



ELSEVIER

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

High-throughput field phenotyping using hyperspectral reflectance and partial least squares regression (PLSR) reveals genetic modifications to photosynthetic capacity

Katherine Meacham-Hensold^{a,b}, Christopher M. Montes^a, Jin Wu^{c,d}, Kaiyu Guan^{e,f}, Peng Fu^b, Elizabeth A. Ainsworth^{a,g}, Taylor Pederson^b, Caitlin E. Moore^b, Kenny Lee Brown^h, Christine Raines^h, Carl J. Bernacchi^{a,b,g,*}

^a Department of Plant Biology, University of Illinois at Urbana-Champaign, USA

^b Carl R Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, USA

^c Environmental & Climate Science Department, Brookhaven National Laboratory, Upton, New York, USA

^d School of Biological Sciences, University of Hong Kong, Pokfulam, Hong Kong

^e Department of Natural Resources and Environmental Sciences, University of Illinois at Urbana-Champaign, IL, USA

^f National Center of Supercomputing Applications, University of Illinois at Urbana-Champaign, USA

^g USDA ARS Global Change and Photosynthesis Research Unit, Urbana, IL, USA

^h Department of Biological Sciences, University of Essex, Colchester CO4 3SQ, United Kingdom

ARTICLE INFO

Keywords:

Hyperspectral reflectance
Partial least squares regression (PLSR)
Photosynthesis
Leaf nitrogen
Food security
Gas exchange
Spectroscopy

ABSTRACT

Spectroscopy is becoming an increasingly powerful tool to alleviate the challenges of traditional measurements of key plant traits at the leaf, canopy, and ecosystem scales. Spectroscopic methods often rely on statistical approaches to reduce data redundancy and enhance useful prediction of physiological traits. Given the mechanistic uncertainty of spectroscopic techniques, genetic modification of plant biochemical pathways may affect reflectance spectra causing predictive models to lose power. The objectives of this research were to assess over two separate years, whether a predictive model can represent natural and imposed variation in leaf photosynthetic potential for different crop cultivars and genetically modified plants, to assess the interannual capabilities of a partial least square regression (PLSR) model, and to determine whether leaf N is a dominant driver of photosynthesis in PLSR models. In 2016, a PLSR analysis of reflectance spectra coupled with gas exchange data was used to build predictive models for photosynthetic parameters including maximum carboxylation rate of Rubisco ($V_{c,max}$), maximum electron transport rate (J_{max}) and percentage leaf nitrogen ([N]). The model was developed for wild type and genetically modified plants that represent a wide range of photosynthetic capacities. Results show that hyperspectral reflectance accurately predicted $V_{c,max}$, J_{max} and [N] for all plants measured in 2016. Applying these PLSR models to plants grown in 2017 resulted in a strong predictive ability relative to gas exchange measurements for $V_{c,max}$, but not for J_{max} and not for genotypes unique to 2017. Building a new model including data collected in 2017 resulted in more robust predictions, with R^2 increases of 17% for $V_{c,max}$ and 13% J_{max} . Plants generally have a positive correlation between leaf nitrogen and photosynthesis, however, tobacco with reduced Rubisco (SSuD) had significantly higher [N] despite much lower $V_{c,max}$. The PLSR model was able to accurately predict both lower $V_{c,max}$ and higher leaf [N] for this genotype suggesting that the spectral based estimates of $V_{c,max}$ and leaf nitrogen [N] are independent. These results suggest that the PLSR model can be applied across years, but only to genotypes used to build the model and that the actual mechanism measured with the PLSR technique is not directly related to leaf [N]. The success of the leaf-scale analysis suggests that similar approaches may be successful at the canopy and ecosystem scales but to use these methods across years and between genotypes at any scale, application of accurately populated physical based models based on radiative transfer principles may be required.

* Corresponding author at: USDA-ARS Global Change and Photosynthesis Research Unit, 1201 W. Gregory Drive, Urbana, IL 61801, USA.

E-mail address: bernacchi@illinois.edu (C.J. Bernacchi).

<https://doi.org/10.1016/j.rse.2019.04.029>

Received 14 February 2019; Received in revised form 16 April 2019; Accepted 27 April 2019

0034-4257/ © 2019 Published by Elsevier Inc.

1. Introduction

Projected population increases, rising global affluence, and mounting pressures from a changing global climate necessitate improvements to global food supply (Tilman et al., 2009; Foley et al., 2011). Yield increases brought about from the ‘Green Revolution’ have plateaued over the last two decades for many crop species (Fischer and Edmeades, 2010; Parry et al., 2010), requiring novel strategies to realize further gains in productivity. Improving photosynthetic potential likely remains the best strategy to increase crop production (Monteith and Moss, 1977; Zhu et al., 2010; Ort et al., 2015), potentially without the need for additional fertilizer and pesticide that were critical to support yield increases associated with the Green Revolution (Long et al., 2006; Evans, 2013).

Despite being one of nature's most conserved processes, photosynthesis has a staggering number of component inefficiencies (Long et al., 2006; Evans, 2013). These inefficiencies inspire current research efforts to improve crop yields through manipulating photosynthetic pathways (Ort et al., 2015; Andralojc et al., 2018) and exploiting natural variation in photosynthetic rates (Lawson et al., 2012; Meacham et al., 2016). Regardless of the means of improvement, the ability to non-destructively sample phenotypic variation in photosynthetic capacity among tens to hundreds of thousands of plants representing genotypic variation within a reasonable time presents a significant phenotyping challenge (Furbank and Tester, 2011).

Remote sensing communities have long used spectral vegetation indices to estimate leaf and canopy properties at the ecosystem scale (Peñuelas et al., 1994; Curran et al., 1997; Martin and Aber, 1997; Ustin et al., 2004; Asner and Martin, 2008). More recently, spectral detection of emission at discreet wavebands corresponding to solar induced fluorescence (SIF) (Frankenberg and Berry, 2018) have been used as a functional proxy for gross primary productivity (GPP) of natural and forest ecosystems (Guanter et al., 2007; Meroni et al., 2009; Yang et al., 2018; Guan et al., 2016) and crops (Miao et al., 2018; Yang et al., 2018). While SIF is valuable in the context of inferring GPP, it does not provide insights into the underlying photosynthetic mechanisms, i.e., maximum Rubisco carboxylation ($V_{c,max}$) and maximum electron transport rate (J_{max}), that are used as indicators of photosynthetic capacity and to model vegetation at leaf- to ecosystem-scales (e.g., Bernacchi et al., 2013; Bagley et al., 2015a,b). Improving photosynthetic productivity for increased global crop yields requires techniques to quantify these parameters, yet traditional methods rely on leaf sampling and analysis under laboratory conditions or using in-field gas exchange systems (Long and Bernacchi, 2003). These provide a wealth of photosynthetic information but are costly and time intensive. However, spectroscopy techniques coupled with regression analysis (Serbin et al., 2012, 2014) have been used to screen for germplasm among cropping species with the highest photosynthetic potential (Ainsworth et al., 2014; Yendrek et al., 2016; Silva-Perez et al., 2017). Because of the challenges with gas exchange measurements, remote sensing of intensive agricultural regions of the planet is limited by ground-truth data and the potential for genetically modified crops to increase in areal extent suggests methods are required that quantify $V_{c,max}$ and J_{max} over larger spatial areas, beyond what is capable with traditional techniques.

Statistical approaches that link spectral reflectance patterns with ‘ground-truth’ measurements from traditional techniques have potential to significantly decrease sampling time by orders of magnitude. The partial least squares regression (PLSR) model (Wold et al., 2001), which relates two data matrices using a linear multivariate model to predict plant properties of interest, is increasingly used for this purpose. This approach has been applied to rapidly collected leaf reflectance spectra and used to predict key photosynthetic parameters in Aspen and Cottonwood trees (Serbin et al., 2012), soybean exposed to ozone treatments (Ainsworth et al., 2014), wheat (Silva-Perez et al., 2017) and maize (Yendrek et al., 2016; Heckmann et al., 2017).

Before combined gas exchange and aerial spectroscopic techniques

can be employed to quantify ecosystem function to feed ecosystem models or to be used as breeding tools, the combination of these techniques must be tested at scales that eliminate confounding factors. Leaf-scale analysis using spectral sensors with an artificial light source on field grown plants, provides an ideal testbed for spectroscopic techniques as it removes many of the issues with spectral measurements at larger scales. While hyperspectral analysis has focused on inter- and intra-specific variation in photosynthetic potential (Ainsworth et al., 2014; Silva-Perez et al., 2017; Yendrek et al., 2016; Heckmann et al., 2017), there is considerable uncertainty whether this technique can be applied to plants in which the photosynthetic machinery has been genetically modified. Genetic modifications can range from optimizing concentrations of existing proteins (e.g., Driever et al., 2017; López-Calcano et al., 2019) to metabolic engineering of novel pathways (e.g., South et al., 2019). As the underlying mechanisms measured with PLSR techniques remain elusive, artificially altering the amount, or introducing novel proteins and pathways may challenge the predictive ability of these models. Estimations of photosynthetic physiology have also been attributed to leaf nitrogen content ([N]) as calculated from key light absorption pigments in the reflectance spectra (Kattge et al., 2009; Rogers, 2014; Walker et al., 2014; Dechant et al., 2017). This also questions whether altered photosynthetic potential independent of changes in leaf [N] would lead to an inability of spectroscopic analysis of photosynthetic potential in genetically modified plants. If leaf-scale spectroscopic techniques are unable to accurately identify natural and/or imposed variation in photosynthetic potential among a diverse collection of crop genotypes, then it would question the potential for this technique at larger spatial scales relevant for breeding and modeling purposes.

In this study a PLSR model is used to predict photosynthetic capacity from leaf hyperspectral reflectance in field grown wild-type cultivars and genetically modified lines of *Nicotiana tabacum* (tobacco) over multiple time periods over two growing seasons. Specifically, the objectives are to determine whether (1) PLSR based spectral models predict photosynthetic capacity in genetically modified plants, (2) the PLSR model can be applied across growing seasons, and (3) PLSR can predict $V_{c,max}$ and J_{max} independent of leaf nitrogen. Tobacco was chosen as a model crop species to test the effectiveness of modifications to the photosynthetic pathway based on the ease of genetic transformation, short growing seasons, and large number of seeds produced (Kromdijk et al., 2016). This allows rapid field trial testing prior to insertion of promising modifications into staple food crops. These objectives were tested on wild type cultivars exhibiting natural variation in photosynthetic capacities and genotypes genetically modified to present increased and decreased photosynthetic potential (Table 1).

2. Methods

2.1. Plant material

In 2016, six genotypes consisting of three transgenic and three wild type lines of *Nicotiana tabacum* (Table 1) were grown under field conditions at the University of Illinois Energy Farm Facility in Urbana, Illinois (40°03'46.4"N 88°12'25.4"W, 215 m above sea level). Genotypes were chosen to exhibit variation in photosynthetic capacity using the three wild type cultivars representing different relative growth rates, two transgenic Rubisco antisense lines with reduced photosynthetic capacity (Hudson et al., 1992), and one transgenic with overexpression of photosynthetic carbon reduction cycle enzymes to increase photosynthetic capacity (Simkin et al., 2015). Plants were germinated in greenhouse conditions and transplanted to the field at the four-leaf stage. High levels of ESN Smart Nitrogen (310 kg/ha, equating to ~150 ppm soil concentration) were applied to the field site two weeks prior to transplanting. A biological pesticide *Bacillus thuringiensis v. kurstaki* (54%) (DiPel PRO), was applied to the prepared field site five days prior to transplant and at biweekly intervals thereafter to control

Table 1
Nicotiana tabacum genotypes used in this study and brief description of transgenic modification, with reference for detailed description of transformation.

Year(s) grown	Genotype	Transgene	Transgene expected function
2016 & 2017	Petite Havana	None (WT)	n/a
2016 & 2017	Samsun	None (WT)	n/a
2016 & 2017	Mammoth	None (WT)	n/a
2016	SFX	Overexpressed photosynthetic carbon reduction cycle enzymes, background: Samsun (Simkin et al., 2015)	Improved photosynthetic capacity, due to increased carbon reduction enzymes.
2016 & 2017	Single Rubisco Knockdown (SSuS)	Rubisco small subunit antisense. 40% of WT Rubisco, background: W38 (Hudson et al., 1992)	Reduced photosynthetic capacity, due to reduced Rubisco
2016 & 2017	Double Rubisco Knockdown (SSuD)	Rubisco small subunit antisense. 10% of WT Rubisco, background: W38 (Hudson et al., 1992)	Reduced photosynthetic capacity, due to reduced Rubisco
2017	200–8	Insertion of two transgenic genes expressing the enzyme Glycolate dehydrogenase and Malate synthase as an alternative pathway to native photorespiration, background: Petite Havana (South et al., 2019)	Increased photosynthetic capacity, by reduction of energy loss associated with photorespiration.
2017	43-OE	Increased PsbS mRNA levels from transformation with <i>Nicotiana benthamiana</i> Psbs coding sequence and 35S promoter, background: Petite Havana (Głowacka et al., 2016, 2018)	Increased photosynthetic capacity, due to increase in the electron transport metabolite pools.
2017	4-KO	Decreased PsbS mRNA levels from transformation with <i>Nicotiana benthamiana</i> Psbs coding sequence and 35S promoter, background: Petite Havana (Głowacka et al., 2016, 2018)	Reduced photosynthetic capacity, due to decreased electron transport metabolite pools.

for tobacco pests. A broad action herbicide, Glyphosate-isopropylammonium (41%) (Killzall; VPG) (15 l at 70 g/l) was applied to all plots once, two days prior to transplanting. Irrigation was provided as needed to eliminate water limitation throughout growth. The experiment consisted of four replicated plots of each genotype with 36 plants per plot arranged in a 6 × 6 grid and spaced 0.38 m apart.

In 2017, the SFX genotype was removed from the experiment and three transgenic lines were added (Table 1). Plants were grown at the same location following the same protocol as 2016. Two newly introduced lines had either reduced or increased photoprotective quenching capacity, and the third had an alternative photorespiratory pathway relative to the wild-type. Field set up, plot design, pesticide and nutrient application followed the same protocol as 2016.

2.2. Leaf reflectance and gas exchange

2.2.1. Leaf reflectance

Leaf spectral reflectance was measured in situ from 400 to 2500 nm using a spectroradiometer (*Fieldspec4, Analytical Spectral Devices - ASD, Boulder, CO USA*), with spectral resolution of 3 nm in the visible and NIR (350–1000 nm) and 8 nm in shortwave-infrared (SWIR; 1000–2500 nm). Measurements were made with a leaf clip attached to the fiberoptic cable. The device contains a radiometrically calibrated light source which was standardized for relative reflectance (white reference) prior to each measurement using a spectralon panel. The last fully expanded leaf on each plant was measured, always keeping its natural orientation, avoiding leaf midrib and leaf edges. Each measurement was the mean of 10 scans at 100 ms scanning speed per scan. Six reflectance spectra were recorded using the leaf clip attachment in different regions of the same leaf, and a total of three leaves were sampled per plot.

A spectral splice correction was applied to each spectrum to align the VIS and SWIR sensor to the NIR sensor, and a bias threshold removed spectra with high light levels at 450 nm as a quality control to ensure the leaf clip was properly fastened onto the leaf during each measurement. The six spectra for a single leaf were then averaged to give a mean spectrum per leaf. Spectra from the six samples with a deviation from the mean greater than 2% reflectance were eliminated using the *FieldSpectra* package in R according to Serbin et al. (2014). Leaves with a remaining number of viable spectra less than 4 repetitions were eliminated from analysis.

Plants were measured at multiple developmental stages to capture a wide variation in maturity and in meteorological conditions for each genotype. Measurements were collected during three date ranges in 2016: June 30 – July 1 (T1), July 19–21 (T2), and August 4–5 (T3). In

2017 measurements were made on four date ranges: June 26–28 (T1), July 6–12 (T2), July 31–August 1 (T3) and August 18 (T4). Measurements were made on clear sky days between 11 am and 2:30 pm local time (Central Daylight Time). Meteorological conditions for measurement periods are summarized in Fig. S1.

2.2.2. Gas exchange

Within 30 min of the spectral measurements, photosynthetic (A) vs. intercellular CO_2 (C_i) response curves were collected to determine $V_{c,max}$ and J_{max} for each leaf to use as ground-truth training for a PLSR model. Curves were measured on the same leaves as the hyperspectral measurements using a portable leaf gas exchange system (*LI-6400, LICOR Biosciences, Lincoln, NE, USA*). Ambient leaf temperature was determined as the mean of three measurements of leaf temperature with a handheld IR gun (*FLIR TG54, FLIR® Systems, Inc., Wilsonville, Oregon, USA*). Block temperature on the gas exchange system was set to match this mean leaf temperature prior to each CO_2 response curve. PAR was set to $1800 \mu\text{mol m}^{-2} \text{s}^{-1}$, and CO_2 concentrations were adjusted stepwise over a range of 50 to $2000 \mu\text{mol mol}^{-1}$ in set increments as follows: 400, 200, 50, 100, 300, 400, 600, 900, 1200, 1500, 1800, 2000. Leaves were acclimated to chamber conditions for a minimum of 300 s prior to initiating each A/C_i curve and a minimum and maximum wait time of 160 s and 200 s, respectively, was incorporated before triggering each individual measurement. Relative humidity inside the chamber was manually controlled to $65 \pm 5\%$ before each curve by adjusting the flow through the desiccant tube integrated into the gas exchange system. $V_{c,max}$ and J_{max} were determined from these A/C_i curves according to the mechanistic model of photosynthesis (Farquhar et al., 1980). A/C_i curves were analyzed using a curve fitting utility developed by Sharkey et al. (2007) with mesophyll conductance (g_m) constrained according to values for tobacco at 25°C reported previously with temperature dependency incorporated from the linear relationship of g_m with temperature where $y = -0.44 + 0.058x$ (Evans and Von Caemmerer, 2013).

2.3. Leaf nitrogen concentration

In 2016, immediately following each A/C_i curve, three 2.01cm^2 leaf disks were destructively harvested from each leaf using a cork borer and dried until constant mass and a subset of ground tissue of known mass ($3 \pm 0.5\text{mg}$) was combusted with oxygen in an elemental analyzer (*Costech 4010; Costech Analytical Technologies*) and calibrated to %N against an acetanilide standard curve.

2.4. Partial least squares regression (PLSR)

Two separate PLSR model build sets were performed for $V_{c,max}$, J_{max} and [N]. One model (Model set 1) was built using data collected during the 2016 growing season, and the second (Model set 2) was built using 75% of the data collected during the 2016 and 2017 growing seasons. Model set 1 was validated against data collected in the following year (2017). Model set 2 data was validated against the 25% of data not used for the model build (Table S1). Model training data sets consisted of pairs of modelled or measured parameters with reflectance spectra measured on the same leaf. For model validation, coefficients output from the PLSR model build were applied at each spectral waveband to collected reflectance spectra to predict the trait of interest.

All models were built following the same PLS build principles (Ollinger and Smith, 2005; Asner and Martin, 2008; Townsend et al., 2003) using previously published methods (Serbin et al., 2014) but modified for *N. tabacum*. An open-source Partial Least Squares (PLS) package (Mevik and Wehrens, 2007) in R (*The R Foundation for Statistical Computing, Wien, Austria*) was used to create the linear model of waveband coefficients by identifying latent variables (LVs) that account for trait variation in the reflectance spectra. It uses a leave-one-out cross validation approach that then makes a prediction for the out-of-sample observation (Siegmann and Jarmer, 2015). The predicted residual sum of squares (PRESS) statistic and lowest root mean square error of prediction from cross validation (RMSEPCV) were used to determine the optimal number of LVs and to prevent overfitting. The PRESS statistic determines the number of LVs to achieve minimum root mean square error (RMSE) between modelled and observed leaf traits (Wold et al., 2001). The RMSEPCV cross validates the model bias and variance (Gowen et al., 2011).

2.4.1. Model set 1

For Model set 1, the $V_{c,max}$ model was built with a training dataset of 113 measurement pairs. The collected reflectance spectra from six genotypes (Fig. 2a) were used as a training dataset for a PLSR model build with six latent variables as determined by the PRESS statistic and RMSE (Fig. S2a, c and e). The J_{max} PLSR model was built independently given that the double Rubisco knockdown plants (SSuD) were found to not be electron transport limited even at high CO_2 concentrations. As such, maximal electron transport rate could not be determined for the double Rubisco knockdown measurements (SSuD) and they were removed from the J_{max} model build, leaving a data training set of 94 measurement pairs, with 9 latent variables (Fig. S2b, d and f). The spectral range for the $V_{c,max}$ and J_{max} models was 500–2400 nm, with a spectral resolution of 3 nm in the visible and NIR (350–1000 nm) and 8 nm in shortwave-infrared (SWIR; 1000–2500 nm). While spectra were collected across the full range (400–2500 nm) we excluded regions below 500 nm and above 2400 nm due to noise. The %N model was built with a training data set of 131 N values measured from leaf tissue samples in the same leaves measured for $V_{c,max}$ and J_{max} . The spectral range for the %N model was also 500–2400 nm. PLSR build statistics for all 2016 models are shown in Figs. S2–4.

The models for $V_{c,max}$, J_{max} and %N were built with varying sample sizes, according to the reliability of leaf level field gas exchange measurements and sampling. For the %N 131 data pairs were used, for $V_{c,max}$ 114, and for J_{max} only 97, given that a model build requires a collected spectrum and accurate ‘ground truth’ measures of leaf N, $V_{c,max}$ or J_{max} . In 2016, a total of 132 plants were measured (132 spectra collected). One of the leaf disks for nitrogen extraction was compromised during transport to the lab and therefore discarded, leaving 131 measurement pairs to build the %N model. CO_2 response curves made with gas exchange systems respond to plant environment and physiology over the ~30mins required per response curve, and thus have a greater margin for error. If the collected curves could not be fit for a reliable value for $V_{c,max}$ and J_{max} , they were removed from the analysis prior to any model build. For example, if stomatal conductance was

shown to be limiting, or the sum of squares from being fit according to the Farquhar et al. (1980) model of photosynthesis was greater than 100, the derivation of $V_{c,max}$ and J_{max} was deemed unreliable and removed from the analysis. Reliable modelled values of J_{max} were further reduced compared with $V_{c,max}$ given the inability to define an electron transport limited state for the plants with reduced Rubisco. No data pairs with reliable ground truth measurements for %N, $V_{c,max}$ and J_{max} were removed from analysis beyond this point for any model builds. No outliers were removed from any models.

2.4.2. Model set 2

The same PLSR model builds protocol as for Model Set 1 was followed for Model Set 2. The $V_{c,max}$ model was built with a training dataset of 186 measurement pairs of collected reflectance spectra and modelled gas exchange $V_{c,max}$ values from 9 genotypes, with 15 latent variables as determined by the PRESS statistic and RMSEPCV (Fig. S5a, c and e). The J_{max} model was built independently, again with the double Rubisco knockdown measurements (SSuD) removed, leaving a data training set of 165 measurement pairs, with 15 latent variables (Fig. S5b, d, f). PLSR scores for both models are presented in Fig. S6 and model residuals in Fig. S7. The spectral range for the $V_{c,max}$ and J_{max} models was 500–2400 nm. Newly generated PLSR coefficients were applied to the spectra from the 25% holdout validation dataset for both $V_{c,max}$ and J_{max} .

3. Results

3.1. Model build set 1

Measurements for the Model Set 1 made over the 2016 growing season (Fig. 1) represented a wide range of meteorological conditions (Fig. S1), which coupled with the different cultivars and genetic modifications, yielded a wide range of values for $V_{c,max}$ (14.7–279.8 $\mu mol CO_2 m^{-2} s^{-1}$) and J_{max} (92.8–323.2 $\mu mol CO_2 m^{-2} s^{-1}$). To build a predictive model for $V_{c,max}$, J_{max} and [N], measurements of each parameter were paired with collected reflectance spectra from the same leaves (Fig. 2).

For Model Set 1, the PLSR model for $V_{c,max}$ ($R^2 = 0.6$; Fig. 3a), J_{max} ($R^2 = 0.59$; Fig. 3b), and [N] ($R^2 = 0.83$; Fig. 3c) all showed strong positive correlations with the measured values. Leaf [N] was very high in the SSuD double Rubisco knockdown leaves which had the lowest $V_{c,max}$ (Fig. 3). Fig. S2 shows the number of latent variables (LV's) chosen for each model build, defined as the LV number that minimizes the predicted residual error sum of squares (PRESS) statistic and root mean square error of prediction from cross validation (RMSECV) from the predictions vs. sample observation fit. In model set 1, six LV's were used for the $V_{c,max}$ build, while nine LV's were used for J_{max} and [N]. PLSR scores for LV's and model residuals are shown in Figs. S3 and S4, respectively, with residuals evenly centered around zero for all model builds, confirming the suitability of a linear regression model.

Model coefficients generated from the model fit center around zero in all model builds (Figs. 4a and b). Model loadings indicate the impact of spectral regions on the model build.

Where loading values are closer to zero, they have a lower impact on the model build and the more the values deviate from zero indicate greater influence on the PLSR predictions. $V_{c,max}$ model loadings across the spectra (Fig. 4c) showed stronger correlation in the chlorophyll bands (450–550 nm and 640–680 nm), across the red edge (680–730 nm), and across the NIR (900–1400 nm). In the short wave infra-red region (177–2400 nm) loading values had a lesser impact than in the visible and near infra-red ranges. The same chlorophyll bands were highly loaded in the J_{max} model, yet the NIR region (800–1400 nm) and SWIR (1700–2400) had a lower loading weight (Fig. 4c). Percentage [N] model loadings were strong across the NIR spectra (700–1300 nm) and from 1500 to 2400 nm in the SWIR, with areas of lower loading weight between 1300–1400 and 1900–2000 nm

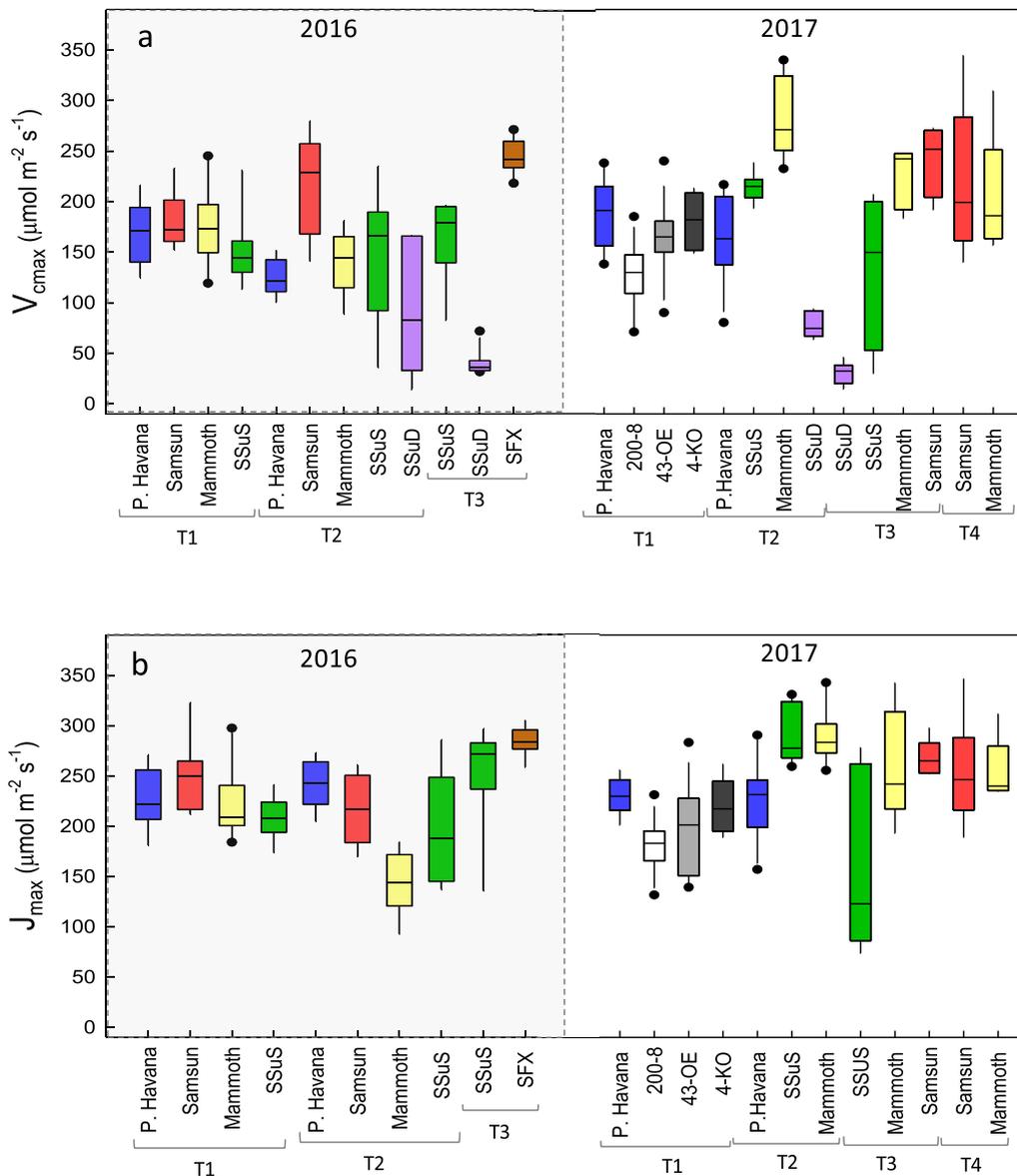


Fig. 1. Box plots for $V_{c,max}$ (a) and J_{max} (b) calculated from photosynthetic- CO_2 response curves for tobacco plants over two growing seasons. The boxes show the interquartile range with the median as solid horizontal line. Whiskers show data outside the interquartile range but within $1.5\times$ the interquartile range. Dots show outliers. Colors are included to assist in comparisons with Figs. 3, 5, 6 and 8. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(Fig. 4d).

Spectral measurements were coupled with gas exchange measurements from leaves at varying growth stages and meteorological conditions during 2017 in the same cultivars to test the model build from 2016 (Fig. 5). With PLSR coefficients applied to these collected leaf reflectance spectra, Model set 1 correlated strongly for measured $V_{c,max}$ ($R^2 = 0.69$) but not for J_{max} ($R^2 = 0.17$) in 2017 (Fig. 4). However, photorespiratory bypass and modified photoprotection transgenic lines (200-8, 43-OE and 4-KO) that were not grown in 2016 reduced the predictive power of $V_{c,max}$ Model set 1 when added to the validation dataset (Fig. 6a). Predicting $V_{c,max}$ for all 2017 genotypes using Model set 1 resulted in lower R^2 values of 0.53 (Fig. 6a). Applying the PLSR coefficients solely to the three newly added 2017 transgenic cultivars showed no predictive power for $V_{c,max}$ (Fig. 6b).

3.2. Model build set 2

Model build set 2 used the data collected in 2016 and 2017 with

75% of the spectra from both years (Fig. 7) used to train a second PLSR model to test against the remaining 25%. The new model showed a strong relationship for $V_{c,max}$ ($R^2 = 0.77$, $y = 0.77x + 40$; Fig. 8a) and J_{max} ($R^2 = 0.72$, $y = 0.72x + 68.1$; Fig. 8b). Both models were built with 15 latent variables (LV's) through determination of the PRESS statistic (Fig. S5). PLSR scores are shown in Fig. S6 and model residuals had an even spread around zero (Fig. S7).

The RMSEP CV was 35.78 and 29.52 $\mu\text{mol m}^{-2} \text{s}^{-1}$ and model bias was -0.06 and 0.17% for $V_{c,max}$ and J_{max} , respectively. With 25% of the data randomly selected and held back for cross validation by application of the PLSR coefficients from the new model (Fig. 9a), both $V_{c,max}$ and J_{max} were faithfully predicted from the reflectance spectra where R^2 for $V_{c,max}$ was 0.61 (Fig. 8c) and for J_{max} was 0.62 (Fig. 8d). Model loadings for model set 2 (Fig. 9b) follow similar patterns to model set 1 (Fig. 4c), with the exception that J_{max} has higher loadings in the NIR region (800–1400 nm).

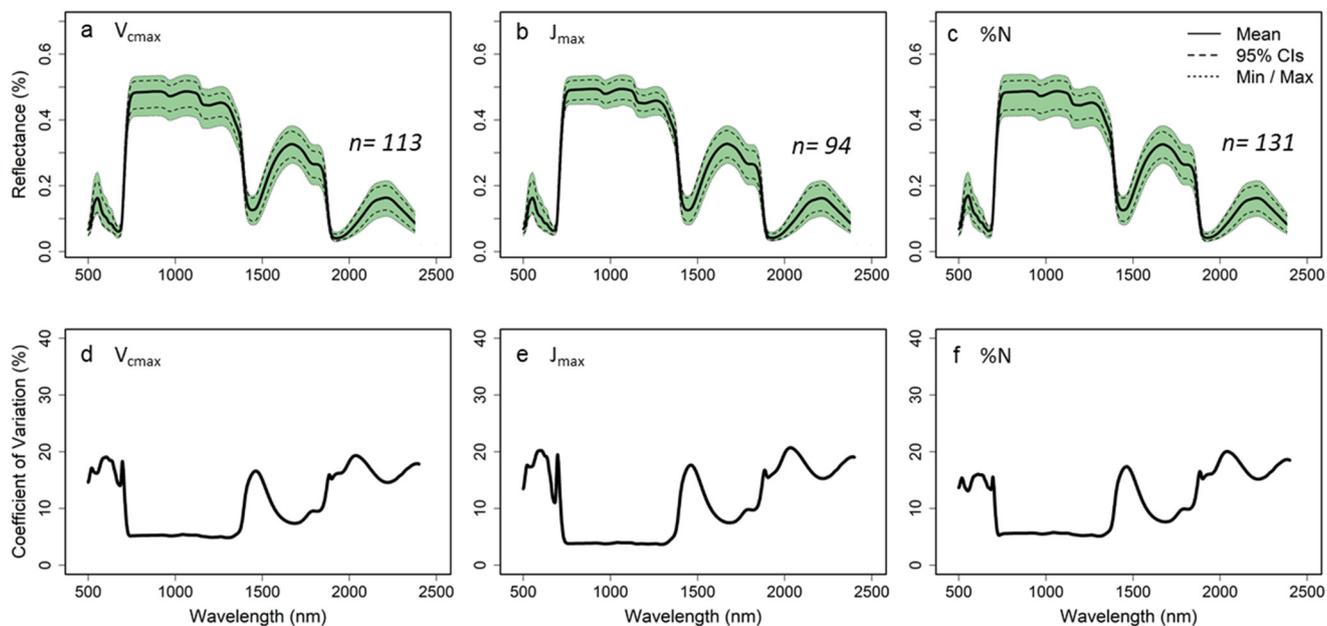


Fig. 2. Mean, 95% confidence intervals, and minimum and maximum leaf reflectance for all leaves of *Nicotiana tabacum* used for the 2016 V_{cmax} (a), J_{max} (b) and %N (c) model builds and the co-efficient of variation across the full spectra for each model build respectively (d, e and f).

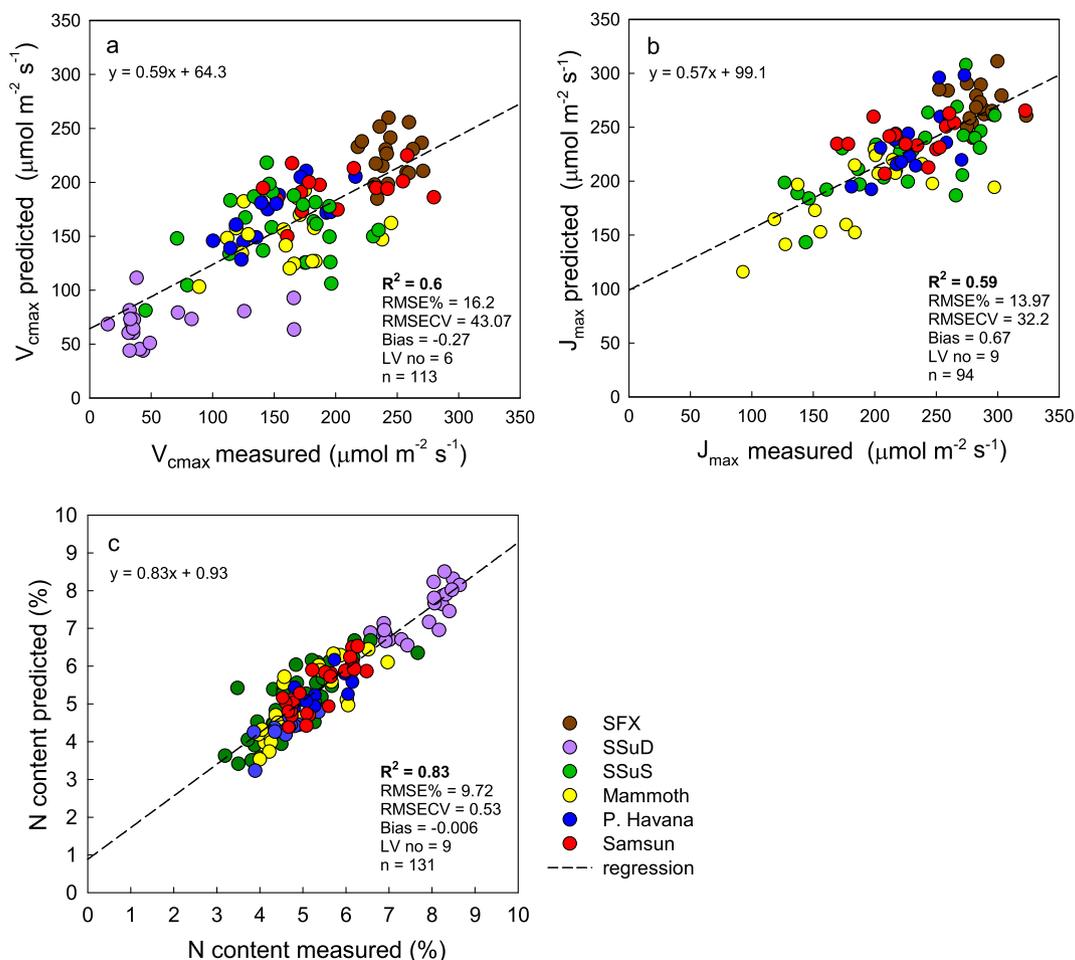


Fig. 3. PLSR predicted from leaf spectral measurement (500–2400 nm) vs. measured using traditional techniques of V_{cmax} (a), J_{max} (b) and [N] (c) from 6 tobacco genotypes in 2016. The dashed line represents a linear regression fit to the data with statistical results are inset.

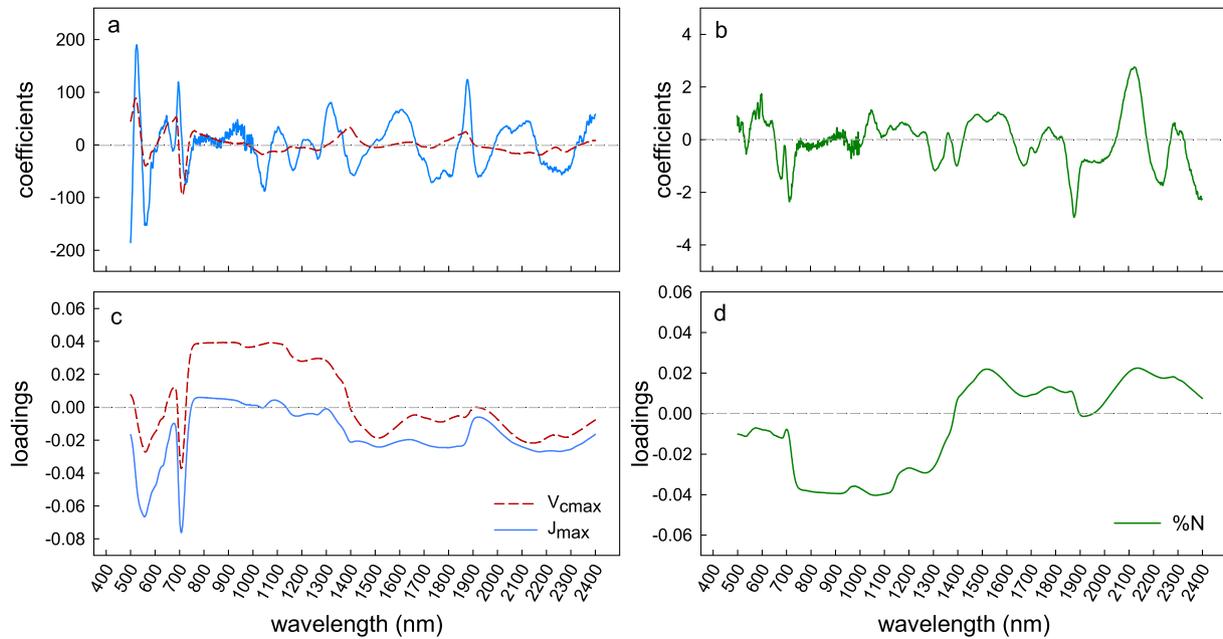


Fig. 4. Model build set 1 spectral-specific coefficients for $V_{c,max}$ and J_{max} (a) and %N [N] (b), with model loadings for $V_{c,max}$ and J_{max} (c) and %N (d).

4. Discussion

This research addressed whether the PLSR method can be applied to genetically modified crops to rapidly and accurately predict $V_{c,max}$ and J_{max} (Prediction 1) and to assess interannual variability in photosynthetic performance (Prediction 2). We further sought to test whether spectra-based $V_{c,max}$ and J_{max} models would be independent of spectra-leaf nitrogen relationships rather than arising from the close inter-relationships between $V_{c,max}$ and J_{max} with leaf nitrogen (Prediction 3).

The PLSR models were able to predict significant variation in $V_{c,max}$ and J_{max} imposed by genetic modification and measured across growing seasons, supporting Prediction 1. A PLSR model built using 2016 data (Model set 1) resulted in a strong ability to predict $V_{c,max}$ in 2017 but not J_{max} , supporting Prediction 2 for $V_{c,max}$, but not for J_{max} . Further, the PLSR Model set 1 was only successful in predictions of $V_{c,max}$ and J_{max} for genotypes used to build the model in 2016, i.e., the model was unable to predict either $V_{c,max}$ or J_{max} for genotypes unique to 2017 using a model built with data from 2016 (Fig. 6a). Finally, the model worked for all species, including a genetically modified genotype showing the lowest $V_{c,max}$ yet the highest [N], reverse for all other genotypes. That the model faithfully predicted both $V_{c,max}$ and J_{max} for

this genotype suggests that the PLSR model is not merely using leaf [N] as a proxy of for photosynthetic physiology, which supported Prediction 3.

Measurements represented a wide range of photosynthetic capacity and leaf nitrogen concentrations, however the full strength of the model was only realized when built using data that captured the full range of interannual and genetic variation (Fig. 8). The application of our 2016 model to the 2017 dataset for the same species but with three additional transgenic lines (Fig. 6) showed that the model may only be applicable to the varieties used to build the model, regardless whether wild-type or genetically modified. PLSR coefficients from Model set 1 predicted 2017 measurements well (Fig. 4) but only for $V_{c,max}$ and for the genotypes present in both years.

The inability of model build set 1 to predict J_{max} in 2017 may be linked with the underlying processes represented by J_{max} compared with $V_{c,max}$. The total amount of the enzyme Rubisco that is present and metabolically active determines $V_{c,max}$ (Bernacchi et al., 2001; Portis, 2003; Suzuki et al., 2009). While there is substantial uncertainty surrounding the mechanism being used to predict $V_{c,max}$, there is no inter-annual variation in the structure and function of Rubisco. Therefore, the performance of the 2016 model in predicting $V_{c,max}$ in 2017 is not

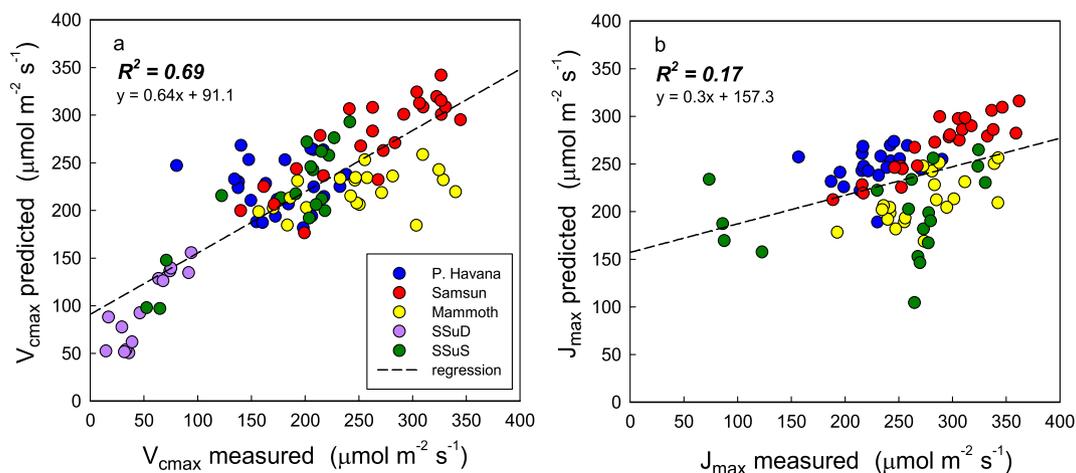


Fig. 5. Validation of model build set 1 for $V_{c,max}$ and J_{max} using the same genotypes measured in 2017. The regression equation and R^2 are inset for each graph.

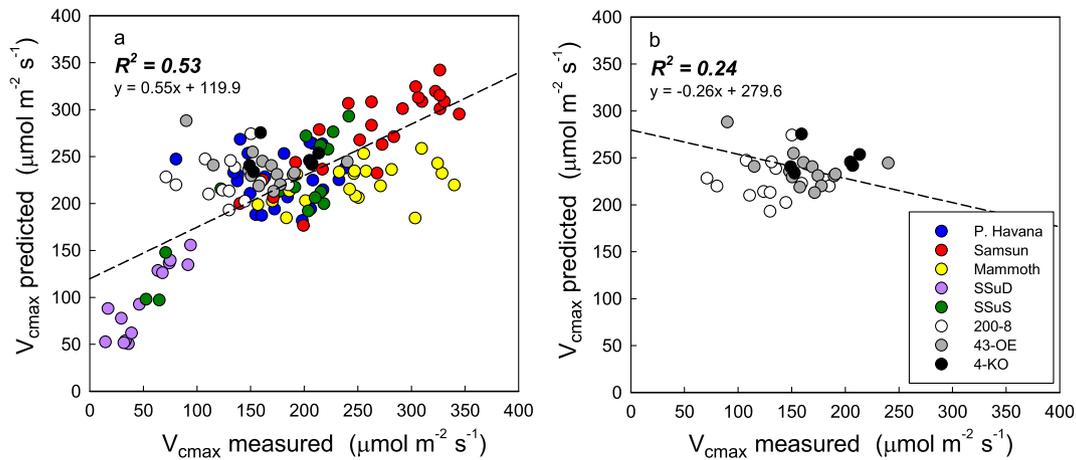


Fig. 6. PLSR coefficients from 2016 model build shown in Fig. 3 applied to reflectance spectra collected in 2017 to predict $V_{c,max}$ in 8 *Nicotiana tabacum* genotypes (a). The 3 newly added transgenic genotypes in 2017 are separated from the dataset, and the same PLSR coefficients applied to predict $V_{c,max}$ in those genotypes alone (b).

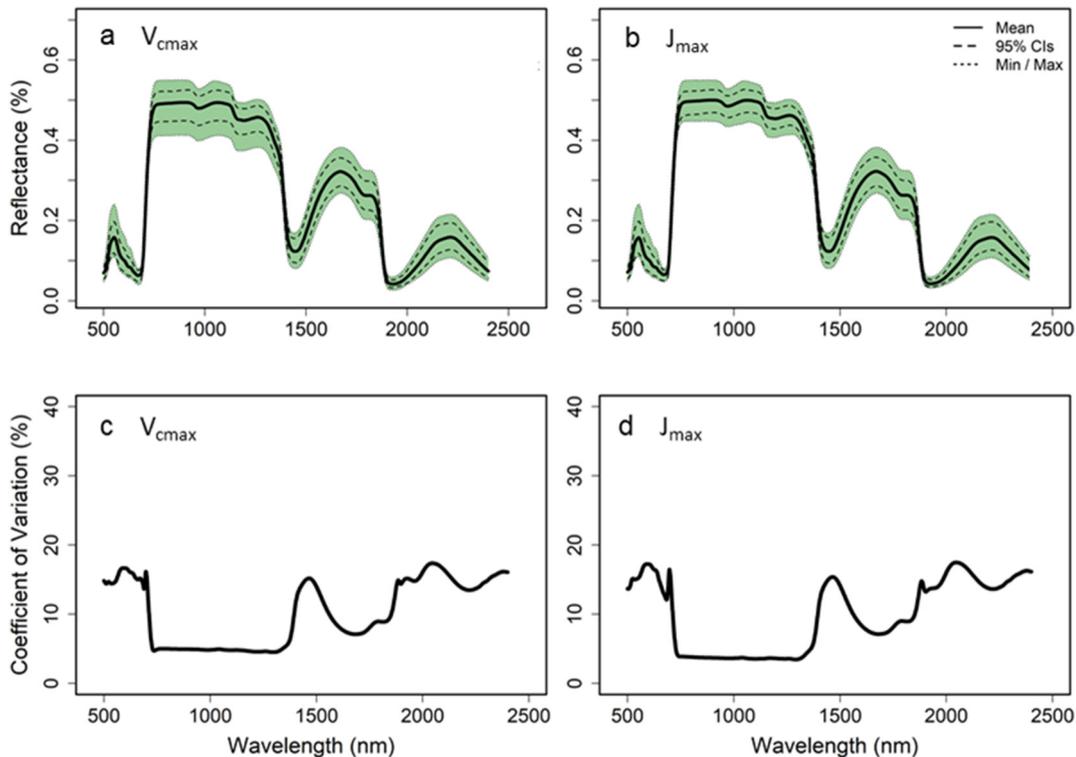


Fig. 7. Mean, 95% confidence Intervals, and minimum and maximum leaf reflectance Model set 2 for $V_{c,max}$ (a) and J_{max} (b) and the co-efficient of variation for the full spectra for both models respectively (c and d).

unexpected. However, J_{max} represents the coordination of a complex series of reactions involving many proteins integrated into the thylakoid membrane in the chloroplast (Farquhar et al., 1980; von Caemmerer, 2000). Furthermore, the estimation of J_{max} may not necessarily reflect rates of electron transport as more recent research suggests J_{max} may be constrained by metabolic reactions involved in the regeneration of RuBP in the photosynthetic ‘dark’ reactions (Raines, 2003; Lefebvre et al., 2005). Therefore, the more complex metabolic processes associated with J_{max} likely leads to its poor performance compared with $V_{c,max}$ (Fig. 5b). A similar challenge is observed when trying to model photosynthesis using J_{max} for plants grown under different environmental conditions (Bernacchi et al., 2003; Köhler et al., 2016). The model built with 2016 and 2017 data (Model set 2) predicted $V_{c,max}$ and J_{max} better (Fig. 8a) than a model built with only 2016

data (Fig. 3). R^2 values were higher and RMSE was reduced in model build set 2 for both $V_{c,max}$ (from 16.2 to 10.8%) and J_{max} (14.0 to 10.1%). Model bias also decreased for both $V_{c,max}$ (from -0.27 to -0.06%) and J_{max} (from 0.67 to 0.17%).

These results suggest that the importance of including environmental variation when building PLSR models (Wu et al., 2017) is matched by the importance of incorporating genetic variation. Combined over both years, a total of nine genotypes each expressing different photosynthetic phenotypes were used in this analysis and the model only performed well when validated against the same genotypes used to build the model. Whether a PLSR model built using dozens or hundreds of genotypes can reliably predict photosynthetic physiology for a unique genotype still needs to be resolved.

The dependence of $V_{c,max}$ and J_{max} on temperature is well

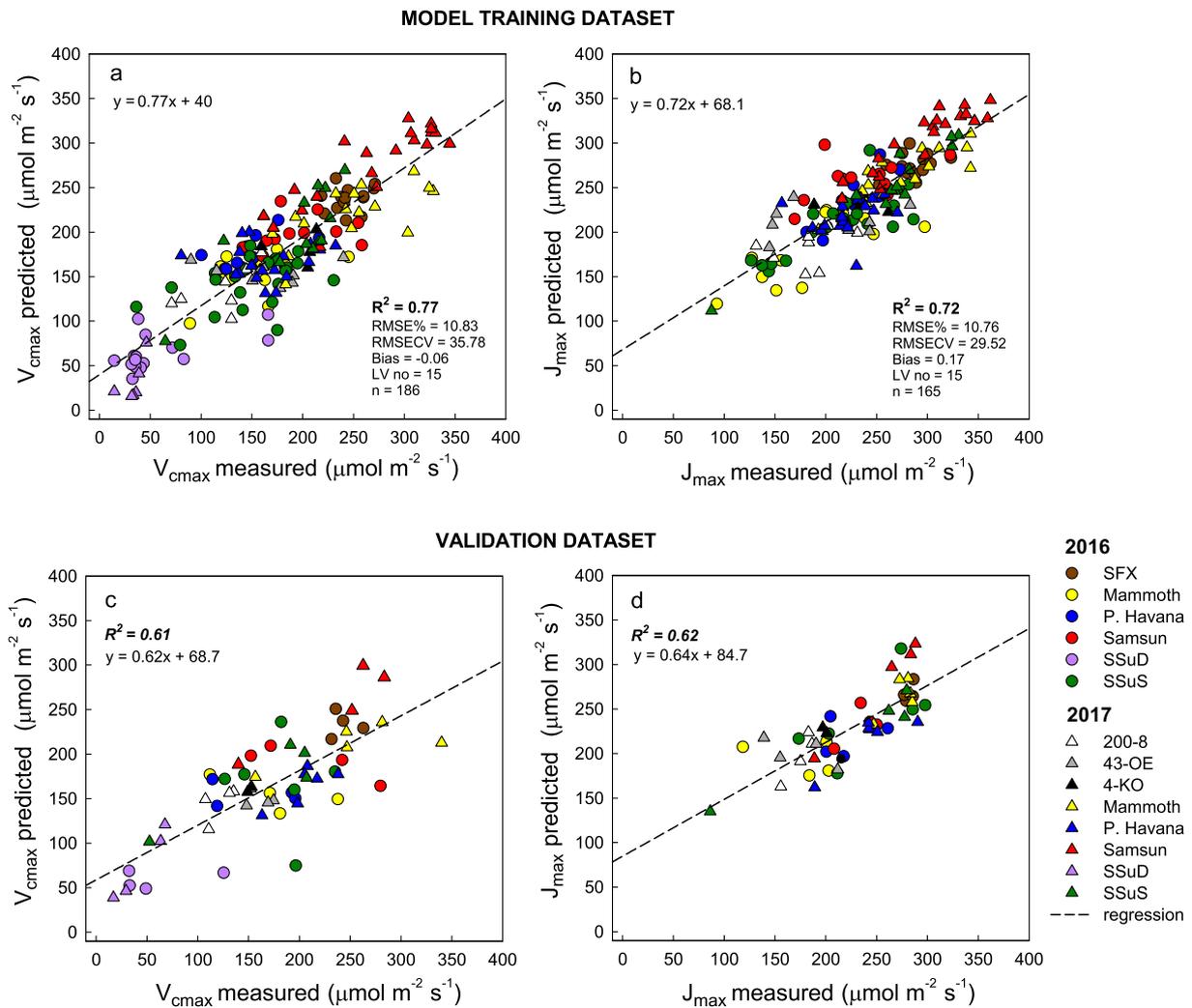


Fig. 8. Measured versus predicted PLSR values of $V_{c,max}$ (a) and J_{max} (b) from PLSR models built with 75% of data collected in 2016 and 2017, randomly selected for model training (Model set 2). Model build statistics are presented in Figs. S5–7.

documented (Farquhar et al., 1980; von Caemmerer, 2000; Bernacchi et al., 2001, 2003). Therefore, the gas exchange measurements were collected at the same temperature as leaf spectral measurements to remove temperature variation between the measured and modelled parameters. Normalizing to a standard temperature might lead to improved regressions but would challenge the application of this technique under field-ambient conditions where temperatures constantly fluctuate.

PLSR model loadings identify regions of the spectrum significant for the trait of interest, by identifying maximum variations common to all spectra in the dataset (Wold et al., 2001; Serbin et al., 2012). Highly loaded regions are linked with physiological indicators, which suggest that using PLSR to mathematically analyze reflectance spectra may be more than purely empirical. The PLSR model loadings across the spectra for $V_{c,max}$ and J_{max} for build 1 show strong correlations between 450 and 680 nm and across the ‘red-edge’ (Fig. 4c). Leaf reflectance spectra between 450 and 680 nm is influenced by photosynthetic pigments (Gates et al., 1965; Knipling, 1970; Rouse Jr et al., 1974) and carotenoid content (Gamon et al., 1992). Similarly, the ‘red-edge’ spectral region (680–730 nm), in which leaf reflectance greatly increases as light in the near infra-red region is no longer absorbed by chlorophyll (Woolley, 1971; Horler et al., 1983) has been shown to correlate with photosystem II function (e.g., Fv/Fm; Zarco-Tejada et al., 2000). These same spectral regions have shown similar loading for photosynthetic PLSR models of other crop species (Serbin et al., 2012;

Yendrek et al., 2016; Silva-Perez et al., 2017). The lower loading weights in the J_{max} model compared with $V_{c,max}$ model in the NIR region (Fig. 4c) are consistent with previous results (Serbin et al., 2012). Model build 2 (Fig. 8) showed a very similar pattern for loading weights (Fig. 9).

While model loadings are generally consistent with previous reports and suggest links with physiological controls, the underlying mechanism is unclear that PLSR models utilize to predict $V_{c,max}$ and J_{max} . While leaf [N] cannot be directly measured by spectral analysis, we show that it can be accurately predicted by PLSR analysis (Fig. 3c) consistent with other studies (Kattge et al., 2009; Serbin et al., 2012; Rogers, 2014; Walker et al., 2014; Yendrek et al., 2016; Dechant et al., 2017; Silva-Perez et al., 2017). Given that photosynthetic enzymes, predominantly Rubisco, account for a large proportion of leaf nitrogen, under favorable environmental conditions photosynthetic rate per unit leaf area (photosynthetic capacity) increases linearly with leaf nitrogen content (Field and Mooney, 1986; Evans, 1989; Poorter et al., 1990; Nakano et al., 1997; Reich et al., 1998). Yet, in the SSuD Rubisco antisense line this relationship is reversed – the SSuD lines have, by far, the lowest $V_{c,max}$ and the highest [N], similar to other studies using tobacco with reduced Rubisco (Quick et al., 1991; Masle et al., 1993). Despite this genotype differing from all others, the $V_{c,max}$ and [N] models faithfully predict each phenotype independently (Fig. 3a and c). This suggests that, despite previous research suggesting the PLSR model is dominated by a nitrogen signal (Kattge et al., 2009; Rogers, 2014;

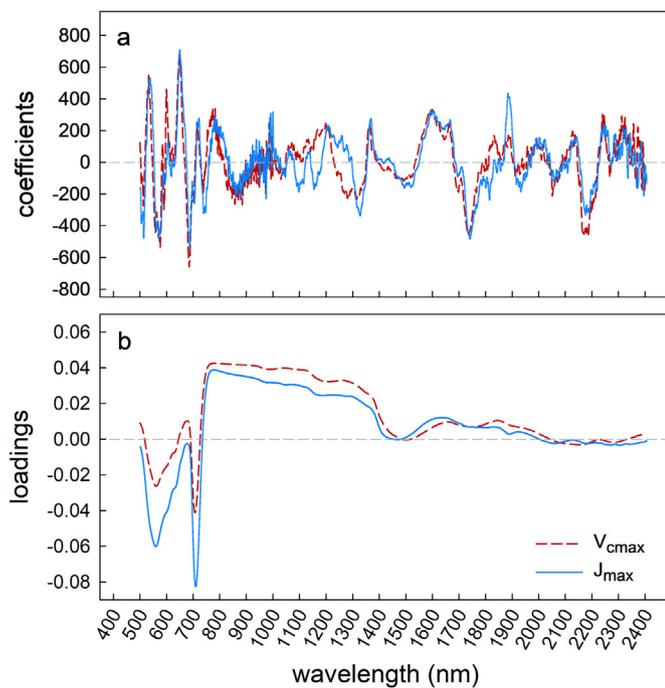


Fig. 9. Model build set 2 generated coefficients (a) and loading weights (b) for V_{cmax} and J_{max} .

Walker et al., 2014; Dechant et al., 2017), the model estimates of the traits are independent. While the results are unable to elucidate a key mechanistic understanding of what the PLSR method is measuring, they are both a valuable step in understanding that the relationship is not simply a proxy for leaf nitrogen and that this method can be very useful in phenotyping both wild-type and genetically modified plants despite the lack of mechanistic understanding.

The model shows strong predictive ability over a wide range of environmental conditions for plants showing a range of phenotypes. Additional experiments using genotypes with distinct alterations to the photosynthetic pathway and/or related metabolic pathways can be further exploited to refine the mechanism associated with this technique. For example, Rubisco active site inhibitors, which reduce the catalytic activity of Rubisco, thereby reducing $V_{c,max}$ without changing the concentration of Rubisco, may provide more insight into the mechanisms behind the PLSR results. Machine learning analysis using plants with genetic modifications that extend beyond photosynthesis may elucidate spectral regions driving PLSR predictions to provide further physiological interpretation.

The results presented here also suggest that caution must be exercised with reflectance-based analysis to infer crop productivity or health for genetically modified plants. Introduction of transgenic modifications to the photosynthetic pathway for the SSuD tobacco line led to shifts in known relationships between leaf chemistry and photosynthetic capacity. Using many established reflectance-based approaches, particularly related to simple indices (e.g., PRI, NDVI), would suggest the high leaf [N] for SSuD might predict higher biomass. Therefore, the widely held and supported view that photosynthesis and leaf N content are highly correlated (Evans, 1989) for wild-type species would suggest photosynthetic capacity and productivity of established techniques may not apply to genetically modified plants.

The full extent of current and potential opportunities to improve photosynthesis in plants extends well beyond the strategies employed in this analysis (Ort et al., 2015). Continued use and development of high throughput techniques using novel genotypes holds potential for further insights into the mechanisms behind spectral shifts and how they relate to photosynthetic physiology. Furthermore, it may prove to be a useful tool as novel breeding strategies are realized. Extending this model to

incorporate spectra collected from imaging systems, thereby removing the need for leaf-clip based measurements, may increase throughput capabilities by orders of magnitude and provide critical information related to canopy-scale variance in photosynthetic physiology. The need to understand agronomic ecosystem functioning has led to substantial efforts to measure productivity over large spatial scales. Using solar induced fluorescence (e.g., Miao et al., 2018) to quantify GPP can lead to a better understanding of ecosystem functioning. However, efforts to improve understanding of ecosystem function extend beyond in situ measurements of photosynthesis and towards understanding the underlying physiology necessitating techniques such as those described here.

5. Conclusions

We show the spectral PLSR method can be applied to genetically modified crops to rapidly and accurately predict photosynthetic capacity. However, models lose predictive power when used interannually on new genetic material not included in model builds. Despite the strengths of the model, the results suggest a need for repopulation of PLSR models annually when dealing with discrete variation in photosynthetic capacity between genotypes of a single species in crop trials. However, the extent that models need to be repopulated in time and space for hyperspectral PLSR models is still uncertain. The need for repopulating the model may also apply when using hyperspectral PLSR to compare photosynthetic physiology between geographical regions or when environmental conditions change dramatically over time. We also deconvolute the relationship between photosynthetic capacity and leaf nitrogen content, which speaks to the potential of full spectral analysis for elucidating biochemical mechanism from spectral reflectance with further work, at any scale.

Our results suggest that a link between remote sensing and photosynthetic physiology can be applied to agricultural species. This provides opportunity to better parameterize agricultural models and to identify variation in photosynthetic physiology for breeding efforts. As future food production will almost certainly depend on genetically modified crops (Ort et al., 2015), we also show that novel techniques will continue to be useful as the agricultural landscape continues to change.

Acknowledgements

The information, data, or work presented herein was funded in parts by (1) Bill and Melinda Gates Foundation grant OPP1060461, titled “RIPE—Realizing increased photosynthetic efficiency for sustainable increases in crop yield”, (2) by the Advanced Research Projects Agency - Energy - U.S. Department of Energy, under Award Number DE-AR0000598, and (3) by the Agricultural Research Service of the United States Department of Agriculture. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof. We thank David Drag and Ben Harbaugh for greenhouse and field plant management. Prof. Susanne von Caemmerer (ARC Centre of Excellence for Translational Photosynthesis Research, Australian National University) kindly provided the Rubisco Antisense *Nicotiana tabacum* and discussion that helped improve this work. Johannes Kromdijk, Katarzyna Glowaka and Stephen P. Long provided transgenic *N. tabacum* lines 43-OE and 4-KO and Paul South and Donald R. Ort provided line 200-8. Co-author J.W. was supported by the United States Department of Energy contract No. DE-SC0012704 to Brookhaven National Laboratory. We also thank Evan Dracup, Justine Brumm, Kyle Coffland, Morgan Prinn, Alyssa Dwyer, Alex Riley, Isaac Howenstein, Jennifer Ward, Amanda Cavanagh, and Elena Pelech for assistance with the field work.

Author contributions

Katherine Meacham-Hensold and Carl Bernacchi designed the experiment. Katherine Meacham-Hensold carried out field work and data analysis and led the development of the manuscript. Christopher M. Montes carried out data analysis and assisted with building of PLSR models. Jin Wu and Kaiyu Guan helped with experimental design, technical equipment set up and advice on data analysis. Peng Fu advised on data analysis and manuscript editing. Taylor Pederson and Caitlin Moore helped with technical aspects of data collection. Elizabeth Ainsworth helped with data analysis and experimental design. Christine Raines and Kenny Lee Brown provided the construct and experimental design for SFX genotype. Carl Bernacchi supervised the work as lab leader, advising on experimental design and data analysis. All authors contributed to editing all drafts of the manuscript.

Competing interest statement

There are no conflicts of interest to declare.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.rse.2019.04.029>.

References

- Ainsworth, E.A., Serbin, S.P., Skoneczka, J.A., Townsend, P.A., 2014. Using leaf optical properties to detect ozone effects on foliar biochemistry. *Photosynth. Res.* 119, 65–76.
- Andralojc, P.J., Carmo-Silva, E., Degen, Gustaf E., Parry, Martin A.J., 2018. Increasing metabolic potential: C-fixation. *Essays Biochem.* 62, 109–118.
- Asner, G.P., Martin, R.E., 2008. Spectral and chemical analysis of tropical forests: scaling from leaf to canopy levels. *Remote Sens. Environ.* 112 (10), 3958–3970.
- Bagley, J.E., Miller, J., Bernacchi, C.J., 2015a. Biophysical impacts of climate-smart agriculture in the Midwest United States. *Plant Cell Environ.* 38, 1913–1930.
- Bagley, J.E., Rosenthal, D.M., Ruiz-Vera, U.M., Siebers, M.H., Kumar, P., Ort, D.R., Bernacchi, C.J., 2015b. The influence of photosynthetic acclimation to rising CO₂ and warmer temperatures on leaf and canopy photosynthesis models. *Glob. Biogeochem. Cycles* 29, 194–206.
- Bernacchi, C.J., Singaas, E.L., Pimentel, C., Portis Jr, A.R., Long, S.P., 2001. Improved temperature response functions for models of Rubisco-limited photosynthesis. *Plant Cell Environ.* 24, 253–259.
- Bernacchi, C., Pimentel, C., Long, S., 2003. In vivo temperature response functions of parameters required to model RuBP-limited photosynthesis. *Plant Cell Environ.* 26, 1419–1430.
- Bernacchi, C.J., Bagley, J.E., Serbin, S.P., Ruiz-Vera, U.M., Rosenthal, D.M., VanLoocke, A., 2013. Modelling C-3 photosynthesis from the chloroplast to the ecosystem. *Plant Cell Environ.* 36, 1641–1657.
- Curran, P.J., Kupiec, J.A., Smith, G.M., 1997. Remote sensing the biochemical composition of a slash pine canopy. *IEEE Trans. Geosci. Remote Sens.* 35 (2), 415–420.
- Dechant, B., Cuntz, M., Vohland, M., Schulz, E., Doktor, D., 2017. Estimation of photosynthesis traits from leaf reflectance spectra: correlation to nitrogen content as the dominant mechanism. *Remote Sens. Environ.* 196, 279–292.
- Driever, S.M., Simkin, A.J., Alotaibi, S., Fisk, S.J., Madgwick, P.J., Sparks, C.A., Jones, H.D., Lawson, T., Parry, M.A., Raines, C.A., 2017. Increased SBPase activity improves photosynthesis and grain yield in wheat grown in greenhouse conditions. *Philos. Trans. R. Soc. B* 372, 20160384.
- Evans, J.R., 1989. Photosynthesis and nitrogen relationships in leaves of C₃ plants. *Oecologia* 78, 9–19.
- Evans, J.R., 2013. Improving photosynthesis. *Plant Physiol.* 162, 1780–1793.
- Evans, J.R., Von Caemmerer, S., 2013. Temperature response of carbon isotope discrimination and mesophyll conductance in tobacco. *Plant Cell Environ.* 36, 745–756.
- Farquhar, G.D., Caemmerer, S., Berry, J.A., 1980. A biochemical model of photosynthetic CO₂ assimilation in leaves of C₃ species. *Planta* 149, 78–90.
- Field, C.H., Mooney, H.A., 1986. Photosynthesis—nitrogen relationship in wild plants. In: *On the Economy of Plant Form and Function: Proceedings of the Sixth Maria Moors Cabot Symposium, Evolutionary Constraints on Primary Productivity, Adaptive Patterns of Energy Capture in Plants*, Harvard Forest, August 1983. Cambridge University Press, Cambridge [Cambridgeshire] (c1986).
- Fischer, R.A.T., Edmeades, G.O., 2010. Breeding and cereal yield progress. *Crop Sci.* 50, 85–98.
- Foley, J.A., Ramankutty, N., Brauman, K.A., Cassidy, E.S., Gerber, J.S., Johnston, M., Mueller, N.D., O'Connell, C., Ray, D.K., West, P.C., 2011. Solutions for a cultivated planet. *Nature* 478, 337.
- Frankenberg, C., Berry, J., 2018. Solar induced chlorophyll fluorescence: origins, relation to photosynthesis and retrieval. In: *Reference Module in Earth Systems and Environmental Sciences: Comprehensive Remote Sensing*. Elsevier, Oxford, pp. 143–162.
- Furbank, R.T., Tester, M., 2011. Phenomics – technologies to relieve the phenotyping bottleneck. *Trends Plant Sci.* 16, 635–644.
- Gamon, J., Penueles, J., Field, C., 1992. A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. *Remote Sens. Environ.* 41, 35–44.
- Gates, D.M., Keegan, H.J., Schleter, J.C., Weidner, V.R., 1965. Spectral properties of plants. *Appl. Opt.* 4 (1), 11–20.
- Głowacka, K., Kromdijk, J., Leonelli, L., Niyogi, K.K., Clemente, T.E., Long, S.P., 2016. An evaluation of new and established methods to determine T-DNA copy number and homozygosity in transgenic plants. *Plant Cell Environ.* 39, 908–917.
- Głowacka, K., Kromdijk, J., Kucera, K., Xie, J., Cavanagh, A.P., Leonelli, L., Leakey, A.D., Ort, D.R., Niyogi, K.K., Long, S.P., 2018. Photosystem II subunit S overexpression increases the efficiency of water use in a field-grown crop. *Nat. Commun.* 9, 868.
- Gowen, A.A., Downey, G., Esquerre, C., O'Donnell, C.P., 2011. Preventing over-fitting in PLS calibration models of near-infrared (NIR) spectroscopy data using regression coefficients. *J. Chemom.* 25, 375–381.
- Guan, K., Berry, J.A., Zhang, Y., Joiner, J., Guanter, L., Badgley, G., Lobell, D.B., 2016. Improving the monitoring of crop productivity using spaceborne solar-induced fluorescence. *Glob. Chang. Biol.* 22, 716–726.
- Guanter, L., Alonso, L., Gómez-Chova, L., Amorós-López, J., Vila, J., Moreno, J., 2007. Estimation of solar-induced vegetation fluorescence from space measurements. *Geophys. Res. Lett.* 34 (8).
- Heckmann, D., Schlüter, U., Weber, A.P.M., 2017. Machine learning techniques for predicting crop photosynthetic capacity from leaf reflectance spectra. *Mol. Plant* 10, 878–890.
- Horler, D.N.H., Dockray, M., Barber, J., 1983. The red edge of plant leaf reflectance. *Int. J. Remote Sens.* 4 (2), 273–288.
- Hudson, G.S., Evans, J.R., von Caemmerer, S., Arvidsson, Y.B., Andrews, T.J., 1992. Reduction of ribulose-1, 5-bisphosphate carboxylase/oxygenase content by antisense RNA reduces photosynthesis in transgenic tobacco plants. *Plant Physiol.* 98, 294–302.
- Kattge, J., Knorr, W., Raddatz, T., Wirth, C., 2009. Quantifying photosynthetic capacity and its relationship to leaf nitrogen content for global-scale terrestrial biosphere models. *Glob. Chang. Biol.* 15, 976–991.
- Knipling, E.B., 1970. Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation. *Remote Sens. Environ.* 1 (3), 155–159.
- Köhler, I.H., Ruiz-Vera, U.M., VanLoocke, A., Thomey, M.L., Clemente, T., Long, S.P., Ort, D.R., Bernacchi, C.J., 2016. Expression of cyanobacterial FBP/SBPase in soybean prevents yield depression under future climate conditions. *J. Exp. Bot.* 68 (3), 715–726.
- Kromdijk, J., Głowacka, K., Leonelli, L., Gabilly, S.T., Iwai, M., Niyogi, K.K., Long, S.P., 2016. Improving photosynthesis and crop productivity by accelerating recovery from photoprotection. *Science* 354, 857–861.
- Lawson, T., Kramer, D.M., Raines, C.A., 2012. Improving yield by exploiting mechanisms underlying natural variation of photosynthesis. *Curr. Opin. Biotechnol.* 23, 215–220.
- Lefebvre, S., Lawson, T., Fryer, M., Zakhleniuk, O.V., Lloyd, J.C., Raines, C.A., 2005. Increased sedoheptulose-1, 7-bisphosphatase activity in transgenic tobacco plants stimulates photosynthesis and growth from an early stage in development. *Plant Physiol.* 138 (1), 451–460.
- Long, S.P., Bernacchi, C.J., 2003. Gas exchange measurements, what can they tell us about the underlying limitations to photosynthesis? Procedures and sources of error. *J. Exp. Bot.* 54, 2393–2401.
- Long, S.P., Zhu, X.-G., Naidu, S.L., Ort, D.R., 2006. Can improvement in photosynthesis increase crop yields? *Plant Cell Environ.* 29, 315–330.
- López-Calcagno, P.E., Fisk, S., Brown, K.L., Bull, S.E., South, P.F., Raines, C.A., 2019. Overexpressing the H-protein of the glycine cleavage system increases biomass yield in glasshouse and field-grown transgenic tobacco plants. *Plant Biotechnol. J.* 17, 141–151.
- Martin, M.E., Aber, J.D., 1997. High spectral resolution remote sensing of forest canopy lignin, nitrogen, and ecosystem processes. *Ecol. Appl.* 7 (2), 431–443.
- Masle, J., Hudson, G.S., Badger, M.R., 1993. Effects of ambient CO₂ concentration on growth and nitrogen use in tobacco (*Nicotiana tabacum*) plants transformed with an antisense gene to the small subunit of ribulose-1, 5-bisphosphate carboxylase/oxygenase. *Plant Physiol.* 103, 1075–1088.
- Meacham, K.G., Sirault, X., Quick, W.P., von Caemmerer, S., Furbank, R.T., 2016. Diurnal solar energy conversion and photoprotection in rice canopies. *Plant Physiol.* 173, 495–508.
- Meroni, M., Rossini, M., Guanter, L., Alonso, L., Rascher, U., Colombo, R., Moreno, J., 2009. Remote sensing of solar-induced chlorophyll fluorescence: review of methods and applications. *Remote Sens. Environ.* 113, 2037–2051.
- Mevik, B.-H., Wehrens, R., 2007. The pls package: principal component and partial least squares regression in R. *J. Stat. Softw.* 18, 1–24.
- Miao, G., Guan, K., Yang, X., Bernacchi, C.J., Berry, J.A., DeLucia, E.H., Wu, J., Moore, C.E., Meacham, K., Cai, Y., 2018. Sun-induced chlorophyll fluorescence, photosynthesis, and light use efficiency of a soybean field. *J. Geophys. Res. Biogeosci.* 123 (2), 610–623.
- Monteith, J.L., Moss, C.J., 1977. Climate and the efficiency of crop production in Britain [and discussion]. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 281, 277–294.
- Nakano, H., Makino, A., Mae, T., 1997. The effect of elevated partial pressures of CO₂ on the relationship between photosynthetic capacity and N content in rice leaves. *Plant Physiol.* 115 (1), 191–198.
- Ollinger, S.V., Smith, M.L., 2005. Net primary production and canopy nitrogen in a temperate forest landscape: an analysis using imaging spectroscopy, modeling and field data. *Ecosystems* 8 (7), 760–778.
- Ort, D.R., Merchant, S.S., Alric, J., Barkan, A., Blankenship, R.E., Bock, R., Croce, R., Hanson, M.R., Hibberd, J.M., Long, S.P., 2015. Redesigning photosynthesis to

- sustainably meet global food and bioenergy demand. *Proc. Natl. Acad. Sci.* 112, 8529–8536.
- Parry, M.A.J., Reynolds, M., Salvucci, M.E., Raines, C., Andralojc, P.J., Zhu, X.-G., Price, G.D., Condon, A.G., Furbank, R.T., 2010. Raising yield potential of wheat. II. Increasing photosynthetic capacity and efficiency. *J. Exp. Bot.* 62 (2), 453–467.
- Peñuelas, J., Gamon, J.A., Fredeen, A.L., Merino, J., Field, C.B., 1994. Reflectance indices associated with physiological changes in nitrogen-and water-limited sunflower leaves. *Remote Sens. Environ.* 48 (2), 135–146.
- Poorter, H., Remkes, C., Lambers, H., 1990. Carbon and nitrogen economy of 24 wild species differing in relative growth rate. *Plant Physiol.* 94 (2), 621–627.
- Portis, A.R., 2003. Rubisco activase–Rubisco's catalytic chaperone. *Photosynth. Res.* 75 (1), 11–27.
- Quick, W.P., Schurr, U., Fichtner, K., Schulze, E.D., Rodermeil, S.R., Bogorad, L., Stitt, M., 1991. The impact of decreased Rubisco on photosynthesis, growth, allocation and storage in tobacco plants which have been transformed with antisense rbcS. *Plant J.* 1, 51–58.
- Raines, C.A., 2003. The Calvin cycle revisited. *Photosynth. Res.* 75 (1), 1–10.
- Reich, P.B., Walters, M.B., Tjoelker, M.G., Vanderklein, D., Buschena, C., 1998. Photosynthesis and respiration rates depend on leaf and root morphology and nitrogen concentration in nine boreal tree species differing in relative growth rate. *Funct. Ecol.* 12 (3), 395–405.
- Rogers, A., 2014. The use and misuse of $V_{c,max}$ in Earth System Models. *Photosynth. Res.* 119, 15–29.
- Rouse Jr., J.W., Haas, R., Schell, J., Deering, D., 1974. Monitoring Vegetation Systems in the Great Plains with ERTS.
- Serbin, S.P., Dillaway, D.N., Kruger, E.L., Townsend, P.A., 2012. Leaf optical properties reflect variation in photosynthetic metabolism and its sensitivity to temperature. *J. Exp. Bot.* 63, 489–502.
- Serbin, S.P., Singh, A., McNeil, B.E., Kingdon, C.C., Townsend, P.A., 2014. Spectroscopic determination of leaf morphological and biochemical traits for northern temperate and boreal tree species. *Ecol. Appl.* 24, 1651–1669.
- Sharkey, T.D., Bernacchi, C.J., Farquhar, G.D., Singaas, E.L., 2007. Fitting photosynthetic carbon dioxide response curves for C3 leaves. *Plant Cell Environ.* 30, 1035–1040.
- Siegmann, B., Jarmer, T., 2015. Comparison of different regression models and validation techniques for the assessment of wheat leaf area index from hyperspectral data. *Int. J. Remote Sens.* 36, 4519–4534.
- Silva-Perez, V., Molero, G., Serbin, S.P., Condon, A.G., Reynolds, M.P., Furbank, R.T., Evans, J.R., 2017. Hyperspectral reflectance as a tool to measure biochemical and physiological traits in wheat. *J. Exp. Bot.* 69 (3), 483–496.
- Simkin, A.J., McAusland, L., Headland, L.R., Lawson, T., Raines, C.A., 2015. Multigene manipulation of photosynthetic carbon assimilation increases CO₂ fixation and biomass yield in tobacco. *J. Exp. Bot.* 66 (13), 4075–4090.
- South, P.F., Cavanagh, A.P., Liu, H.W., Ort, D.R., 2019. Synthetic glycolate metabolism pathways stimulate crop growth and productivity in the field. *Science* 363 (6422), 9077.
- Suzuki, Y., Miyamoto, T., Yoshizawa, R., Mae, T., Makino, A., 2009. Rubisco content and photosynthesis of leaves at different positions in transgenic rice with an over-expression of RBCS. *Plant Cell Environ.* 32 (4), 417–427.
- Tilman, D., Socolow, R., Foley, J.A., Hill, J., Larson, E., Lynd, L., Pacala, S., Reilly, J., Searchinger, T., Somerville, C., 2009. Beneficial biofuels—the food, energy, and environment trilemma. *Science* 325, 270–271.
- Townsend, P.A., Foster, J.R., Chastain, R.A., Currie, W.S., 2003. Application of imaging spectroscopy to mapping canopy nitrogen in the forests of the central Appalachian Mountains using Hyperion and AVIRIS. *IEEE Trans. Geosci. Remote Sens.* 41 (6), 1347–1354.
- Ustin, S.L., Roberts, D.A., Gamon, J.A., Asner, G.P., Green, R.O., 2004. Using imaging spectroscopy to study ecosystem processes and properties. *AIBS Bull.* 54 (6), 523–534.
- von Caemmerer, S., 2000. *Biochemical Models of Leaf Photosynthesis*. CSIRO Publishing, Colingwood VIC 3066, Australia.
- Walker, A.P., Beckerman, A.P., Gu, L., Kattge, J., Cernusak, L.A., Domingues, T.F., Scales, J.C., Wohlfahrt, G., Wullschlegel, S.D., Woodward, F.I., 2014. The relationship of leaf photosynthetic traits – $V_{c,max}$ and J_{max} – to leaf nitrogen, leaf phosphorus, and specific leaf area: a meta-analysis and modeling study. *Ecol. Evol.* 4, 3218–3235.
- Wold, S., Sjöström, M., Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* 58, 109–130.
- Woolley, J.T., 1971. Reflectance and transmittance of light by leaves. *Plant Physiol.* 47 (5), 656–662.
- Wu, J., Chavana-Bryant, C., Prohaska, N., Serbin, S.P., Guan, K., Albert, L.P., Yang, X., van Leeuwen, W.J., Garnello, A.J., Martins, G., 2017. Convergence in relationships between leaf traits, spectra and age across diverse canopy environments and two contrasting tropical forests. *New Phytol.* 214, 1033–1048.
- Yang, K., Ryu, Y., Dechant, B., Berry, J.A., Hwang, Y., Jiang, C., Kang, M., Kim, J., Kimm, H., Kornfeld, A., Yang, X., 2018. Sun-induced chlorophyll fluorescence is more strongly related to absorbed light than to photosynthesis at half-hourly resolution in a rice paddy. *Remote Sens. Environ.* 216, 658–673.
- Yendrek, C., Tomaz, T., Montes, C.M., Cao, Y., Morse, A.M., Brown, P.J., McIntyre, L., Leahey, A., Ainsworth, E., 2016. High-throughput phenotyping of maize leaf physiology and biochemistry using hyperspectral reflectance. *Plant Physiol.* 173, 614–626.
- Zarco-Tejada, P.J., Miller, J.R., Mohammed, G.H., Noland, T.L., 2000. Chlorophyll fluorescence effects on vegetation apparent reflectance: I. Leaf-level measurements and model simulation. *Remote Sens. Environ.* 74, 582–595.
- Zhu, X.-G., Long, S.P., Ort, D.R., 2010. Improving photosynthetic efficiency for greater yield. *Annu. Rev. Plant Biol.* 61, 235–261.