Unique contributions of chlorophyll and nitrogen to predict crop photosynthetic capacity from leaf spectroscopy

Sheng Wang[1,2*], Kaiyu Guan[1,2,3*], Zhihui Wang[4], Elizabeth A. Ainsworth[1,2,5,6], Ting Zheng[4], Philip A. Townsend[4], Kaiyuan Li[1,2], Christopher Moller[5], Genghong Wu[1,2], Chongya Jiang[1,2]

1. College of Agricultural, Consumer and Environmental Sciences, University of Illinois at Urbana Champaign, Urbana, IL 61801, USA
2. Center for Advanced Bioenergy and Bioproducts Innovation, Institute for Sustainability, Energy, and Environment, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
3. National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
4. Department of Forest and Wildlife Ecology, University of Wisconsin-Madison, 1630 Linden Drive, Madison, WI 53706, USA
5. Department of Plant Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
6. USDA ARS Global Change and Photosynthesis Research Unit, Urbana, IL 61801, USA

***Corresponding author**: shengwang12@gmail.com; kaiyug@illinois.edu Tel: +1-2174197661

Author Email ID:

Sheng Wang: shengwang12@gmail.com, Kaiyu Guan: kaiyug@illinois.edu, Zhihui Wang zwang896@wisc.edu, Elizabeth A. Ainsworth: lisa.ainsworth@usda.gov, Ting Zheng tzheng39@wisc.edu, Philip A. Townsend ptownsend@wisc.edu, Kaiyuan Li kaiyuan9@illinois.edu, Christopher Moller: moller3@illinois.edu, Genghong Wu: gw8@illinois.edu, Chongya Jiang: chongya@illinois.edu

Highlights:

Leaf chlorophyll content (visible spectra) and nitrogen concentration (infrared signals) have key and unique contributions to predict maize photosynthetic capacity. RTM accurately predicts chlorophyll, while generalized PLSR estimates nitrogen better.

Abstract:

The photosynthetic capacity or CO_2-saturated photosynthetic rate (V_{max}), chlorophyll, and nitrogen are closely linked leaf traits that determine C_4 crop photosynthesis and yield. Accurate, timely, rapid, and nondestructive approaches to predict leaf photosynthetic traits from hyperspectral reflectance are urgently needed for high-throughput crop monitoring to ensure food and bioenergy security. Therefore, this study thoroughly evaluated the state-of-the-art physically-based radiative transfer models (RTMs), data-driven partial-least-squares regression (PLSR), and generalized PLSR (gPLSR) models to estimate leaf traits from leaf-clip hyperspectral reflectance, which was collected from maize (*Zea mays L.*) bioenergy plots with diverse genotypes, growth stages, treatments of nitrogen fertilizers and ozone stresses in three growing seasons. Results show that leaf RTMs considering bidirectional effects can give accurate estimates of chlorophyll content (Pearson correlation r = 0.95), while gPLSR enabled retrieval of leaf nitrogen concentration (r = 0.85). Using PLSR with field measurements for training, the cross-validation indicates that V_{max} can be well predicted from spectra (r = 0.81). The integration of chlorophyll content (strongly related to visible spectra) and nitrogen concentration (linked to shortwave infrared signals) can provide better predictions of V_{max} (r = 0.71) than only using either chlorophyll or nitrogen individually. This study highlights leaf chlorophyll content and nitrogen concentration have key and unique contributions to V_{max} prediction.

Keywords: Hyperspectral leaf reflectance; the CO_2 saturated photosynthetic rate; chlorophyll; nitrogen; partial-least-squares regression; radiative transfer model; maize; bioenergy crop

1. Introduction

Photosynthesis captures and converts solar radiation into chemical energy to drive CO_2 fixation into carbohydrates that ultimately power ecosystems and feed humanity (Ainsworth, 2018). The conservation of photosynthetic proteins and enzymes has aided the mathematical modeling of photosynthetic processes (Farquhar *et al.*, 1980; Caemmerer and Furbank, 1999). In C_4 photosynthesis models, carbon assimilation is limited by PEP carboxylation in mesophyll and Rubisco carboxylation of bundle sheath cells (Von Caemmerer, 2000; Sage & Kubien, 2007). With the ability to concentrate CO_2 around Rubisco, the photosynthesis of C_4 crops (e.g. *Zea mays, Miscanthus sinensis, and Panicum virgatum*) in current atmospheric conditions is often limited by the CO_2 saturated photosynthesis rate V_{max}, which corresponds to the maximal Rubisco carboxylation rate (Leakey *et al.*, 2019). However, due to limited *in-situ* measurements and knowledge of V_{max}, most crop and terrestrial ecosystem models ignore such variability and specify a fixed value of V_{max} for each plant functional type (Kattge *et al.*, 2009). Inaccurate temporal and spatial representation of V_{max} can cause significant uncertainties in photosynthesis models and crop yield predictions (Hu *et al.*, 2014). Thus, accurate, timely, rapid, nondestructive, and cost-effective approaches to estimate V_{max} are highly needed for yield forecasting, bioenergy production, and agricultural management.

V_{max} is sensitive to leaf nitrogen, temperature, ozone, and pathogens, and shows spatial, temporal and developmental variability (Bernacchi *et al.*, 2001; Ainsworth *et al.*, 2014; Kucharik *et al.*, 2016). Leaf nitrogen is often cited as the primary mechanism controlling V_{max}, as multiple studies have shown that V_{max} standardized to a certain temperature shows a strong relationship with leaf total nitrogen content or concentration (Walker *et al.*, 2014; Dechant *et al.*, 2017; Yendrek *et al.*, 2017). Other studies, however, reveal that the V_{max} and nitrogen relationships are complicated. For instance, plants in soils with low nitrogen availability can achieve high V_{max} per leaf nitrogen (Ainsworth and Rogers, 2007). Miner and Bauerle (2019) found nitrogen content and the Rubisco carboxylation rate were not correlated for Sunflower. The Rubisco activity of soybean did not significantly correlate to leaf nitrogen due to the excessive nitrogen storage in leaves (Koester *et al.*, 2016). For tree species aspen, maple and ash, Croft et al. (2017) found that Rubisco carboxylation rates were more sensitive to Chl than leaf nitrogen content. The high correlation between Chl and V_{max} could be explained by the nitrogen resource optimality allocation (Dewar,

1996; Evans and Clarke, 2019; Smith et al., 2019). As light and dark reactions should be well coordinated to maximize leaf photosynthesis, nitrogen investment in light harvesting Chl and Rubisco should be optimized. Although these studies imply strong connections among nitrogen, Chl and V_{max}, it is still debated whether the use of Chl (Houborg et al., 2015; Croft et al., 2017), nitrogen content/concentration (Yendrek et al., 2017; Dechant et al., 2017), or their derivatives (e.g. Chl*carotenoids, Chou et al., 2020) as the proxy for V_{max}.

Traditional methods to estimate leaf V_{max} by A/Ci (net photosynthesis / intercellular CO_2 concentration) curves from leaf gas exchange experiments (Von Caemmerer, 2013) provide accurate measurements, but are time-consuming and not suitable for high-throughput crop monitoring in the context of field phenotyping and precision agriculture. Sensing techniques such as optical reflectance, solar-induced fluorescence, or thermal infrared data are rapid, non-destructive, and cost-effective ways to quantify crop traits (Houborg *et al.*, 2013; Serbin *et al.*, 2015; Guan *et al.*, 2017). Particularly, spectroscopy can exploit spectral information of the entire optical range (400 - 2500 nm) through either physically-based radiative transfer models (RTMs, e.g. Jacquemoud and Baret, 1990; Vilfan *et al.*, 2019) or data-driven methods (Serbin *et al.*, 2012; e.g. Ainsworth *et al.*, 2014; Yendrek *et al.*, 2017) to estimate traits. These estimates include photosynthetic traits (Chl, nitrogen and V_{max}), structural parameters, chemical composition, and photo-protective pigments (Townsend *et al.*, 2003; Weber *et al.*, 2012; Singh et al., 2015). The information of diverse traits from spectroscopy provides opportunities to interpret the linkage of leaf traits and V_{max} to evaluate using Chl, nitrogen, or other traits as the proxy of V_{max}.

The RTM approaches, for instance the pre-calibrated generalized plate-based turbid medium PROSPECT (Jacquemoud and Baret, 1990) or PROSPECT-DyN (Wang *et al.*, 2015, 2018*b*) models, have merits to operationally predict foliar traits such as Chl, nitrogen and water content across species, growth stages, and environmental conditions. However, the accuracy and number of predictable traits from RTMs are limited (Verrelst *et al.*, 2019). For instance, due to the weak absorption features of Rubisco protein, existing RTMs do not include the spectral absorption coefficients of Rubisco protein to directly predict V_{max}. The prediction of V_{max} through RTMs often relies on statistical regression with RTM based traits (Houborg *et al.*, 2013; Croft *et al.*, 2017; Dechant *et al.*, 2017) or detecting photosynthetic functioning (Vilfan *et al.*, 2019; Zheng and Chen,

2017; Bayat *et al.*, 2018). Additionally, leaf RTMs are developed with hemispherical reflectance measured with integrating spheres. These reflectance measurements are different from conical reflectance (Schaepman-Strub *et al.*, 2006) collected by the leaf-clip, which is the high-throughput approach for spectra collection in fields. To solve the above issues, the close-range spectral imaging of leaves model (COSINE, Jay *et al.*, 2016), which accounts for bi-directional reflectance factors, needs to be combined with PROSPECT to simulate leaf conical reflectance. Conversely, data-driven approaches, e.g. partial-least-squares regression (PLSR), can flexibly fit leaf spectra with diverse measured traits with high predictive performance. For instance, a few studies have demonstrated that V_{max} can be accurately estimated from leaf spectra (Serbin *et al.*, 2012; Yendrek *et al.*, 2017; Wu *et al.*, 2019). However, the performance of PLSR models can vary significantly depending on species, plant growth stages, and sensor configurations (Wang *et al.*, 2019). In addition, PLSR requires sufficient samples of measured traits for model training, which is less operational compared with RTMs. The pre-trained generalized PLSR (gPLSR) models, which were developed from a large database of *in-situ* observations, can be promising for applications lacking measured trait data for modeling training (Wang *et al.*, 2020). The traits Chl, nitrogen, and V_{max} can be retrieved from leaf spectra through RTMs, PLSR, or gPLSR, but it remains uncertain for the performance comparison of these approaches. A comprehensive evaluation of these approaches to quantify Chl, nitrogen, and V_{max} for high-throughput crop monitoring is highly needed.

Maize (*Zea mays L.*) is one of the major nitrogen-deficient staple and bioenergy crops, which represents a model for species with the C_4 photosynthesis pathway. Due to environmental factors or management strategies (e.g. shortage of nitrogen fertilizers), maize V_{max} is often suppressed and the average yield reaches only 64% of maximum potential globally (Neumann *et al.*, 2010). In this study, we collected leaf gas exchange measurements, leaf-level hyperspectral reflectance, nitrogen and chlorophyll data from maize experimental plots with various genotypes, growth stages, treatments of nitrogen fertilizers and ozone stress during three growing seasons. The objective was to develop and evaluate spectroscopy approaches for estimating photosynthetic traits from leaf-clip spectra and to understand the relationship among Chl, nitrogen, and V_{max}. Two key questions were addressed: (1) Among RTMs, PLSR, and gPLSR approaches, which method performs best to estimate Chl and nitrogen from the leaf-clip reflectance? (2) Can we utilize leaf spectra or

spectra-based traits to accurately estimate V_{max}? If so, what are the key spectra or traits for V_{max} prediction? By answering these questions, this study aimed to identify the operational approaches to predict photosynthetic traits using leaf hyperspectral reflectance and understand the linkage among Chl, nitrogen, and V_{max} prediction.

2. Materials and methods

2.1 Leaf spectra and photosynthetic trait measurements

The maize experimental plots of diverse genotypes were treated with nitrogen fertilizers (maize nitrogen plots, Fig. S1) and ozone (SoyFACE, Fig. S1) in Champaign, Illinois, and measured during the growing seasons of 2014, 2015 and 2019. The ozone experiment was previously described by Yendrek et al. (2017), along with hyperspectral reflectance, gas exchange, and biochemical data from the ozone experiments. The nitrogen experiment was designed with management practices of different nitrogen fertilization amounts (0, 50, 100, 150, 200 and 250 pounds per acre), time (planting, V6, and V10 stages) and approaches (middle-row injection and in-row dribble). The ozone and nitrogen fertilization experiments provided test cases to evaluate the approaches to retrieve photosynthetic traits from leaf-clip spectra, and further to identify the relationship among leaf V_{max}, nitrogen, and Chl.

Leaf reflectance spectra (500 - 2400 nm) were acquired from the central section of the leaf adaxial surface using ASD FieldSpec 4 Standard Res full-range spectroradiometers (Analytical Spectral Devices Inc., Colorado, the USA) equipped with an illuminated leaf-clip contact probe. A/Ci curves were measured with LI-COR 6400 and 6800 portable photosynthesis systems (LI-COR Inc., Nebraska, the USA) after measurements of leaf-clip reflectance. The A/Ci measurements were conducted with the leaf positioned in the chamber with air humidity of 55% and leaf temperature close to the ambient conditions (25-32 °C). The leaf adaxial side was placed facing the light source with an intensity of 2000 $\mu mol \cdot m^{-2} \cdot s^{-1}$. For each A/Ci curve, the ambient CO_2 concentrations were set to the sequences of 400, 50, 100, 150, 250, 350, 500, 700, 900, and 1200 ppm. To ensure the accuracy of measured V_{max} from gas-exchange measurements, we have tested the reproducibility of LiCOR machines in our experiments. This study followed the Yendrek et al. (2017) protocol to use the horizontal asymptote of a four-parameter non-rectangular hyperbolic function to process the measured A/Ci curves to estimate V_{max}. Furthermore, we collected V_{max} measurements of the

same leaves with different leaf temperatures to quantify the V_{max} temperature response curve using the Q10 formula and temperature inhibition curves (Leuning, 2002). By using the fitted temperature response curve of maize, V_{max} measurements of this study were standardized to the reference temperature. 25°C is a commonly used reference temperature to normalize the temperature impacts on V_{max}. However, this study selected 27 °C as the reference temperature due to the following reasons. 60% of the field V_{max} measurements were collected at 27 °C and using $V_{max,27}$ could reduce uncertainties of temperature normalization. Furthermore, as 27 °C is close to ambient temperature in the peak growing season of our study site, using $V_{max,27}$ agrees with our desire to estimate biochemical limitations to photosynthesis close to growth temperature. Additionally, with the measured temperature correction curve of Fig. S3, all $V_{max,27}$ relationships of this study can be converted to $V_{max,25}$ ($V_{max,25} = 0.875V_{max,27}$).

After A/Ci measurements, leaf tissues were sampled using a cork borer and stored in liquid nitrogen. The wet laboratory experiments were conducted to measure Chl content and nitrogen concentration of leaf samples. For Chl, one leaf disc (approximately 1.4 cm^2) was incubated in 96% (v/v) ethanol to determine Chl content using the equations of Lichtenthaler and Wellburn (1983). Three leaf discs were dried in an oven (60 °C) for three weeks to determine leaf dry mass. An analytical balance (ME204TE/00, Mettler Toledo Inc., Ohio, the USA) was used to measure the dry matter weight per area (Cdm, g cm^{-2}) of leaf samples. The dried leaf tissues were ground to a fine powder and combusted with oxygen in an elemental analyzer (Costech 4010, Costech Analytical Technologies Inc., California, the USA). The nitrogen per mass (N_{mass}, %) was determined by comparing experimental samples with an acetanilide standard curve. In total, we collected 460 leaf spectra, 297 leaf V_{max} measurements, 177 leaf Chl measurements, 350 leaf N_{mass} and Cdm measurements. Raw data and experiment sources of these measurements can be found in the supplementary Dataset S1. A correlation matrix among the measured photosynthetic traits of $V_{max,27}$, Chl, N_{mass}, and nitrogen per area (N_{area}, N_{mass}× Cdm, mg cm^{-2}), was calculated to characterize their relationships.

2.2 Models to predict traits from leaf spectra

2.2.1 Radiative transfer modeling

The PROSPECT models (Jacquemoud and Baret, 1990) were employed in this study to simulate leaf hemispherical reflectance over the optical domain from 400 to 2500 nm to retrieve the photosynthetic traits Chl and nitrogen. PROSPECT-D can utilize leaf reflectance to estimate multiple traits, e.g. leaf structure parameter (N), leaf Chl content (Chl, $\mu g/cm^2$), equivalent water thickness (Cw, cm), leaf mass per area (Cdm, g/cm^2), and the senescent (brown) materials (Féret *et al.*, 2017). The PROSPECT-DyN model was utilized to incorporate protein, cellulose, and lignin by recalibrating spectral absorption coefficients (Wang *et al.*, 2015, 2018*b*). As leaf protein strongly linearly correlates with nitrogen (Yeoh and Wee, 1994), estimated protein content was converted to the nitrogen content. PROSPECT was developed to simulate the hemispherical reflectance of leaves, but the leaf-clip reflectance collected in this study was conical. To convert the conical reflectance to the hemispherical, the COSINE model (Jay *et al.*, 2016) was implemented with PROSPECT models. Detailed information about the parameters of these three models can be found in Table 1.

The retrieval of foliar traits was conducted through a numerical inversion of RTMs by minimizing the root mean square deviation (RMSD) between the measured leaf-clip and simulated reflectance. The numerical optimization procedure used the same constrained Powell's line-search method as Féret et al. (2017). As nitrogen is sensitive to the shortwave infrared, a two-step retrieval following Wang et al. (2015) was performed. The first step was to use the entire optical domain 500-2400 nm to invert the leaf structural parameter (N) and Chl content (Chl). Then we applied the shortwave infrared domain 2100-2300 nm to invert PROSPECT-DyN to estimate the protein content (Cp), which was further converted to nitrogen content using the ratio of 4.43 (Wang *et al.*, 2018*b*).

2.2.2 Partial Least Squares Regression (PLSR)

The PLSR approach has been widely applied to process hyperspectral reflectance with high collinearity. PLSR can minimize predictor variables to a few orthogonal latent components (Geladi and Kowalski, 1986; Wold *et al.*, 2001). In this study, we selected PLSR to develop models to predict photosynthetic traits (i.e. Chl, nitrogen, and V_{max}) from the measured leaf spectra. We conducted four-fold cross-validation to split the collected spectra and traits into training and testing. In each training data set, the model between leaf spectra and traits was developed. Then this model was tested using the independent testing data set. The uncertainty analysis of the PLSR models

was conducted by splitting the training dataset via 100 permutations and generating the new model coefficients following Meerdink et al. (2016). Then we used the ensemble mean of PLSR models to predict the photosynthetic traits. By doing so, we obtained both predictive values and uncertainties for traits. Notably, the leaf reflectance from 500 to 2400 nm was utilized to develop the PLSR models between spectra and V_{max} or Chl. However, as nitrogen is well known to be linked to the shortwave infrared (Curran, 1988; Serbin *et al.*, 2014; Yendrek *et al.*, 2017), we used leaf reflectance of 1500-2400 nm to generate PLSR models to predict nitrogen. To avoid overfitting between spectra and foliar traits, we optimized the number of PLSR components by minimizing the prediction residual sum of squares (PRESS) statistic (e.g. Meerdink *et al.*, 2016). PRESS of successive model components was calculated through a cross-validation analysis. We selected model components corresponding to the minimum PRESS statistic until successive PLSR components did not significantly increase the model predictive accuracy (Serbin *et al.*, 2014).

2.2.3 Generalized Partial Least Squares Regression (gPLSR)

The generation of PLSR models requires sufficient measured traits to be collected for modeling training, which could limit the applicability of PLSR in a fast and operational manner. To deal with such limitations, we tested the pre-trained gPLSR models (Wang *et al.*, 2020) to predict leaf Chl and nitrogen. The gPLSR models were generated from a database of leaf spectra and traits of 40 species (including maize) across NEON field sites in the Eastern U.S (data available from doi:10.21232/e2jt-5209 and model code at ecosml.org). The pre-trained gPLSR model has advantages of free calibration and only requires leaf spectra data to predict foliar traits. In this study, we tested whether such gPLSR models can be applied to agricultural sites with different environmental conditions and sampling time.

2.3 Model application and evaluation

Our workflow to compare the predictive ability of RTMs, PLSR, and gPLSR to estimate leaf Chl, nitrogen, and V_{max} is illustrated in Fig. 1. We evaluated the performance of the PLSR, gPLSR, and RTMs to estimate leaf Chl and nitrogen. Then we conducted a comparison of various approaches to estimate V_{max}. The first approach utilized the leaf Chl or nitrogen to develop the linear regression models to estimate V_{max}. We conducted four-fold cross-validation to evaluate the performance of these linear regression models. Then we tested the accuracy of using the leaf reflectance data to

develop a PLSR model to estimate V_{max}. We also applied RTMs and gPLSR to estimate various leaf traits. Then through these estimated ten traits (N, Chl, Car, Ant, Cs, Cw, Cdm, B, N_{mass} and Ccl), we developed the trait-based PLSR model to predict V_{max}. Furthermore, as Chl and N_{mass} are two commonly used variables to proxy V_{max}, we also compared the performance of using only Chl, N_{mass}, and their multiplication to predict V_{max}. We used the comparison of spectra-based and trait-based PLSR models to identify the accurate and robust approaches to estimate V_{max}.

To comprehensively evaluate the estimated crop traits from leaf spectra, the Taylor diagram (Taylor, 2001) was used to present these three complementary statistics with a triangle-cosine-law relationship: the Pearson correlation coefficient (r), normalized standard deviation (NSTD, as Eq. 1), and normalized unbiased root-mean-square deviation (NubRMSD, Eq. 2). The radial distance stands for the NSTD and the angle in the polar plot represents r. The reference point on the X-axis with r=1, NSTD=1 and NubRMSD=0 refers to the observation. The distance from the simulation point to the reference point represents NubRMSD of simulations and stands for the integrated performance for the simulation. The closer distance from simulation points to the reference point indicates better simulation performance.

$$NSTD_{sim} = STD_{sim}/STD_{obs} \quad \text{Eq. (1)}$$

$$NubRMSD^2_{obs,sim} = NSTD^2_{obs} + NSTD^2_{sim} - 2NSTD_{obs}NSTD_{sim}\, cos\, r_{obs,sim} \quad \text{Eq. (2)}$$

Where *sim* and *obs* represent the simulation results and the observations, respectively. *N* is the total number, with the subscript *i* representing the number *i* of simulations or observations. $r_{obs,sim}$ refers to the correlation coefficient between simulations and observations and NSTD is the normalized standard deviation.

2.4 Analysis of spectral signatures on predicting traits

To identify the contribution of spectral wavelengths to the prediction of Chl, nitrogen and V_{max}, both physical model and statistical method based sensitivity approach were applied. Through the comparison of these two approaches, this study can get a comprehensive understanding of the relationship between spectra wavelengths and the prediction of leaf traits.

In the physical model based approach, we conducted a global sensitivity analysis of the PROSPECT-DyN-COSINE using the Sobol method (Sobol, 2001; Saltelli *et al.*, 2004), which is based on analysis of variance decomposition to calculate the sensitivity of coupled inputs. The Sobol analysis can quantify the contribution of model parameters (leaf traits) to the wavelengths of leaf reflectance. The first order Sobol sensitivity quantifies the independent contribution from each input to the output variables, while the second-order sensitivity quantifies interactions between every two inputs to the output variable. The Sobol analysis is sensitive to the configuration of the model parameter range and distribution. As this study focused on the maize photosynthetic traits, we utilized the collected 460 leaf spectra to invert RTMs to obtain the parameter distribution. Then, the kernel density sampling method was applied to generate the input data for sensitivity analysis. The kernel density sampling method has the advantage of resembling the distribution of the sampled dataset (Wang *et al.*, 2018*a*). According to the kernel density distribution of model parameters, 20,000 samples were generated to assess the sensitivity of simulated leaf spectra to traits. Additionally, to our best knowledge, this study is the first one to integrate PROSPECT-DyN and COSINE for the retrieval of foliar traits. Such sensitivity analysis can also give insights into evaluating the impacts of incorporating COSINE for reflectance simulation.

In the statistical approaches, the PLSR loading, coefficients, and Variable Influence on the Projection (VIP) scores (Wold *et al.*, 2001) were computed. The wavelengths with high absolute values of loading, coefficients, and VIP scores indicate a high contribution to the leaf trait prediction. The similarity and difference of the model loadings, coefficients and VIP scores of the PLSR models to predict Chl, N_{mass}, and V_{max} were compared to explore the spectral linkage among these key photosynthetic traits. Notably, this study did not employ RTMs to directly predict V_{max}. The sensitivity of spectral wavelength to V_{max} prediction focused on the PLSR approach. In addition, we also analyzed the PLSR loading of using these estimated ten traits (N, Chl, Car, Ant, Cs, Cw, Cdm, B, N_{mass} and Ccl) to predict V_{max}. The VIP scores and loading analysis can show the linkages of these ten traits to V_{max}.

3. Results

3.1 Measured leaf spectra and photosynthetic traits

The measured leaf spectra (Fig. 2a) followed a general pattern of low reflectance in the visible region (500-700 nm), high reflectance in the near-infrared region (700-1300 nm) and two water absorption features in the shortwave infrared region (1300-2400 nm). The coefficient of variation (CV) of the spectral data (Fig. 2b) indicated that the visible, red edge (700-750 nm), and shortwave infrared regions, which are strongly linked to leaf pigments and biochemical traits, have greater variability compared to the near-infrared, which reflects the structural parameters of leaves (e.g. leaf thickness and dry matter content).

After V_{max} was standardized to 27 °C using the fitting temperature response curve in Fig S2 ($V_{max,25}$ = $0.875V_{max,27}$), the measured leaf biochemical and photosynthetic traits, Chl, N_{mass}, N_{area}, and $V_{max,27}$, were all highly correlated (Fig. 3). Nonetheless, there were differences in the strength of trait correlations. Among all pairs, Chl and N_{area} had the highest linear correlation (r = 0.89), and this high correlation indicated that maize tended to allocate leaf total nitrogen to Chl at a relatively constant rate. Both Chl and N_{mass} were highly correlated with $V_{max,27}$ (r = 0.77 and 0.75, respectively), confirming previous findings that $V_{max,27}$ of maize is highly correlated with Chl (Houborg *et al.*, 2013; Croft *et al.*, 2017) and N_{mass} (Yendrek *et al.*, 2017). The large variabilities of measured photosynthetic traits can serve a robust dataset for testing the model performance to predict traits.

3.2 Predicted photosynthetic traits from leaf spectra

The results of comparing RTM, PLSR and gPLSR (Fig. 1) to estimate leaf photosynthetic traits are shown in the Taylor Diagram (Fig. 4). For leaf Chl, PLSR achieved the highest r of around 0.95, lowest NubRMSD of about 0.33, and NSTD close to 1. The RTM approach also achieved high performance with r around 0.95 and NubRMSD around 0.45. The gPLSR approach can obtain good performance with r of 0.88 and NubRMSD of 0.48. For N_{mass}, the PLSR method showed the highest r of around 0.96 and NubRMSD of 0.28. The gPLSR approach can also obtain a relatively good prediction of nitrogen with r of about 0.85 and NubRMSD of 0.56. The predictive power of the RTM (PROSPECT-DyN-COSINE) was weaker with r of around 0.60. Detailed scatterplots of predicting Chl and N_{mass} are illustrated in Fig. S4.

For $V_{max,27}$ predictions, the best performance was achieved by the spectra based PLSR model with r of 0.81, NubRMSD of around 0.61, and NSTD close to 1. The trait-based PLSR model utilizing ten spectra based traits (N, Chl, Car, Ant, Cs, Cw, Cdm, B, N_{mass} and Ccl) to predict $V_{max,27}$ also demonstrated a good predictive skill with r of about 0.72 and NubRMSD of 0.70. These two PLSR models showed better performance than the linear regression models based on either Chl or N_{mass}. The linear regression models based on Chl or N_{mass} achieved similar and moderate prediction performance with r of around 0.6. However, the predictive performance of the linear model significantly improved by using Chl*N_{mass}. The linear regression model between Chl*N_{mass} and V_{max} can achieve r of around 0.71 and NubRMSD of 0.70, which is close to the performance of the trait-based model (Fig. S5). This result indicates that Chl and N_{mass} play a major role in the prediction of $V_{max,27}$ in the trait-based PLSR model.

3.3 Contribution of spectral signatures on predicting traits

In the RTM based spectra contribution analysis, this study retrieved parameter distribution (Fig. S3) from the collected 460 leaf spectra. Then, the global sensitivity analysis results of PROSPECT-COSINE and PROSPECT-DyN-COSINE for the case of simulating maize leaf-clip reflectance were conducted as Fig. 5. In the visible region, pigments including Chl, Car, Ant, and Cs contributed to the reflectance variation (Fig. 5a), with red edge and green wavelengths (500-750 nm) influenced primarily by Chl. The leaf structural parameter N, which indicates the leaf thickness, and dry matter content (protein, cellulose and lignin in PROSPECT-DyN, Fig. 5b) contributed to the variability of reflectance in near-infrared and shortwave infrared regions that was not explained by B. In particular, the shortwave infrared 1500-1900 nm and 2000-2400 nm are the main wavelengths exhibiting the nitrogen signal (Cp on Fig. 5b). The parameter B representing the bidirectional reflectance factor of leaves showed a significant contribution to the spectral variability across visible, near-infrared and shortwave infrared, especially in blue and red wavelengths and the water absorption feature around 1900 nm. This high contribution indicated the importance of considering the bidirectional effects of leaf reflectance collected from a handheld leaf-clip spectroradiometer (Li *et al.*, 2019). In general, from the model-based contribution analysis, the visible information (500-750 nm) has strong implications for Chl estimation, while the shortwave infrared bands (1500-1900 and 2000-2400 nm) are important for nitrogen prediction.

In the statistical analysis, the VIP scores, loading, and coefficients of the spectra based PLSR models were compared to analyze the similarity and difference of using spectra to predict $V_{max,27}$, Chl and N_{mass} (Fig. 6 a-c). In general, the visible wavelengths associated with green reflectance and red absorption (550 and 710 nm) contributed most significantly to the prediction of Chl, while the SWIR wavelengths in the 1700-1900 nm and 2100-2200 nm SWIR regions were most important to the prediction of nitrogen. These findings also agree with the model-based sensitivity analysis (Fig. 5) and confirm the robust performance of RTMs. The shaded grey regions in Fig. 6 correspond to the high absolute values of VIP scores for predicting $V_{max,27}$. In the visible part of the spectrum (500 - 750 nm), the VIP scores, loadings, and coefficients of $V_{max,27}$ and Chl were very similar. Specifically, the green and red edge (550 and 710 nm) largely contributed to the prediction of $V_{max,27}$. In the shortwave infrared region, the patterns of VIP scores, loadings, and coefficients for $V_{max,27}$ were close to those for N_{mass}. The PLSR models of $V_{max,27}$ and N_{mass} shared key wavelengths such as 1590, 1830, 1910, 2030 and 2110 nm. These results indicate that the spectra signals of Chl and N_{mass} have complementary contributions to the prediction of $V_{max,27}$. However, notably, there are also unique wavelengths such as 1500 nm, 2200 nm, and 2300 nm contributing to the prediction of $V_{max,27}$ that are not strongly related to Chl or N_{mass}.

Similar to the analysis of spectra-based PLSR models, the VIP scores and loading of the trait-based PLSR model also supported the findings on the large contribution of Chl and N_{mass} to $V_{max,27}$ predictions. In the VIP scores of trait-based PLSR model (Fig. 7a), Chl and N_{mass} were the two strongest contributors to the prediction of $V_{max,27}$. The analysis of the components 1 and 2 of PLSR loading (Fig. 7b) showed that Chl largely contributed to the first component of PLSR loading. N_{mass} had a contribution to the first component but also provided unique information in the second component. This analysis indicated that Chl and N_{mass} had shared but also unique contributions to the prediction of $V_{max,27}$. In the VIP scores for the trait-based PLSR model, Car and Cw showed a high contribution to the model prediction following Chl and N_{mass}. This contribution was likely due to the high correlation between Chl and Car (Kopsell *et al.*, 2004). Under drought conditions, low water availability can alter nitrogen uptake and thus results in a high correlation between leaf water content and V_{max} (Camino *et al.*, 2019).

4. Discussion

We provided a comprehensive evaluation of spectroscopy methods to retrieve Chl, $N_{mass,}$ and V_{max}. These analyses could be helpful for the model selection to estimate leaf photosynthetic traits in high-throughput crop monitoring. First, for pre-trained approaches, the PROSPECT-D coupled with the COSINE model showed a strong ability to predict Chl, while gPLSR predicted leaf nitrogen better. With field measurements for model training, PLSR showed the best performance to predict foliar traits. Second, the spectra-based or trait-based PLSR models can provide accurate and effective means to predict V_{max}. We also found that Chl and N_{mass}, which are strongly linked to visible and shortwave infrared signals respectively, showed shared and unique contributions to the prediction of V_{max}. Measurement and model uncertainties, implications on RTM and PLSR model selection, and mechanisms of controlling V_{max} based on these results are further discussed.

4.1 Uncertainty for photosynthetic capacity prediction

Compared with Chl and N_{mass} prediction, achieving high accuracy to predict V_{max} through spectroscopy has more challenges. These challenges are partially due to the limited amount and weak absorption features of Rubisco enzyme. Furthermore, uncertainties in field V_{max} measurements and models may also contribute to the performance of V_{max} prediction.

This study used commercial gas exchange systems to obtain the A/Ci curves to fit the horizontal asymptote of a four-parameter non-rectangular hyperbolic function (Yendrek et al., 2017) to quantify V_{max}. However, gas leakage, chamber edge effects, and lateral flux through leaf air space could bring uncertainties for A/Ci curves, when operating systems in fields (Long and Bernacchi, 2003). In addition, the Rubisco capacity V_{max} derived from gas-exchange measurements is not always equal to the amount of Rubisco protein present (Crafts-Brandner & Salvucci, 2000). However, this study carefully screened all A/Ci curves to exclude the bad fitting of measurement curves as Kauwe et al. (2016). As multiple machines were employed, we have also tested the reproducibility of machines to ensure similar A-Ci curves obtained from different machines for the same leaf. Furthermore, our V_{max} measurements are comparable to estimates reported in previous studies (Houborg et al., 2013; Yendrek et al., 2017; Miner and Bauerle et al., 2019). All these strategies ensure the high accuracy of the measured V_{max} data for this study.

To diagnose the performance of spectra-$V_{max,27}$ model, we further analyzed the relationships between model prediction residuals with leaf conditions, environmental stressors, experiment year, and genotypes (Fig. 8). The comparison between model residuals and leaf $V_{max,27}$ (Fig. 8a) show model overestimation of $V_{max,27}$ when leaf $V_{max,27}$ is low, while the model underestimates $V_{max,27}$ when leaves have high $V_{max,27}$. We also found the model residuals exhibit dependence on O_3 treatment (Fig. 8b), which indicates that O_3 can alter the leaf spectra and traits relationship (Yendrek et al., 2017). Similarly, the spectra-trait model also shows a large difference when applying to different genotypes (Fig. 8d). However, we did not find a significant difference for model performances in different year data (Fig. 8c), which demonstrates the transferability of PLSR models for plants across growth stages (Wang et al., 2019).

4.2 Selection of physically-based and data-driven approaches

RTMs are developed based on physically based radiative transfer processes and thus have high accuracy to utilize the observed leaf spectra to accurately predict traits with strong absorption features, such as pigments. For instance, this study demonstrated the high accuracy of PROSPECT-COSINE to estimate Chl ($r = 0.94$) in maize. Compared to pigments, protein has relatively weaker absorption features in shortwave infrared and RTM showed moderate accuracy to estimate N_{mass} (Fig. 4). Data-driven methods such as PLSR have the advantage of exploiting spectral signatures to link reflectance with *in-situ* measurements to accurately predict traits such as N_{mass} ($r = 0.96$). However, the development of PLSR models requires collecting a large data set of foliar traits for model training, and models may not be applicable outside the conditions of *in-situ* collections. In practical applications, the pre-trained gPLSR models, which can be implemented without field measured traits, have high flexibility and accuracy to predict traits such as N_{mass} ($r = 0.85$).

Regarding V_{max} prediction, the spectra-based PLSR model in this study achieved the highest accuracy ($r = 0.81$). The trait-based PLSR model achieved slightly worse but still reasonably good performance ($r = 0.72$). The integration of Chl and N_{mass} can also achieve good predictive performance ($r = 0.71$). For leaf scale applications, the spectra based models show great potential. However, such leaf spectra models have challenges to be directly applied to the canopy scale, as spectra vary significantly across leaf and canopy scales. The upscaling of reflectance from leaf to

canopy is also highly nonlinear, due to light scattering throughout the canopy profile, sensor viewing angles, solar radiation angles, fraction of sunlit and shaded leaves (Verhoef, 1984). The upscaling of V_{max} from leaf to canopy is also highly nonlinear, but the process is influenced by nitrogen allocation throughout the canopy profile and within leaves (Wright and Hammer, 1994; Evans and Clarke, 2019). Thus, the leaf scale spectra-Vmax relationship could be hardly used to the canopy spectra, as different mechanisms involved in the upscaling of spectra and V_{max} from leaf to canopy.

To predict V_{max} across leaf, canopy, regional, or global scales, the trait-based V_{max} model has more flexibility (Houborg et al., 2013; Luo et al., 2019). For instance, Houborg et al. (2013) showed that using the leaf Chl-V_{max} relationship along with satellite-derived chlorophyll content, the community land model achieved an improved estimation of canopy GPP. Similarly, Luo et al. (2019) applied such leaf Chl-V_{max} relationship to the global scale to derive terrestrial photosynthesis. In these studies, leaf traits were retrieved from the canopy reflectance through RTMs (Jacquemoud *et al.*, 2009) and then the trait-V_{max} relationship were applied to derive photosynthetic capacity.

4.3 Foliar nitrogen allocation and photosynthetic capacity prediction

Photosynthesis requires a large number of proteins, e.g., Rubisco and light-harvesting complex, which account for 69-75% of the nitrogen in leaves (Makino and Osmond, 1991; Onoda *et al.*, 2017). Around 25-31% nitrogen is allocated to the non-photosynthetic components such as cell walls, mitochondria, peroxisomes, and the cytosol, as shown in Fig. 9 (Mu *et al.*, 2016; Evans and Clarke, 2019). The nitrogen allocation to Rubisco and other components show strong variability depending on species, growth stages and environmental conditions (Evans and Clarke, 2019). For instance, Onoda et al. (2017) found that when leaves increased leaf dry mass per area, the fraction of leaf nitrogen allocated to Rubisco declined to compensate for the increased allocation to the cell wall materials. Due to the greater photosynthetic rate per unit leaf nitrogen in young leaves, V_{max} showed strong variations with leaf ages (Albert *et al.*, 2018; Wu *et al.*, 2019). The proportion of photosynthetic proteins in maize showed large variations with treatments of nitrogen fertilizers (Mu *et al.*, 2016). Understanding leaf nitrogen allocation is important for V_{max} prediction.

514 The proposed approach in this study (Fig. 9), which estimates Chl and total nitrogen through the
515 visible and shortwave infrared spectra respectively, can integrate Chl and nitrogen information to
516 infer nitrogen allocation to predict V_{max}. Compared to the remote sensing approaches utilizing
517 either Chl or total nitrogen to approximate V_{max} (Houborg *et al.*, 2013; Dechant *et al.*, 2017), this
518 proposed approach has greater potential for V_{max} retrieval. For instance, chlorophyll deficit
519 tobaccos have a much lower Chl-to-V_{max} ratio than normal species (Meacham-Hensold *et al.*,
520 2019). Using a universal Chl and V_{max} relationship may underestimate V_{max} in such species.
521 However, with additional nitrogen information, the prediction of V_{max} could be improved.
522 Likewise, use of total nitrogen to predict V_{max} may result in low correlations for species such as
523 soybean (Koester *et al.*, 2016) due to excessive nitrogen storage. The additional information of
524 Chl could thus be vital to improving the prediction of soybean V_{max}. Moreover, the sensing
525 techniques provide estimates of the pool sizes for leaf nitrogen components, e.g., Chl or total
526 nitrogen. To further constrain V_{max} prediction, the optimality theories on plant resource allocation
527 (Smith *et al.*, 2019) can be leveraged to combine with the retrieved nitrogen components from
528 sensing techniques. For natural ecosystems or nitrogen deficit crops, plants tend to maximize
529 carbon gains with improving nitrogen allocation among leaf nitrogen pools (Quebbeman and
530 Ramirez, 2016). With such information about nitrogen allocation, the prediction of V_{max} could be
531 further improved. Towards operational prediction of V_{max} from hyperspectral reflectance with less
532 dependency on model training, the integration of RTM derived Chl and gPLSR derived N_{mass} to
533 develop the generalized model for V_{max} prediction shows great potential.

534 **5. Conclusion**

535 The accurate, fast, nondestructive, and cost-effective approaches to estimate photosynthetic traits,
536 such as CO_2-saturated photosynthesis rate (V_{max}), chlorophyll, and nitrogen, are highly needed for
537 crop monitoring. This study comprehensively evaluated radiative transfer models (RTMs), partial
538 least-squares regression (PLSR), and generalized PLSR (gPLSR) to retrieve photosynthetic traits
539 from leaf-clip reflectance collected in diverse maize plots with different genotypes, growth stages,
540 treatments of nitrogen fertilizers and ozone pollution in three growing seasons. This study led to
541 the following conclusions: (i) Both pre-trained RTM and gPLSR methods have great potential to
542 estimate photosynthetic traits. RTMs can achieve a high performance to retrieve foliar pigments
543 such as chlorophyll content ($r = 0.95$). gPLSR can be used to estimate foliar nitrogen concentration

(r = 0.85). (ii) With model training, PLSR methods can exploit leaf reflectance in conjunction with field samples to achieve high accuracy to predict traits. The PLSR models based on spectra (r = 0.81) or the spectra retrieved traits (r = 0.72) can provide good predictions of V_{max}. In particular, the trait-based V_{max} model has the ability to be applied across spatial scales, i.e. using either leaf or canopy level data. (iii) We found that leaf chlorophyll content and nitrogen concentration showed complementary contributions to the prediction of V_{max}. The integration of leaf chlorophyll and total nitrogen information, which indicates leaf chlorophyll nitrogen and total nitrogen pool sizes respectively, can significantly improve V_{max} prediction (r = 0.71) than that using only chlorophyll or nitrogen. The information on nitrogen allocation among nitrogen pools is vital for V_{max} predictions.

This study provided new insights into improving V_{max} prediction by sensing both chlorophyll and nitrogen for maize. Such approaches could also be applied to other crops, e.g. perennial bioenergy C_4 grasses. Further, applying estimated photosynthetic traits from such approaches into the terrestrial ecosystem models could significantly improve the ability to predict crop yields and carbon cycles. Leveraging the advanced imaging spectroscopy approaches on towers, unmanned or manned airborne systems, or satellites such as PRISMA (launched in 2019), HISUI (launched in 2019), EnMAP (expected launch in 2021), and NASA SBG and ESA CHIME (expected launches in late 2020s), we can extend the leaf retrieval to the canopy and regional scale for high-throughput and large-scale agricultural monitoring to improve food and bioenergy production.

Supplementary data

Supplementary data are available at JXB online.

Fig. S1. Overview of the study site.

Fig. S2. Fitted V_{max} temperature correction curve for Maize.

Fig. S3. Retrieved distribution of the PRO-COSINE and PRODyN-COSINE parameters from the measured 470 maize leaf reflectance.

Fig. S4. Scatterplots of predicting (a-c) leaf chlorophyll content and (d-f) nitrogen concentration from leaf-clip reflectance.

Fig. S5. Scatterplots of predicting V_{max} from leaf spectra or spectra based traits.

Dataset S1. Measured leaf traits, measured reflectance, and generated spectra-trait PLSR models.

Acknowledgment

This work was supported by the DOE Center for Advanced Bioenergy and Bioproducts Innovation (U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018420). Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the U.S. Department of Energy. We would also like to thank the support from the seed funding from the Center for Digital Agriculture, National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign. K.G. is funded by NASA New Investigator Award and Carbon Monitoring System program (NNX16AI56G and 80NSSC18K0170) managed by the NASA Terrestrial Ecology Program, and USDA National Institute of Food and Agriculture (NIFA) Foundational Program award (2017-67013-26253, 2017-68002-26789, 2017-67003-28703). Z.W., T.Z. and P.T. received support through NASA Jet Propulsion Laboratory award 1638464, NSF Macrosystems Biology grant 1638720 and USDA Hatch award WIS01874. We would also like to thank the editor Dr. Tracy Lawson and two anonymous referees for suggestions and comments that improved the paper.

Data Availability

All data supporting the findings of this study are available within the paper and within its supplementary materials published online.

Author contributions

S.W., K.G., Z.W., E.A.A. and P.T. conceived the project. K.G., E.A.A. and P.T. contributed to funding acquisition. S.W., Z.W., T.Z., K.L., C.M. and G.W. performed the experiments and data collection. S.W. conducted data processing and analysis. S.W., K.G., Z.W., E.A.A., P.T. and C.J.

contributed to data interpretation and discussion. S.W. wrote the original draft of the manuscript. All authors have revised and approved the final manuscript.

Conflict of interest

The authors declare that they have no conflicts of interest.

Reference

Ainsworth EA. 2018. Agroclimatology: Linking Agriculture to Climate. 1–23.

Ainsworth EA, Rogers A. 2007. The response of photosynthesis and stomatal conductance to rising [CO2]: Mechanisms and environmental interactions. Plant, Cell and Environment **30**, 258-270.

Ainsworth EA, Serbin SP, Skoneczka JA, Townsend PA. 2014. Using leaf optical properties to detect ozone effects on foliar biochemistry. Photosynthesis Research **119**, 65–76.

Albert LP, Wu J, Prohaska N, *et al.* 2018. Age-dependent leaf physiology and consequences for crown-scale carbon uptake during the dry season in an Amazon evergreen forest. New Phytologist **219**, 870-884.

Bayat B, van der Tol C, Verhoef W. 2018. Integrating satellite optical and thermal infrared observations for improving daily ecosystem functioning estimations during a drought episode. Remote Sensing of Environment **209**, 375-394.

Bernacchi CJ, Singsaas EL, Pimentel C, Portis AR, Long SP. 2001. Improved temperature response functions for models of Rubisco-limited photosynthesis. Plant, Cell and Environment **24**, 253-259.

Caemmerer S von, Furbank RT. 1999. Modeling C4 Photosynthesis. C4 Plant Biology.

Von Caemmerer S. 2013. Steady-state models of photosynthesis. Plant, Cell and Environment **36**, 1617-1630.

Camino C, Gonzalez-Dugo V, Hernandez P, Zarco-Tejada PJ. 2019. Radiative transfer Vcmax estimation from hyperspectral imagery and SIF retrievals to assess photosynthetic performance in rainfed and irrigated plant phenotyping trials. Remote Sensing of Environment **231**, 111186.

Chou S, Chen B, Chen J, Wang M, Wang S, Croft H, Shi Q. 2020. Estimation of leaf photosynthetic capacity from the photochemical reflectance index and leaf pigments. Ecological Indicators, **110**, 10, 58-67.

Crafts-Brandner SJ, Salvucci ME. 2000. Rubisco activase constrains the photosynthetic potential of leaves at high temperature and CO_2. Proceedings of the National Academy of Sciences, **97**, 13430-13435.

Croft H, Chen JM, Luo X, Bartlett P, Chen B, Staebler RM. 2017. Leaf chlorophyll content as a proxy for leaf photosynthetic capacity. Global Change Biology **23**, 3513-3524.

Curran PJ. 1988. The semivariogram in remote sensing: An introduction. Remote Sensing of Environment **24**, 493–507.

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.

642 Remote Sensing of Environment **196**, 279–292.
643 **Evans JR, Clarke VC**. 2019. The nitrogen cost of photosynthesis. Journal of Experimental
644 Botany **70**, 7-15.
645 **Farquhar GD, von Caemmerer S, Berry JA**. 1980. A biochemical model of photosynthetic
646 CO2 assimilation in leaves of C3 species. Planta **149**, 78–90.
647 **Féret JB, Gitelson AA, Noble SD, Jacquemoud S**. 2017. PROSPECT-D: Towards modeling
648 leaf optical properties through a complete lifecycle. Remote Sensing of Environment **193**, 204–
649 215.
650 **Geladi P, Kowalski BR**. 1986. Partial least-squares regression: a tutorial. Analytica chimica acta
651 **185**, 1-17.
652 **Guan K, Wu J, Anderson MC, Kimball J, Frolking S, Li B, Lobell D**. 2017. The shared and
653 unique value of optical, fluorescence, thermal and microwave satellite data for estimating large-
654 scale crop yields. Remote Sensing of Environment **199**, 333–349.
655 **Hammer GL, Wright GC.** 1994. A theoretical analysis of nitrogen and radiation effects on
656 radiation use efficiency in peanut. Australian Journal of Agricultural Research **45**, 575-589.
657 **Houborg R, Cescatti A, Migliavacca M, Kustas WP**. 2013. Satellite retrievals of leaf
658 chlorophyll and photosynthetic capacity for improved modeling of GPP. Agricultural and Forest
659 Meteorology **117**, 10–23.
660 **Houborg R, McCabe MF, Cescatti A, Gitelson AA**. 2015. Leaf chlorophyll constraint on
661 model simulated gross primary productivity in agricultural systems. International Journal of
662 Applied Earth Observation and Geoinformation **43**, 160-176.
663 **Hu S, Mo X, Lin Z**. 2014. Optimizing the photosynthetic parameter Vcmax by assimilating
664 MODIS-fPAR and MODIS-NDVI with a process-based ecosystem model. Agricultural and
665 Forest Meteorology **198**, 320–334.
666 **Jacquemoud S, Baret F**. 1990. PROSPECT: A model of leaf optical properties spectra. Remote
667 Sensing of Environment **34**, 75–91.
668 **Jacquemoud S, Verhoef W, Baret F, Bacour C, Zarco-Tejada PJ, Asner GP, François C,**
669 **Ustin SL**. 2009. PROSPECT + SAIL models: A review of use for vegetation characterization.
670 Remote Sensing of Environment **113**, S56–S66.
671 **Jay S, Bendoula R, Hadoux X, Féret JB, Gorretta N**. 2016. A physically-based model for
672 retrieving foliar biochemistry and leaf orientation using close-range imaging spectroscopy.
673 Remote Sensing of Environment **177**, 220-236.
674 **Kattge J, Knorr W, Raddatz T, Wirth C**. 2009. Quantifying photosynthetic capacity and its
675 relationship to leaf nitrogen content for global-scale terrestrial biosphere models. Global Change
676 Biology **15**, 976-991.
677 **Kauwe D, Martin G, Lin Y, Wright I, Belinda EM, Kristine YC, David SE**, **Maire V**. 2016.
678 A test of the 'one-point method' for estimating maximum carboxylation capacity from field-
679 measured, light-saturated photosynthesis. New Phytologist **210**, 1130-1144.
680 **Koester RP, Nohl BM, Diers BW, Ainsworth EA**. 2016. Has photosynthetic capacity increased
681 with 80years of soybean breeding? An examination of historical soybean cultivars. Plant Cell
682 and Environment **39**, 1058–1067.
683 **Kopsell DA, Kopsell DE, Lefsrud MG, Curran-Celentano J, Dukach LE**. 2004. Variation in
684 lutein, β-carotene, and chlorophyll concentrations among Brassica oleracea cultigens and
685 seasons. HortScience **39**, 361-364.
686 **Kucharik CJ, Mork AC, Meehan TD, Serbin SP, Singh A, Townsend PA, Whitney KS,**
687 **Gratton C**. 2016. Evidence for compensatory photosynthetic and yield response of soybeans to

aphid herbivory. Journal of Economic Entomology **109**, 1177-1187.
Leakey ADB, Ferguson JN, Pignon CP, Wu A, Jin Z, Hammer GL, Lobell DB. 2019. Water Use Efficiency as a Constraint and Target for Improving the Resilience and Productivity of C3 and C4 Crops . Annual Review of Plant Biology **70**, 781-808.
Leuning R. 2002. Temperature dependence of two parameters in a photosynthesis model. Plant, Cell and Environment **25**, 1205-1210.
Li D, Tian L, Wan Z, Jia M, Yao X, Tian Y, Zhu Y, Cao W, Cheng T. 2019. Assessment of unified models for estimating leaf chlorophyll content across directional-hemispherical reflectance and bidirectional reflectance spectra. Remote Sensing of Environment **231**, 111240.
Lichtenthaler HK, Wellburn AR. 1983. Determinations of total carotenoids and chlorophylls a and b of leaf extracts in different solvents. Biochemical Society Transactions **11**, 591-592.
Long SP, Bernacchi CJ. 2003. Gas exchange measurements, what can they tell us about the underlying limitations to photosynthesis? Procedures and sources of error. Journal of experimental botany **54**, 2393-2401.
Luo X, Croft H, Chen JM, He L, Keenan TF. 2019. Improved estimates of global terrestrial photosynthesis using information on leaf chlorophyll content. Global change biology **25**, 2499-2514.
Makino A, Osmond B. 1991. Solubilization of ribulose-1,5-bisphosphate carboxylase from the membrane fraction of pea leaves. Photosynthesis Research **29**, 79-85.
Meacham-Hensold K, Montes CM, Wu J, *et al.* 2019. High-throughput field phenotyping using hyperspectral reflectance and partial least squares regression (PLSR) reveals genetic modifications to photosynthetic capacity. Remote Sensing of Environment **231**, 111176.
Meerdink SK, Roberts DA, King JY, Roth KL, Dennison PE, Amaral CH, Hook SJ. 2016. Linking seasonal foliar traits to VSWIR-TIR spectroscopy across California ecosystems. Remote Sensing of Environment **186**, 322-338.
Miner GL, Bauerle WL. 2019. Seasonal responses of photosynthetic parameters in maize and sunflower and their relationship with leaf functional traits. Plant, cell & environment, **42**, 1561-1574
Mu X, Chen Q, Chen F, Yuan L, Mi G. 2016. Within-leaf nitrogen allocation in adaptation to low nitrogen supply in maize during grain-filling stage. Frontiers in plant science **7**, 699.
Neumann K, Verburg PH, Stehfest E, Müller C. 2010. The yield gap of global grain production: A spatial analysis. Agricultural Systems, **103**, 316-326.
Onoda Y, Wright IJ, Evans JR, Hikosaka K, Kitajima K, Niinemets Ü, Poorter H, Tosens T, Westoby M. 2017. Physiological and structural tradeoffs underlying the leaf economics spectrum. New Phytologist **214**, 1447-1463.
Quebbeman JA, Ramirez JA. 2016. Optimal allocation of leaf-level nitrogen: Implications for covariation of Vcmax and Jmax and photosynthetic downregulation. Journal of Geophysical Research: Biogeosciences, **12**, 2464-2475.
Saltelli A, Tarantola S, Campolongo F, Ratto M. 2004. *Sensitivity Analysis in practice: a guide to assessing scientific models* (Vol. 1). New York: Wiley.
Schaepman-Strub G, Schaepman ME, Painter TH, Dangel S, Martonchik J V. 2006. Reflectance quantities in optical remote sensing-definitions and case studies. Remote Sensing of Environment **103**, 27-42.
Serbin SP, Dillaway DN, Kruger EL, Townsend PA. 2012. Leaf optical properties reflect variation in photosynthetic metabolism and its sensitivity to temperature. Journal of Experimental Botany **63**, 489-502.

Serbin SP, Singh A, Desai AR, Dubois SG, Jablonski AD, Kingdon CC, Kruger EL, Townsend PA. 2015. Remotely estimating photosynthetic capacity, and its response to temperature, in vegetation canopies using imaging spectroscopy. Remote Sensing of Environment **167**, 78–87.

Serbin SP, Singh A, McNeil BE, Kingdon CC, Townsend PA. 2014. Spectroscopic determination of leaf morphological and biochemical traits for northern temperate and boreal tree species. Ecological Applications **24**, 1651-1669.

Smith NG, Keenan TF, Colin Prentice I, *et al.* 2019. Global photosynthetic capacity is optimized to the environment. Ecology Letters **22**, 506-517.

Singh A, Serbin SP, McNeil BE, Kingdon CC, Townsend PA. 2015. Imaging spectroscopy algorithms for mapping canopy foliar chemical and morphological traits and their uncertainties. Ecological Applications **25**, 2180–2197.

Sobol IM. 2001. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Mathematics and Computers in Simulation.

Taylor KE. 2001. Summarizing multiple aspects of model performance in a single diagram. Journal of Geophysical Research Atmospheres **106**, 7183-7192

Townsend PA, Foster JR, Chastain RA, Currie WS. 2003. Application of imaging spectroscopy to mapping canopy nitrogen in the forest of the central Appalachian mountains using Hyperion and AVIRIS. IEEE Transactions on Geoscience and Remote Sensing **41**, 1347-1354.

Verrelst J, Malenovský Z, Van der Tol C, Camps-Valls G, Gastellu-Etchegorry JP, Lewis P, North P, Moreno J. 2019. Quantifying Vegetation Biophysical Variables from Imaging Spectroscopy Data: A Review on Retrieval Methods. Surveys in Geophysics **40**, 589-629.

Vilfan N, van der Tol C, Verhoef W. 2019. Estimating photosynthetic capacity from leaf reflectance and Chl fluorescence by coupling radiative transfer to a model for photosynthesis. New Phytologist **223**, 487-500.

Walker AP, Beckerman AP, Gu L, Kattge J, Cernusak LA, Domingues TF, Scales JC, Wohlfahrt G, Wullschleger SD, Woodward FI. 2014. The relationship of leaf photosynthetic traits - Vcmax and Jmax - to leaf nitrogen, leaf phosphorus, and specific leaf area: A meta-analysis and modeling study. Ecology and Evolution **4**, 3218-3235.

Wang S, Ibrom A, Bauer-Gottwein P, Garcia M. 2018*a*. Incorporating diffuse radiation into a light use efficiency and evapotranspiration model: An 11-year study in a high latitude deciduous forest. Agricultural and Forest Meteorology **248**, 479–493.

Wang Z, Chlus A, Geygan R, Ye Z, Zheng T, Singh A, Couture J, Cavender-Bares J, Kruger EL, Townsend PA. 2020. Foliar functional traits from imaging spectroscopy across biomes in eastern North America. New Phytologist.

Wang Z, Skidmore AK, Darvishzadeh R, Wang T. 2018*b*. Mapping forest canopy nitrogen content by inversion of coupled leaf-canopy radiative transfer models from airborne hyperspectral imagery. Agricultural and Forest Meteorology **253**, 247-260.

Wang Z, Skidmore AK, Wang T, Darvishzadeh R, Hearne J. 2015. Applicability of the PROSPECT model for estimating protein and cellulose+ lignin in fresh leaves. Remote sensing of environment **168**, 205–218.

Wang Z, Townsend PA, Schweiger AK, Couture JJ, Singh A, Hobbie SE, Cavender-Bares J. 2019. Mapping foliar functional traits and their uncertainties across three years in a grassland experiment. Remote Sensing of Environment **221**, 405-416.

Weber VS, Araus JL, Cairns JE, Sanchez C, Melchinger AE, Orsini E. 2012. Prediction of

grain yield using reflectance spectra of canopy and leaves in maize plants grown under different water regimes. Field Crops Research **128**, 82-90.
Verhoef W. 1984. Light scattering by leaf layers with application to canopy reflectance modeling: The SAIL model. Remote sensing of environment, **16**, 125-141.
Wold S, Sjöström M, Eriksson L. 2001. PLS-regression: A basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems.
Wu J, Rogers A, Albert LP, Ely K, Prohaska N, Wolfe BT, Oliveira RC, Saleska SR, Serbin SP. 2019. Leaf reflectance spectroscopy captures variation in carboxylation capacity across species, canopy environment and leaf age in lowland moist tropical forests. New Phytologist **224**, 663-674.
Yendrek CR, Tomaz T, Montes CM, Cao Y, Morse AM, Brown PJ, McIntyre LM, Leakey ADB, Ainsworth EA. 2017. High-throughput phenotyping of maize leaf physiological and biochemical traits using hyperspectral reflectance. Plant Physiology **173**, 614-626.
Yeoh HH, Wee YC. 1994. Leaf protein contents and nitrogen-to-protein conversion factors for 90 plant species. Food Chemistry **49**, 245-250.
Zheng T, Chen JM. 2017. Photochemical reflectance ratio for tracking light use efficiency for sunlit leaves in two forest types. ISPRS Journal of Photogrammetry and Remote Sensing **123**, 47–61.

Table

Table 1 Model parameters and their typical ranges for PROSPECT-D, PROSPECT-DyN and COSINE leaf radiative transfer models.

Model	Parameter	Description and unit	Typical range
PROSPECT-D	N	Leaf structure parameter [unitless]	0 – 5
	Chl	Chlorophyll content [μg cm^{-2}]	0 – 100
	Car	Carotenoids content [μg cm^{-2}]	0 – 60
	Ant	Anthocyanin content [μg cm^{-2}]	0 – 5
	Cs	Senescent (brown) materials [unitless]	0 – 5
	Cw	Leaf water thickness [cm]	0 – 0.1
	Cdm	Dry matter content [g cm^{-2}]	0 – 0.02
PROSPECT-DyN	Cp	Protein content [g cm^{-2}]	0 – 0.02
	Ccl	Cellulose and lignin content [g cm^{-2}]	0 – 0.02
COSINE	θs	Sensor view angle [°]	0 – 180
	θi	Light incident angle [°]	0 – 90
	B	Specular term to account for the bidirectional reflectance factor [unitless]	-0.2 – 0.6

Figure legends

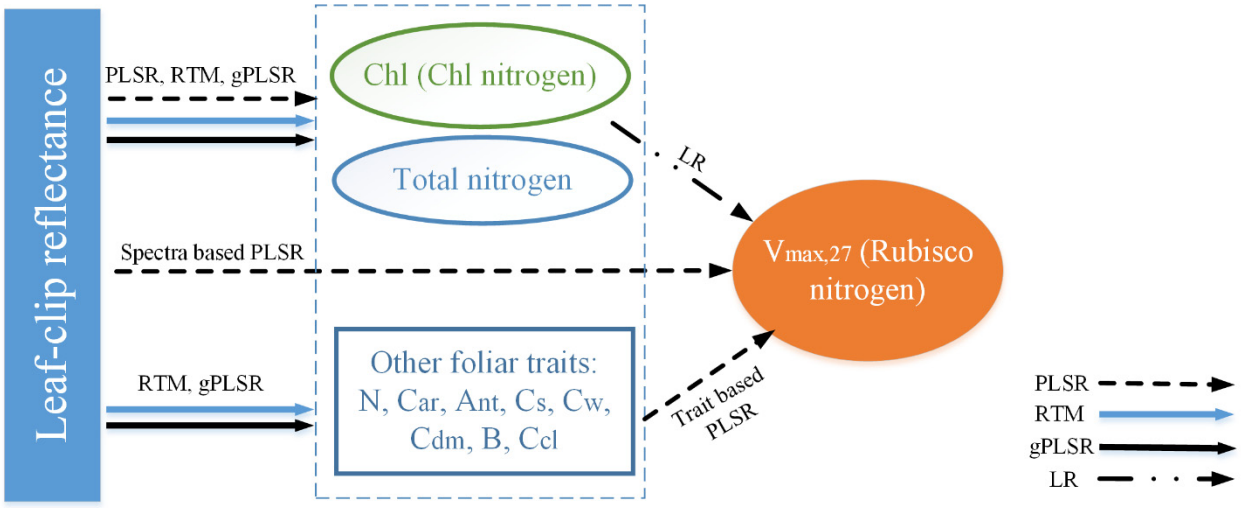


Fig. 1. Comparison of approaches to estimate leaf chlorophyll, nitrogen, and $V_{max,27}$. PLSR: partial-least-squares regression; gPLSR: generalized PLSR; RTM: radiative transfer model; LR: linear regression. Chl: leaf chlorophyll content; $V_{max,27}$: leaf maximum carboxylation rate standardized to 27 °C; N: leaf thickness parameter; Car: carotenoids; Ant: anthocyanins; Cs: senescent material fraction; Cw: leaf water content; Cdm: leaf dry matter content; B: the parameter to account for the leaf bidirectional reflectance; Ccl: leaf cellulose and lignin content; The dashed line indicates that methods require model training, while the solid lines are calibration-free approaches. This study compared three approaches to retrieve leaf chlorophyll and total nitrogen content, and four approaches to retrieve leaf $V_{max,27}$. This figure is available in colour at *JXB* online.

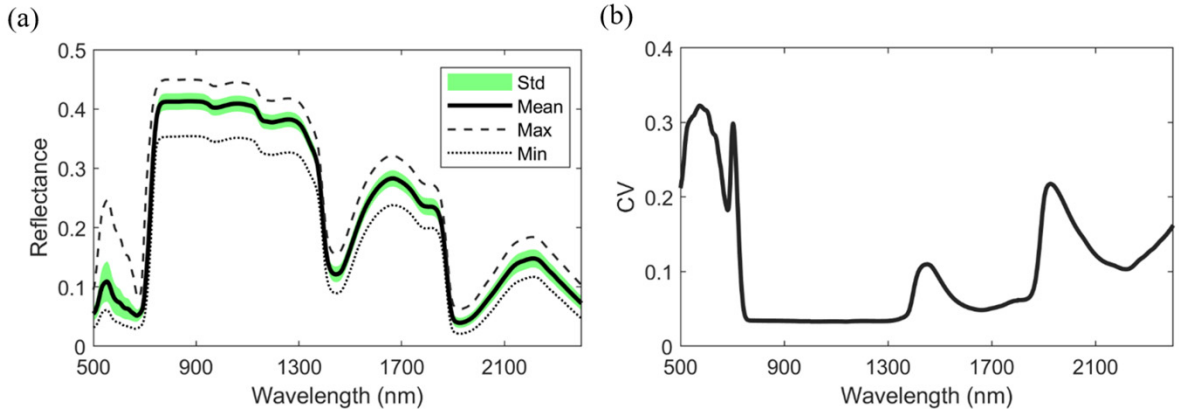


Fig. 2. Mean, maximum, minimum, standard deviation, and coefficient of variation (CV) of the measured leaf reflectance for maize. This figure is available in colour at *JXB* online.

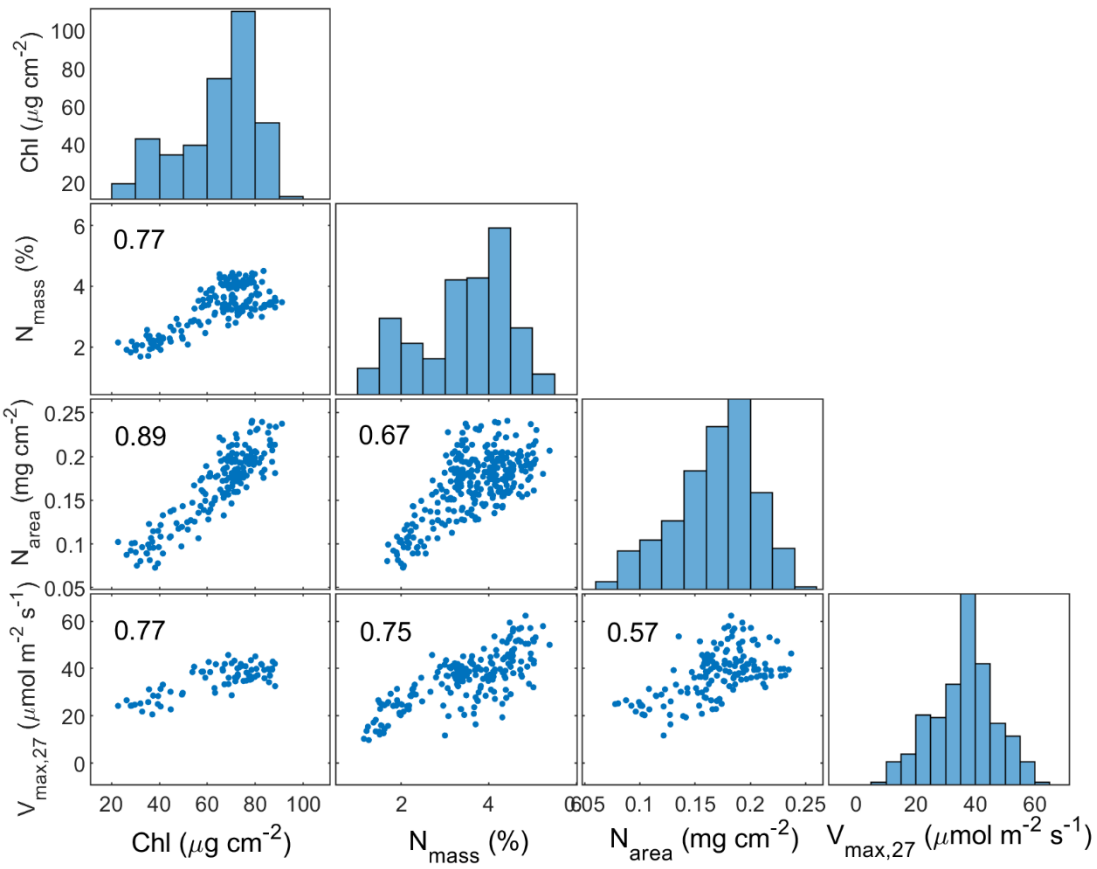


Fig. 3. Correlation matrix for measured leaf photosynthetic traits. $V_{max,27}$: the carboxylation rate at 27 °C (μmol $m^{-2}s^{-1}$); Chl: leaf chlorophyll content (μg/cm^2); N_{mass}: leaf nitrogen per mass (%); N_{area}: leaf nitrogen per area (mg/cm^2). The statistics in plots refer to the Pearson correlation coefficients.

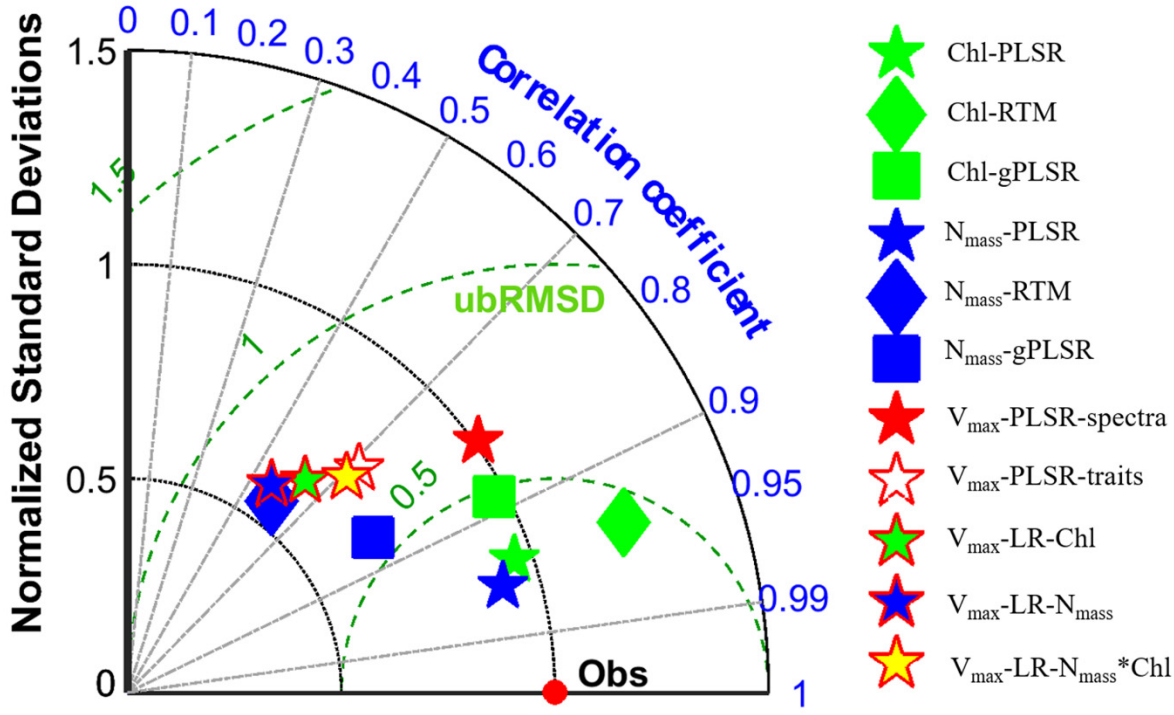


Fig. 4. Taylor diagram to present the performance of estimating leaf chlorophyll, nitrogen and V_{max}. The pentagrams represent PLSR and LR (linear regression) methods. The diamonds are the RTM approaches, which refer to the PROSPECT-COSINE and PROSPECT-DyN-COSINE. The squares indicate the gPLSR method. The markers with the green color represent chlorophyll related predictions. The markers with the blue color are nitrogen related predictions. The markers with the red edge indicate V_{max} related predictions. The radial coordinate represents the normalized standard deviation, which is equal to 1 for the observations. The angular coordinate indicates the correlation coefficient, which refers to 1 for the observations. The concentric green dashed semi-circles represent the normalized unbiased RMSD. In the Taylor diagram, the closer points to the observation point refer to higher predictive ability for the models. This figure is available in colour at *JXB* online.

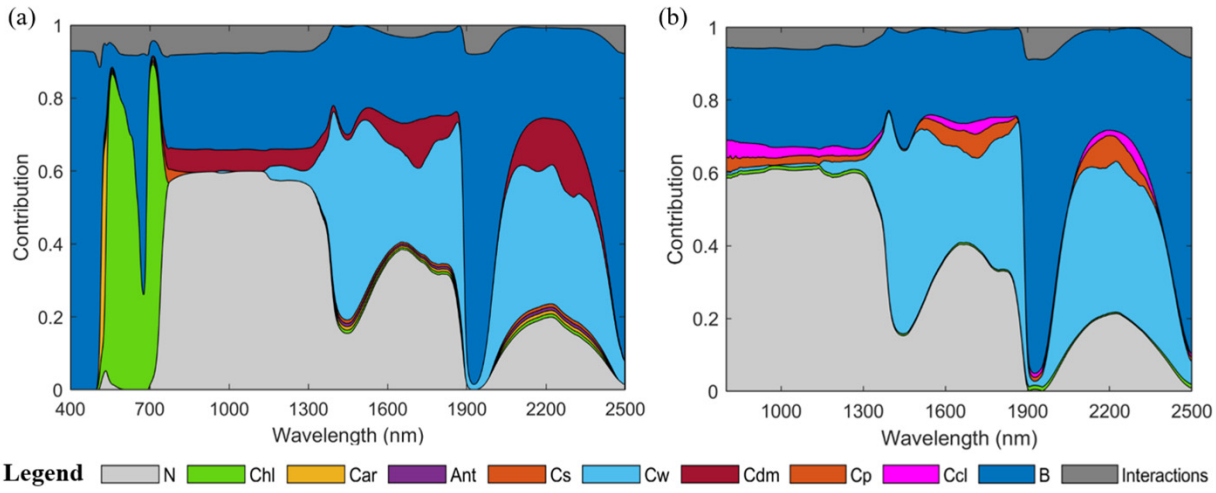


Fig. 5. Global sensitivity analysis of radiative transfer models. (a) PROSPECT-COSINE (b) PROSPECT-DyN-COSINE. In the legend, the variables N, Chl, Car, Ant, Cs, Cw, Cdm, Cp, Ccl, B and Interactions refer to leaf thickness structure parameter, chlorophyll, carotenoids, anthocyanin, senescent materials, water content, dry matter content, protein, cellulose and lignin, leaf bidirectional reflectance factors, and interactions for the parameter sensitivities, respectively. This figure is available in colour at *JXB* online.

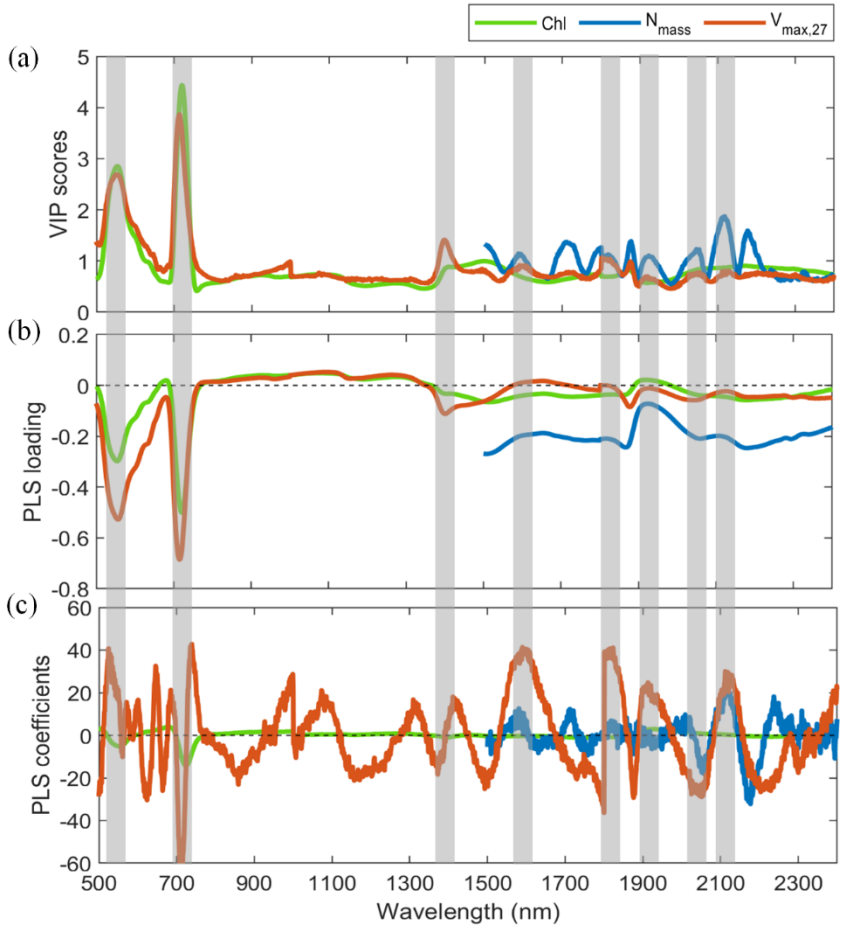


Fig. 6. (a) VIP scores, (b) loading and (c) coefficients of the spectra based PLSR model for $V_{max,27}$, chlorophyll, and nitrogen predictions. The orange curve shows leaf $V_{max,27}$ predictions. The green curve refers to the leaf chlorophyll content prediction. The blue curve represents leaf nitrogen per mass predictions. The shaded grey region indicates the key wavelengths for $V_{max,27}$ predictions. This figure is available in colour at *JXB* online.

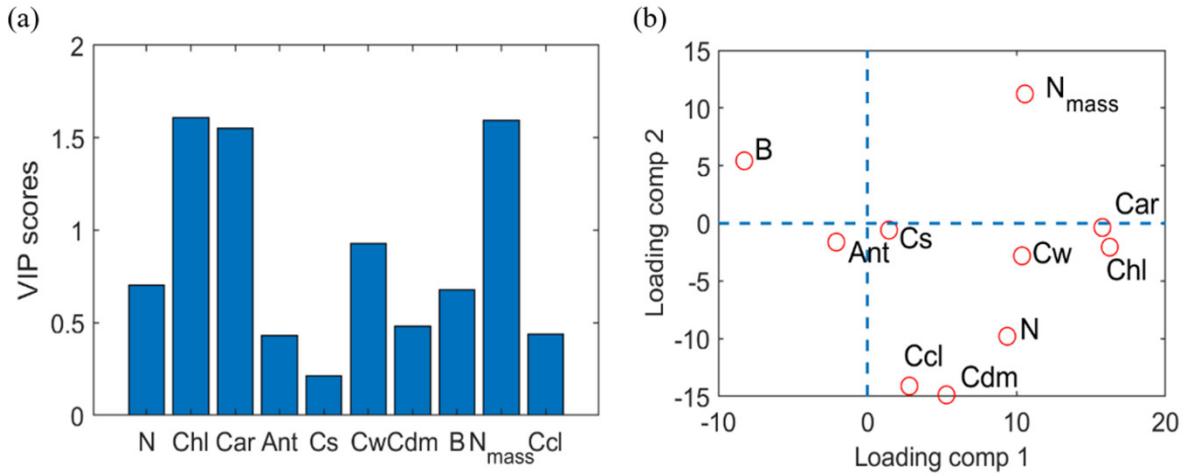


Fig. 7. (a) VIP scores and (b) loading components 1 and 2 of the trait-based PLSR model for the $V_{max,27}$ prediction. The traits N, Chl, Car, Ant, Cs, Cw, Cdm, B, N_{mass} and Ccl refer to leaf thickness structure parameter, chlorophyll, carotenoids, anthocyanin, senescent materials, water content, dry matter content, leaf bidirectional reflectance factor, nitrogen per mass, and cellulose and lignin content, respectively. These foliar traits are from calibration-free approaches. The estimated N_{mass} is from gPLSR, due to its high accuracy. Other traits are from RTMs.

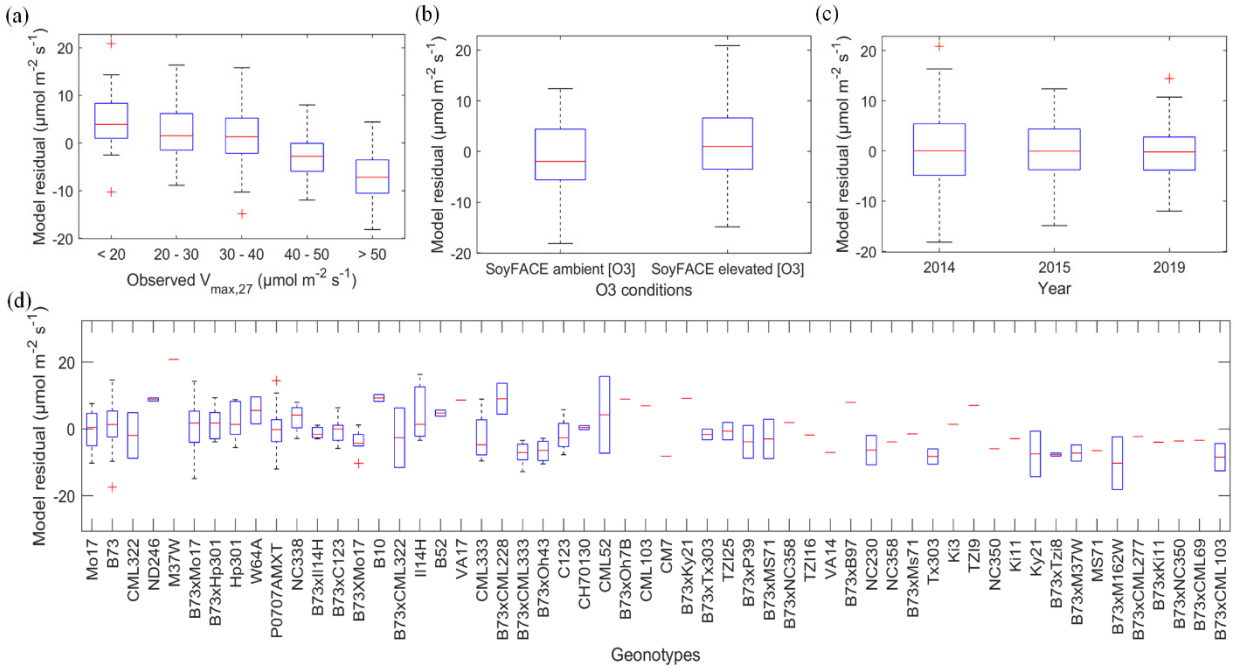
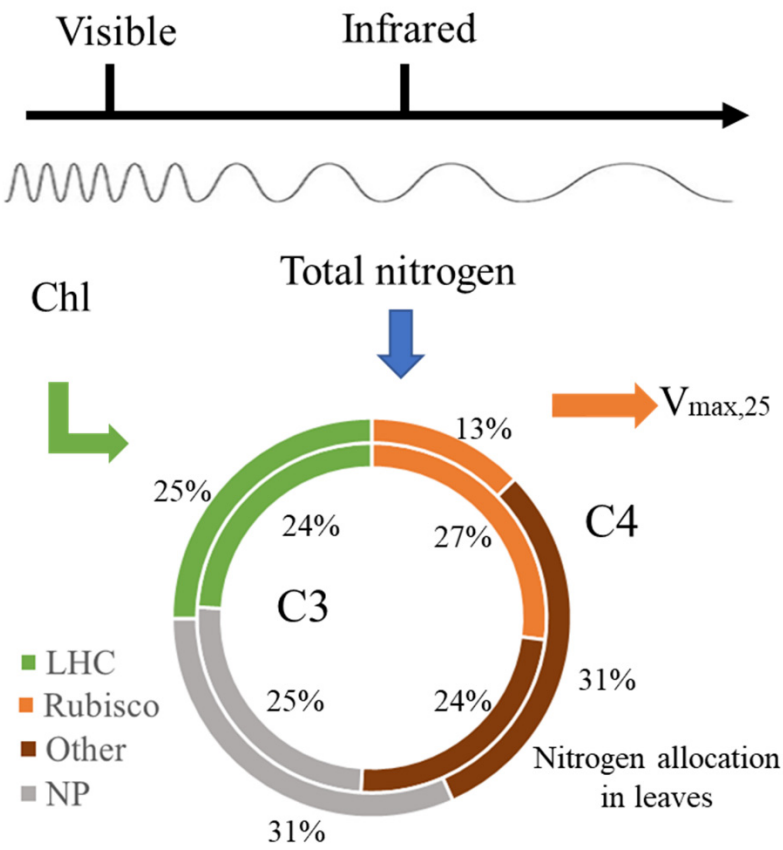


Fig. 8. Analysis of the performance of spectra-$V_{max,27}$ model by (a) leaf condition, (b) environmental stressor, (c) experiment year, and (d) genotype.



867

868 **Fig. 9.** Methodology to integrate the visible and infrared hyperspectral reflectance to quantify
869 nitrogen allocation to estimate V_{max}. The inner and outer circles refer to the typical nitrogen
870 allocation in C_3 leaves and C_4 leaves, respectively. The data of nitrogen allocation for C_3 and C_4
871 leaves are from Evans and Clarke (2019) and Mu et al. (2016), respectively. Notably, the allocation
872 rates vary with environmental conditions, species and growth stages. LHC refers to nitrogen in the
873 light-harvesting complex. Rubisco represents nitrogen in the Rubisco protein. Other stands for
874 nitrogen in other photosynthetic proteins. NP means non-photosynthetic proteins, e.g. cell wall,
875 mitochondria, and cytosol. This figure is available in colour at *JXB* online.