). DO was two-point calibrated using air-saturated seawater (100% saturation) and sodium sulfate-saturated water (0% saturation). At 2 hour intervals, a solenoid valve (SMC Pneumatics Pty. Ltd.) was used to stop aeration for 10 minutes to allow for observations of net photosynthesis.

Alkalinity and DIC was measured twice a day closely following the Standard Operating Procedures (SOP) outline in [3]. Approximately 30 mL of *N. oceanica* media was titrated against 0.1 M hydrochloric acid on an auto-titrator (800 Dosino; Metrohm AG) [SOP3b Open-cell titration]. The Total Alkalinity and DIC were calculated from the output of the auto-titrator (volume of HCl delivered, pH) by

calculating the pH as a function of the volume of the acid delivered (see SOP3a Annexe 1).

#### 4.2.2 Data model: Data treatment, distributions and measurement error

Gas valve, temperature, light (normalised to 0/1) and dilution rates were used to force the model. Dissolved oxygen, pH, dissolved inorganic carbon and total alkalinity observations for 4 days post acclimation were assimilated. While pH observations were calibrated and corrected, it was visible that  $O_2$  observations were not completely calibrated and experienced some sensor drift during the experiment.

The data model assigned log normally distributed observation errors for each instrument;  $O_{2_{obs}} \sim \text{Log}\mathcal{N}(\log(O_2), \sigma_{O_2})$ ,  $pH_{obs} \sim \text{Log}\mathcal{N}(\log(pH), \sigma_{pH})$ ,  $DIC_{obs} \sim \text{Log}\mathcal{N}(\log(DIC), \sigma_{DIC})$ ,  $TA_{obs} \sim \text{Log}\mathcal{N}(\log(TA), \sigma_{DIC})$ , where the standard deviations ( $\sigma_{O_2}$ ,  $\sigma_{pH}$ ,  $\sigma_{DIC}$ ) were unknown parameters to be estimated as part of the assimilating model. Dissolved inorganic carbon and total alkalinity measurements were obtained from the same instrument thus the error is shared between these states.

Due to different instruments used in data collection,  $O_2$  observations were collected most frequently (6008 data-points), followed by  $pH$  observations (1179 data-points), with  $DIC$  and  $TA$  collected intermittently (11 data-points). Initial posterior runs with these observation densities resulted in the MCMC chain unable to mix properly due to the high density of observations. In an attempt to improve MCMC mixing, the denser observation sets were thinned/sampled down. The  $pH$  observations were thinned to approximately 5% of the full  $pH$  dataset by taking every 30th observation. The  $O_2$  observations were thinned to approximately 15% of the original  $O_2$  dataset by only taking those consecutive observations that had at least  $1 \mu\text{M L}^{-1}$  difference (Figure 4.1).

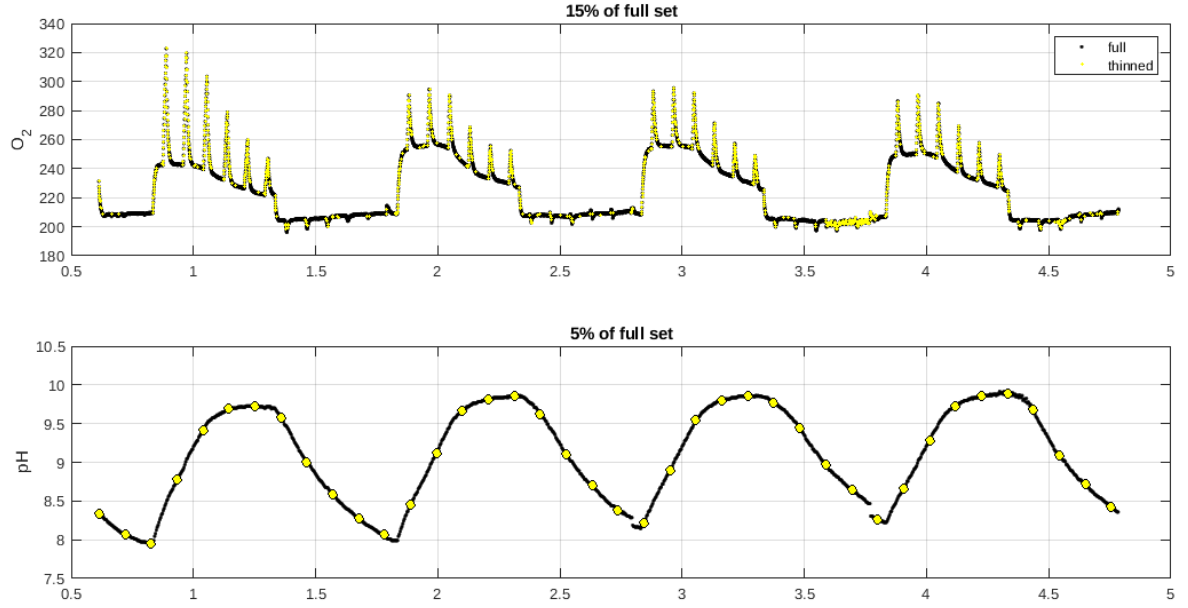


Figure 4.1 : Full  $O_2$  and  $pH$  datasets with thinned  $O_2$  and  $pH$  observations.

### 4.2.3 Process model: Carbon chemistry

[ Chris : Carbon in seawater summary ] main equations and rate constants role of TA This consists of 4(?) equations with (4?) unknowns

Ideally CO2SYS [7] would have been used to calculate the carbon chemistry of the photo-bioreactor. CO2SYS is a program developed for  $CO_2$  system calculations (CO2SYS) that calculates and returns a detailed state of the carbonate system of oceanographic water samples in seawater and freshwater [7]. A thorough explanation of CO2SYS, and seawater carbon chemistry in general, can be found in Zeebe and Wolf-gladrow [12]. A concise summary of the equations can be found in SOP 3a, Annexe 1 of [3].

Using two of the four measurable carbonate system parameters (total alkalinity, total inorganic  $CO_2$ , pH, and either fugacity  $fCO_2$  or partial pressure of  $CO_2$ ) to calculate the other two parameters at a set of input conditions (temperature and



pressure). In the model the state variables TA and DIC are given and the pH and pCO<sub>2</sub> are calculated.

To incorporate CO2SYS into LiBbi for solving carbon chemistry on the timescale of the microalgae model, we explicitly defined 3 iterations (Eq. 4.4 - 4.12) of the Newton-Raphson method for finding approximations to roots of real valued functions. The Newton-Raphson method is an iterative process  $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$  considering a function  $f(x_n)$ , its derivative  $f'(x_n)$  and an initial starting value  $x_0$ . The approximate root  $x_{n+1}$  converges to the exact solution very quickly if a close initial starting value is picked. To ensure the quick convergence of the Newton-Raphson method, an approximating equation for  $pH_0$  (Eq. 4.3) was obtained by fitting a stepwise regression with interactions to a range of simulated CO2SYS input parameters (temperature: 20-30, salinity: 30-40, DIC: 200-2500, and alkalinity: 1500-3000). A range of initial conditions and parameter values were tested, and each converged with RMSE < 0.01 across  $pH$ ,  $HCO_3^-$ ,  $CO_2$ , and  $CO_3$ , DIC,  $O_2$ , TA by the 3rd iteration (Figure 4.2 and Table 4.1).

### ***CO2SYS constants and iterative solution for pH, $HCO_3^-$ , $CO_2$ , and $CO_3$***

Total Sulfur:

$$\begin{aligned}
 TS &= \frac{0.14}{96.062} * \frac{S}{1.8065} \\
 IS &= 19.924 \frac{S}{(1000 - 1.005S)} \\
 KS_{int} &= -\frac{4276.1}{T_K} + 141.328 - 23.093 \log(T_K) + \left(-\frac{13856.0}{T_K} + 324.57 \right. \\
 &\quad \left. - 47.986 \log(T_K)\right) \sqrt{IS} + \left(\frac{35474}{T_K} - 771.54 + 114.723 \log(T_K)\right) IS \\
 &\quad - \frac{2698}{T_K} IS^{1.5} + \frac{1776}{T_K} IS^2 \\
 KS &= (1 - 0.001005S) e^{(KS_{int})}
 \end{aligned}$$

Fluorine:

$$TF = \frac{\frac{0.000067S}{18.9984}}{1.80655}$$

$$KF = e^{(-(-\frac{874.0}{T_K} - 0.111\sqrt{S} + 9.68))}$$

$$SW S_{2T} = \frac{(1 + \frac{TS}{KS})}{(1 + \frac{TS}{KS} + \frac{TF}{KF})}$$

$$Free_{2T} = 1 + \frac{TS}{KS}$$

H2O dissoc:

$$KW = e^{(148.9802 - \frac{13847.26}{T_K} - 23.6521 \log(T_K) + (\frac{118.67}{T_K} - 5.977 + 1.0495 \log(T_K))\sqrt{S} - 0.01615S)}$$

Boron:

$$KB = \exp\left(\frac{(-8966.90 - 2890.53\sqrt{S} - 77.942S + 1.728S\sqrt{S} - 0.0996S^2)}{T_K}\right)$$

$$+ 148.0248 + 137.1942\sqrt{S} + 1.62142S$$

$$- (24.4344 + 25.085\sqrt{S} + 0.2474S)\log(T_K) + 0.053105T_K\sqrt{S})$$

$$TB = 0.0004326 \frac{S}{35}$$

Choice of carbonate dissociation constants  $K_1$  and  $K_2$  were Mehrbach [8] (refit by Dickson and Millero [2]) with  $1.23K_1$  and  $0.53K_2$  measured experiment specific adjustments:

$$K_1 = 10^{(-(\frac{3633.86}{T_K} - 61.2172 + 9.6777 \log(T_K) - 0.011555S + 0.0001152S^2))} * 1.23 \quad (4.1)$$

$$K_2 = 10^{(-(\frac{471.8}{T_K} + 25.9290 - 3.16967 \log(T_K) - 0.01781S + 0.0001122S^2))} * 0.53 \quad (4.2)$$

Approximating equation for the starting value of  $pH$ :

$$pH_0 = 12.26 - 0.0030605DIC - 0.043752T - 0.013625S + 0.00011315TA$$

$$+ 1.3463e - 5DIC * T + 5.2215e - 7DIC * TA \quad (4.3)$$

Newton-Raphson iterations:

$$h_n = 10^{-pH_n} \quad (4.4)$$

$$h_{n_{free}} = \frac{h_n}{Free_{2T}} \quad (4.5)$$

$$\begin{aligned} f_n = & (DIC * 1e - 6 * \frac{K_1 h_n + 2K_1 K_2}{h_n^2 + K_1 h_n + K_1 K_2} \\ & - h_{n_{free}} + \frac{KW}{h_n} - TA * 1e - 6 + \frac{TB}{1 + \frac{h_n}{KB}}) * 1e6 \end{aligned} \quad (4.6)$$

$$\begin{aligned} df_n = & (DIC * 1e - 6 * \frac{K_1 + 2K_1 K_2}{h_n^2 + K_1 h_n + K_1 K_2} \\ & - DIC * 1e - 6 * \frac{(K_1 h_n + 2K_1 K_2)}{(h_n^2 + K_1 h_n + K_1 K_2)^2} (2h_n + K_1) \\ & - TB \frac{1}{(1 + \frac{h_n}{KB})^2} / KB \\ & - \frac{KW}{h_n^2} - \frac{1}{Free_{2T}}) * 1e6 * (-\log(10) * 10^{-pH}) \end{aligned} \quad (4.7)$$

$$pH_{n+1} = pH_n - \frac{f_n}{df_n} \quad (4.8)$$

$$H_{n+1} = 10^{-pH_{n+1}} \quad (4.9)$$

$$CO_{2n+1} = \frac{H_{n+1}^2 DIC}{H_{n+1}^2 + K_1 H_{n+1} + K_1 K_2} \quad (4.10)$$

$$HCO_{3n+1} = \frac{H_{n+1} K_1 DIC}{H_{n+1}^2 + K_1 H_{n+1} + K_1 K_2} \quad (4.11)$$

$$CO_{3n+1} = \frac{K_1 K_2 DIC}{H_{n+1}^2 + K_1 H_{n+1} + K_1 K_2} \quad (4.12)$$

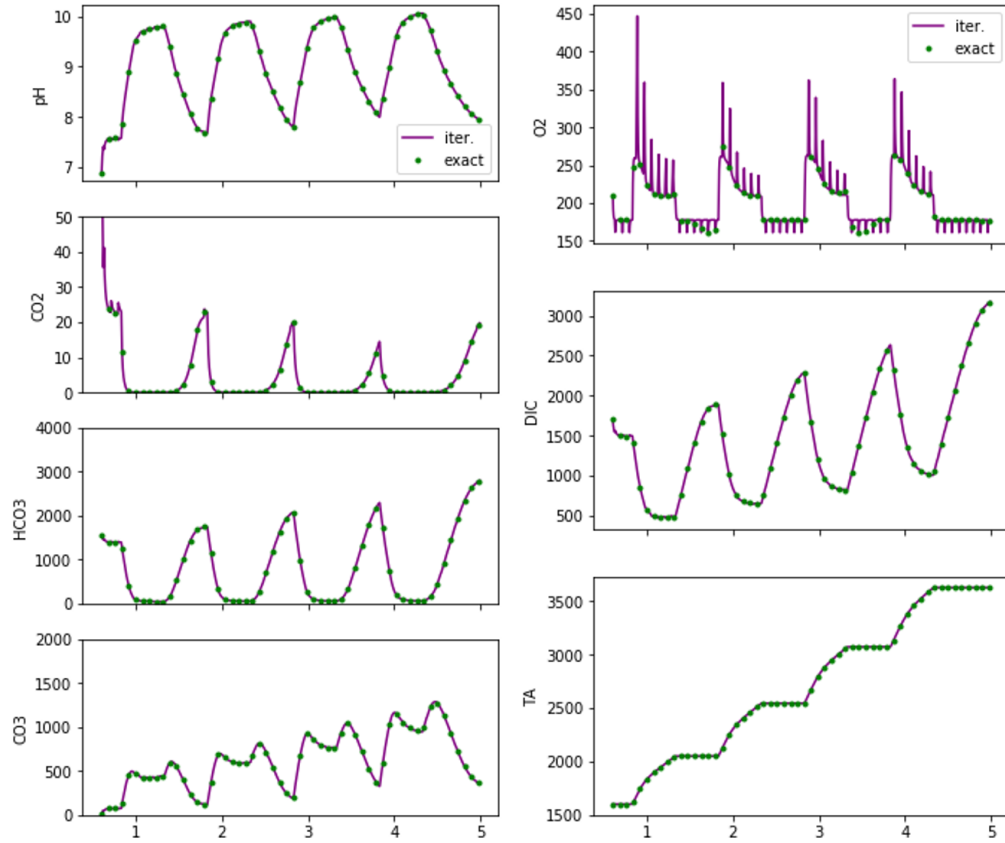


Figure 4.2 : Iterative (3rd iteration) vs exact solution for carbon chemistry  $CO_2$ ,  $HCO_3$ ,  $CO_3$ ,  $pH$  and state variables  $O_2$ ,  $DIC$ , and  $TA$ .

Variable	Iter. 1	Iter. 2	Iter. 3	Iter. 4	Iter. 5
$pH$	0.036092734	0.002355758	1.41E-05	6.93E-06	6.93E-06
$CO_2$	2.109401968	0.145719349	0.001222812	0.000866728	0.000866727
$HCO_3$	19.81869214	1.21021115	0.008016765	0.001025002	0.001025139
$CO_3$	20.89660704	1.307061652	0.00867642	0.001102278	0.001102434
$DIC$	16.78775711	0.958511825	0.005229318	0.002305054	0.002333411
$O_2$	0.308389964	0.016044284	4.18E-05	6.89E-05	7.59E-05
$TA$	2.607767674	0.160897272	0.000688102	0.001257725	0.001218981

Table 4.1 : RMSE for 5 iterations of the Newton-raphson carbon chemistry iterative solution.

#### 4.2.4 Process model: Gas transfer equilibrium concentrations for $O_2$ and $CO_2$

The equilibrium concentration for  $CO_2$  solubility in water  $CO_{2H}$  ( $\mu\text{mol/L}$ ) is calculated using Henry's law,

$$CO_{2H} = K0_{CO_2} * fCO_2 * 1.0220 * 1e6 \quad (4.13)$$

where  $fCO_2$  (atm) is the fugacity or approximately the partial pressure of  $CO_2$ , 1.0220 is the density of seawater (kg/L) at salinity 34 ppt and temperature  $27^\circ\text{C}$  [10] [4].  $K0_{CO_2}$  (mol/kg<sub>soln</sub>/atm) is the solubility of gas in seawater [BM: ask Chris: solubility of gas? is this right] and is calculated from the fitted van't Hoff equation and the logarithmic Setchenow salinity dependence [11],

$$K0_{CO_2} = e^{(-60.2409 + 93.4517 \frac{100}{T_K} + 23.3585 * \log(\frac{T_K}{100}) + S(0.023517 - 0.023656 \frac{T_K}{100} + 0.0047036 (\frac{T_K}{100})^2))} \quad (4.14)$$

where  $T_K$  is the temperature (K) and  $S$  is salinity (ppt). Similarly the equilibrium concentration for  $O_2$  solubility in water  $O_{2H}$  is calculated using Henry's law,

$$O_{2H} = K0_{O_2} * fO_2 * 1.0220 * 1e - 6 \quad (4.15)$$

where  $fO_2$  (atm) is the fugacity or approximately the partial pressure of  $O_2$ , 1.0220 is the density of seawater (kg/L) at salinity 34 ppt and temperature  $27^\circ\text{C}$  [10] [4], and  $K0_{O_2}$  (mol/kg<sub>soln</sub>/atm) is the solubility of oxygen in seawater with an adjusted salinity dependence [1],

$$K0_{O_2} = \frac{e^{(-1282.8704 + \frac{36619.96}{T_K} + 223.1396 \log(T_K) - 0.354707 T_K + S(5.957e - 3 - \frac{3.7353}{T_K}) + 3.68e - 6 S^2)}}{0.2094e - 6} \quad (4.16)$$

where  $T_K$  is the temperature (K) and  $S$  is salinity (ppt). The equilibrium concentrations for  $O_2$  and  $CO_2$  are modelled together with the gas turning on and off during the experiment, as

$$Q^{air} kLa_{O_2}^{air} (O_{2H} - O_2) \quad (4.17)$$

$$Q^{air} kLa_{CO_2}^{air} (CO_{2H} - CO_2) \quad (4.18)$$

where  $Q^{air}$  is the gas state (1= on, 0= off),  $kLa_{O_2}^{air}$  and  $kLa_{CO_2}^{air}$  are the mass transfer coefficients for air ( $d^{-1}$ ), and 0.893 is the ratio between measured  $O_2$  and  $CO_2$  mass transfer constants [5].

#### 4.2.5 Process model: Flux into cells by photosynthesis and respiration

The carbon flux into cells is measured by net photosynthesis,

$$\frac{\partial DIC}{\partial t} = -(P_1 I \frac{HCO_3^-}{K_m + HCO_3^-} - R_1) \quad (4.19)$$

$$\frac{\partial O_2}{\partial t} = \frac{1}{(RQ_d I + RQ_n(1 - I))} (P_1 I \frac{HCO_3^-}{K_m + HCO_3^-} - R_1) \quad (4.20)$$

$$\frac{\partial TA}{\partial t} = R_R (P_1 I \frac{HCO_3^-}{K_m + HCO_3^-}) \quad (4.21)$$

$$\frac{\partial C_p}{\partial t} = (P_1 I \frac{HCO_3^-}{K_m + HCO_3^-} - R_1) \quad (4.22)$$

where the parameters common to all states are  $P_1$ , the maximum photosynthesis rate ( $\mu M L^{-1} \text{ hour}^{-1}$ ),  $I$  the light indicator (0/1), and the michaelis menton term  $\frac{HCO_3^-}{K_m + HCO_3^-}$  representing the photosynthetically active carbon used for photosynthesis. This can be  $CO_2$ ,  $HCO_3^-$ , or a combination of both,  $CO_2 + HCO_3^-$  if the microalgae are using both carbon dioxide and bicarbonate for photosynthesis.  $K_m$  is the carbon restriction term ( $\mu M L^{-1}$ ).

The respiration rate  $R_1$  ( $\mu M L^{-1} \text{ hour}^{-1}$ ), is present in the net photosynthesis calculation for dissolved inorganic carbon ( $DIC$ ), oxygen ( $O_2$ ), and carbon concentration in the form of cells ( $C_p$ ). Oxygen also accounts for the day ( $RQ_d$ ) and night respiratory quotients ( $RQ_n$ ), the ratio of  $CO_2$  produced and  $O_2$  consumed by a cell. Total alkalinity ( $TA$ ) only increases due to photosynthesis while accounting for the Redfield ratio ( $R_R$ ).

[BM: Chris do I need more explanation here? Or references?]

#### 4.2.6 Process model: Dilution

To maintain a semi-batch culturing system, manual dilutions that occurred once per day over the period of the experiment were extracted to measure total alkalinity and dissolved inorganic carbon (described in more detail in Section 4.2.1). This also keeps the algae in the growth stage of the biomass curve, resetting it every day so that it never reaches the logit plateau. In the ODE's this affects every state variable, dissolved organic carbon ( $DIC$ ), oxygen ( $O_2$ ), total alkalinity ( $TA$ ) and amount of carbon in the form of cells ( $C_p$ ),

$$\frac{\partial DIC}{\partial t} = \frac{Q^M}{V}(DIC^M - DIC) \quad (4.23)$$

$$\frac{\partial O_2}{\partial t} = \frac{Q^M}{V}(O_2^M - O_2) \quad (4.24)$$

$$\frac{\partial TA}{\partial t} = \frac{Q^M}{V}(TA^M - TA) \quad (4.25)$$

$$\frac{\partial C_p}{\partial t} = \frac{Q^M}{V}(\quad - C_p) \quad (4.26)$$

where  $Q^M$  is the measured dilution rate ( $\text{ml day}^{-1}$ ) used to force the model,  $V$  is the volume of the photo-bioreactor fixed at 500 ml, and  $DIC^M$ ,  $O_2^M$ , and  $TA^M$  are the respective concentrations of the media. The media concentrations were calculated using CO2SYS (at temperature = 27 and salinity = 34) and set to be fixed throughout the experiment as  $DIC^M = 1724.20 \mu\text{M L}^{-1}$ ,  $O_2^M = 226.65 \mu\text{M L}^{-1}$ , and  $TA^M = 1797.90 \mu\text{M L}^{-1}$ .

#### 4.2.7 Process model summary and parameter model

A summary of the ODE's that make up the process model described in previous sections:

$$\begin{aligned}
 \frac{\partial DIC}{\partial t} &= \text{Rate} & \text{flux into cells} & \text{gas transfer} & \text{dilution} \\
 &= -(P_1 I \frac{HCO_3^-}{K_m + HCO_3^-} - R_1) & + \hat{Q}^{air} kLa_{CO_2}^{air} (CO_2^{air} - CO_2) & + \frac{Q^M}{V} (DIC^M - DIC) \\
 \frac{\partial O_2}{\partial t} &= \frac{(P_1 I \frac{HCO_3^-}{K_m + HCO_3^-} - R_1)}{(RQ_d I + RQ_n(1 - I))} & + \hat{Q}^{air} kLa_{O_2}^{air} (O_2^{air} - O_2) & + \frac{Q^M}{V} (O_2^M - O_2) \\
 \frac{\partial TA}{\partial t} &= R_R (P_1 I \frac{HCO_3^-}{K_m + HCO_3^-}) & & + \frac{Q^M}{V} (TA^M - TA) \\
 \frac{\partial C_p}{\partial t} &= (P_1 I \frac{HCO_3^-}{K_m + HCO_3^-} - R_1) & & + \frac{Q^M}{V} (-C_p)
 \end{aligned} \tag{4.27}$$

	Symbol	Description	Prior / Value	Unit
Initial conditions	$DIC^0$	Dissolved inorganic carbon	$\text{Log}\mathcal{N}(\log(1300), 0.2)$	$\mu\text{M L}^{-1}$
	$O_2^0$	Oxygen	$\text{Log}\mathcal{N}(\log(225), 0.2)$	$\mu\text{M L}^{-1}$
	$TA^0$	Total alkalinity	$\text{Log}\mathcal{N}(\log(1750), 0.1)$	$\mu\text{M L}^{-1}$
	$C_p^0$	Carbon in the form of cells	$\text{Log}\mathcal{N}(\log(300), 0.2)$	$\mu\text{M L}^{-1}$
	$pH^0$	-	$\text{Log}\mathcal{N}(\log(8.5), 0.2)$	$\log_{10}(-\text{mol/L H}^+)$
	$CO_2^0$	Carbon dioxide	$\text{Log}\mathcal{N}(\log(5), 0.4)$	$\mu\text{M L}^{-1}$
	$HCO_3^{-0}$	Bicarbonate	$\text{Log}\mathcal{N}(\log(1500), 0.3)$	$\mu\text{M L}^{-1}$
	$CO_3^{2-0}$	Carbonate	$\text{Log}\mathcal{N}(\log(100), 0.4)$	$\mu\text{M L}^{-1}$
Flux into cells	$P_1$	Maximum photosynthesis rate	*	$\mu\text{M L}^{-1} \text{ hour}^{-1}$
	$R_1$	Respiration rate	*	$\mu\text{M L}^{-1} \text{ hour}^{-1}$
	$K_m$	Carbon restriction	*	$\mu\text{M L}^{-1}$
	$RQ_d$	Daytime respiratory quotient	*	-
	$RQ_n$	Night respiratory quotient	*	-
	$R_R$	Redfield ratio	*	-



	$I$	Light indicator	forcing (0/1)	-
Gas transfer terms	$\hat{Q}^{air}$	Indicator for flow in air line	forcing (0/1)	-
	$x_{CO_2}^{air}$	Mole fraction of CO <sub>2</sub> atmosphere	400	ppm
	$CO_{2H}$	Equilibrium CO <sub>2</sub> concentration	Eq. 4.13	$\mu\text{M L}^{-1}$
	$CO_2^{air}$	Sat CO <sub>2</sub> conc with atmosphere	$x_{CO_2}^{air} CO_{2H}$	
	$kLa_{CO_2}^{air}$	Mass transfer coefficient for CO <sub>2</sub>	$0.893kLa_{O_2}^{air}$	$\text{day}^{-1}$
	$x_{O_2}^{air}$	Mole fraction of O <sub>2</sub> atmosphere	0.2094	atm
	$O_{2H}$	Equilibrium O <sub>2</sub> concentration	Eq. 4.15	$\mu\text{M L}^{-1}$
	$O_2^{air}$	Sat O <sub>2</sub> conc with atmosphere	$x_{O_2}^{air} O_{2H}$	
	$\tau$	half-life of $kLa_{O_2}^{air}$	range(2-20)	$\text{min}^{-1}$
	$kLa_{O_2}^{air}$	Mass transfer coefficient for O <sub>2</sub>	$\ln(2) * 24 * 60 / \tau$	$\text{day}^{-1}$
Dilution terms	$Q^M$	Dilution rate	forcing	$\text{ml day}^{-1}$
	$V$	Volume of the reactor	500	ml
	$DIC^M$	Media dissolved inorganic carbon	1724.20	$\mu\text{M L}^{-1}$
	$O_2^M$	Media oxygen concentration	226.65	$\mu\text{M L}^{-1}$
	$TA^M$	Media total alkalinity	1797.90	$\mu\text{M L}^{-1}$

Table 4.2 : State variable, parameter and forcing definitions with units, and their assignments: either fixed values, priors on initial condition or priors on parameters (\*) defined later in respective sections.

#### 4.2.8 Design and setup of data assimilation model with both simulated and experimental data

A simulated dataset was created by running the process model described in Section 4.2.7 with a fixed set of parameters ( $P_1 = 200$ ,  $R_1 = 30$ ,  $kLa_{O_2}^{air} = 200$ ,  $K_m = 150$ ,  $RQ_d = 0.85$ ,  $RQ_n = 0.95$ ,  $R_R = 0.075$ ) and initial conditions ( $O_2^0 = 225$ ,  $DIC^0 = 1250$ ,  $TA^0 = 1750$ ). [BM: Chris- noise on simulated dataset? sigmas?]

The set of experimental runs with simulated observations are

1. All parameters constant through time.
2. P and R as random walks through time.

The first set of results assimilated the simulated observations and tested how well the system recovered the true parameter values (Section 4.3.1 and 4.3.2). Then the following results assimilated the experimental data with various formulations of model parameters (Section ?? and ??). [BM: update these to their ref labels to reference each section]

The set of experimental runs with experimental data are

1. All parameters constant through time.
2. P and R as random walks through time.
3. P and R as random walks through time, with daily and nightly respiratory quotients as random walks through time.
4. Adding in an offset.
5. Thinning out the observations further.
6. Constraining the respiratory quotients inside biologically justifiable bounds.

Each result section will include the treatment of parameters, priors and proposal distributions for clarity.

### 4.3 Results and Discussion

State posteriors are visualised by plotting the median and shading 95% credible intervals, while parameter priors and posteriors are displayed by histograms. Interval estimates (25%, 75%) and (5%, 95%) of parameter posterior distributions are also quoted. Where simulated data was assimilated the true parameter values are

indicated by solid green lines.

#### 4.3.1 Posterior results with simulated data where photosynthesis and respiration are constant through time

$P_1$ ,  $R_1$ ,  $kLA_{O_2}$ ,  $K_m$ ,  $R_R$ ,  $RQ_d$ , and  $RQ_n$  were all treated as parameters constant through time but unknown with prior distributions and proposal distributions defined in Table 4.3. The other model parameters were defined earlier in Table 4.2.7. The data model assigned log normally distributed observation errors for each instrument (Section 4.2.2) with observation error standard deviation values cited in Table 4.3.

Parameter	Prior	Proposal
$P_1$	$\text{Log}\mathcal{N}(\log(250.0), 0.8)$	$\text{Log}\mathcal{N}(\log(P_1), 0.08)$
$R_1$	$\text{Log}\mathcal{N}(\log(20.0), 0.8)$	$\text{Log}\mathcal{N}(\log(R_1), 0.08)$
$kLA_{O_2}$	$\text{Log}\mathcal{N}(\log(200.0), 0.3)$	$\text{Log}\mathcal{N}(\log(kLA_{O_2}), 0.03)$
$K_m$	$\text{Log}\mathcal{N}(\log(200.0), 0.6)$	$\text{Log}\mathcal{N}(\log(K_m), 0.06)$
$R_R$	$\text{Uniform}(0, 0.2)$	$\text{Trun}\mathcal{N}(R_R, 0.01, \text{lower} = 0, \text{upper} = 0.2)$
$RQ_d$	$\text{Uniform}(0.6, 1)$	$\text{Trun}\mathcal{N}(RQ_d, 0.005, \text{lower} = 0.6, \text{upper} = 1.0)$
$RQ_n$	$\text{Uniform}(0.6, 1)$	$\text{Trun}\mathcal{N}(RQ_n, 0.005, \text{lower} = 0.6, \text{upper} = 1.0)$
$\sigma_{O_2}$	0.3	*
$\sigma_{pH}$	0.3	*
$\sigma_{DIC}$	0.3	*

Table 4.3 : Table of Parameters, their priors and proposal distributions (\* indicates the parameter was held fixed).

Parameter	Quantiles (25%, 75%)	Quantiles (5%, 95%)	True value
$P_1$	(237.2167, 302.0565)	(200.8971, 365.1751)	200
$R_1$	(17.5370, 32.3501)	(10.6537, 45.3096)	30

$kLA_{O_2}^{air}$	(177.1383, 222.9641)	(148.2838, 267.5934)	200
$K_m$	(182.6643, 362.6714)	(118.4125, 643.8518)	150
$R_R$	(0.0504, 0.1041)	(0.0196, 0.1458)	0.075
$RQ_d$	(0.7529, 0.8718)	(0.6894, 0.9666)	0.85
$RQ_n$	(0.7469, 0.9338)	(0.6400, 0.9920)	0.95

Table 4.4 : True values for parameters used to create the simulated observations, and posterior (25%, 75%), (5%, 95%) quantiles for parameters after assimilating observations.

After running the PMMH for 50,000 samples (10,000 discarded as burn-in) with 1024 particles, the true values of parameters  $R_1$ ,  $kLA_{O_2}^{air}$ ,  $R_R$ , and  $RQ_d$  were recovered to the 25th and 75th percentile (Table 4.4). While the true values of  $K_m$  and  $RQ_n$  did not lie within the 25th and 75th quantile, they were captured by the 5th and 95th percentiles (Table 4.4).  $P_1$  was the only parameter whose true value lied on the cusp of the 5th percentile. Parameters  $P_1$ ,  $R_1$ ,  $kLA_{O_2}^{air}$ , and  $R_R$  mixed well, while parameters  $K_m$ ,  $RQ_d$ , and  $RQ_n$  did not mix as well, some autocorrelation between samples was present (Figure A.3).

Observed state variables  $O_2$ ,  $DIC$ ,  $TA$ , and observed  $pH$  posteriors were in excellent agreement with observations, with all observations fitting within the 95% credible interval posteriors (Figure 4.3, Figure 4.4).

Unobserved state variable  $C_p$  .. (Figure 4.4).

Carbon chemistry variables  $HCO_3$ ,  $CO_2$ , and  $CO_3$  ... (Figure 4.5).

[BM: Chris what to say about  $C_p$  and carbon chem]

Log-likelihood stopped rising (Figure A.2). Acceptance rate was XX.

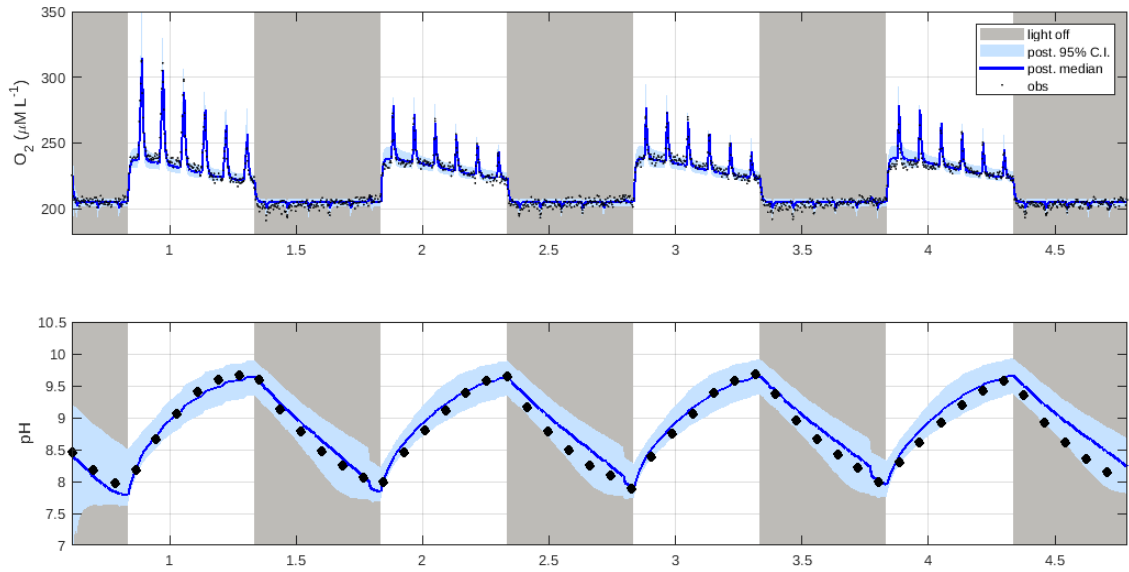


Figure 4.3 : Posterior medians (solid blue line), 95% credible intervals (shaded blue), and simulated observations (black) for  $O_2$  and  $pH$  across 4 days.

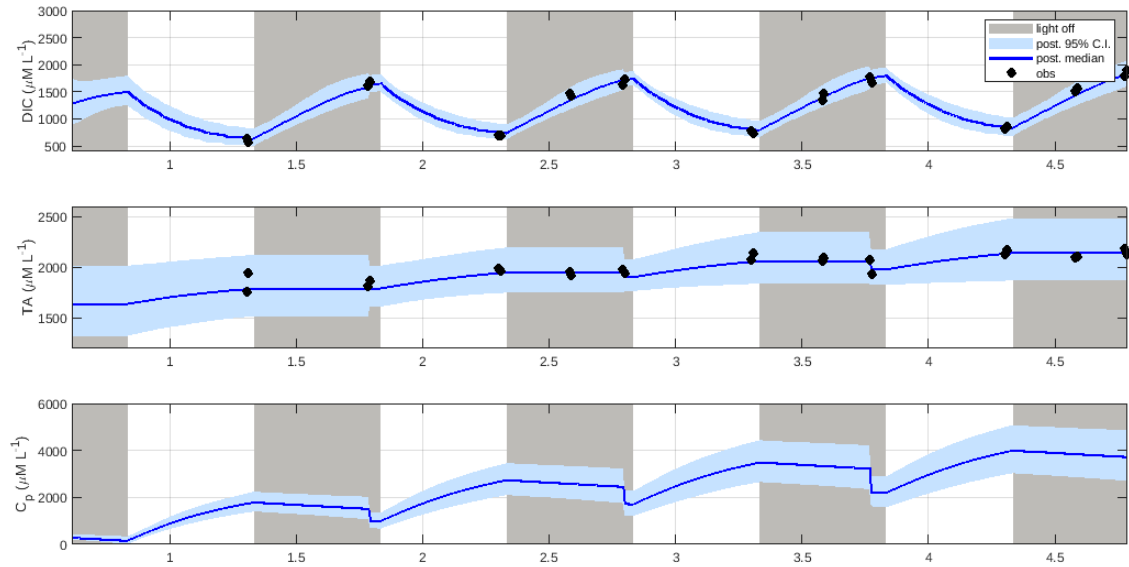


Figure 4.4 : Posterior medians (solid blue line), 95% credible intervals (shaded blue), and simulated observations (black) for  $DIC$ ,  $TA$  and  $C_p$  across 4 days.

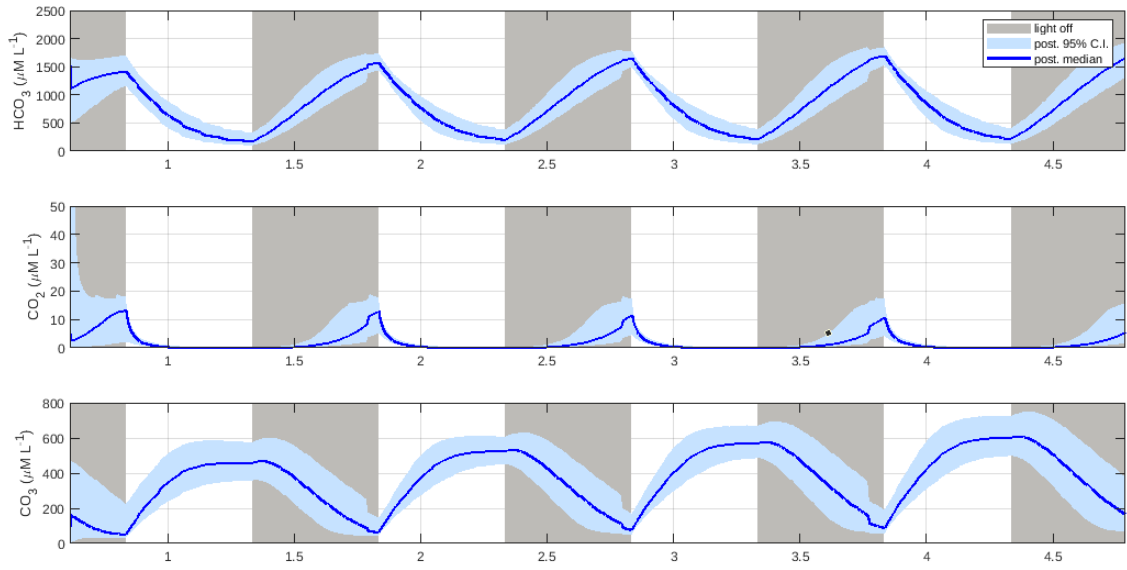


Figure 4.5 : Posterior medians (solid blue line) and 95% credible intervals (shaded blue) for  $HCO_3$ ,  $CO_2$  and  $CO_3$  across 4 days.

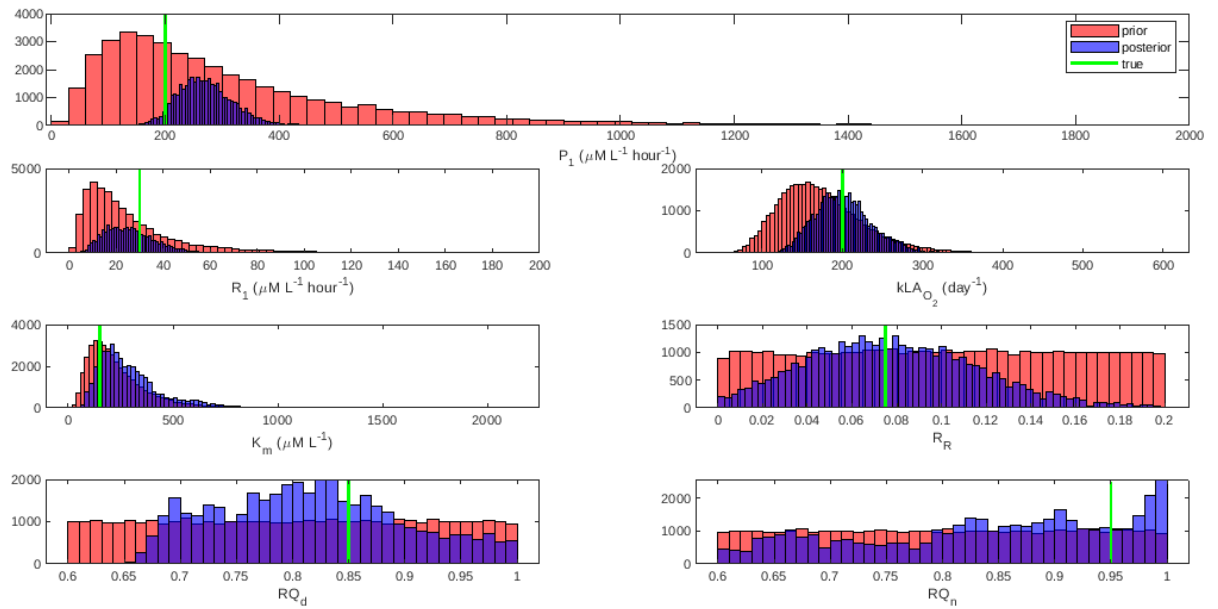


Figure 4.6 : Priors (pink), posteriors (purple) and true values (green) for model parameters.

### 4.3.2 Posterior results with simulated data where photosynthesis and respiration are changing through time

**In this run:**

$kLA_{O_2}$ ,  $K_m$ ,  $R_R$ ,  $RQ_d$ , and  $RQ_n$  were all treated as parameters constant through time but unknown with prior distributions and proposal distributions defined in Table 4.5. Photosynthesis ( $P_1$ ) and respiration ( $R_1$ ) were both modelled as random walks, by taking  $P$  and  $R$ , previously constant parameters, and replacing them by  $P_1(t)$  and  $R_1(t)$ . Here, we take  $P_1(t)$  and  $R_1(t)$  to be such that

$$P_1(t + \Delta t) = P(t) + r_P$$

$$R_1(t + \Delta t) = R(t) + r_R$$

where  $r_P \sim N(0, \sigma_{r_P})$ ,  $r_R \sim N(0, \sigma_{r_R})$ , and  $\Delta t$  is the length of discrete time-step. For the purpose of the Bayesian analysis here,  $\sigma_{r_P}$  and  $\sigma_{r_R}$  are treated as parameters to be inferred. The other model parameters were defined earlier in Table 4.2.7. The data model assigned log normally distributed observation errors for each instrument (Section 4.2.2) with observation error standard deviation values cited in Table 4.5.

Parameter	Prior	Proposal
$kLA_{O_2}$	$\text{Log}\mathcal{N}(\log(200.0), 0.3)$	$\text{Log}\mathcal{N}(\log(kLA_{O_2}), 0.03)$
$K_m$	$\text{Log}\mathcal{N}(\log(200.0), 0.6)$	$\text{Log}\mathcal{N}(\log(K_m), 0.06)$
$R_R$	$\text{Uniform}(0, 0.2)$	$\text{Trun}\mathcal{N}(R_R, 0.01, \text{lower} = 0, \text{upper} = 0.2)$
$RQ_d$	$\text{Uniform}(0.6, 1)$	$\text{Trun}\mathcal{N}(RQ_d, 0.005, \text{lower} = 0.6, \text{upper} = 1.0)$
$RQ_n$	$\text{Uniform}(0.6, 1)$	$\text{Trun}\mathcal{N}(RQ_n, 0.005, \text{lower} = 0.6, \text{upper} = 1.0)$
$\sigma_{r_P}$	$\mathcal{N}(0.01, 0.001)$	$\mathcal{N}(\sigma_{r_P}, 0.0001)$
$\sigma_{r_R}$	$\mathcal{N}(0.01, 0.001)$	$\mathcal{N}(\sigma_{r_R}, 0.0001)$
$\sigma_{O_2}$	0.3	*
$\sigma_{pH}$	0.3	*
$\sigma_{DIC}$	0.3	*

Table 4.5 : Table of Parameters, their priors and proposal distributions (\* indicates the parameter was held fixed).

Parameter	Quantiles (25%, 75%)	Quantiles (5%, 95%)	True value
$kLA_{O_2}^{air}$	(170.9854, 216.0578)	(145.6899, 253.4652)	200
$K_m$	(187.3586, 386.4103)	(93.9670, 641.1432)	150
$R_R$	(0.0521, 0.1075)	(0.0192, 0.1544)	0.075
$RQ_d$	(0.7906, 0.8965)	(0.6833, 0.9711)	0.85
$RQ_n$	(0.6654, 0.8334)	(0.6164, 0.9739)	0.95

Table 4.6 : True values for parameters used to create the simulated observations, and posterior (25%, 75%), (5%, 95%) quantiles for parameters after assimilating observations.

After running the PMMH for 50,000 samples (10,000 discarded as burn-in) with 1024 particles, the true values of parameters  $kLA_{O_2}^{air}$ ,  $R_R$ , and  $RQ_d$  were recovered to the 25th and 75th percentile (Table 4.6). While the true values of  $K_m$  and  $RQ_n$  did not lie within the 25th and 75th quantile, they were captured by the 5th and 95th percentiles (Table 4.6). Parameters  $\sigma_{rP}$ ,  $\sigma_{rR}$ ,  $kLA_{O_2}^{air}$ , and  $R_R$  mixed well, while parameters  $K_m$ ,  $RQ_d$ , and  $RQ_n$  did not mix as well, some autocorrelation between samples was present (Figure 4.12). The true values of  $P_1$  (200) and  $R_1$  (30) lie within the posterior 95% credible intervals (Figure 4.11).

Observed state variables  $O_2$ ,  $DIC$ ,  $TA$ , and observed  $pH$  posteriors were in excellent agreement with observations, with all observations fitting within the 95% credible interval posteriors (Figure 4.7, Figure 4.8).

Unobserved state variable  $C_p$  .. (Figure 4.8).



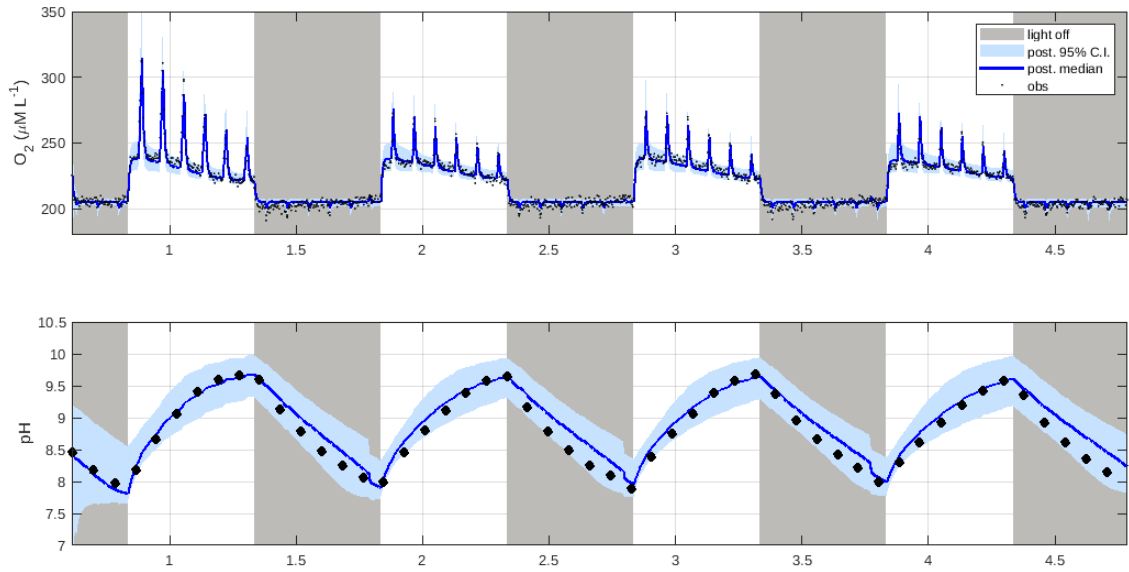


Figure 4.7 : Posterior medians (solid blue line), 95% credible intervals (shaded blue), and simulated observations (black) for  $O_2$  and  $pH$  across 4 days.

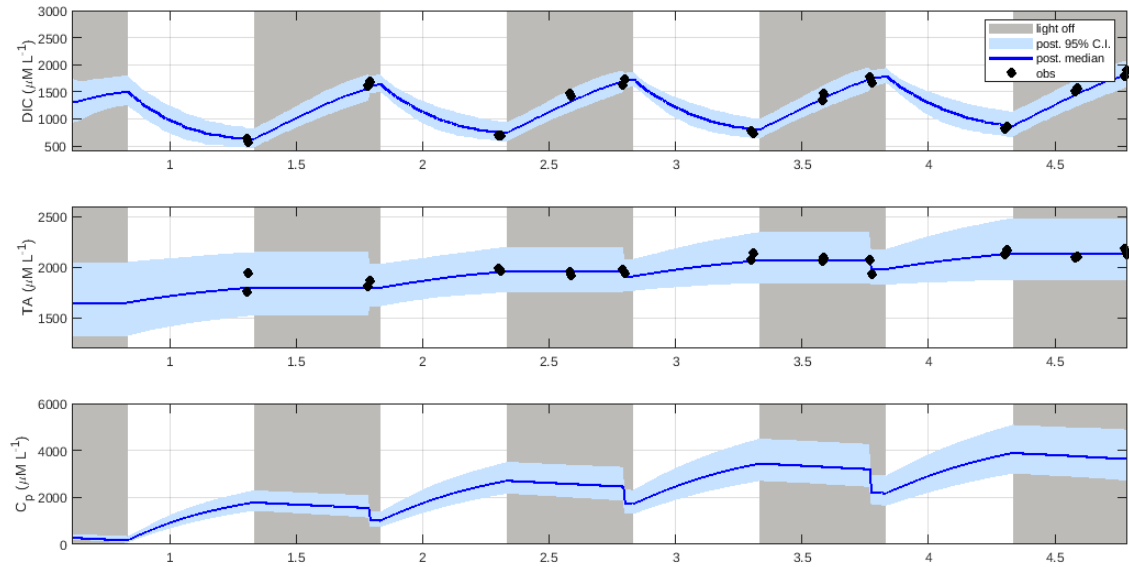


Figure 4.8 : Posterior medians (solid blue line), 95% credible intervals (shaded blue), and simulated observations (black) for  $DIC$ ,  $TA$  and  $C_p$  across 4 days.

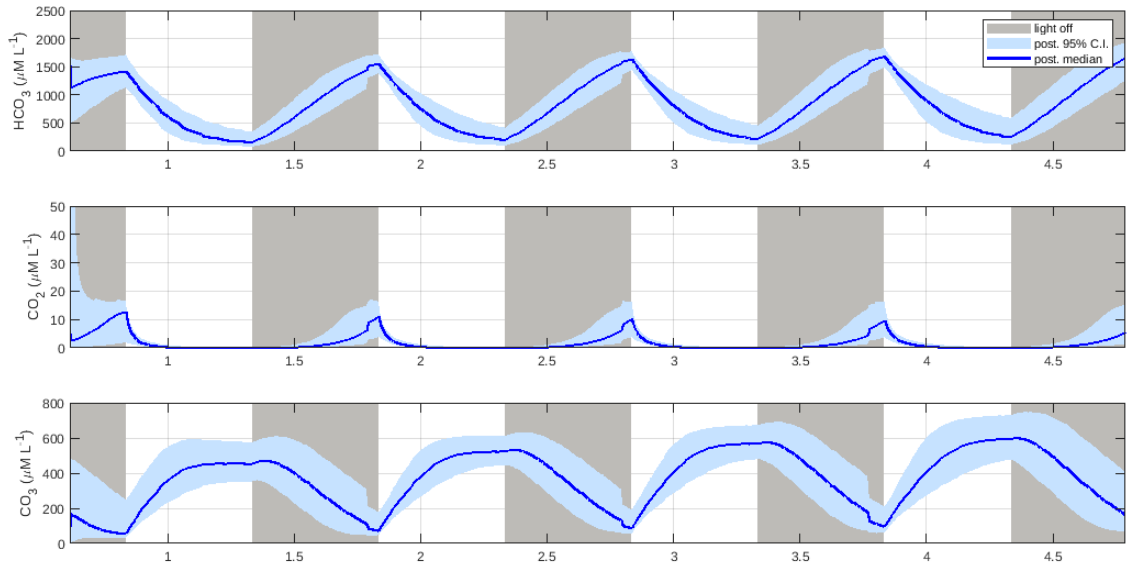


Figure 4.9 : Posterior medians (solid blue line) and 95% credible intervals (shaded blue) for  $HCO_3$ ,  $CO_2$  and  $CO_3$  across 4 days.

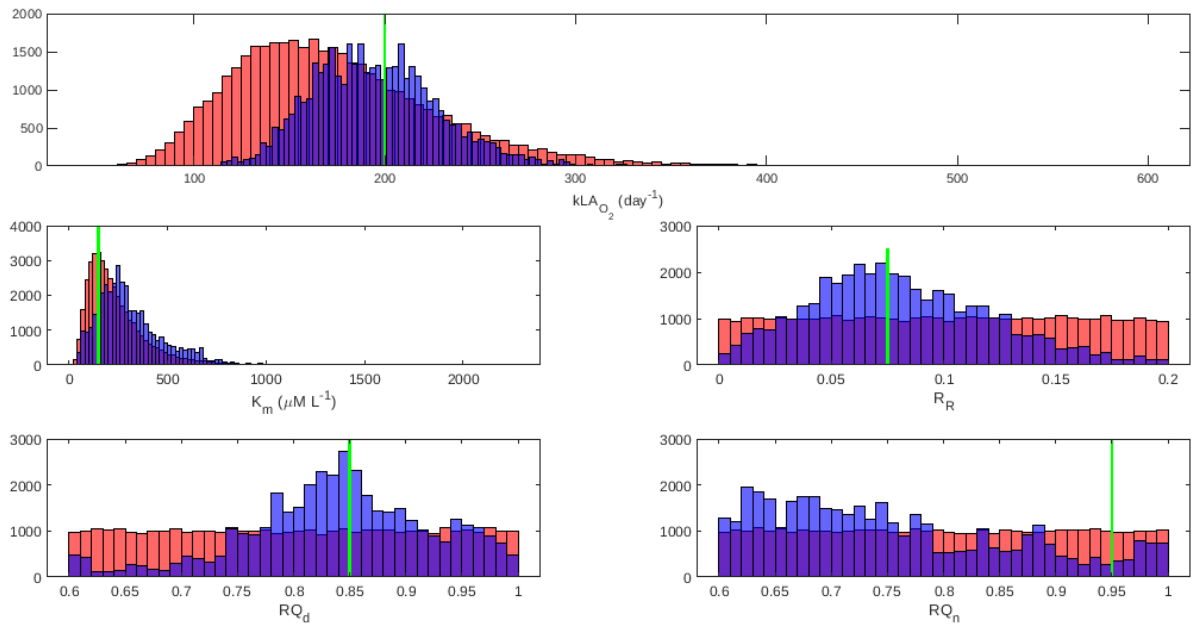


Figure 4.10 : Priors (pink), posteriors (purple) and true values (green) for model parameters.

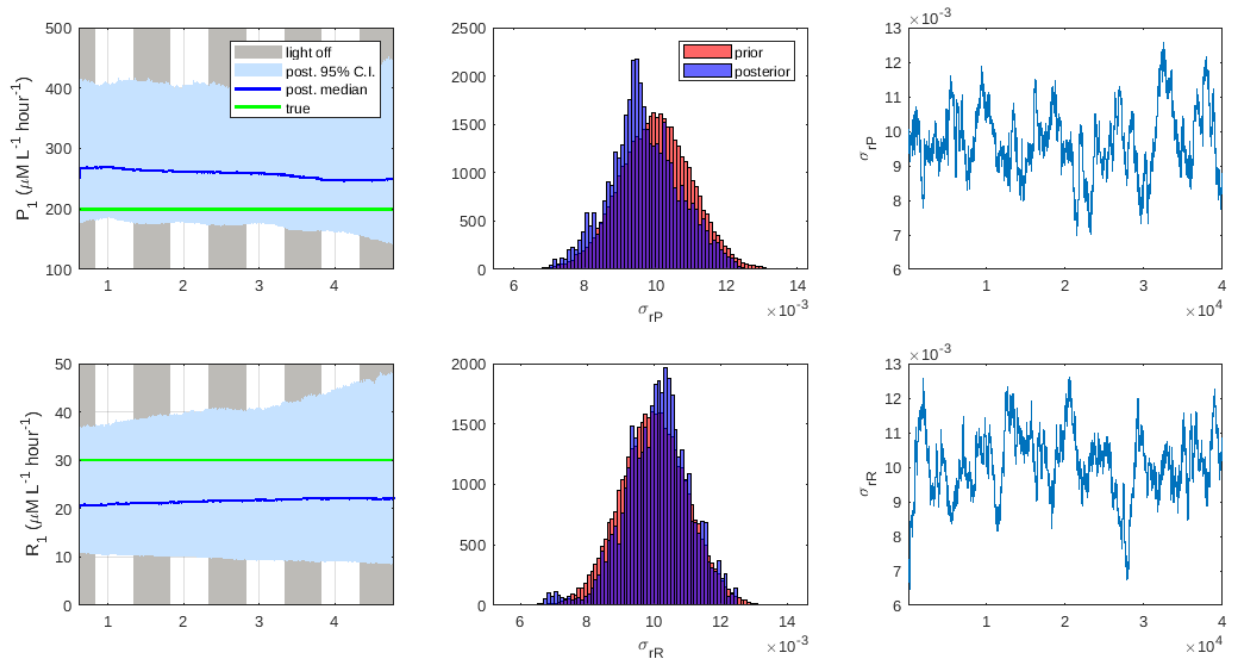


Figure 4.11 : Random walk posteriors  $P_1$  and  $R_1$  medians (solid blue), 95% credible intervals (shaded blue), and true values (green).  $\sigma_{rP}$  and  $\sigma_{rR}$  priors (pink), posteriors (purple), true values (green) and traces.

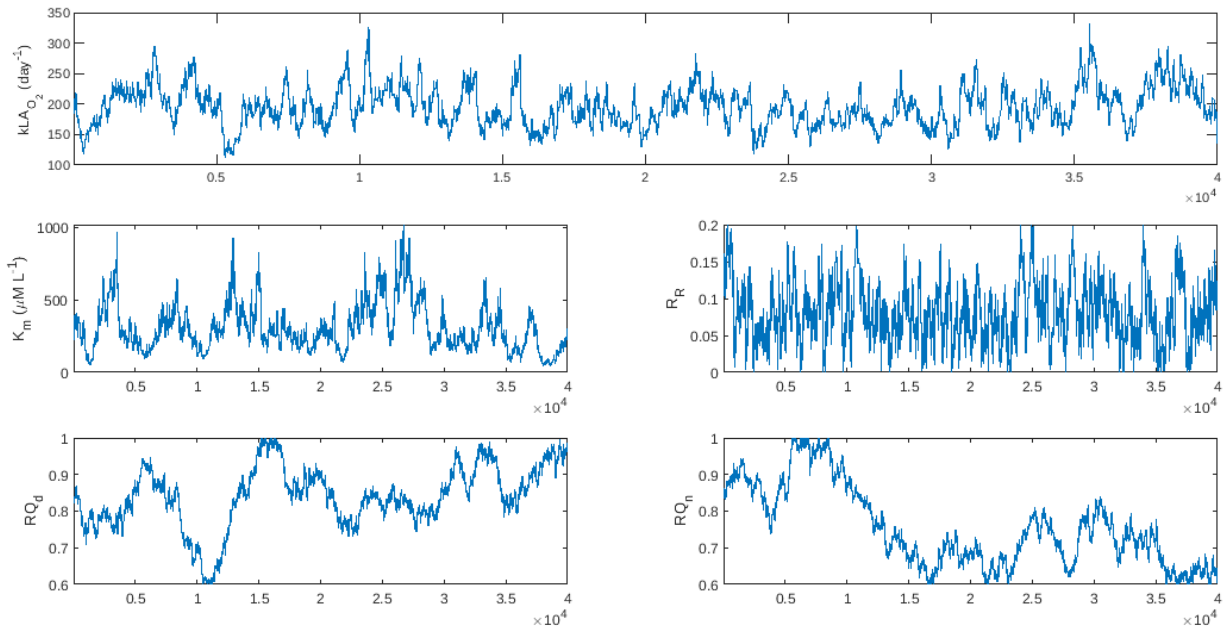


Figure 4.12 : Traces for model parameters.

Carbon chemistry variables  $HCO_3$ ,  $CO_2$ , and  $CO_3$  ... (Figure 4.9).

Log-likelihood stopped rising (Figure ??). Acceptance rate was XX.

### 4.3.3 Posterior results with experimental data where photosynthesis, respiration and respiratory quotients are changing through time

Photosynthesis ( $P_1$ ) and respiration ( $R_1$ ) were both modelled as random walks, by taking  $P$  and  $R$ , previously constant parameters, and replacing them by  $P_1(t)$  and  $R_1(t)$ . Here, we take  $P_1(t)$  and  $R_1(t)$  to be such that

$$P_1(t + \Delta t) = P(t) + r_P$$

$$R_1(t + \Delta t) = R(t) + r_R$$

where  $r_P \sim N(0, \sigma_{r_P})$ ,  $r_R \sim N(0, \sigma_{r_R})$ , and  $\Delta t$  is the length of discrete time-step. For the purpose of the Bayesian analysis here,  $\sigma_{r_P}$  and  $\sigma_{r_R}$  are treated as parameters to be inferred. The respiratory quotients  $RQ_d$  and  $RQ_n$  were also treated as random walks with  $rP$  and  $rR$  as wiener processes. Parameters  $kLA_{O_2}$ ,  $K_m$ ,  $R_R$ ,  $\sigma_{rP}$ ,  $\sigma_{rR}$  were all treated as constant through time but unknown with prior distributions and proposal distributions defined in Table 4.7. The data model assigned log normally distributed observation errors for each instrument (Section 4.2.2) with observation error standard deviation values cited in Table 4.7.

Parameter	Prior	Proposal
$kLA_{O_2}$	$\text{Log}\mathcal{N}(\log(200.0), 0.3)$	$\text{Log}\mathcal{N}(\log(kLA_{O_2}), 0.03)$
$K_m$	$\text{Log}\mathcal{N}(\log(200.0), 0.6)$	$\text{Log}\mathcal{N}(\log(K_m), 0.06)$
$R_R$	$\text{Uniform}(0, 0.2)$	$\text{Trun}\mathcal{N}(R_R, 0.005, \text{lower} = 0, \text{upper} = 0.2)$
$\sigma_{rP}$	$\mathcal{N}(0.02, 0.002)$	$\mathcal{N}(\sigma_{r_P}, 0.0002)$
$\sigma_{rR}$	$\mathcal{N}(0.01, 0.001)$	$\mathcal{N}(\sigma_{r_R}, 0.0001)$
$\sigma_{O_2}$	0.3	*
$\sigma_{pH}$	0.3	*

$\sigma_{DIC}$	0.5	*
----------------	-----	---

Table 4.7 : Table of Parameters, their priors and proposal distributions (\* indicates the parameter was held fixed).

Parameter	Quantiles (25%, 75%)	Quantiles (5%, 95%)
$kLA_{O_2}^{air}$	(139.5979, 170.8102)	(120.9171, 204.5474)
$K_m$	(168.1931, 378.8001)	(104.2598, 599.6881)
$R_R$	(0.0815, 0.1512)	(0.0284, 0.1844)
$\sigma_{r_P}$	(0.0178, 0.0202)	(0.0159, 0.0222)
$\sigma_{r_R}$	(0.0094, 0.0113)	(0.0081, 0.0121)

Table 4.8 : Posterior (25%, 75%), (5%, 95%) quantiles for parameters after assimilating observations.

The PMMH was run for 40,000 samples with 1024 particles, where 10,000 samples were discarded as burn-in. 40,000 samples was all that could complete in the maximum time allocation on the hpc cluster.

Observed state variables  $DIC$  and  $TA$  posteriors perform well with all observations lying within the 95% credible intervals (Figure 4.14).  $O_2$  posteriors tracked the observations well while the light was on, with all observations falling inside tightly constrained 95% credible intervals. During times when there was no light, most days the posteriors would fit the observations well to start and then potentially there was a sensor drift causing increasing observations that the model was not accounting for (Figure 4.13).  $pH$  captured most observations within the 95% credible intervals except for 2 points on the 1st day while the light was off (Figure 4.13).

The photosynthesis rate ( $P_1$ ) centred around  $240 \mu\text{M L}^{-1} \text{ hour}^{-1}$  with small oscillations during daylight. The 95% credible intervals lay approximately between

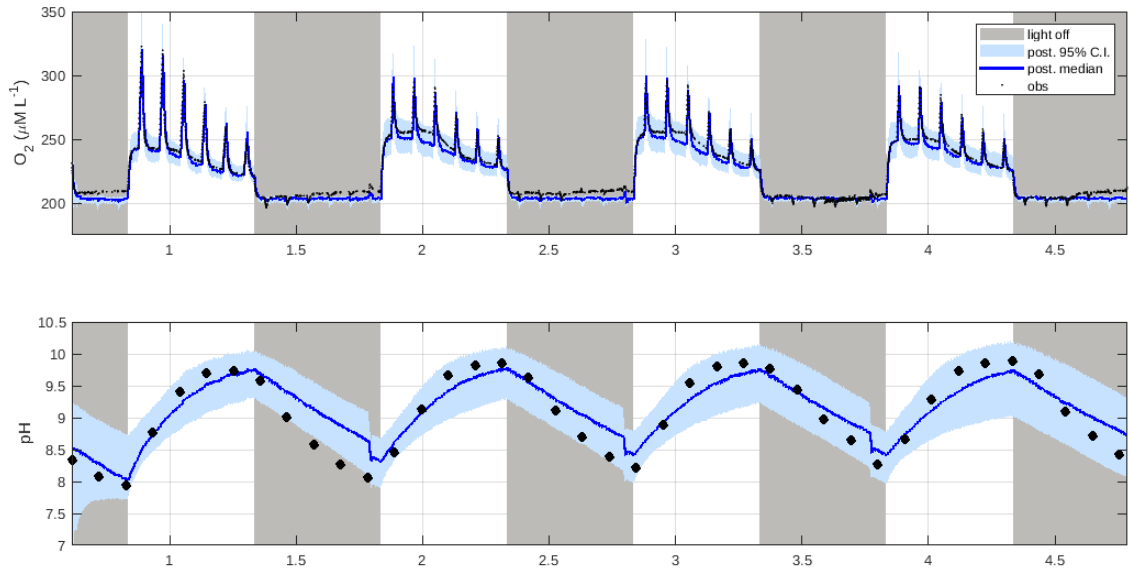


Figure 4.13 : Posterior medians (solid blue line), 95% credible intervals (shaded blue), and simulated observations (black) for  $O_2$  and  $pH$  across 4 days.

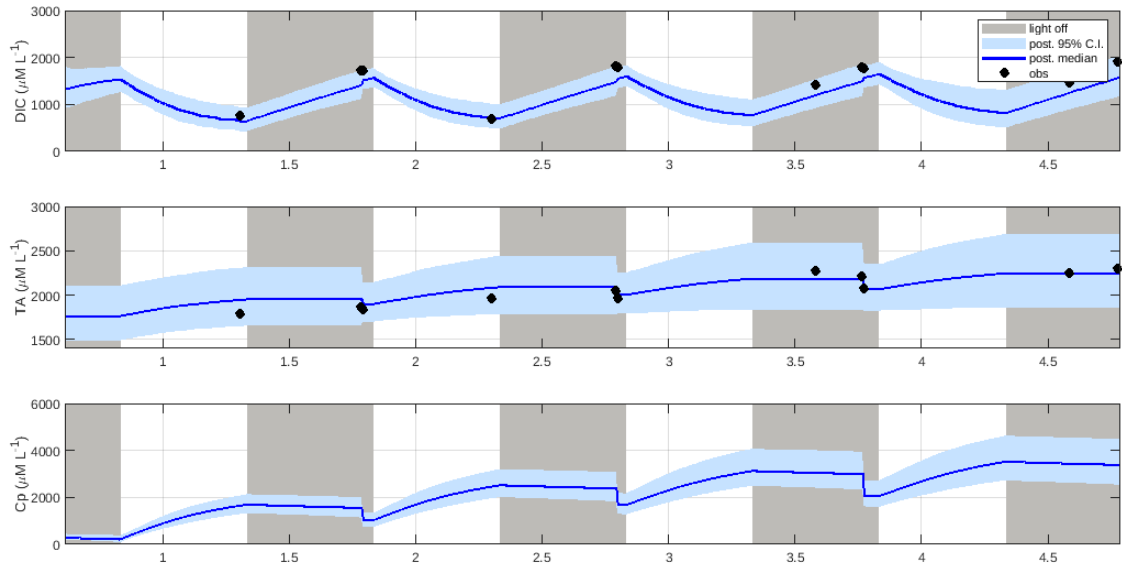


Figure 4.14 : Posterior medians (solid blue line), 95% credible intervals (shaded blue), and simulated observations (black) for  $DIC$ ,  $TA$  and  $C_p$  across 4 days.

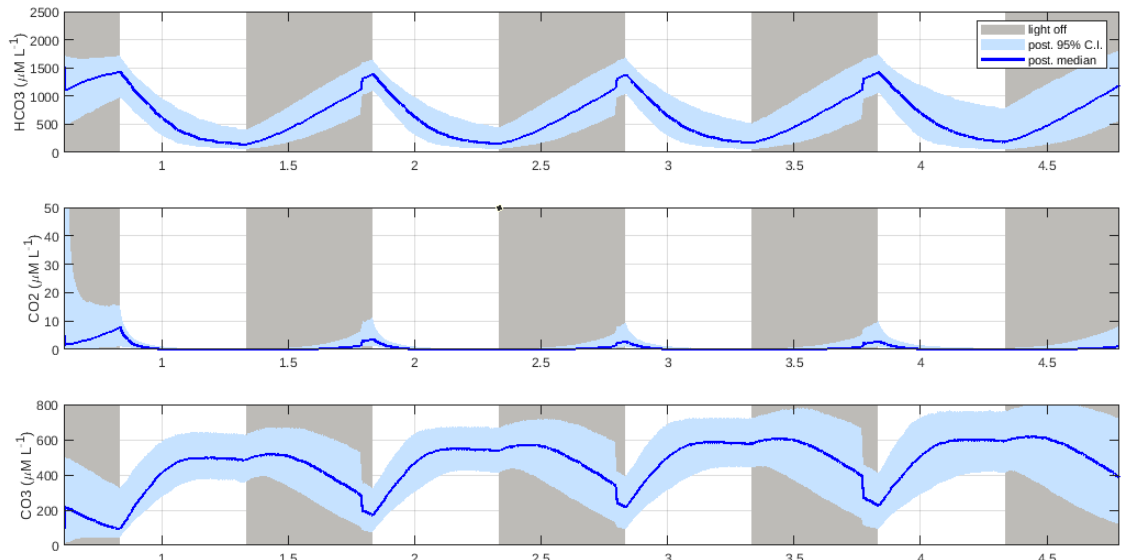


Figure 4.15 : Posterior medians (solid blue line) and 95% credible intervals (shaded blue) for  $HCO_3^-$ ,  $CO_2$  and  $CO_3$  across 4 days.

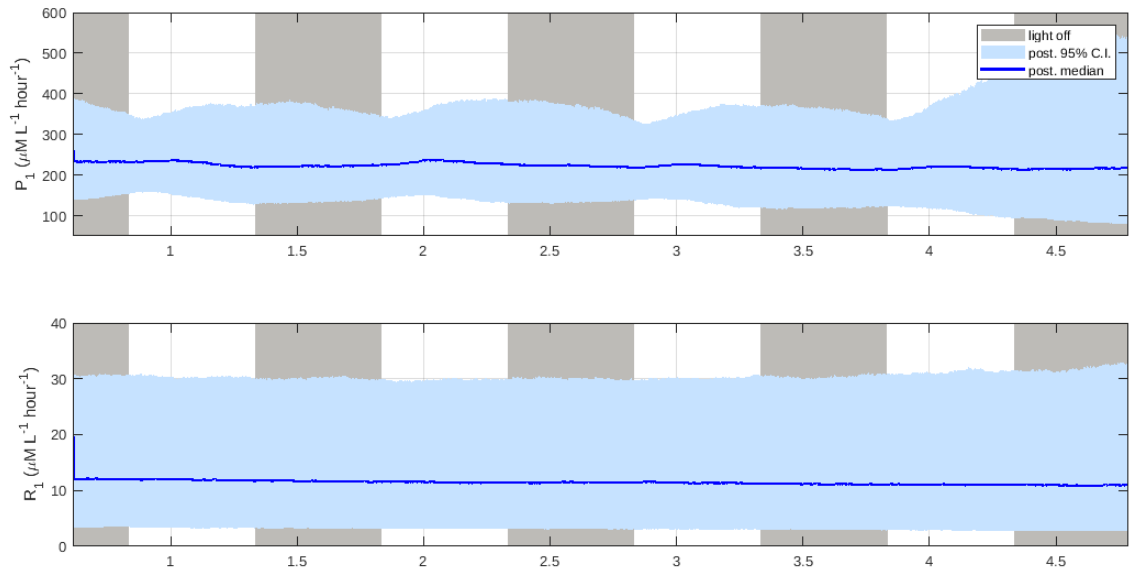


Figure 4.16 : Posterior medians (solid blue line) and 95% credible intervals (shaded blue) for photosynthesis  $P_1$  and respiration  $R_1$ .

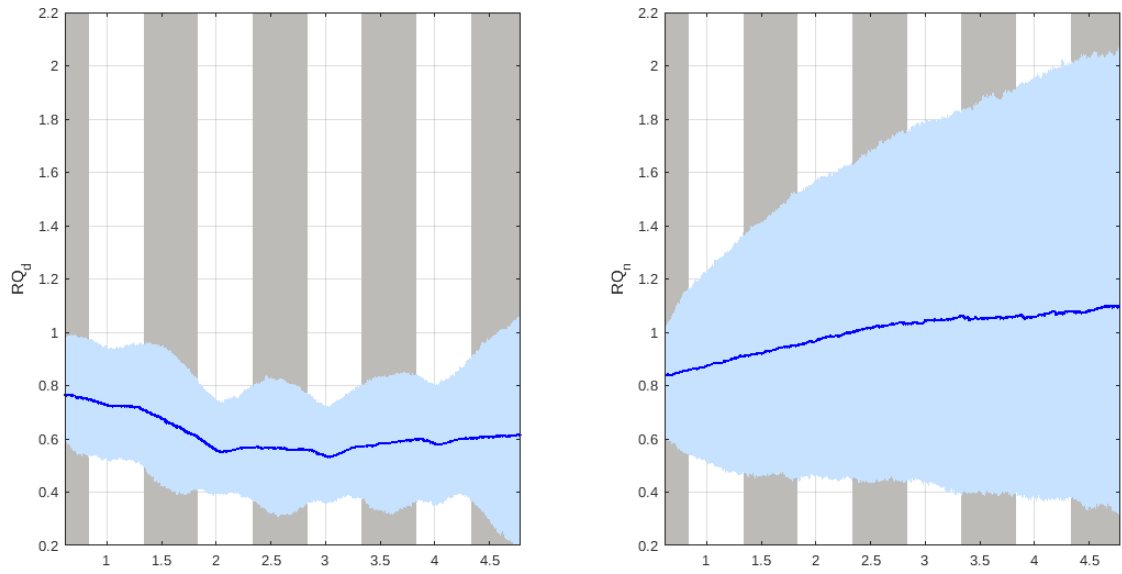


Figure 4.17 : Posterior medians (solid blue line) and 95% credible intervals (shaded blue) for  $RQ_d$  and  $RQ_n$ .

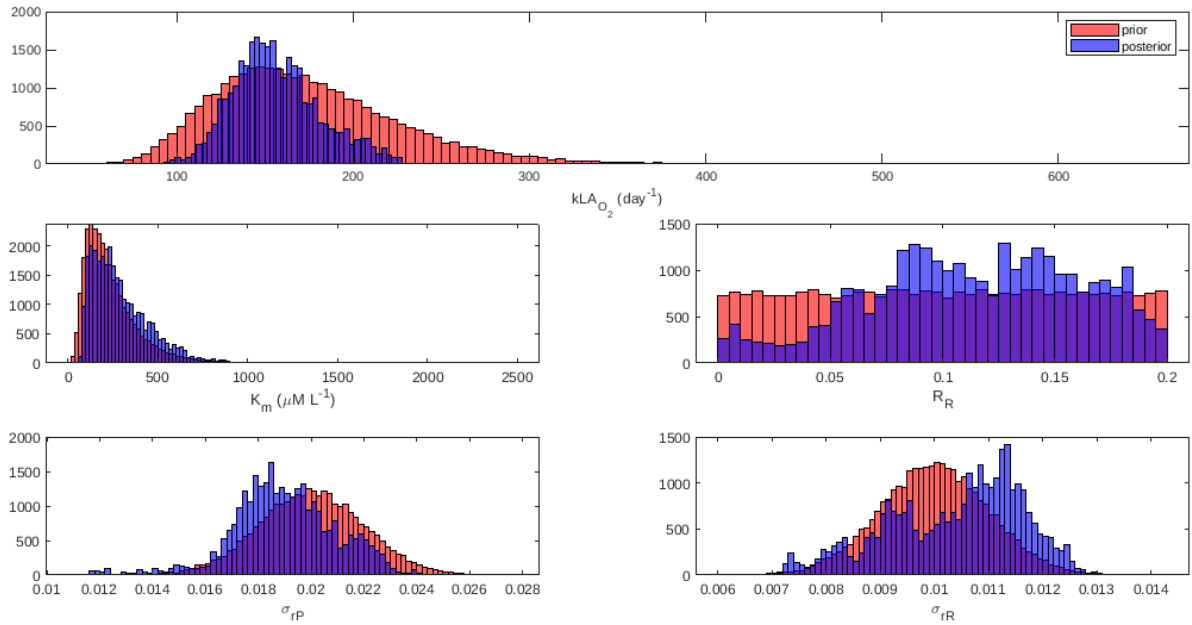


Figure 4.18 : Priors (pink) and posteriors (purple) for model parameters.



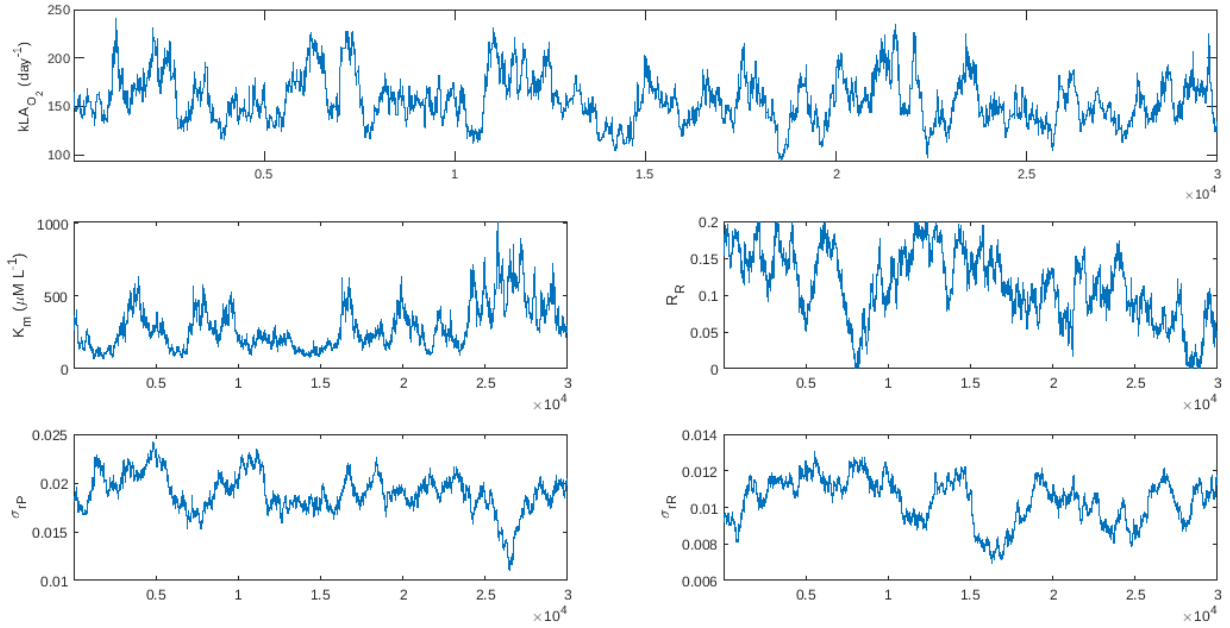


Figure 4.19 : Traces for model parameters.

$150 \mu\text{M L}^{-1} \text{ hour}^{-1}$  and  $400 \mu\text{M L}^{-1} \text{ hour}^{-1}$  with tighter intervals at the start of each day (Figure 4.16). The respiration rate ( $R_1$ ) remained centred around  $12 \mu\text{M L}^{-1} \text{ hour}^{-1}$  with 95% credible intervals between  $4 \mu\text{M L}^{-1} \text{ hour}^{-1}$  and  $30 \mu\text{M L}^{-1} \text{ hour}^{-1}$  through the course of the experiment (Figure 4.16).

The day-time respiratory quotient  $RQ_d$  centred around 0.75 with a 95% credible interval of 0.6-1 during the first day of the experiment. The centre dropped to 0.6 during the 2nd, 3rd and 4th days, with 95% credible intervals of 0.4-0.8 (Figure 4.17). The night-time respiratory quotient  $RQ_n$  was initially centred around 0.85 and slowly rose to 1.1 by the end of the experiment. The 95% credible intervals started with 0.6-1 and increased its span steadily to 0.4-2 by the end of the experiment (Figure 4.17).

parameters

mixing

acceptance rate was XX.

[BM: Values  $>1$  are not really realistic for respiratory quotients but this can be compensation for not having an offset on the  $O_2$  obs, talk about this in the discussion]

#### 4.3.4 [To update] Posterior results with experimental data where photosynthesis, respiration and respiratory quotients are changing through time and an offset on $O_2$ is introduced

Photosynthesis ( $P_1$ ) and respiration ( $R_1$ ) were both modelled as random walks, by taking  $P$  and  $R$ , previously constant parameters, and replacing them by  $P_1(t)$  and  $R_1(t)$ . Here, we take  $P_1(t)$  and  $R_1(t)$  to be such that

$$P_1(t + \Delta t) = P(t) + r_P$$

$$R_1(t + \Delta t) = R(t) + r_R$$

where  $r_P \sim N(0, \sigma_{r_P})$ ,  $r_R \sim N(0, \sigma_{r_R})$ , and  $\Delta t$  is the length of discrete time-step. For the purpose of the Bayesian analysis here,  $\sigma_{r_P}$  and  $\sigma_{r_R}$  are treated as parameters to be inferred. The respiratory quotients  $RQ_d$  and  $RQ_n$  were also treated as random walks with  $r_P$  and  $r_R$  as wiener processes. Parameters  $kLA_{O_2}$ ,  $K_m$ ,  $R_R$ ,  $\sigma_{r_P}$ ,  $\sigma_{r_R}$ , and  $offset_{O_2}$  were all treated as constant through time but unknown with prior distributions and proposal distributions defined in Table 4.9. The data model assigned log normally distributed observation errors for each instrument (Section 4.2.2) with observation error standard deviation values cited in Table 4.9.

Parameter	Prior	Proposal
$kLA_{O_2}$	$\text{Log}\mathcal{N}(\log(200.0), 0.3)$	$\text{Log}\mathcal{N}(\log(kLA_{O_2}), 0.03)$
$K_m$	$\text{Log}\mathcal{N}(\log(200.0), 0.6)$	$\text{Log}\mathcal{N}(\log(K_m), 0.06)$
$R_R$	$\text{Uniform}(0, 0.2)$	$\text{Trun}\mathcal{N}(R_R, 0.005, \text{lower} = 0, \text{upper} = 0.2)$
$\sigma_{r_P}$	$\mathcal{N}(0.02, 0.002)$	$\mathcal{N}(\sigma_{r_P}, 0.0002)$

$\sigma_{r_R}$	$\mathcal{N}(0.01, 0.001)$	$\mathcal{N}(\sigma_{r_R}, 0.0001)$
$offset_{O_2}$	$\mathcal{N}(0, 5.0)$	$\mathcal{N}(offset_{O_2}, 0.5)$
$\sigma_{O_2}$	0.1	*
$\sigma_{pH}$	0.1	*
$\sigma_{DIC}$	0.2	*

Table 4.9 : Table of Parameters, their priors and proposal distributions (\* indicates the parameter was held fixed).

Parameter	Quantiles (25%, 75%)	Quantiles (5%, 95%)
$kLA_{O_2}^{air}$	(152.9413, 167.1751)	(134.8441, 171.2411)
$K_m$	(253.4041, 297.8397)	(227.8899, 372.5976)
$R_R$	(0.0532, 0.0744)	(0.0346, 0.0833)
$\sigma_{r_P}$	(0.0170, 0.0177)	(0.0168, 0.0188)
$\sigma_{r_R}$	(0.0093, 0.0095)	(0.0092, 0.0097)
$offset_{O_2}$	(-0.7349, 2.3508)	(-1.3238, 5.7816)

Table 4.10 : Posterior (25%, 75%), (5%, 95%) quantiles for parameters after assimilating observations.

The PMMH was run for 20,000 samples with 2048 particles, where 10,000 samples were discarded as burn-in. 20,000 samples were all that could complete with 2048 particles in the maximum time allocation on the hpc cluster. The number of particles was doubled to test whether this improved mixing.

[UPDATE text below]

Observed state variables  $DIC$  and  $TA$  posteriors perform well with all observations lying within the 95% credible intervals (Figure 4.21).  $O_2$  posteriors tracked the observations excellently while the light was on, with all observations falling inside

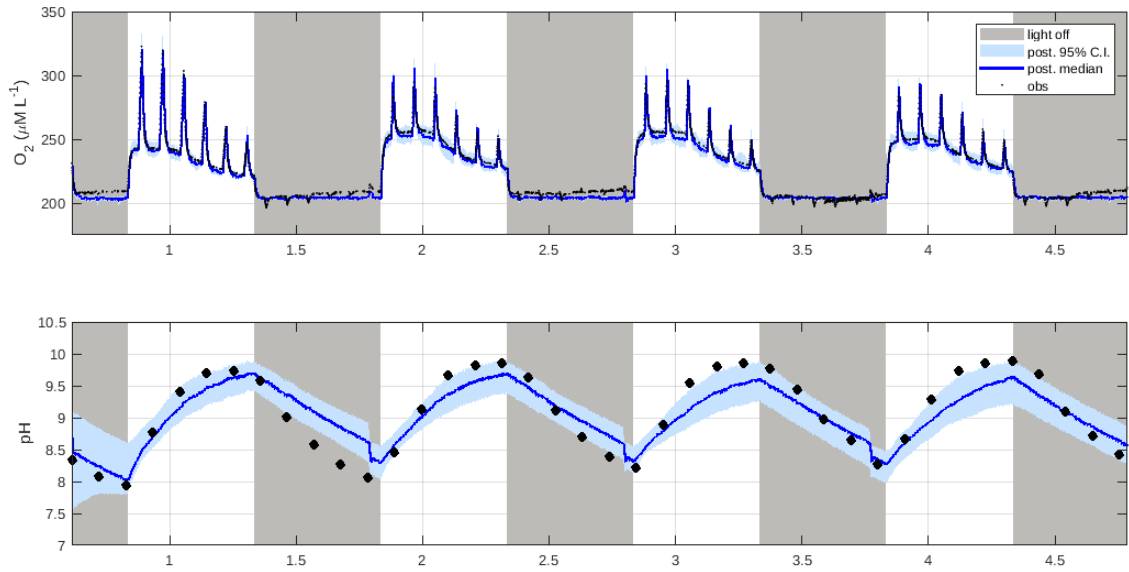


Figure 4.20 : Posterior medians (solid blue line), 95% credible intervals (shaded blue), and simulated observations (black) for  $O_2$  and  $pH$  across 4 days.

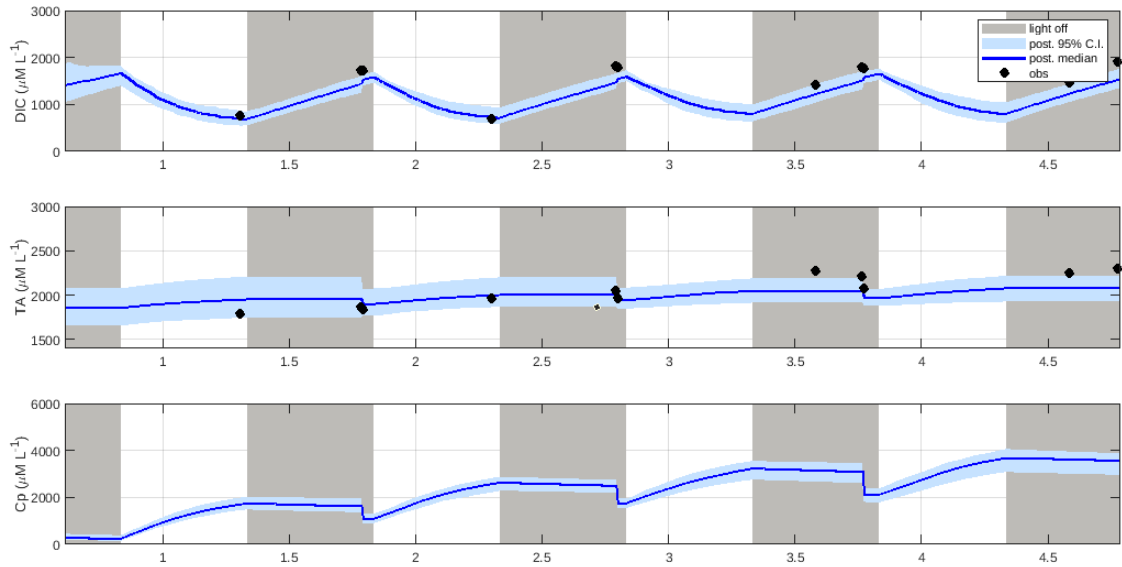


Figure 4.21 : Posterior medians (solid blue line), 95% credible intervals (shaded blue), and simulated observations (black) for  $DIC$ ,  $TA$  and  $C_p$  across 4 days.

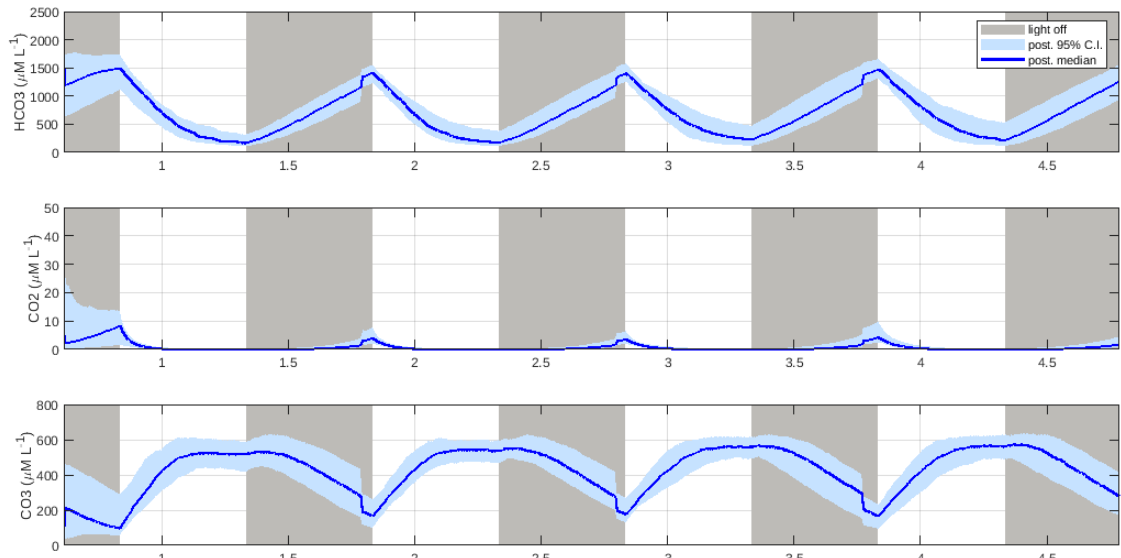


Figure 4.22 : Posterior medians (solid blue line) and 95% credible intervals (shaded blue) for  $HCO_3^-$ ,  $CO_2$  and  $CO_3$  across 4 days.

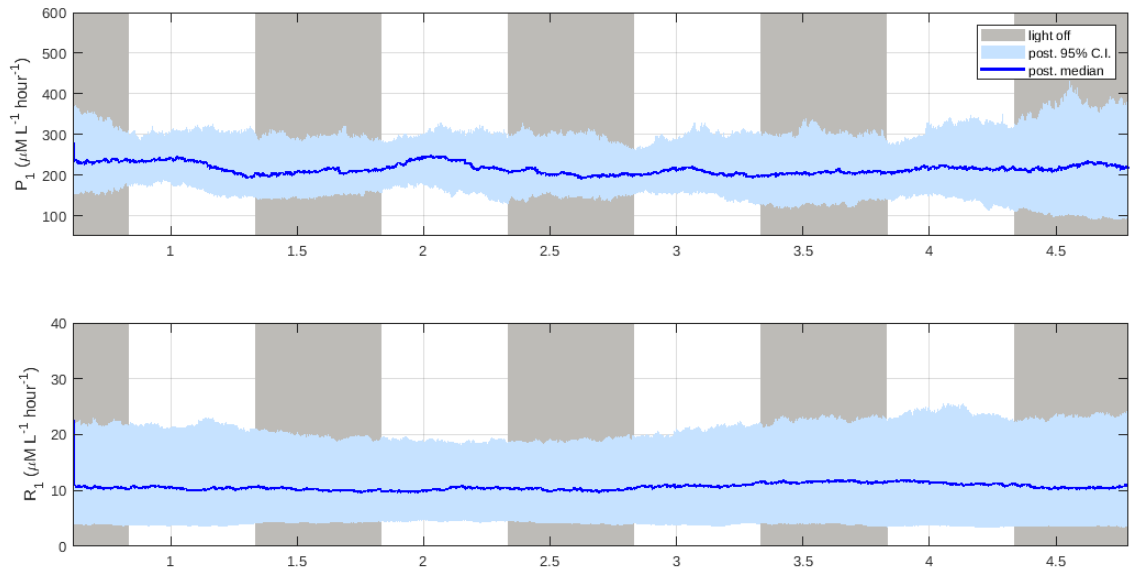


Figure 4.23 : Posterior medians (solid blue line) and 95% credible intervals (shaded blue) for photosynthesis  $P_1$  and respiration  $R_1$ .

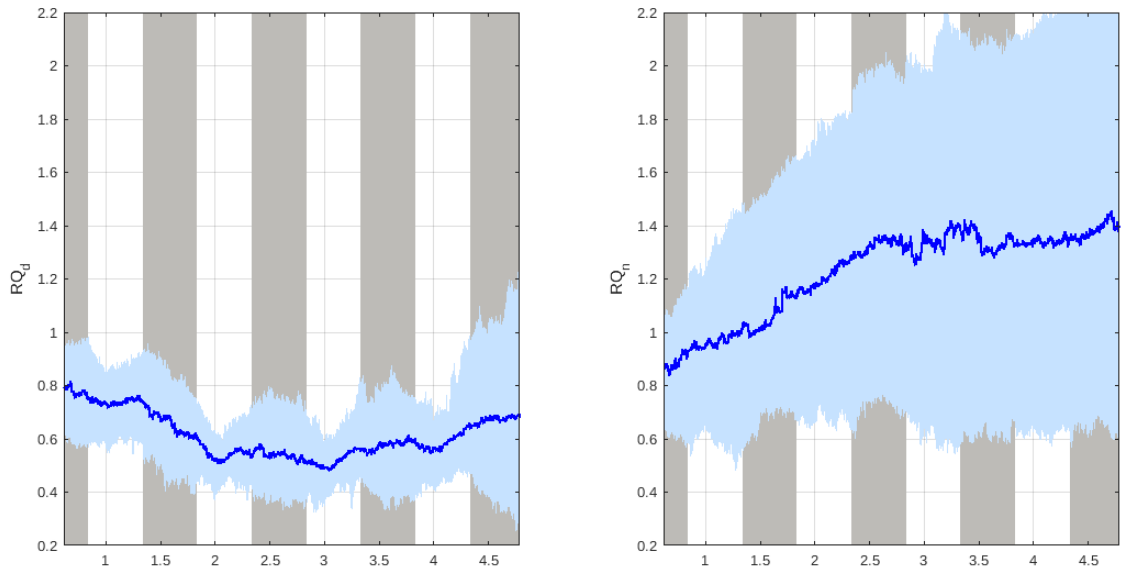


Figure 4.24 : Posterior medians (solid blue line) and 95% credible intervals (shaded blue) for  $RQ_d$  and  $RQ_n$ .

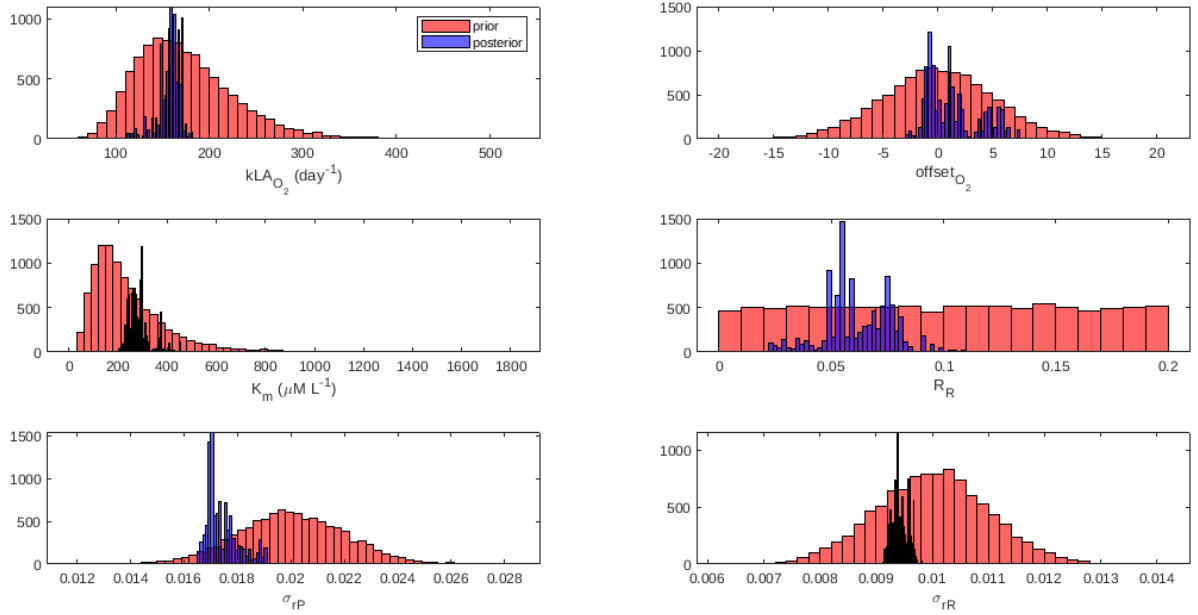


Figure 4.25 : Priors (pink) and posteriors (purple) for model parameters.

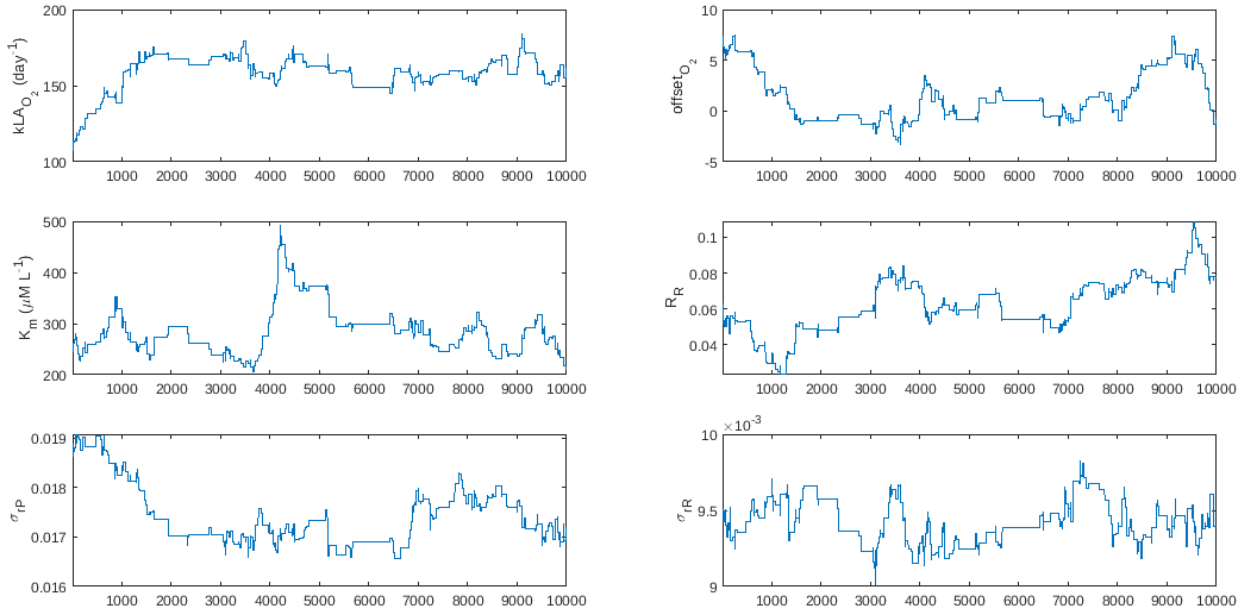


Figure 4.26 : Traces for model parameters.

tightly constrained 95% credible intervals. During times when there was no light, most days the posteriors would fit the observations well to start and then potentially there was a sensor drift causing increasing observations that the model was not accounting for (Figure 4.20). *pH* captured most observations within the 95% credible intervals except for 2 points on the 1st day while the light was off (Figure 4.20).

The photosynthesis rate ( $P_1$ ) centred around  $240 \mu\text{M L}^{-1} \text{ hour}^{-1}$  with small oscillations during daylight. The 95% credible intervals lay approximately between  $150 \mu\text{M L}^{-1} \text{ hour}^{-1}$  and  $400 \mu\text{M L}^{-1} \text{ hour}^{-1}$  with tighter intervals at the start of each day (Figure 4.23). The respiration rate ( $R_1$ ) remained centred around  $12 \mu\text{M L}^{-1} \text{ hour}^{-1}$  with 95% credible intervals between  $4 \mu\text{M L}^{-1} \text{ hour}^{-1}$  and  $30 \mu\text{M L}^{-1} \text{ hour}^{-1}$  through the course of the experiment (Figure 4.23).

The day-time respiratory quotient  $RQ_d$  centred around 0.75 with a 95% credible interval of 0.6-1 during the first day of the experiment. The centre dropped to

0.6 during the 2nd, 3rd and 4th days, with 95% credible intervals of 0.4-0.8 (Figure 4.24). The night-time respiratory quotient  $RQ_n$  was initially centred around 0.85 and slowly rose to 1.1 by the end of the experiment. The 95% credible intervals started with 0.6-1 and increased its span steadily to 0.4-2 by the end of the experiment (Figure 4.24).

parameters offset

mixing, not great. try thinning out the observations further (results in next section).

acceptance rate was 3.7%.

#### 4.3.5 [To update] Posterior results with experimental data where photosynthesis, respiration and respiratory quotients are changing through time, $O_2$ has an offset and the $O_2$ observations were thinned further

[BM: UPDATE THESE RESULTS WHEN THE RUNS COMPLETE]

Because we can see that the MCMC chain is mixing poorly, we tested whether thinning down the  $O_2$  observations even further would improve the results (Figure 4.27).

#### Results:

20,000 samples run with 2048 particles (20,000 was all that could complete in the maximum time allocation on the hpc cluster). 10,000 discarded as burn-in, acceptance rate was 14%.



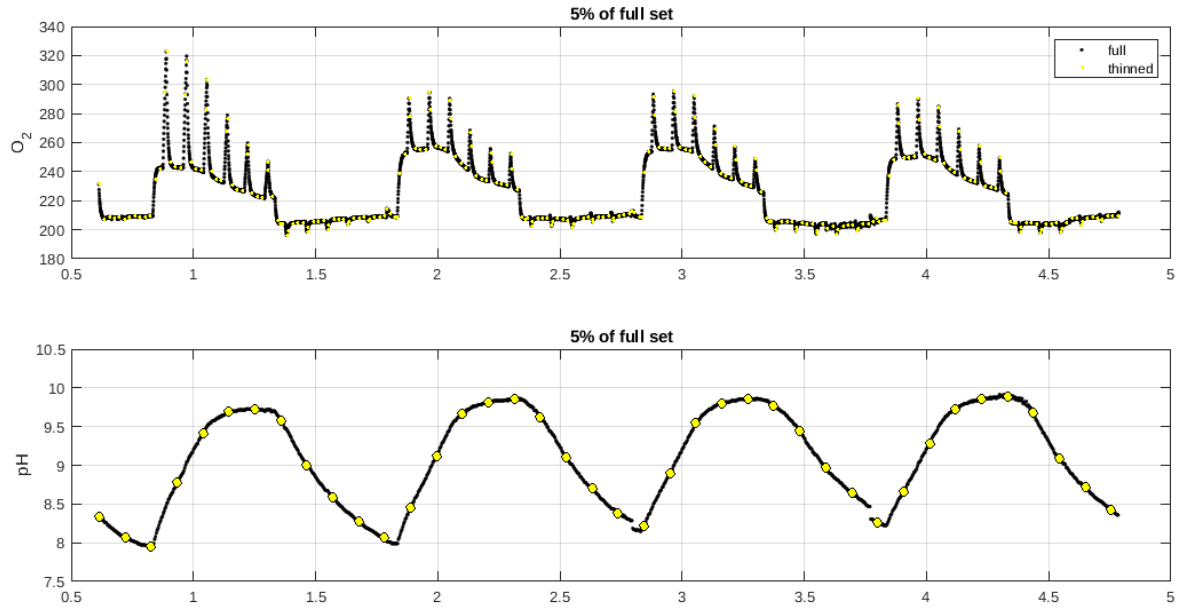


Figure 4.27 : Full  $O_2$  and  $pH$  datasets with further thinned  $O_2$  and  $pH$  observations.

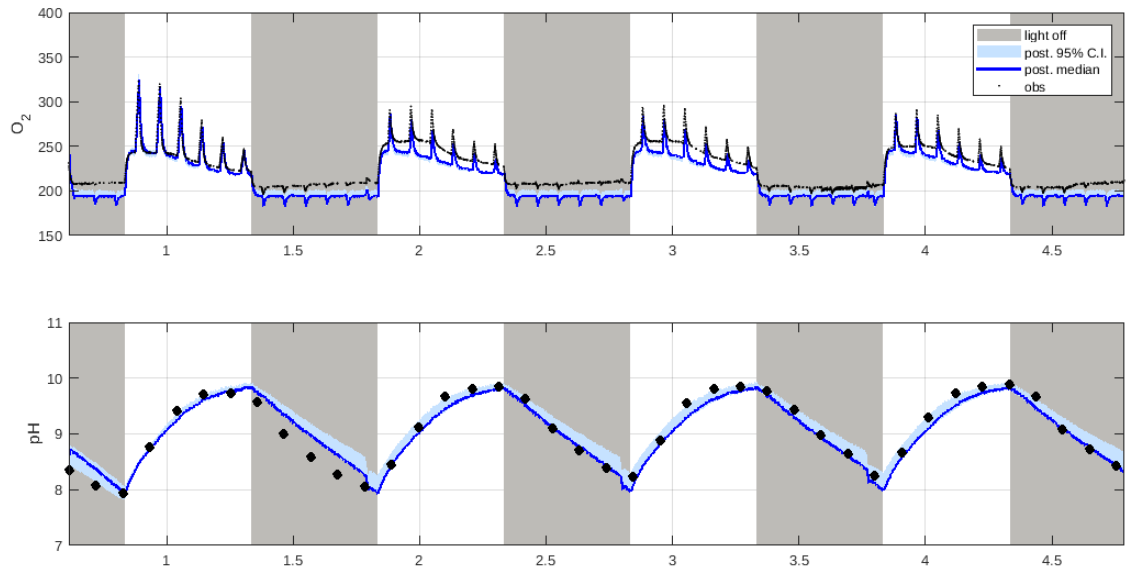


Figure 4.28 : Posterior medians (solid blue line), 95% credible intervals (shaded blue), and simulated observations (black) for  $O_2$  and  $pH$  across 4 days.

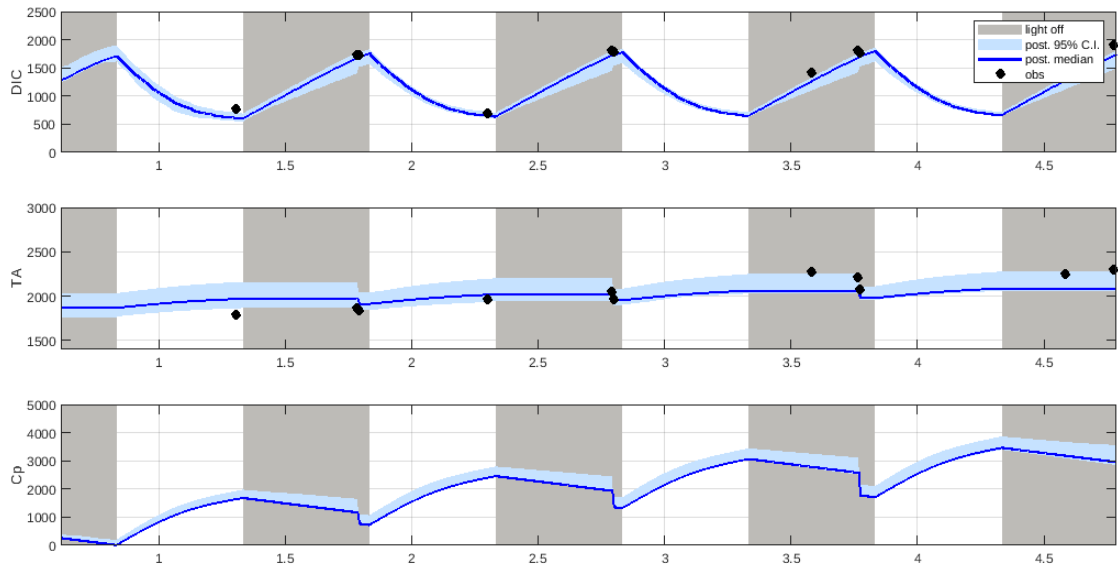


Figure 4.29 : Posterior medians (solid blue line), 95% credible intervals (shaded blue), and simulated observations (black) for  $DIC$ ,  $TA$  and  $C_p$  across 4 days.

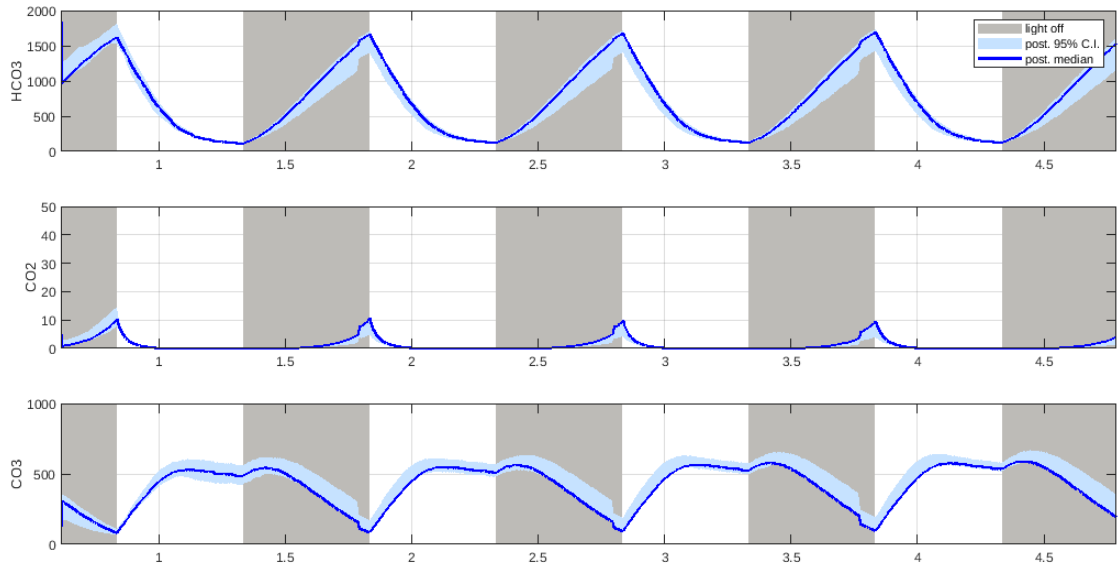


Figure 4.30 : Posterior medians (solid blue line) and 95% credible intervals (shaded blue) for  $HCO_3$ ,  $CO_2$  and  $CO_3$  across 4 days.

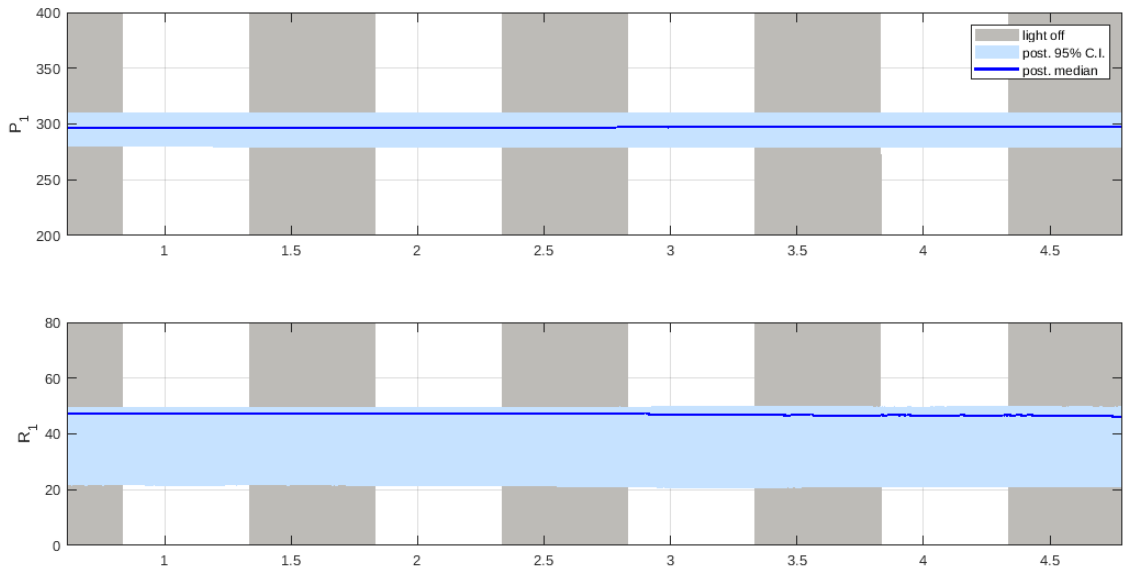


Figure 4.31 : Posterior medians (solid blue line) and 95% credible intervals (shaded blue) for photosynthesis  $P_1$ .

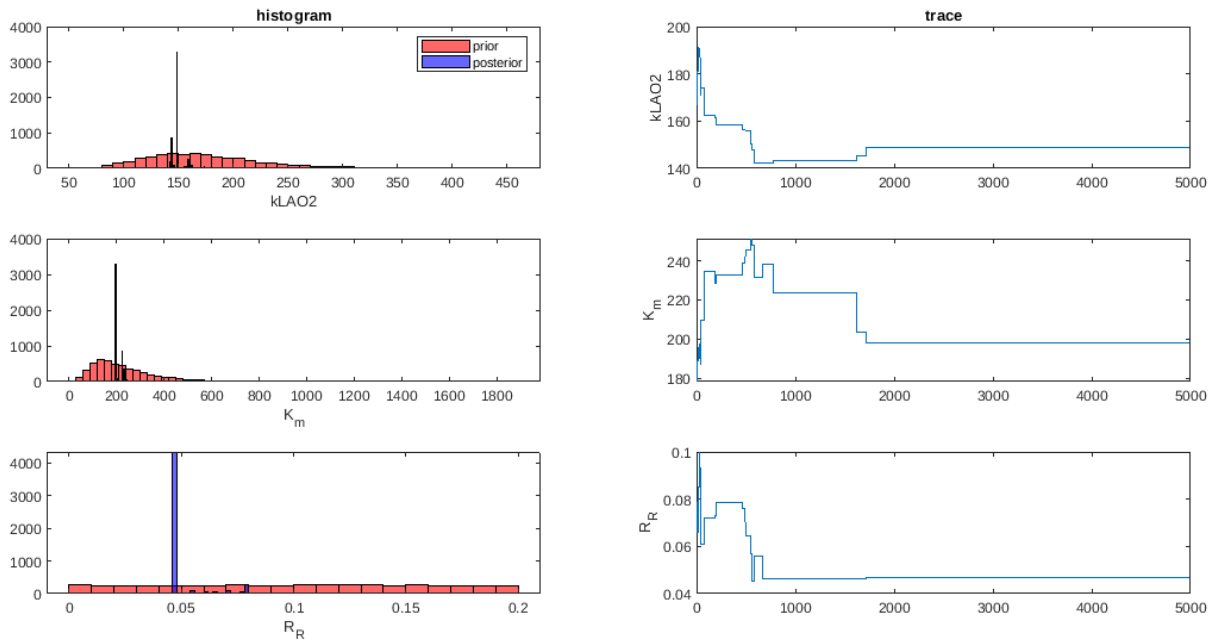


Figure 4.32 : Priors (pink) and posteriors (purple) for model parameters.

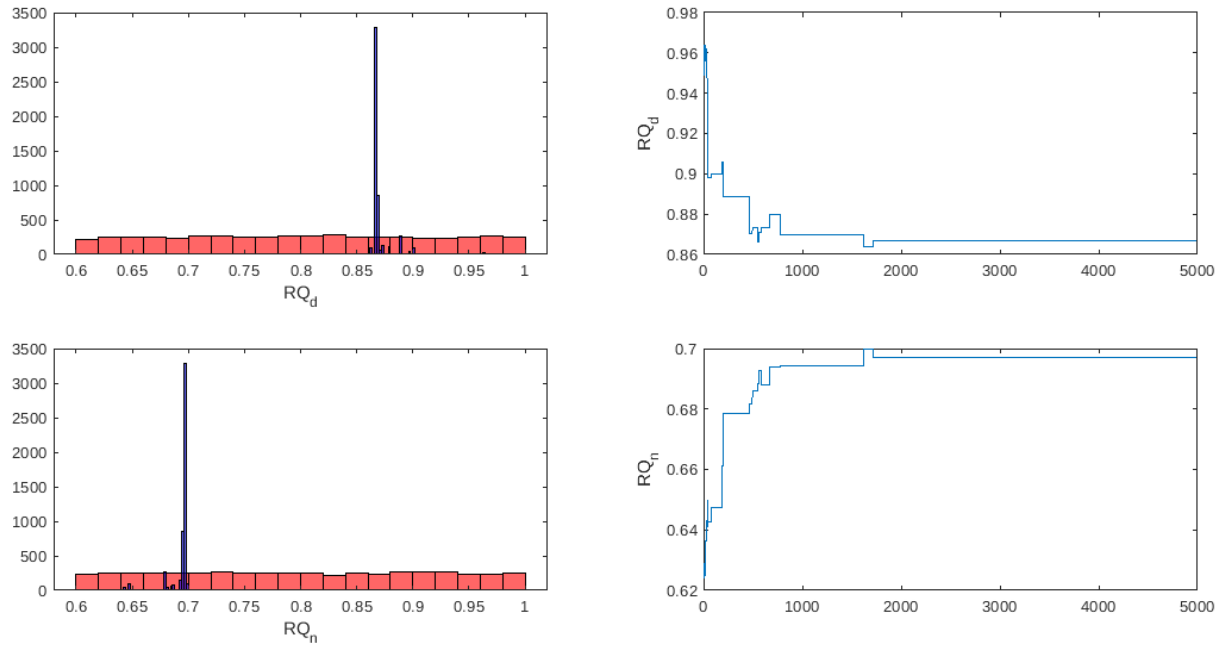


Figure 4.33 : Priors (pink) and posteriors (purple) for model parameters.

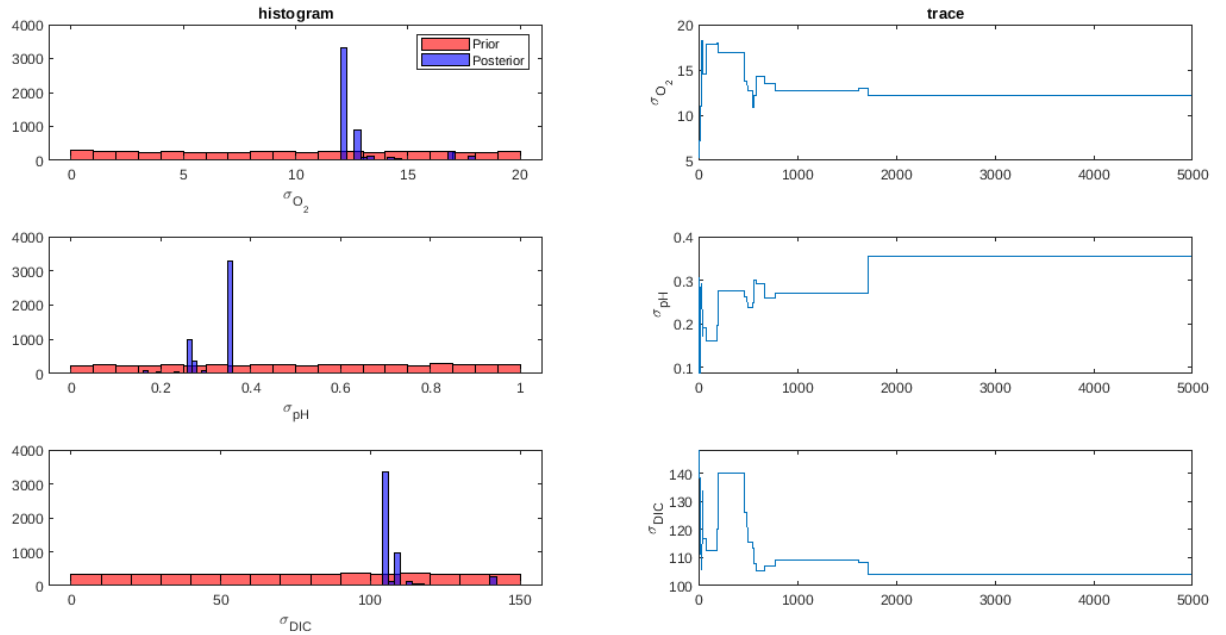


Figure 4.34 : Priors (pink) and posteriors (purple) for model parameters.

#### 4.3.6 [To update] Posterior results with experimental data where respiratory quotients are changing through time and bounded to biologically feasible values

[BM: currently running an updated version of this, if it works out then update the plots and results here. if it doesn't then keep this run but move it to the appendix.]

##### In this run:

Photosynthesis ( $P_1$ ) and respiration ( $R_1$ ) are both modelled as random walks, by taking  $P$  and  $R$ , previously constant parameters, and replacing them by  $P_1(t)$  and  $R_1(t)$ . Here, we take  $P_1(t)$  and  $R_1(t)$  to be such that

$$P_1(t + \Delta t) = P(t) + r_P$$

$$R_1(t + \Delta t) = R(t) + r_R$$

where  $r_P \sim N(0, \sigma_{r_P})$ ,  $r_R \sim N(0, \sigma_{r_R})$ , and  $\Delta t$  is the length of discrete time-step. For the purpose of the Bayesian analysis here,  $\sigma_{r_P}$  and  $\sigma_{r_R}$  are treated as parameters to be inferred. The respiratory quotients ( $RQ_d$  and  $RQ_n$ ) are treated as normally distributed noisy states, truncated between 0.6 and 1, where the mean and standard deviation are unknown parameters to be estimated.

$$RQ_d \sim \text{truncated}\mathcal{N}(\mu_{RQ_d}, \sigma_{RQ_d}, \text{lower} = 0.6, \text{upper} = 1.0)$$

$$RQ_n \sim \text{truncated}\mathcal{N}(\mu_{RQ_n}, \sigma_{RQ_n}, \text{lower} = 0.6, \text{upper} = 1.0)$$

$kLA_{O_2}$ ,  $K_m$ ,  $R_R$ ,  $\sigma_{r_P}$ ,  $\sigma_{r_R}$ ,  $\mu_{RQ_d}$ ,  $\sigma_{RQ_d}$ ,  $\mu_{RQ_n}$ ,  $\sigma_{RQ_n}$  are all treated as parameters constant through time but unknown.

Parameter	Prior	Proposal
$kLA_{O_2}$	$\text{Log}\mathcal{N}(\log(200.0), 0.3)$	$\text{Log}\mathcal{N}(\log(kLA_{O_2}), 0.03)$
$K_m$	$\text{Log}\mathcal{N}(\log(200.0), 0.6)$	$\text{Log}\mathcal{N}(\log(K_m), 0.06)$
$R_R$	$\text{Uniform}(0, 0.2)$	$\text{Trun}\mathcal{N}(R_R, 0.01, \text{lower} = 0, \text{upper} = 0.2)$

$\mu_{RQ_d}$	Uniform(0.6, 1)	$\mathcal{N}(\mu_{RQ_d}, 0.01)$
$\mu_{RQ_n}$	Uniform(0.6, 1)	$\mathcal{N}(\mu_{RQ_n}, 0.01)$
$\sigma_{RQ_d}$	Uniform(0, 0.5)	$\mathcal{N}(\sigma_{RQ_d}, 0.01)$
$\sigma_{RQ_n}$	Uniform(0, 0.5)	$\mathcal{N}(\sigma_{RQ_n}, 0.01)$
$\sigma_{r_P}$	$\mathcal{N}(0.01, 0.001)$	$\mathcal{N}(\sigma_{r_P}, 0.0001)$
$\sigma_{r_R}$	$\mathcal{N}(0.01, 0.001)$	$\mathcal{N}(\sigma_{r_R}, 0.0001)$
$\sigma_{O_2}$	0.4	*
$\sigma_{pH}$	0.4	*
$\sigma_{DIC}$	0.4	*

Table 4.11 : Table of Parameters, their priors and proposal distributions (\* indicates the parameter was held fixed).

### Results:

10,000 samples run (no burn-in was discarded) with 1024 particles. Updated version:  
20,000 (10,000 discarded as burn-in) samples with 2048 particles.

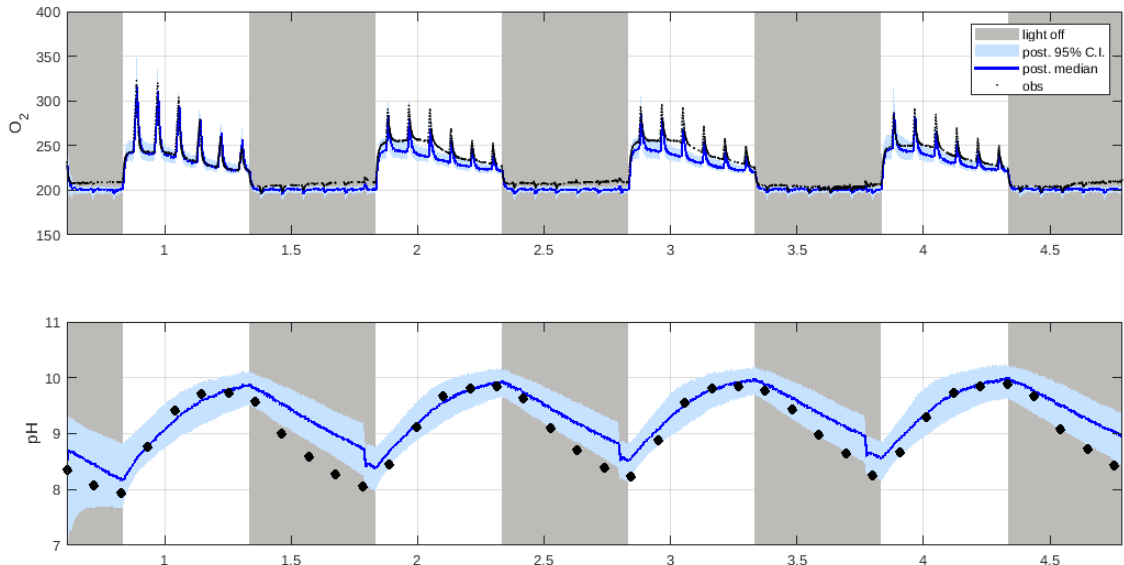


Figure 4.35 : Posterior medians (solid blue line), 95% credible intervals (shaded blue), and simulated observations (black) for  $O_2$  and  $pH$  across 4 days.

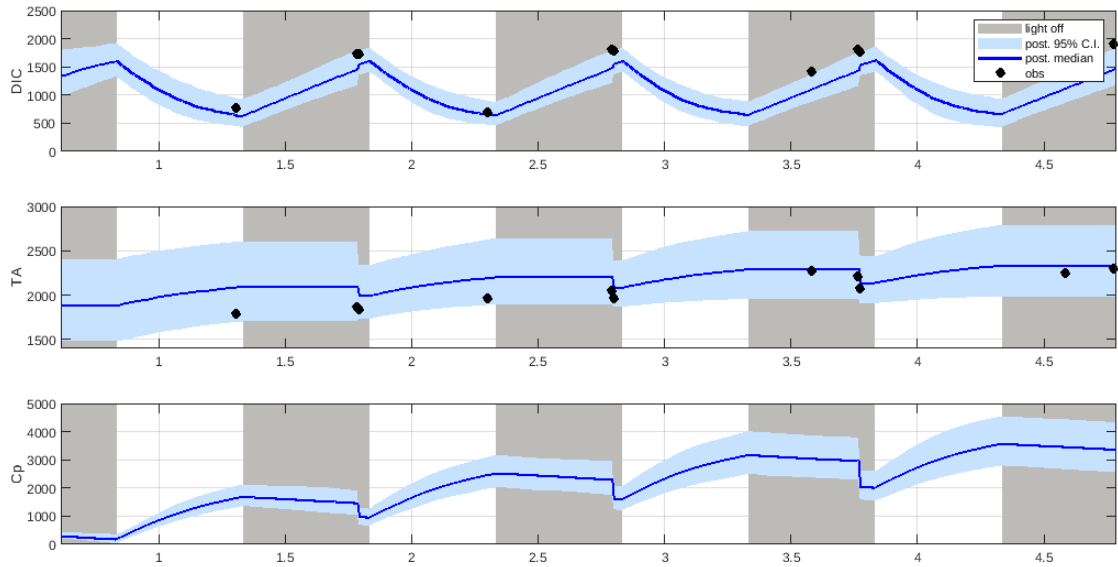


Figure 4.36 : Posterior medians (solid blue line), 95% credible intervals (shaded blue), and simulated observations (black) for  $DIC$ ,  $TA$  and  $C_p$  across 4 days.

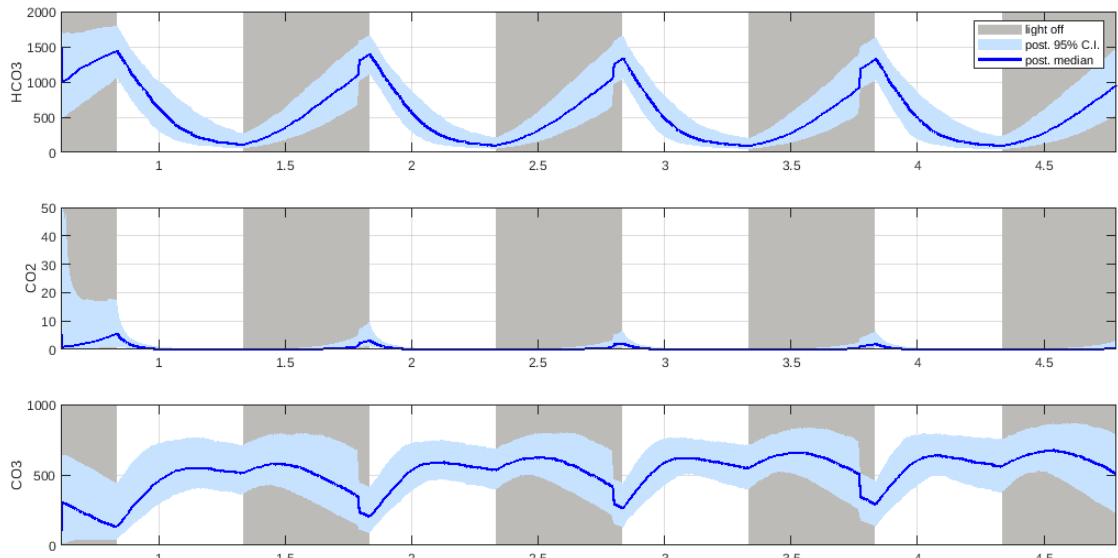


Figure 4.37 : Posterior medians (solid blue line) and 95% credible intervals (shaded blue) for  $HCO_3$ ,  $CO_2$  and  $CO_3$  across 4 days.

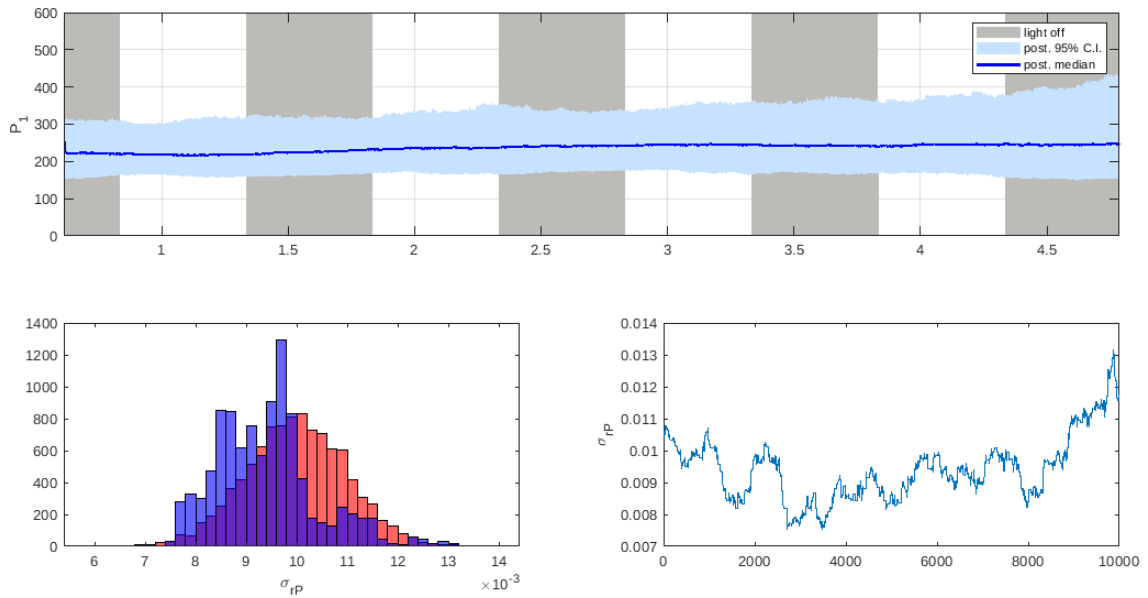


Figure 4.38 : Posterior medians (solid blue line) and 95% credible intervals (shaded blue) for photosynthesis  $P_1$ .



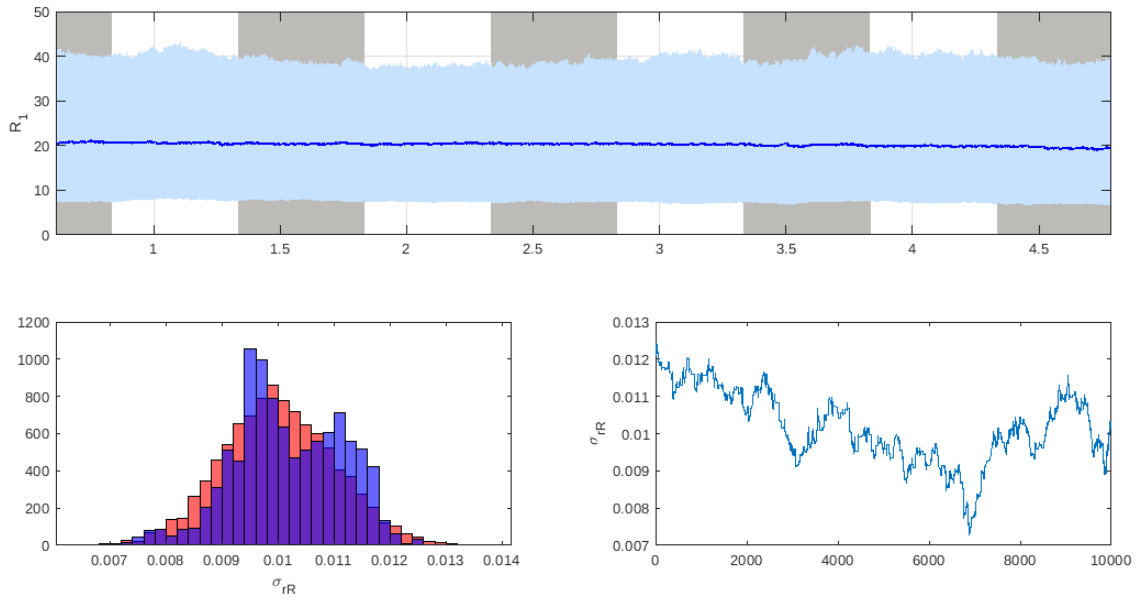


Figure 4.39 : Posterior medians (solid blue line) and 95% credible intervals (shaded blue) for respiration  $R_1$ .

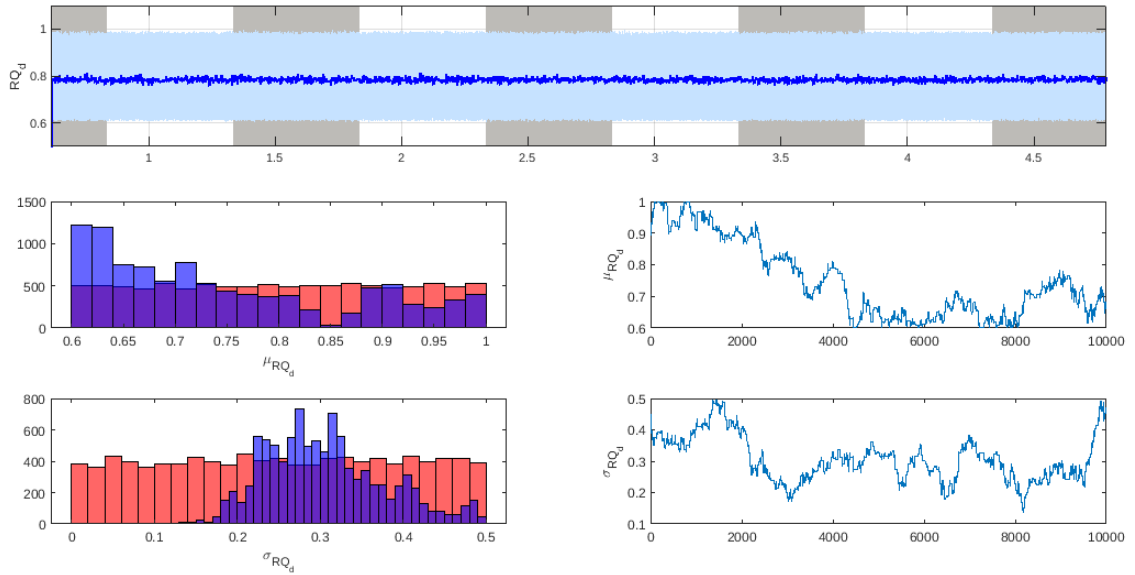


Figure 4.40 : Posterior medians (solid blue line) and 95% credible intervals (shaded blue) for  $RQ_d$ .

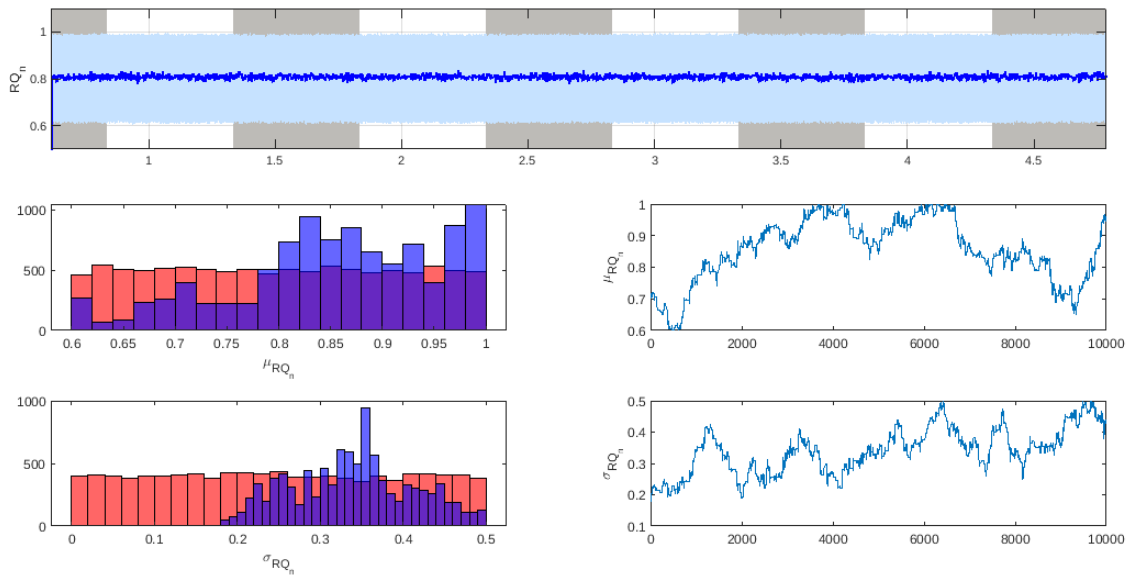


Figure 4.41 : Posterior medians (solid blue line) and 95% credible intervals (shaded blue) for  $RQ_n$ .

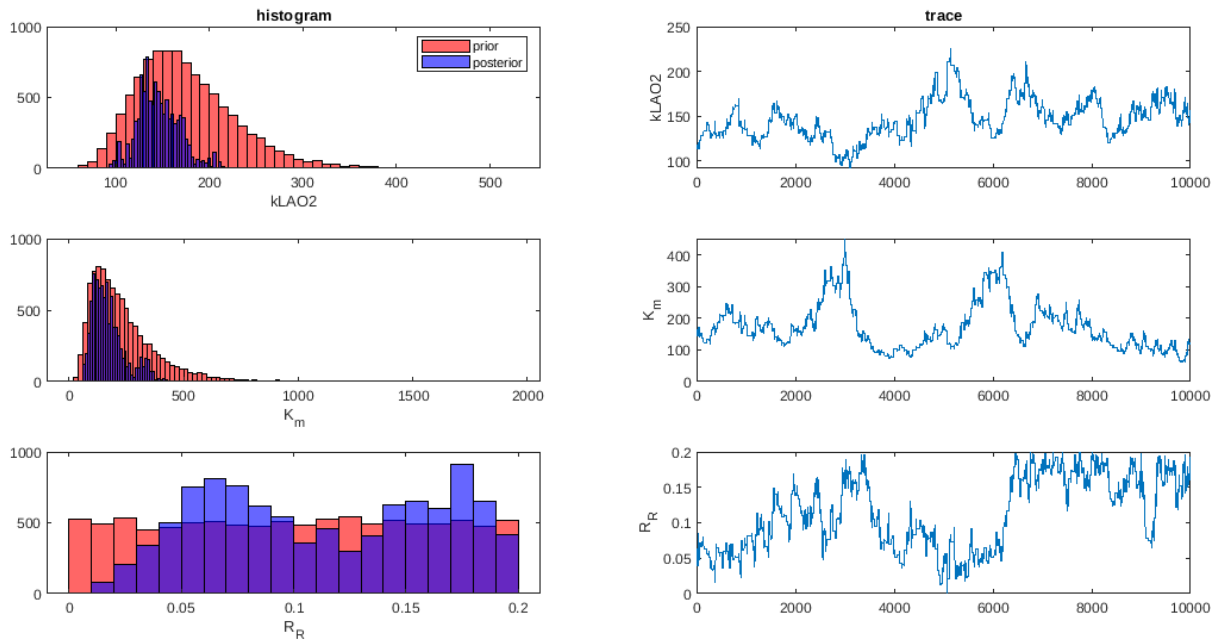


Figure 4.42 : Priors (pink) and posteriors (purple) for model parameters.

## 4.4 Discussion

Non destructive measurements allows us to use sensors to obtain these measurements instead of destructive sampling techniques. Meaning we can get informative estimates from well calibrated sensor measurements that combined with a data assimilating model can

Model states are sensitive to small perturbations in parameters. This is why instrument calibration for measurements is essential. [BM: maybe show a python run with different values of temp, constants etc to show the sensitivity? Or maybe just talk about it.]

Possibilities for RQn going well outside what the biology would allow. More processes taking place than what we currently have in the model description. We have a parameter  $RQ_n$  that can compensate for the input  $O_2$  gas line may have dropped or increased. In the model  $O_2$  concentration is treated as fixed when in reality this is an unmodelled process. If it's taking air in during the night time in the lab, it might go higher because there are less people there diurnal changes a possibility also, therefore can go up can go down. Henry's law shift future work: can measure this too.

when you constrain those two to be between 0.6 and 1, we see things like in the other runs where they slam up against the edge values of the priors and the sampling and mixing goes to shit.

sigma starts going to 0, for estimating the obs error.

bias in instrument errors in completely unmeasured, where the noise on the instrument error is easily measurable, bias on the other hand is not.

Big achievement for this chapter: putting the CO2SYS stuff (if you don't get CO2SYS right then you start getting all sorts of crazy values like negative values

etc. CO2SYS is especially sensitive because CO2 is what's controlling the fluxes, must be accurate, HCO3 is in the thousands so the errors need to be tiny because we have huge numbers for HCO3 and tiny numbers for CO2 need to be tiny, need to be really accurate. then pH which is on a log scale means that making a mistake at one part is completely different to making a mistake at another part.) into libbi which opens it up to details and nonlinear processes. getting this to work in the libbi constraints which are using gpus and very simple statistical models- letting us expand this in the mathematical sense for more complex models and equations. this would be the same limitation as stan on the gpu complexity of model side of things.

Discuss increasing the number of particles and changing the proposal distribution but how this had little effect on the mixing. Discuss mixing and correlation overall in reference to each results section.

Unbalanced weights of observations-  $O_2$  and  $pH$  have much denser datasets than  $DIC$  and  $TA$ . When these densely populated datasets are too frequent, this causes problems in the likelihood estimation [This issue of how to combine fuse different measurements with huge differences in sampling density is an interesting one]

Discuss failed experiments and why. Maybe add them to the appendix and reference them here.

## Appendix A

### LiBbi model code

LiBbi model file: micro\_iterative.bi

```

1 model micro_iterative {
2
3   const F02          = 0.2094
4   const FC02         = 397e-6
5   const S            = 34.0
6   const V            = 500.0           // volume of the
      reactor
7   const DIC_M        = 1724.20         // calculated with
      CO2SYS[ DIC_M = 1724.20, Alk = 1797.90, T = 27, S = 34]
8   const O_2_M        = 226.65
9   const alk_M         = 1797.90
10  const tau          = 6.0
11  const kLA02_m       = log(2.0)*24.0*60.0/tau
12
13  param kLA02
14  param Km
15  param RR
16  param RQ_d
17  param RQ_n
18  param sigma_O_2
19  param sigma_pH
20  param sigma_DIC

```

```
21 param offset_0_2
22
23 input I           // light intensity
24 input T           // temperature (C)
25 input gas         // gas on/off
26 input dil         // dilution rate
27
28 state DIC // state variables
29 state O_2
30 state pH
31 state Cp
32 state mich_ment
33 state O2H_pr
34 state CO2H_pr
35 state R
36 state R1
37 state P
38 state P1
39 state alk
40 state CO2
41 state HCO3
42 state CO3
43 state O_2H
44 state CO2H
45 state h_3
46 state h_free_3
47
48 noise r_R
49 noise r_P
```

```

50
51 /* random walk parameter */
52 param sigma_r_R
53 param sigma_r_P
54
55 obs O2_obs
56 obs pH_obs
57 obs DIC_obs
58 obs alk_obs
59
60 sub parameter { /* prior distribution over parameters */
61 Km      ~ log_normal(log(100.0), 0.5)
62 kLA02   ~ log_normal(log(kLA02_m), 0.3)
63 RR      ~ uniform(0.0001, 0.2)
64 RQ_d    ~ uniform(0.66, 1.0)
65 RQ_n    ~ uniform(0.66, 1.0)
66
67 sigma_O_2 ~ log_normal(log(0.03), 0.5)
68 sigma_pH  ~ log_normal(log(0.03), 0.5)
69 sigma_DIC ~ log_normal(log(0.03), 0.5)
70
71 offset_O_2 ~ normal(0, 2.0)
72
73 sigma_r_R      ~ normal(0.01, 0.001)
74 sigma_r_P      ~ normal(0.05, 0.01)
75 }
76
77 const prop_std = 0.1;
78 sub proposal_parameter {

```

```

79 Km      ~ log_normal(log(Km), 0.5*prop_std)
80 kLA02   ~ log_normal(log(kLA02), 0.3*prop_std)
81 RR      ~ truncated_normal(RR, 0.2*prop_std, lower = 0.0001,
      upper = 0.2)
82 RQ_d    ~ truncated_normal(RQ_d, 0.2*prop_std, lower = 0.66,
      upper = 1.0)
83 RQ_n    ~ truncated_normal(RQ_n, 0.2*prop_std, lower = 0.66,
      upper = 1.0)
84
85
86 sigma_0_2 ~ log_normal(log(sigma_0_2), 0.5*prop_std)
87 sigma_pH  ~ log_normal(log(sigma_pH), 0.5*prop_std)
88 sigma_DIC ~ log_normal(log(sigma_DIC), 0.5*prop_std)
89
90 offset_0_2 ~ normal(offset_0_2, 2.0*prop_std)
91
92 sigma_r_R ~ normal(sigma_r_R, 0.001*prop_std)
93 sigma_r_P ~ normal(sigma_r_P, 0.01*prop_std)
94 }
95
96 sub initial {/* prior distribution over initial conditions,
      given parameters */
97 // specify the initial condition model
98 R      ~ normal(log(20.0), 0.4)
99 R1     ~ log_normal(log(20.0), 0.4)
100 P      ~ normal(log(200.0), 0.4)
101 P1     ~ log_normal(log(200.0), 0.4)
102
103 Cp     ~ log_normal(log(300.0), 0.2)

```



```

104 alk      ~ log_normal(log(1750.0), 0.1)
105 DIC      ~ log_normal(log(1300.0), 0.2)
106 O_2      ~ log_normal(log(225.0), 0.2)
107 pH       ~ log_normal(log(8.5), 0.2)
108 CO2      ~ log_normal(log(3.0), 0.4)
109 HCO3     ~ log_normal(log(1000.0), 0.3)
110 CO3      ~ log_normal(log(300.0), 0.4)
111 O_2H     ~ log_normal(log(200.0), 0.2)
112 CO2H     ~ log_normal(log(10.0), 0.2)
113 }
114
115
116 //sub transition(delta = 0.0023) { // obs are in days ie
      delta=1.0 for daily solving. delta=0.00069 for solving
      every minute, 0.0014 for every 2 mins, 0.0021 for 3 mins,
      0.0028 for 4mins, delta=0.000011574 for solving every
      second
117 sub transition(delta = 0.0021) {
118
119 /* processes */
120
121 inline TK      = T + 273.15           // temp in kelvin
122 inline K0_CO2 = exp(-60.2409 + 93.4517*(100.0/TK) + 23.3585*
      log(TK/100.0)+ S*(0.023517 - 0.023656*(TK/100) +
      0.0047036*(TK/100.0)*(TK/100.0)))
123 CO2H          <- K0_CO2*FCO2*1.0220*1e6
124
125 inline K0_O2   = (exp(-1282.8704 + 36619.96/TK + 223.1396*log
      (TK) -0.354707*TK + S*(5.957e-3 -3.7353/TK) + 3.68e-6*S*S)

```

```

    )/(0.2094e-06)

126 O_2H          <- K0_02*F02*1.0220*1e-6
127
128 inline PAC      = HC03          //PAC=photosynthetically
    active carbon. if the phyto are just using CO2 to
    photosynthesise then PAC=CO2
129 inline mm       = PAC/(Km + PAC)
130
131 // CO2SYS iterative solution
132 // set up all the constants
133
134 inline logTK     = log(TK)
135 inline S2        = S*S
136 inline sqrtS     = sqrt(S)
137
138 // total sulphur
139
140 inline TS        = (0.14/96.062)*(S/1.80655)
141 inline IS        = 19.924*S/(1000.0 - 1.005*S)
142
143 inline KS_int    = -4276.1/TK + 141.328 - 23.093*logTK +
    (-13856.0/TK + 324.57 - 47.986*logTK)*sqrt(IS) + (
    35474.0/TK - 771.54 + 114.723*logTK)*IS - 2698.0/TK*IS
    **1.5 + 1776.0/TK*IS**2
144 inline KS       = exp(KS_int)*(1 - 0.001005*S)
145
146 // Fluorine
147
148 inline TF        = 0.000067*S/18.9984/1.80655

```

```

149 inline KF          = exp(-(-874.0/TK - 0.111*sqrtS + 9.68))
150 inline SWS_2_T      = (1.0 + TS/KS)/(1.0 + TS/KS + TF/KF)
151 inline Free_2_T     = 1.0 + TS/KS
152
153 // H2O dissoci
154
155 inline KW = exp(148.9802 - 13847.26/TK - 23.6521*logTK +
    (118.67/TK - 5.977 + 1.0495*logTK)*sqrtS - 0.01615*S)
156
157 // Boron
158
159 inline KB = exp((-8966.90 - 2890.53*sqrtS - 77.942*S + 1.728*
    S*sqrtS - 0.0996*S2)/TK + 148.0248 + 137.1942*sqrtS +
    1.62142*S - (24.4344 + 25.085*sqrtS + 0.2474*S)*logTK +
    0.053105*sqrtS*TK)
160 inline TB = 0.0004326*S/35.0
161
162 // Carbon eq constants
163
164 inline K1 = 10**(-(3633.86/TK - 61.2172 + 9.6777 *logTK -
    0.011555*S + 0.0001152*S**2))*1.23 //1.23 experiment
    specific and measured
165 inline K2 = 10**(-( 471.8/TK + 25.9290 - 3.16967*logTK -
    0.01781*S + 0.0001122*S**2))*0.53 //0.53 experiment
    specific and measured
166
167 // end all the constants
168
169 // intial guess at the pH (use the approximating equation)

```

```

170
171 inline pH_init = 12.26 -0.0030605*DIC -0.043752*T -0.013625*S
      + 0.00011315*alk + 1.3463e-05*DIC*T + 5.2215e-07*DIC*alk
172
173 // iteration 1
174
175 inline h_1      = 10.0**(-pH_init)
176 inline h_free_1 = h_1/Free_2_T
177 inline f0_1      = (DIC*1e-6*(K1*h_1 + 2.0*K1*K2)/(h_1*h_1 +
      K1*h_1 + K1*K2) - h_free_1 + KW/h_1 - alk*1e-6 + TB/(1.0 +
      h_1/KB))*1e6
178 inline df0_1     = (DIC*1e-6*(K1 + 2.0*K1*K2)/(h_1**2.0 + K1*
      h_1 + K1*K2) - DIC*1e-6*(K1*h_1 + 2.0*K1*K2)/(h_1**2.0 +
      K1*h_1 + K1*K2)**2.0*(2.0*h_1 + K1) - TB*1.0/(1.0 + h_1/KB
      )**2.0/KB - KW/h_1**2.0 - 1.0/Free_2_T)*1e6*(-log(10.0)
      *10.0**(-pH_init))
179 inline pH_1      = pH_init - f0_1/df0_1
180
181 // iteration 2
182
183 inline h_2      = 10.0**(-pH_1)
184 inline h_free_2 = h_2/Free_2_T
185 inline f0_2      = (DIC*1e-6*(K1*h_2 + 2.0*K1*K2)/(h_2*h_2 +
      K1*h_2 + K1*K2) - h_free_2 + KW/h_2 - alk*1e-6 + TB/(1.0 +
      h_2/KB))*1e6
186 inline df0_2     = (DIC*1e-6*(K1 + 2.0*K1*K2)/(h_2**2.0 + K1*
      h_2 + K1*K2) - DIC*1e-6*(K1*h_2 + 2.0*K1*K2)/(h_2**2.0 +
      K1*h_2 + K1*K2)**2.0*(2.0*h_2 + K1) - TB*1.0/(1.0 + h_2/KB
      )**2.0/KB - KW/h_2**2.0 - 1.0/Free_2_T)*1e6*(-log(10.0)

```

```

      *10.0**(-pH_1))
187 inline pH_2      = pH_1 - f0_2/df0_2
188
189 // iteration 3
190
191 h_3                <- 10.0**(-pH_2)
192 h_free_3           <- h_3/Free_2_T
193 inline f0_3        = (DIC*1e-6*(K1*h_3 + 2.0*K1*K2)/(h_3*h_3 +
      K1*h_3 + K1*K2) - h_free_3 + KW/h_3 - alk*1e-6 + TB/(1.0 +
      h_3/KB))*1e6
194 inline df0_3       = (DIC*1e-6*(K1 + 2.0*K1*K2)/(h_3**2.0 + K1*
      h_3 + K1*K2) - DIC*1e-6*(K1*h_3 + 2.0*K1*K2)/(h_3**2.0 +
      K1*h_3 + K1*K2)**2.0*(2.0*h_3 + K1) - TB*1.0/(1.0 + h_3/KB
      )**2.0/KB - KW/h_3**2.0 - 1.0/Free_2_T)*1e6*(-log(10.0)
      *10.0**(-pH_2))
195 pH                <- pH_2 - f0_3/df0_3
196
197 // iteration 4
198
199 //      inline h_4      = 10.0**(-pH_3)
200 //      inline h_free_4 = h_4/Free_2_T
201 //      inline f0_4      = (DIC*1e-6*(K1*h_4 + 2.0*K1*K2)/(h_4
      *h_4 + K1*h_4 + K1*K2) - h_free_4 + KW/h_4 - alk*1e-6 + TB
      /(1.0 + h_4/KB))*1e6
202 //      inline df0_4     = (DIC*1e-6*(K1 + 2.0*K1*K2)/(h_4
      **2.0 + K1*h_4 + K1*K2) - DIC*1e-6*(K1*h_4 + 2.0*K1*K2)/(
      h_4**2.0 + K1*h_4 + K1*K2)**2.0*(2.0*h_4 + K1) - TB
      *1.0/(1.0 + h_4/KB)**2.0/KB - KW/h_4**2.0 - 1.0/Free_2_T)
      *1e6*(-log(10.0)*10.0**(-pH_3))

```

```

203 //      inline pH_4      = pH_3 - f0_4/df0_4
204
205 //      pH                <- pH_4
206
207 // calculate the final concentrations
208
209 inline H      = 10.0**(-pH)
210 inline H2     = H*H
211 inline denom  = (H2 + K1*H + K1*K2)
212 CO2          <- DIC*H2/denom
213 HCO3         <- DIC*H*K1/denom
214 CO3          <- DIC*K1*K2/denom
215
216 // end CO2SYS iterative solution
217
218
219 /* R and P as random walks */
220
221 r_R      ~ normal(0.0, sigma_r_R)
222 R        <- R + r_R
223 R1       <- exp(R)
224
225 r_P      ~ normal(0.0, sigma_r_P)
226 P        <- P + r_P
227 P1       <- exp(P)
228
229 ode(h = 0.1, atoler = 1.0e-6, rtoler = 1.0e-6, alg = 'RK4(3)
      '){

```

```

230 dDIC/dt = -P1*24.0*I*mm + R1*24.0
          + gas*0.893*kLA02*(C02H - C02)          + dil/V*(
          DIC_M - DIC)
231 dO_2/dt = (P1*24.0*I*mm - R1*24.0)/(RQ_d*I + RQ_n*(1.0-I))
          + gas*kLA02*(O_2H - O_2)          + dil/V*(O_2_M
          - O_2) + offset_O_2
232 dalk/dt = RR*P1*24.0*I*mm
          + dil/V*(
          alk_M - alk)
233 dCp/dt = (P1*24.0*I*mm - R1*24.0)
          + dil/V*(
          Cp)
234
235 }
236
237 mich_ment <- mm
238 O2H_pr <- O_2H
239 C02H_pr <- C02H
240
241 }
242
243
244 sub observation {
245
246 O2_obs ~ log_normal(log(O_2), sigma_O_2)
247 pH_obs ~ log_normal(log(pH), sigma_pH)
248 DIC_obs ~ log_normal(log(DIC), sigma_DIC)
249 alk_obs ~ log_normal(log(alk), sigma_DIC)
250 }

```

251 }

### LiBbi prior sampling file: prior.conf

```
1 --target prior
2 --model-file micro_iterative.bi
3 --nsamples 500
4 --start-time 0.61304
5 --end-time 4.7866
6 --noutputs 6049
7 --input-file data/input_all_2018_normalised.nc
8 --output-file results/prior_micro_iterative.nc
```

### LiBbi posterior sampling file: posterior.conf

```
1 --target posterior
2 --model-file micro_iterative.bi
3 --input-file data/input_all_2018_normalised.nc
4 --obs-file data/obs_all_2018.nc
5 --nsamples 500
6 --nparticles 1024
7 --start-time 0.61304
8 --end-time 4.7866
9 --noutputs 6049
10 --output-file results/posterior_micro_iterative.nc
11 --with-transform-initial-to-param
```



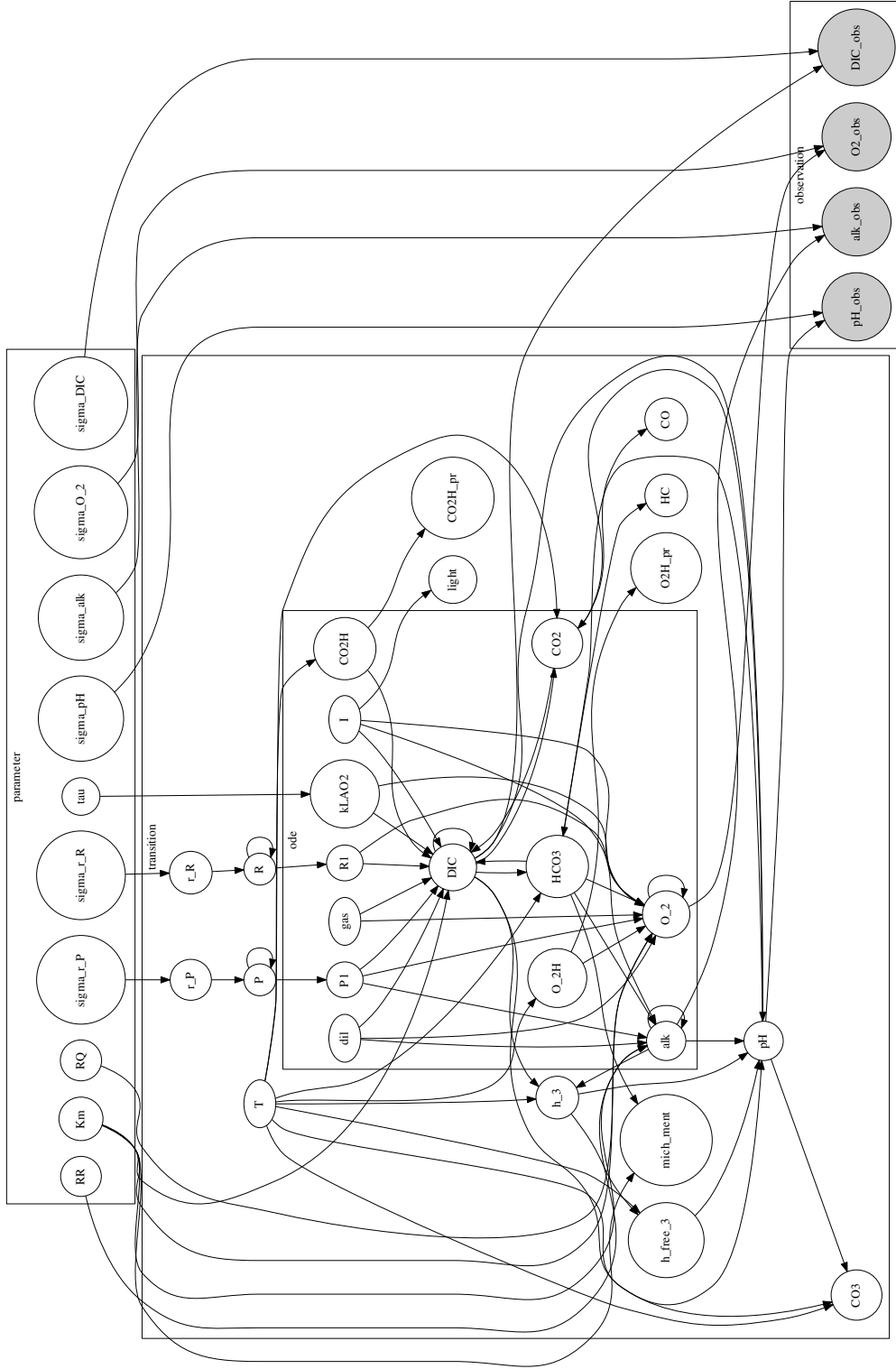


Figure A.1 : Directed Acyclic Graph of the LiBbi model file micro\_iterative.bi

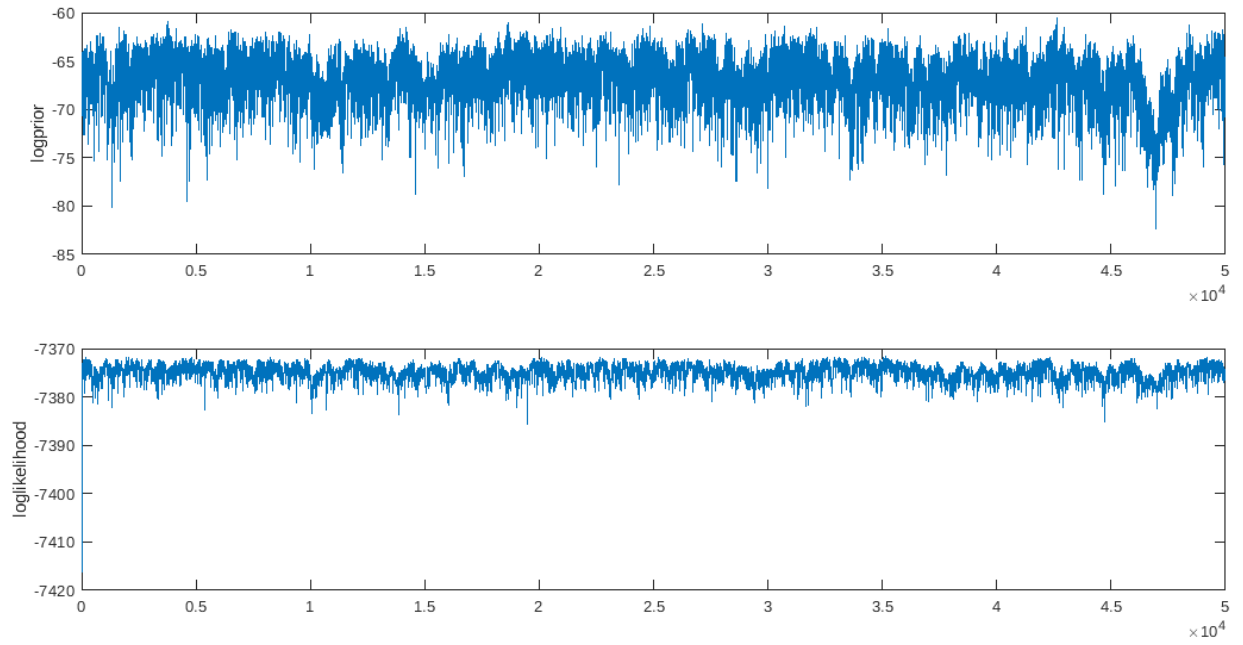


Figure A.2 : Log-prior and log-likelihood for the simulated data experiment.

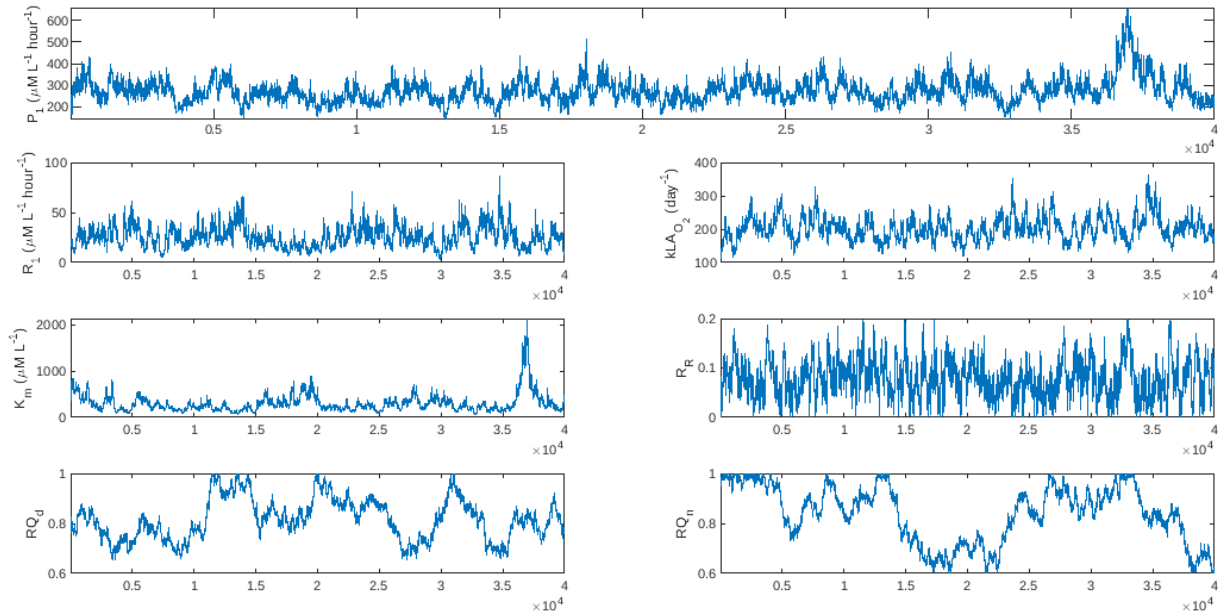


Figure A.3 : Parameter posterior traces for the simulated data experiment.

## Appendix B

### Posteriors with experimental data (photosynthesis, respiration and respiratory quotients are random walks, and an offset, and estimating obs error)

[BM: this is using the further thinned out observations too]

State posteriors are visualised by plotting the median and shading 95% credible intervals, while parameter priors and posteriors are displayed by histograms.

#### In this run:

Photosynthesis ( $P_1$ ) and respiration ( $R_1$ ) are both modelled as random walks, by taking  $P$  and  $R$ , previously constant parameters, and replacing them by  $P_1(t)$  and  $R_1(t)$ . Here, we take  $P_1(t)$  and  $R_1(t)$  to be such that

$$P_1(t + \Delta t) = P(t) + r_P$$

$$R_1(t + \Delta t) = R(t) + r_R$$

where  $r_P \sim N(0, \sigma_{r_P})$ ,  $r_R \sim N(0, \sigma_{r_R})$ , and  $\Delta t$  is the length of discrete time-step. For the purpose of the Bayesian analysis here,  $\sigma_{r_P}$  and  $\sigma_{r_R}$  are treated as parameters to be inferred. The respiratory quotients  $RQ_d$  and  $RQ_n$  were also treated as random walks with  $rP$  and  $rR$  as wiener processes.  $kLA_{O_2}$ ,  $K_m$ ,  $R_R$ ,  $\sigma_{rP}$ ,  $\sigma_{rR}$  are all treated as parameters constant through time but unknown. The data model assigned log normally distributed observation errors for each instrument;

$$O_{2_{obs}} \sim \text{Log}\mathcal{N}(\log(O_2), \sigma_{O_2})$$

$$pH_{obs} \sim \text{Log}\mathcal{N}(\log(pH), \sigma_{pH})$$

$$DIC_{obs} \sim \text{Log}\mathcal{N}(\log(DIC), \sigma_{DIC})$$

$$TA_{obs} \sim \text{Log}\mathcal{N}(\log(TA), \sigma_{DIC})$$

Parameter	Prior	Proposal
$kLA_{O_2}$	$\text{Log}\mathcal{N}(\log(200.0), 0.3)$	$\text{Log}\mathcal{N}(\log(kLA_{O_2}), 0.03)$
$K_m$	$\text{Log}\mathcal{N}(\log(200.0), 0.6)$	$\text{Log}\mathcal{N}(\log(K_m), 0.06)$
$R_R$	$\text{Uniform}(0, 0.2)$	$\text{Trun}\mathcal{N}(R_R, 0.005, \text{lower} = 0, \text{upper} = 0.2)$
$\sigma_{r_P}$	$\mathcal{N}(0.02, 0.002)$	$\mathcal{N}(\sigma_{r_P}, 0.0002)$
$\sigma_{r_R}$	$\mathcal{N}(0.01, 0.001)$	$\mathcal{N}(\sigma_{r_R}, 0.0001)$
$offset_{O_2}$	$\mathcal{N}(0, 5.0)$	$\mathcal{N}(offset_{O_2}, 0.5)$
$\sigma_{O_2}$	0.1	*
$\sigma_{pH}$	0.1	*
$\sigma_{DIC}$	0.2	*

Table B.1 : Table of Parameters, their priors and proposal distributions. \* indicates the parameter was held fixed.

Parameter	Quantiles (25%, 75%)	Quantiles (5%, 95%)
$kLA_{O_2}^{air}$	()	()
$K_m$	()	()
$R_R$	()	()
$\sigma_{r_P}$	()	()
$\sigma_{r_R}$	()	()
$offset_{O_2}$	()	()
$\sigma_{O_2}$	()	()
$\sigma_{pH}$	()	()
$\sigma_{DIC}$	()	()

Table B.2 : Posterior (25%, 75%), (5%, 95%) quantiles for parameters after assimilating observations.

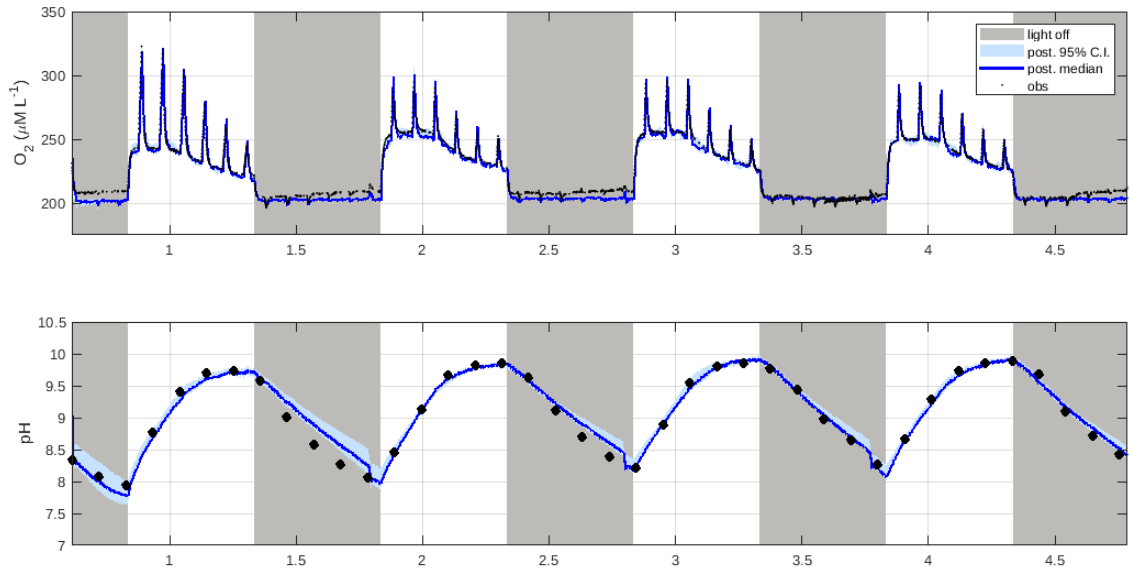


Figure B.1 : Posterior medians (solid blue line), 95% credible intervals (shaded blue), and simulated observations (black) for  $O_2$  and  $pH$  across 4 days.

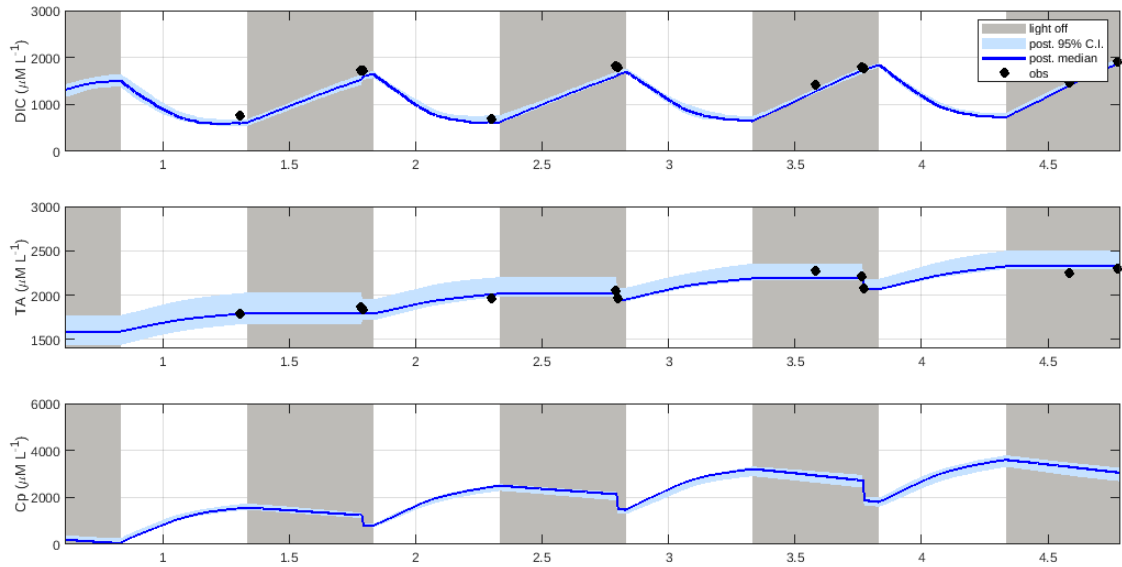


Figure B.2 : Posterior medians (solid blue line), 95% credible intervals (shaded blue), and simulated observations (black) for  $DIC$ ,  $TA$  and  $C_p$  across 4 days.

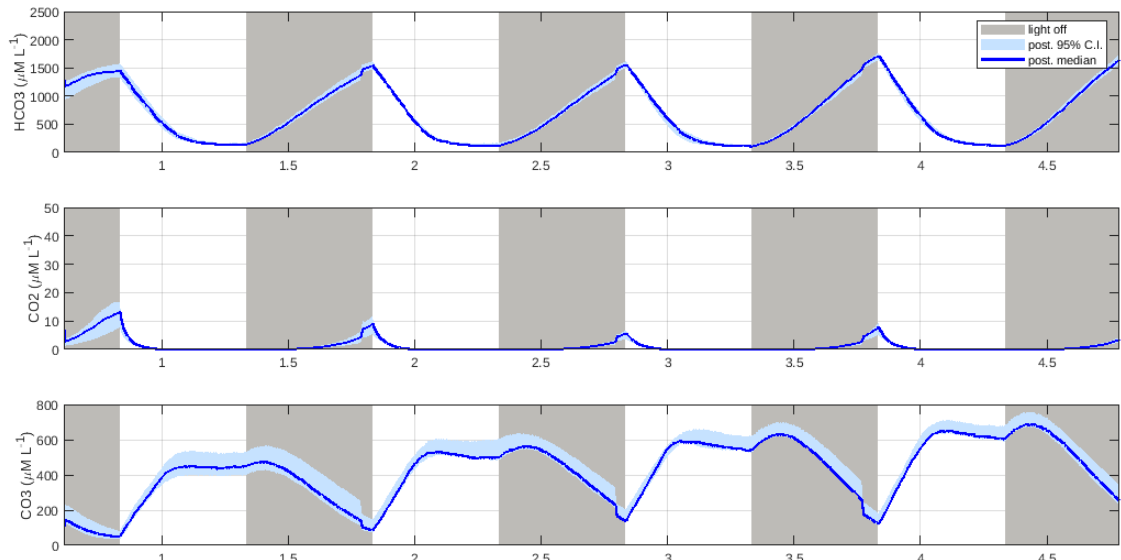


Figure B.3 : Posterior medians (solid blue line) and 95% credible intervals (shaded blue) for  $HCO_3^-$ ,  $CO_2$  and  $CO_3$  across 4 days.

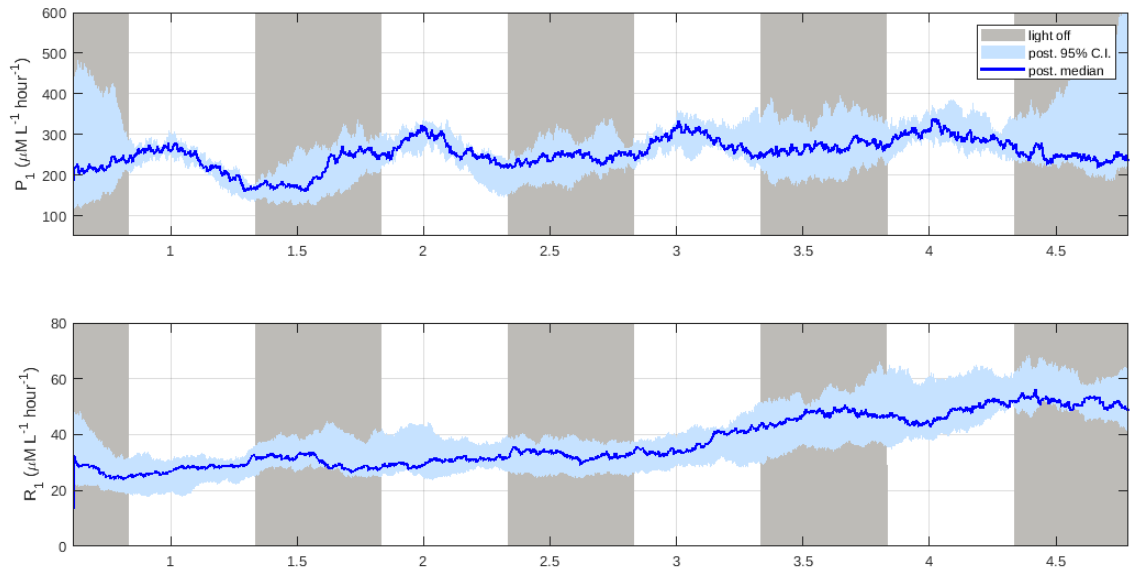


Figure B.4 : Posterior medians (solid blue line) and 95% credible intervals (shaded blue) for photosynthesis  $P_1$  and respiration  $R_1$ .

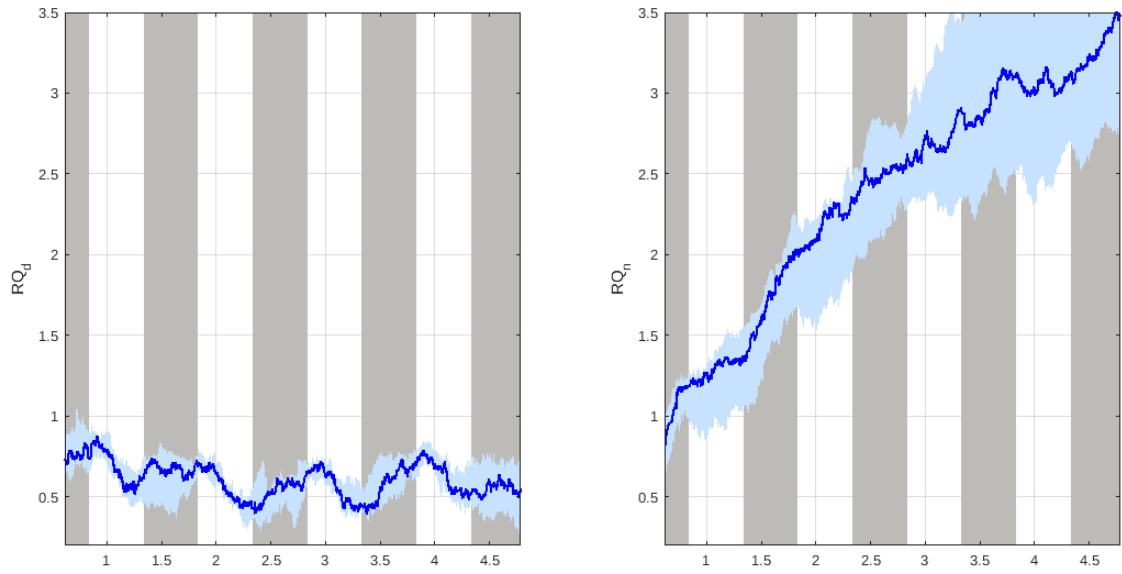


Figure B.5 : Posterior medians (solid blue line) and 95% credible intervals (shaded blue) for  $RQ_d$  and  $RQ_n$ .

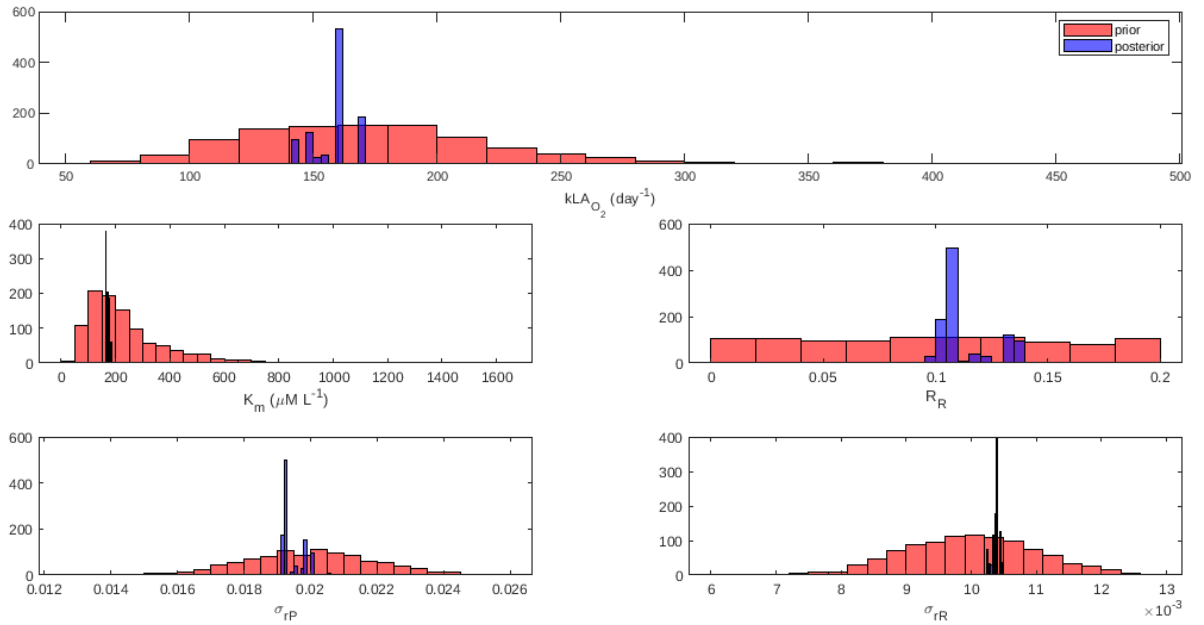


Figure B.6 : Priors (pink) and posteriors (purple) for model parameters.

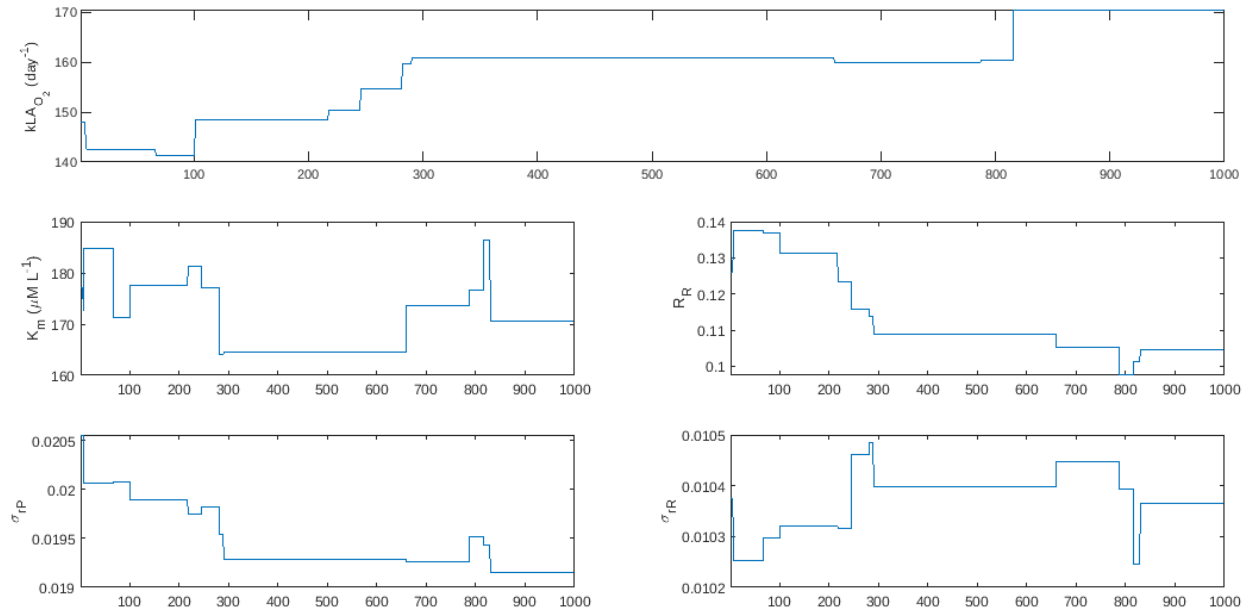


Figure B.7 : Traces for model parameters.

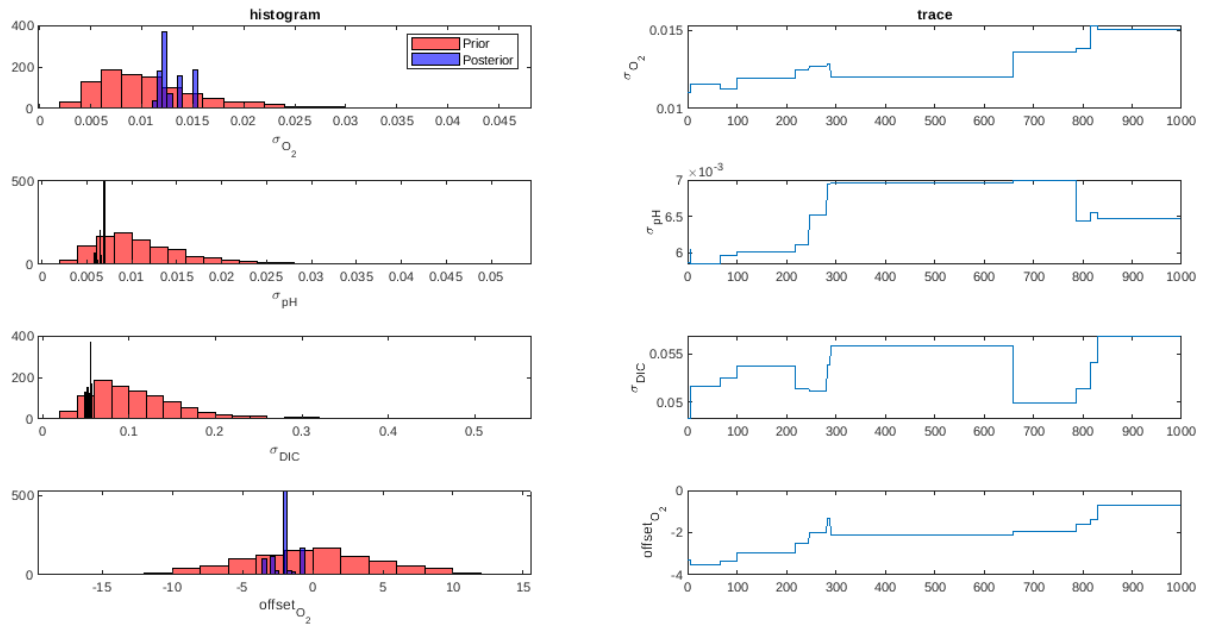


Figure B.8 : Priors (pink), posteriors (purple), and traces for observation error and offset model parameters.



**Results:**

The further thinned observations were used to test 1,000 samples with 2048 particles, returning an acceptance rate of 1.3%.

## Bibliography

- [1] Rubin Battino, Timothy R Rettich, and Toshihiro Tominaga. The solubility of oxygen and ozone in liquids. *Journal of physical and chemical reference data*, 12(2):163–178, 1983.
- [2] AG Dickson and FJ Millero. A comparison of the equilibrium constants for the dissociation of carbonic acid in seawater media. *Deep Sea Research Part A. Oceanographic Research Papers*, 34(10):1733–1743, 1987.
- [3] Andrew Gilmore Dickson, Christopher L Sabine, and James Robert Christian. *Guide to best practices for ocean CO<sub>2</sub> measurements*. Number PICES Special Publication 3. North Pacific Marine Science Organization, 2007.
- [4] Arnold E Greensberg, LS Clesceri, Andrew D Eaton, and MAH Franson. Standard methods for the examination of water and wastewater. *American Public Health Asociation, Whashington, DC*, 1992.
- [5] E Molina Grima, JA Sánchez Pérez, F Ía Garc Camacho, and A Robles Medina. Gas-liquid transfer of atmospheric co<sub>2</sub> in microalgal cultures. *Journal of Chemical Technology & Biotechnology*, 56(4):329–337, 1993.
- [6] Robert RL Guillard and John H Ryther. Studies of marine planktonic diatoms: I. cyclotella nana hustedt, and detonula confervacea (cleve) gran. *Canadian journal of microbiology*, 8(2):229–239, 1962.
- [7] Ernie Lewis, Doug Wallace, and Linda J Allison. Program developed for co {sub 2} system calculations. Technical report, Brookhaven National Lab., Dept. of Applied Science, Upton, NY (United States . . . , 1998.

- [8] Carl Mehrbach, CH Culberson, JE Hawley, and RM Pytkowicz. Measurement of the apparent dissociation constants of carbonic acid in seawater at atmospheric pressure<sup>1</sup>. *Limnology and Oceanography*, 18(6):897–907, 1973.
- [9] Clemens Posten. Design principles of photo-bioreactors for cultivation of microalgae. *Engineering in Life Sciences*, 9(3):165–177, 2009.
- [10] Niels Ramsing and Jens Gundersen. Seawater and gases. *Limnol. Oceanogr*, 37:1307–1312, 2011.
- [11] R.F Weiss. Carbon dioxide in water and seawater: the solubility of a non-ideal gas. *Marine chemistry*, 2(3):203–215, 1974.
- [12] Richard E Zeebe and Dieter Wolf-Gladrow. *CO<sub>2</sub> in seawater: equilibrium, kinetics, isotopes*. Number 65. Gulf Professional Publishing, 2001.