---

**Article:**
Xiao, Y., Sloan, J. orcid.org/0000-0003-0334-3722, Hepworth, C. et al. (4 more authors) (2021) Estimating uncertainty: A Bayesian approach to modelling photosynthesis in C3 leaves. Plant, Cell & Environment, 44 (5). pp. 1436-1450. ISSN 0140-7791

https://doi.org/10.1111/pce.13995

---

# Estimating Uncertainty: a Bayesian Approach to Modelling Photosynthesis in C3 Leaves

Yi Xiao[1,4], Jen Sloan[2,4], Chris Hepworth[2], Colin P. Osborne[2], Andrew J. Fleming[2], Xingyuan Chen[3], Xin-Guang Zhu[1]

[1]Center of Excellence for Molecular Plant Science, Institute of Plant Physiology and Ecology, Chinese Academy of Sciences, Shanghai, 200032, China

[2]Department of Animal and Plant Sciences, University of Sheffield, Western Bank, Sheffield, UK

[3]Pacific Northwest National Laboratory, Richland, WA, 99354, USA

[4]Joint first author

Author for Correspondence:

Xin-Guang Zhu

Center of Excellence for Molecular Plant Science, Institute of Plant Physiology and Ecology, Chinese Academy of Sciences, Shanghai, 200032, China

Email: zhuxg@sippe.ac.cn

Fax:

Tel:

**Abstract**

The Farquhar-von Caemmerer-Berry (FvCB) model is extensively used to model photosynthesis from gas exchange measurements. Since its publication, many methods have been developed to measure, or more accurately estimate, parameters of this model. Here we have created a tool that uses Bayesian statistics to fit photosynthetic parameters using concurrent gas exchange and chlorophyll fluorescence measurements whilst evaluating the reliability of the parameter estimation. We have tested this tool on synthetic data and experimental data from rice leaves. Our results indicate that reliable parameter estimation can be achieved whilst only keeping one parameter, $K_m$, i.e., Michaelis constant for $CO_2$ by Rubisco, prefixed. Additionally, we show that including detailed low $CO_2$ measurements at low light levels increases reliability, and suggest this as a new standard measurement protocol. By providing an estimated distribution of parameter values, the tool can be used to evaluate the quality of data from gas exchange and chlorophyll fluorescence measurement protocols. Compared to earlier model fitting methods, the use of a Bayesian statistics-based tool minimises human interaction during fitting, reducing the subjectivity which is essential to most existing tools. A user friendly, interactive Bayesian tool script is provided.

**Key Words:** leaf photosynthesis, parameter estimation, Bayesian statistics, mesophyll conductance

Interactive Bayesian Tool: https://github.com/xiaoyizz78/FvCB-JAGS

## Introduction

Since its publication in 1980, the Farquhar-von Caemmerer-Berry (FvCB) model has been widely used to model leaf photosynthesis in $C_3$ plants (Farquhar, von Caemmerer & Berry 1980, 2001; von Caemmerer 2013). By assuming leaf photosynthetic rate ($A_n$) (see **Table 1** for definition of terms) is either limited by the Rubisco catalysed carboxylation rate or the regeneration rate of ribulose 1,5-bisphosphate (RuBP), the model derives an elegant and powerful expression of $A_n$ in response to environmental $CO_2$ conditions. Estimating the parameters in the model brings insight into the processes limiting photosynthetic gas exchange. Parameters such as $V_{cmax}$, $K_m$ and $\Gamma^*$ improve our understanding of photosynthetic limitations from Rubisco. Calculation of $J$ allows an estimation of the conversion efficiency from light to RuBP regeneration, while $g_m$ can be used to quantify the diffusive resistance of $CO_2$ from the substomatal cavity to the chloroplast stroma. Such modelling is crucial for a number of aspects of photosynthesis research including, for example, understanding acclimation or adaptation of photosynthesis to environmental change (e.g. Bernacchi *et al.*, 2005), and the response of photosynthetic properties to targeted genetic manipulation (e.g. Simkin *et al.*, 2015; Perveen *et al.*, 2020). The accuracy of the estimations made via the FvCB model also inform many predictions in ecological studies where the FvCB model has been widely incorporated and used to predict instantaneous $CO_2$ assimilation at the canopy level or integrated carbon assimilation over an entire growing season (Humphries & Long 1995; Wang *et al.* 2015; Golaz *et al.* 2019; Lawrence *et al.* 2019).

63 **The challenges of parameterising the FvCB model**

64 Despite its widespread use, challenges remain in parameterising the model, and various
65 analytical and numerical methods have been developed using data from gas exchange
66 and/or chlorophyll fluorescence measurements (Harley, Loreto, Di Marco & Sharkey
67 1992; Ethier & Livingston 2004; Ethier, Livingston, Harrison, Black & Moran 2006;
68 Dubois, Fiscus, Booker, Flowers & Reid 2007; Sharkey, Bernacchi, Farquhar &
69 Singsaas 2007; Yin & Struik 2009; Gu, Pallardy, Tu, Law & Wullschleger 2010;
70 Bellasio, Beerling & Griffiths 2015; Sharkey 2016; Moualeu-Ngangue, Chen & Stützel
71 2017). These methods differ in the measurements taken for parameter estimation, the
72 assumptions made during estimation and the fitting strategies used.

73 Basic models use only $A_n$-$C_i$ curves as input data, with fixed $K_m$ and $\Gamma^*$ values, requiring
74 input from the user on the limitation status for each data point, with many parameters,
75 including $V_{cmax}$, $J$ and $g_m$, being fitted (Ethier & Livingston 2004; Ethier *et al.* 2006;
76 Sharkey *et al.* 2007; Sharkey 2016). For a more complete estimation, concurrent
77 measurements of $Y(II)$-$C_i$ curves can be included in the parameter estimation (Bongi &
78 Loreto 1989; Di Marco et al., 1990; Harley et al., 1992). Then the accuracy of parameter
79 estimation is particularly affected by the method used to calculate $J$ from measured
80 $Y(II)$. Harley *et al.* (1992) calibrated the calculation of $J$ with concurrent measurement
81 of $A_n$ and $Y(II)$ under non-photorespiratory conditions; this calibration, which requires
82 additional measurements, was simplified in many subsequent studies (e.g. Bernacchi et
83 al., 2002; Sun et al., 2014) with implicit assumptions. For example, although the leaf

light absorption coefficient $\alpha$ can be measured with an integrating sphere, the light partition coefficient $\beta$ cannot, and is usually fixed at 0.5, i.e., assuming that PSI and PSII each receive half of the absorbed photons. When $s$ ($\alpha$ multiplied by $\beta$) is measured, values range between 0.35 to 0.45 (Valentini *et al.* 1995) and 0.42 to 0.6 (Laisk & Loreto 1996), depending on the species investigated.

To correct the assumptions made on values of $s$, $K_m$ and $\Gamma^*$, Laisk *et al.* (2002, 2006) made additional measurements of the initial slopes of $A_n$-$C_i$ and $Y(II)$-$C_i$ under different $O_2$ levels, followed by an iterative fitting strategy for the estimation of $s$, $K_m$, $\Gamma^*$ and the estimation of $V_{cmax}$, $J$, and $g_m$. Yin *et al.* (2009a, 2009b) developed an alternative strategy in which $s$, $K_m$ and $\Gamma^*$ were first estimated from concurrent $A_n$-$C_i$ and $Y(II)$-$C_i$ measurements at a low $C_i$ range under both normal and low $O_2$ levels. A regression method, similar to Dubois *et al.* (2007), was then applied to fit the concurrent $A_n$-$C_i$ and $Y(II)$-$C_i$ measurements under saturating light and normal $O_2$, thus obtaining the remaining parameters, including $V_{cmax}$, $g_m$ and $J$. This pipeline was later incorporated into an Excel worksheet by Bellasio *et al.* (2015), where input data required both $CO_2$ response curves and light response curves from concurrent $A_n$-$C_i$ and $Y(II)$-$C_i$ measurements under both ambient and low oxygen levels. These more recent methods clearly improve model accuracy, but data collection becomes increasingly time-consuming, limiting their widespread application. Finally, although it has long been recognized that the performance of these different methods is affected by the choices of fixed parameters and the different fitting strategies implemented (Manter & Kerrigan

2004; Miao, Xu, Lathrop & Wang 2009; Gu & Sun 2014), to date the reliability and robustness of parameter estimation of the FvCB model have not been systematically evaluated.

The accuracy of parameter estimation can be tested using synthetic data. For example, Gu *et al.* (2010) generated $A_n$-$C_i$ data without variation using the FvCB model and applied a parameter fitting technique, demonstrating that their method was able to predict an unbiased parameter estimation from error-free measurements. However, for data with sample variation or error, and for estimation with concurrent $A_n$-$C_i$ and $Y(II)$-$C_i$ measurements, quantification of the accuracy and robustness of the parameter fitting of different methods has not been performed.

**The Bayesian Approach**

Bayesian estimation is a powerful statistical approach to address many of the issues described above. It uses prior (already known) data to create a series of possible estimations (the posterior), which in turn are used to shape future estimations. Comparing the variability within the estimated parameter range gives an indication of the reliability of the estimation. The popularity of Bayesian inference is linked to the Markov Chain Monte Carlo (MCMC) method, which provides an effective sampling strategy to approximate the posterior distribution (Gelfand & Smith 1990; Smith & Roberts 1993; Tierney & Mira 1999; Andrieu, de Freitas, Doucet & Jordan 2003), especially for models with small sets of parameters, such as the FvCB model. A Bayesian framework with MCMC methods has been applied successfully in many

126 fields (Clark & Gelfand 2006; Chen X., Rubin, Ma & Baldocchi 2008), and it has also
127 been applied in large scale ecological studies which estimate photosynthetic parameters
128 with $A_n$-$C_i$ and/or $A_n$-$I$ data collected from tens of species across a growing season
129 (Patrick, Ogle & Tissue 2009; Feng & Dietze 2013; Han et al., 2020). Here we apply
130 the Bayesian approach to the leaf-by-leaf photosynthetic parameter estimation with
131 concurrent $A_n$-$C_i$ and $Y(II)$-$C_i$ measurements.
132 In this study, each parameter of the FvCB model is initially constrained to a range of
133 possible values represented by an *a priori* probability distribution, then Bayesian
134 statistics are used to calculate the posterior probability of parameters based on the prior
135 information, the observation, and a probability model of observation. We use synthetic
136 concurrent $A_n$-$C_i$ and $Y(II)$-$C_i$ with 5 replicates, as well as experimental data from rice
137 leaves, to test the performance of the new Bayesian estimation tool. In particular, using
138 100 synthetic datasets of widely varied photosynthetic parameters, our tool is compared
139 with a simple fitting method to highlight the ability of the new tool to calculate the
140 reliability of the estimated parameters. In addition, we evaluate the trade-off between
141 time-consuming measurements and the accuracy of parameter estimation, comparing
142 estimation using only high light $A_n$-$C_i$ and $Y(II)$-$C_i$ data with estimation using additional
143 low $CO_2$ and low light data. Finally, we report on the number of parameters which must
144 be known and fixed in order to give accurate estimations using the synthetic data.
145 Testing the estimation technique with both synthetic and experimental data not only
146 increases our understanding of how accurate the Bayesian estimation is (given that the

model is true), but also gives insight into how well the FvCB model reflects reality. The overall aim of the investigation is to establish a Bayesian framework to estimate parameter values, as well as to evaluate the robustness and reliability of parameter estimation of the FvCB model. Factors limiting the accuracy of estimation are identified. The tool is incorporated into an easy-to-use Bayesian parameter estimation script for use with concurrent $A_n$-$C_i$ and $Y(II)$-$C_i$ measurements.

**Computational Methods:**

**The Farquhar-von Caemmerer-Berry model for $CO_2$ assimilation rate and quantum efficiency of PSII**

An adapted version of the FvCB model is used, after von-Caemmerer (2000) and Gu *et al.*, 2010 (Eqns 1-3). Leaf net photosynthesis rate ($A_n$) is modeled as the minimum of the Rubisco-limited carboxylation rate ($W_c$) and the ribulose 1,5-bisphosphate (RuBP) regeneration-limited carboxylation rate ($W_j$). To make the analysis of the relationship between Bayesian estimation and the model structure clearer, triose phosphate utilisation (TPU) limitation, which occurs in some plants (McClain & Sharkey, 2019), is not included here. Bayesian estimation of the FvCB model with TPU limitation is discussed in more detail later.

$$A_n = \min\{W_c,\ W_j\}\left(1-\frac{\Gamma^*}{C_c}\right) - R_d \qquad \text{Eqn 1}$$

$$W_c = \frac{V_{c\max}}{C_c + K_m} \qquad \text{Eqn 2}$$

$$W_j = \frac{J}{4C_c + 8\Gamma^*} \quad \text{Eqn 3}$$

In practice, $C_c$ cannot be measured directly. Instead, a response curve of $A_n$ to intercellular $CO_2$ concentration ($C_i$) is usually recorded. To model the $A_n$-$C_i$ curve, mesophyll conductance ($g_m$) is introduced to the FvCB model.

$$C_c = C_i - A / g_m \quad \text{Eqn 4}$$

Measured *Y(II)* based on chlorophyll fluorescence has a linear relationship with whole chain electron transport rate $J_f$.

$$J_f = I \cdot \alpha \cdot \beta \cdot Y(II) = I \cdot s \cdot Y(II) \quad \text{Eqn 5}$$

where $\alpha$ is light absorption by PSI and PSII, $\beta$ is the fraction of the incident irradiance (*I*) absorbed by PSII, $s$ is $\alpha$ multiplied by $\beta$ representing a combined effect of light absorption and partitioning. Here we are interested in the robustness and reliability of parameter estimation using the fewest possible model parameters, thus $\alpha \cdot \beta$ is estimated as a whole and represented by $s$. We assume that true electron transport rate $J_A$ calculated from $CO_2$ assimilation rate (Eqn 6) equals $J_f$. Therefore other synthetic processes which consume electrons, such as nitrate reduction, the Mehler reaction and malate-oxaloacetate shuttling (von Caemmerer 2000; Yin *et al.* 2009; Selinski & Scheibe 2019), contribute to the observed fluorescence signal and will be combined into an apparent $s$.

$$J_A = \frac{(A_n + R_d)(4C_c + 8\Gamma^*)}{C_c - \Gamma^*} \quad \text{Eqn 6}$$

To synthesise data under different light levels, a non-rectangular hyperbola light

response curve of potential electron transport rate $J$ is adopted (Farquhar & Wong, 1984),

$$J_i = I \cdot s \cdot Y(II)_{LL} \qquad \text{Eqn 7}$$

$$J = \frac{(J_i + J_{\max}) - \sqrt{(J_i + J_{\max})^2 - 4\theta J_i J_{\max}}}{2\theta} \qquad \text{Eqn 8}$$

Where $Y(II)_{LL}$ is the initial slope of $Y(II)$-$I{\cdot}s$ curve, i.e. the maximum quantum efficiency of PSII under low light. $\theta$ is the curvature index of the $J$-$I$ curve.

**Synthesise physiological measurements with sample variation and measurement errors**

Variance of observation ( $A^o_{n(i)}$, $Y(II)^o_{(i)}$ ) in the synthetic data is modeled by sample variation and systematic error (Eqn 9, **Fig. 1**). Sample variation corresponds to the variance explained by error-free measurement of each replicate ( $A^s_{n(i)}$, $Y(II)^s_{(i)}$ ) used in the experiment, while systematic error ($\varepsilon_A$, $\varepsilon_Y$) means the variance of data due to systematic or random factors other than biological variability, such as the accuracy of the Infra-Red Gas Analyser (IRGA), the measurement protocol or human operation.

$$\begin{bmatrix} A^o_{n(i)} \\ Y(II)^o_{(i)} \end{bmatrix} = \begin{bmatrix} A^s_{n(i)} \\ Y(II)^s_{(i)} \end{bmatrix} + \begin{bmatrix} \varepsilon_A \\ \varepsilon_Y \end{bmatrix} \qquad \text{Eqn 9}$$

To generate the synthetic data used in this manuscript (**Fig. 1, Fig. 3 & Fig. S4**), error-free observations $(A^s_{n(i)},\ Y(II)^s_{(i)})$ of the $i^{\text{th}}$ replicate are randomly generated from the FvCB model with slightly varied input parameters (**Fig. S1**).

$$\begin{bmatrix} A^s_{n(i)} \\ Y(II)^s_{(i)} \end{bmatrix} = FvCB\left(C_i, I, K_m, \Gamma^*, g_m, V_{c\max(i)}, J_{\max(i)}, R_{d(i)}, s_{(i)}, Y_{(II)LL(i)}\right) \qquad \text{Eqn 10}$$

For each replicate, we assume that $K_m$, $\Gamma^*$ and $g_m$ are constant across all replicates, and

the remaining parameters $\boldsymbol{X}$ (including $V_{cmax}$, $J_{max}$, $R_d$, $s$ and $Y(II)_{LL}$) are randomly generated from a generic (two-sided truncated) normal distribution.

$$\mathcal{X}_{(i)} \sim TruncNorm(\overline{\mathcal{X}}, \sigma_{\mathcal{X}}) \qquad \text{Eqn 11}$$

The two-sided truncation here is to avoid randomly derived extreme values. Specifically, standard deviation ($\sigma_x$) of the normal distribution of any variable $x$ is equal to 5% of its mean value ($\mu_x$), then the intervals of truncation are all equal to ($\mu_x$-2$\sigma_x$, $\mu_x$+2$\sigma_x$), i.e. outliers smaller than $\mu_x$-2$\sigma_x$ or larger than $\mu_x$+2$\sigma_x$ will be discarded.

Systematic error of $A_n$ and $Y(II)$ measurements are also generated from generic (two-sided truncated) normal distributions, considering that outliers of measurements will be excluded in practice. The standard deviation of each $A_n$ measurement ($\sigma_{\varepsilon\_A}$) is assumed as 0.1 μmol $m^{-2}$ $s^{-1}$ and the standard deviation of each $Y(II)$ measurement ($\sigma_{\varepsilon\_Y}$) is assumed as 0.01. The intervals of truncation for error of both $A_n$ and $Y(II)$ measurements are from -3$\sigma$ to 3$\sigma$.

$$\varepsilon_A \sim TruncNorm(0, \sigma_{\varepsilon\_A}) \qquad \text{Eqn 12}$$

$$\varepsilon_Y \sim TruncNorm(0, \sigma_{\varepsilon\_Y}) \qquad \text{Eqn 13}$$

It is worth mentioning that although different replicates have varied $V_{cmax}$, $J_{max}$, $R_d$, $s$ and $Y(II)_{LL}$, the purpose of parameter estimation is still to estimate one representative value for each parameter. With a limited number of replicates, the mean values of replicates will likely bias from the mean of generic normal distributions. Here in this paper, we are not going to discuss this layer of biases, therefore the results of Bayesian estimation and traditional fitting are all compared with the mean values of

photosynthetic parameters of all replicates.

**Probability model of physiological measurements and Bayesian estimation**

Bayesian theorem (Eqn 14) calculates the joint posterior distribution based on the likelihood of observations and prior information. This equation estimates $A_n$ and $Y(II)$ observations with replicates. Some parameters are prefixed during estimation, with $\boldsymbol{X}$ representing the remaining parameters.

$$p(\mathcal{X} \mid A^o_{n(i)}, Y(II)^o_{(i)}) \propto p(A^o_{n(i)}, Y(II)^o_{(i)} \mid \mathcal{X}) p(\mathcal{X}) \qquad \text{Eqn 14}$$

The Markov Chain Monte Carlo (MCMC) method is used to approximate this joint posterior distribution numerically. Its calculation requires a process model describing the observations, equations of the likelihood of observations, and prior distributions of estimated parameters.

The process model for both the $A_n$-$C_i$ and $Y(II)$-$C_i$ data used in this study is the FvCB model described above. The likelihood of measured $A_n$-$C_i$ and $Y(II)$-$C_i$ data is calculated based on the likelihood of each observation. We assumed that $A_n$ and $Y(II)$ signals can be described by a normal distribution. Thus for the $i^{th}$ observation ($i$ = 1, 2,…, $N$):

$$A^o_{n(i)} \sim \text{Normal}(\overline{A}_{n(i)}, \sigma_{obs\_A}) \qquad \text{Eqn 15}$$

$$Y(II)^o_{(i)} \sim \text{Normal}(\overline{Y(II)}_{(i)}, \sigma_{obs\_Y}) \qquad \text{Eqn 16}$$

where $\overline{A}_{n(i)}$ and $\overline{Y(II)}_{(i)}$ are the error-free mean values of $A_n$ and $Y(II)$ signals, and $\sigma_{obs\_A}$ and $\sigma_{obs\_Y}$ are the standard deviations describing variability of observations. A uniform distribution was set as the prior for each parameter in the FvCB model for the Bayesian estimation. $V_{cmax}$ ranged from 10 to 200 μmol m$^{-2}$ s$^{-1}$, $J$ was from 20 to

400 μmol $m^{-2}$ $s^{-1}$, $R_d$ was from -5 to 5 μmol $m^{-2}$ $s^{-1}$, $K_m$ was from 100 to 1000 μbar, $\Gamma^*$ was from 10 to 50 μbar, $g_m$ was from 0.02 to 50 mol $m^{-2}$ $s^{-1}$ $bar^{-1}$, and $s$ was from 0.2 to 0.8. $r_m$ is the reciprocal of $g_m$ ($r_m = 1/g_m$) and as such is restricted between 0.02 and 50 $mol^{-1}$ $m^2$ s bar. Using such relatively large ranges as prior allows the convergence of Bayesian estimation, whilst ensuring that the estimated joint posterior distribution is not influenced by these prior distributions. The MCMC method is implemented in a software package JAGS (Just Another Gibbs Sampler, Plummer 2003). The rationale of the MCMC method is to construct a series of sampling points in the parameter space, where each sampling point is an array of possible values for the estimated parameters. A Markov chain is constructed where the next sampling point is dependent on the current sampling point and a transition probability. Three parallel MCMC chains are run for 20,000 iterations each, and the Gelman-Rubin convergence diagnostic (Gelman & Rubin 1992) is used to check the convergence of the MCMC algorithm, i.e. all potential scale reduction factor (Rhat) values from JAGS are smaller than 1.1. This is an indicator of the reliability of the estimation technique. The first 10,000 samples are discarded as burn-in samples and the remaining 10,000 samples are used to approximate the joint posterior distribution.

The Bayesian estimation is coded in R. The script is open-source and implemented through a web-based interactive platform, Jupyter Notebook. The details of the fitting method used for comparison with the Bayesian estimation can be found in the Supplementary Methods.

270 **Experimental Methods:**

271 **Plant material and growth conditions**

272 Rice (*Oryza sativa* var. indica; IR64) plants were grown in a controlled growth chamber
273 (Controlled Environments Ltd, Winnipeg, MB, Canada) at 1000 µmol $m^{-2}$ $s^{-1}$
274 photosynthetic photon flux density (*PPFD*), with a 12-h/12-h light/dark cycle, ambient
275 $CO_2$ (410-420ppm), 60% humidity and a day/night temperature of 28/24°C. Seeds were
276 germinated on filter paper with 15 ml water, and seedlings transferred after 7 days to
277 13D pots (0.88 l) filled with 71% Kettering Loam (Boughton, UK), 23.5% Vitax John
278 Innes No. 3 (Leicester, UK), 5% silica sand and 0.5% Osmocote Extract Standard 5–6
279 month slow-release fertilizer (ICL, Ipswich, UK) by volume, saturated with water.

280 **Gas exchange and chlorophyll fluorescence measurements**

281 Gas exchange and chlorophyll fluorescence were measured simultaneously on the fully
282 expanded 6$^{th}$ true leaf of 28-day old plants using a Licor 6800 (LI-COR Inc., Lincoln,
283 NE, USA) and attached Multiphase Flash Fluorometer (6800-01A). For each replicate,
284 the leaf remained in the IRGA chamber for the duration of all high light and low light
285 curves. Relative humidity was maintained at c. 60% with the chamber flow rate set at
286 300 $\mu$mol $s^{-1}$ and leaf temperature set at 28°C. For $A_n$-$C_i$ curves, saturating *PPFD* was
287 held at 2000 $\mu$mol $m^{-2}$ $s^{-1}$ and the following $[CO_2]_{ref}$ were used: 500, 350, 200, 110, 80,
288 60, 30, 500, 700, 900, 1100, 1300, 1500 ppm. Leaves were held at each $[CO_2]$ for a
289 minimum and maximum of 90 and 180 seconds for the first 7 $[CO_2]$ and 180-300
290 seconds for the last 5 $[CO_2]$. For the 8$^{th}$ $[CO_2]$, leaves were held until stable. IRGAs

291 were matched at every [$CO_2$]. Low light curves were performed without unclipping the

292 leaf, under 3 different *PPFD* levels, 300, 200 and 100 $\mu$mol $m^{-2}s^{-1}$. All of the low light

293 curves used the following $[CO_2]_{ref}$: 110, 90, 80, 70, 50, 30 ppm. Leaves were held at

294 each [$CO_2$] for a minimum and maximum of 90 and 180 seconds. The fluorometer was

295 set to measure Fs Fm' Fo', with a light mod rate of 50kHz, flash mod rate of 250kHz,

296 and flash type: Multiphase.

297

298 **Results**

299 **Bayesian estimation to evaluate the uncertainty of parameter estimation**

300 Having created a new photosynthetic metabolism parameter estimation tool using

301 concurrent $A_n$-$C_i$ and $Y(II)$-$C_i$ data (described in Methods and implemented through

302 Jupyter Notebook (https://github.com/xiaoyizz78/FvCB-JAGS)), we proceeded to

303 analyse the uncertainty under different input data and different prefixed parameters

304 during estimation. In particular, we compared the Bayesian estimation method with a

305 traditional fitting method developed using the Levenberg-Marquardt algorithm (see

306 Supplementary Methods for details on the fitting method).

307 We first used synthetic input data to explore the model. Thus, $A_n$-$C_i$ (**Fig. 1 A**) and $Y(II)$-

308 $C_i$ (**Fig. 1 B**) curves under saturating irradiance were created from a set of synthetic

309 data, with inbuilt sample variation and systematic error such as that caused by machine

310 or human operation during measurement. Data was also synthesised at low $CO_2$ and

311 low light levels (**Fig. 1 C, D**). For these curves, the joint posterior distribution

312 calculated from Bayesian estimation was approximated by 3 MCMC chains and 10000
313 sampling points for each chain. Each point thus represents an array of possible
314 estimated parameters, creating an estimate of the "true" $A_n$-$C_i$ and $Y(II)$-$C_i$ curves. We
315 then used the synthetic data to test the effect of fixing different parameters during
316 Bayesian estimation.

317 With $K_m$, $\Gamma^*$ and $s$ prefixed at their true values, Bayesian estimation was applied to the
318 synthetic concurrent $A_n$-$C_i$ and $Y(II)$-$C_i$ data under saturating light (**Fig. 2 A-B**). The
319 marginal posterior distribution of each estimated parameter is shown by the grey
320 regions in **Fig. 2 C-F**. As can be seen, both the best fitted values and the mode of
321 distribution (or the value with the highest probability) of $V_{cmax}$, $J_{1500}$, $R_d$ and $r_m$ are very
322 close to the true values. Grey regions in **Fig. 2 A-B** show $A_n$-$C_i$ and $Y(II)$-$C_i$ curves
323 predicted based on the joint posterior distribution. These predicted $A_n$-$C_i$ and $Y(II)$-$C_i$
324 curves only vary within a very small range compared to the synthetic observations with
325 error (red line with error bars). This seems counterintuitive to the deviation of estimated
326 $V_{cmax}$, $J$, $R_d$ and $r_m$ (**Fig. 2 C-F**). However, it is worth mentioning here that estimated
327 parameters are not necessarily independent, as shown by the bivariate marginal
328 distributions (**Fig. S2**).

329 The standard deviation (std) of the marginal posterior distribution based on 30,000
330 sampling points gives an indication of the accuracy of the estimation for each estimated
331 parameter (**Table 2).** $V_{cmax}$ and $J_{1500}$ reach a high precision when the $K_m$, $\Gamma^*$ and $s$ values
332 are pre-fixed, as shown by the small std values (within 2% of the true value). For $R_d$

and $r_m$, std values are equal to 0.22 μmol $m^{-2}$ $s^{-1}$ and 0.43 $mol^{-1}$ $m^2$ s bar respectively, which is 26% and 8.7% relative to the true values (**Table 2**).

Bayesian estimation was also applied to the same synthetic data with different prefixed parameters, with the stds of the marginal posterior distributions listed in **Table 2**. We found that if we loosened the constraint of $s$ while still keeping the prefixed $K_m$ and $\Gamma^*$ at the true values, std of estimated $s$ was still reasonable (1.2% relative to its true value) with little change in the stds of the other estimated parameters. However, if we further loosened the constraints of $\Gamma^*$ and/or $K_m$, much larger stds were observed (**Table 2).**

It seems that the synthetic $A_n$-$C_i$ and $Y(II)$-$C_i$ data cannot support the identifiability of all photosynthetic parameters in the FvCB model (a model is identifiable if it is theoretically possible to learn the true values of this model's underlying parameters after obtaining an infinite number of observations from it). This identifiability issue of parameter estimation is not evident from the fitted $A_n$-$C_i$ curve, where fitness is usually the focus rather than uncertainty of fitted values. Taking the scenario with no prefixed values as an example, we found that the corresponding $A_n$-$C_i$ curves calculated based on the joint posterior distribution also fit the $A_n$-$C_i$ curve well. However, the best estimated values are strongly biased from the true values (**Table 2**, **Fig. S3**).

**Bayesian estimation, verified with synthetic data with inbuilt error, provides an uncertainty analysis which is unavailable in traditional fitting methods**

It is a good practice to verify new methods of parameter estimation with synthetic data.

354 Therefore, we further tested the Bayesian estimation on more synthetic datasets with
355 inbuilt error, and compared the estimation with fitting using a traditional method. Thus,
356 having characterised the model with the synthetic dataset described in **Fig. 1**, we
357 synthesised a further 100 datasets with varying photosynthetic parameters to mimic the
358 natural variation and systematic error in biological data. In each dataset, concurrent $A_n$-
359 $C_i$ and $Y(II)$-$C_i$ measurements similar to **Fig. 1 A, B** were generated. We then compared
360 the performance of the Bayesian estimation with a fitting method (Dubois *et al.* 2007;
361 **Supplementary Methods**).
362 In the initial comparison, we kept only $K_m$ prefixed at the true values and focused on $R_d$
363 (**Fig. 3 A-D**) and $r_m$ (**Fig. 3 E-H**) as they are the most difficult parameters to fit or
364 estimate (other estimated parameters - **Fig. S4 A-P)**. Significant biases to the true values
365 of $R_d$ and $r_m$ were observed for both the traditional fitting method and Bayesian
366 estimation (**Fig. 3 A-B, E-F**). When the true values for these parameters were very small,
367 the best fitted values for these parameters were particularly inaccurate, with bias from
368 the true value as high as 100% to 160% (**Fig. 3 C-D, G-H**). The Bayesian estimation
369 method also showed a high deviation at low true values (up to 350% and 100% for $R_d$
370 and $r_m$, respectively), and the accuracy of estimation dramatically improved as the true
371 value of $R_d$ and $r_m$ increased (**Fig. 3 C, G**). The z-score of the true values, measured in
372 terms of standard deviations from the mean, was within 1.5 standard deviations of the
373 mean of the posterior distribution across the range of true values (**Fig. 3 D, H**), implying
374 that the value estimated by the Bayesian method is highly related to the true value. The

375 remaining parameters were more accurately fitted and estimated than $R_d$ and $r_m$ (**Fig.**
376 **S4 A-P).**
377 When both *Km* and $\Gamma^*$ were prefixed in the traditional fitting and Bayesian estimation,
378 the best fitted values and the mean of the marginal posterior distribution were much
379 less biased from the true values, both for $R_d$ and $r_m$ (**Fig. 3 I-L, M-P**) and the other
380 parameters estimated (**Fig. S4 Q-AB**). Indeed, fitted values for $V_{cmax}$, $J$ and $s$ were less
381 than 1% biased relative to the true values (**Fig. S4 R, V & Z**). Overall, the stds of
382 posterior distributions from Bayesian estimation with $K_m$ and $\Gamma^*$ prefixed were much
383 smaller than the results with only $K_m$ prefixed (**Fig. 3** and **Fig. S4**), with Z-scores of
384 true values lying mostly within -0.5 and 0.5 (**Fig. 3 L&P**, **Fig. S4 T, X&AB**). Parameter
385 estimation is often very sensitive to biological variation and error in data collection.
386 The uncertainty analysis provided by std values in the Bayesian estimation tool allows
387 quantification of this sensitivity, a capacity that is lacking in traditional fitting methods.
388
389 **Including additional concurrent $A_n$-$C_i$ and Y(II)-$C_i$ measurements at low $CO_2$ and**
390 **light levels improves estimation of photosynthetic parameters**
391 Inspired by Laisk's method for the estimation of $\Gamma^*$ and $R_d$ (Laisk, 1977; Brooks &
392 Farquhar, 1985) and to improve the parameter estimation (especially of $R_d$ and $r_m$), we
393 synthesised additional detailed low $CO_2$ concurrent $A_n$-$C_i$ and $Y(II)$-$C_i$ curves under
394 three different low light levels based on the FvCB model (**Fig. 1 C-D**) for each
395 replication. Bayesian estimation can deal with data from different signals and

conditions to calculate the joint posterior distribution (**Table 3**). With $K_m$ prefixed at its true value, these additional low light measurements improved the identifiability of $\Gamma^*$. Moreover, they decreased the variability of estimated $R_d$ and $r_m$ from 72% and 23% of the true value to much smaller ranges (16% and 14%, respectively) (**Table 2**, **Table 3**). Although these additional low light, low $CO_2$ experimental measurements improved parameter estimation, they are time consuming and rely heavily on the stability of the instrument. It also involves having the leaf clamped in the IRGA chamber for a long time, potentially causing stress to the plant. Therefore, we explored the reliability of Bayesian estimation using fewer low light measurements (**Table 3**). The results showed that parameter estimation using data from low $CO_2$ curves at two low light levels gave an estimation of similar accuracy to that obtained using three low light levels. Specifically, the combination of light at PPFD of 50 & 200 μmol $m^{-2}$ $s^{-1}$ performs better than the combination of 50 & 100 μmol $m^{-2}$ $s^{-1}$ or 100 & 200 μmol $m^{-2}$ $s^{-1}$. With 100 & 200 μmol $m^{-2}$ $s^{-1}$ as low light levels, the std of $R_d$ increases from 16% of its true value (data from three low light curves) to 27%. With only the lowest light level, 50 μmol $m^{-2}$ $s^{-1}$, in combination with the high light $A_n$-$C_i$ and $Y(II)$-$C_i$ data, estimation of $R_d$ is actually more accurate, with a std of 21%, but using only 100 μmol $m^{-2}$ $s^{-1}$or 200 μmol $m^{-2}$ $s^{-1}$ increases the std of the estimated parameters further (**Table 3**).

**Sensitivity analysis of error in the synthetic data with prefixed parameters**

There are two different types of error modelled in the synthetic data, which include

417 sample variation and systematic error (see **Methods** for more details). We conducted a
418 sensitivity analysis of the effect of sample variation in the synthetic concurrent $A_n$-$C_i$
419 and $Y(II)$-$C_i$ data (with $K_m$ prefixed at its true value) on the accuracy of parameter
420 estimation. Synthetic data with different levels of sample variation were generated by
421 scaling up the difference of photosynthetic parameters to the true values in each
422 replicate (Eqn 10 & 11). Bayesian estimation was then conducted on this data (**Fig. 4**).
423 For a given scale factor, new photosynthetic parameters $X_i$ of *i*th replicate were
424 calculated from the original $X_i$ by scaling the deviation (Eqn. 10 & 11). The marginal
425 posterior distribution is plotted as a column of pixels in **Fig. 4**, with different colours
426 representing the probability density. The lower the scale factor of sample variation, the
427 closer the approximations of different replicates are to the true values. As can be seen
428 in **Fig. 4**, stds of the marginal posterior distribution increase linearly with increasing
429 scale factor of sample variability, however, the mean value of the marginal posterior
430 distribution approximated the true value at all levels of variability. Using data without
431 sample variation (represented by a single replicate; Rep No.1 of the synthetic
432 concurrent $A_n$-$C_i$ and $Y(II)$-$C_i$ data from **Fig. 1**) as input observations to Bayesian
433 estimation dramatically decreased the stds of all estimated parameters (**Table 3**).
434 As we have shown, parameter identifiability by Bayesian estimation is only possible
435 with a prefixed $K_m$ value. We therefore conducted a sensitivity analysis on the effect of
436 decreasing and increasing the prefixed $K_m$ value by 50% from its true value (**Fig. S5**).
437 The estimated marginal posterior distribution of $V_{cmax}$ was very sensitive to the prefixed

value of $K_m$. The other parameters were less sensitive, with good parameter identifiability, especially with $K_m$ values prefixed higher than the true value (**Fig. S5**). We also tested the effect of systematic error on the accuracy of parameter estimation. Here only synthetic data of replicate No.1 were used and its systematic error was manually controlled ($\varepsilon_A$ and $\varepsilon_Y$ in Eqn 9). For a given scale factor, new observations of replicate No.1 were calculated from default observation by scaling the deviation $\varepsilon_A$ and $\varepsilon_Y$ in Eqn 9. Results of the Bayesian estimation showed that stds of the marginal posterior distributions increased linearly with increased systematic error (**Fig. S6**).

**Parameter estimation with experimental data from rice leaves under several light levels**

To further characterise the utility of the Bayesian tool for photosynthetic parameter estimation, experimental datasets comprising concurrent $A_n$-$C_i$ and $Y(II)$-$C_i$ measurements under several light levels were taken from rice leaves from four independent plants (Experimental Dataset available on https://github.com/xiaoyizz78/FvCB-JAGS; replicates are numbered from 1 to 4). Bayesian estimation was conducted with data for each of the four replicates separately and with all data combined. The marginal posterior distributions of each estimated parameter from these five scenarios showed different values and probabilities of each peak, suggesting that these different replicates result in different levels of estimation uncertainty (**Fig. 5**). The corresponding means and stds are shown in **Table 4**.

Estimation using data from sample No. 3 consistently had the smallest stds of all of the individual datasets, while parameter values estimated from sample No. 4 were generally closer to that estimated from analysis of the combined datasets. Bayesian estimation with the combined datasets showed a higher accuracy than most individual estimations, with the exception of sample No. 3. **Fig. 6** shows the Bayesian estimation for sample No. 3 calculated from the joint posterior distributions (see **Fig. S7-9** for the remaining three replicates). The curves under saturating light (**Fig. 6 A,B**) showed less bias than the curves at the three low light levels (**Fig. 6 C,D**), possibly reflecting the use of normal distributions in the error term of the probability model during Bayesian estimation (Eqn 15&16). Interestingly, the $Y(II)$-$C_i$ data showed a slight discrepancy between measured values and estimation (panel D in **Fig. 6, Fig. S7-S9**), with measured values going up with increasing $C_i$ but estimated values remaining level.

## Discussion

Most methods developed for parameter estimation of the Farquhar-von Caemmerer-Berry model lack evaluation of robustness and reliability of the estimated parameters. In this paper, we report a Bayesian parameter estimation framework which not only estimates the photosynthetic parameters of the FvCB model, but also gives the standard deviation of the parameters, which represents their robustness and reliability. Using synthetic concurrent $A_n$-$C_i$ and $Y(II)$-$C_i$ measurements at high light, we show that multiple prefixed parameters are needed to enable accurate estimation of the parameter

480 mean and to estimate the standard deviation. We dramatically improved the accuracy
481 of prediction of the mean and decreased the std of the estimated parameters by including
482 additional detailed low $CO_2$ concurrent $A_n$-$C_i$ and $Y(II)$-$C_i$ measurements under low
483 light, keeping only $K_m$ prefixed.
484 Sensitivity analysis showed that sample variation and systematic error (from human
485 mistakes or technical issues) are the major limits to the accuracy of parameter
486 estimation. We have shown that using the correct prefixed value for $K_m$ is crucial to
487 robust estimations. We have tested these ideas and the Bayesian framework *in planta*,
488 on rice leaves, with the corresponding pipeline of Bayesian estimation provided as a
489 user-friendly interactive script in Jupyter Notebook (see supplementary code or
490 https://github.com/xiaoyizz78/FvCB-JAGS).
491
492 **Bayesian estimation enables objective evaluation of robustness and reliability of**
493 **the estimated parameters**
494 The performance of various methods of photosynthetic parameter estimation is usually
495 evaluated by $r^2$ of the fitted values or the sum of squared error between observation and
496 prediction. However, an $r^2$ close to 1 or a predicted curve very close to the observation
497 does not necessarily imply that the estimated values are accurate. This is evident in our
498 Bayesian estimation with the synthetic $A_n$-$C_i$ and $Y(II)$-$C_i$ data under saturating light in
499 **Fig. 1 A & B**. Even with no parameters prefixed, the predicted $A_n$-$C_i$ curve appears to
500 be very close to the true curve (**Fig. S3 A**). However, the parameter estimations are

clearly biased from the true values, as shown by the posterior distributions and difference between the true value and the mode of the posterior distribution (**Table 2**, **Fig. S3 C-I**). Extending this parameter estimation to 100 synthetic datasets, comparison of a traditional fitting method and our Bayesian estimation confirmed that while neither method can accurately estimate the parameter values without sufficient prefixed parameters (**Fig. 3 A-H, Fig. S4 A-P**), the Bayesian framework is able to accurately estimate the reliability of parameters through the posterior distributions.

**Factors influencing identifiability and accuracy of parameter estimation of FvCB model**

Accuracy of parameter estimation is generally affected by how well the model reflects the observations, the calibre of the data, and the quality of the estimation assuming the model to be perfect. We analysed factors influencing the accuracy of current photosynthetic parameter estimation with the FvCB model using synthetic datasets. First, for given synthesised measurements, reasonable prefixed parameters are required to estimate the remaining parameters in the FvCB model (**Table 2 & 3**). We have shown that identifiability of photosynthetic parameters from saturating light $A_n$-$C_i$ and $Y(II)$-$C_i$ data is only possible if $K_m$ and $\Gamma^*$ are prefixed at their true value (**Table 2**). With concurrent $A_n$-$C_i$ and $Y(II)$-$C_i$ data under saturating light, and additional focused low $CO_2$ data from multiple low light levels, a comparable accuracy of parameter estimation is achieved, with only $K_m$ needing to be prefixed (**Table 3**).

522 Secondly, we have shown that bias in these prefixed parameters also affects the
523 parameter estimation. To obtain an accurate parameter estimation using all of the
524 synthetic measurements in **Fig. 1**, $K_m$ needs to be prefixed. Without a known $K_m$,
525 sensitivity analysis of prefixed $K_m$ values demonstrated that $V_{cmax}$ estimation is very
526 sensitive to a biased prefixed $K_m$, while $J$, $R_d$ and $\Gamma^*$ are much less sensitive (**Fig. S6**).
527 Technically, *in vivo* determinations of $K_m$ should be done with transgenic plants with
528 decreased amounts of Rubisco (von Caemmerer et al., 1994). A small number of
529 measurements from one location on a normal leaf, where only a fraction of the $A_n$-$C_i$
530 curve is Rubisco-limited, is not sufficient to support the estimation of $K_m$.
531 Thirdly, we show that the accuracy of parameter estimation in the FvCB model was
532 greatly affected by sample variation and systematic error in the data. Sample variation
533 is especially poorly acknowledged in many previous studies. In practice, errors due to
534 sample variation are inevitable, given the heterogeneity existing among biological
535 replicates and even at different positions along the same leaf (Chen C.P., Zhu & Long
536 2008; Xiong *et al.* 2015). Among all the photosynthetic parameters, $R_d$ and $r_m$ are most
537 sensitive to these errors in the data. If Bayesian estimation is applied to one synthetic
538 replicate (data with no sample variation) instead of all replicates, the std of the estimated
539 $R_d$ decreased from 16% to 6.3% relative to the true value, while the std of the estimated
540 $r_m$ decreased from 14% to 4.5% relative to the true value (**Table 3**). These remaining
541 stds (6.3% and 4.5%) are still large considering the fact that the systematic error of the
542 $A_n$ signal and the *Y(II)* signal was very small in the synthetic data ($\sigma_{\varepsilon\_A}$ = 0.1 μmol m$^{-2}$

543 $s^{-1}$ and $\sigma_{\varepsilon\_Y}$ = 0.01 in Eqn. 15 & 16). Previously, Gu *et al.* (2010) showed that their
544 optimum fitting method predicts unbiased photosynthetic parameters using completely
545 error-free synthetic $A_n$-$C_i$ data. Our sensitivity analysis of the sample variation (**Fig. 4**)
546 and systematic error (**Fig. S6**) leads to a similar conclusion, i.e., that with a decrease of
547 both types of error, marginal posterior distributions converge to the true values. The
548 high sensitivity of estimation of $R_d$ and $r_m$ to these errors is a property of the structure
549 of the FvCB model.
550

551 **Evaluating the quality of experimental data and the experimental protocols**

552 Using the std values as an indicator of estimation accuracy, it is possible to use the
553 Bayesian tool to instantly analyse data quality. With the above analysis, concurrent
554 measurements of $A_n$-$C_i$ and $Y(II)$-$C_i$ under high light and three low light levels were
555 conducted on rice leaves (**Fig. 5**) with four replicates. It is possible to compare the stds
556 of parameters from each replication to assess whether the data is robust. The range of
557 stds in the above data, where Bayesian estimation with the combined dataset led to
558 smaller predicted stds for all estimated parameters compared with estimation using data
559 from individual samples No.1, 2 or 4 (**Fig. 5, Table 4**), but using only data from sample
560 No.3 led to smaller stds for all estimated parameters than with combined data,
561 demonstrates this potential (**Table 4, Supplementary code**).
562 Since this Bayesian approach can be used to estimate parameters from a single sample
563 and provide an estimate of the confidence interval, our Bayesian tool also allows for

rapid assessment of protocol quality. There are many options in Infra-Red Gas Analysis that can be varied during data collection for $A_n$-$C_i$ and $Y(II)$-$C_i$ curves, such as the number and spacing of $[CO_2]_{ref}$ values in the curves, maximum and minimum wait times for stability at each $[CO_2]_{ref}$ point, and low light levels used. This capacity of the tool can be used in the initial planning stages of an experiment, which could potentially eliminate many time-consuming and unnecessary measurements.

**Using Bayesian statistics to quantify more accurate respiration and mesophyll conductance *in planta***

The Bayesian estimation with data from the "best" biological sample in our analysis had a standard deviation of the marginal posterior distribution of estimated $r_m$ equal to 0.46 mol $m^{-2}$ $s^{-1}$ $bar^{-1}$, which is about 44% of the mean value (**Table 4**). This uncertainty is much larger than that predicted using the synthetic data (**Table 3**), where estimation of $r_m$ from one replicate has a std of 0.23 mol $m^{-2}$ $s^{-1}$ $bar^{-1}$ (4.5% of the mean). As we discussed above, systematic error is part of the reason for this difference. Meanwhile, biological systems tend to be inherently complex and noisy, which cannot be fully represented by the simplified model used to synthesise this data. For example, $r_m$ potentially varies in a real leaf under different light or $CO_2$ levels, variation of which is embedded in the measurements and estimation (Flexas *et al.* 2007; Tholen & Zhu 2011; Tholen, Ethier, Genty, Pepin & Zhu 2012; Tholen, Éthier & Genty 2014; Evans & von Caemmerer 2013; Xiao & Zhu 2017).

585 From the $A_n$-$C_i$ and $Y(II)$-$C_i$ curves calculated based on posterior distribution, the FvCB
586 model with a constant $r_m$ generally fitted well with the experimental data (**Fig. 6, Fig.**
587 **S7-S9**). However, an interesting observation from using the experimental data for
588 Bayesian estimation is an apparent mismatch between predicted and measured $Y(II)$-$C_i$
589 curves under low light (panel D in **Fig. 6, Fig. S7-S9**). Specifically, the measured $Y(II)$-
590 $C_i$ curves under low light showed a slight trend to increase with $C_i$, while the FvCB
591 model predicted a constant $Y(II)$ under different $C_i$. As we use the variable $J$ method for
592 calculating $r_m$, this discrepancy is most likely attributed to a varying $r_m$ (Flexas *et al.*
593 2007; Tholen & Zhu 2011; Tholen, Ethier, Genty, Pepin & Zhu 2012; Tholen, Éthier &
594 Genty 2014; Evans & von Caemmerer 2013; Xiao & Zhu 2017). However, there are
595 several additional factors or processes related to this mismatch: 1) a varying $V_{cmax}$ due
596 to activation of Rubisco under low light (von Caemmerer & Edmondson 1986); 2) a
597 larger $R_d$ under low light due to the Kok effect (Kok 1948, 1956; Hoch, Owens & Kok
598 1963); 3) a varying $s$ due to change of cyclic or alternative electron transport in the
599 whole electron transport chain (Yin *et al.* 2004, 2009b); 4) a varying expression of $4C_c$
600 $+ 8\Gamma^*$ in Eqn. 3 due to RuBP regeneration being limited by insufficient NADPH or by
601 insufficient ATP (von Caemmerer 2000); 5) a varying $\Gamma^*$ due to competition of electron
602 flow from nitrogen fixation with carboxylation and photorespiration (Busch, Sage &
603 Farquhar 2018).
604 Together, these six factors are pertinent when dissecting factors underlying the slight
605 increasing trend of low light $Y(II)$-$C_i$ curves (**Fig. 6 D, Fig. S7-S9**). Practically, from

606 the perspective of Bayesian estimation, an accurate quantification of the varying $r_m$
607 would require extending the current FvCB model or developing new models to include
608 these factors, both of which have been attempted a number of times (e.g. Yin *et al.* 2004,
609 2009b; Tholen *et al.* 2012; Busch *et al.* 2018). At the same time, it would also require
610 that variables representing these factors or processes are identifiable from experimental
611 data, which could be evaluated by using the same Bayesian statistical framework.
612

613 **A generic framework of uncertainty evaluation for estimating photosynthetic**
614 **parameters with various physiological measurements and models**

615 In this paper, the identifiability issue and accuracy of parameter estimation shown in
616 the analysis is limited by the FvCB model used (Eqns 1-5) and corresponding
617 physiological measurements taken. However, many variants of the FvCB model and
618 various related experimental protocols have been developed during the past decades.
619 For example, measurements under low oxygen achieve a non-photorespiratory
620 condition under which $\Gamma^*$ and $R_d$ are better estimated compared to Laisk's method
621 (Laisk *et al.* 2002, 2006; Yin *et al.* 2009b, 2011; Bellasio *et al.* 2015). In addition to
622 Rubisco and RuBP regeneration, photosynthesis can be limited by triose phosphate
623 utilisation (TPU) in many species (McClain and Sharkey, 2019). Our measured $A_n$-$C_i$
624 and $Y(II)$-$C_i$ curves in rice also seems to show a decreasing trend especially under high
625 $C_i$ (**Panel A,B** in **Fig. 6**, **Fig. S7-S9**) which implies potential TPU limitation, this is not
626 reflected in the estimated data. An alternative process model could be used to better

627 incorporate TPU limitation, perhaps improving the identifiability of the model
628 parameters. The number of data points falling within the range of $C_i$ whereby this
629 limitation is relevant would also need to be increased. The Bayesian statistics shown in
630 our study would still apply to scenarios such as this with different physiological
631 measurements and different variants of the FvCB model.
632 When comparing the performance of multiple identifiable models, in Bayesian statistics,
633 information criterion is calculated for each model based on the posterior distribution
634 approximated by MCMC, quantifying the likelihood of a model with penalty to its
635 complexity, i.e. number of parameters. There are several information criteria proposed
636 (Gelman, Hwang & Vehtari, 2014), such as AIC (Akaike information criterion), DIC
637 (deviance information criterion) and WAIC (Watanabe-Akaike information criterion),
638 among which DIC is computed by default in the software JAGS used here (Plummer,
639 2003). Alternative code, for other scenarios with different experimental measurements
640 or models is available in the Jupyter Notebook.
641
642 **Using the new Bayesian parameter estimation tool**
643 The interactive Bayesian tool includes full instructions on how to download the
644 prerequisites and run the program: https://github.com/xiaoyizz78/FvCB-JAGS. Three
645 optional $K_m$ values for rice, tobacco and Arabidopsis are provided. Consideration should
646 be given to the low light levels used. Ideally all low light curves should be RuBP
647 regeneration limited, and each should be separate from the others. Data is inputted in

one csv file of all replicates for each treatment. Bayesian estimation is performed on each individual replicate (for quality control) and on all replicates (combined data) for parameter estimation. The output is a series of graphs of the raw data, the Bayesian posterior distributions for each estimated parameter, a table of estimated parameters (including the stds of posterior distributions for each parameter), traceplots to ensure the estimation is stable, and posterior distribution $C_i$ response curves. Compared to many existing parameter fitting tools, this is extremely simple to use and has a user-friendly output. The tool is able to handle a large amount of data extremely quickly and removes much of the human interaction which can potentially affect parameter estimation.

## Conclusion

Bayesian estimation not only predicts the most likely parameter values, but also provides the standard deviation of marginal posterior distributions, a measure of estimation accuracy. Systematic analyses with synthetic data have highlighted important factors influencing this. The Bayesian system enables evaluation of the quality of experimental data and the reliability of experimental protocols. The addition of concurrent measurements of $A_n$-$C_i$ and $Y(II)$-$C_i$ curves within a focused low $CO_2$ range, at two or more low light levels produces much more reliable parameter estimation than saturating light concurrent $A_n$-$C_i$ and $Y(II)$-$C_i$ curves alone, which should be considered in future experimental protocol development. Ultimately, accurate

estimation of photosynthetic parameters is limited by physiological parameter variability within the samples and measurement error introduced by human or machine. We have highlighted the importance of striving to minimise these sources of error. Finally, Bayesian estimation can capture the mismatch between theoretical models and experimental data, which can help to direct systems level studies towards more accurate quantification of photosynthetic related processes.

**Acknowledgements**

This work was supported in part by Chinese Strategic Leading project category B (XDB27020105), National Natural Science Foundation of China (31700201, 31900304) to XGZ and YX; BBSRC (BB/N013719/1) Newton Fund Rice Research and Royal Society (CH\R1\180027) Challenge-led grant to AJF. XYC is supported by the U.S. Department of Energy (DOE), Office of Biological and Environmental Research (BER), as part of BER's Subsurface Biogeochemical Research Program (SBR), through the SBR Scientific Focus Area (SFA) at the Pacific Northwest National Laboratory (PNNL). PNNL is operated for the DOE by Battelle Memorial Institute under contract DE-AC05-76RL01830. The authors also thank Dr Matthew Wilson for comments on the manuscript and declare no competing interests.

**References**

Andrieu C., de Freitas N., Doucet A. & Jordan M.I. (2003) An introduction to MCMC

for machine learning. *Machine Learning* **50**, 5–43.

Bellasio C., Beerling D.J. & Griffiths H. (2015) An Excel tool for deriving key photosynthetic parameters from combined gas exchange and chlorophyll fluorescence: Theory and practice. *Plant, Cell & Environment*.

Bernacchi C.J., Morgan P.B., Ort D.R. & Long S.P. (2005) The growth of soybean under free air $CO_2$ enrichment (FACE) stimulates photosynthesis while decreasing in vivo Rubisco capacity. *Planta* **220**.

Bernacchi C.J., Portis A.R., Nakano H., von Caemmerer S. & Long S.P. (2002) Temperature response of mesophyll conductance. Implications for the determination of Rubisco enzyme kinetics and for limitations to photosynthesis in vivo. *Plant Physiology* **130**, 1992–1998.

Bongi G. & Loreto F. (1989) Gas-exchange properties of salt-stressed Olive (*Olea europea L.*) leaves. *Plant Physiology* **90**, 1408–1416.

Brooks A. & Farquhar G.D. (1985) Effect of temperature on the $CO_2/O_2$ specificity of ribulose-1,5-bisphosphate carboxylase/oxygenase and the rate of respiration in the light: Estimates from gas-exchange measurements on spinach. *Planta* **165**, 397–406.

Busch F.A., Sage R.F. & Farquhar G.D. (2018) Plants increase $CO_2$ uptake by assimilating nitrogen via the photorespiratory pathway. *Nature Plants* **4**, 46–54.

von Caemmerer S. (2000) *Biochemical models of leaf photosynthesis*. CSIRO Publishing, Collingwood, VIC, Australia.

von Caemmerer S. (2013) Steady-state models of photosynthesis. *Plant, Cell & Environment* **36**, 1617–1630.

von Caemmerer S. & Edmondson D.L. (1986) Relationship between steady-state gas exchange, in vivo ribulose bisphosphate carboxylase activity and some carbon reduction cycle intermediates in *Raphanus sativus*. *Functional Plant Biology* **13**, 669–688.

Caemmerer S., Evans J.R., Hudson G.S. & Andrews T.J.J. (1994) The kinetics of ribulose-1,5-bisphosphate carboxylase/oxygenase in vivo inferred from measurements of photosynthesis in leaves of transgenic tobacco. *Planta* **195**, 88–97.

Chen C.P., Zhu X.-G. & Long S.P. (2008) The effect of leaf-level spatial variability in photosynthetic capacity on biochemical parameter estimates using the Farquhar model: a theoretical analysis. *Plant Physiology* **148**, 1139–47.

Chen X., Rubin Y., Ma S. & Baldocchi D. (2008) Observations and stochastic modeling of soil moisture control on evapotranspiration in a Californian oak savanna. *Water Resources Research* **44**.

Clark J.S. & Gelfand A.E. (2006) A future for models and data in environmental science. *Trends in Ecology & Evolution* **21**, 375–380.

Di Marco G., Manes F., Tricoli D. & Vitale E. (1990) Fluorescence parameters measured concurrently with net photosynthesis to investigate chloroplastic $CO_2$ concentration in leaves of *Quercus ilex* L. *Journal of Plant Physiology* **136**, 538–

543.

Dubois J.-J.B., Fiscus E.L., Booker F.L., Flowers M.D. & Reid C.D. (2007) Optimizing the statistical estimation of the parameters of the Farquhar–von Caemmerer–Berry model of photosynthesis. *New Phytologist* **176**, 402–414.

Ethier G.J. & Livingston N.J. (2004) On the need to incorporate sensitivity to $CO_2$ transfer conductance into the Farquhar-von Caemmerer-Berry leaf photosynthesis model. *Plant, Cell & Environment* **27**, 137–153.

Ethier G.J., Livingston N.J., Harrison D.L., Black T.A. & Moran J.A. (2006) Low stomatal and internal conductance to $CO_2$ versus Rubisco deactivation as determinants of the photosynthetic decline of ageing evergreen leaves. *Plant, Cell & Environment* **29**, 2168–2184.

Evans J.R. & von Caemmerer S. (2013) Temperature response of carbon isotope discrimination and mesophyll conductance in tobacco. *Plant, Cell & Environment* **36**, 745–756.

Farquhar G.D., von Caemmerer S. & Berry J. (1980) A biochemical model of photosynthetic $CO_2$ assimilation in leaves of $C_3$ species. *Planta* **149**, 78–90.

Farquhar G.D., von Caemmerer S. & Berry J. (2001) Models of photosynthesis. *Plant Physiology* **125**, 42–45.

Farquhar G.D. & Wong S.C. (1984) An empirical model of stomatal conductance. *Australian Journal of Plant Physiology* **11**, 191.

Feng X. & Dietze M. (2013) Scale dependence in the effects of leaf ecophysiological

traits on photosynthesis: Bayesian parameterization of photosynthesis models. *New Phytologist* **200**, 1132–1144.

Flexas J., Diaz-Espejo A., Galmés J., Kaldenhoff R., Medrano H. & Ribas-Carbo M. (2007) Rapid variations of mesophyll conductance in response to changes in $CO_2$ concentration around leaves. *Plant, Cell & Environment* **30**, 1284–1298.

Gelfand A.E. & Smith A.F.M. (1990) Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**, 398–409.

Gelman A., Hwang J. & Vehtari A. (2014) Understanding predictive information criteria for Bayesian models. *Statistics and Computing* **24**, 997–1016.

Gelman A. & Rubin D.B. (1992) Inference from iterative simulation using multiple sequences. *Statistical Science* **7,** 457–472.

Golaz J.-C., Caldwell P.M., Van Roekel L.P., Petersen M.R., Tang Q., Wolfe J.D., … Zhu Q. (2019) The DOE E3SM coupled model version 1: Overview and evaluation at standard resolution. *Journal of Advances in Modeling Earth Systems* **11**, 2089–2129.

Gu L., Pallardy S.G., Tu K., Law B.E. & Wullschleger S.D. (2010) Reliable estimation of biochemical parameters from $C_3$ leaf photosynthesis-intercellular carbon dioxide response curves. *Plant, Cell & Environment* **33**, 1852–1874.

Gu L. & Sun Y. (2014) Artefactual responses of mesophyll conductance to $CO_2$ and irradiance estimated with the variable J and online isotope discrimination methods. *Plant, Cell & Environment* **37**, 1231–1249.

Han T., Zhu G., Ma J., Wang S., Zhang K., Liu X., … Huang C. (2020) Sensitivity analysis and estimation using a hierarchical Bayesian method for the parameters of the FvCB biochemical photosynthetic model. *Photosynthesis Research* **143**, 45–66.

Harley P.C., Loreto F., Di Marco G. & Sharkey T.D. (1992) Theoretical considerations when estimating the mesophyll conductance to $CO_2$ flux by analysis of the response of photosynthesis to $CO_2$. *Plant Physiology* **98**, 1429–36.

Hoch G., Owens O. v. H. & Kok B. (1963) Photosynthesis and respiration. *Archives of biochemistry and biophysics* **101**, 171–180.

Humphries S.W. & Long S.P. (1995) WIMOVAC: a software package for modelling the dynamics of plant leaf and canopy photosynthesis. *Computer applications in the biosciences : CABIOS* **11**, 361—371.

Kok B. (1948) A Critical consideration of the quantum yield of *Chlorella*-photosynthesis. *Enzymologia* **13**, 1–56.

Kok B. (1956) On the inhibition of photosynthesis by intense light. *Biochimica et Biophysica Acta* **21**, 234–244.

Laisk A.K. (1977) *Kinetics of photosynthesis and photorespiration of $C_3$ in plants*. Moscow, Russia; Nauka

Laisk A., Eichelmann H., Oja V., Rasulov B. & Rämma H. (2006) Photosystem II cycle and alternative electron flow in leaves. *Plant and Cell Physiology* **47**, 972–983.

Laisk A. & Loreto F. (1996) Determining photosynthetic parameters from leaf $CO_2$

exchange and chlorophyll fluorescence. *Plant Physiology* **110**, 903–912.

Laisk A., Oja V., Rasulov B., Rämma H., Eichelmann H., Kasparova I., … Vapaavuori E. (2002) A computer-operated routine of gas exchange and optical measurements to diagnose photosynthetic apparatus in leaves. *Plant, Cell & Environment* **25**, 923–943.

Lawrence D.M., Fisher R.A., Koven C.D., Oleson K.W., Swenson S.C., Bonan G., … Zeng X. (2019) The Community Land Model version 5: description of new features, benchmarking, and impact of forcing uncertainty. *Journal of Advances in Modeling Earth Systems* **11**, 4245–4287.

Manter D.K. & Kerrigan J. (2004) $A/C_i$ curve analysis across a range of woody plant species: influence of regression analysis parameters and mesophyll conductance. *Journal of Experimental Botany* **55**, 2581–2588.

McClain A.M. & Sharkey T.D. (2019) Triose phosphate utilization and beyond: From photosynthesis to end product synthesis. *Journal of Experimental Botany* **70**, 1755–1766.

Miao Z., Xu M., Lathrop R.G. & Wang Y. (2009) Comparison of the $A$-$C_c$ curve fitting methods in determining maximum ribulose 1.5-bisphosphate carboxylase/oxygenase carboxylation rate, potential light saturated electron transport rate and leaf dark respiration. *Plant, Cell & Environment* **32**, 109–22.

Moualeu-Ngangue D.P., Chen T.-W. & Stützel H. (2017) A new method to estimate photosynthetic parameters through net assimilation rate−intercellular space $CO_2$

816 concentration ($A-C_i$) curve and chlorophyll fluorescence measurements. *New*
817 *Phytologist* **213**, 1543–1554.
818 Patrick L.D., Ogle K. & Tissue D.T. (2009) A hierarchical Bayesian approach for
819 estimation of photosynthetic parameters of $C_3$ plants. *Plant, Cell & Environment*
820 **32**, 1695–1709.
821 Perveen S., Qu M., Chen F., Essemine J., Khan N., Lv M.A., Chang T., Song Q., Chen
822 G-Y, and Zhu X-G (2020) Over-expression of maize transcription factor mEmBP-
823 1 increases photosynthesis, biomass and yield in rice. *Journal of Experimental*
824 *Botany* **71**, 4944-4957.
825 Plummer M. (2003) JAGS : A program for analysis of Bayesian graphical models.
826 *Proceedings of the 3rd International Workshop on Distributed Statistical*
827 *Computing (DSC 2003)*.
828 Selinski J. & Scheibe R. (2019) Malate valves: old shuttles with new perspectives. *Plant*
829 *Biology* **21**, 21–30.
830 Simkin A.J., McAusland L., Headland L.R., Lawson T. & Raines C.A. (2015)
831 Multigene manipulation of photosynthetic carbon assimilation increases $CO_2$
832 fixation and biomass yield in tobacco. *Journal of Experimental Botany* **66**, 4075–
833 4090.
834 Sharkey T.D. (2016) What gas exchange data can tell us about photosynthesis. *Plant,*
835 *Cell & Environment* **39**, 1161–1163.
836 Sharkey T.D., Bernacchi C.J., Farquhar G.D. & Singsaas E.L. (2007) Fitting

photosynthetic carbon dioxide response curves for $C_3$ leaves. *Plant, Cell & Environment* **30**, 1035–40.

Smith A.F.M. & Roberts G.O. (1993) Bayesian computation via the Gibbs sampler and related Markov Chain Monte Carlo methods. *Journal of the Royal Statistical Society. Series B (Methodological)* **55**, 3–23.

Sun Y., Gu L., Dickinson R.E., Pallardy S.G., Baker J., Cao Y., … Winter K. (2014) Asymmetrical effects of mesophyll conductance on fundamental photosynthetic parameters and their relationships estimated from leaf gas exchange measurements. *Plant, Cell & Environment* **37**, 978–994.

Tholen D., Éthier G. & Genty B. (2014) Mesophyll conductance with a twist. *Plant, Cell & Environment* **37**, 2456–2458.

Tholen D., Ethier G., Genty B., Pepin S. & Zhu X.-G. (2012) Variable mesophyll conductance revisited: theoretical background and experimental implications. *Plant, Cell & Environment* **35**, 2087–103.

Tholen D. & Zhu X.-G. (2011) The mechanistic basis of internal conductance: A theoretical analysis of mesophyll cell photosynthesis and $CO_2$ diffusion. *Plant Physiology* **156**, 90–105.

Tierney L. & Mira A. (1999) Some adaptive Monte Carlo methods for Bayesian inference. *Statistics in Medicine* **18**, 2507–2515.

Valentini R., Epron D., De Angelis P., Matteucci G. & Dreyer E. (1995) In situ estimation of net $CO_2$ assimilation, photosynthetic electron flow and

858 photorespiration in Turkey oak (*Q. cerris* L.) leaves: diurnal cycles under different
859 levels of water supply. *Plant, Cell & Environment* **18**, 631–640.
860 Wang D., Jaiswal D., Lebauer D.S., Wertin T.M., Bollero G.A., Leakey A.D.B. & Long
861 S.P. (2015) A physiological and biophysical model of coppice willow (*Salix* spp.)
862 production yields for the contiguous USA in current and future climate scenarios.
863 *Plant, Cell & Environment* **38**, 1850–1865.
864 Xiao Y. & Zhu X.-G. (2017) Components of mesophyll resistance and their
865 environmental responses: A theoretical modelling analysis. *Plant, Cell &*
866 *Environment* **40**, 2729–2742.
867 Xiong D., Yu T., Liu X., Li Y., Peng S. & Huang J. (2015) Heterogeneity of
868 photosynthesis within leaves is associated with alteration of leaf structural features
869 and leaf N content per leaf area in rice. *Functional Plant Biology* **42**, 687–696.
870 Yin X., Van Oijen M. & Schapendonk a. H.C.M. (2004) Extension of a biochemical
871 model for the generalized stoichiometry of electron transport limited $C_3$
872 photosynthesis. *Plant, Cell & Environment* **27**, 1211–1222.
873 Yin X. & Struik P.C. (2009a) Theoretical reconsiderations when estimating the
874 mesophyll conductance to $CO_2$ diffusion in leaves of $C_3$ plants by analysis of
875 combined gas exchange and chlorophyll fluorescence measurements. *Plant, Cell*
876 *& Environment* **32**, 1513–1524.
877 Yin X., Struik P.C., Romero P., Harbinson J., Evers J.B., Van Der Putten P.E.L. & Vos
878 J. (2009b) Using combined measurements of gas exchange and chlorophyll

fluorescence to estimate parameters of a biochemical $C_3$ photosynthesis model: A critical appraisal and a new integrated approach applied to leaves in a wheat (*Triticum aestivum*) canopy. *Plant, Cell & Environment* **32**, 448–464.

Yin X., Sun Z., Struik P.C. & Gu J. (2011) Evaluating a new method to estimate the rate of leaf respiration in the light by analysis of combined gas exchange and chlorophyll fluorescence measurements. *Journal of Experimental Botany* **62**, 3489–3499.

## Tables

**Table 1.** List of model variables and their units

| Variables | Definition | Units |
|---|---|---|
| $A_n$ | Net photosynthesis rate | μmol $m^{-2}$ $s^{-1}$ |
| $A_c$ | Rubisco limited net photosynthesis rate | μmol $m^{-2}$ $s^{-1}$ |
| $A_j$ | RuBP regeneration limited net photosynthesis rate | μmol $m^{-2}$ $s^{-1}$ |
| $C_i$ | Intercellular $CO_2$ partial pressure | μbar |
| $C_c$ | Chloroplastic $CO_2$ partial pressure | μbar |
| $g_m$ | Mesophyll conductance | mol $m^{-2}$ $s^{-1}$ $bar^{-1}$ |
| $I$ | Incident irradiance | μmol $m^{-2}$ $s^{-1}$ |
| $J$ | Potential linear electron transport rate through PSII | μmol $m^{-2}$ $s^{-1}$ |
| $J_i$ | Maximum electron transport rate through PSII limited by $I$ | μmol $m^{-2}$ $s^{-1}$ |
| $J_A$ | Electron transport rate calculated from $CO_2$ assimilation rate | μmol $m^{-2}$ $s^{-1}$ |
| $J_f$ | Electron transport rate calculated from chlorophyll fluorescence measurement | μmol $m^{-2}$ $s^{-1}$ |
| $J_m$ | Maximum electron transport rate through PSII limited by capacity of photosystem | μmol $m^{-2}$ $s^{-1}$ |
| $K_m$ | Michaelis-Menten constant of Rubisco for $CO_2$ in the presence of $O_2$ | μbar |
| $R_d$ | Day respiration | μmol $m^{-2}$ $s^{-1}$ |
| $r_m$ | Mesophyll resistance, i.e. the reciprocal of $g_m$ | $mol^{-1}$ $m^2$ s bar |
| $s$ | Product of $\alpha$ and $\beta$; a combined parameter of light absorption and partition to PSII | Unitless |
| $V_{cmax}$ | $CO_2$ saturated Rubisco carboxylation rate | μmol $m^{-2}$ $s^{-1}$ |
| $W_c$ | Rubisco limited carboxylation rate | μmol $m^{-2}$ $s^{-1}$ |
| $W_j$ | RuBP regeneration limited carboxylation rate | μmol $m^{-2}$ $s^{-1}$ |
| $Y(II)$ | Quantum efficiency of absorbed photons on PSII | Unitless |
| $Y(II)_{LL}$ | Initial slope of $Y(II)$-$I \cdot s$ curve | Unitless |
| $\theta$ | Curvature index of $J$ to $I$ | Unitless |
| $\alpha$ | Light absorption | Unitless |
| $\beta$ | Partition of total absorbed light to PSII | Unitless |
| $\Gamma^*$ | $C_c$-based $CO_2$ compensation point in the presence of $R_d$ | μbar |

**Table 2.** Standard deviations (std) of the marginal posterior distributions of estimated parameters using the synthetic concurrent $A_n$-$C_i$ and $Y(II)$-$C_i$ data shown in Fig. 1 A & B. Four scenarios with different prefixed parameters are conducted here. Prefixed parameters are fixed at the true values used to generate the synthetic data. The remaining parameters are estimated with Bayesian estimation. Numbers in brackets show the std as a percentage of the true values. ** indicates parameter values with sample variation.

| Prefixed parameters | | $V_{cmax}$ (μmol m$^{-2}$ s$^{-1}$) | $J_{1500}$ (μmol m$^{-2}$ s$^{-1}$) | $R_d$ (μmol m$^{-2}$ s$^{-1}$) | $r_m$ (mol$^{-1}$ m$^2$ s bar) | $K_m$ (μbar) | $\Gamma^*$ (μbar) | $s$ (unitless) |
|---|---|---|---|---|---|---|---|---|
| | true | 73.0** | 142.3** | 0.84** | 5.0 | 535.3 | 38.5 | 0.368** |
| $K_m$, $\Gamma^*$, $s$ | mean | 73.2 | 142.0 | 0.81 | 5.1 | / | / | / |
| | std | 1.3 (1.7%) | 0.8 (0.5%) | 0.22 (26%) | 0.43 (8.7%) | 0 (0%) | 0 (0%) | 0 (0%) |
| $K_m$, $\Gamma^*$ | mean | 73.4 | 142.1 | 0.83 | 5.1 | / | / | 0.369 |
| | std | 1.6 (2.2%) | 1.3 (0.9%) | 0.28 (33%) | 0.43 (8.7%) | 0 (0%) | 0 (0%) | 0.0044 (1.2%) |
| $K_m$ | mean | 72.7 | 142.1 | 0.76 | 4.9 | / | 39.7 | 0.369 |
| | std | 6.5 (8.9%) | 1.6 (1.2%) | 0.60 (72%) | 1.2 (23%) | 0 (0%) | 4.5 (12%) | 0.0054 (1.5%) |
| $\Gamma^*$ | mean | 75.4 | 142.2 | 0.84 | 5.2 | 551 | / | 0.369 |
| | std | 9.5 (13.0%) | 1.4 (1.0%) | 0.30 (36%) | 0.59 (12%) | 77 (14%) | 0 (0%) | 0.0044 (1.2%) |
| / | mean | 77.1 | 142.0 | 0.62 | 4.8 | 590 | 40.8 | 0.369 |
| | std | 10.3 (14.9%) | 1.9 (1.4%) | 0.87 (104%) | 1.2 (24%) | 112 (21%) | 6.0 (16%) | 0.0057 (1.6%) |

**Table 3.** Standard deviations (std) of the marginal posterior distributions of estimated parameters with synthetic concurrent $A_n$-$C_i$ and $Y(II)$-$C_i$ data under one saturating light (1500 PPFD) and three low light levels (50, 100 and 200 PPFD) in Fig. 1A-D. Bayesian estimation is applied to data with different low light (LL) measurements. Data from Rep. No. 1 alone is also shown. All these estimations prefix $K_m$ at the true value used to generate the synthetic data. ** indicates parameter values with sample variation.

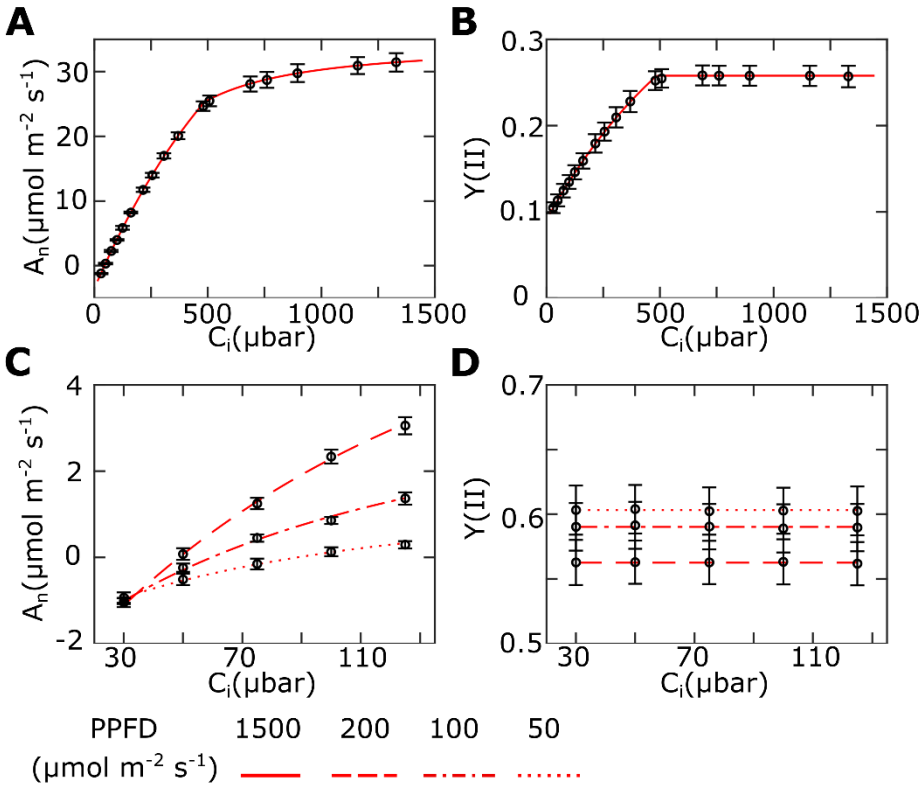
| Input data | | $V_{cmax}$ (μmol m$^{-2}$ s$^{-1}$) | $J_{1500}$ (μmol m$^{-2}$ s$^{-1}$) | $R_d$ (μmol m$^{-2}$ s$^{-1}$) | $r_m$ (mol$^{-1}$ m$^2$ s bar) | $K_m$ (μbar) | $\Gamma^*$ (μbar) | $s$ |
|---|---|---|---|---|---|---|---|---|
| | true | 73.0** | 142.3** | 0.84** | 5.0 | 535.3 | 38.5 | 0.368** |
| LL=50, 100, 200 | mean | 73.0 | 142.2 | 0.80 | 5.0 | / | 39.3 | 0.369 |
| | std | 3.4 (4.7%) | 0.7 (0.5%) | 0.14 (16%) | 0.72 (14%) | 0 (0%) | 2.0 (5.2%) | 0.0044 (1.2%) |
| LL=50, 100 | mean | 72.5 | 142 | 0.80 | 4.9 | / | 39.6 | 0.369 |
| | std | 3.8 (5.2%) | 0.8 (0.6%) | 0.15 (18%) | 0.84 (17%) | 0 (0%) | 2.4 (6.3%) | 0.0043 (1.2%) |
| LL=50, 200 | mean | 72.7 | 142.2 | 0.79 | 4.9 | / | 39.4 | 0.369 |
| | std | 3.5 (4.8%) | 0.8 (0.6%) | 0.16 (19%) | 0.75 (15%) | 0 (0%) | 2.2 (5.7%) | 0.0042 (1.1%) |
| LL=100, 200 | mean | 72.5 | 142.2 | 0.76 | 4.9 | / | 39.6 | 0.369 |
| | std | 4.1 (5.6%) | 0.8 (0.6%) | 0.23 (27%) | 0.85 (17%) | 0 (0%) | 2.8 (7.2%) | 0.0043 (1.2%) |
| LL=50 | mean | 72.4 | 142.2 | 0.79 | 4.8 | / | 39.6 | 0.369 |
| | std | 4.0 (5.4%) | 0.9 (0.6%) | 0.18 (21%) | 0.90 (18%) | 0 (0%) | 2.7 (6.9%) | 0.0041 (1.1%) |
| LL=100 | mean | 72.2 | 142.2 | 0.76 | 4.8 | / | 39.9 | 0.369 |
| | std | 4.5 (6.1%) | 0.9 (0.6%) | 0.25 (30%) | 0.97 (19%) | 0 (0%) | 3.1 (8.1%) | 0.0042 (1.1%) |
| LL=200 | mean | 72.2 | 142.1 | 0.71 | 4.8 | / | 40.0 | 0.369 |
| | std | 5.2 (7.1%) | 1.0 (0.7%) | 0.40 (47%) | 1.01 (20%) | 0 (0%) | 3.8 (9.7%) | 0.0044 (1.2%) |
| Rep. No.1 & LL=50, 100, 200 | true | 74.2 | 144.3 | 0.83 | 5.0 | 535.3 | 38.5 | 0.361 |
| | mean | 73.7 | 144.0 | 0.85 | 4.8 | / | 38.8 | 0.361 |
| | std | 0.97 (1.3%) | 0.3 (0.21%) | 0.052 (6.3%) | 0.23 (4.5%) | 0 (0%) | 0.7 (1.8%) | 0.0010 (0.28%) |

**Table 4.** Standard deviations of marginal posterior distributions of estimated parameters with experimental concurrent $A_n$-$C_i$ and $Y(II)$-$C_i$ data under one saturating light and three low light levels on rice leaves. Bayesian estimation is conducted with data of all replicates and with data of each replicate. For all estimations, $K_m$ is prefixed at 427.7 μbar (von Caemmerer 2000).

| Input exp. data | | $V_{cmax}$ (μmol m$^{-2}$ s$^{-1}$) | $J_{1500}$ (μmol m$^{-2}$ s$^{-1}$) | $R_d$ (μmol m$^{-2}$ s$^{-1}$) | $r_m$ (mol$^{-1}$ m$^2$ s bar) | $K_m$ (μbar) | $\Gamma^*$ (μbar) | $s$ |
|---|---|---|---|---|---|---|---|---|
| all reps. | mean | 101.1 | 179.5 | 1.67 | 0.51 | 427.7 | 33.5 | 0.300 |
| | std | 5.0 (4.9%) | 2.8 (1.6%) | 0.39 (23%) | 0.57 (81%) | 0 | 2.7 (7.6%) | 0.0060 (2.0%) |
| rep. No.1 | mean | 107.0 | 157.6 | 2.67 | 0.62 | 427.7 | 28.2 | 0.281 |
| | std | 8.6 (8.0%) | 4.5 (2.9%) | 0.64 (24%) | 0.57 (92%) | 0 | 3.7 (13%) | 0.0090 (3.2%) |
| rep. No.2 | mean | 109.6 | 196.9 | 2.71 | 0.60 | 427.7 | 35.7 | 0.305 |
| | std | 8.3 (7.5%) | 6.3 (3.2%) | 0.75 (28%) | 0.56 (94%) | 0 | 4.8 (13%) | 0.011 (3.7%) |
| rep. No.3 | mean | 101.6 | 168.8 | 1.24 | 1.05 | 427.7 | 32.1 | 0.308 |
| | std | 5.5 (5.4%) | 1.5 (0.9%) | 0.28 (23%) | 0.46 (44%) | 0 | 2.2 (6.9%) | 0.0044 (1.4%) |
| rep. No.4 | mean | 110.9 | 187.7 | 1.78 | 1.03 | 427.7 | 31.1 | 0.296 |
| | std | 11 (9.9%) | 3.3 (1.7%) | 0.58 (33%) | 0.76 (74%) | 0 | 4.0 (13%) | 0.0063 (2.1%) |

916 **Figures**

917 **Fig 1**




918

919 **Fig 1.** Synthetic concurrent gas exchange (A, C) and chlorophyll fluorescence (B, D)

920 measurements under saturating light level, i.e. 1500 PPFD (A, B) and three low light

921 levels, i.e. 50, 100 and 200 PPFD (C, D). Red lines are predicted lines calculated from

922 true values. Error bars plot five replicates containing sample variation and systematic

923 error as described in Eqn. S1.

924

**Fig 2**

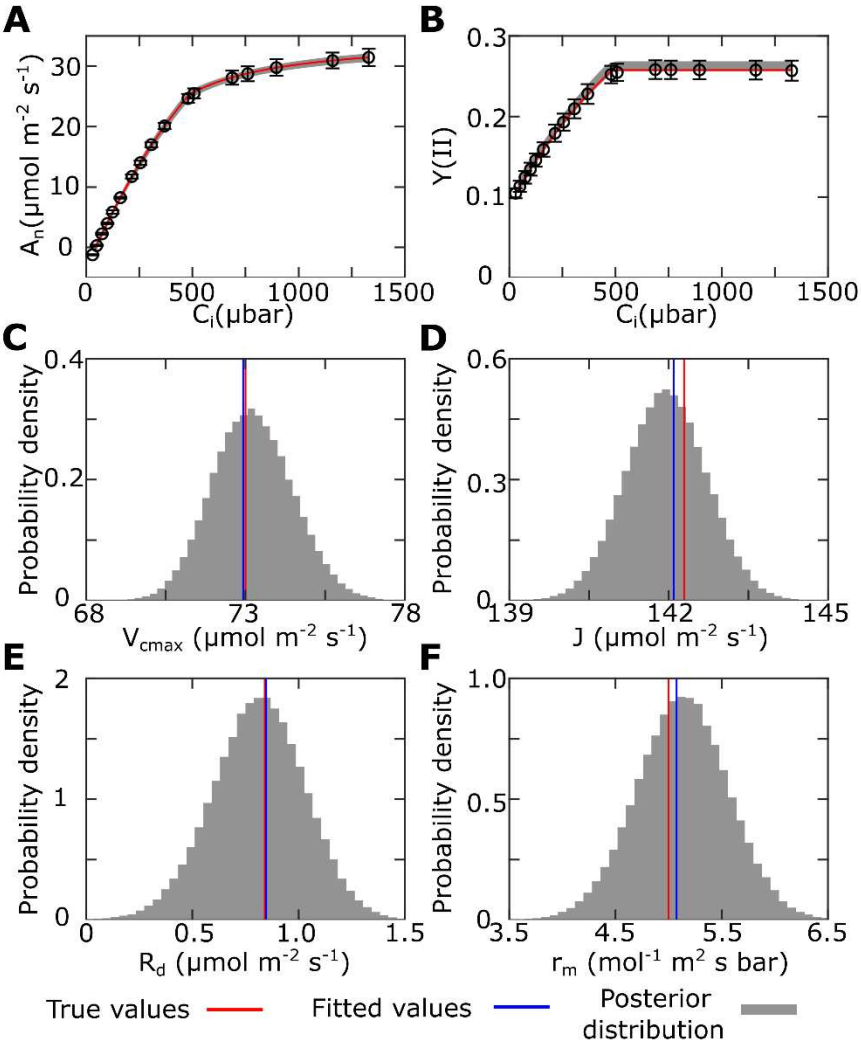
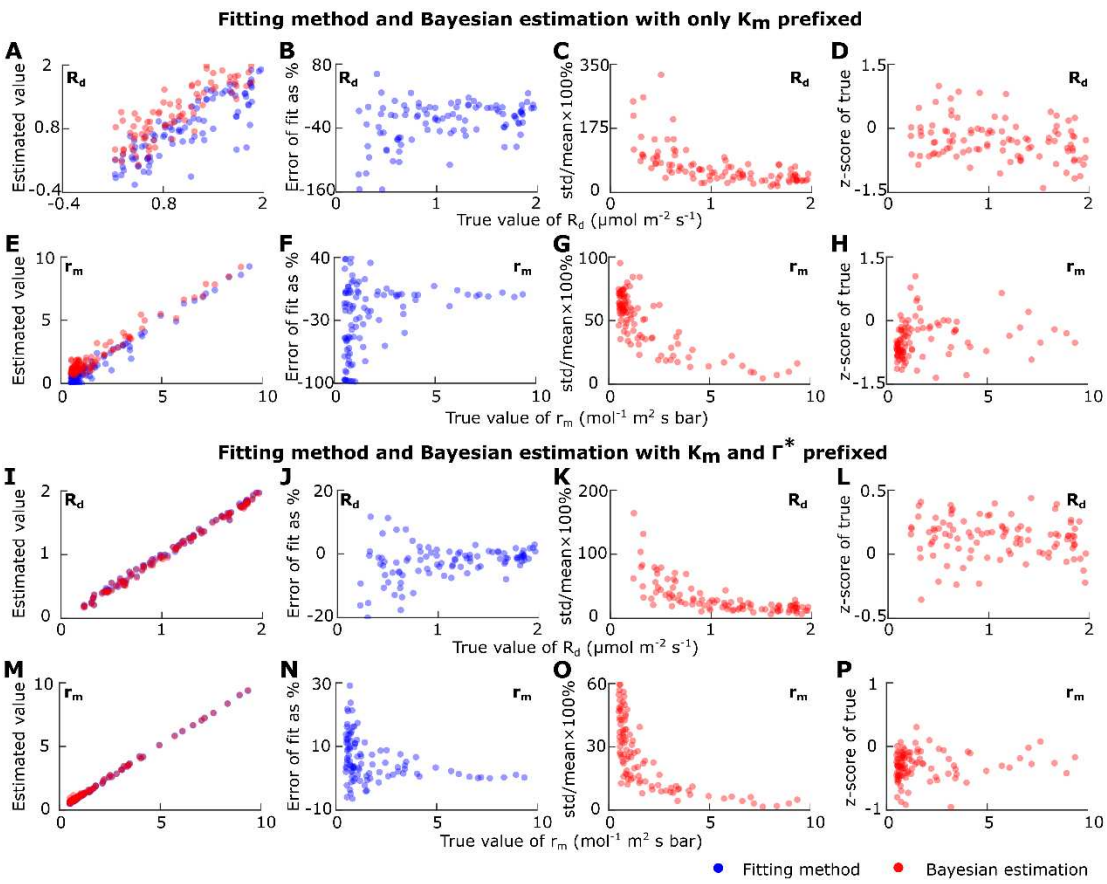



**Fig 2. Bayesian estimation with $A_n$-$C_i$ and $Y(II)$-$C_i$ data in Fig. 1 A & B under prefixed $K_m$, $\Gamma^*$ and $s$.**

(A) $A_n$-$C_i$ observations and Bayesian estimation. Red lines and error bars here are observations from Fig. 1 A. Grey regions are $A_n$-$C_i$ curves calculated from the joint posterior distribution. (B) $Y(II)$-$C_i$ observations (red lines and error bars) and Bayesian estimation (grey region). (C-F) Marginal posterior distributions of estimated parameters. Grey regions are the marginal posterior distribution. Blue solid lines indicate best fitted

value by a simultaneous fitting algorithm. Red solid lines are the mean of the five replicates (Eqn S2-S3).

936 **Fig 3**




937

938 **Fig 3. Fitting vs Bayesian estimation of $R_d$ and $r_m$.**

939 Comparison between fitting method and Bayesian estimation with 100 random

940 synthetic datasets. Both methods are conducted with $K_m$ (A-H) or $K_m$ & $\Gamma^*$ (I-P)

941 prefixing at true values. (A, E, I & M) Estimated values from both methods versus true

942 values. Blue dots represent best fitted values, while red dots represent mean values of

943 estimated marginal posterior distributions. (B, F, J & N) Error of fitted values relative

944 to the true values as percentage. (C, G, K & O) Standard deviations of estimated

945 marginal posterior distributions relative to the true values as percentage. (D, H, L & P)

946 z-scores of the true value against the marginal posterior distributions, which equals the

difference between the true value and the mean of distribution divided by the standard deviation of the distribution.

**Fig. 4**

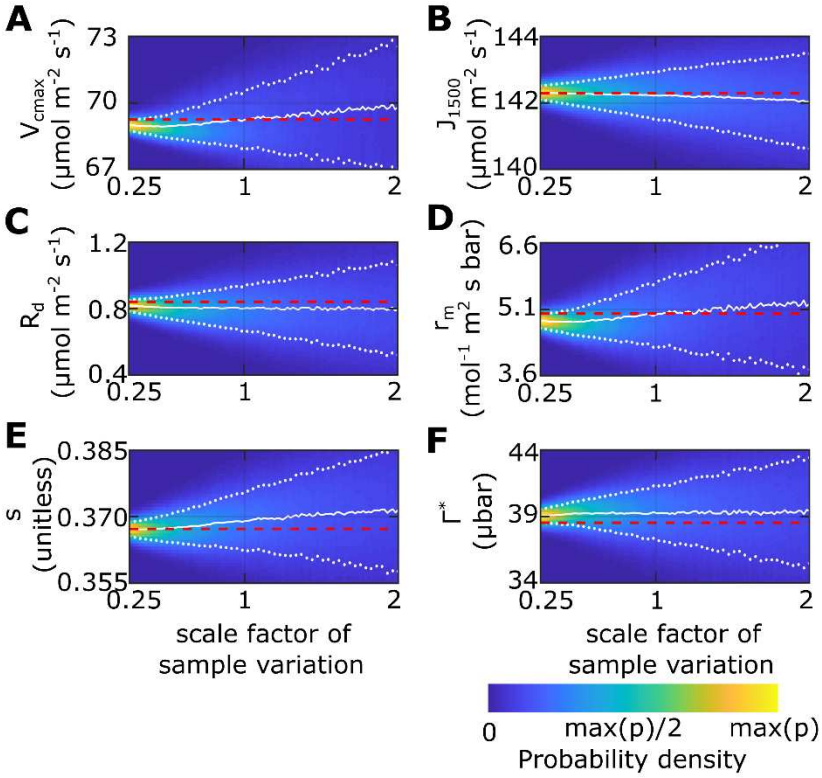



**Fig. 4: Sensitivity analysis of sample variation on parameter estimation.**

Based on synthetic data in Fig. 1, data with different amounts of sample variation are generated by scaling up the difference between photosynthetic parameters and true values in each replicate (Eqn 10 & 11). Each column of pixels represents a marginal posterior distribution. The colour bar represents the probability density. White solid lines represent the mean value of the marginal posterior distribution. White dashed lines represent the standard deviation of the marginal posterior distributions. Red dashed lines are the true values. For $V_{cmax}$, $J$, $R_d$, and $s$ with sample variation, red dashed lines actually represent the mean value of five replicates.

**Fig. 5**

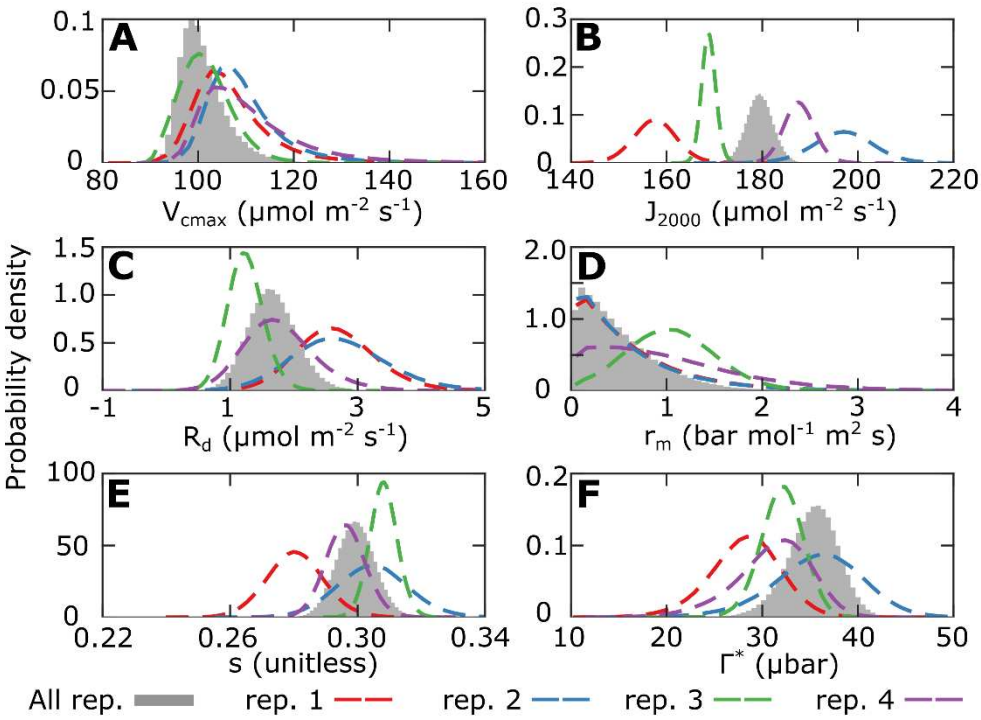
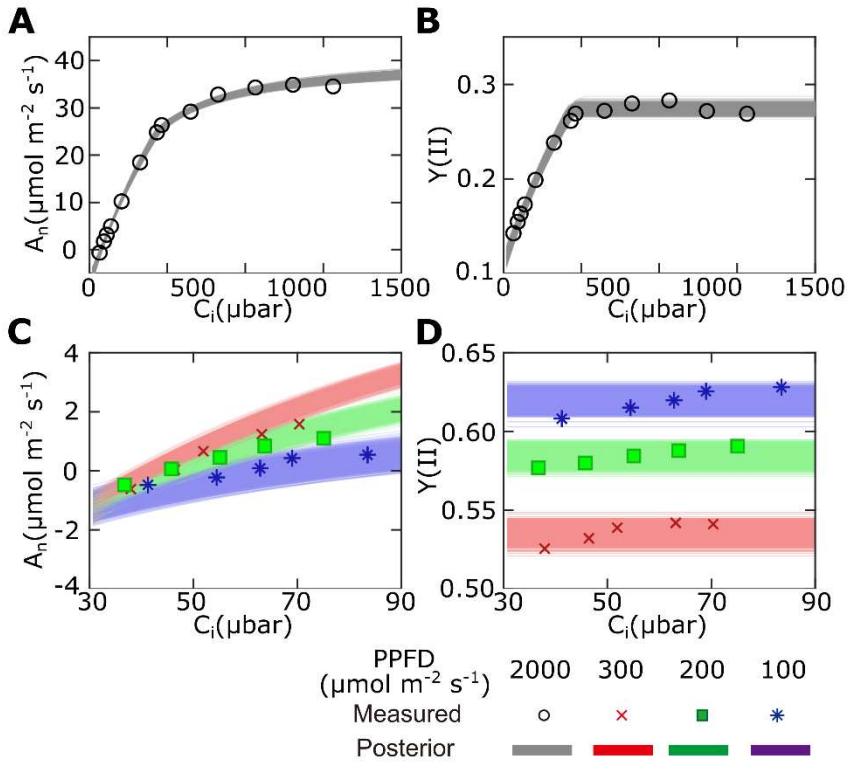



**Fig 5. Marginal posterior distributions estimated from experimental data on rice leaves.** Dashed lines in different colours represent the posterior distributions estimated for each replicate. Grey regions represent the posterior distribution estimated using the data from all the replicates.

970 **Fig. 6**




971

972 **Fig 6. $A_n$-$C_i$ and *Y(II)*-$C_i$ curves predicted from Bayesian estimation with one**

973 **experimental replicate.**

974 The data from replicate No. 3 (green dash lines in Fig. 5) is used for Bayesian estimation.

975 Open circles represent $A_n$-$C_i$ and *Y(II)*-$C_i$ curves under saturating light. Blue asterisks,

976 green squares and red crosses represent $A_n$-$C_i$ and *Y(II)*-$C_i$ curves under three different

977 low light levels. Grey regions are $A_n$-$C_i$ and *Y(II)*-$C_i$ curves under saturating light

978 calculated from the joint posterior distribution. Blue, green and red regions are $A_n$-$C_i$

979 and *Y(II)*-$C_i$ curves under three low light levels calculated from the joint posterior

980 distributions.

981

982