Modelling photosynthetic parameters via mixed effects models and the Farquhar-van Cammerer-Berry model

Stephen D. Coleman

April 14, 2019

Abstract

The study of the photosynthetic process frequently involves analysis of net assimilation - intercellular CO_2 concentration (ACI) curves. These are used in the estimation of key parameters associated with the Farquhar-van Caemmerer-Berry (FvCB) [8] model:

- $V_{c_{max}}$: the rate of maximum Rubisco carboxylation;
- J: electron transport rate;
- R_d: daytime respiration; and
- g_m: mesophyll conductance.

Accurate, unbiased estimation of these parameters is a non-trivial exercise with the optimal method still a matter of debate within the field [14][26][16]. The problem of selecting suitable starting values for a non-linear model is thus complicated by the lack of unanimity on what is a "good" estimate of these. The core model we are using to estimate the parameters is based on Yin et al. [27]. This method offers several steps to estimate various parameters rather than attempting estimation in a single step.

The data collected for the purpose of modelling steady-state photosynthesis is inevitably repeated measurements on individual plants. This leads to correlation between measurements, a fact that violates the assumptions in the fixed effects models used in estimating the parameters of interest [4][2]. We introduce the use of non-linear mixed effects models in keeping with Qian et al. [16] to overcome this issue, and extend on their work by considering more modern versions of the FvCB model.

Keywords— ACI curves, parameter estimation, FvCB model, non-linear mixed effects models

1 Theory

1.1 FvCB model

The brilliant realisation of Farquhar et al. [8] was that the net photosynthesis rate, A, of C_3 plants could be accurately modeled at the leaf level as being in one of two steady-state systems. In *Rubisco activity limited net photosynthesis rate* (A_c), the net rate of photosynthesis is limited by the properties of the enzyme ribulose 1.5 biphosphate carboxylase/oxygenase (Rubisco) (we assume saturating levels of the substrate, RuBP). In the second we assume that it is the substrate that determines the rate of assimilation; specifically its regeneration rate. Hence the second state is referred to as RuBP-regeneration limited or *electron transport limited net photosynthesis rate* (A_j)[27]. We prefer the latter name for being a more clear difference for readers unfamiliar with the biochemical constituents of the photosynthesis cycle.

1.1.1 Some assumptions

For parsimony's sake, we make several simplifying assumptions regarding the homogeneity of conditions within the leaf. We disregard any gradient of temperature (T) or O_2 within the leaf. According to von Caemmerer [22], these, specifically within C_3 species, are unlikely to be important. Due to the thinness of leaves we also assume an equal distribution of light to all chloroplasts within the leaf; however, it has been shown that under certain conditions these assumptions and their corollary, that photosynthesis itself is homogeneous across the leaf, do not hold. Non-uniform photosynthesis can occur and this will affect the gas-exchange measurements and their interpretation [21].

We account for the *Kok effect* in our modelling. This is a physiological effect where in low light (typically $0 - 20\mu\text{mol}\,\text{m}^{-2}\,\text{s}^{-1}$ [20]) conditions the quantum yield of net photosynthesis is higher than at higher levels of incident light. Thus, the net assimilation recorded tends to escape description by our model. This is a well documented effect [20][19][13] and hence in some the low light data the first point (we will often consider increasing intercellular carbon dioxide (C_i) or light intensity (I) as pseudo-time in our descriptions) will be ignored in our models.

Several parameters require estimation, and some measurements are not feasible for the majority of experiments. For these we need either a starting value that is our initial best guess (feasible in C_3 species, less so in C_4) or a correction to some related measurement. We will show the reasoning behind our choices below.

1.1.2 Rubisco activity limited net photosynthesis rate

The photosynthetic carbon reduction (PCR) and photorespiratory carbon oxidation (PCO) cycles are linked by an enzyme common to both, RuBP [8]. In the PCO cycle, 0.5 mol of CO_2 is released, thus:

$$A = V_c - 0.5V_o - R_d \tag{1}$$

Where:

- V_c is the rate of carboxylation; and
- V_o is the rate of oxygenation.

We have from Farquhar [9] (assuming RuBP binding first in the reaction):

$$V_c = V_{c_{max}} \frac{C_c}{C_c + K_c\,(1 + O/K_o)} \cdot \frac{R}{R + K_r'} \tag{2}$$

$$V_o = V_{o_{max}} \frac{O}{O + K_o\,(1 + C_c/K_c)} \cdot \frac{R}{R + K_r'} \tag{3}$$

Where:

- $V_{c_{max}}$ is the maximum velocity of the carboxylase;
- $V_{o_{max}}$ is the maximum velocity of the oxygenase;
- O is the partial pressure of oxygen in the chloroplast;
- K_c is the Michaelis-Menten constant for CO_2;
- K_o is the Michaelis-Menten constant for O;
- R is the concentration of free RuBP; and
- K_r' is the effective Michaelis-Menten constant for RuBP.

This leads to the ratio of oxygenation to carboxylation:

$$\phi = \frac{V_o}{V_c} = \frac{V_{o_{max}}}{V_{c_{max}}} \cdot \frac{O/K_o}{C_c/K_c} \tag{4}$$

We also have that the velocity of carboxylation, at saturating levels of RuBP, (W_c) is given by [10]:

$$W_c = V_{c_{max}} \frac{C_c}{C_c + K_c(1 + O/K_o)} \tag{5}$$

Note that this arises from setting the RuBP component of (2) to 1. The compensation point for C_c at which there is no net assimilation, Γ_*, is given by [8]:

$$\Gamma_* = \frac{K_c O}{2K_o} \cdot \frac{V_{o_{max}}}{V_{c_{max}}} \tag{6}$$

Relating (1), (4), (5) and (6) we have:

$$A_c = V_{c_{max}} \frac{C_c - \Gamma_*}{C_c + K_c(1 + O/K_o)} - R_d \tag{7}$$

This is the Rubisco-limited rate of assimilation, assuming saturation of RuBP.

1.1.3 Electron transport limited net photosynthesis rate

The physiology of this model is somewhat more complex, involving analysis of the cycle for the production of NADPH [8]. A thorough understanding of this is beyond the scope of this project, but we recommend the reader to the original paper [8] and von Caemmerer [22]. We are content to state that when A is limited by RuBP regeneration:

$$A_j = J \frac{C_c - \Gamma_*}{4C_c + 8\Gamma^*} - R_d \tag{8}$$

Where J is the rate of electron transport (and hence electron transport limited net photosynthesis rate). This equation arises using assumptions that RuBP regeneration is limited because of insufficient NADPH and implicitly assumes 100% linear e^- transport. Sharkey et al. [18] recommend the use of 4 and 8 in the denominator based on the number of electrons required for $NADP^+$ reduction; however this is not always the case; values of 4.5 and 10.5 also occur in the literature and less frequently we see 4 and 9.33 (see [18][27][25]). These different values arise from different assumptions about the complexity of linear electron (e^-) rate. The disagreement over pairs $\{4.5, 10.5\}$ or $\{4, 9.33\}$ arise due to further disagreement about deeper levels of this system that are a source of some discussion in the literature ([22][27]) but are less of interest to us. To allow for this disagreement we use default values of 4 and 8 in our R model, but the functions incorporating this aspect of the model allow the user to input their own version of these coefficients.

1.1.4 TPU limited net photosynthesis rate

Other models extend to different cases (such as C_4 species) or to encompass more physiologically limiting factors (such as triose phosphate utilization (TPU) in the synthesis of starch and sucrose). TPU is required at one third the rate of CO_2 fixation. If this is not acquired, the level of free phosphate declines and the rate of photosynthesis is limited. When this is limit is imposed, it can be seen that photosynthesis becomes independent of the ratio of oxygenation to carboxylation [17]. Sharkey et al. [18] model this *TPU limited net photosynthesis rate* (A_p) using the following equation:

$$A_p = 3TPU - R_d \tag{9}$$

Where TPU is the rate of triose phosphate utilization. This extension is only relevant under high C_i and I environments and is normally modeled as a upper bound on A. We do not include this in our model, but recommend extending the model to this scenario in later versions. From von Caemmerer [22] we know that this model is too simple for the reality, please see section 2.4.3 for an extension as formulated by Harley and Sharkey [12]. As (9) is only apparent in extreme cases and we ignore it for now but recommend inclusion of this additional case in any extensions to our work.

1.1.5 FvCB summary

For the final model, our actual assimilation rate is calculated by combining these individual curves. The curve fitted is:

$$A = \min\left\{A_c, A_j\right\} \tag{10}$$

Where:

- A_c is the Rubisco activity limited net photosynthesis rate calculated from (7); and
- A_j is the electron transport limited net photosynthesis rate calculated from (8).

1.1.6 Relating C_i to C_c

One of the driving factors in net assimilation rate (A) is the level of CO_2 present at the Rubisco site of carboxylation in the chloroplasts (C_c) [23]. Estimation of the intercellular CO_2 partial pressure (C_i) is based on gas exchange measurements and is not a matter of controversy. Less well established is estimation of C_c [26]. We can relate C_i to C_c using Flick's first law relating the diffusive flux to the concentration under steady-state assumptions. We have:

$$J = -D\frac{\mathrm{d}\phi}{\mathrm{d}x} \tag{11}$$

Where:

- J is the diffusion flux;
- D is the diffusion coefficient;
- ϕ is the concentration; and
- x is position.

Please note that the notation used in (11) is the generic notation used normally in describing Flick's Law and will not be referring to these terms after this subsection. From this we can consider A as our flux, and the difference between C_c and C_i as our gradient in concentration:

$$A = -D(C_c - C_i) \tag{12}$$

The diffusion conductance between the substomatal cavities and the chloroplasts is g_m, the mesophyll conductance [15]. Hence we can write (12) as:

$$A = g_m(C_i - C_c) \tag{13}$$

Or equivalently:

$$C_c = C_i - \frac{A}{g_m} \tag{14}$$

Thus we can calculate C_c, the variable relevant to our models, using C_i.

1.1.7 Mesophyll conductance

To estimate values of C_c we need to have an estimate of g_m. Harley et al. [11] propose a model for this:

$$g_m = \frac{A}{C_i - \frac{\Gamma_*[J+8(A+R_d)]}{J-4(A+R_d)}} \tag{15}$$

1.1.8 Light dependence of electron transport rate

We can relate J_{max} to the incident irradiance by empirical equation:

$$\theta J^2 - J\left(I_2 + J_{max}\right) + I_2 J_{max} = 0 \tag{16}$$

Where:

- I_2 is the useful light absorbed by PSII (photosystem two);
- J is the electron transport;
- J_{max} is the maximum electron transport; and
- θ is an empirical curvature factor (often around 0.7 [6]).

I_2 is related to total incident irradiance I (sometimes referred to as the Photosynthetically Active Radiation, PAR, in the literature) by:

$$I_2 = I \times \text{abs}(1 - f)/2 \tag{17}$$

The coefficient of I here is often collapsed into a lump term, α which is estimated. Where:

- abs is the absorptance of leaves (commonly around 0.85 [22]);
- f is the correction for the spectral quality of light (approximately 0.15 [7]); and
- The 2 in the denominator is due to the split of light between photosystems I and II (PSI and PSII respectively).

We can thus solve for J in the usual way ([16] [25]):

$$J = \frac{\alpha \cdot I + J_{max} - \sqrt{\left(\alpha \cdot I + J_{max}\right)^2 - 4 \cdot \alpha \cdot \theta \cdot I \cdot J_{max}}}{2 \cdot \theta} \tag{18}$$

Equations of this form are referred to as *non-rectangular hyperbola functions*.

1.1.9 Lump parameter, s

Yin et al. [27] propose use of a lump parameter, s, in their photosynthesis model. This parameter is a function of several variables describing the photosystems I and II. Specifically it is:

$$s = \rho_2 \beta \left(1 - \frac{f_{pseudo(b)}}{1 - f_{cyc}}\right) \tag{19}$$

Where:

- ρ_2 is the proportion of light absorbed by photosynthetic pigments partitioned to PSII (Yin et al. [27] suggest this is in the range of 0.5 but suggest some caution in using this value. We suspect that the 2 in the denominator of (17) corresponds to this coefficient. Again the literature finds a new name and symbol for a pre-existing concept);
- β is the absorptance by leaf photosynthetic pigments (abs in (17));
- $f_{pseudo(b)}$ is the fraction of electrons at PSI that follow the basal pseudocyclic e^- flow; and
- f_{cyc} is the fraction of electrons at PSI that follow cyclic transport around PSI.

Note that f in (17) corresponds to $\frac{f_{pseudo(b)}}{1-f_{cyc}}$.

This highlights the problem rampant in the photosynthesis modeling literature - this appears to be another way of representing the previously defined lump parameter α with more detail (more

aspects of the photosystems are accounted for). However, it is a completely different set of symbols and no acknowledgment of overlap despite α being far more common in preceding papers. Possibly this is not problematic for plant physiologists, but we suspect that this has lead to confusion and could be avoided by either continuity in symbols or else directly acknowledging past preferences in the current form.

1.1.10 Relative CO_2/O_2 specificity factor for Rubisco, $S_{c/o}$

The relative CO_2/O_2 specificity factor for Rubisco, $S_{c/o}$, is used in calculating both Γ_* and J in the linear part of the light response curve [27]. It is estimated using the linear part of the high and low oxygen ACI curves. Yin et al. [27] propose a relationship between the net rate of photosynthesis in the linear part of the high and low oxygen ACI curves defined by:

$$A_h = (A_l + R_d)\frac{b_h}{b_l} - \left(\frac{O_h - O_l}{2S_{c/o}} + C_{i_l} - C_{i_h}\right) b_h - R_d \tag{20}$$

By solving for $S_{c/o}$ we find:

$$S_{c/o} = \frac{(O_h - O_l)}{2((A_l + R_d)/b_l - (A_h + R_d)/b_h - (C_{i_l} - C_{i_h}))} \tag{21}$$

In these equations the subscripts refer to which dataset the parameter is from, l for the low oxygen dataset and h for the high oxygen set respectively. An explanation of these different datasets is sketched below in section 2. b_h and b_l are the slope of the linear part of the ACI curve under the different oxygen conditions.

1.1.11 CO_2 compensation point in the absence of R_d

The C_i-based CO_2 compensation point in the absence of R_d, Γ_*, is point at which the plant's production of CO_2 has a net value of 0; i.e. it is consuming as much CO_2 in photosynthesis as it produces through photorespiration (as R_d, the day respiration, is respiratory CO_2 release other than by photorespiration). We let:

$$\Gamma_* = \frac{O}{2S_{c/o}} \tag{22}$$

1.1.12 Temperature dependency

The dependency of the rate of carboxylation and oxygenation of Rubisco is reflected in the temperature dependency of A. As a rate dependent upon temperature, we will be using Arrhenius functions, equations of the form:

$$x(T) = k \cdot \exp\left(-\frac{E_a}{R \cdot T}\right) \tag{23}$$

Where:

- $x(T)$: the temperature dependent rate we are interested in;
- k: the rate coefficient;
- E_a: the activation energy, the kinetic energy of substrate required for the reaction to occur;
- R: the universal gas constant ($8.314\,\mathrm{J\,K^{-1}\,mol^{-1}}$); and
- T: temperature (in Kelvin).

As most photosynthesis measurement are recorded in °C, and specifically use a default value of 25 °C, we transform the equation to this scale. In this case we have:

$$x(T') = k' \cdot \exp\left(-\frac{(25 - T')E_a}{298.15R \cdot (273.15 + T')}\right) \tag{24}$$

Where T', k' are the relevant form of T, k for a scale centred on $T = 25\,°C$. As we will not be using the Kelvin scale again, we denote $T = T'$, $k = k'$. Many reactions in photosynthesis are reversible, with differing activation energies depending on the direction of the reaction. Thus the net activation energy may vary depending on the ratio of forward to backward reactions. This means that the Arrhenius function (24) is only semi-empirical, but it does allow easy comparison between studies.

Another frequently used method to describe a dependency on temperature is the Q_{10} *temperature coefficient*. This is a measure of of the rate of change of a system as a result of raising the temperature by 10 °C. It is calculated:

$$Q_{10} = \left(\frac{R(T_2)}{R(T_1)}\right)^{10\,°C/(T_2 - T_1)} \tag{25}$$

Where:

- Q_{10} is the factor by which the reaction rate increases when temperature is raised by 10 °C;
- T_i is the temperature for the i^{th} measurement; and
- $R(T_i)$ is the rate of the reaction at temperature T_i.

Alternatively we can write this in the form:

$$R(T_2) = R(T_1)Q_{10}^{(T_2 - T_1)/10\,°C} \tag{26}$$

Again, this general form is of less interest as we have a specific, default temperature of 25 °C for which we are quite well informed. Hence, we use:

$$R(T) = R(25\,°C)Q_{10}^{(T-25)/10} \tag{27}$$

This function allows us to relate values given at the temperature $T = 25\,°C$ to more general temperatures.

1.1.13 Model parameters

Many parameters can be assigned *a priori*, leaving only the estimation of a small number of key variables. The kinetic constants of rubisco vary very little among C_3 species such that one can use the same K_c, K_o and Γ_* across all members of this category [22]. This means that the only parameter requiring estimation for Rubisco is the maximal Rubisco activity, $V_{c_{max}}$.

For RuBP-limited photosynthesis, we need to solve for J_{max} as we can relate this to J by equation (18). According to Walcroft et al. [24], the ratio $J_{max} : V_{c_{max}}$ is expected to vary from 2 to 1.4 across the range $[8\,°C, 30\,°C]$.

Evans and von Caemmerer [5] recommend setting $g_m = 0.0045V_{c_{max}}$. Other values can be seen, with an assumption of $g_m = \infty$ used frequently, but this is controversial and can lead to biased estimates of $V_{c_{max}}$ and J_{max} [26]. Specific plants in specific conditions can see diverging values, but we feel this is a sufficient initial value.

This means that if we have an empirically driven initial value for $V_{c_{max}}$ we can initialise all our parameters to some default temperature, however these relationships vary with temperature, the relationship of which we state below.

We use table 1 for our initial values of photosynthetic parameters in our model. We can combine these with (27) to calculate initial values for our model across a range of temperatures. Many of the values in table 1 can be found in von Caemmerer [22].

Parameter	*unit*	Value	E ($\mathrm{kJ\,mol^{-1}}$)	$Q_{10}(T = 25\,^{\circ}\mathrm{C})$
K_c	µbar	260 or 404[a]	59.36[b]	2.24
K_o	mbar	179 or 248[a]	35.94	1.63
Γ_*	µbar	38.6 or 37[a]	23.4	1.37
$V_{c_{max}}$	$\mu\mathrm{mol\,m^{-2}\,s^{-1}}$	80	58.52	2.21
$V_{o_{max}}$	$\mu\mathrm{mol\,m^{-2}\,s^{-1}}$	$0.25 \times V_{c_{max}}$	58.52	2.21
R_d	$\mu\mathrm{mol\,m^{-2}\,s^{-1}}$	$0.01 - 0.02 \times V_{c_{max}}$	66.4	2.46
J_{max}	$\mu\mathrm{mol\,m^{-2}\,s^{-1}}$	$1.4 - 2.0 \times V_{c_{max}}$	37	1.65
g_m	$\mu\mathrm{mol\,m^{-2}\,s^{-1}}$	$0.0045 {V_{c_{max}}}^{c}$ or ∞		
θ		0.7		

[a] The first value is appropriate when an internal diffusion conductance is included; the second value should be used if the internal conductance is not included (i.e. $g_m = \infty$) and C_c is assumed to equal C_i.
[b] From Farquhar et al. [8]
[c] From Evans and von Caemmerer [5]

Table 1: Photosynthetic parameters and their activation energy for $T = 25\,^{\circ}\mathrm{C}$

The assumption that $g_m = \infty$, equivalent to $C_c = C_i$ is now considered redundant but would severely change these values. We include a sample of this impact to enable comparison with older papers and research.

1.2 Mixed-effects models

1.2.1 Linear mixed-effects models

Traditional parametric models incorporate only *fixed effects*. That is, they have a set of parameters describing with the entire population. For example, consider a system of $n \in \mathbb{N}$ observations of dependent variable, $Y = (y_1, \ldots, y_n)$, and the $n \times p$ matrix of associated independent variable measurements, $X = \{x_{i,j}\}$ for $i \in [1, n], j \in [1, p]$ for some $p \in \mathbb{N}$, and a p-vector of weights, β, represented:

$$Y = X\beta + \epsilon \tag{28}$$

Here X is assumed to contain a column of 1's in the first position (hence the +1 in the description of its dimensionality) and ϵ is the n-vector of associated errors. We assume that $\epsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$ for $i = (1, \ldots, n)$. If we consider the intuitive case of *biological* and *technical replicates*, where we have n biological replicates and for the i^{th} sample m_i associated technical replicates. As each of the m_i measurements is on the same sample we expect there to be a non-negligible correlation between the m_i technical replicates for each i. The model in (28) does not allow for the within group effects due to the correlation between technical replicates. Consider the simplest possible model for the fixed effects model including only the intercept:

$$y_{i,j} = \beta + \epsilon_{i,j}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, m_i \tag{29}$$

Some of this model's limitations become apparent if one considers the case of unbalanced data. Continuing the previous example of biological and technical replicates, consider the case that $m_i \neq m_j$ for any $i, j \in (1, \ldots, n)$. In this case the model is skewed by the within sample data rather than by the true observations, the biological replicates. A possible solution is the inclusion of a individual intercept for each group of technical replicates, accommodating the within-sample variability, or *random effects*:

$$y_{i,j} = \beta_i + \epsilon_{i,j}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, m_i \tag{30}$$

While this better describes the observed data, it has certain inherent flaws; most notably the number of parameters scales linearly with the number of observations and the model only describes the measurements included in the sample. Consider as a solution a combination of these models, containing both a sample mean, β, and a random variable for each group representing the deviation from the population mean, b_i, i.e. a *mixed effects* model:

$$y_{i,j} = \beta + b_i + \epsilon_{i,j}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, m_i \tag{31}$$

The linear mixed effects model described in (31) contains information at both a population level (in the fixed effects) but also at the individual level (in the random effects).

For now we assume $b_i \overset{iid}{\sim} \mathcal{N}(0, \sigma_b^2)$ for $i = 1, \ldots, n$. This means that the variance of the observations is divided into two parts, σ_b^2 for the biological variability and σ^2 for the technical variability:

$$b_i \overset{iid}{\sim} \mathcal{N}(0, \sigma_b^2), \quad \epsilon_{i,j} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2) \tag{32}$$

The assumption of normality can be modified if deemed inappropriate and it is possible to generalise the model to allow for heteroscedasticity.

The b_i are called *random effects* as they are associated with the experimental unit and selected at random from the population of interest (at least in theory, obviously there are limitations on this particularly in the area of medicine) They represent that the effect of choosing the sample i is to shift the mean expression of Y from β to $\beta + b_i$ - i.e. they effect a deviation from an overall mean. Technical replicates share the same random effect b_i and are correlated. The covariance between technical replicates on the same experimental unit is σ_b^2; this corresponds to a correlation of $\frac{\sigma_b^2}{(\sigma_b^2 + \sigma^2)}$.

1.2.2 Non-linear mixed effects models

Non-linear mixed effects models, also known as *non-linear hierarchical models*, are an extension to the more traditional linear mixed effects models. They are used in scenarios where all of the following features are present [3]:

1. Repeated observations of a continuous variable on each of several *experimental units* (in our case these are individuals, thus individual is considered equivalent to empircial unit in the following section) over time or another condition (e.g. measurements at given heights on a tree) (the *condition variable*);
2. We expect the relationship between the response variable and the condition variable to vary across individuals; and
3. Availability of a scientifically relevant model characterising the behaviour of the individual response in terms of meaningful parameters that vary across individuals and dictate variation in patterns of condition-response (for us this will be the Farquhar-van Cammerer-Berry model).

The final point from the list in 1.2.2 is where the non-linear aspect is introduced. This is often a mechanistic function describing a physical or chemical system (for example in toxicokineitcs physiologically-based pharmacokinetics models are used or HIV dynamics in "precision medicine"); that is the model is described by by meaningful, interpretable parameters rather than being an empirical best fit. It is expected that the mechanistic model will better describe data beyond the range of the measurements used here (whereas an empirical fit might describe the data over the range captured in the measurement data but then misbehave woefully beyond these boundaries).

The analysis tends to have the goal of understanding one or more of the following:

1. The "typical" behaviour of the phenomena (i.e. mean or median values) represented by the model parameters;
2. The variation of these parameters, and hence the phenomena, between individuals; and

3. If some of the variation is inherently associated with individual characteristics.

Individual level prediction can also be of interest (e.g. in medical treatment with highly individual reaction), but is less relevant to this project. From the individual level we are interested in investigating the level of variation between individuals and questioning if this is sufficiently small to allow the "all-purpose" models generally used in photosynthesis describing the parameters and systems of interest.

Consider an experiment involving repeated measurements of some response variable, Y, across a condition variable T for n individuals. Each individual's characteristics are recorded in A (this is assumed to be time-independent measurements, like an individual's height over the course of a drugs trial) and possible additional initial conditions in U (for example the dose of a drug the individual was receiving at time $t = 0$). Let $y_{i,j}$ denote the jth measurement of the response under condition $t_{i,j}$, $j = 1, \ldots, n_i$, for individual i, $i = 1, \ldots, n$ with additional initial conditions u_i. Thus:

$$\begin{array}{ll} Y & = (y_1, \ldots, y_n) \\ y_i & = \left(y_{i,1}, \ldots, y_{i,n_i}\right) \\ T & = (t_1, \ldots, t_n) \\ t_i & = \left(t_{i,1}, \ldots, t_{i,n_i}\right) \\ A & = (a_1, \ldots, a_n) \\ U & = (u_1, \ldots, u_n) \end{array} \tag{33}$$

We denote $x_{i,j} = \left(t_{i,j}, u_i\right)$. Often T is time and $U = \emptyset$, but it might be the case that each $t_{i,j}$ is a p_1-vector of measurements and u_i is a p_2-vector such that $p_1 + p_2 = p$, returning to the notation of (31). The assumption that the triplets (y_i, u_i, a_i) are independent across i is often included to reflect the belief that individuals are unrelated (this will hold for us, but might require more thought in other situations). For some function f regulating the within-individual behaviour defined by a vector of parameters β_i unique to individual i, we have:

$$y_{i,j} = f\left(x_{i,j}, \beta_i\right) + \epsilon_{i,j}, \quad j = 1, \ldots, n_i \tag{34}$$

For us this will be the Farquhar-van Cammerer-Berry model of steady-state photosynthesis. We assume that $\mathbb{E}\left(\epsilon_{i,j} | u_i, \beta_i\right) = 0$ and $\epsilon_{i,j} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$ for all i, j (the assumption of normality can be relaxed, but this extension exceeds the reach of this project). This model in (34) is called the *individual level model*. To model the population parameters we consider d, a p-dimensional function depending on an r-vector of fixed parameters, or *fixed effects*, β, and a k-vector of *random effects*, b_i, associated with individual i:

$$\beta_i = d\left(a_i, \beta, b_i\right), \quad i = 1, \ldots, n \tag{35}$$

Here, the *population model* in (35) describes how β_i varies among individuals due to both individual attributes a_i and biological variation in b_i. We assume that the b_i are independent of the a_i, i.e.:

$$\begin{array}{l} \mathbb{E}(b_i | a_i) = \mathbb{E}(b_i) = 0 \\ \mathbb{V}ar(b_i | a_i) = \mathbb{V}ar(b_i) = D \end{array} \tag{36}$$

Here, D is an unstructured covariance matrix and is common to all i. It characterises the degree of unexplained variation in the elements of β_i and associations among them; the ubiquitous assumption is $b_i \sim \mathcal{N}(0, D)$. However, if this set of assumptions regarding the conditional distribution of b_i on a_i is found to be insufficient, then $b_i \sim \mathcal{N}\left(0, D(a_i)\right)$ is frequently used.

In (35), β_i is considered to have an associated random effect, reflecting the belief that each component varies non-negligibly in the population even after systematic relationships with subject characteristics are accounted for. It may happen that "unexplained" variance in a component of β_i may be very small in magnitude relative to that in the remaining elements. In this situation it is common to drop the negligible quantity entirely. This lacks biological sense as each parameter is part of

the "scientifically relevant model" and thus is unlikely to have no associated unexplained variation. Hence, one must recognise that this omission of an element of β_i is adopted to achieve numerical stability in fitting rather than to reflect belief in perfect biological consistency across individuals and analyses in the literature to determine whether elements of β_i are fixed or random effects should be interpreted so.

1.3 Our model

Much of the preceding concepts and theory make for very pleasant reading, but we consider it useful to explicitly state our mixed-effects model and it's assumptions in one place.

First, recall the FvCB model (10):

$$A = \min\left\{A_c, A_j\right\} \tag{37}$$

$$= FvCB(C_c, \Gamma_*, I, K_c, K_o, O, J_{max}, V_{c_{max}}, \alpha, \theta, R_{d_i}) \tag{38}$$

Where A_c and A_j are two curves describing separate biochemical limits on the rate of photosynthesis, from (7) and (8) respectively:

$$A_c = V_{c_{max}} \frac{C_c - \Gamma_*}{C_c + K_c(1 + O/K_o)} - R_d \tag{39}$$

$$A_j = J \frac{C_c - \Gamma_*}{4C_c + 8\Gamma_*} - R_d \tag{40}$$

Furthermore, from (18):

$$J = \frac{\alpha \cdot I + J_{max} - \sqrt{(\alpha \cdot I + J_{max})^2 - 4 \cdot \alpha \cdot \theta \cdot I \cdot J_{max}}}{2 \cdot \theta} \tag{41}$$

We plug this into (34) for each of our experimental effects (in Rachel Schipper's data, SIDE and TREATMENT) in line with Bates and Pinheiro [1]. Thus, we have:

- Our response variable, Y, is net photosynthesis rate, A;
- Our individual level function, $f(\cdot)$, is $FvCB(\cdot)$;
- Our measured condition variables, X, are $C_c, \Gamma_*, I, K_c, K_o$, and O; and
- Our fixed effects, β, are $J_{max}, V_{c_{max}}, \alpha, \theta$ and R_d.

Notice that $U = \emptyset$. Therefore our individual-level model is:

$$\begin{aligned} &A_{i,j} = FvCB(C_{c_{i,j}}, \Gamma_{*_{i,j}}, I_{i,j}, K_c, K_o, O_{i,j}, J_{max_i}, V_{c_{max_i}}, \alpha_i, \theta_i, R_{d_i}) + \epsilon_{i,j} \\ &\epsilon_{i,j} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2) \\ &\mathbb{E}\left(\epsilon_{i,j} | J_{max_i}, V_{c_{max_i}}, \alpha_i, \theta_i, R_{d_i}\right) = 0 \end{aligned} \tag{42}$$

For our specific version of (35), we cannot know in advance what kind of model we will use. This will be based on the ability of different models to converge. We expect, given the paucity of data that we will have a diagonal covariance structure with 0's in the non-diagonal entries imposed on our mixed effects model (as we expect that there is not enough data for more flexible formats to converge). This corresponds to a statement of independence between random effects which may not be the ideal assumption, and with this in mind we will attempt other formats. However we do know that our photosynthetic parameters of interest ($J_{max}, V_{c_{max}}, \alpha, \theta$ and R_d) will all be represented in this aspect of the model to consider each individual plant with unique photosynthetic parameters.

2 Methods

The methods entailed hereafter are based upon those described by Yin et al. [27] except for the inclusion of mixed-effects models which is, insofar as the author is aware, only previously used in [16]. The method described in [27] uses several different datasets to estimate the parameters of interest over several steps. We acknowledge the flaw of this disjoint model which does not accumulate error but instead uses the point estimates in each succeeding step with no error accumulation. If this work is extended upon this is probably the area which most requires improvement, i.e. conversion to a full joint model. We suspect that this should also enable ease of convergence as the point estimates used in any given step might be wrong and restrict the current estimate from moving to the correct subspace of the possible solution space.

We have five datasets available for this method. For each dataset we have the same 12 individual plants present.

1. *LRC1*: Repeated measurements for plants acclimatised to increasing values of light intensity (I) at constant intercellular CO_2 (C_i) and atmospheric (i.e. 210 µmol) O_2;
2. *LRC2*: Repeated measurements for plants acclimatised to increasing values of I at constant CO_2 and low (i.e. 20 µmol) O_2;
3. *LRC2-F*: as LRC2 with the addition of measurements taken using fluorescence;
4. *ACI1*: Repeated measurements for plants acclimatised to increasing values of C_i at constant I and atmospheric O_2; and
5. *ACI2*: Repeated measurements for plants acclimatised to increasing values of C_i at constant I and low O_2.

The method is to calculate or estimate various parameters in individual steps.

1. Calculate lump parameter S and R_d using the datasets LRC2 and LRC2-F;
2. Using the results of the previous step and LRC2-F, estimate J for the low light data;
3. Calculate the relative CO_2/O_2 specificity factor for Rubisco, $S_{c/o}$, using the ACI1 and ACI2 datasets by (21);
4. Using the value of $S_{c/o}$ from the previous step, calculate Γ_* by (22);
5. Calculate g_m using (15);
6. Calculate C_c values for each observation via (14); and
7. Using all of the values previously estimated, estimate $V_{c_{max}}$ and J_{max} using the ACI1 and LRC1 datasets and the model described in 1.3.

2.1 Estimating R_d and s

We use the relationship proposed by Yin et al. [27] allowing for individual values for each plant. This regresses A on $\frac{1}{4}\phi I$ using a linear mixed effect (LME) model. This combines information from LRC2 and LRC2F (the ϕ parameter is measured used flourescence, hence the need for LRC2F). The intercept of the model is R_d and the slope is s. It is assumed that R_d and s are constant for each plant regardless of conditions and can thus be used in all datasets for each plant.

2.2 Estimating J in low-light conditions

J can be estimated in low light conditions for the linear part of the LCR curve. The calculation is:

$$J = sI\phi_{PSII} \tag{43}$$

3 Results

3.1 R_d and s

The estimates of R_d and s are shown in figures 1 and 2 respectively. We notice that the values estimated are generally positive, which is required for these parameters. However, the standard error (s.e.) on the estimate of R_d does include negative values. We notice that the estimate of R_d is not precise in comparison to s (the point estimates are on the same order of magnitude, but the associated errors are not).

Side	Treatment	R_d	R_d s.e	s	s s.e
AB	A1	0.09	0.20	0.36	0.02
AD	A1	0.10	0.26	0.46	0.03
AB	A2	0.30	0.40	0.38	0.04
AD	A2	0.38	0.20	0.40	0.04

Table 2: Estimate of parameters, R_d and s

Only consider the linear part of LRC2 and LRC2F we were obliged to include only points for $I \in [30, 100]$ (see Appendix A). This reduced the dataset for each individual to three points as shown in figures 3 and 4.

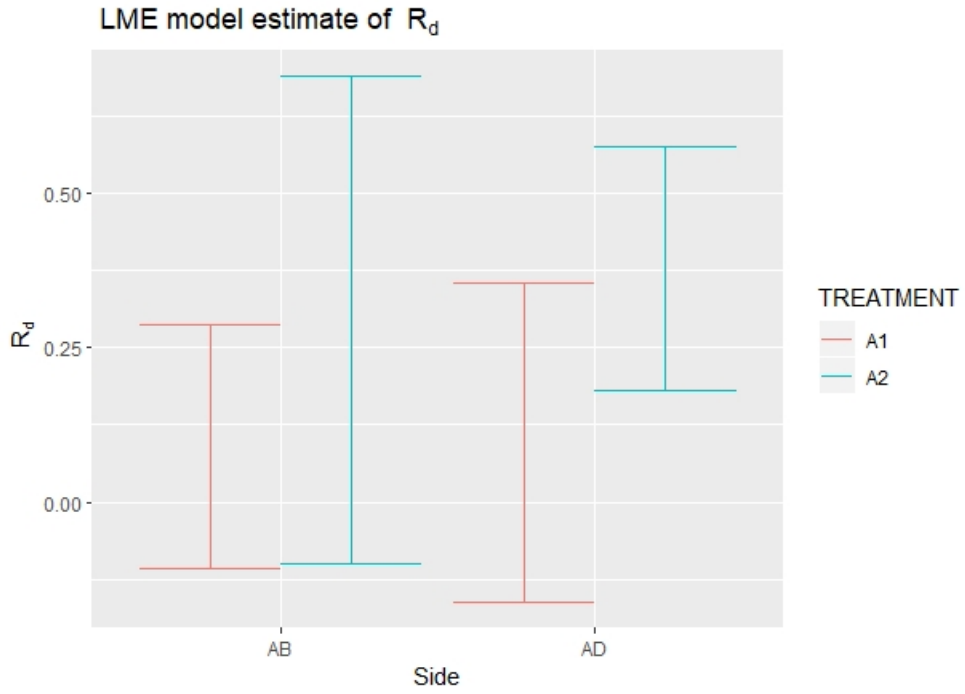

Figure 1: LME model estimates of R_d for treatment and side effects.

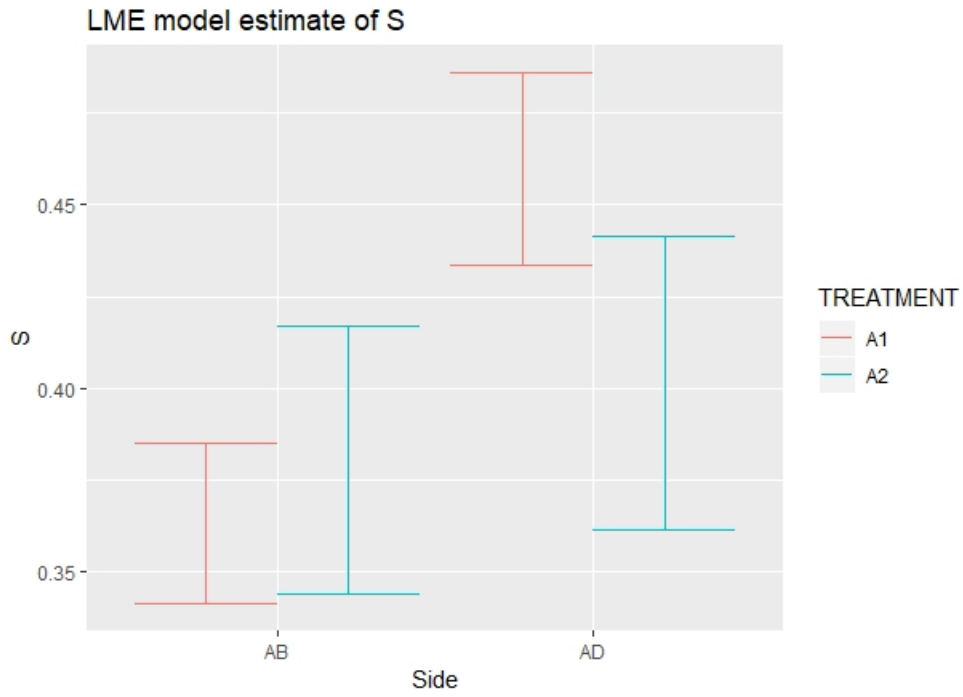

Figure 2: LME model estimates of s for treatment and side effects.

3.2 $S_{c/o}$

We include boxplots showing the range of values calculated for $S_{c/o}$ across different side and treatment effects for each individual in figures 5 and 6. The median values are given in table 3.

Side	Treatment	$S_{c/o}$
AB	A1	1.46
AD	A1	1.71
AB	A2	1.45
AD	A2	1.60

Table 3: Median estimated values of $S_{c/o}$ for each side / treatment combination

3.3 g_m

Using the previous results g_m is estimated. We use a restricted range of $C_i \in [250, 300]$ based on correspondence with Elias Konig. This gives a median estimate of g_m of 0.03.

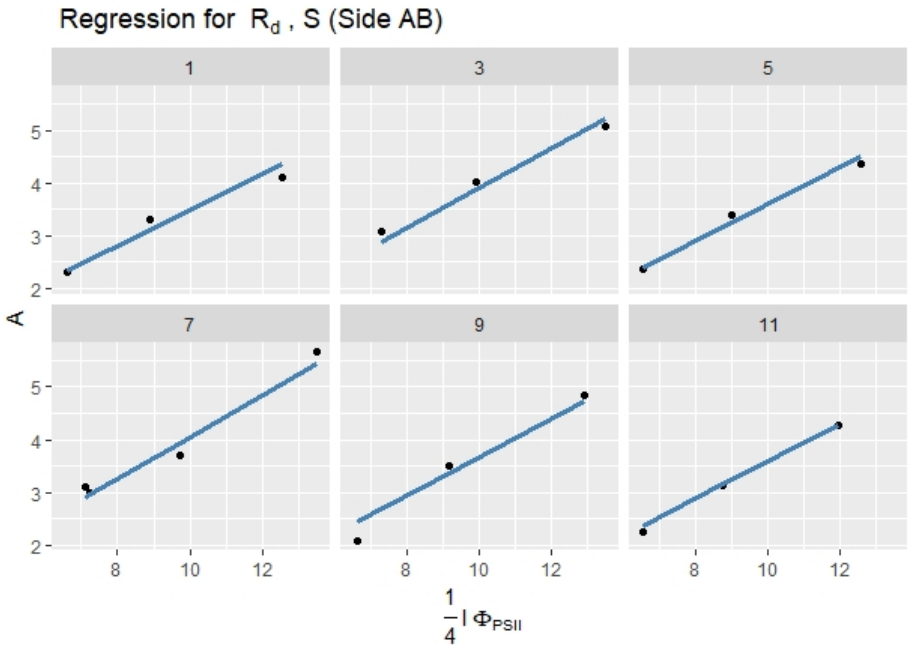

Figure 3: Model fit for LME model regressing A on $\frac{1}{4}I\phi_{PSII}$ for side AB and treatment A1 faceted by individual plant ID.

3.4 Non linear fixed effects model

We attempt to fit a non-linear fixed effect model using the values estimated and allowing $V_{c_{max}}$, J_{max} and θ to vary. Our results for the combinations of side and treatment (A1, AB) and (A2, AD) are found in table 4; other side and treatment combinations could not converge.

Side	Treatment	$V_{c_{max}}$	$V_{c_{max}}$ s.e.	J_{max}	J_{max} s.e.	θ	θ s.e
AB	A1	53.54	1.94	166.153	212.99	-3.75	12.23
AD	A2	44.38	2.90	192.00	300.98	-9.25	26.29

Table 4: Estimate of parameters, R_d and s, using fixed effect models

3.5 Non-linear mixed effects model

Unfortunately no on-linear mixed effects models could converge, so we have no results for this section.

4 Conclusion and recommendations

If we consider the results here, the values of $S_{c/o}$ are approximately half the expected value based upon the literature, g_m is an order of magnitude smaller than what we expect to see. This suggests either our analysis is flawed; we expect that this partially why no nlme model would converge, giving no esitmates of the value for $V_{c_{max}}$ or J_{max}. Based on figure 1, R_d is more strongly affected

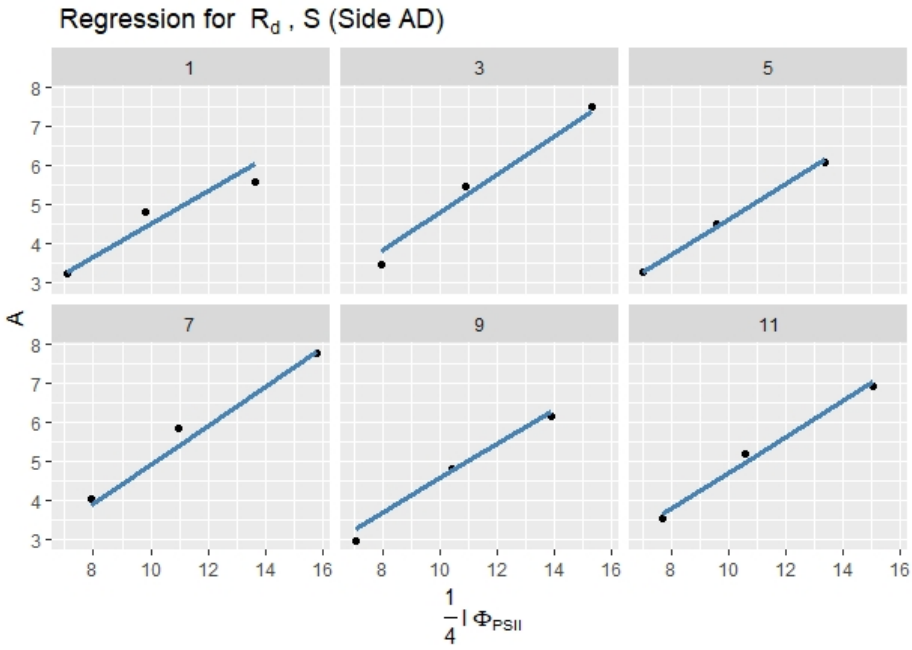

Figure 4: Model fit for LME model regressing A on $\frac{1}{4}I\phi_{PSII}$ for side AD and treatment A1 faceted by individual plant ID.

by side than treatment, as is s (see table 2 and figure 2) and $S_{c/o}$ (see table 3 and figures 5 and 6). However, as our results are strange and limited we have not implemented a full analysis.

The main realisation of this project is that there is a severe lack of transparency in the field of plant physiology. Different modelling techniques and underlying, unstated assumptions mean that many papers cannot be directly compared in their results. As a small example of unstated assumptions, no paper we read explicitly stated accounting for the Kok effect, however it appears to be standard practice based on correspondence and conversation with plant scientists. This is a serious flaw, as many models use the same name (i.e. the FvCB model), but actually refer to different versions. Beyond this all examples we could find in the literature use models that do not account for correlation within the data despite the use of repeated measurements (another unwritten assumption). This, coupled with a lack of reporting of errors on parameter estimates, leaves many published results highly opaque. Consider the table 4. Here the point estimates of $V_{c_{max}}$ and J_{max} appear sensible in comparison with the published literature. However, the error on J_{max} is so large as to disfigure the result entirely. Furthermore, θ is negative (which is not expected) and also has a large error. If this result is not unique to this paper, many published results become questionable. Without explicit statement of the error on these estimates it is hard to comment on them.

With so many parameters, with so many methods available (either for estimating photosynthetic parameters as mentioned above or in estimating the Michaelis-Menten constants) and an understanding that dropping certain data points is dependent solely upon the modeler's discretion (which points are under the Kok effect, which points are part of the linear part of the ACI and LCR curves) it leaves the final results difficult to interpret. Arguably there is sufficient room to play with choices of relevant data, models used, auxiliary parameters to ensure that the point estimates found for the key parameters are within a reasonable distance of those previously published. We do not know if this flexibility is taken advantage of to produce results that are "sensible", but it does create difficulty in reproducing results or recreating a given method. We recommend that plant scientists be explicit in

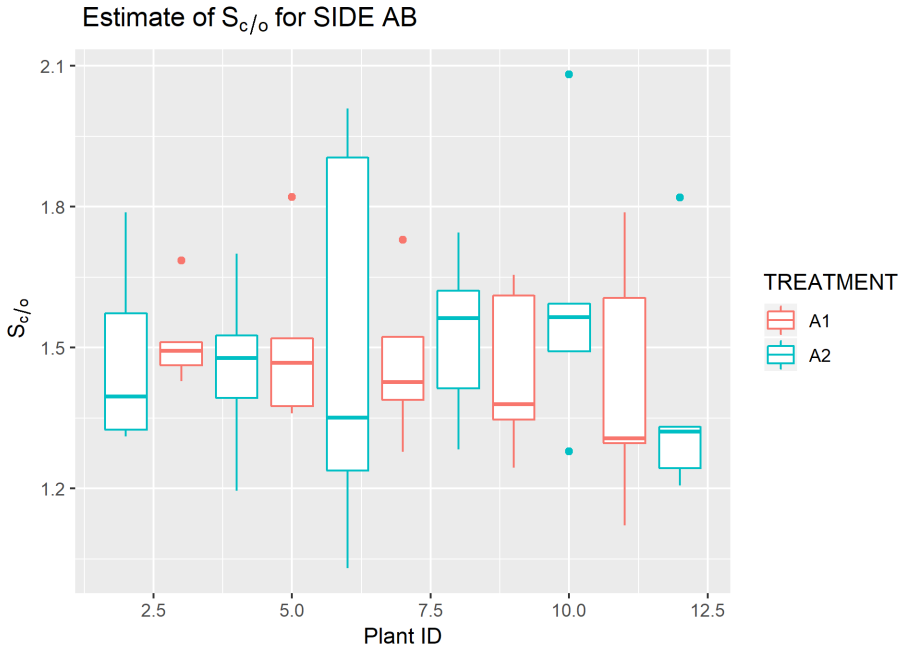

Figure 5: Range of values for $S_{c/o}$ for side measurement of AB.

their modelling assumptions, in their data selection and in their estimated error values. A thorough review paper comparing the performance of different variations of the FvCB model, and within each model comparing the use of mixed effect and fixed effect models could be of great interest. If such a paper included error estimations and an explicit statement of the assumptions and data selection process for each model, it could be of great benefit.

As there is a lack of sufficient data and there exists strong domain knowledge, we also recommend use of Bayesian methods. Using a model such as that described by Yin et al. [27] also allows one to use the initial steps to produce priors on the parameters such as s and R_d while allowing the model to fit all variables simultaneously as in the work published by Qian et al. [16]. This would be advantageous as there is greater flexibility in fitting all of the parameters simultaneously and also the error associated with each parameters is present in the final model. Furthermore, the Bayesian methods allow principled quantification of the uncertainty associated with each parameter, so not only is the credible range on the parameter included, but we have the full posterior distribution.

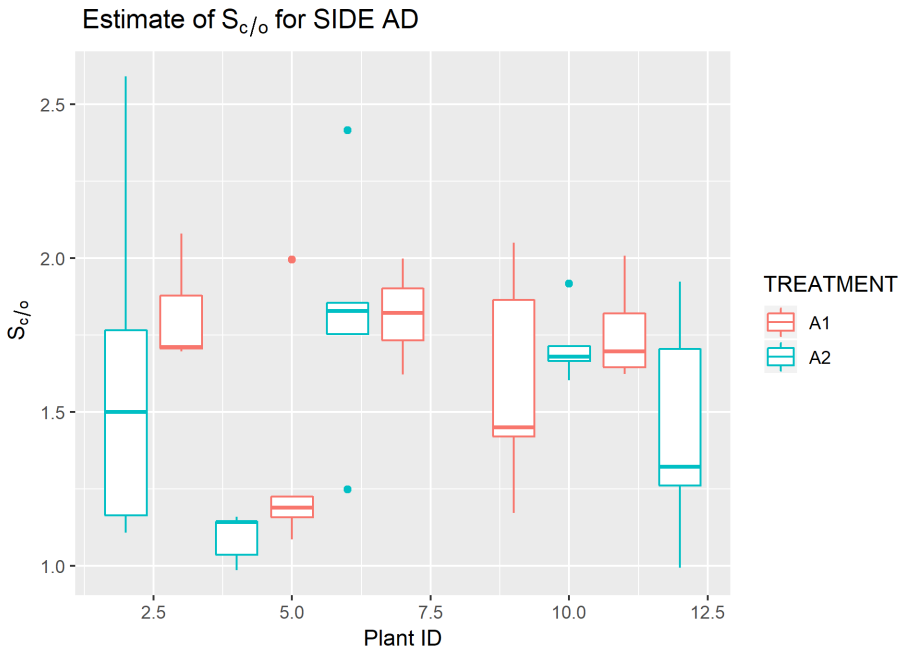

Figure 6: Range of values for $S_{c/o}$ for side measurement of AD.

A Linear light curves

We include plots of the light response curves to show which points are relevant to the linear part of the curve and not influenced by the Kok effect.

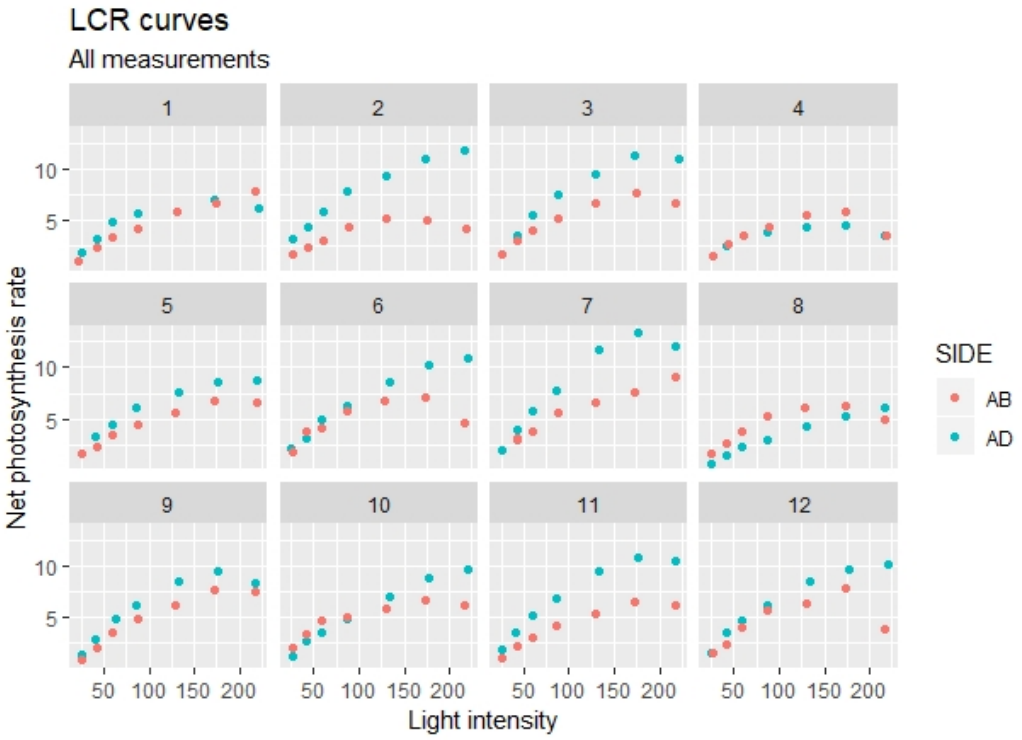
LCR curves
All measurements
1
2
3
4
5
6
7
8
9
10
11
12
10
5
Net photosynthesis rate
50 100 150 200
Light intensity
SIDE
AB
AD

Figure 7: LCR.

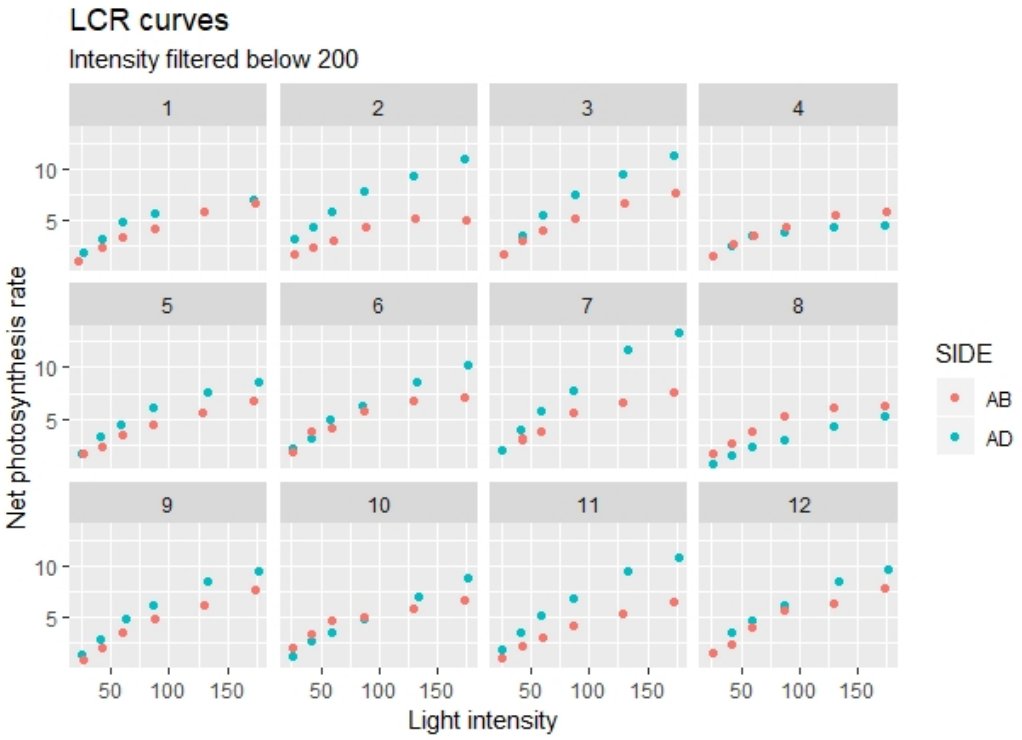

Figure 8: LCR.

References

[1] Douglas M Bates and Jose C Pinheiro. Computational Methods for Multilevel Modelling. page 30.

[2] Chandra Bellasio, David J Beerling, and Howard Griffiths. An Excel tool for deriving key photosynthetic parameters from combined gas exchange and chlorophyll fluorescence: Theory and practice: Descriptive modelling of gas exchange data. *Plant, Cell & Environment*, 39(6):1180–1197, June 2016. ISSN 01407791. doi: 10.1111/pce.12560.

[3] Marie Davidian and David M. Giltinan. Nonlinear models for repeated measurement data: An overview and update. *Journal of Agricultural, Biological, and Environmental Statistics*, 8(4):387–419, December 2003. ISSN 1085-7117, 1537-2693. doi: 10.1198/1085711032697.

[4] Remko A. Duursma. Plantecophys - An R Package for Analysing and Modelling Leaf Gas Exchange Data. *PLOS ONE*, 10(11):e0143346, November 2015. ISSN 1932-6203. doi: 10.1371/journal.pone.0143346.

[5] J. R. Evans and S. von Caemmerer. Carbon Dioxide Diffusion inside Leaves. *Plant Physiology*, 110(2):339–346, February 1996. ISSN 0032-0889, 1532-2548. doi: 10.1104/pp.110.2.339.

[6] John R Evans. Photosynthesis and nitrogen relationships in leaves of Ca plants. *Oceologica*, 78 (1):9–19, 1989.

[7] Jr Evans. The Dependence of Quantum Yield on Wavelength and Growth Irradiance. *Australian Journal of Plant Physiology*, 14(1):69, 1987. ISSN 0310-7841. doi: 10.1071/PP9870069.

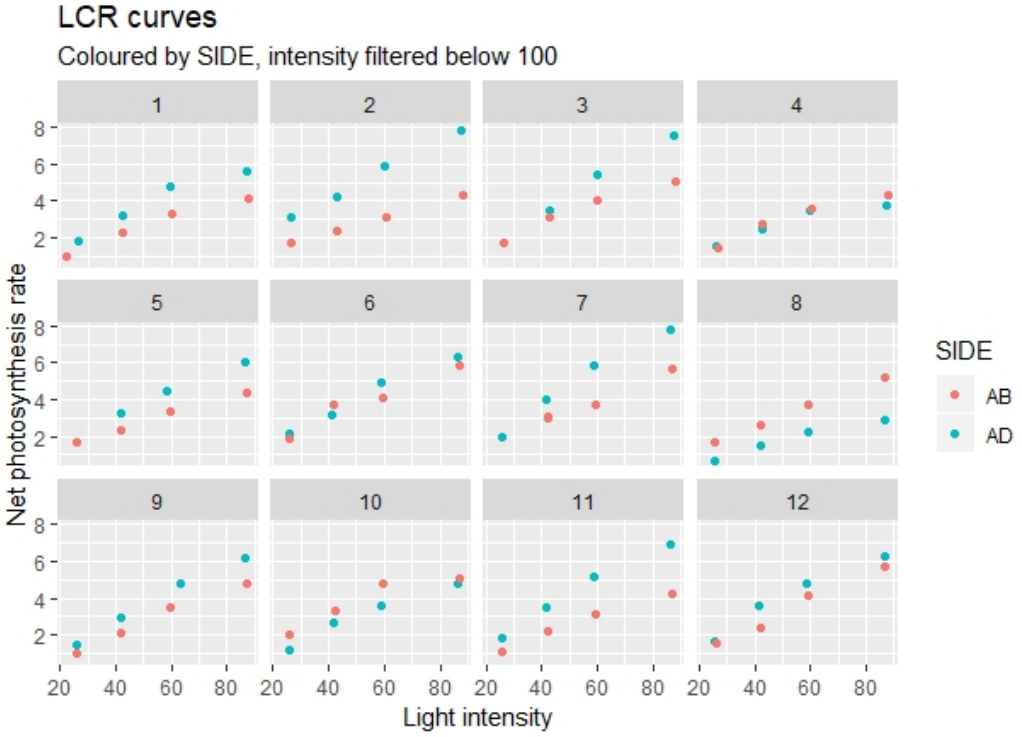

Figure 9: LCR.

[8] G. D. Farquhar, S. von Caemmerer, and J. A. Berry. A biochemical model of photosynthetic CO2 assimilation in leaves of C3 species. *Planta*, 149(1):78–90, June 1980. ISSN 0032-0935, 1432-2048. doi: 10.1007/BF00386231.

[9] Graham D. Farquhar. Models describing the kinetics of ribulose biphosphate carboxylase-oxygenase. *Archives of Biochemistry and Biophysics*, 193(2):456–468, April 1979. ISSN 00039861. doi: 10.1016/0003-9861(79)90052-3.

[10] R.F Grant. Test of a simple biochemical model for photosynthesis of maize and soybean leaves. *Agricultural and Forest Meteorology*, 48(1-2):59–74, October 1989. ISSN 01681923. doi: 10.1016/0168-1923(89)90007-5.

[11] P. C. Harley, F. Loreto, G. Di Marco, and T. D. Sharkey. Theoretical Considerations when Estimating the Mesophyll Conductance to CO2 Flux by Analysis of the Response of Photosynthesis to CO2. *PLANT PHYSIOLOGY*, 98(4):1429–1436, April 1992. ISSN 0032-0889, 1532-2548. doi: 10.1104/pp.98.4.1429.

[12] Peter C Harley and Thomas D Sharkey. An improved model of C3 photosynthesis at high CO2: Reversed 02 sensitivity explained by lack of glycerate reentry into the chloroplast. 27(169):10, 1991. doi: https://doi-org.ezproxy.library.wur.nl/10.1007/BF00035838.

[13] Mary A. Heskel, Owen K. Atkin, Matthew H. Turnbull, and Kevin L. Griffin. Bringing the Kok effect to light: A review on the integration of daytime respiration and net ecosystem exchange. *Ecosphere*, 4(8):art98, August 2013. ISSN 2150-8925. doi: 10.1890/ES13-00120.1.

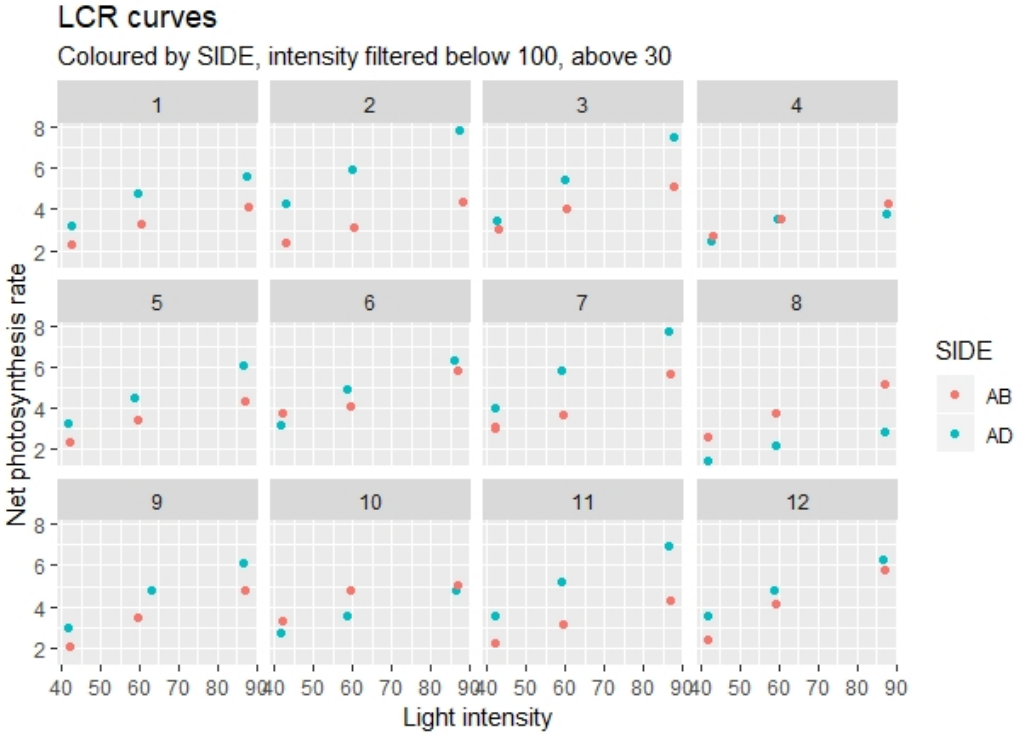

Figure 10: LCR.

[14] Dany P. Moualeu-Ngangue, Tsu-Wei Chen, and Hartmut Stützel. A new method to estimate photosynthetic parameters through net assimilation rate-intercellular space CO_2 concentration (A - C_i) curve and chlorophyll fluorescence measurements. *New Phytologist*, 213(3):1543–1554, February 2017. ISSN 0028646X. doi: 10.1111/nph.14260.

[15] Ülo Niinemets, Antonio Díaz-Espejo, Jaume Flexas, Jeroni Galmés, and Charles R. Warren. Importance of mesophyll diffusion conductance in estimation of plant photosynthesis in the field. *Journal of Experimental Botany*, 60(8):2271–2282, May 2009. ISSN 1460-2431, 0022-0957. doi: 10.1093/jxb/erp063.

[16] T. Qian, A. Elings, J.A. Dieleman, G. Gort, and L.F.M. Marcelis. Estimation of photosynthesis parameters for a modified Farquhar–von Caemmerer–Berry model using simultaneous estimation method and nonlinear mixed effects model. *Environmental and Experimental Botany*, 82: 66–73, October 2012. ISSN 00988472. doi: 10.1016/j.envexpbot.2012.03.014.

[17] Thomas D. Sharkey. Photosynthesis in Intact Leaves of C3 Plants: Physics, Physiology and Rate Limitations. *Botanical Review*, 51(1):53–105, 1985.

[18] Thomas D. Sharkey, Carl J. Bernacchi, Graham D. Farquhar, and Eric L. Singsaas. Fitting photosynthetic carbon dioxide response curves for C_3 leaves. *Plant, Cell & Environment*, 30(9): 1035–1040, September 2007. ISSN 01407791, 13653040. doi: 10.1111/j.1365-3040.2007.01710.x.

[19] R. E. Sharp, M. A. Matthews, and J. S. Boyer. Kok Effect and the Quantum Yield of Photosynthesis : Light Partially Inhibits Dark Respiration. *PLANT PHYSIOLOGY*, 75(1):95–101, May 1984. ISSN 0032-0889, 1532-2548. doi: 10.1104/pp.75.1.95.

[20] Guillaume Tcherkez, Paul Gauthier, Thomas N. Buckley, Florian A. Busch, Margaret M. Barbour, Dan Bruhn, Mary A. Heskel, Xiao Ying Gong, Kristine Crous, Kevin L. Griffin, Danielle A. Way, Matthew H. Turnbull, Mark A. Adams, Owen K. Atkin, Michael Bender, Graham D. Farquhar, and Gabriel Cornic. Tracking the origins of the Kok effect, 70 years after its discovery. *New Phytologist*, 214(2):506–510, April 2017. ISSN 0028646X. doi: 10.1111/nph.14527.

[21] Ichiro Terashima. Anatomy of non-uniform leaf photosynthesis. *Photosynthesis Research*, 31(3): 195–212, March 1992. ISSN 0166-8595, 1573-5079. doi: 10.1007/BF00035537.

[22] Susanne von Caemmerer. *Biochemical Models of Leaf Photosynthesis*. Number 2 in Techniques in Plant Science. CSIRO, Collingwood, 2000. ISBN 978-0-643-06379-2. OCLC: 247732918.

[23] Susanne Von Caemmerer. Steady-state models of photosynthesis: Steady-state models of photosynthesis. *Plant, Cell & Environment*, 36(9):1617–1630, September 2013. ISSN 01407791. doi: 10.1111/pce.12098.

[24] A. S. Walcroft, D. Whitehead, W. B. Silvester, and F. M. Kelliher. The response of photosynthetic model parameters to temperature and nitrogen concentration in Pinus radiata D. Don. *Plant, Cell and Environment*, 20(11):1338–1348, November 1997. ISSN 0140-7791, 1365-3040. doi: 10.1046/j.1365-3040.1997.d01-31.x.

[25] X. Yin and P.C. Struik. C3 and C4 photosynthesis models: An overview from the perspective of crop modelling. *NJAS - Wageningen Journal of Life Sciences*, 57(1):27–38, December 2009. ISSN 15735214. doi: 10.1016/j.njas.2009.07.001.

[26] Xinyou Yin and Paul C. Struik. Theoretical reconsiderations when estimating the mesophyll conductance to CO_2 diffusion in leaves of C_3 plants by analysis of combined gas exchange and chlorophyll fluorescence measurements. *Plant, Cell & Environment*, 32(11):1513–1524, November 2009. ISSN 01407791, 13653040. doi: 10.1111/j.1365-3040.2009.02016.x.

[27] Xinyou Yin, Paul C. Struik, Pascual Romero, Jeremy Harbinson, Jochem B. Evers, Peter E. L. Van Der Putten, and Jan Vos. 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(5):448–464, May 2009. ISSN 01407791, 13653040. doi: 10.1111/j.1365-3040.2009.01934.x.