

## RESEARCH ARTICLE

# Rapid estimation of photosynthetic leaf traits of tropical plants in diverse environmental conditions using reflectance spectroscopy

Julien Lamour<sup>1\*</sup>, Kenneth J. Davidson<sup>1,2</sup>, Kim S. Ely<sup>1</sup>, Jeremiah A. Anderson<sup>1</sup>, Alistair Rogers<sup>1</sup>, Jin Wu<sup>3</sup>, Shawn P. Serbin<sup>1</sup>

**1** Environmental & Climate Sciences Department, Brookhaven National Laboratory, Upton, NY, United States of America, **2** Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY, United States of America, **3** Division of Ecology and Biodiversity, School of Biological Sciences, University of Hong Kong, Hong Kong, China

\* [jlamour.sci@gmail.com](mailto:jlamour.sci@gmail.com)



## OPEN ACCESS

**Citation:** Lamour J, Davidson KJ, Ely KS, Anderson JA, Rogers A, Wu J, et al. (2021) Rapid estimation of photosynthetic leaf traits of tropical plants in diverse environmental conditions using reflectance spectroscopy. *PLoS ONE* 16(10): e0258791. <https://doi.org/10.1371/journal.pone.0258791>

**Editor:** Robert E. Sharwood, Western Sydney University, AUSTRALIA

**Received:** May 9, 2021

**Accepted:** October 6, 2021

**Published:** October 19, 2021

**Copyright:** © 2021 Lamour et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are available from NGEE Tropics Data Collection at the following DOIs: [10.15486/ngt/1781003](https://doi.org/10.15486/ngt/1781003), [10.15486/ngt/1781004](https://doi.org/10.15486/ngt/1781004) or in [Supporting Information](#).

**Funding:** This work was supported by the Next-Generation Ecosystem Experiments (NGEE Tropics) project that is supported by the Office of Biological and Environmental Research in the Department of Energy, Office of Science, and through the United States Department of Energy

## Abstract

Tropical forests are one of the main carbon sinks on Earth, but the magnitude of CO<sub>2</sub> absorbed by tropical vegetation remains uncertain. Terrestrial biosphere models (TBMs) are commonly used to estimate the CO<sub>2</sub> absorbed by forests, but their performance is highly sensitive to the parameterization of processes that control leaf-level CO<sub>2</sub> exchange. Direct measurements of leaf respiratory and photosynthetic traits that determine vegetation CO<sub>2</sub> fluxes are critical, but traditional approaches are time-consuming. Reflectance spectroscopy can be a viable alternative for the estimation of these traits and, because data collection is markedly quicker than traditional gas exchange, the approach can enable the rapid assembly of large datasets. However, the application of spectroscopy to estimate photosynthetic traits across a wide range of tropical species, leaf ages and light environments has not been extensively studied. Here, we used leaf reflectance spectroscopy together with partial least-squares regression (PLSR) modeling to estimate leaf respiration ( $R_{\text{dark}25}$ ), the maximum rate of carboxylation by the enzyme Rubisco ( $V_{\text{cmax}25}$ ), the maximum rate of electron transport ( $J_{\text{max}25}$ ), and the triose phosphate utilization rate ( $T_{\text{p}25}$ ), all normalized to 25°C. We collected data from three tropical forest sites and included leaves from fifty-three species sampled at different leaf phenological stages and different leaf light environments. Our resulting spectra-trait models validated on randomly sampled data showed good predictive performance for  $V_{\text{cmax}25}$ ,  $J_{\text{max}25}$ ,  $T_{\text{p}25}$  and  $R_{\text{dark}25}$  (RMSE of 13, 20, 1.5 and 0.3  $\mu\text{mol m}^{-2} \text{s}^{-1}$ , and  $R^2$  of 0.74, 0.73, 0.64 and 0.58, respectively). The models showed similar performance when applied to leaves of species not included in the training dataset, illustrating that the approach is robust for capturing the main axes of trait variation in tropical species. We discuss the utility of the spectra-trait and traditional gas exchange approaches for enhancing tropical plant trait studies and improving the parameterization of TBMs.

contract No. DE-SC0012704 to Brookhaven National Laboratory.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

Tropical forests absorb more CO<sub>2</sub> than any other biome and current estimates suggest that they are responsible for 60% of the annual, global CO<sub>2</sub> assimilation [1]. Terrestrial biosphere models (TBMs) are commonly used to forecast the response of vegetation carbon sequestration to environmental change, yet their ability to adequately capture the responses of CO<sub>2</sub> cycling is currently limited by the poor representation of key leaf-level physiological processes. In particular, the parameterization of TBMs associated with the underlying representation of photosynthesis and respiration drives some of the greatest uncertainty in key TBM outputs [2,3]. Importantly, the parameterization of photosynthesis in TBMs also has a strong influence on the projected response to rising atmospheric CO<sub>2</sub> concentration and temperature [4,5]. Therefore, accurate representation of these key parameters is central to projecting CO<sub>2</sub> exchange, and determining the response of tropical forests to a changing climate.

Most TBMs utilize some form of the Farquhar, von Caemmerer and Berry (FvCB) representation of leaf-level photosynthesis ([6]), or a version proposed by ([3,7,8]). In the FvCB model, the net CO<sub>2</sub> assimilation rate of leaves corresponds to the minimum of three potential rates of three processes taken independently: the Rubisco limited assimilation rate ( $A_c$ ), which depends on the maximum carboxylation rate of Rubisco ( $V_{cmax}$ ), the electron transport limited assimilation rate ( $A_j$ ), which depends on the maximum rate of electron transport ( $J_{max}$ ), and the export limited assimilation rate ( $A_p$ ), which depends on the rate of triose phosphate utilization ( $T_p$ ) for the synthesis of starch and sucrose [6,9,10]. The FvCB model also considers the release of CO<sub>2</sub> by dark respiration ( $R_{dark}$ ) which lowers the rate of gross CO<sub>2</sub> assimilation (photosynthesis, *stricto sensu*). These four parameters ( $V_{cmax}$ ,  $J_{max}$ ,  $T_p$  and  $R_{dark}$ ), are all temperature dependent [5,11–14] and are expressed at a common reference temperature, typically 25°C ( $V_{cmax25}$ ,  $J_{max25}$ ,  $T_{p25}$  and  $R_{dark25}$ ). Providing robust estimates of these parameters and rates is central to improving parameterization and ability of TBMs to simulate the response of tropical forests to climate change.

The “gold standard” for determining leaf  $V_{cmax25}$ ,  $J_{max25}$ ,  $T_{p25}$  and  $R_{dark25}$  relies on laborious, time consuming and logistically challenging procedures. For example, the traditional approach for estimating  $V_{cmax25}$ ,  $J_{max25}$  and  $T_{p25}$  uses a combination of direct measurements of the response of photosynthesis to CO<sub>2</sub> in a light acclimated leaf (commonly referred to as an  $A-C_i$  curve), followed by fitting the data to a mechanistic model to estimate  $V_{cmax25}$ ,  $J_{max25}$  and  $T_{p25}$  (e.g. [15–18]). Measurement of an  $A-C_i$  curve followed by dark adaptation and measurement of dark adapted  $R_{dark25}$  can take two hours for a single leaf. In addition, these key photosynthetic parameters also vary with a range of other biotic and abiotic covariates, including plant species, leaf age, season, environment gradients and stress [19–24]. Therefore, the use of traditional gas exchange approaches has been a significant bottleneck in providing key leaf-level physiological data, thus limiting our ability to characterize variation in key photosynthetic parameters across these axes of biotic and abiotic variation.

Leaf reflectance spectroscopy has shown promise for enabling the rapid estimation of a wide-range of leaf traits, and in many cases replacing slower traditional methods. For example, spectroscopic approaches have already been used to replace chemical assays for leaf chlorophyll extraction (e.g., Dualex: FORCE-A, Orsay, France; SPAD: Minolta Camera Co., Osaka, Japan) using simple spectral indices. Leaf traits involving more complex relationships with reflectance at different wavelengths [25] can be inferred using leaf-level radiation transfer models (RTMs), such as the PROSPECT model [26,27]. Another common approach consists of using empirical methods such as Partial Least Square Regression (PLSR) [28] to model the relationship between reflectance and leaf traits such as leaf mass per area, carbohydrate content and nitrogen content [29–31]. In addition, the PLSR modeling approach has been shown

capable of estimating a range of physiological parameters currently impossible to estimate with leaf RTMs, including the parameters needed to model net photosynthesis,  $V_{\text{cmax}25}$ ,  $J_{\text{max}25}$ ,  $T_{\text{p}25}$  and  $R_{\text{dark}25}$  [32–39].

In the past, PLSR models have been successfully applied to individual species [32–36]. However, studies exploring more globally-applicable relationships between photosynthetic parameters and reflectance are sparse [37–39]. For tropical forests, some work has been done [39], but has been limited to the use of the spectra-trait approach on a small dataset ( $n = 40$  leaves, 11 species), which also showed low overall model performance ( $R^2 = 0.39$  and  $\text{RMSE} = 36 \mu\text{mol m}^{-2} \text{s}^{-1}$  for  $V_{\text{cmax}25}$ ,  $R^2 = 0.52$  and  $\text{RMSE} = 39 \mu\text{mol m}^{-2} \text{s}^{-1}$  for  $J_{\text{max}25}$  and  $R^2 = 0.48$  and  $\text{RMSE} = 0.52 \mu\text{mol m}^{-2} \text{s}^{-1}$  for  $R_{\text{dark}25}$ ). More recently, the possibility to estimate  $V_{\text{cmax}25}$  from leaf reflectance spectra across 21 tropical tree species was explored and showed much higher model performance ( $R^2 = 0.89$  and  $\text{RMSE} = 6.6 \mu\text{mol m}^{-2} \text{s}^{-1}$ ,  $n = 216$  leaves) [38].

However, to effectively replace or complement the gold standard measurements used to estimate a variety of plant traits, the spectra-trait PLSR approach needs to be accurate across the many axes of variation that include the plant material and environmental conditions that are the target of a given study. Traditional gas exchange methods, while slow, are capable of inferring photosynthetic capacity across a wide-range of plants and environments, and reflectance spectroscopy would need to show similar generality to serve as a viable means for supplementing them. Despite the potential shown in past studies, the possibility of using PLSR models for predicting photosynthetic parameters across plant species and conditions is still an open question. In most cases, PLSR models have been trained on a small number of species and conditions relative to the large number of species present in a tropical forest. For example, the predictive performance of PLSR models for  $V_{\text{cmax}25}$  applied to independent species or conditions was not robust ( $R^2 = 0.23$ ) when compared with application of those models to species and conditions represented in the training dataset ( $R^2 = 0.90$ ) [38].

In order to evaluate the potential to use spectroscopy to enable and accelerate extensive data collection for parameterization of TBMs, the objectives of the study were: (i) to measure the four net photosynthesis parameters across an expanded number of tropical species and conditions to better fill the optical property and trait space, (ii) to test the ability of PLSR models to predict the photosynthetic parameters on external species or sites not included in the training dataset and, (iii) to quantify the uncertainty associated with estimation of our target leaf traits using spectroscopy and traditional gas exchange approaches. To address these objectives, we measured these parameters and reflectance spectra of leaves of a wide variety of species encountered in a tropical forest in Panama at different elevations inside the canopy and at different leaf phenological stages. We increased the size and scope of this database by combining our data with measurements from previous work [38,40] to allow us to capture variation over several years (2012, 2013, 2016, 2017, 2020) and sites in Panama and Brazil, comprising 53 different species.

## Materials and methods

### Study site

Between January and March 2020, we sampled leaves in the area surrounding the Smithsonian Tropical Research Institute (STRI) canopy crane site located in the San Lorenzo Protected Area in the Province of Colon, Republic of Panama ( $9.281^\circ\text{N}$ ,  $79.974^\circ\text{W}$ , 130 m above sea level). This location represents a wet evergreen tropical forest composed of a large diversity of lianas, epiphytes and evergreen trees [41]. The climate in this area is tropical and characterized by a low variation in the mean monthly air temperature ( $25^\circ\text{C}$ ), but a high seasonal variation in precipitation. The rainy season is from May to December with a mean monthly

precipitation of 370 mm. The dry season is from January to April with a mean monthly precipitation of 80 mm.

### Plant material

In this study, we were interested in evaluating the potential for reflectance spectroscopy to accurately estimate photosynthetic and respiratory traits under a wide range of leaf conditions. Therefore, branches were selected across a wide range of species, canopy position and leaf age. Branches were sampled randomly from ten vertical profiles at five different heights, equally distributed from the top of the canopy to the ground. In addition to samples taken from the vertical profiles, other branches from six species (*Brosimum utile*, *Cecropia insignis*, *Guatteria dumetorum*, *Miconia minutiflora*, *Terminalia amazonia* and *Vochysia ferruginea*) representing different growth strategies (early, medium or late successional) were also sampled. In total, branches from 39 species were collected (Table 1). On the branches, when possible, the measurements were made on leaves at different stages of leaf age development, from young to old, by sampling leaves at different positions on the branch and choosing leaves with different colors and texture (young leaves being lighter green and at the tip of the branches, older leaves being further down the branch, thicker and often darker, [42]). Branches were excised from the trees before dawn, immediately put in buckets filled with water and re-cut under water at least 20 cm up from the original cut, toward the tip of the branch to avoid cavitation as described previously [38,43]. Importantly, measurements made on cut branches show no statistically significant differences from *in situ* measurements on intact branches when precautions are taken to wait for acclimation and to avoid cavitation [44–46]. Gas exchange measurements and reflectance spectra were measured in a nearby shaded area.

### Gas exchange measurements

We used five LI-6400XT Portable Photosynthesis Systems and one LI-6800 Portable Photosynthesis System (LI-COR, Lincoln, Nebraska, USA). Instruments were zeroed using a common nitrogen standard at the beginning of the study. Before beginning each measurement, a single leaf was placed in the leaf chamber under stable conditions (irradiance, CO<sub>2</sub>, temperature, flow rate) and allowed to acclimate to these conditions for a minimum of 20 minutes. Measurements of the response of photosynthesis to irradiance and carbon dioxide concentration followed an established protocol [47]. The response to irradiance was measured by sequentially lowering the irradiance as follows: 1800, 1400, 1200, 1000, 800, 600, 400, 300, 200, 120, 80, 50, 30, 20, 10 and finally 0  $\mu\text{mol m}^{-2} \text{s}^{-1}$ . For all measurements, the color spectrum of the light was 90% red, 10% blue. The temperature of the leaves was maintained constant at 30, 31 or 32°C depending on ambient conditions at the time of measurement. The CO<sub>2</sub> concentration at the surface of the leaf was maintained at 400  $\mu\text{mol CO}_2 \text{ mol}^{-1}$ . For the five LI-6400 XTs, the humidity was not controlled and fluctuated with the ambient humidity at the time of the measurement with values generally above 60%. On the LI-6800, the humidity was controlled and set at 70%. After each light response curve, the saturating irradiance was estimated from visual assessment of the light response curve, and the leaf was acclimated at this irradiance for another 20 minutes, allowing the stomatal conductance ( $g_{sw}$ ) and the CO<sub>2</sub> assimilation rate ( $A$ ) to stabilize at the new irradiance. The  $g_{sw}$  and  $A$  were compared against the rates measured prior to starting the light response to ensure that the leaves fully recovered. The CO<sub>2</sub> concentration entering the leaf chamber was then sequentially modified as follows 400, 300, 225, 150, 100, 75, 50, 400, 475, 575, 675, 800, 1000, 1400, 1800  $\mu\text{mol mol}^{-1}$  air. The humidity and temperature settings were controlled in the same manner as in the light response curves. Upon completion of the  $A-C_i$  curve the chamber illumination was turned off and the leaf and branch

Table 1. Number of leaves measured by species in the different datasets.

Species Name	Brazil	Panama (2016 2017)	Panama (2020)
<i>Albizia adinocephala</i>	-	4	-
<i>Anacardium excelsum</i>	-	4	-
<i>Annona spraguei</i>	-	-	1
<i>Apeiba membranacea</i>	-	5	2
<i>Aspidosperma spruceanum</i>	-	-	5
<i>Bixa orellana</i>	-	-	1
<i>Brosimum utile</i>	-	-	11
<i>Calycophyllum candidissimum</i>	-	3	-
<i>Carapa guianensis</i>	-	3	1
<i>Castilla elastica</i>	-	1	-
<i>Cecropia insignis</i>	-	-	17
<i>Cecropia obtusifolia</i>	-	-	1
<i>Cespedesia spathulata</i>	-	-	1
<i>Chamaecrista xinguensis</i>	12	-	-
<i>Clusia rosea</i>	-	-	2
<i>Cordia alliodora</i>	-	4	-
<i>Cupania scrobiculata</i>	-	-	1
<i>Dendropanax arboreus</i>	-	-	2
<i>Erisma uncinatum</i>	24	-	-
<i>Ficus insipida</i>	-	5	-
<i>Guatteria dumetorum</i>	-	23	7
<i>Inga multijuga</i>	-	-	2
<i>Lonchocarpus heptaphyllus</i>	-	-	1
<i>Luehea seemannii</i>	-	5	1
<i>Manilkara bidentata</i>	-	-	1
<i>Manilkara elata</i>	3	-	-
<i>Manilkara zapota</i>	-	-	3
<i>Maranthes panamensis</i>	-	-	1
<i>Marila laxiflora</i>	-	-	1
Melastomaceae family*	-	-	6
<i>Mezilaurus itauba</i>	5	-	-
<i>Miconia minutiflora</i>	-	29	1
<i>Pera arborea</i>	-	-	3
<i>Persea americana</i>	-	-	1
<i>Philodendron fragrantissimum</i>	-	-	3
<i>Philodendron grandipes</i>	-	-	2
<i>Pittoniotis trichantha</i>	-	4	-
<i>Pourouma bicolor</i>	-	-	1
<i>Protium panamense</i>	-	-	2
<i>Salacia multiflora</i>	-	-	1
<i>Sloanea meianthera</i>	-	-	1
<i>Symphonia globulifera</i>	-	-	2
<i>Tachigali cf. chrysophylla</i>	1	-	-
<i>Tachigali versicolor</i>	-	7	2
<i>Tapirira guianensis</i>	-	-	12
<i>Terminalia amazonia</i>	-	21	-
<i>Tocoyena pittieri</i>	-	10	-
<i>Tovomita longifolia</i>	-	-	3

(Continued)

Table 1. (Continued)

Species Name	Brazil	Panama (2016 2017)	Panama (2020)
<i>Tovomita stylosa</i>	-	-	4
<i>Virola elata</i>	-	-	1
<i>Virola multiflora</i>	-	-	1
<i>Vochysia ferruginea</i>	-	17	1
<i>Xylopia macrantha</i>	-	-	2

\* The species for the leaves corresponding to this family could not be identified.

<https://doi.org/10.1371/journal.pone.0258791.t001>

were covered with a black cloth to ensure full darkness. The CO<sub>2</sub> entering the cuvette was set to 400 μmol mol<sup>-1</sup> air and the flow was reduced to 350 μmol s<sup>-1</sup> air to increase the signal to noise ratio. After 45 minutes of dark adaptation and when the CO<sub>2</sub> exchange was stable, the CO<sub>2</sub> efflux from the leaves was recorded every second for one minute and averaged to estimate R<sub>dark</sub>.

### Leaf reflectance spectra measurements

Leaf reflectance was measured shortly after completion of the gas exchange measurements (typically immediately after completion or within one hour) as described previously [30,38], using a full range spectrometer (PSR 3500+, Spectral Evolution, Inc., Lawrence, MA, USA; spectral range: 350–2500 nm; spectral resolution: 2.8 nm at 700 nm, 8 nm at 1500 nm, and 6 nm at 2100 nm) together with an LC-RP-Pro leaf clip foreoptic (Spectra Vista Corporation, Poughkeepsie, New York) containing an internal, full-spectrum calibrated light source. The integration time was set to 2 seconds at the lowest light intensity. A white reference was taken before measuring the leaf reflectance using a white Spectralon® standard. To account for potential heterogeneity of the leaf sample, 3–4 reflectance measurements were taken on different sections of the leaf. They were then averaged to provide one mean leaf spectrum for each leaf.

### Additional data

In order to increase the number of samples and to improve the robustness of the PLSR, we also included the data described in [38,40] with our new data collection to generate one dataset that was used for the analysis described below. The data collection from [38,40] added diversity to our data and included, in addition to measurements made in the same Panamanian forest as described above, samples measured at the drier Parque Natural Metropolitano site, located in Panama, as well as data measured during a previous campaign in Tapajos National Forest, near Santarem, Para, Brazil. Samples from these campaigns included different leaf age classes, from young to old leaves. The leaves collected in Panama [38] were sunlit leaves whereas the data acquired in Brazil [40] also included shaded leaves. In total, 53 species with different growth strategies and shade tolerance were present in the overall dataset, 39 from this study in 2020 and 14 additional species from the previous studies [38,40] (Table 1). The complete instrument output from the original A-C<sub>i</sub> curves presented in [38,40] was used and refitted following the same procedure as the A-C<sub>i</sub> measured in this study. This was done to eliminate any potential bias in the estimation of the parameters that could occur using different equations and fitting methods [3]. The data from [38,40] did not include measurements of R<sub>dark25</sub> so this dataset was only used to improve the PLSR prediction of V<sub>cmax25</sub>, J<sub>max25</sub> and T<sub>p25</sub>.



## Fitting of the photosynthetic parameters from the gas exchange data

From each  $A-C_i$  curve, the parameters  $J_{\max 25}$ ,  $V_{\max 25}$  and  $T_{p25}$  and their confidence intervals were estimated at a reference temperature of 25 °C by fitting the Farquhar model [6] using the modified Arrhenius function published by [11] to correct for the difference between the measurement temperature and the reference temperature. The parameters used to model the temperature dependencies were those summarized in [48] (Table B3, PSN temperature dependencies) using the activation energy parameters from [12,14,49]. The modified Arrhenius function and corresponding parameters were also used to estimate  $R_{\text{dark}25}$  from  $R_{\text{dark}}$ . Our approach assumed an infinite mesophyll conductance and therefore estimates of photosynthetic parameters should be considered apparent. The fitting was done using the ‘mle2’ function from the R package ‘bbmle’ [50,51] which allowed the estimation of all the parameters as well as their standard errors and confidence interval. Note that the transitions between the  $A_c$ ,  $A_j$  and  $A_p$  limitations were not assigned manually but automatically by the fitting procedure so the set of parameters maximized the likelihood of the estimation for each  $A-C_i$  curve.

Triose phosphate utilization limitation ( $A_p$ ) is not always present in  $A-C_i$  curves. To determine if the consideration of  $A_p$  (from which the parameter  $T_{p25}$  is associated) was relevant to improve the quality of the  $A-C_i$  fitting, the Akaike Information Criterion (AIC) of the models with and without  $A_p$  was compared. This criterion was used to compare both models and to assess if the addition of  $A_p$  improved the fit of the model. Ultimately, the model with the lower AIC was kept and, as a result, not all curves have a  $T_{p25}$  parameter value. The code for processing  $A-C_i$  data is available online [52].

It is important to note that in the fitting procedure, a respiration parameter was also fitted. Other methods to estimate the respiration value are often preferred and, in this study, we did not use this value to build the PLSR models. Instead, we used the measurement of the respiration in the dark ( $R_{\text{dark}25}$ ) presented in the section ‘Gas exchange measurements’.

## Development and validation of the spectra-trait PLSR models on different validation datasets

The PLSR models were built for each response variable separately, i.e., for  $V_{\max 25}$ ,  $J_{\max 25}$ ,  $T_{p25}$  and  $R_{\text{dark}25}$ . We followed a double cross validation procedure which uses distinct training and validation datasets to train and then assess the predictive performance of the PLSR models. Importantly, we used special care to define the training and validation datasets to test the generality of the PLSR models and to validate if they could be used to predict observations for species or sites that were different from those used to train the PLSR models. For this purpose, we built three different training and validation datasets with different characteristics allowing us to carry out three different tests of model generality. In the first test, we randomly sampled 80% of all the leaves measured during the three campaigns (i.e., this study campaign and the campaigns from [38,40]) to build the training dataset. The validation dataset consisted of the remaining 20% of the samples. We call this test the ‘random split’ test, hereafter. This test corresponded to the type of validation usually made in previous studies [38,39] and leaves from potentially the same species and even the same plants as those used in the training dataset can be used for validation. In the second test, we randomly sampled 30% of the new species measured in 2020 (i.e., 10 species) to create the validation dataset. All observations made on the other species constituted the training dataset. This test is called the ‘species split’ test hereafter and contrary to the random split test, this test ensures that different species are used to train and validate the PLSR models. In the third test, we separated the data into the training and validation datasets according to the forest site. We used all the observations made in Panama to

build the training dataset and all the observations made in Brazil to build the validation dataset. We call this test the 'site split' test hereafter. This last test was the most challenging in terms of prediction as it involved different spectrometers, different species and a different site from the one used to build the models. This test was only made on  $V_{\text{cmax}25}$  and  $J_{\text{max}25}$  since the data acquired in Brazil did not include  $R_{\text{dark}25}$  and only included a small number of  $T_{\text{p}25}$  observations.

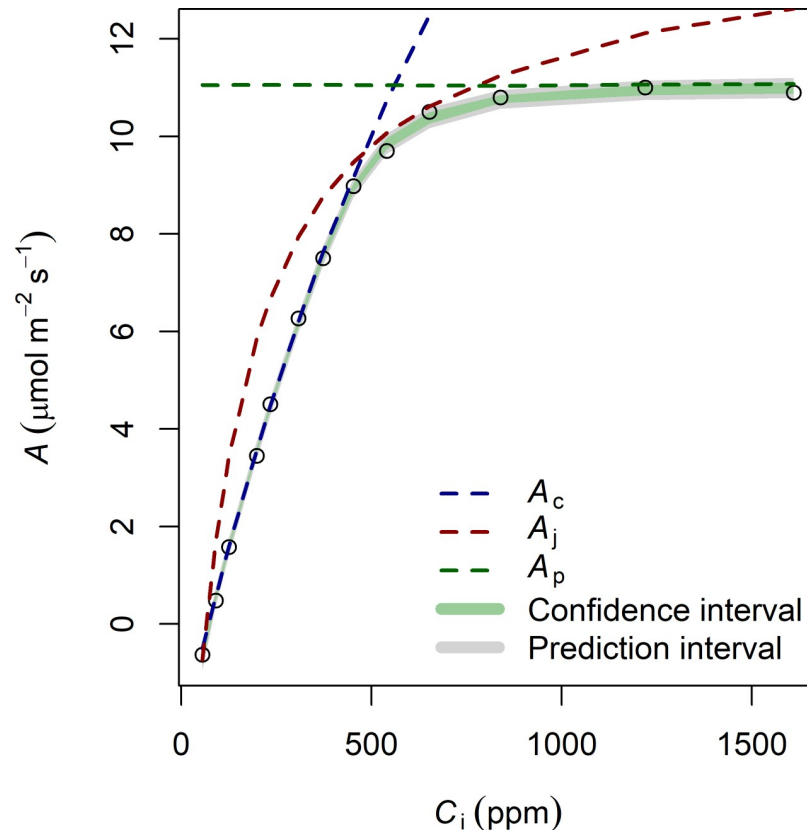
Once the training and validation datasets were created, we used the guide and code from [53] to train and validate the PLSR models. In this guide, two steps are considered: 1) Statistically selecting the most parsimonious number of PLSR components which balances model performance. We did this by creating 1000 random subsets of the training dataset, each one containing 70% of the training observations, and assessing the performance of the sub models on the 30% of the remaining training observations (internal validation). We assessed the performance of each of the 1000 sub models across the selected range of possible PLSR components using the prediction residual sum of square (PRESS). We then identified the minimum number of components as the number which was still less than one standard error away from the overall best model. This method is designed to avoid overfitting, a common issue with PLSR. Given the high number of repetitions (1000), it also reduces the effect of the sub sampling on the determination of the number of components. Note that in order to be consistent between the tests, the number of components for each PLSR variable was chosen during the random split test and was not changed for the species and site split tests. 2) The second step consisted of validating the final calibrated model on the withheld validation dataset (external validation), i.e., comparing the predicted values with the observed values based on the validation dataset observations. To do so, the 1000 sub models were used to predict the response of the validation dataset. The mean of the response could therefore be predicted as well as the confidence and prediction intervals.

Here, we considered the random split test and the PLSR models resulting from this test as the final PLSR models because they included the widest conditions for training and followed the same validation procedure (random split) as previous studies. We added the species split and site split tests to investigate if the data was general enough to train PLSR models that can be used on new species and on a new site. We used three different metrics to quantify the performance of the PLSR obtained by the three tests on their validation datasets: the coefficient of determination ( $R^2$ ), the root mean square error (RMSE), and the RMSE standardized by the trait range across the whole dataset including the three campaigns of measurement (%RMSE). It is important to note that  $V_{\text{cmax}25}$  and  $J_{\text{max}25}$  were square root transformed to increase the symmetry of their distributions before performing the PLSR as advised in [28].

## Interpretation of the PLSR models

In order to find the spectral area that had the greatest importance in the PLSR models to predict the response variables, we used the coefficients of the PLSR models and the VIP (Variable Importance in the Projection) scores [28,54]. The coefficients non significantly different from zero have no effect on the prediction of the response variable, whereas coefficients with high values (negative or positive) have a strong effect, so the 'peaks' in the coefficient plots give information to interpret the PLSR model. We also used the VIP scores as they are complementary to the coefficients [28,54]. Contrary to the coefficients, the VIP scores also consider the importance of each particular wavelength to predict the reflectance spectrum itself, in addition to the effect on the prediction of the response variable. So, in the end, the most important wavelengths have both a high VIP (in practice, above one) and a high coefficient values (positive or negative).





**Fig 1. Example of fitting of an  $A-C_i$  curve.** The observations are represented by points. The Rubisco carboxylation rate ( $A_c$ ), the electron transport rate ( $A_j$ ) and the triose phosphate utilization rate ( $A_p$ ) of net photosynthesis are shown. The three limiting rates of photosynthesis ( $A_c$ ,  $A_j$  and  $A_p$ ) depend on the parameters  $V_{cmax25}$ ,  $J_{max25}$  and  $T_{p25}$ , respectively. The confidence interval of the mean  $A$  is shown in green (i.e., the 95% interval where the mean  $A$  is expected) as well as the interval of prediction in grey (i.e., the 95% interval where a measure of  $A$  should fall).

<https://doi.org/10.1371/journal.pone.0258791.g001>

## Results

### Variation in leaf net photosynthetic traits estimated with gas exchange measurements

An example of our  $A-C_i$  fitting approach is provided in Fig 1. In our approach, we estimated the three potentially limiting processes which determine  $A$  at a given  $CO_2$  concentration: the Rubisco carboxylation rate ( $A_c$ ) which depends on  $V_{cmax25}$ , the electron transport rate ( $A_j$ ) which depends on  $J_{max25}$  and the triose phosphate utilization rate ( $A_p$ ) which depends on  $T_{p25}$ . Our fitting approach also allowed us to calculate the confidence interval of each parameter and to calculate the confidence interval of the mean assimilation rate (i.e., the 95% interval where the mean  $A$  is expected, in green) as well as the prediction interval (i.e., the 95% interval where a measure of  $A$  should fall, in grey). The uncertainty associated with the fitting of the  $A-C_i$  curves and with the dark respiration measurement are presented in Table 2 and are always below 10%.

The standard deviation for the gas exchange measurement corresponds to the average standard deviation from the  $A-C_i$  curve fittings ( $V_{cmax25}$ ,  $J_{max25}$  and  $T_{p25}$ ) or to the average standard deviation from the dark acclimated measurements of the respiration ( $R_{dark25}$ ). The standard deviation for the PLSR predictions corresponds to the average standard deviation of the validation points of the random validation test. Note that the standard deviation for the

Table 2. Uncertainties associated with the gas exchange measurements and the PLSR predictions.

Variable	Standard deviation gas exchange	Relative standard deviation gas exchange	Standard deviation PLSR prediction	Relative standard deviation PLSR prediction
$V_{\text{cmax}25}$	1.49	3.0%	14.0	36.6%
$J_{\text{max}25}$	1.64	1.8%	21.6	28.1%
$T_{\text{p}25}$	0.123	1.9%	1.53	27.8%
$R_{\text{dark}25}$	0.038	8.8%	0.31	47.3%

<https://doi.org/10.1371/journal.pone.0258791.t002>

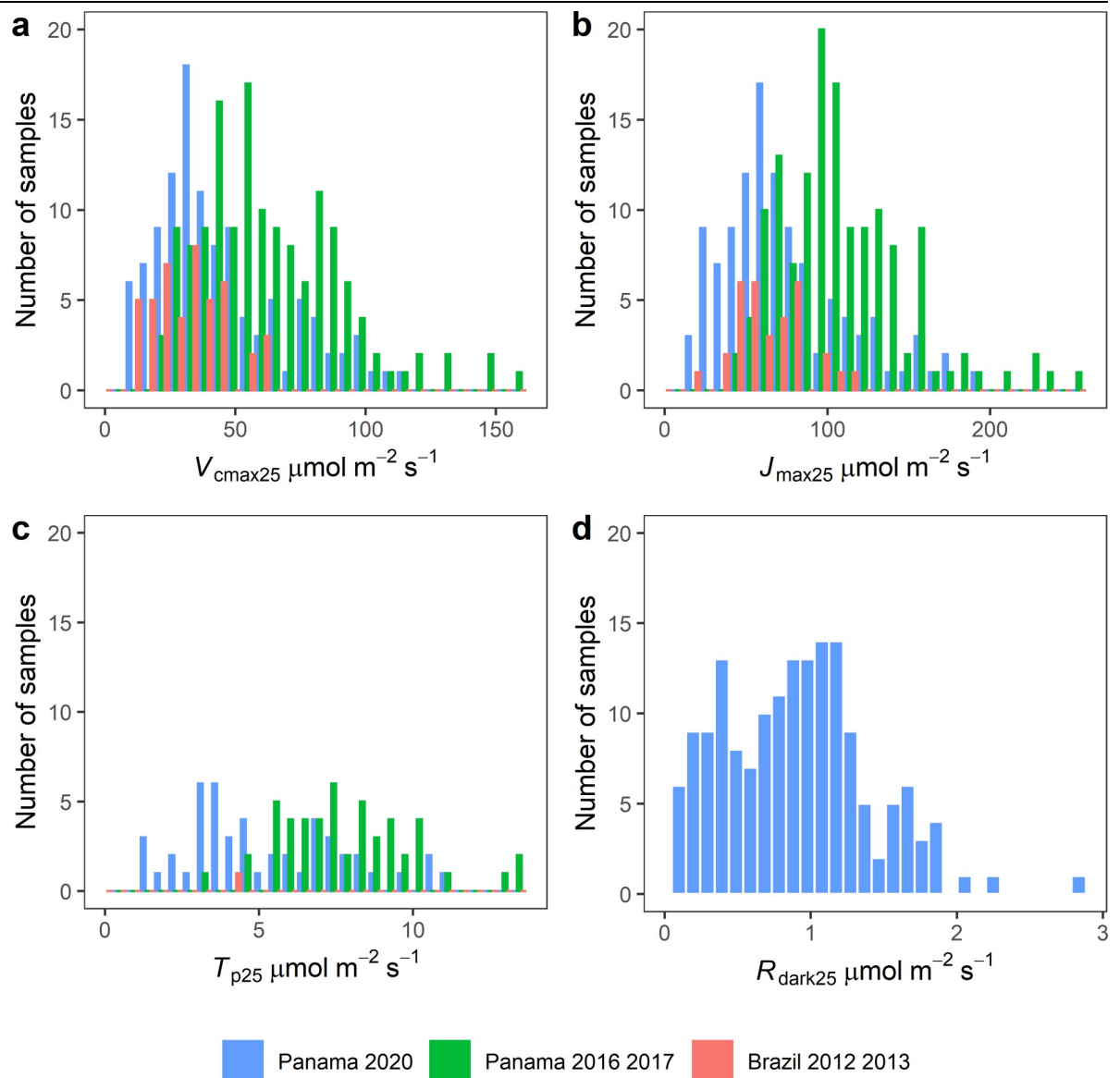
PLSR prediction includes the model uncertainty, i.e., the error associated with the estimation by the 1000 PLSR sub models, and the residual error. The relative standard deviation is the ratio of the standard deviation to the mean.

The resulting distributions of  $V_{\text{cmax}25}$ ,  $J_{\text{max}25}$ ,  $R_{\text{dark}25}$  and  $T_{\text{p}25}$  are shown in Fig 2. These distributions show that our dataset captures a large degree of variation in leaf physiological traits. This variation is not surprising given the nature of the leaf sampling, which was designed to select leaves across a wide range of drivers of leaf trait variation, including species ( $n = 53$  species), different leaf development stages, five different canopy positions on vertical profiles, and three observation site locations in Panama and Brazil. Due to the nature of parameter estimation, we ended up with fewer observations of  $T_{\text{p}25}$  compared with  $V_{\text{cmax}25}$  and  $J_{\text{max}25}$  observations because only 32% of the  $A-C_i$  curves showed evidence of a  $A_p$  limitation, as evaluated using model AIC. This resulted in a lower number of points used to train the  $T_{\text{p}25}$  PLSR model than for  $V_{\text{cmax}25}$ ,  $J_{\text{max}25}$  and  $R_{\text{dark}25}$ . Note that the  $A-C_i$  curves measured in Brazil only showed  $A_p$  limitation for one  $A-C_i$  curve.

### Performance of the PLSR models over the different validation datasets

The performance of PLSR models expressed in terms of  $R^2$ , RMSE and %RMSE for each parameter  $V_{\text{cmax}25}$ ,  $J_{\text{max}25}$ ,  $T_{\text{p}25}$  and  $R_{\text{dark}25}$  are shown in Table 3, as well as the final number of PLSR components. In this table, the performance using the different training and validation datasets (random split, species split and site split) is also shown.

Overall, our validation results for the random split (Table 3) showed good prediction performance, with  $R^2$  above 0.60 for the variables  $V_{\text{cmax}25}$ ,  $J_{\text{max}25}$  and  $T_{\text{p}25}$ , and a lower quality of prediction for  $R_{\text{dark}25}$  ( $R^2 = 0.58$ ). The performance of the models was similar with the species split (Table 3). Generally, the  $R^2$  increased slightly and the RMSE and %RMSE decreased slightly, suggesting that reflectance spectroscopy provides a robust approach for the prediction of photosynthetic parameters on independent species not used to train the PLSR model. Note that the number of observations in the species split validation dataset was smaller than in the random split validation dataset, and that the number of observations in the training dataset was higher which could explain the slightly better performance (Table 3). A change in the partitioning of the number of observations in the training and validation datasets between the random split and the species split was expected given that the number of samples measured for each species was variable and that we did not constrain the choice of the species to populate the validation dataset (Table 1). In the site split test, the RMSE and %RMSE for  $J_{\text{max}25}$  and  $V_{\text{cmax}25}$  decreased slightly as compared to the random split (Table 3). This was despite using data acquired on a different site and different species to train and validate the models. We observed a lower  $R^2$  but given that the range of  $V_{\text{cmax}25}$  and  $J_{\text{max}25}$  at the Brazil site was low ( $10\text{--}62 \mu\text{mol m}^{-2} \text{s}^{-1}$  and  $21\text{--}114 \mu\text{mol m}^{-2} \text{s}^{-1}$ , respectively, Fig 2) compared with the observations from Panama ( $10\text{--}159 \mu\text{mol m}^{-2} \text{s}^{-1}$  and  $17\text{--}254 \mu\text{mol m}^{-2} \text{s}^{-1}$ , respectively, Fig 2) with mostly smaller values, it was expected that a smaller ratio of the total variability could be explained ( $R^2$ ).



**Fig 2. Histogram of the parameters  $V_{\text{cmax}25}$ ,  $J_{\text{max}25}$ ,  $R_{\text{dark}25}$  and  $T_{\text{p}25}$ .**  $V_{\text{cmax}25}$  (a),  $J_{\text{max}25}$  (b),  $R_{\text{dark}25}$  (c) and  $T_{\text{p}25}$  (d) correspond to the fitted values from the A-C<sub>i</sub> curves and to measurements of dark acclimated leaves sampled in a large variety of tropical species, leaf ages and canopy positions.

<https://doi.org/10.1371/journal.pone.0258791.g002>

The observed versus predicted values obtained on the random split test are presented in Fig 3. It is clear that the uncertainty associated with the PLSR model fitting are higher than with the gas exchange measurement. By comparison, the standard deviation of the PLSR predictions (Table 2) were between 8 and 13 times higher than standard deviation obtained with the traditional gas exchange model fitting approach.

### Dominant spectral regions for leaf physiological traits

The PLSR regression coefficients (random split dataset) are provided in Fig 4. In addition, we have overlaid grey shaded bars associated with variable importance of projection (VIP) values greater than one. Overall, these results show that strong peaks are present across the entire (0.5 to 2.4 microns) spectral region, illustrating that our PLSR models utilized spectral information

**Table 3. Description of the PLSR models and their performance on the different validation datasets (random split, species split and site split).**

Variable	Dataset split	N <sub>obs</sub>	N <sub>val</sub>	Transformation	N <sub>comp</sub>	RMSE	Range	%RMSE	R <sup>2</sup>
V <sub>cmax25</sub>	Random	302	59	sqrt	18	13.1	149.3	8.8%	0.74
V <sub>cmax25</sub>	Species	302	34	sqrt	18	11.2	149.3	7.5%	0.86
V <sub>cmax25</sub>	Site	302	43	sqrt	18	8.2	149.3	5.5%	0.66
J <sub>max25</sub>	Random	286	56	sqrt	18	19.8	237.3	8.4%	0.73
J <sub>max25</sub>	Species	286	34	sqrt	18	17.6	237.3	7.4%	0.86
J <sub>max25</sub>	Site	286	30	sqrt	18	16.6	237.3	7.0%	0.38
T <sub>p25</sub>	Random	99	18	-	3	1.52	12.3	12.4%	0.64
T <sub>p25</sub>	Species	99	19	-	3	1.44	12.3	11.7%	0.74
R <sub>dark25</sub>	Random	168	49	-	11	0.27	2.8	9.5%	0.58
R <sub>dark25</sub>	Species	168	12	-	11	0.23	2.8	8.4%	0.50

N<sub>obs</sub> represents the total number of observations in the training and validation datasets. N<sub>val</sub> represents the number of observations in the validation dataset from which the Root Mean Square Error (RMSE), the percent RMSE (%RMSE) and R<sup>2</sup> are calculated. N<sub>comp</sub> is the number of components of the PLSR model. The range corresponds to the difference between the maximum value of the parameter and its minimum over the whole dataset and is used to calculate the %RMSE.

<https://doi.org/10.1371/journal.pone.0258791.t003>

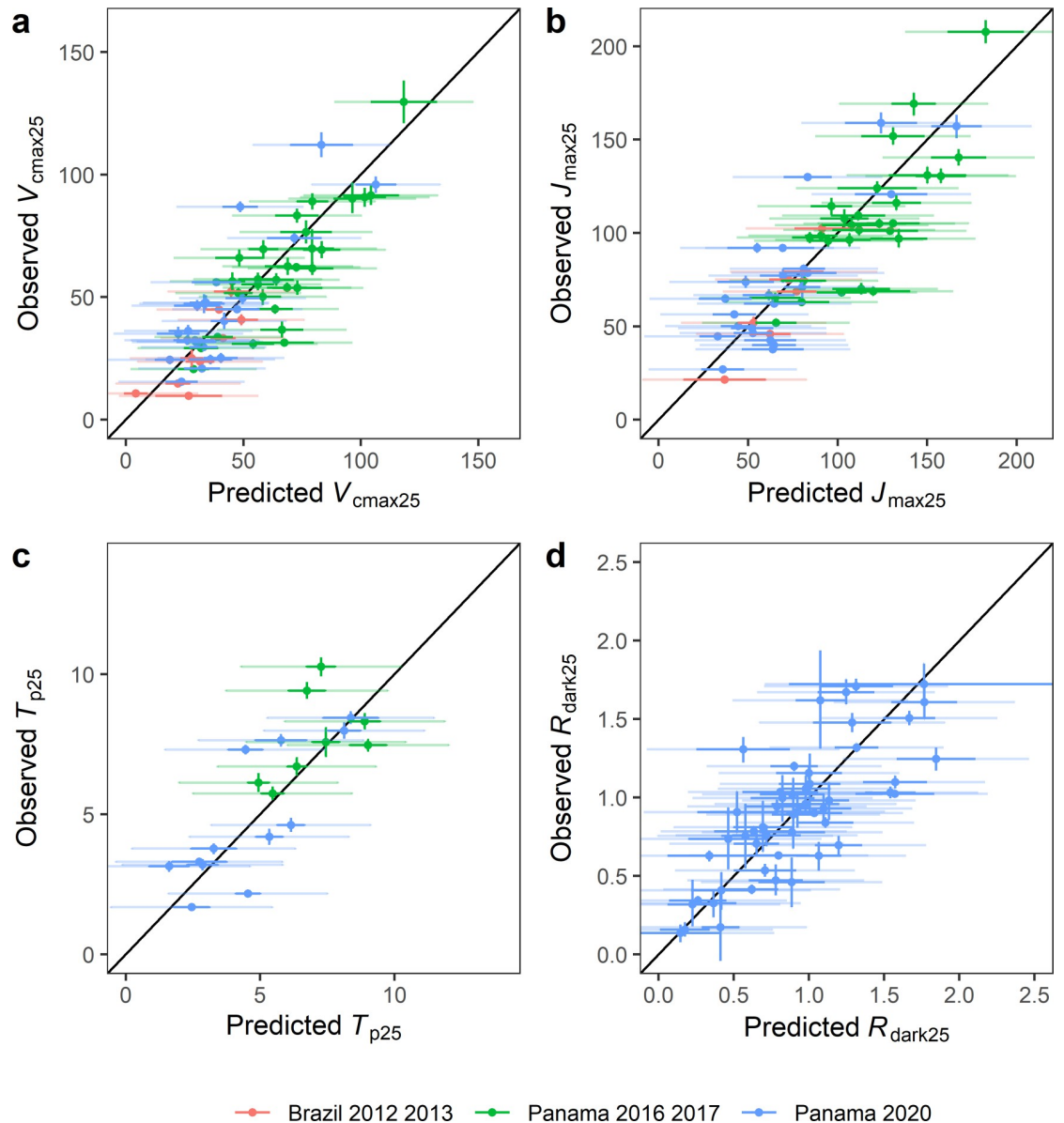
in the visible, near infrared (NIR) and also in the shortwave infrared (SWIR). Clear similarities across traits are also present in the coefficients and VIP values, notably between V<sub>cmax25</sub>, J<sub>max25</sub>, and R<sub>dark25</sub> under 1100 nm and around 1100, 1400 1880 and 2200 nm. We also saw spectral areas where some of the physiological traits show VIP above 1 and others do not. For example, for T<sub>p25</sub>, the VIP in the visible region was below one whereas it was above one for the other traits. Moreover, the coefficients for V<sub>cmax25</sub> and J<sub>max25</sub> were found to show similar patterns, but, despite this strong correspondence (Pearson correlation coefficient = 0.82), we also found significant differences in the models, for example at 550, 670, 1910 and 2155 nm wavelengths (Fig 5).

## Discussion

In this study we combined leaf-level reflectance spectroscopy and traditional gas exchange estimates of four key physiological parameters (J<sub>max25</sub>, V<sub>cmax25</sub>, T<sub>p25</sub> and R<sub>dark25</sub>) used in the FvCB model of leaf net photosynthesis [15]. We explored the ability of reflectance spectroscopy to estimate these parameters across a wide range of biotic and abiotic sources of variation, in three different tropical forest sites over multiple years. We demonstrated that reflectance spectroscopy could be used to estimate these parameters with high accuracy for J<sub>max25</sub>, V<sub>cmax25</sub> and T<sub>p25</sub> and with moderate accuracy for R<sub>dark25</sub>. Importantly and contrary to previous studies in the tropics [38,39], we also showed that spectroscopy could be used to infer the photosynthetic parameters of leaves from species and sites that were not used in the training dataset, showing that the models from this study could potentially be applied pan tropically. These results suggest that spectroscopy and PLSR models could be a viable alternative for collecting much larger physiological datasets using spectral measurements alone, where individual leaf trait estimation accuracy is not the primary requirement. This could significantly enrich trait databases by harnessing spectral measurements to fill in the key observational gaps in tropical forests.

## Comparison of PLSR model performance with other studies

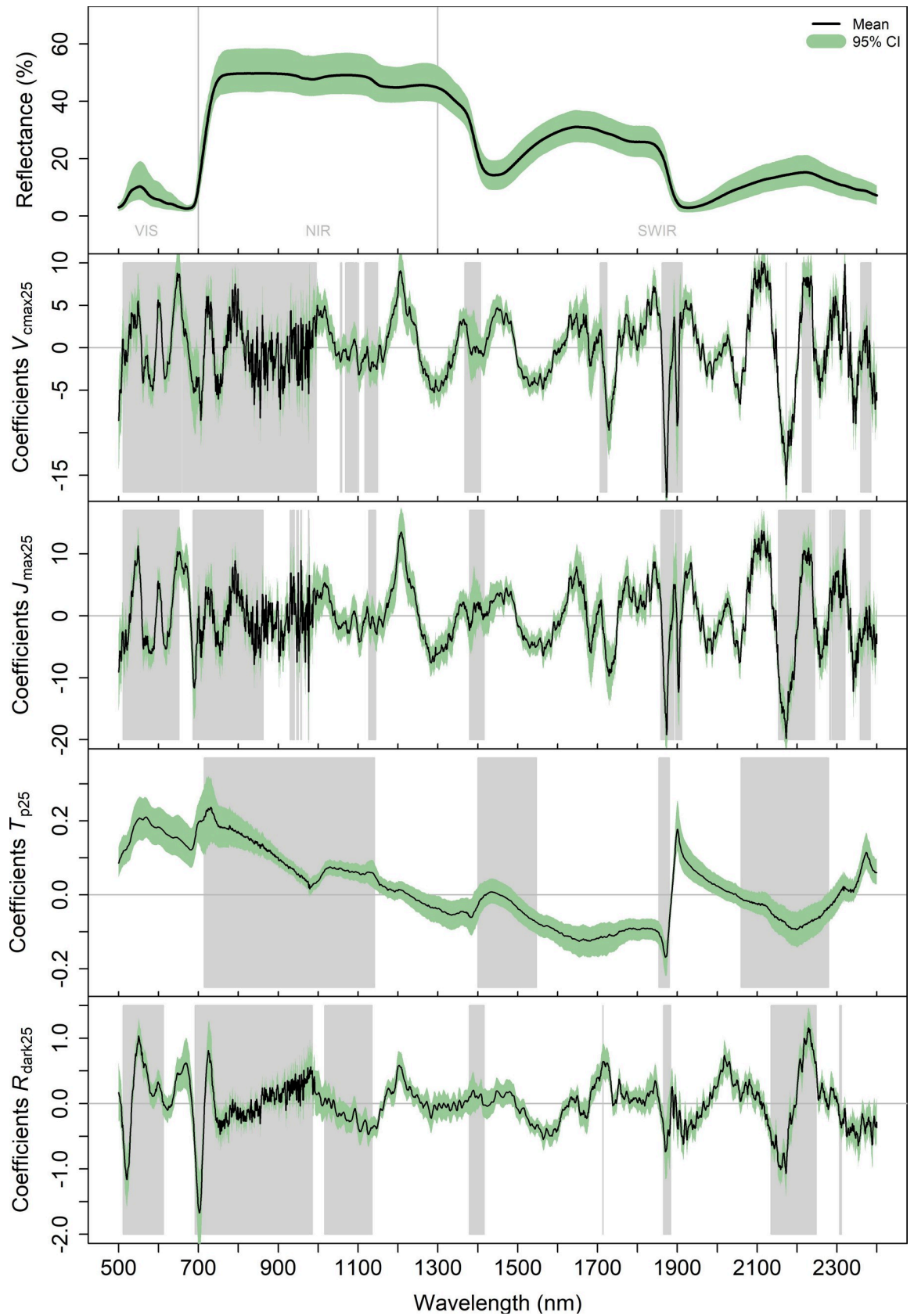
The accuracy of the PLSR models we obtained were comparable with other studies. For example, past work exploring V<sub>cmax25</sub> [37] on 37 temperate deciduous tree species found a R<sup>2</sup> of 0.64 and a RMSE of 17.36 μmol m<sup>-2</sup> s<sup>-1</sup> using 15 PLSR model components. For tropical studies,



**Fig 3. Observed vs predicted value of the core photosynthetic parameters using PLSR models on the random split validation dataset.** The points correspond to the mean value of the variables, the vertical lines correspond to the interval of confidence of the parameters estimated by the fitting of the  $A-C_i$  curves. The horizontal dark lines correspond to the interval of confidence of the PLSR models, and the light lines correspond to the interval of the prediction. The black line corresponds to  $y = x$ .

<https://doi.org/10.1371/journal.pone.0258791.g003>

the model from [38] built on 21 species had a good prediction performance, with a  $R^2$  of 0.89 and low RMSE of  $6.6 \mu\text{mol m}^{-2} \text{s}^{-1}$  using 11 components. This accuracy was obtained when it was applied on the same species as in their training dataset, i.e., similar to our use of random split. That team, however, obtained a lower accuracy when tested on independent species and site ( $R^2 = 0.23$ ;  $\text{RMSE} = 38.8 \mu\text{mol m}^{-2} \text{s}^{-1}$ ). We applied their PLSR model to our data acquired in 2020 and also found a low  $R^2$  ( $R^2 = 0.05$ ), which suggests that the model was not valid outside the conditions of their training dataset, perhaps due to limitations of the spectra or trait data ranges [55]. Comparison of this study with past examples [e.g., 39] both demonstrates the challenges of using empirically-derived PLSR models, but also highlights the value in





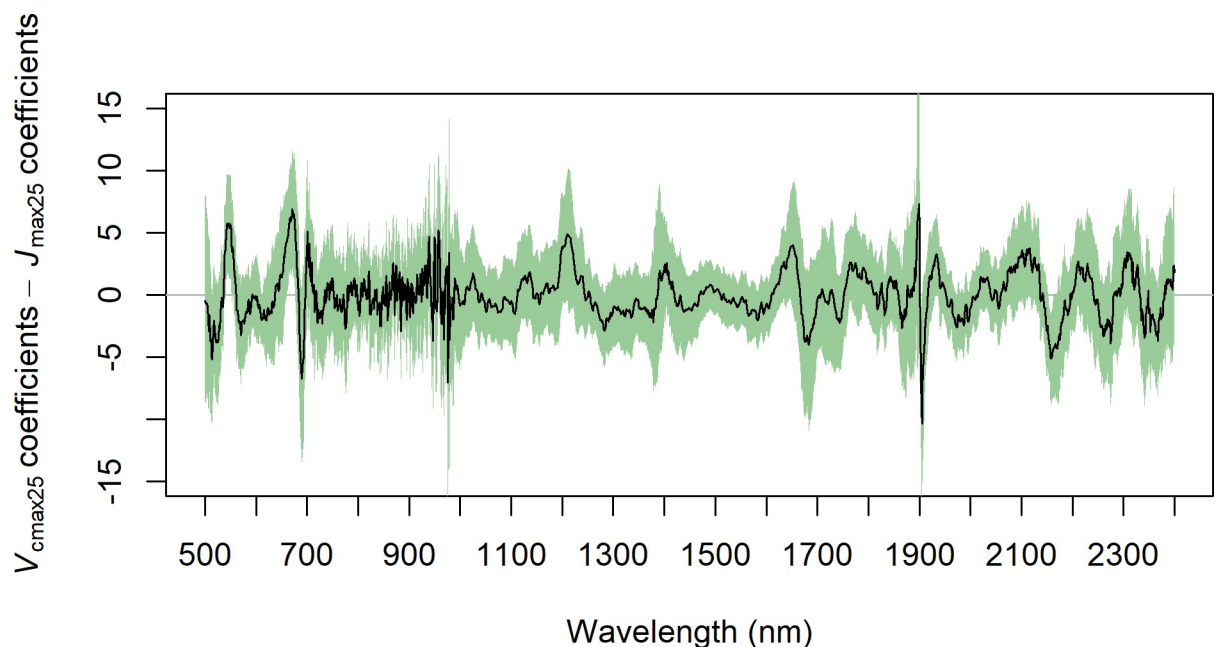
**Fig 4. Coefficients of the PLSR model of  $V_{\text{cmax}25}$ ,  $J_{\text{max}25}$ ,  $R_{\text{dark}25}$  and  $T_{\text{p}25}$ .** The reflectance spectra are given in panel (a) and the coefficients are given below (b, c, d and e). The black curve corresponds to the mean and the green area corresponds to 95% of the values. The grey area corresponds to spectral region where the VIP score is above one.

<https://doi.org/10.1371/journal.pone.0258791.g004>

preserving the underlying data to enable continued model development that can encompass greater variation in leaf traits and ultimately improve model fidelity.

For  $J_{\text{max}25}$ , the quality of prediction and number of components was similar to the model developed by [37] on temperate species who obtained a  $R^2$  of 0.70 and a RMSE of  $27.77 \mu\text{mol m}^{-2} \text{s}^{-1}$  with 20 components. On poplar species, [36] showed high PLSR model accuracy for  $J_{\text{max}25}$  ( $R^2 = 0.93$ ,  $\text{RMSE} = 18.67 \mu\text{mol m}^{-2} \text{s}^{-1}$ ) using 13 components. In addition, other previous work [56] used two-channel spectral vegetation indices to estimate  $J_{\text{max}25}$  of oak trees with good accuracy ( $R^2 = 0.84$ ). Our results are similar but for a dataset that includes a much greater range of biotic and abiotic variation. This suggests the spectroscopy approach is robust and could be adapted to estimate  $J_{\text{max}25}$  of tropical plants.

For  $R_{\text{dark}25}$ , a similar  $R^2$  of 0.56 and RMSE of  $0.14 \mu\text{mol m}^{-2} \text{s}^{-1}$  was reported for wheat leaves [57]. A similar performance with a  $R^2$  of 0.48 was also obtained in the tropics [39] using a relatively small dataset ( $n = 40$ ). The lower quality of prediction for  $R_{\text{dark}25}$  compared to the other photosynthetic parameters of this study was expected due to several methodological limitations. Measurement of the dark respiration via the use of gas exchange systems is known to be imprecise [58]. Such systems measure the flux of  $\text{CO}_2$  and not the  $\text{O}_2$  consumed by the respiration, and since the amount of  $\text{CO}_2$  produced by the dark respiration is low the sensitivity limits of the gas exchange sensors make this measurement less precise. Secondly, the release of  $\text{CO}_2$  in leaves can result from multiples reactions and include several metabolic pathways and enzymes [59] and even transport from other plant organs [60] making the identification of a spectral signature challenging.



**Fig 5. Difference between  $V_{\text{cmax}25}$  and  $J_{\text{max}25}$  PLSR coefficients.** The black line corresponds to the mean of the difference and the green band corresponds to 95% of the values. The areas where the green band does not overlap with the grey horizontal line highlight wavelengths where the difference is statistically different from zero.

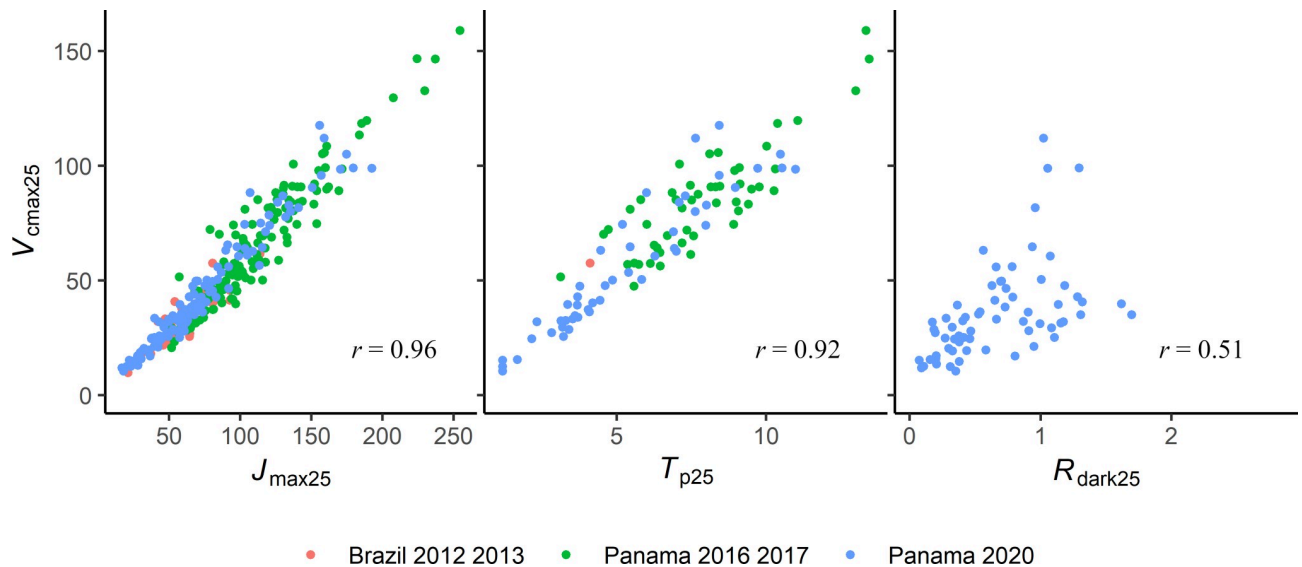
<https://doi.org/10.1371/journal.pone.0258791.g005>

To our knowledge, no previous studies have explored the capacity to estimate  $T_{p25}$  using reflectance spectroscopy. The PLSR for  $T_{p25}$  used a surprisingly low number of components (3) and had good prediction quality ( $R^2$  greater than 0.64), suggesting that the properties of the leaves governing the  $T_{p25}$  could be simpler to detect spectrally than those involved in  $V_{cmax25}$  prediction. It is important to note however that the sample size was lower for  $T_{p25}$  prediction with notably few samples from the Brazil study site (Fig 2) and that more measurements and sites will be necessary to validate this model. The difference in the training dataset could explain a lower number of components than for  $V_{cmax25}$ . In TBMs, the use of the  $A_p$  limitation has been shown to be an unnecessary complexity [61], however a rapid estimation of  $T_{p25}$  could be of interest in other models [62].

### Interpretation of the PLSR models and their ability to predict the net photosynthesis parameters of new species or conditions

We showed in this study that the PLSR models could be used to predict four parameters required to model leaf net photosynthesis in new species not included in the training dataset. The mechanisms behind this capability must be examined to fully understand the limits and potential caveats of those models. The capacity to predict these parameters on new species could be due to two possibilities (assumptions 1 and 2 hereafter). Assumption 1), as with chlorophyll or other pigments [63,64], the PLSR models could specifically identify the spectral signature of the enzymes or molecules involved in the different net photosynthesis reactions ( $A_c$ ,  $A_j$ ,  $A_p$  and  $R_{dark}$ ) and the model thus could allow the prediction of their quantity and therefore the potential rates of the reactions. This would be the most desirable property for the PLSR models as it would be then straightforward to predict net photosynthesis parameters in new leaves. However, the assumption that the PLSR specifically captures the absorption wavelengths of the enzymes and molecules corresponding to each net photosynthesis reaction is unlikely given our results, the processes involved and the PLSR method. Indeed, the coefficients of the PLSR models can be used to understand the spectral features (Figs 4 & 5) that are most influential to the prediction of the response variables. In our study, these spectral features showed very similar shapes for  $V_{cmax25}$  and  $J_{max25}$  and to a lesser extent with  $R_{dark25}$ . This strong similarity advocates against the assumption (1) as these two physiological variables are mechanistically linked with the activity of different enzymes in photosynthesis [9] which have different absorption wavelengths [37,65,66]. Moreover, the specific signals associated with Rubisco ( $V_{cmax}$ ) or, cytochrome b6f ( $J_{max}$ ), are likely to be a minor component of the overall reflectance signal associated with other constituents of the leaves [37]. Assumption 2), a more likely assumption, is that the PLSR captures a range of direct and indirect correlations among a host of leaf traits and properties also known to influence leaf reflectance [67] to enable the estimation of the net photosynthetic parameters. That is, the PLSR method develops a model based on the strongest covariance between the reflectance spectra and response variable so direct or indirect correlations between the response variable and other traits are likely to be equally captured. This would explain the similarity in PLSR coefficients between  $V_{cmax25}$  and  $J_{max25}$  and to a lower extent with  $R_{dark25}$  as the net photosynthesis parameters are known to be strongly correlated in leaves and this was also the case in this study (Fig 6). Note that for  $T_{p25}$  and  $R_{dark25}$  it is less relevant to compare their PLSR coefficients with those of  $V_{cmax25}$  as we did not use the same number of components to model them.

[33,35,37] previously discussed the assumptions (1) and (2). Specifically, they analyzed if the nitrogen and the chlorophyll content, which have a strong spectral signature, were the constituents that enabled prediction of  $V_{cmax25}$  and  $J_{max25}$  with PLSR (assumption 2). They also



**Fig 6. Scatter plot between  $V_{cmax25}$  and  $J_{max25}$ ,  $T_{p25}$  and  $R_{dark25}$ .** The coefficient of correlation ( $r$ ) is indicated on each plot.

<https://doi.org/10.1371/journal.pone.0258791.g006>

analyzed to what extent the PLSR models were also capturing more mechanistic links with the net photosynthesis enzymes (assumption 1).

The fact that PLSR models used to estimate net photosynthesis parameters are at least partially driven by correlations with a host of underlying traits does not minimize their applicability in plant science. Indeed, we empirically showed that our spectra approach was accurate even when applied to new species or sites. We also think that our PLSR models could be applied more generally in the tropics, as there is a strong coordination among the traits of leaves and their function; known as the leaf economic spectrum [68,69]. This strong coordination could explain why a PLSR model trained on a sufficient number of species and conditions allowed that PLSR model to be sufficiently general for other species of the same biome [29,55]. However, the net photosynthesis parameters of species that depart the most from the leaf economic spectrum could be less predictable as would leaves with a transient change in their composition due to a strong stress.

### Utility of the spectra-trait and traditional gas exchange approaches for enhancing tropical plant trait studies

The calculation of the uncertainty of the photosynthetic parameter estimation by the  $A-C_i$  fitting procedure and by the spectroscopy was important in this study as it allowed us to define which applications are compatible with the estimated levels of uncertainty. When using leaf level gas exchange, the precision in the estimation of the photosynthetic parameters was, as expected, very good. When using the PLSR models, prediction uncertainty for an individual leaf was much higher than compared with the leaf gas exchange technique. This was expected, but also, importantly, the accuracy of the PLSR models were similar to previous studies. This limits the use of PLSR models for studies requiring very precise physiological parameter estimates on individual leaves. On the other hand, studies aiming to estimate average parameters over a forest or large population of individuals, like those used to inform TBMs, represent a much more appropriate use of the spectroscopy approach as a means to obtain a large number of estimates quickly (a reflectance measurement takes less than one minute as compared to around two hours for gas exchange measurements). The benefit of the spectra-trait approach

also lies in the ability to estimate a large number of traits from the measurement of the reflectance i.e.  $V_{\text{cmax}25}$ ,  $J_{\text{max}25}$ ,  $T_{\text{p}25}$  and  $R_{\text{dark}25}$  as done in this study but also other traits (nitrogen, chlorophyll, leaf mass per area etc. [29–31]). From past studies, however, it was difficult to take full advantage of this measurement technique because the PLSR predictions of photosynthetic traits were only valid within the trait space included in the training dataset which was relatively narrow compared to the variation in the tropics. The result of this study is important because we showed that our models had the same accuracy outside the training dataset for new species and a new site. In the future, additional efforts will be needed to cover broader axes of variation, associated with, for example, different levels of water, heat and nutrient stress. Therefore, we advocate for the continued combination of traditional leaf gas exchange measurements and spectroscopy to further enable the development of generalized PLSR models that can be applied across a broader range of plant material and environments. We also strongly advocate for the use of standardized data reporting formats [70], and preservation of the underlying data in spectra-trait libraries, such as EcoSIS (<https://ecosis.org/>), which can be used by the community to continually expand and improve trait prediction using PLSR models.

## Supporting information

### S1 File. PLSR coefficients.

(ZIP)

## Acknowledgments

The authors would like to thank Benjamin Dechant for sharing information with us and the authors of the study [40] for having shared their raw  $A-C_i$  data and spectra online.

## Author Contributions

**Conceptualization:** Julien Lamour, Kenneth J. Davidson, Alistair Rogers, Shawn P. Serbin.

**Data curation:** Julien Lamour, Kenneth J. Davidson, Kim S. Ely, Shawn P. Serbin.

**Formal analysis:** Julien Lamour.

**Investigation:** Julien Lamour, Kenneth J. Davidson, Kim S. Ely, Jeremiah A. Anderson, Alistair Rogers, Jin Wu, Shawn P. Serbin.

**Supervision:** Alistair Rogers, Shawn P. Serbin.

**Writing – original draft:** Julien Lamour.

**Writing – review & editing:** Julien Lamour, Kenneth J. Davidson, Kim S. Ely, Alistair Rogers, Jin Wu, Shawn P. Serbin.

## References

1. Mitchard ETA. The tropical forest carbon cycle and climate change. *Nature*. 2018; 559: 527–534. <https://doi.org/10.1038/s41586-018-0300-2> PMID: 30046067
2. Ricciuto D, Sargsyan K, Thornton P. The Impact of Parametric Uncertainties on Biogeochemistry in the E3SM Land Model. *J Adv Model Earth Syst*. 2018; 10: 297–319. <https://doi.org/10.1002/2017ms000962>
3. Rogers A, Medlyn BE, Dukes JS, Bonan G, Von Caemmerer S, Dietze MC, et al. A roadmap for improving the representation of photosynthesis in Earth system models. *New Phytol*. 2017; 213: 22–42. <https://doi.org/10.1111/nph.14283> PMID: 27891647

4. Sullivan MJP, Lewis SL, Affum-Baffoe K, Castilho C, Costa F, Sanchez AC, et al. Long-term thermal sensitivity of Earth's tropical forests. *Science*. 2020; 368: 869. <https://doi.org/10.1126/science.aaw7578> PMID: 32439789
5. Kumarathunge DP, Medlyn BE, Drake JE, Tjoelker MG, Aspinwall MJ, Battaglia M, et al. Acclimation and adaptation components of the temperature dependence of plant photosynthesis at the global scale. *New Phytol*. 2019; 222: 768–784. <https://doi.org/10.1111/nph.15668> PMID: 30597597
6. Farquhar GD, von Caemmerer S, Berry JA. A biochemical model of photosynthetic CO<sub>2</sub> assimilation in leaves of C<sub>3</sub> species. *Planta*. 1980; 149: 78–90. <https://doi.org/10.1007/BF00386231> PMID: 24306196
7. Collatz GJ, Ball JT, Grivet C, Berry JA. Physiological and environmental regulation of stomatal conductance, photosynthesis and transpiration: a model that includes a laminar boundary layer. *Agric For Meteorol*. 1991; 54: 107–136. [https://doi.org/10.1016/0168-1923\(91\)90002-8](https://doi.org/10.1016/0168-1923(91)90002-8)
8. Walker AP, Johnson AL, Rogers A, Anderson J, Bridges RA, Fisher RA, et al. Multi-hypothesis comparison of Farquhar and Collatz photosynthesis models reveals the unexpected influence of empirical assumptions at leaf and global scales. *Glob Change Biol*. 2020. <https://doi.org/10.1111/gcb.15366> PMID: 33037690
9. von Caemmerer S, Farquhar GD, Berry JA. Biochemical model of C<sub>3</sub> photosynthesis In *Photosynthesis in Silico. Understanding Complexity from Molecules to Ecosystems*. Edited by Laissac A., Nedbal L. and Govindjee. Springer, Dordrecht, The Netherlands; 2009.
10. Yin X, Struik PC. C<sub>3</sub> and C<sub>4</sub> photosynthesis models: An overview from the perspective of crop modeling. *NJAS—Wagening J Life Sci*. 2009; 57: 27–38. <https://doi.org/10.1016/j.njas.2009.07.001>
11. Leuning R. Temperature dependence of two parameters in a photosynthesis model. *Plant Cell Environ*. 2002; 25: 1205–1210. <https://doi.org/10.1046/j.1365-3040.2002.00898.x>
12. Bernacchi CJ, Pimentel C, Long SP. In vivo temperature response functions of parameters required to model RuBP-limited photosynthesis. *Plant Cell Environ*. 2003; 26: 1419–1430.
13. McClain AM, Sharkey TD. Triose phosphate utilization and beyond: from photosynthesis to end product synthesis. *J Exp Bot*. 2019; 70: 1755–1766. <https://doi.org/10.1093/jxb/erz058> PMID: 30868155
14. Bernacchi CJ, Singaas EL, Pimentel C, Portis AR Jr, Long SP. Improved temperature response functions for models of Rubisco-limited photosynthesis. *Plant Cell Environ*. 2001; 24: 253–259.
15. Bernacchi CJ, Bagley JE, Serbin SP, Ruiz-Vera UM, Rosenthal DM, Vanlooche A. Modelling C<sub>3</sub> photosynthesis from the chloroplast to the ecosystem. *Plant Cell Environ*. 2013; 36: 1641–1657. <https://doi.org/10.1111/pce.12118> PMID: 23590343
16. Feng X, Dietze M. Scale dependence in the effects of leaf ecophysiological traits on photosynthesis: Bayesian parameterization of photosynthesis models. *New Phytol*. 2013; 200: 1132–1144. <https://doi.org/10.1111/nph.12454> PMID: 23952643
17. Sharkey TD, Bernacchi CJ, Farquhar GD, Singaas EL. Fitting photosynthetic carbon dioxide response curves for C<sub>3</sub> leaves. *Plant Cell Environ*. 2007; 30: 1035–1040. <https://doi.org/10.1111/j.1365-3040.2007.01710.x> PMID: 17661745
18. Xiao Y, Sloan J, Hepworth C, Osborne CP, Fleming AJ, Chen X, et al. Estimating uncertainty: A Bayesian approach to modelling photosynthesis in C<sub>3</sub> leaves. *Plant Cell Environ*. 2021; n/a. <https://doi.org/10.1111/pce.13995> PMID: 33410527
19. Carswell FE, Meir P, Wandelli EV, Bonates LCM, Kruijt B, Barbosa EM, et al. Photosynthetic capacity in a central Amazonian rain forest. *Tree Physiol*. 2000; 20: 179–186. <https://doi.org/10.1093/treephys/20.3.179> PMID: 12651470
20. Lloyd J, Patiño S, Paiva RQ, Nardoto GB, Quesada CA, Santos AJB, et al. Optimisation of photosynthetic carbon gain and within-canopy gradients of associated foliar traits for Amazon forest trees. *Bio-geosciences*. 2010; 7: 1833–1859. <https://doi.org/10.5194/bg-7-1833-2010>
21. Walker AP, Beckerman AP, Gu L, Kattge J, Cernusak LA, Domingues TF, et al. The relationship of leaf photosynthetic traits—V<sub>cmax</sub> and J<sub>max</sub>—to leaf nitrogen, leaf phosphorus, and specific leaf area: a meta-analysis and modeling study. *Ecol Evol*. 2014; 4: 3218–3235. <https://doi.org/10.1002/ece3.1173> PMID: 25473475
22. Dusenge ME, Wallin G, Gvaardesten J, Niyonzima F, Adolfsson L, Nsabimana D, et al. Photosynthetic capacity of tropical montane tree species in relation to leaf nutrients, successional strategy and growth temperature. *Oecologia*. 2015; 177: 1183–1194. <https://doi.org/10.1007/s00442-015-3260-3> PMID: 25694041
23. Wu J, Albert LP, Lopes AP, Restrepo-Coupe N, Hayek M, Wiedemann KT, et al. Leaf development and demography explain photosynthetic seasonality in Amazon evergreen forests. *Science*. 2016; 351: 972–976. <https://doi.org/10.1126/science.aad5068> PMID: 26917771
24. Scafaro AP, Xiang S, Long BM, Bahar NHA, Weerasinghe LK, Creek D, et al. Strong thermal acclimation of photosynthesis in tropical and temperate wet-forest tree species: the importance of altered



- Rubisco content. *Glob Change Biol.* 2017; 23: 2783–2800. <https://doi.org/10.1111/gcb.13566> PMID: 27859952
25. Curran PJ. Remote sensing of foliar chemistry. *Remote Sens Environ.* 1989; 30: 271–278. [https://doi.org/10.1016/0034-4257\(89\)90069-2](https://doi.org/10.1016/0034-4257(89)90069-2)
  26. Jacquemoud S, Baret F. PROSPECT: A model of leaf optical properties spectra. *Remote Sens Environ.* 1990; 34: 75–91. [https://doi.org/10.1016/0034-4257\(90\)90100-z](https://doi.org/10.1016/0034-4257(90)90100-z)
  27. Féret J-B, Gitelson AA, Noble SD, Jacquemoud S. PROSPECT-D: Towards modeling leaf optical properties through a complete lifecycle. *Remote Sens Environ.* 2017; 193: 204–215. <https://doi.org/10.1016/j.rse.2017.03.004>
  28. Wold S, Sjöström M, Eriksson L. PLS-regression: a basic tool of chemometrics. *Chemom Intell Lab Syst.* 2001; 58: 109–130. [https://doi.org/10.1016/S0169-7439\(01\)00155-1](https://doi.org/10.1016/S0169-7439(01)00155-1)
  29. Serbin SP, Wu J, Ely KS, Kruger EL, Townsend PA, Meng R, et al. From the Arctic to the tropics: multi-biome prediction of leaf mass per area using leaf reflectance. *New Phytol.* 2019; 224: 1557–1568. <https://doi.org/10.1111/nph.16123> PMID: 31418863
  30. Ely KS, Burnett AC, Lieberman-Cribbin W, Serbin SP, Rogers A. Spectroscopy can predict key leaf traits associated with source–sink balance and carbon–nitrogen status. *J Exp Bot.* 2019; 70: 1789–1799. <https://doi.org/10.1093/jxb/erz061> PMID: 30799496
  31. Vigneau N, Ecartot M, Rabatel G, Roumet P. Potential of field hyperspectral imaging as a non destructive method to assess leaf nitrogen content in Wheat. *Field Crops Res.* 2011; 122: 25–31. <https://doi.org/10.1016/j.fcr.2011.02.003>
  32. Barnes ML, Breshears DD, Law DJ, Leeuwen WJD van, Monson RK, Fojtik AC, et al. Beyond greenness: Detecting temporal changes in photosynthetic capacity with hyperspectral reflectance data. *PLOS ONE.* 2017; 12: e0189539. <https://doi.org/10.1371/journal.pone.0189539> PMID: 29281709
  33. Meacham-Hensold K, Montes CM, Wu J, Guan K, Fu P, Ainsworth EA, et al. High-throughput field phenotyping using hyperspectral reflectance and partial least squares regression (PLSR) reveals genetic modifications to photosynthetic capacity. *Remote Sens Environ.* 2019; 231: 111176. <https://doi.org/10.1016/j.rse.2019.04.029> PMID: 31534277
  34. Yendrek CR, Tomaz T, Montes CM, Cao Y, Morse AM, Brown PJ, et al. High-Throughput Phenotyping of Maize Leaf Physiological and Biochemical Traits Using Hyperspectral Reflectance. *Plant Physiol.* 2017; 173: 614–626. <https://doi.org/10.1104/pp.16.01447> PMID: 28049858
  35. Wang S, Guan K, Wang Z, Ainsworth EA, Zheng T, Townsend PA, et al. Unique contributions of chlorophyll and nitrogen to predict crop photosynthetic capacity from leaf spectroscopy. Lawson T, editor. *J Exp Bot.* 2020. <https://doi.org/10.1093/jxb/eraa432> PMID: 32937655
  36. Serbin SP, Dillaway DN, Kruger EL, Townsend PA. Leaf optical properties reflect variation in photosynthetic metabolism and its sensitivity to temperature. *J Exp Bot.* 2011; 63: 489–502. <https://doi.org/10.1093/jxb/err294> PMID: 21984647
  37. Dechant B, Cuntz M, Vohland M, Schulz E, Doktor D. Estimation of photosynthesis traits from leaf reflectance spectra: Correlation to nitrogen content as the dominant mechanism. *Remote Sens Environ.* 2017; 196: 279–292. <https://doi.org/10.1016/j.rse.2017.05.019>
  38. Wu J, Rogers A, Albert LP, Ely K, Prohaska N, Wolfe BT, et al. Leaf reflectance spectroscopy captures variation in carboxylation capacity across species, canopy environment and leaf age in lowland moist tropical forests. *New Phytol.* 2019; 224: 663–674. <https://doi.org/10.1111/nph.16029> PMID: 31245836
  39. Doughty CE, Asner GP, Martin RE. Predicting tropical plant physiology from leaf and canopy spectroscopy. *Oecologia.* 2011; 165: 289–299. <https://doi.org/10.1007/s00442-010-1800-4> PMID: 20963611
  40. Albert LP, Wu J, Prohaska N, Camargo PB de, Huxman TE, Tribuzy ES, et al. Age-dependent leaf physiology and consequences for crown-scale carbon uptake during the dry season in an Amazon evergreen forest. *New Phytol.* 2018; 219: 870–884. <https://doi.org/10.1111/nph.15056> PMID: 29502356
  41. Basset Y. Studying Forest Canopies from Above: The International Canopy Crane Network. Smithsonian Tropical Research Institute; 2003. Available: <https://books.google.com/books?id=qyimAAAACAAJ>.
  42. Wu J, Chavana-Bryant C, Prohaska N, Serbin SP, Guan K, Albert LP, et al. Convergence in relationships between leaf traits, spectra and age across diverse canopy environments and two contrasting tropical forests. *New Phytol.* 2017; 214: 1033–1048. <https://doi.org/10.1111/nph.14051> PMID: 27381054
  43. Sperry J. Cutting-edge research or cutting-edge artefact? An overdue control experiment complicates the xylem refilling story. *Plant Cell Environ.* 2013; 36: 1916–1918. <https://doi.org/10.1111/pce.12148> PMID: 23763611



44. Leakey ADB, Bernacchi CJ, Ort DR, Long SP. Long-term growth of soybean at elevated [CO<sub>2</sub>] does not cause acclimation of stomatal conductance under fully open-air conditions. *Plant Cell Environ.* 2006; 29: 1794–1800. <https://doi.org/10.1111/j.1365-3040.2006.01556.x> PMID: 16913868
45. Wolz KJ, Wertin TM, Abordo M, Wang D, Leakey ADB. Diversity in stomatal function is integral to modelling plant carbon and water fluxes. *Nat Ecol Evol.* 2017; 1: 1292–1298. <https://doi.org/10.1038/s41559-017-0238-z> PMID: 29046531
46. Verryckt LT, Van Langenhove L, Ciais P, Courtois EA, Vicca S, Peñuelas J, et al. Coping with branch excision when measuring leaf net photosynthetic rates in a lowland tropical forest. *Biotropica.* 2020; 52: 608–615. <https://doi.org/10.1111/btp.12774>.
47. Rogers A, Serbin SP, Ely KS, Wullschlegel SD. Terrestrial biosphere models may overestimate Arctic CO<sub>2</sub> assimilation if they do not account for decreased quantum yield and convexity at low temperature. *New Phytol.* 2019; 223: 167–179. <https://doi.org/10.1111/nph.15750> PMID: 30767227
48. Bonan GB, Lawrence PJ, Oleson KW, Levis S, Jung M, Reichstein M, et al. Improving canopy processes in the Community Land Model version 4 (CLM4) using global flux fields empirically inferred from FLUXNET data. *J Geophys Res.* 2011; 116. <https://doi.org/10.1029/2010jg001593>
49. Harley PC, Thomas RB, Reynolds JF, Strain BR. Modelling photosynthesis of cotton grown in elevated CO<sub>2</sub>. *Plant Cell Environ.* 1992; 15: 271–282. <https://doi.org/10.1111/j.1365-3040.1992.tb00974.x>.
50. Bolker B, R Development Core Team. *bbmle: Tools for General Maximum Likelihood Estimation.* 2020. Available: <https://CRAN.R-project.org/package=bbmle>.
51. R Core Team. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing; 2020. Available: <https://www.R-project.org/>.
52. Lamour Julien, Serbin SP. TESTgroup-BNL/LeafGasExchange: Updated package release. Zenodo; 2021. <https://doi.org/10.5281/zenodo.4545818>
53. Burnett AC, Anderson J, Davidson KJ, Ely KS, Lamour J, Li Q, et al. A best-practice guide to predicting plant traits from leaf-level hyperspectral data using partial least squares regression. *J Exp Bot.* 2021 [cited 26 Jul 2021]. <https://doi.org/10.1093/jxb/erab295> PMID: 34131723
54. Chong I-G, Jun C-H. Performance of some variable selection methods when multicollinearity is present. *Chemom Intell Lab Syst.* 2005; 78: 103–112. <https://doi.org/10.1016/j.chemolab.2004.12.011>
55. Schweiger AK. *Spectral Field Campaigns: Planning and Data Collection.* In: Cavender-Bares J, Gamon JA, Townsend PA, editors. *Remote Sensing of Plant Biodiversity.* Cham: Springer International Publishing; 2020. pp. 385–423. [https://doi.org/10.1007/978-3-030-33157-3\\_15](https://doi.org/10.1007/978-3-030-33157-3_15)
56. Stylinski CD, Oechel WC, Gamon JA, Tissue DT, Miglietta F, Raschi A. Effects of lifelong [CO<sub>2</sub>] enrichment on carboxylation and light utilization of *Quercus pubescens* Willd. examined with gas exchange, biochemistry and optical techniques. *Plant Cell Environ.* 2000; 23: 1353–1362. <https://doi.org/10.1046/j.1365-3040.2000.00638.x>
57. Coast O, Shah S, Ivakov A, Gaju O, Wilson PB, Posch BC, et al. Predicting dark respiration rates of wheat leaves from hyperspectral reflectance. *Plant Cell Environ.* 2019; 42: 2133–2150. <https://doi.org/10.1111/pce.13544> PMID: 30835839
58. Duursma RA, Blackman CJ, Lopéz R, Martin-StPaul NK, Cochard H, Medlyn BE. On the minimum leaf conductance: its role in models of plant water use, and ecological and environmental controls. *New Phytol.* 2019; 221: 693–705. <https://doi.org/10.1111/nph.15395> PMID: 30144393
59. O'Leary BM, Asao S, Millar AH, Atkin OK. Core principles which explain variation in respiration across biological scales. *New Phytol.* 2019; 222: 670–686. <https://doi.org/10.1111/nph.15576> PMID: 30394553
60. Teskey RO, Saveyn A, Steppe K, McGuire MA. Origin, fate and significance of CO<sub>2</sub> in tree stems. *New Phytol.* 2008; 177: 17–32. <https://doi.org/10.1111/j.1469-8137.2007.02286.x> PMID: 18028298
61. Rogers A, Kumarathunge DP, Lombardozi DL, Medlyn BE, Serbin SP, Walker AP. Triose phosphate utilization limitation: an unnecessary complexity in terrestrial biosphere model representation of photosynthesis. *New Phytol.* 2021; 230: 17–22. <https://doi.org/10.1111/nph.17092> PMID: 33217768
62. Gregory LM, McClain AM, Kramer DM, Pardo JD, Smith KE, Tessmer OL, et al. The triose phosphate utilization limitation of photosynthetic rate: Out of global models but important for leaf models. *Plant Cell Environ.* 2021;n/a. <https://doi.org/10.1111/pce.14153> PMID: 34278582
63. Ustin SL, Gitelson AA, Jacquemoud S, Schaepman M, Asner GP, Gamon JA, et al. Retrieval of foliar information about plant pigment systems from high resolution spectroscopy. *Imaging Spectrosc Spec Issue.* 2009; 113: S67–S77. <https://doi.org/10.1016/j.rse.2008.10.019>
64. Richardson AD, Duigan SP, Berlyn GP. An evaluation of noninvasive methods to estimate foliar chlorophyll content. *New Phytol.* 2002; 153: 185–194. <https://doi.org/10.1046/j.0028-646x.2001.00289.x>
65. Elvidge CD. Visible and near infrared reflectance characteristics of dry plant materials. *Int J Remote Sens.* 1990; 11: 1775–1795. <https://doi.org/10.1080/01431169008955129>

66. Nelson N, Neumann J. Isolation of a Cytochrome b6-f Particle from Chloroplasts. *J Biol Chem.* 1972; 247: 1817–1824. [https://doi.org/10.1016/s0021-9258\(19\)45547-6](https://doi.org/10.1016/s0021-9258(19)45547-6) PMID: 5012764
67. Serbin SP, Townsend PA. Scaling Functional Traits from Leaves to Canopies. *Remote Sensing of Plant Biodiversity.* Springer International Publishing; 2020. pp. 43–82. [https://doi.org/10.1007/978-3-030-33157-3\\_3](https://doi.org/10.1007/978-3-030-33157-3_3)
68. Wright IJ, Reich PB, Westoby M, Ackerly DD, Baruch Z, Bongers F, et al. The worldwide leaf economics spectrum. *Nature.* 2004; 428: 821–827. <https://doi.org/10.1038/nature02403> PMID: 15103368
69. Chen X, Sun J, Wang M, Lyu M, Niklas KJ, Michaletz ST, et al. The Leaf Economics Spectrum Constrains Phenotypic Plasticity Across a Light Gradient. *Front Plant Sci.* 2020; 11: 735. <https://doi.org/10.3389/fpls.2020.00735> PMID: 32595665
70. Ely KS, Rogers A, Agarwal DA, Ainsworth EA, Albert LP, Ali A, et al. A reporting format for leaf-level gas exchange data and metadata. *Ecol Inform.* 2021; 61: 101232. <https://doi.org/10.1016/j.ecoinf.2021.101232>