sustainability

MDPI

Article

Parameter Estimation of the Farquhar—von Caemmerer—Berry Biochemical Model from Photosynthetic Carbon Dioxide Response Curves

Qingguo Wang [1,2], Jong Ahn Chun [2,*], David Fleisher [3], Vangimalla Reddy [3], Dennis Timlin [3] and Jonathan Resop [4]

1 Wye Research and Education Center, University of Maryland, Queenstown, MD 21658, USA; qingguo_wang@yahoo.ca
2 APEC Climate Center, 12 Centum 7-ro, Haeundae-gu, Busan 48058, Korea
3 USDA-ARS Crop Systems and Global Change Lab, Beltsville, MD 20705, USA; David.Fleisher@ARS.USDA.GOV (D.F.); Vangimalla.Reddy@ARS.USDA.GOV (V.R.); Dennis.Timlin@ARS.USDA.GOV (D.T.)
4 Department of Geographical Sciences, University of Maryland, College Park, MD 20742, USA; resop@umd.edu
* Correspondence: jachun@apcc21.org; Tel.: +82-51-745-3957

Received: 26 April 2017; Accepted: 19 July 2017; Published: 24 July 2017

Abstract: The Farquhar—von Caemmerer—Berry (FvCB) biochemical model of photosynthesis, commonly used to estimate CO_2 assimilation at various spatial scales from leaf to global, has been used to assess the impacts of climate change on crop and ecosystem productivities. However, it is widely known that the parameters in the FvCB model are difficult to accurately estimate. The objective of this study was to assess the methods of Sharkey et al. and Gu et al., which are often used to estimate the parameters of the FvCB model. We generated A_n/C_i datasets with different data accuracies, numbers of data points, and data point distributions. The results showed that neither method accurately estimated the parameters; however, Gu et al.'s approach provided slightly better estimates. Using Gu et al.'s approach and datasets with measurement errors and the same accuracy as a typical open gas exchange system (i.e., *Li-6400*), the majority of the estimated parameters—V_{cmax} (maximal Rubisco carboxylation rate), K_{co} (effective Michaelis-Menten coefficient for CO_2), g_m (internal (mesophyll) conductance to CO_2 transport) and Γ^* (chloroplastic CO_2 photocompensation point)—were underestimated, while the majority of R_d (day respiration) and α (the non-returned fraction of the glycolate carbon recycled in the photorespiratory cycle) were overestimated. The distributions of T_p (the *r*ate of triose phosphate export from the chloroplast) were evenly dispersed around the 1:1 line using both approaches. This study revealed that a high accuracy of leaf gas exchange measurements and sufficient data points are required to correctly estimate the parameters for the biochemical model. The accurate estimation of these parameters can contribute to the enhancement of food security under climate change through accurate predictions of crop and ecosystem productivities. A further study is recommended to address the question of how the measurement accuracies can be improved.

Keywords: food security; CO_2 assimilation; photosynthesis; biochemical model; climate change

1. Introduction

The FvCB leaf photosynthesis model for C_3 plants [1,2] is fundamental for the prediction of leaf responses to environmental variation [3]. This model has been widely used to simulate CO_2 assimilation and the response of plants to climate change for different spatiotemporal scales [4–11], due to its solid theoretical basis and simplicity [12]. It is also frequently used in reverse to quantify the

underlying biochemical properties (i.e., the model parameters) of leaves under different environmental conditions [13–17]. These parameters are often considered easier to estimate from gas exchange measurements rather than making the required in vitro measurements to quantify enzyme activity. This is because it is difficult to extract functional enzymes from many species [18], and in vitro conditions seldom represent those experienced in vivo [19]. According to the different versions of FvCB model [1,20–23], up to 8 parameters (V_{cmax}, K_{co}, J, T_p, α, g_i, R_d, and Γ*) can be estimated from an analysis of the response of the net assimilation rate (A_n) to intercellular CO_2 concentration (C_i) if enough accurate data points are available [2].

There are numerous publications discussing the different parameterization methods associated with A_n/C_i curves [2,12,24–29]. Each method relies on different assumptions and has technical limitations. Most approaches assume that α = 0 and that the K_{co} and Γ* can be chosen a priori from estimates in previous studies to determinate V_{cmax}, $J_{\max}$, T_p, R_d and g_m [2,12,22,26,27]. Gu et al. [2] extended the method of Ethier and Livingston [12] to propose an exhaustive dual optimization (EDO) approach to estimate the parameters from fitting A_n/C_i curves. All of the curve-fitting methods minimize an objective function; e.g., the sum of the square of errors (SSE), based on the nonlinear FvCB model, with a limited number of measurements (typically 8–12, [2,24,26,30]) and the expected accuracy of a measured A_n/C_i curve (see [2,25,26] for comprehensive reviews). Depending on the equations used for fitting the parameters, two major groups of methods can be distinguished: Group I directly fits parameters with the FvCB model [14,25–27,31] and Group II fits parameters with a quadratic equation [2,12]. The implementation is sensitive to the methods used. The estimated parameters can be substantially different when using different A_n/C_i curve-fitting methods on the same dataset [2,25,26,29]. It is frequently difficult to determine which method is superior based on the measures of SSE because of the characteristics of the FvCB model [2], the assumption that the assigned values of the kinetic properties of Rubisco are the correct ones, and the number of parameters to be estimated [24].

Some studies compared different fitting methods to determine the parameters of the FvCB model. Miao et al. [26] made a comparison of six different methods using 160 randomly selected A_n/C_i datasets from 4 shrubby indicator species, including highbush blueberry (*Vaccinium corymbosum* L.), dangleberry (*Gaylusaccia frondosa* L.), coastal fetterbush (*Eubotrys racemosa* L.) and sweet pepperbush (*Clethra alnifolia* L.). They concluded that the method developed by Sharkey et al. [27] was among the 'best', based on the lowest minimum of the root mean square error. Gu et al. [2] stated that their approach could estimate reliable FvCB parameters using error-free synthetic A_n/C_i curves and predicted limited states that matched chlorophyll fluorescence patterns from actual datasets.

An erroneous determination of the FvCB model parameters can lead to inaccurate predictions of ecosystem productivity, because the potential errors can worsen when moving to larger temporal and spatial scales (e.g., from field measurements over short periods to ecosystem predictions over long periods) [32,33]. However, to the best of our knowledge, there is little information on a comprehensive test of these different fitting methods using common generated datasets superimposed by possible measurement errors. The objective for this study was to assess the two approaches for fitting the FvCB model: group I (Sharkey et al.'s method) and group II (Gu et al.'s method).

2. Materials and Methods

2.1. The FvCB Model and Characteristics

A_n/C_i curves are fitted with the FvCB model for C_3 leaves [1] that accounts for g_m and whereby A_n is given as

$$A_n = min\{W_c, W_j, W_p\}\left(1 - \frac{\Gamma^*}{C_c}\right) - R_d, \text{ for } C_c > (1 + 3\alpha)\Gamma^* \quad (1)$$

$$A_n = min\{W_c, W_j, W_p\}\left(1 - \frac{\Gamma^*}{C_c}\right) - R_d, \text{ for } C_c > (1 + 3\alpha)\Gamma^*, \quad (2)$$

$$A_n = min\{W_c, W_j\}\left(1 - \frac{\Gamma^*}{C_c}\right) - R_d, \text{ for } \Gamma^* \leq C_c \leq (1+3\alpha)\Gamma^*, \tag{3}$$

$$A_n = W_c\left(1 - \frac{\Gamma^*}{C_c}\right) - R_d, \text{ for } C_c < \Gamma^*, \tag{4}$$

$$W_c = V_{cmax}\frac{C_c}{C_c + K_{co}}, \tag{5}$$

$$W_j = J\frac{C_c}{4C_c + 8\Gamma^*}, \tag{6}$$

where $min\{\}$ denotes 'the minimum of'.

α in Equation (6) is often set to zero, and Equation (6) can be then reduced to

$$W_p = 3T_p\frac{C_c}{C_c - \Gamma^*}, \; for \; C_c > \Gamma^*. \tag{7}$$

In the above equations, W_c, W_j and W_p are the carboxylation rates limited by Rubisco (A_c state), the Ribulose 1,5-bisphosphate (RuBP) regeneration (A_j state) and triose phosphate utilization (TPU) (A_p state), respectively; Equation (3) is one of the points where the A_j state is equal to the A_n state (see next the section for an explanation); C_c is the chloroplastic CO_2 partial pressure and can be estimated by

$$C_c = C_i - \frac{A_n}{g_m}. \tag{8}$$

A typical condition for assessing A_n/C_i curves at saturating light levels is to assume that J approaches J_{max}, the maximum rate of electron transport. If light is not saturating at the time of measurement, J_{max} must be calculated from J [24]. In this study, we assume that $J_{max} = J$.

Combining Equations (1)–(8), the relationship between A_n and C_i in the A_c, A_j and A_p states can be expressed as three segment hyperbolic curves [2,12], respectively;

$$\begin{cases} {A_{nc}}^2 - g_m C_{ic} A_{nc} + p_c A_{nc} + q_c C_{ic} + u_c = 0, \\ {A_{nj}}^2 - g_m C_{ij} A_{nj} + p_j A_{nj} + q_j C_{ij} + u_j = 0, \\ {A_{np}}^2 - g_m C_{ip} A_{np} + p_p A_{np} + q_p C_{ip} + u_p = 0, \end{cases} \tag{9}$$

where the subscripts c, j and p refer to the A_c, A_j and A_p states, respectively; g_m, p, q and u are 4 'coefficients' in each segment of an A_n/C_i curve. The general form of Equation (9) is

$${A_n}^2 - g_m C_i A_n + pA_n + qC_i + u = 0 \tag{10}$$

In the A_c state,

$$\begin{cases} g_m = g_m, \\ p_c = -[(V_{cmax} - R_d) + g_m K_{co}], \\ q_c = (V_{cmax} - R_d\ )g_m, \\ u_c = -[V_{cmax}\Gamma^* + K_{co}R_d]g_m. \end{cases} \tag{11}$$

In the A_j state,

$$\begin{cases} g_m = g_m, \\ p_j = -\left[\left(\frac{J_{max}}{4} - R_d\right) + 2\Gamma^* g_m\right], \\ q_j = \left(\frac{J_{max}}{4} - R_d\right)g_m, \\ u_j = -\left(\frac{J_{max}}{4} + 2R_d\right)\Gamma^* g_m. \end{cases} \tag{12}$$

In the A_p state,

$$\begin{cases} g_m = g_m, \\ p_p = -\left[\left(3T_p - R_d\right) - (1+3\alpha)\Gamma^* g_m\right], \\ q_p = \left(3T_p - R_d\right) g_m, \\ u_p = -\left[3T_p - (1+3\alpha)R_d\right]\Gamma^* g_m. \end{cases} \tag{13}$$

When $\alpha = 0$, Equation (9) in the A_p state is reduced to

$$A_p = 3T_p - R_d \tag{14}$$

Mathematically, there are two positive roots for the quadratic Equation (10),

$$A_n = \frac{g_m C_i - p \pm \sqrt{(g_m C_i - p)^2 - 4(u - 4qC_i)}}{2}, \tag{15}$$

but only one root meets the constraints of the FvCB model.

$$A_n = \frac{g_m C_i - p - \sqrt{(g_m C_i - p)^2 - 4u + 4qC_i}}{2}. \tag{16}$$

The above equation describes A_n as a function, C_i [12]. Equation (16) consists of up to three segments (states). The three states share one common coefficient g_m, while the other coefficients (p, q and u) are combinations of the common parameters g_m, Γ^* and R_d and the state specific parameters V_{cmax} and K_{co} in the A_c state, J_{max} in the A_j state, and T_p and α in the A_p state.

2.2. Data Generation

2.2.1. Constraints for the Parameters

To generate datasets in order to test the methods, the following constraints for the parameters are required. All parameters are greater than zero to keep their biological meanings. The FvCB model is a monotonic increasing function in both A_c and A_j states and a monotonically decreasing function when $\alpha > 0$, or it is a constant in the A_p state when $\alpha = 0$. The three states follow a fixed order along the C_i axis [2].

If an A_n/C_c curve consists of both A_c and A_j states, there could, mathematically, be two conditions: (1) The two states could be exactly the same if $J_{max}/4 = V_{cmax}$ and $K_{co} = 2\Gamma^*$. Since $K_{co} > 2\Gamma^*$ [2,22], however, the two states cannot be the same. (2) There are two intersection points where $A_{nc} = A_{nj}$. The first point is at $C_c = \Gamma^*$; $A_{nj} < A_{nc}$ when $C_c < \Gamma^*$. We define $A_n = A_{nc}$ when $C_c < \Gamma^*$ to create the fixed order of A_c and A_j states (Equation (3)). The second point is defined as the transition point C_{cc_cj} by combining Equations (4) and (5),

$$C_{cc\_cj} = \frac{J_{max} - 4V_{cmax}2\Gamma^*/K_{co}}{4V_{cmax} - J_{max}}. \tag{17}$$

Since $2\Gamma^*/K_{co} < 1$, the constraints for V_{cmax} and J_{max} can be expressed as:

$$4V_{cmax} > J_{max}. \tag{18}$$

When A_j and A_p states coexist, the constraint for J_{max} as a function of T_p is [2]:

$$J_{max} > 12T_p. \tag{19}$$

When A_c and A_p states coexist, the constraint for V_{cmax} as a function of T_p is

$$V_{cmax} > 3T_p. \tag{20}$$

When three states coexist, the constrains are given by

$$\begin{cases} V_{cmax},\ J_{max}, T_p, K_{co},\ \Gamma^*, R_d,\ g_m, \alpha > 0, \\ K_{co} > 2\Gamma^*, \\ C_c \geq (1+3\alpha)\Gamma^*, \text{ in } \mathrm{A_p} \text{ state}, \\ A_n = A_{nc} \text{ when } C_c \leq \Gamma^*, \\ 4V_{cmax} > J_{max} > 12T_p. \end{cases} \tag{21}$$

2.2.2. Criteria for Deriving Parameters

An involved parameter is a parameter that is presented in the given dataset. A non-resolvable parameter is defined as an involved parameter whose value cannot be correctly derived. It should be noted that resolvable, non-resolvable, and noninvolved parameters were defined by Gu et al. [2]. More detailed information on these definitions can be found in Gu et al. [2]. The equality of an estimated parameter against its "true value" is defined by an equivalency at the accuracy of three decimal places if the lowest precision of the A_n/C_i dataset is three or higher, or at the lowest precision of the A_n/C_i dataset if the precision of A_n/C_i is less than three decimal places. Since there are limitations due to measurement accuracy, the number of data points of the A_n/C_i dataset, and the fitting methods used, an estimated parameter might be equal to its "true value" by chance. To stop this result from biasing our analysis, a correctly estimated parameter is defined here as one that is obtained only if all resolvable parameters in this dataset are equal to their "true values". If the numbers of data points in A_c, A_j and A_p states are x, y and z, respectively, the data point distribution is written as (x, y, z). More detailed information on the theoretical resolvability of parameters can be found in Gu et al. [2].

In practice, all of the measurements are subject to measurement errors (Appendix A). Therefore, the measured points lie near to, but not on, the theoretical curve. A number of techniques are needed to obtain the optimal approximation of the "true value". The common methods for estimating the parameters of the FvCB model are based on the minimization of an objective function which characterizes a goodness-of-fit of a particular curve with respect to the given set of data points. For example, the parameters are estimated by minimizing the SSE [2,29]:

$$f = \sum_{i=1}^{nc}(A_{ci} - A_{cmi})^2 + \sum_{j=1}^{nj}(A_{jj} - A_{jmj})^2 + \sum_{k=1}^{np}\left(A_{pk} - A_{pmk}\right)^2, \tag{22}$$

where A_{ci}, A_{jj} and A_{pk} are the calculated net assimilation rates at point i in the A_c, A_j, A_p states respectively; A_{cmi}, A_{jmj} and A_{pmk} are the measured counterparts, respectively; and subscripts nc, nj and np are the numbers of counterpart points, respectively. The objective equation used in the methods of Sharkey et al. and Gu et al. are similar to Equation (22); however, the calculation procedures are different.

2.2.3. Generation of Datasets

The parameters were randomly chosen from selected parameter ranges with three decimal places. V_{max} varied from 20.000 to 160.000 μmol m^{-2} s^{-1}, J_{max} from 20.000 to 250.000 μmol m^{-2} s^{-1}, T_p from 5.000 to 15.500 μmol m^{-2} s^{-1}, R_d from 0.010 to 5.000 μmol m^{-2} s^{-1}, g_m from 0.100 to 30.000 μmol m^{-2} s^{-1} Pa^{-1}, Γ^* from 0.100 to 5.000 Pa, K_{co} from 20.000 to 100.000 Pa, and α from 0.001 to 1.000. As a special case, α = 0, with the requirement of the constraints mentioned in the previous section (Equation (21)). C_i ranges from 5 to 150 Pa in both cases. We assumed that the datasets were collected at a leaf temperature of 25 °C and an air pressure of 100 Pa. Each set of parameters (except for α) was used to generate datasets with either α = 0 or α > 0. There were 200 datasets within each accuracy level, of which 100 used α = 0 and 100 with α > 0.

There were three accuracy subgroups (high, normal, and varied). Firstly, a high accuracy implies the accuracy of the generated A_n and C_i to eight decimal places. Secondly, a normal accuracy dataset was defined as a dataset which is rounded off from a high accuracy dataset. Normal precision is the same as a typical open gas exchange system, e.g., *Li-6400* (Li-Cor, Inc., Lincoln, NE, USA). In this case, A_n is rounded to three decimal places if A_n < 1.000 μmol m^{-2}s^{-1}, to two decimal places if 1.00 μmol s^{-1} m^{-1} < A_n < 10.00 μmol s^{-1} m^{-1}, and to one decimal place if A_n > 10.0 μmol m^{-2} s^{-1}; C_i is rounded to one decimal place if C_i < 100 μmol mol^{-1} and to an integer if C_i > 100 μmol mol^{-1}. Datasets can commonly have measurement errors. The errors in A_n and C_i were calculated according to Equations (A5) and (A6), respectively, in the appendix. The precision of a dataset from the measurements is the same as a normal accuracy dataset. Varied accuracy datasets were generated without measurement errors, either with varied accuracy or varied data points. The parameters used to generate varied accuracy datasets were: V_{cmax} of 45.572 μmol m^{-2} s^{-1}, J_{max} of 111.315 μmol m^{-2} s^{-1}, T_p of 7.871 μmol m^{-2} s^{-1}, R_d of 1.381 μmol m^{-2} s^{-1}, g_m of 1.897 μmol m^{-2} s^{-1} Pa^{-1}, Γ^* of 4.396 Pa, K_{co} of 43.616 Pa, and α of 0.352. The maximum number of data points in each state was nine. There were two subgroups in this group:

(i) Datasets with varied data points. The accuracy of this dataset was eight decimal places; and the numbers of data points were 4, 5, 6, 7, 8, 9, and 12. The varied data point dataset was used to evaluate the impact of the number of data points on parameterization. It should be noted that these datasets are included in the high accuracy dataset.
(ii) Datasets with varied accuracy. These datasets were with either eight or 12 data points, and accuracies were from one to eight decimal places. The varied accuracy dataset was used to identify the impact of accuracy on parameterization.

2.3. Fitting Methods

2.3.1. Gu et al.'s Method

Gu et al. [2] developed a four-step method to estimate the parameters of the FvCB model.

(1) *The enumeration of all possible data point distributions of three states of a given dataset.* The three limited states must follow a certain pattern along the C_i axis in an order dictated by the FvCB model. The minimum numbers of data points (3 or 0, 2 or 0, 3 or 0) and the number of data points higher than the number of parameters to be derived are required for resolvable parameters. Under these conditions, the resolvable parameters are defined as Gu et al.'s resolvable parameters to differentiate them from resolvable parameters as a general case. Thus, the minimum numbers of data points (3, 2, 3) and a minimum number of nine observed data points are required for all eight parameters to be resolvable. We refer to these as Gu et al.'s requirements for all eight resolvable parameters. If only one or two states are Gu et al.'s resolvable states, the dataset is Gu et al.'s partially resolvable dataset and the resolvable parameters are Gu et al.'s partially resolvable parameters. If a state does not meet the minimum data point requirement of Gu et al., Gu et al.'s method forces the dataset to meet the requirements by moving data points from one state to another. If the number of observed data points is zero in the A_p state, α = 0 and T_p = (asymptote of A_n + R_d)/three.
(2) *Fitting the FvCB model to each limited state distribution separately.* In this step, the transition points are never calculated and the carboxylation rates in different states are never compared. The A_n is calculated with the submodel of the limited state to which the data point is assigned.
(3) *Detection and correction of inadmissible fits.* Gu et al., [2] defined "inadmissible fits" as cases where the limitation states of the points in the A/C_i curve have not consistently or correctly identified the derived parameters. This step is only used for a dataset that contains multiple limited states. If the calculated limited state distribution is the same as the assigned limited state distribution,

then the fit is admissible; otherwise, the fit is inadmissible. If the fit is inadmissible, the fit will be corrected via a penalization strategy.

(4) *Section of best fit.* The best fit for an observed set of data is the method that gives the smallest value for the minimized objective function. If the values of the minimized objective function are equal when comparing across different limited state distributions, the one with fewer parameters is selected.

2.3.2. Sharkey et al.'s Method

Sharkey et al.'s method requires an initial set of estimated parameter values to be assigned and iteratively improves this set. The algorithm starts with two initial transition values for A_c and A_j states and A_j and A_p states; it then changes the values until the objective function is minimized. Since the parameters K_{co}, α and Γ^* are assigned a priori, a maximum of five parameters are estimated. The three states share common parameters—g_m and R_d —with state specific parameters of V_{cmax} in the A_c state, J_{max} in the A_j state, and T_p in the A_p state. Thus, the minimum data point distribution is (1, 1, 1) and the minimum number of points is five for all five parameters to be resolvable. If the data points are only distributed in one state or in two states, the minimum number of data points is three or four, respectively, for all involved parameters to be resolvable. It should be noted that, for the same dataset, the number of Gu et al.'s resolvable parameters is different from that of Sharkey et al.'s, and both are different from the resolvable parameters in the general case.

2.3.3. Parameter Calculations

The parameter estimations were conducted in May 2011 to March 2012. The generated datasets were uploaded to the website http://Leafweb.ornl.gov to estimate the parameters by Gu et al.'s method. The detailed procedures can be found on the website. The Excel spreadsheet-software created by Sharkey et al. [27] (A_n/C_i curve fitting utility version 1.1) was used to test Sharkey et al.'s method.

3. Results

3.1. High Accuracy Dataset

Using Gu et al.'s method, all eight parameters were correctly retrieved in 23 datasets for high accuracy datasets with $\alpha > 0$. In fact, there were a total of 33 datasets with all eight resolvable parameters, in which ten of the data sets did not meet the requirements set by Gu et al. for the minimum data point distribution (3, 2, 3) to retrieve all eight parameters. For datasets with $\alpha = 0$, all eight parameters were correctly derived in 24 out of a total of 25 datasets that met the requirements of Gu et al. for all eight resolvable parameters. There was one exception, where a (3, 2, 10) dataset resulted in poor parameter estimation. These results imply that, when the dataset met Gu et al.'s requirements for all resolvable parameters in high accuracy datasets, Gu et al.'s method obtained a full set of parameters for a dataset with $\alpha > 0$, while the method might not correctly estimate all eight parameters for a dataset with $\alpha = 0$.

For datasets with Gu et al.'s partially solvable parameters, the resolvable parameters may or may not be correctly estimated. For example, two datasets had the same data distribution (9, 0, 6). One dataset was able to correctly estimate all Gu et al.'s partially resolvable parameters, but another could not correctly estimate any parameter. T_p and α were non-involved parameters in this dataset, where the missing A_p state was incorrectly estimated because T_p was calculated by fitting the A_n/C_i curve with a sigmoid function and fixed $\alpha = 0$. For example, one dataset (0, 15, 0) was identified as (5, 10, 0), and the non-involved parameters V_{cmax}, K_{co}, T_p and α were incorrectly derived.

For a dataset with one or two states which did not meet the minimum data point requirements set by Gu et al. for resolvable parameters, the specific parameters in this state could not be correctly estimated. For example, the method forced a dataset distribution of (7, 6, 2) to (7, 5, 3), leading to an incorrect estimation of all eight parameters.

The numbers of estimated parameters were larger than the resolvable parameters, which in turn were larger than the correctly estimated parameters. For example, for V_{cmax}, J_{max} and T_p, when $\alpha > 0$, the numbers of estimated parameters were 91, 74 and 100; the numbers of resolvable parameters were 71, 72 and 61; and the numbers of correctly estimated parameters were 52, 49, and 42, respectively (Table 1); when $\alpha = 0$, the total numbers parameters were 86, 92 and 100 for estimated, 44, 62 and 45 for resolvable, and 37, 46 and 30 for correctly estimated parameters, respectively (Table 1). Obviously, g_m was a resolvable parameter in any dataset, but could not always be correctly estimated.

Using Gu et al.'s method, more than half of the values of V_{cmax}, K_{co}, J_{max}, g_m, T_p and Γ^* ranged within ±10% of error (Table 1). Some estimated values of g_m could be very large; up to 100,000 μmol m^{-2} s^{-1} Pa^{-1} (Figure 1E). The estimated R_d values were more variable in comparison with other parameters, and many of the values were larger than the upper limit of 5.000 μmol m^{-2} s^{-1} used to generate datasets (Figure 1D). Most estimates of α were zero or very close to their "true values" when $\alpha > 0$ (Figure 1H). The incorrectly estimated values of V_{cmax}, K_{co}, J_{max}, Γ^* and α showed a tendency to be underestimated, while R_d tended to be overestimated. About half of the incorrectly estimated g_m values (except for extreme values) were overestimated, while T_p was evenly distributed around the 1:1 line ($y = 1.002x$, $R^2 = 0.941$). The uneven distributions of the estimated parameters V_{cmax}, K_{co}, Γ^*, R_d and α around the 1:1 line implied that the averages of these parameters may not be close to their "true values".

Table 1. Summary of the parameterization of Gu et al.'s and Sharkey et al.'s methods using 15 data points A_n/C_i curves. Resolvable: the parameter can be correctly estimated by an appropriate method; Correctly estimated: the estimated parameter by a specific method with the same value as "true value"; Total estimated: total estimated parameters including correctly and incorrectly estimated parameters; Error within ±10%: the ranges of the error of estimated parameters within ±10% of the "true value".

α	Dataset	Method	Number of Parameters	g_m	Γ^*	R_d	V_{cmax}	K_{co}	J_{max}	T_p	α
$\alpha > 0$	HDS	Gu et al.	Resolvable [a]	100	90	90	71	80	72	61	61
			Correctly estimated	72	62	62	52	62	49	42	42
			Total estimated	100	100	100	91	91	74	100	100
			Error within ±10%	89	80	75	71	70	69	89	55
		Sharkey et al.	Resolvable [a]	100	NA	100	88	NA	79	72	NA
			Correctly estimated [b]	0	NA	0	0	NA	0	0	NA
			Total estimated [d]	100	NA	100	100	NA	100	100	NA
			Error within ±10%	41	NA	6	22	NA	65	7	NA
	NDS	Gu et al.	Correctly estimated [e]	0	0	0	0	0	0	0	0
			Total estimated	100	100	100	98	98	76	100	100
			Error within ±10%	26	48	16	55	40	65	70	26
		Sharkey et al.	Error within ±10%	6	NA	6	24	NA	70	57	NA
	DSE	Gu et al.	Total estimated	100	100	100	98	98	86	100	100
			Error within ±10%	7	36	7	30	10	61	64	23
		Sharkey et al.	Error within ±10%	8	NA	5	24	NA	70	60	NA
$\alpha = 0$	HDS	Gu et al.	Resolvable [a]	98	62	62	44	72	62	45	45
			Correctly estimated	72	46	46	37	63	46	30	30
			Total estimated	100	100	100	86	86	92	100	100
			Error within ±10%	82	61	57	64	65	68	74	100
		Sharkey et al.	Resolvable [a]	100	NA	100	88	NA	95	79	NA
			Error within ±10%	63	NA	5	23	NA	64	8	NA
	NDS	Gu et al.	Total estimated	100	100	100	93	93	90	100	100
			Error within ±10%	20	42	12	43	27	64	67	99
		Sharkey et al.	Error within ±10%	8	NA	8	22	NA	59	51	NA
	DSE	Gu et al.	Total estimated	100	100	100	98	98	93	100	100
			Error within ±10%	5	23	8	22	7	50	51	78
		Sharkey et al.	Error within ±10%	4	NA	7	26	NA	61	63	NA

[a] The number of resolvable parameters is the same for all datasets (HDS: high accuracy dataset, NDS: normal accuracy dataset and DSE: dataset with measurement errors) when $\alpha > 0$ (or $\alpha = 0$) using the method of Gu et al. [2], and using the method of Sharkey et al. [27], respectively; [b] The number of correctly estimated parameters is the same for datasets using the method of Sharkey et al. [27]; [c] For values of α, the differences within 0.1 were listed; [d] The number of estimated parameters were the same for all datasets using the method of Sharkey et al. [27]; [e] The number of correctly estimated parameters is the same for all datasets (using the method of Gu et al. [2]).

Sharkey et al.'s method [27] was unable to correctly estimate any parameter for any dataset with both $\alpha = 0$ and $\alpha > 0$ (Table 1). All five unknown parameters—V_{cmax}, J_{max}, g_m, T_p and R_d—could be estimated for any dataset, and even for some noninvolved parameters in some datasets (Table 1), since Sharkey et al.'s method incorrectly forced data points into a missing state to minimize the objective function. Within ±10% of error ranges, the numbers for V_{cmax}, J_{max}, g_m, T_p and R_d were 22, 65, 41, 7 and 6, respectively, when $\alpha > 0$, and were 23, 64, 63, 8, and 5 when $\alpha = 0$, respectively. V_{cmax} and J_{max} had extreme values in some datasets. Many values of g_m reached their upper limit value of 30 μmol m^{-2} s^{-1} Pa^{-1}. Overall, V_{cmax}, and g_m were overestimated, about 40% of the R_d values were zero, J_{max} was about evenly distributed around the 1:1 line ($y = 1.016x$, $R^2 = 0.462$) with a few extreme large values, and T_p was about evenly distributed around the 1:1 line ($y = 0.987x$, $R^2 = 0.866$) (Figure 1).

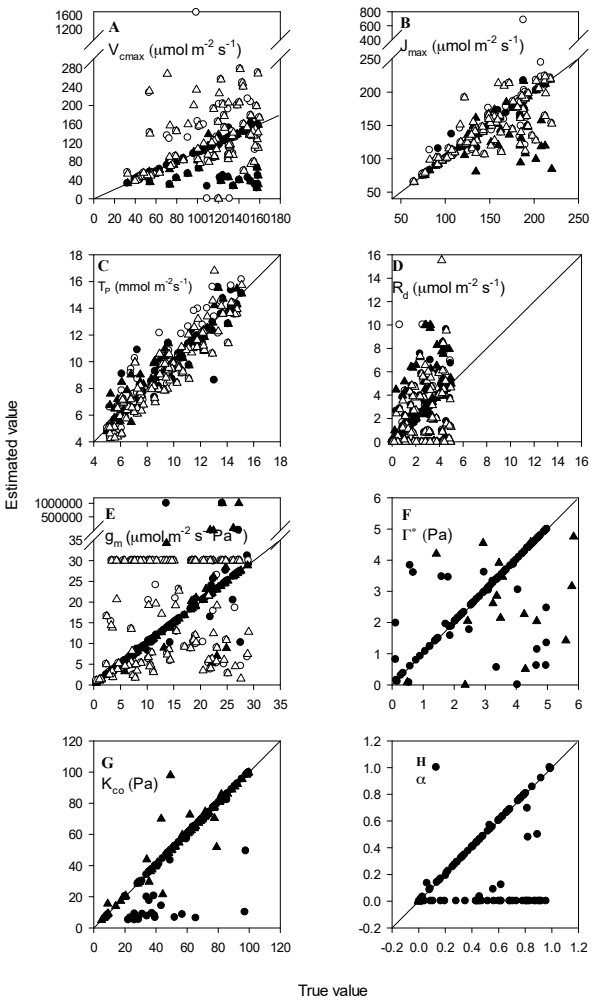

Figure 1. Comparison of the estimated parameters by Gu et al.'s method (black circle for $\alpha > 0$ and black triangle for $\alpha = 0$) and Sharkey et al.'s method (open circle for $\alpha > 0$ and open triangle for $\alpha = 0$) vs. true parameter values for synthetic A_n/C_i curves, 100 datasets with $\alpha > 0$ and 100 with $\alpha = 0$. The datasets consisted of 15 data points with high accuracy (eight decimal points). The points in each figure may have less than 100 values because some datasets do not contain all three states.

3.2. Normal Accuracy Datasets

For normal accuracy datasets with both $\alpha > 0$ and $\alpha = 0$, no parameters were correctly estimated using Gu et al.'s method. In comparison with the high accuracy dataset results, this suggests that the accuracy of A_n/C_i data was important for correctly determining the parameters. The total number of estimated parameters was more than the total number of resolvable parameters (except for g_m)

(Table 1). The number of estimated parameters was more than those utilizing high accuracy datasets, since more non-involved parameters were incorrectly estimated. Compared to the high accuracy datasets, more datasets with a missing state were incorrectly identified as datasets with all three states available by changing a missing state to an available one. For example, a dataset (3, 0, 12) was assigned to (4, 2, 9). The total number of values within ±10% of their "true values" for V_{cmax}, K_{co}, J_{max}, T_p, g_m, R_d, Γ^* and α were 55, 40, 65, 70, 26, 16, 48 and 26 for $\alpha > 0$; and were 43, 27, 64, 67, 20, 12, 42 and 99 for $\alpha = 0$, respectively (Table 1). These numbers were fewer than the corresponding parameters in the high accuracy dataset. Overall, V_{cmax}, K_{co}, was underestimated, R_d was overestimated and g_m was generally underestimated. In some cases, g_m was overestimated when the estimated value was higher than 20 μmol m^{-2} s^{-1} Pa^{-1}. Most of the estimated parameters Γ^*, J_{max} and T_p were evenly distributed around the 1:1 line ($y = 0.952x$ and $R^2 = 0.531$ for $J_{\max}$, $y = 1.001x$ and $R^2 = 0.958$ for T_p) (Figure 2).

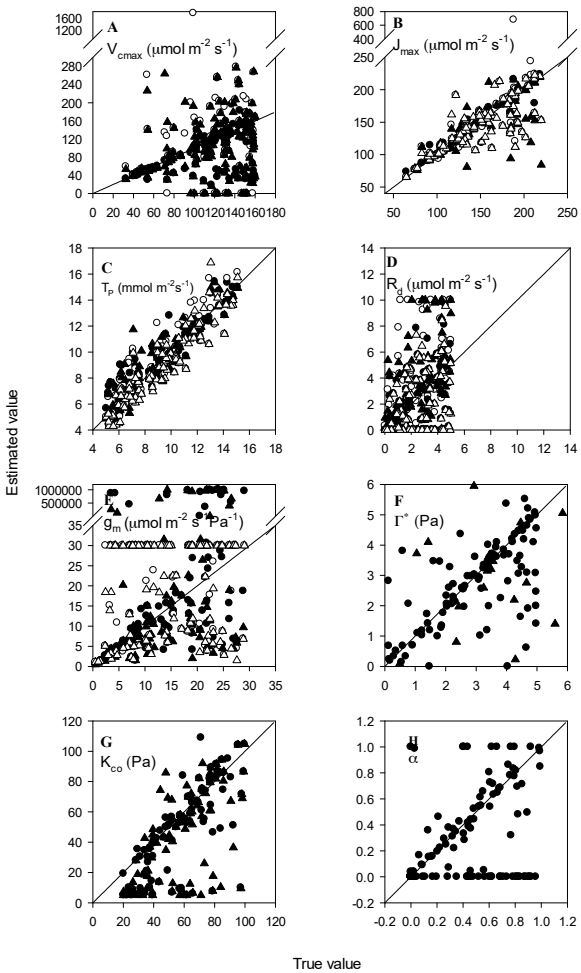

Figure 2. The same as in Figure 1; here applied to normal accuracy datasets.

Using Sharkey et al.'s method [27], there were no correct parameters for both $\alpha = 0$ and $\alpha > 0$ (Table 1). All five unknown parameters—V_{cmax}, J_{max}, g_m, T_p and R_d—were estimated for any dataset (Table 1) within ±10% of error. The values of V_{cmax}, J_{max}, g_m, T_p and R_d were 24, 70, 6, 57 and 6, respectively, when $\alpha > 0$; and 22, 59, 8, 51, and 8, respectively, when $\alpha = 0$. V_{cmax} and J_{max} had extreme values in some datasets. Many values of g_m reached their upper constrained value of 30 μmol m^{-2} s^{-1} Pa^{-1} for both $\alpha = 0$ and $\alpha > 0$. Overall, the parameters V_{cmax} and g_m were underestimated, J_{max} and T_p were evenly distributed around the 1:1 line ($y = 0.991x$ and $R^2 = 0.572$ for J_{max}, $y = 0.976x$ and $R^2 = 0.800$ for T_p), and many values of R_d were over their upper limit of 5.000 μmol m^{-2} s^{-1} when generated (Figure 2).

3.3. Datasets with Measurement Errors

Datasets with measurement errors are representative of observed data. Using Gu et al.'s method, the number of correctly estimated parameters was zero. The common parameters g_m, R_d and Γ^*, and the specific parameters T_p and α were estimated for all 100 datasets. The specific parameter V_{cmax}, K_{co}, or J_{max} was estimated whenever its state was identified (Table 1). The number of estimated parameters was higher than that in normal accuracy datasets, indicating that more parameters would be estimated when using a less accurate dataset. The total numbers of values within ±10% of their "true values" for V_{cmax}, K_{co}, J_{max}, T_p, g_m, R_d, Γ^* and α were 30, 10, 61, 64, 7, 7, 36 and 23 when $\alpha > 0$; and 22, 7, 50, 51, 5, 8, 23 and 78 when $\alpha = 0$, respectively (Table 1). The distributions of estimated parameters (except for T_p) were more scattered than that of normal accuracy datasets. Most of the values for V_{cmax}, K_{co}, g_m and Γ^* were underestimated (Figure 3A,G,E,F), while most of R_d and α were overestimated (Figure 3D,H). About half of the values of T_p and J_{max} were overestimated.

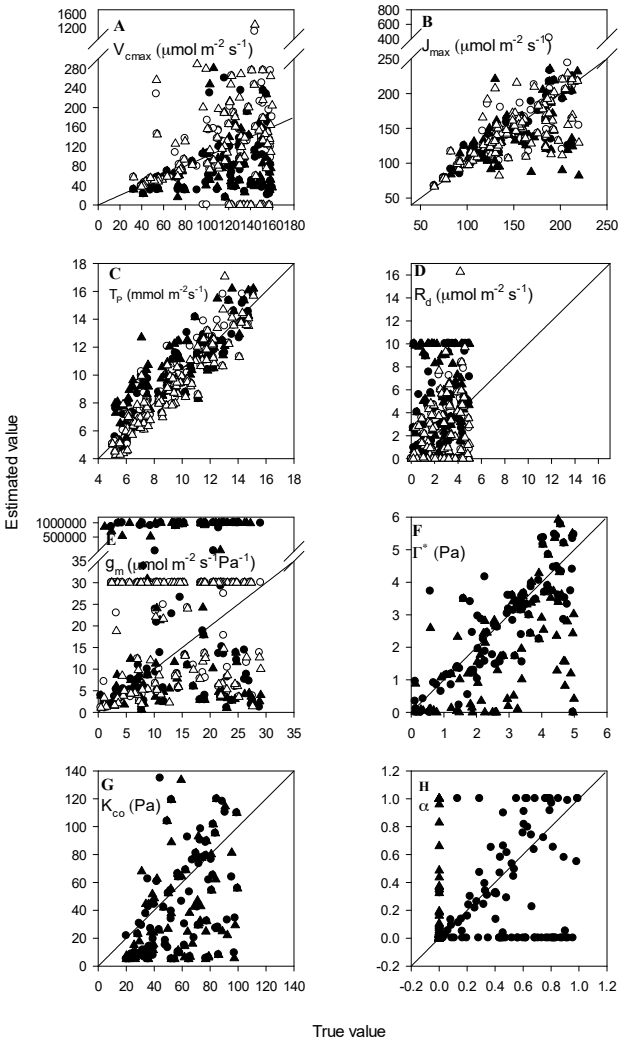

Figure 3. The same as in Figure 2; here applied to datasets with measurement errors.

Using Sharkey et al.'s method, there were no correctly estimated parameters for both $\alpha > 0$ and $\alpha = 0$ (Table 1). The numbers of parameters g_m, R_d, V_{cmax}, J_{max}, T_p, within ±10% errors were 8, 5, 24, 70 and 60 when $\alpha > 0$, and 4, 7, 26, 61 and 63 when $\alpha = 0$, respectively. There were similar numbers of estimated parameters within ±10% errors between datasets with measurement errors having $\alpha > 0$ and $\alpha = 0$ and for normal accuracy and measurement error datasets. This observation suggests that

the impacts of the value of α and measurement errors on the estimated parameters were insignificant in datasets with large errors when using Sharkey et al.'s method. Overall, the distributions of the estimated parameters were similar to those of a normal accuracy dataset (Figure 3).

3.4. Datasets with Varied Data Points

Table 2 summarizes the results of the two methods using high accuracy datasets with a varied number of data points. Gu et al.'s method was unable to guarantee the correct estimation of any parameter values when the number of data points was eight or less. For datasets with nine and 12 points, all eight parameters could be correctly estimated if a dataset met the requirements set by Gu et al. for all eight resolvable parameters. However, the actual number of datasets with all eight resolvable parameters was much higher than the number of datasets that met the requirements of Gu et al. for all eight resolvable parameters. Gu et al.'s method could estimate the parameters g_m, R_d, Γ^*, T_p and α in any dataset (Table 2), though many of them were non-resolvable. Generally, the ratio of correctly estimated parameters to estimated parameters increased (except for K_{co} and Γ^*) with an increasing number of observed data points for the A_n/C_i curve.

Table 2. Effects of the number of data points of A_n/C_i curves on the quality of the parameter estimates obtained from Gu et al.'s and Sharkey et al.'s methods. The "true values" of the parameters are V_{cmax} = 45.572 µmol m^{-2} s^{-1}, J_{max} = 111.32 µmol m^{-2} s^{-1}, R_d = 1.381 µmol m^{-2} s^{-1}, g_m = 1.897 µmol m^{-2} s^{-1} Pa^{-1}, T_p = 7.871 µmol m^{-2} s^{-1}, Γ^* = 4.396 Pa, K_{co} = 43.616 Pa and α = 0.352. The descriptions of the terms are the same as in Table 1.

Number of Data Points	Number of Datasets	Methods	Number of Values	g_m	Γ^*	R_d	V_{cmax}	K_{co}	J_{max}	T_p	α
4	15	Gu et al.	Resolvable	3	1	1	0	1	0	0	0
			Correctly estimated	1	0	0	0	1	0	0	0
			Total estimated	15	15	15	2	2	12	15	15
		Sharkey et al.	Resolvable	7	NA	7	5	NA	5	2	NA
			Correctly estimated [a]	0	0	0	NA	0	0	NA	NA
			Total estimated	15	NA	15	11	NA	13	10	NA
5	21	Gu et al.	Resolvable	9	3	3	0	3	3	0	0
			Correctly estimated	3	1	1	0	0	1	0	0
			Total estimated	21	21	21	10	10	19	21	21
		Sharkey et al.	Resolvable	13	NA	13	10	NA	10	5	NA
			Total estimated	21	NA	21	21	21	NA	19	NA
6	28	Gu et al.	Resolvable	20	10	10	3	8	10	3	3
			Correctly estimated	4	3	3	2	3	3	0	0
			Total estimated	28	28	28	16	16	25	28	28
		Sharkey et al.	Resolvable	21	NA	21	17	NA	17	14	NA
			Total estimated	28	NA	28	27	NA	27	27	NA
7	36	Gu et al.	Resolvable	32	20	20	9	15	19	8	8
			Correctly estimated	7	5	5	3	4	5	1	1
			Total estimated	36	36	36	23	23	31	36	36
		Sharkey et al.	Resolvable	28	NA	28	23	NA	23	20	NA
			Total estimated	28	NA	28	28	NA	28	27	NA
8	45	Gu et al.	Resolvable	45	33	33	20	26	26	20	20
			Correctly estimated	7	6	6	4	4	5	3	3
			Total estimated	45	45	45	31	31	40	45	45
		Sharkey et al.	Resolvable	37	NA	37	31	NA	31	28	NA
			Total estimated	45	NA	44	45	NA	45	45	NA
9	55	Gu et al.	Resolvable	55	43	43	28	34	39	28	28
			Correctly estimated	15	13	13	10	11	11	8	8
			Total estimated	55	55	55	37	37	49	55	55
		Sharkey et al.	Resolvable	47	NA	47	40	NA	40	37	NA
			Total estimated	55	NA	55	53	NA	55	55	NA
12	73	Gu et al.	Resolvable	73	71	71	56	57	64	56	56
			Correctly estimated	30	30	30	25	25	26	24	24
			Total estimated	73	73	73	61	61	67	73	73
		Sharkey et al.	Resolvable	72	NA	72	64	NA	65	65	NA
			Total estimated	73	NA	73	73	NA	73	73	NA

[a] The same values for all parameters using the method of Sharkey et al.

All parameters from the datasets were not correctly estimated by Sharkey et al.'s method (Table 2). The percentage of the number of estimated parameters with a ±10% error relative to the number of estimated parameters for J_{max} and T_p ranged from about 33% when the number of data points was four. When the number of data points was 12 for g_m, R_d and Γ*, this ranged from 0 to 20%. The estimated parameters were unevenly distributed around their "true values".

3.5. Datasets with Varied Accuracy

Gu et al.'s method was unable to correctly estimate any parameter when the accuracy of datasets was four or fewer decimal places. For the datasets with decimal places 5, 6, 7 and 8, the numbers of datasets with correctly estimated parameters were 1, 4, 4, and 9 for datasets with eight data points, and 1, 9, 24 and 30 for datasets with 12 data points, respectively. If the number of data points was eight or fewer, there was no correct estimate for all eight parameters and no guarantee to correctly estimate any parameter. The numbers of datasets with a correctly estimated parameter increased with an increasing number of decimal places. For a dataset with 12 data points, it was possible to obtain all eight correct parameters when the accuracy was five or higher decimal places; however, to guarantee all eight parameters to be correctly estimated, a dataset must meet the requirements of Gu et al. for all eight resolvable parameters, and the accuracy must be seven or higher decimal places. In addition, in Gu et al.'s method, V_{cmax} was underestimated with an error of about 13–17% for datasets with eight data points and 11–15% for datasets with 12 data points. J_{max} was slightly overestimated, with an error of about 0–7% for datasets with eight data points and 1–4% for datasets with 12 data points. T_p was overestimated, with an error of about 8–16% for for datasets with eight data points and 7–11% for for datasets with 12 data points. The estimated parameters g_m, K_{co} R_d, α and Γ* had relatively large errors; for example, the errors of the mean values of R_d were from 80 to 243%.

Sharkey et al.'s method was unable to correctly estimate any parameter for datasets with varied accuracy. The estimated parameters did not change from their initial values when the accuracies of the datasets were five or more decimal places. Compared to the results using Gu et al.'s method, all the estimated parameters had less variation, as indicated by their smaller standard deviations (data not shown). The mean values of V_{cmax} were overestimated, with errors of 18% for datasets with eight data points, and 12–16% for datasets with 12 points. The mean values of J_{max} were close to the "true value", with errors of 5% for the datasets with eight data points, and 5–7% for datasets with 12 points. The means of T_p were overestimated with errors of 12–13% for datasets with eight data points, and 0–2% for datasets with with 12 points. Parameters g_m and R_d had relatively large errors. The ranges of the relative changes of the mean g_m and R_d were from 164 to 694%, and from −43 to 159%, respectively.

4. Discussion

Using high accuracy datasets, Gu et al.'s method was unable to correctly estimate the resolvable parameters that did not meet the requirements set by Gu et al. for resolvable parameters. This is because Gu et al.'s method forced these datasets to have the minimum number of data points as (0 or 3, 0 or 2, 0 or 3), which was an altered distribution. Gu et al.'s method could only correctly derive the resolvable parameters if the dataset satisfied the requirements of Gu et al. for resolvable parameters. However, since any observed dataset distribution is unknown and there are many fitted parameters, one cannot identify which set of parameters are correct.

Using normal accuracy datasets with measurement errors, Gu et al.'s and Sharkey et al.'s methods were unable to correctly estimate any parameter. As shown in Figures 1–3, each set of estimated parameters from a high accuracy dataset, a normal accuracy dataset or a dataset with measurement errors were different. One of the main reasons for these differences was the accuracy of the datasets. Both methods were based on the standard nonlinear regression, which assumes that the error of A_{n} is a random variable whose population mean is zero and variance is constant, and that C_{i} is an independent variable without any error. Since only a few data points are available in each limited state, the sample error will vary considerably, simply by chance. A point with a larger error would tend to

have a larger deviation from the curve and so would have a larger impact on the SSE. In contrast, a point with a smaller error would have a smaller influence. Minimizing the SSE would be inappropriate for datasets with a few data points with relative large errors.

The conditions for correctly estimating all eight parameters using Gu et al.'s method is a dataset with (1) a minimum of three data points for the limited state of Gu et al. for all resolvable parameters; (2) a minimum of nine data points for the A_n/C_i curve; (3) an accuracy of at least seven decimal places; and (4) $\alpha > 0$. These requirements were necessary and sufficient conditions for Gu et al.'s method. If a dataset does not meet these conditions, Gu et al.'s method will be unable to guarantee a fit for any parameter. For example, in the dataset (2, 4, 2) with high accuracy, all the parameters were resolvable; however, Gu et al.'s method incorrectly identified it as (3, 5, 0), leading to all the parameters being incorrect (data not shown).

Sharkey et al.'s method (and other extant methods, see [26]) fits V_{cmax}, J_{max}, T_p, R_d and g_m simultaneously, using all the data points of an A_n/C_i curve, by fitting Equations (1)–(5) and (14), with fixed K_{co} and Γ^*, and assuming $\alpha = 0$. This approach simplifies the fitting method, but may introduce more errors to the estimated parameters if the wrong fixed values are used. Sharkey et al.'s method was unable to correctly estimate any parameters using all examined datasets. One of the major reasons for this could be the use of an incorrect value for the fixed parameters K_{co}, Γ^*, and α. There are different K_{co} and Γ^* values to choose from in the literature [12,27]. In addition, K_{co} changes across diverse species and environmental conditions. One can see from Equations (9)–(13) that a change in one or more parameters may lead to changes in all the other parameters; for example, if a dataset is only in the A_c state, by combining Equations (1), (4) and (8) and assuming independent of all involved parameters, one can have

$$V_{cmax} = \frac{(C_i + K_{co})g_m - A_c}{(C_i - \Gamma^*)g_m - A_c}(A_c + R_d). \tag{23}$$

Thus, the direct impacts of Γ^* and K_{co} on V_{cmax} are

$$\frac{\partial V_{cmax}}{\partial \Gamma^*} = \frac{(C_i + K_{co})g_m - A_c}{[(C_i - \Gamma^*)g_m - A_c]^2}(A_c + R_d)g_m, \tag{24}$$

and

$$\frac{\partial V_{cmax}}{\partial K_{co}} = \frac{(A_c + R_d)g_m}{(C_i - \Gamma^*)g_m - A_c}. \tag{25}$$

From the above equations, one can see that the direct effects of errors in V_{cmax} depend on the errors in K_{co} and Γ^* and the values of A_c, C_i, g_m and R_d. Thus, incorrect values for K_{co} and/or Γ^* will inevitably lead to incorrect parameter estimates. This result is in agreement with that of Ethier and Livingston [12].

In Sharkey et al.'s methods, the CO_2 partial pressure inside the chloroplast C_{cs} was estimated by (a similar estimation is also in the methods of Dubois et al. [24] and Miao et al. [26]):

$$C_{cs} = C_i - \frac{A_{nm}}{g_m}, \tag{26}$$

where A_{nm} is the measured net assimilation rate. Equation (26) is not identical to Equation (8); A_n in Equation (8) is the calculated value in the algorithm. Thus, a minimization of SSE based on Equation (26) or Equation (8) is different. The parameterization of a four-point curve in the A_j state (Table 3) illustrates this problem. The dataset with eight decimal places was generated using the same value of Γ^* as in Sharkey et al.'s method to eliminate the effects of different values of Γ^*. The two sets of estimated parameters were different from their "true values" and modeled A_{jj}s were slightly different from the 'measured' values. One can see that Sharkey et al.'s method was unable to correctly derive the parameters, even using a dataset with a high accuracy and the same fixed value as in the method.

Because the distribution of data points is unknown, it is possible for Sharkey et al.'s method to assign data points to the wrong states to minimize SSE, especially if a state is missing. Table 4 shows an example in which there was one dataset with 12 data points only in the A_c and A_j states. If the data points were assigned to the same states as when it was generated (I), the SSE was larger than if the last three data points were assigned to the A_p state, which was larger than if the transitional point between A_c and A_j was also adjusted. The estimated parameters were different among the three conditions (except for R_d and g_m) and were different from their "true values." It is worth noting that this is an intrinsic problem in Sharkey et al.'s method, because of the limited accuracy and number of data points of an A_n/C_i curve. It is very easy to miss a state, especially the A_p state [24]. This result can be explained by considering that the incorrect identification of the distribution of data points led to incorrect parameter estimation.

Table 3. Parameters estimated from three datasets fitted by Sharkey et al.'s method. The dataset was generated with the fixed value of Sharkey et al.'s method (Γ^* = 3.743 Pa^{-1}) at a leaf temperature of 25 °C and an air pressure of 100 Pa. A_{jj}I and A_{jj}II and estimated parameters I and II are values estimated using two different sets of the initial values of the parameters. A_{jmj} and A_{jj} were measured (generated) values using Equation (8) and modeled values using Equation (26), respectively.

C_i (μmol mol^{-1})	A_{jmj} (μmol m^{-2} s^{-1})	A_{jj}I (μmol m^{-2} s^{-1})	A_{jj}II (μmol m^{-2} s^{-1})	Parameter/SSE	True Parameter	Estimated Parameter I	Estimated Parameter II
373.56422385	20.52707868	20.53126975	20.52533184	J_{max} (μmol m^{-2} s^{-1})	120.494	126.278	117.536
559.64521334	22.90931588	22.90327060	22.91250945	R_d (μmol m^{-2} s^{-1})	1.674	3.014	1.039
672.56218564	23.76820169	23.76423592	23.7701254	g_m (μmol m^{-2} s^{-1}Pa^{-1})	9.564	30.000	7.170
909.96541253	24.92146362	24.92728744	24.91809557	SSE (μmol m^{-2} s^{-1})^2	-	0.000	0.000

Table 4. Comparison of fitting results from a 12-point A_n/C_i using Sharkey et al.'s method by assigning a different transition point between the A_c, A_j and A_p states, as indicated by I, II, III or IV. The dataset was generated as containing only A_c and A_j states, and the transitional point of C_i is between 201 and 284 μmol mol^{-1}.

A_n	C_i	Parameter/SSE	True Value	I [a]	II [b]	III [c]
2.17	38.9	V_{cmax} (μmol m^{-2} s^{-1})	99.4	108.9	108.9	112.0
4.47	58.1	J_{max} (μmol m^{-2} s^{-1})	136.2	143.1	146.1	145.5
9.60	99.4	T_p (μmol m^{-2} s^{-1})	-		10.1	10.1
15.3	150	R_d (μmol m^{-2} s^{-1})	1.1	0.0	0.0	0.0
20.6	201	g_m (μmol m^{-2} s^{-1} Pa^{-1})	3.7	30.0	30.0	30.0
25.3	284	K_{co} (Pa)	42.9	-	-	-
27.0	371	Γ^* (Pa)	1.9	-	-	-
28.2	415	SSE	-	9.939	7.315	7.266
29.4	552	-	-	-	-	-
30.1	673	-	-	-	-	-
30.1	730	-	-	-	-	-
30.6	908	-	-	-	-	-

[a] The transition point was the same as the corresponding error-free data set generated by true parameters; [b] is the same as [a], except for the fact that the last 3 points were assigned to A_p state; [c] minimum of SSE where the transition point of C_i between A_c and A_j is 150 and 201 μmol mol^{-1}, and between A_j and A_p it is between 552 to 673 μmol mol^{-1}.

Both the methods of Sharkey et al. and Gu et al. have assigned constraints for some parameters based on prior knowledge; for example, $g_m \leq 30$ μmol m^{-2} s^{-1} Pa^{-1} in Sharkey et al.'s method, and $R_d \leq 10$ μmol m^{-2} s^{-1}, $g_m \leq 1{,}000{,}000$ μmol m^{-2} s^{-1} Pa^{-1}, and $\Gamma^* \geq 0$ Pa in Gu et al.'s method. There were also different constraints found in other methods in the literature, such as the constraint of $-3 < R_d < 50$ μmol m^{-2} s^{-1} in Dubois et al. [24]. Firstly, if a parameter is estimated within the range of its constraint, the local minimum must be achieved in the range of the constraint. Secondly, the constraints are subjective choices which are probably not realistic; for example, we obtained R_d

as zero by Sharkey et al.'s method (data not shown), and Γ^* as zero (data not shown) by Gu et al.'s method. Thirdly, if a parameter is equal to its constraint, which is likely to be incorrect, this incorrect parameter may substantially affect the estimates of other parameters. The problems of finding a local minimum and non-uniqueness of the parameters is intrinsic to nonlinear regression; for example, in Sharkey et al.'s method, the estimates of the parameters were sensitive to the initial values (Table 3). There are similar problems in the method of Dubois et al. [24] and Miao et al. [26], as stated by Gu et al. [2].

However, the best fit but unreliable parameter set may be used to predict A_n from C_i if the SSE is small, as argued by Ethier and Livingston [12]. The four A_n/C_i curves modeled by the three sets of parameters with 15 data points (7, 6, 2) were compared. The three sets of parameters were high accuracy, including measurement errors, and derived from curves generated using the parameters fitted to the dataset with measurement errors by methods of Gu et al. and Sharkey et al., respectively. All curves are superposed.

The distributions of the estimated parameters (except for T_p) by both methods using a normal accuracy dataset and a dataset with measurement errors were very scattered. The majority of the estimated parameters were either overestimated or underestimated (Table 1, Figures 2 and 3), implying that the mean values of these estimated parameters could not represent the "true values". This can be explained by low data accuracies and by the intrinsic problems in both methods. The estimated T_p was evenly distributed around the 1:1 line, indicating that the mean T_p was close to its "true value"; this was because T_p is almost equal to $\frac{A_n}{3}$ (Equations (1), (6), and (7)) since $A_n >> R_d$ in the A_p state.

5. Conclusions

In this study, we tested Sharkey et al.'s [21] commonly-used A_n/C_i curve fitting method, and the method developed recently by Gu et al. [2], using datasets with a number of A_n/C_i curve points from four to 15 and accuracies from one to eight decimal places. The generated datasets were conservative. In the literature, the typical number of data points of an A_n/C_i curve is eight to 12, which is considered enough for estimation [2,24,26,30]. The accuracy of the measured data in a typical open gas exchange system is one decimal place; e.g., *Li-6400*. The error level of the generated datasets was lower than normally seen in practice, because only one source of measurement error, e.g., the random noise of CO_2 $\pm$0.2 µmol mol^{-1}, was imposed as an error to the datasets; other error sources were not included [34].

To correctly estimate the parameters of an A_n/C_i curve, we have to consider data accuracy and the number and distribution of data points as well as error distribution. Based on the results using different generated datasets, we concluded that Sharkey et al.'s method failed to correctly estimate the parameters, while Gu et al.'s method was unable to correctly estimate the parameters using a dataset with a number of data points fewer than five or with an accuracy of four or fewer decimal places. At least eight data points were required for Gu et al.'s method to correctly estimate all eight parameters. For the datasets with measurement errors and the same accuracy of a typical open gas exchange system—i.e., *Li-6400*—using Gu et al.'s approach, the parameters V_{cmax}, K_{co}, g_m and Γ^* were underestimated, while R_d and α were often overestimated. The distributions of T_p were evenly dispersed around the 1:1 line using both approaches. Using Sharkey et al.'s approach, the parameters J_{max} was overestimated, V_{cmax} and g_m were underestimated, and many values of R_d were over their upper limit of 5.000 µmol m^{-2} s^{-1}. The mean values of all estimated parameters, except for T_p, were not close to their "true values".

This failure of parameterization was due to two types of problems. One was the limited number of data points and the limited accuracy of datasets in both methods that did not meet the assumptions for nonlinear regression (measurement errors in both A_n and C_i). The other problem was the failure to identify correct parameter estimates using Gu et al.'s method, due to the unknown data point distribution.

This study revealed that high accuracy A_n/C_i and enzyme kinetic measurements are required to correctly estimate these parameters, even when sufficient data points are provided. An accurate

estimation of the parameters can contribute to the enhancement of food security under climate change by reducing potential errors when the biological and biophysical processes of CO_2 assimilation are correctly spatially and temporally scaled-up for ecosystem studies. This study does not address the question of how these measurement accuracies can be improved. It is recommended that this question be addressed in a further study.

Acknowledgments: This research was supported by the APEC Climate Center.

Author Contributions: Qingguo Wang, David H. Fleisher and Jong Ahn Chun conceived and designed the experiments; Qingguo Wang and Jong Ahn Chun performed the experiments; Qingguo Wang, Jonathan P. Resop and Jong Ahn Chun analyzed the data; Dennis Timlin and Vangimalla Reddy contributed reagents/materials/analysis tools; Qingguo Wang and Jong Ahn Chun wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Nomenclature

A_c	Net assimilation rate assuming Rubisco-limited state, μmol m^{-2} s^{-1}
A_{cmi}	Measured net assimilation rates in the A_c state point i, μmol m^{-2} s^{-1}
A_{jj}	Calculated net assimilation rates in the A_j state point j, μmol m^{-2} s^{-1}
A_n	Net assimilation rate, μmol m^{-2} s^{-1}
A_{nj}	Net assimilation rate at A_j state, μmol m^{-2} s^{-1}
A_p	Net assimilation rate assuming triose phosphate utilization (TPU) limited state, μmol m^{-2} s^{-1}
A_{pk}	Calculated net assimilation rates in the A_p state point k, μmol m^{-2} s^{-1}
C_c	Chloroplastic CO_2 partial pressure, Pa
C_i	Intercellular CO_2 partial pressure, Pa
g_m	Internal (mesophyll) conductance to CO_2 transport, μmol m^{-2} s^{-1} Pa^{-1}
J_{max}	Maximum rate of electron transport, μmol m^{-2} s^{-1}
K_o	Michaelis–Menten constant for Rubisco for O_2, Pa
R_d	Day respiration, μmol m^{-2} s^{-1}
T_p	Rate of triose phosphate export from the chloroplast, μmol m^{-2} s^{-1}
W_c	Maximal Rubisco carboxylation rate, μmol m^{-2} s^{-1}
W_p	TPU-limited carboxylation rate, μmol m^{-2} s^{-1}
A_{ci}	Calculated net assimilation rates in the A_c state point i, μmol m^{-2} s^{-1}
A_j	Net assimilation rate assuming RuBP regeneration limited state, μmol m^{-2} s^{-1}
A_{jmj}	Measured net assimilation rates in the A_j state point j, μmol m^{-2} s^{-1}
A_{nc}	Net assimilation rate at A_c state, μmol m^{-2} s^{-1}
A_{nm}	Measured net assimilation rate, μmol m^{-2} s^{-1}
Γ^*	Chloroplastic CO_2 photocompensation point, Pa
A_{pmk}	Measured net assimilation rates in the A_p state point k, μmol m^{-2} s^{-1}
C_{cc_cj}	Chloroplastic CO_2 partial pressure at transition point between A_c and A_j state, Pa
C_s	Intercellular CO_2 partial pressure, Pa
J	Potential electron transport rate at the measurement light level, μmol m^{-2} s^{-1}
K_c	Michaelis-Menten constant for Rubisco for CO_2, Pa
K_{co}	Effective Michaelis–Menten coefficient for CO_2, $K_{co} = K_c(1 + O/K_o)$, Pa
O	Oxygen partial pressure, Pa
V_{cmax}	Maximal Rubisco carboxylation rate, μmol m^{-2} s^{-1}
W_j	RuBP regeneration-limited carboxylation rate, μmol m^{-2} s^{-1}
α	Non-returned fraction of the glycolate carbon recycled in the photorespiratory cycle (dimensionless)

Appendix A. Errors Superimposed to an Ideal Dataset

Both A_n and C_i are subject to measurement errors. The NORMINV function in Microsoft Excel 2007 (Microsoft, Seattle, WA, USA) was used to generate these errors. There are three arguments for the functions (Probability, Mean, and Standard Deviation). For this study, the RAND function was used for probability and the Mean and Standard Deviation values were 0 and 0.2, respectively. The values of A_n and C_i are not directly measured in an open gas exchange system (e.g., *Li-6400*). They

are calculated from the measured environmental conditions of a leaf enclosed in the leaf chamber, assuming the leaf is homogeneous, implying that the photosynthesis of every portion of the measured leaf is in the same limited state in the same time. The environmental conditions are provided by some signal processing algorithm; for example, CO_2 concentration was calculated from the signal of the infrared gas analyzer (IRGA). Under such conditions, the A_n/C_i curve can be described by the FvCB model if the measurements are without error. However, in practice, all physical measurements are subject to measurement errors. The accuracy of the measurements could be affected by many factors, including, but not limited to, random noise in the CO_2 and water vapor sensors, leaf biological and environmental heterogeneity, leaf temperature variation, gas leaks, dark transpiration and respiration from the leaf under the gaskets, signal processing algorithms, and the calibration of the instrument. For simplicity, we only focus on the errors from CO_2 sensors. Following von Caemmerer and Farquhar [23] and ignoring the dilution effects of water vapor, the gas exchange rate A_n and C_i in an open system can be simply expressed as:

$$A_n = \frac{F}{S}(C_r - C_s) \tag{A1}$$

$$C_i = C_s - \frac{A_n}{g_{tc}} \tag{A2}$$

where g_{tc} is the total conductance to CO_2; F is the incoming air flow rate; C_r and C_s are reference IRGA and sample IRGA CO_2 concentrations, respectively; and S is the leaf chamber area. From Equations (A1) and (A2), one can see that C_i is not independent of A_n. The uncertainties of A_n will transfer to C_i.

It is reasonable to assume that errors in g_{tc}, F, S, Cr and C_s are independent and randomly distributed, because the data were recorded at steady state. According to the principle of the propagation of errors, the uncertainties of A_n and C_i are given, respectively, by

$$\varepsilon_{A_n} = \frac{1}{S}\sqrt{(F)^2\left[(\varepsilon C_r)^2 + (\varepsilon C_s)^2\right] + (C_r - C_s)^2\left[(\varepsilon F)^2 + \frac{F^2}{S^2}(\varepsilon S)^2\right]} \tag{A3}$$

$$\varepsilon_{C_i} = \sqrt{\left[1 - \left(\frac{F}{Sg_{tc}}\right)\right]^2(\varepsilon C_s)^2 + \left(\frac{F}{Sg_{tc}}\right)^2(\varepsilon C_r)^2 + \left(\frac{C_r - C_s}{g_{tc}S}\right)^2\left[(\varepsilon F)^2 + \left(\frac{F}{S}\right)^2(\varepsilon S)^2 + \frac{F^2}{(g_{tc})^2}(\varepsilon g_{tc})^2\right]} \tag{A4}$$

The bigger the area of the leaf chamber and/or the less flow rate and/or the closer the CO_2 difference between the reference and the sample, the smaller measurement errors of A_n (Equation (A3)) and C_i (Equation (A4)). Each gas analysis system has its own specific errors for F, C_r and C_s; for example, the *Li-6400* we used had errors of C_r ± 0.36 μmol mol^{-1} and errors of C_S ± 0.12 μmol mol^{-1} at 370 μmol mol^{-1} [34].

For simplicity, we assume that the errors in S, F and g_{tc} are ignorable, and $\varepsilon C_r = \varepsilon C_s = \varepsilon C$ Equations (A3) and (A4) become to

$$\varepsilon_{A_n} = \frac{\sqrt{2}F}{S}\varepsilon C \tag{A5}$$

$$\varepsilon_{C_i} = \varepsilon C\sqrt{\left[1 - 2\left(\frac{F}{Sg_{tc}}\right) + 2\left(\frac{F}{Sg_{tc}}\right)^2\right]} \tag{A6}$$

For the typical area of the leaf chamber, S is 600 mm^2, the flow rate is 400 μmol s^{-1} and the δC = ±0.2 μmol mol^{-1} [32], the error of A_n is approximately ±0.2 μmol m^{-2} s^{-1}. The uncertainties of C_i also depend on g_{tc}, which is mainly derived from stomatal conductance. For generating the A_n/C_i points, we assume that g_s is linear with CO_2 concentration and the range of g_s is from 0.2 mol CO_2 m^{-2} s^{-1} at C_i of 0 μmol mol^{-1} to 0.1 mol CO_2 m^{-2} s^{-1} at C_s of 2000 μmol mol^{-1}.

References

1. Farquhar, G.D.; Von Caemmerer, S.; Berry, J.A. A biochemical model of photosynthetic CO_2 assimilation in leaves of C_3 species. *Planta* **1980**, *149*, 78–90. [CrossRef] [PubMed]
2. Gu, L.; Pallardy, S.G.; Tu, K.; Law, B.E.; Wullschleger, S.D. Reliable estimation of biochemical parameters from C_3 leaf photosynthesis-intercellular carbon dioxide response curves. *Plant Cell Environ.* **2010**, *33*, 1852–1874. [CrossRef] [PubMed]
3. Long, S.P. Modification of the response of photosynthetic productivity to rising temperature by atmospheric CO_2 concentrations: Has its importance been underestimated? *Plant Cell Environ.* **1991**, *14*, 729–739. [CrossRef]
4. Amthor, J. Scaling CO_2-photosynthesis relationships from the leaf to the canopy. *Photosynth. Res.* **1994**, *39*, 321–350. [CrossRef] [PubMed]
5. Field, C.B.; Avissar, R. Bidirectional interactions between the biosphere and the atmosphere—Introduction. *Glob. Chang. Biol.* **1998**, *4*, 459–460. [CrossRef]
6. Jones, J.W.; Hoogenboom, G.; Porter, C.H.; Boote, K.J.; Batchelor, W.D.; Hunt, L.A.; Wilkens, P.W.; Singh, U.; Gijsman, A.J.; Ritchie, J.T. The DSSAT Cropping System Model. *Eur. J. Agron.* **2003**, *18*, 235–265. [CrossRef]
7. De Pury, D.G.G.; Farquhar, D.G. Simple scaling of photosynthesis from leaves to canopy without the errors of bigleaf models. *Plant Cell Environ.* **1997**, *20*, 537–557. [CrossRef]
8. Sellers, P.J.; Bounoua, L.; Collatz, G.J.; Randall, D.A.; Dazlich, D.A.; Los, S.O; Berry, J.A.; Fung, I.; Tucker, C.J.; Field, C.B.; et al. Comparison of radiative and physiological effects of doubled atmospheric CO_2 on climate. *Science* **1996**, *271*, 1402–1406. [CrossRef]
9. Sellers, P.J.; Dickinson, R.E.; Randall, D.A.; Betts, A.K.; Hall, F.G.; Berry, J.A; Collatz, G.J.; Denning, A.S.; Mooney, H.A.; Nobre, C.A.; et al. Modeling the exchanges of energy, water, and carbon between continents and the atmosphere. *Science* **1997**, *275*, 502–509. [CrossRef] [PubMed]
10. Wang, Y.P.; Leuning, R. A two-leaf model for canopy conductance, photosynthesis and portioning of available energy I: Model description and comparison with a multi-layered model. *Agric. For. Meteorol.* **1998**, *91*, 89–111. [CrossRef]
11. Wittig, V.E.; Bernacchi, C.J.; Zhu, X.-G.; Calfapietra, C.; Ceulemans, R.; Deangelis, P.; Gielen, B.; Miglietta, F.; Morgan, P.B.; Long, S.P. Gross primary production is stimulated for three Populus species grown under free-air CO_2 enrichment from planting through canopy closure. *Glob. Chang. Biol.* **2005**, *11*, 644–656. [CrossRef]
12. Ethier, G.J.; Livingston, N.J. On the need to incorporate sensitivity to CO_2 transfer conductance into the Farquhar-von Caemmerer-Berry leaf photosynthesis model. *Plant Cell Environ.* **2004**, *27*, 137–153. [CrossRef]
13. Bunce, J.A. Acclimation of photosynthesis to temperature in eight cool and warm climate herbaceous C_3 species: Temperature dependence of parameters of a biochemical photosynthesis model. *Photosynth. Res.* **2000**, *63*, 59–67. [CrossRef] [PubMed]
14. Harley, P.C.; Loreto, F.; Di Marco, G.; Sharkey, T.D. Theoretical consideration when estimating the mesophyll conductance to CO_2 flux by analysis of the response of Photosynthesis to CO_2. *Plant Physiol.* **1992**, *98*, 1429–1436. [CrossRef] [PubMed]
15. Harley, P.C.; Thomas, R.B.; Reynolds, J.F.; Strain, B.R. Modelling photosynthesis of cotton grown in elevated CO_2. *Plant Cell Environ.* **1992**, *15*, 271–282. [CrossRef]
16. Hikosaka, K.; Ishikawa, K.; Borjigidai, A.; Muller, O.; Onoda, Y. Temperature acclimation of photosynthesis: Mechanisms involved in the changes in temperature dependence of photosynthetic rate. *J. Exp. Bot.* **2006**, *57*, 291–302. [CrossRef] [PubMed]
17. Possell, M.; Hewitt, C.N. Gas exchange and photosynthetic performance of the tropical tree Acacia nigrescens when grown in different CO_2 concentrations. *Planta* **2009**, *229*, 837–846. [CrossRef] [PubMed]
18. Von Caemmerer, S.; Berry, J.; Farquhar, G.D. Biochemical model of C_3 photosynthesis. In *Photosynthesis in Silico: Understanding Complexity from Molecules to Ecosystems*; Laisk, A., NedbalGovindjee, L., Eds.; Springer Science + Business Media, B.V.: Dordrecht, The Netherlands, 2009; pp. 209–230.
19. Bernacchi, C.J.; Bagley, J.E.; Serbin, S.P.; Ruiz-Vera, U.M.; Rosenthal, D.M.; Vanloocke, A. Modelling C_3 photosynthesis from the chloroplast to the ecosystem. *Plant Cell Environ.* **2013**, *36*, 1641–1657. [CrossRef] [PubMed]

20. Farquhar, G.D.; von Caemmerer, S. Modelling of photosynthetic responses to environmental conditions. In *Physiological Plant Ecology II. Encyclopedia of Plant Physiology*; Lange, O.L., Nobel, P.S., Osmond, C.B., Ziegler, H., Eds.; New Series; Springer: Heidelberg, Germany, 1982; Volume 12B, pp. 550–587.
21. Sharkey, T.D.; Berry, J.A.; Raschke, K. Starch and sucrose synthesis in Phaseolus vulgaris as affected by light, CO_2, and abscisic acid. *Plant Physiol.* **1985**, *77*, 617–620. [CrossRef] [PubMed]
22. Von Caemmerer, S. *Biochemical Models of Leaf Photosynthesis*; Techniques in Plant Sciences No. 2; CSIRO Publishing: Collingwood, Australia, 2000.
23. Von Caemmerer, S.; Farquhar, G.D. Some relationships between the biochemistry of photosynthesis and the gas exchange of leaves. *Planta* **1981**, *153*, 376–387. [CrossRef] [PubMed]
24. Dubois, J.J.B.; Fiscus, E.L.; Booker, F.L.; Flowers, M.D.; Reid, C.D. Optimizing the statistical estimation of the parameters of the Farquhar-von Caemmerer-Berry model of photosynthesis. *New Phytol.* **2007**, *176*, 402–414. [CrossRef] [PubMed]
25. Manter, D.K.; Kerrigan, J. A/C_i curve analysis across a range of woody plant species: Influence of regression analysis parameters andmesophyll conductance. *J. Exp. Bot.* **2004**, *55*, 2581–2588. [CrossRef] [PubMed]
26. Miao, Z.W.; Xu, M.; Lathrop, R.G.; Wang, Y. Comparison of the A-Cc 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 Environ.* **2009**, *32*, 109–122. [CrossRef] [PubMed]
27. Sharkey, T.D.; Bernacchi, C.J.; Farquhar, G.D.; Singsaas, E.L. Fitting photosynthetic carbon dioxide response curves for C-3 leaves. *Plant Cell Environ.* **2007**, *30*, 1035–1040. [CrossRef] [PubMed]
28. Su, Y.H.; Zhu, G.F.; Miao, Z.W.; Feng, Q.; Chang, Z. Estimation of parameters of a biochemically based model of photosynthesis using a genetic algorithm. *Plant Cell Environ.* **2009**, *32*, 1710–1723. [CrossRef] [PubMed]
29. Yin, X.; Struik, P.C. 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 Environ.* **2009**, *32*, 1513–1524. [CrossRef] [PubMed]
30. *Using the Li-6400/Li-6400XT Portable Photosynthesis System*, version 6; Li-Cor, Inc.: Lincoln, NE, USA, 2008.
31. Flexas, J.; Ribas-Carbó, M.; Díaz-Espejo, A.; Galmés, J.; Medrano, H. Mesophyll conductance to CO_2: Current knowledge and future prospects. *Plant Cell Environ.* **2008**, *31*, 602–621. [CrossRef] [PubMed]
32. Hanson, P.J.; Amthor, J.S.; Wullschleger, S.D.; Wilson, K.B.; Grant, R.F.; Hartley, A.; Hui, D.; Hunt, E.R., Jr.; Johnson, D.W.; Kimballet, J.S.; et al. Oak forest carbon and water simulations: Model intercomparisons and evaluations against independent data. *Ecol. Monogr.* **2004**, *74*, 443–489. [CrossRef]
33. Harley, P.C.; Baldocchi, D.D. Scaling carbon dioxide and water vapour exchange from leaf to canopy in a deciduous forest. I. leaf model parameterization. *Plant Cell Environ.* **1995**, *18*, 1146–1156. [CrossRef]
34. Wang, Q.; Fleisher, D.; Reddy, V.R.; Timlin, D.; Chun, J.A. Quantifying the Measurement Errors in a Potable Open Gas Exchange System and Their Effects on the Parameterization of Farquhar et al. model for C3 Leaves. *Photosynthetica* **2012**, *50*, 22–238. [CrossRef]