PLOS ONE

RESEARCH ARTICLE

Using hyperspectral leaf reflectance to estimate photosynthetic capacity and nitrogen content across eastern cottonwood and hybrid poplar taxa

Thu Ya Kyaw[1]*, Courtney M. Siegert[1], Padmanava Dash[2], Krishna P. Poudel[1], Justin J. Pitts[1¤], Heidi J. Renninger[1]

1 Department of Forestry, Forest and Wildlife Research Center, Mississippi State University, Starkville, Mississippi, United States of America, **2** Department of Geosciences, Mississippi State University, Starkville, Mississippi, United States of America

¤ Current address: Agronomy Department, University of Florida, Gainesville, Florida, United States of America
* tk758@msstate.edu

OPEN ACCESS

Citation: Kyaw TY, Siegert CM, Dash P, Poudel KP, Pitts JJ, Renninger HJ (2022) Using hyperspectral leaf reflectance to estimate photosynthetic capacity and nitrogen content across eastern cottonwood and hybrid poplar taxa. PLoS ONE 17(3): e0264780. https://doi.org/10.1371/journal.pone.0264780

Editor: Janusz J. Zwiazek, University of Alberta, CANADA

Received: November 4, 2021

Accepted: February 16, 2022

Published: March 10, 2022

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pone.0264780

Data Availability Statement: Our data can be accessed from Scholars Junction: Mississippi State University's Institutional Repository at the

Abstract

Eastern cottonwood (*Populus deltoides* W. Bartram ex Marshall) and hybrid poplars are well-known bioenergy crops. With advances in tree breeding, it is increasingly necessary to find economical ways to identify high-performing *Populus* genotypes that can be planted under different environmental conditions. Photosynthesis and leaf nitrogen content are critical parameters for plant growth, however, measuring them is an expensive and time-consuming process. Instead, these parameters can be quickly estimated from hyperspectral leaf reflectance if robust statistical models can be developed. To this end, we measured photosynthetic capacity parameters (Rubisco-limited carboxylation rate (V_{cmax}), electron transport-limited carboxylation rate (J_{max}), and triose phosphate utilization-limited carboxylation rate (TPU)), nitrogen per unit leaf area (N_{area}), and leaf reflectance of seven taxa and 62 genotypes of *Populus* from two study plantations in Mississippi. For statistical modeling, we used least absolute shrinkage and selection operator (LASSO) and principal component analysis (PCA). Our results showed that the predictive ability of LASSO and PCA models was comparable, except for N_{area} in which LASSO was superior. In terms of model interpretability, LASSO outperformed PCA because the LASSO models needed 2 to 4 spectral reflectance wavelengths to estimate parameters. The LASSO models used reflectance values at 758 and 935 nm for estimating V_{cmax} (R^2 = 0.51 and RMSPE = 31%) and J_{max} (R^2 = 0.54 and RMSPE = 32%); 687, 746, and 757 nm for estimating TPU (R^2 = 0.56 and RMSPE = 31%); and 304, 712, 921, and 1021 nm for estimating N_{area} (R^2 = 0.29 and RMSPE = 21%). The PCA model also identified 935 nm as a significant wavelength for estimating V_{cmax} and J_{max}. Therefore, our results suggest that hyperspectral leaf reflectance modeling can be used as a cost-effective means for field phenotyping and rapid screening of *Populus* genotypes because of its capacity to estimate these physicochemical parameters.

following DOI: https://doi.org/10.54718/BACR5952.

Funding: This work was supported by the National Institute of Food and Agriculture, U.S. Department of Agriculture under award numbers 2018-67020-27934 to HJR and CMS, and 2018-68005-27636 to HJR and CMS, as well as U.S. Department of Agriculture McIntire Stennis Program under accession numbers: MISZ-067050 to HJR, MISZ-032100 to CMS, and MISZ-0621210 to KPP, respectively. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. There was no additional external funding received for this study.

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

Introduction

Production of bioenergy through direct combustion [1] and/or the manufacture of liquid biofuels [2] can reduce the consumption of fossil fuels and associated emissions [3]. In the USA, bioenergy contributed about 7.3% to the energy sector in 2019, which could increase to an additional 18–55% by 2050 [4]. To bolster this increasing trend, short rotation woody crops (SRWCs) can be planted as a biomass feedstock for bioenergy production [5, 6]. Currently in the USA, *Populus* feedstock production for bioenergy makes up a small portion of land use, but there is the potential for large-scale plantations for bioenergy in the future because *Populus* is a promising bioenergy crop. Eastern cottonwood (*Populus deltoides* W. Bartram ex Marshall) and hybrid poplars have preferable traits for producing high biomass yield within a short period of time [7–12]. The biomass yield of *Populus* genotypes can reach 7.5 to 15.2 Mg/ha/year [13]. Due to their rapid growth and high biomass productivity, the United States Department of Energy classifies *Populus* as a potential SRWC for bioenergy production [14]. A *Populus* plantation can be managed with a five-year harvesting cycle [10, 15–17], and their excellent coppicing ability [18–20] eliminates the need for replanting after each harvest. Since *Populus* has a wide range of genetic diversity [21, 22], clonal propagation [23, 24] and the use of genetically improved plant material [3, 25] can maximize biomass yield and will accelerate the development of the bioenergy sector [3, 25]. Moreover, *Populus* feedstocks have desirable attributes needed for producing ethanol and other biofuels, such as high cellulose content, moderate lignin and hemicellulose content, and low ash and extractives [26].

Tree breeding has been used to produce many *Populus* genotypes with the aim of maximizing biomass yield [27, 28]. Tree breeding is usually followed by field trials to compare growth performance across developed genotypes. Since field experiments are laborious, costly, and time-consuming [29], identifying superior genotypes for further testing is needed as an initial screening step to optimize field trials. To this end, photosynthetic capacity can be used as a selection criterion for clonal screening because individuals with enhanced photosynthetic rate under field conditions are likely to result in higher biomass yield [30]. To evaluate photosynthetic capacity, photosynthetic CO_2 response curves can be used to calculate key photosynthetic capacity parameters, such as (1) maximum carboxylation rate limited by ribulose 1-5-bisphosphate carboxylase/oxygenase (RuBisCO) activity (V_{cmax}), (2) electron transport-limited carboxylation rate (J_{max}), and (3) triose phosphate utilization-limited carboxylation rate (TPU) [31, 32]. In addition to photosynthetic capacity parameters, nutrient content in leaves is another important parameter for plant growth because limitations of nutrients can decrease photosynthesis. In plants, nitrogen (N) is a limiting macronutrient [33–35]. N deficiency can reduce leaf area [36], leaf durability, and leaf chlorophyll content. Additionally, leaf N content tends to be linearly correlated with photosynthetic capacity [37, 38], and consequently, variations in leaf N content can affect biomass production [39–41].

Photosynthetic capacity parameters and leaf N content provide valuable information about leaf-level processes. However, measuring photosynthetic CO_2 response curves in the field takes about 15–20 minutes per leaf. Analogously, measuring leaf N content requires a laboratory assay to generate results. Therefore, these measurement techniques are not cost-effective for high-throughput screening and large-scale field phenotyping. On the other hand, hyperspectral leaf reflectance can potentially be used as an alternative to quickly estimate photosynthetic capacity parameters and leaf N content. Generally, leaf spectra have low reflectance in the visible region mainly due to chlorophyll absorption and high reflectance in the near-infrared region due to internal leaf scattering with little to no absorption [42]. Therefore, there is a physiological relationship between leaf reflectance and photosynthetic capacity parameters [43–45]. For example, according to Qian et al. [44], when V_{cmax} increases, the reflectance at

the red edge position and the green peak decreases. Leaf reflectance can also be correlated with leaf N content because N-containing bonds cause variation in spectral features [46]. Ye et al. [47] found that when leaf N content increased, spectral reflectance decreased, and this relationship was more significantly observed in the green to near infrared regions. Due to these relationships, if robust statistical models can be developed, hyperspectral leaf reflectance can be used as a proxy for rapid assessment of photosynthetic capacity [48–50], as well as non-destructive and real-time monitoring of leaf N content [51, 52]. On the other hand, creating empirical models for hyperspectral data requires sophisticated model development techniques, and model robustness depends on the quality of sample data measured in the field [53].

While existing vegetation indices (e.g., normalized difference vegetation index, photochemical reflectance index, etc.) might be sufficient for estimating physicochemical parameters [43, 54–56], many studies have also developed more robust models using multiple regression methods to improve the prediction capacity for certain plant species [50, 57, 58]. Hyperspectral data have very narrow bandwidths with tremendous amounts of redundant information leading to potential issues with multicollinearity and high-dimensionality. In this respect, statistical methods, such as stepwise regression, shrinkage (e.g., ridge regression and least absolute shrinkage and selection operator (LASSO)), and dimension reduction (e.g., principal component analysis (PCA)) may be useful for handling hyperspectral data. Hyperspectral data also contain hundreds of spectral bands, and therefore hundreds of predictor variables. If the number of predictor variables is greater than the sample size, stepwise selection of multiple linear regression likely chooses nuisance variables rather than true variables, and consequently, performs poorly for out-of-sample accuracy [59, 60]. Ridge regression includes all predictor variables in the final model, which is disadvantageous when analyzing data containing a large number of predictor variables [59]. To solve these issues, many studies suggested LASSO [61–63] and PCA [64–66] methods. The PCA method is easy and computationally efficient. Its major drawback is a difficulty in interpreting principal components because they represent linear combinations of the original features, and thus, each component uses the collection of all predictor variables [67]. LASSO can overcome this limitation of PCA. Compared to PCA, LASSO is likely to generate a simpler model because LASSO uses regularization, shrinkage, and variable selection, and the final model includes a smaller set of predictor variables [59]. However, due to being a supervised feature extraction method [68–70], LASSO is more computationally demanding than PCA. Both LASSO and PCA methods have the capacity for analyzing hyperspectral data, but each method has its own strengths and weaknesses [71–74]. Therefore, comparing both methods on the same data can help to elucidate these tradeoffs, determine where the methods provide convergent information, and identify the method with better model performance.

In this study, we established two *Populus* plantations for field trials in upland sites of Mississippi each containing over 100 genotypes of *Populus* and its hybrids. We postulated that utilizing the large variability in leaf physicochemical parameters across different *Populus* genotypes and sites would result in robust models for *Populus* screening programs. Therefore, using leaf-level data from 62 *Populus* genotypes across seven taxa measured at two study plantations, we developed hyperspectral leaf reflectance models to estimate photosynthetic capacity parameters and leaf N content. Our main objectives were: (1) to assess whether hyperspectral leaf reflectance can adequately model photosynthetic capacity parameters and leaf N content of eastern cottonwood and hybrid poplar, (2) to compare the performance of models developed using LASSO and PCA methods, and (3) to identify spectral wavelengths that are the most sensitive to these physicochemical parameters. Successful development of predictive models from hyperspectral data can be applied to clonal screenings, enable high-throughput field phenotyping [50, 75], and thus, reduce the costs of *Populus* bioenergy research. Moreover,

photosynthetic capacity parameters are input parameters in many terrestrial biosphere models because plant productivity depends on photosynthetic processes [76–79]. Therefore, information about these photosynthetic capacity parameters, which may be cost-effectively estimated from hyperspectral reflectance data, can be used as inputs for process-based earth system models to simulate terrestrial carbon fluxes [80–82] for improving our understanding of regional and global carbon cycles.

Materials and methods

Field and laboratory analysis

Study area. The study area consisted of two plantations containing over 100 *Populus* genotypes located in Monroe County (88˚ 17' W, 33˚ 51' N) and Pontotoc County (88˚ 59' W, 34˚ 8' N), MS. The plantation at the Monroe site was established in April 2018 at 2.7 × 1.8 m spacing. The average height of trees was 6.2 m at the end of the second growing season in 2019. The Monroe site has moderately well-drained, Prentiss fine sandy loam soils [83] and is surrounded by, and previously used for, row crop agriculture with 0–2 percent slopes. This site has soil carbon, soil nitrogen, organic matter, and dry bulk density of 2.29 kg/m^2, 0.21 kg/m^2, 1.08%, and 1.60 g/cm^3, respectively. Based on climate data from a nearby weather station (USW00093862), the Monroe site has an annual temperature range of 11–23˚C, and mean annual rainfall of 1,397 mm [84].

The plantation at the Pontotoc site was established in April, 2019 at 2.7 × 1.8 m spacing. The average height of trees was 5.3 m at the end of the second growing season in 2020. The Pontotoc site has well-drained, Atwood silt loam soils [83] and is also surrounded by, and previously used for, row crop agriculture with 0–2 percent slopes. This site has soil carbon, soil nitrogen, organic matter, and dry bulk density of 3.61 kg/m^2, 0.44 kg/m^2, 1.83%, and 1.52 g/cm^3, respectively. Based on climate data from the Pontotoc Experimental Station, MS (USC00227111), the Pontotoc site has an annual temperature range of 10–22˚C, and mean annual rainfall of 1,482 mm [84].

Photosynthetic A/C$_i$ curves. Gas exchange measurements were conducted in the field using a LI-COR 6400 XTP portable photosynthesis system (LI-COR Biosciences Inc., Lincoln, NE, USA) with a red/blue light source attached. At the Pontotoc site, leaves were within reach, and measurements were made on live leaves still attached to branches on the mid canopy sunlit side of trees. At the Monroe site, leaves could not be accessed due to the height of trees, and a 7.8 m long pole pruner was used to cut branches from the mid to upper parts of crowns on the sunlit side of trees. The cut branches were immediately submerged in water, and approximately 20 cm of the branch end was recut underwater to minimize embolisms [85, 86]. Then, sample leaves attached to cut branches were measured in a sunlit location outside of the field plot. Fully expanded leaves with minimal disease located near the top of branches were selected for these measurements. Photosynthetic A/C$_i$ response curves (A = net CO_2 assimilation and C$_i$ = sub-stomatal CO_2 concentration inside the leaf) of the sample leaves were measured by setting ambient CO_2 levels at 400 ppm and then, 300, 200, 100, and 50 ppm before being increased to 400, 600, and 800 ppm. During measurements, stomatal conductance was monitored. If low stomatal conductance occurred as a result of stress due to removal from the tree (in the case of the Monroe site), a new branch was collected and measured. Measurements were made between 10:00 AM and 2:00 PM on clear sunny days. For each individual during each measurement period, photosynthetic photo flux density, leaf temperature, and relative humidity inside the leaf chamber were controlled under saturating light conditions at 1500 μmol/m^2/s, 30˚C, and 40 to 60%, respectively [85–87]. From the A/C$_i$ curves, we calculated three main photosynthetic capacity parameters–V_{cmax}, J_{max}, and TPU in a Microsoft

Excel solver program that contained non-linear curve-fitting equations from Sharkey et al. [88], which was based on Farquhar et al. [31]. During calculation, the parameters were scaled to a standardized temperature of 25˚C.

We measured a total of 105 leaves (n = 105). Among them, we measured 57 leaves at the Monroe site in July and early September in 2019 and 48 leaves at the Pontotoc site in July in 2020. In total, about 85% of leaf samples were measured in July and 15% in early September. September measurements were made on previously measured genotypes from July because some spectral response curves from leaves measured in July were more variable than the majority of collected data, and thus, these genotypes were re-measured. At the Monroe site, we measured leaf samples from 21 genotypes and five taxa, including *P. deltoides* × *P. deltoides* (D×D, or eastern cottonwood), *P. deltoides* × *P. maximowiczii* A. Henry (D×M), *P. deltoides* × *P. nigra* L. (D×N), *P. deltoides* × *P. trichocarpa* Torr. & Gray (D×T), and *P. trichocarpa* × *P. deltoides* (T×D) (S1 Table). At the Pontotoc site, we measured leaf samples from 48 genotypes and six taxa, including D×D, D×M, D×N, the three-way hybrid *Populus deltoides* × *Populus nigra* × *Populus maximowiczii* ((D×N)×M), D×T, and T×M (*P. trichocarpa* × *P. maximowiczii*) (S1 Table).

Hyperspectral leaf reflectance. Immediately after gas exchange measurements of a leaf, hyperspectral leaf reflectance was measured on the same leaf so that both measurements were made in close proximity temporally and under the same environmental and light conditions. The spectral measurements were made with a portable, single-beam GER 1500 field spectroradiometer (Spectra Vista Corp., Poughkeepsie, NY, USA). The GER 1500 can detect a spectral range of 286.90 to 1089.62 nm and contains 512 spectral bands. It has full-width at half-maximum of 3.20 nm and a nominal bandwidth of approximately 1.50 nm covering parts of the ultraviolet, visible, and near-infrared wavelength regions of the electromagnetic spectrum. Decimal values of spectral wavelengths were rounded to integer values when reporting results.

To calculate the remote sensing reflectance of leaves, we collected two types of spectral data—radiance of the leaf and irradiance of the sky under sunny, cloud-free conditions. From approximately 40 cm above the leaf, radiances were measured on the leaf adaxial surface. Excluding the midrib, four radiance measurements were taken from various locations on each leaf lamina, and these readings were then averaged to obtain mean radiance for individual leaves. Irradiance of the sky was measured by holding the spectroradiometer and its attached fiber-optic cable with diffuser vertically toward the cloud-free sky. Six irradiance readings were recorded between individual radiance measurements, and these readings were then averaged to obtain mean irradiance of the sky. To calculate the reflectance of leaves, mean sky irradiance measured in close temporal proximity to individual radiance readings was chosen. The remote sensing reflectance for each individual leaf was then calculated by dividing mean radiance of the leaf with mean irradiance of the sky.

Leaf N content. Leaves measured for both photosynthetic capacity parameters and hyperspectral reflectance were placed in sealed plastic bags, kept inside a cooler, and transported to the laboratory for N determination. Fresh leaf area was measured using a LI-3100C Area Meter (LI-COR Biosciences Inc., Lincoln, NE, USA) in the laboratory. Leaves were then oven-dried at 60˚C for at least 72 h until all water was evaporated from samples, and their dry weight was measured. Next, dried leaves were ground to a fine powder that could pass through a 0.25 mm sieve and analyzed for N content via elemental combustion analysis (ECS 4010 CHNOS Elemental Analyzer, Costech Analytical Technologies Inc., Valencia, CA, USA). Standards, duplicates, and blanks were included in the analysis for quality assurance. Since all photosynthetic capacity parameters were calculated on an area basis, we calculated leaf N content on a per unit area basis (N_{area}; g N/m^2) by using leaf mass per area (LMA; calculated as the ratio of

dried leaf weight and fresh leaf area) as follows:

$$N_{\text{area}} = \frac{\text{leaf N weight}}{\text{leaf sample weight}} \times \text{LMA}$$

Tree biomass. To determine if leaf physicochemical parameters were correlated with woody biomass and productivity, the measured *Populus* trees were harvested to estimate their biomass at the end of the second growing season. The freshly harvested trees were separated into leaf and wood components with fresh wood weights measured in the field and fresh leaf weights measured in the laboratory. A wood sample from each tree was also returned to the laboratory and its fresh weight was recorded. Wood samples were oven-dried at 65˚C, and their dry weight was measured and used to calculate the dry mass of the field-weighed wood samples to determine dry woody biomass (kg) for each tree.

Statistical analysis

Taxa comparisons and correlations across parameters. We made statistical comparisons at the taxa level. We grouped the measured leaf sample data based on taxa. We added T×D genotypes into D×T, and the(D×N)×M genotype into D×N due to their small sample sizes (T×D, $n = 2$ leaf samples, and (D×N)×M, $n = 1$ leaf sample) (S1 Table). Therefore, we had five taxa groups. Since analysis of variance (ANOVA) assumptions, such as normal distribution of model residuals and homogeneity of variance across groups, were not met, a non-parametric Kruskal-Wallis test was performed to compare the differences in photosynthetic capacity parameters (V_{cmax}, J_{max}, and TPU), N_{area}, and tree biomass among the five taxa groups and between the Monroe and Pontotoc sites using R version 4.0.4 [89]. If taxa groups differed significantly, a post-hoc Dunn's Kruskal-Wallis multiple comparison test using the Bonferroni method was then performed to identify significant differences among the groups. To determine the relationship among V_{cmax}, J_{max}, TPU, N_{area}, and tree biomass, we used all measured leaf sample data ($n = 105$) and calculated a Pearson correlation coefficient matrix across these parameters.

Model development: Least absolute shrinkage and selection operator. We used all measured leaf sample data ($n = 105$) to develop models for estimating leaf physicochemical parameters from hyperspectral data across measured *Populus* taxa and genotypes. We followed the least absolute shrinkage and selection operator (LASSO) model development method to create hyperspectral leaf reflectance models for each leaf physicochemical parameter. To address the high-dimensional features of hyperspectral data and to reach a sparse solution, LASSO cross validation [90] (i.e., an L1-regularized regression) was applied for variable selection. Predictor variables were hyperspectral wavelength bands, and the response variables were photosynthetic capacity parameters (V_{cmax}, J_{max}, and TPU) and N_{area}.

All statistical analyses were performed in R version 4.0.4 [89]. Using R package 'glmnet' version 2.0.18 [91], *k*-fold cross validation for 'glmnet' was performed by specifying the number of folds as 10. Deviance, which used mean square error for Gaussian models, was assigned as the loss for cross validation. From the cross validation output, the tuning parameter lambda (λ) at minimum mean cross-validated error was selected. Once the optimal λ was determined, data were fit with a generalized linear model through penalized maximum likelihood using LASSO regularization [91]. LASSO minimized overfitting by shrinking the coefficients of variables that were not related with the response variable to zero and thus, the predictor variables with non-zero coefficients were selected for the linear model. Then, R^2, root mean square error (RMSE), and root mean square percentage error (RMSPE; calculated as $(\text{RMSE}/\bar{X}) \times 100\%$) of the selected models were calculated.

For model validation, the data were divided into training and testing datasets using the 'trainControl' function available in R package 'caret' version 6.0–84 [92]. Training and testing data included 70% and 30% of the entire data, respectively. The control parameters for cross validation were defined as 10-fold cross validation with 20 repetitions [93]. Then, the averaged values of repeated 10-fold cross-validated RMSE and R^2 of the model were calculated.

Model development: Principal component analysis. We used all measured leaf sample data (n = 105) for principal component analysis (PCA) model development. Using the 'prcomp' function in R, we conducted PCA to create hyperspectral leaf reflectance models for each physicochemical parameter. An unrotated P-mode PCA was conducted on the data matrix (spectral bands × leaves) to produce orthogonal principal components (PCs), of which those with eigenvalues greater than 1.0 were retained [94]. Two PCs were retained as they explained 98.2% of the variance (PC1 = 85.8% and PC2 = 12.4%). Next, linear regression was performed using PC1 and PC2 as predictor variables for each of the photosynthetic capacity parameters and N_{area}. Then, the performance of these linear models was evaluated with training and testing data using the methods described above for repeated 10-fold cross validation.

Results

Photosynthetic parameters, N_{area}, biomass, and leaf reflectance

V_{cmax} ranged from 67.8 to 368.8 with a mean of 184.4 ± 15.7 μmol/m^2/s; J_{max} ranged from 71.5 to 402.4 with a mean of 193.5 ± 17.3 μmol/m^2/s; TPU ranged from 5.3 to 25.8 with a mean of 13.2 ± 1.2 μmol/m^2/s; N_{area} ranged from 1.1 to 4.7 with a mean of 2.5 ± 0.1 g/m^2; and age two, whole-tree dry woody biomass ranged from 779.4 to 7288.4 with a mean of 4001.9 ± 341.7 g. These were the results calculated from all measured leaf sample data (n = 105).

Among the five taxa groups, D×N had 31% lower V_{cmax} than D×T; D×D had 32% lower J_{max} than D×T; and D×D had 31% lower TPU than D×T. For tree biomass, D×N and T×M were 39% and 54% lower than D×T, and 40% and 54% lower than D×D, respectively (Fig 1). There was no difference in N_{area} among the taxa (Fig 1). Generally, T×M had lower leaf reflectance especially when compared with D×N and D×T (Fig 2). The positions of peak and trough leaf reflectance were mainly observed within the range of 700–800 nm and 900–1000 nm, respectively (Fig 2 and S1 Fig).

Trees at the Monroe site had higher photosynthetic capacity and leaf N content than those at the Pontotoc site. Trees at the Monroe site had 37% higher V_{cmax}, 44% higher J_{max}, 45% higher TPU, 12% higher N_{area}, and 34% higher tree biomass than those at the Pontotoc site (Fig 1). Tree biomass was positively correlated (P-value < 0.05) with V_{cmax}, J_{max}, and TPU, but not with N_{area} (P-value = 0.22). N_{area} was positively correlated with V_{cmax}, J_{max}, and TPU (P-value < 0.05) (Table 1).

Variable selection and model performance

The LASSO model selected the same two wavelengths, 758 and 935 nm to estimate both V_{cmax} and J_{max} (Table 2 and S1 Fig). The TPU model selected three wavelengths, 687, 746, and 757 nm; and the N_{area} model selected four wavelengths, 304, 712, 921, and 1021 nm (Table 2 and S1 Fig). The LASSO model for N_{area} had the lowest RMSPE (21%), followed by TPU (31%), V_{cmax} (31%), and J_{max} (32%). In terms of R^2, the LASSO model for TPU was the highest (R^2 = 0.56), followed by J_{max} (R^2 = 0.54), V_{cmax} (R^2 = 0.51), and N_{area} (R^2 = 0.29) (Table 2 and Fig 3).

In the PCA method, PC loading in each spectral band indicated that PC1 did not have much variability, however, distinguishability mainly occurred in PC2 (Fig 4). For PC2, the wavelengths that had the highest positive loadings were 429, 431, 437, 439, and 414 nm, and those that had the highest negative loadings were 931, 928, 930, 934, and 927 nm. Both PC1

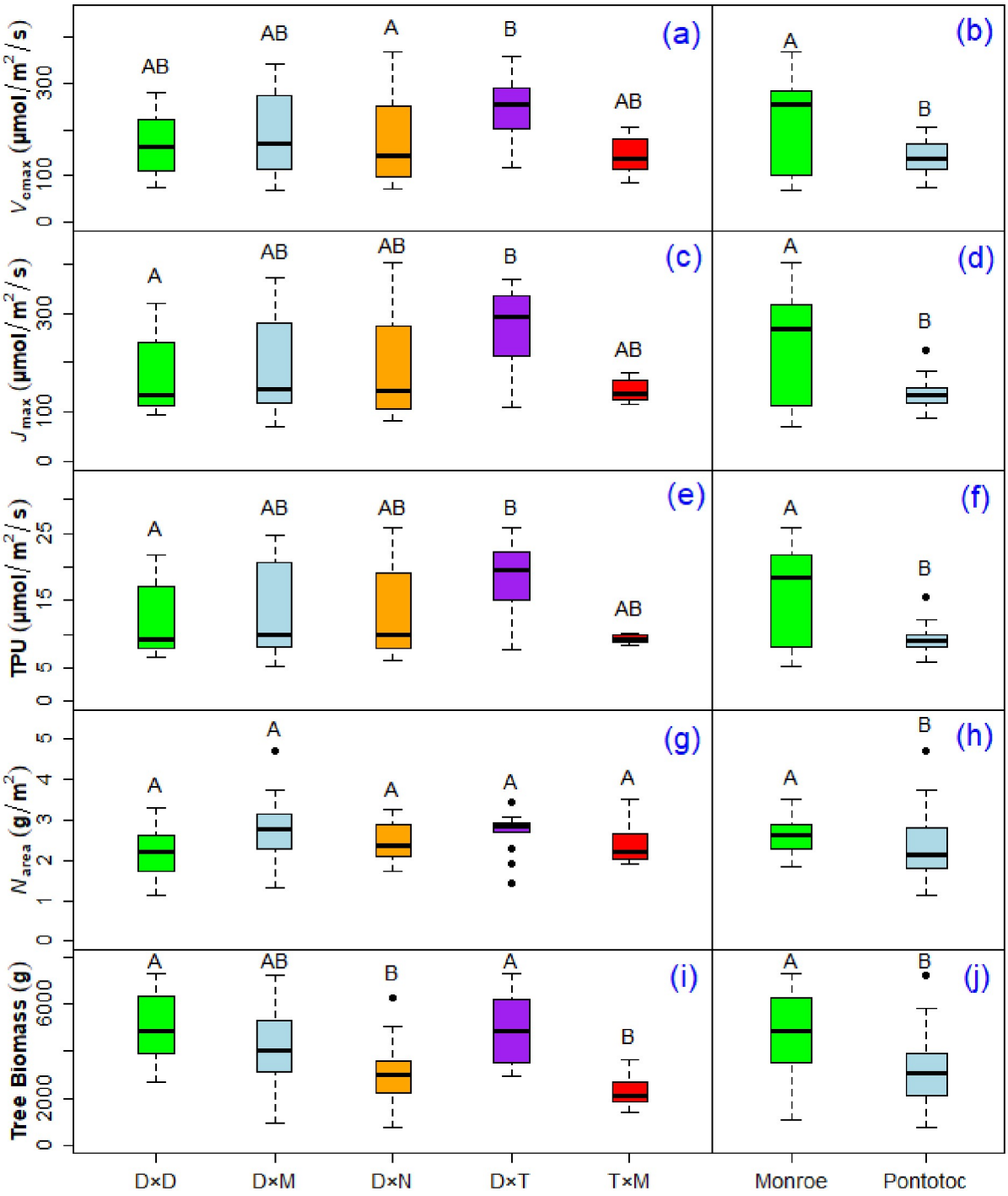

Fig 1. Box plots showing the variability among five different *Populus* taxa (left panels) and between the Monroe and Pontotoc sites (right panels). Bars with different letters indicate statistically significant differences (P-value < 0.05). Parameters include maximum carboxylation rate of Rubisco (V_{cmax}) by (a) taxa and (b) site, maximum electron transport limited carboxylation rate (J_{max}) by (c) taxa and (d) site, triose phosphate utilization limited carboxylation rate (TPU) by (e) taxa and (f) site, nitrogen per unit leaf area (N_{area}) by (g) taxa, (h) site, and tree biomass by (i) taxa and (j) site. D×T and T×D taxa were combined and shown as D×T, and (D×N)×M was included in D×N (S1 Table). D×D = *P. deltoides* × *P. deltoides*, D×M = *P. deltoides* × *P. maximowiczii*, D×N = *P. deltoides* × *P. nigra*, D×N×M = *P. deltoides* × *P. nigra* × *P. maximowiczii*, D×T = *P. deltoides* × *P. trichocarpa*, T×D = *P. trichocarpa* × *P. deltoides*, and T×M = *P. trichocarpa* × *P. maximowiczii*.

https://doi.org/10.1371/journal.pone.0264780.g001

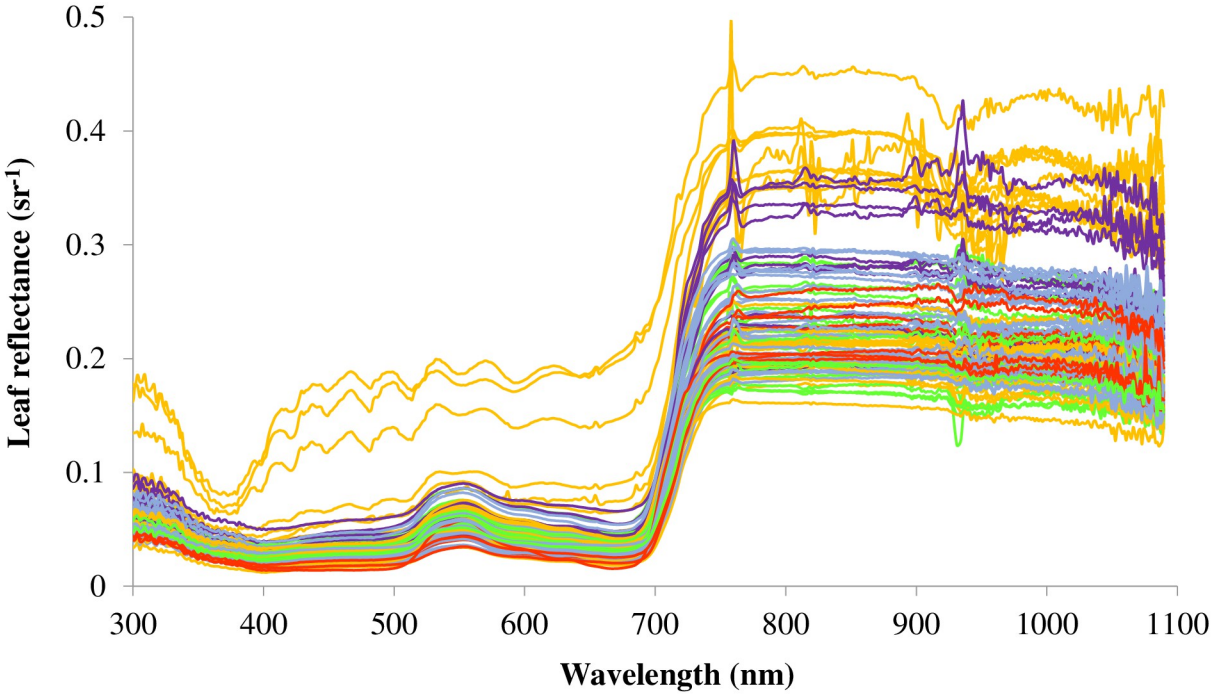

Fig 2. Leaf reflectance curves of five different *Populus* taxa. D×T and T×D taxa were combined and shown as D×T, and taxa (D×N)×M was included in D×N (S1 Table). D×D = *P. deltoides* × *P. deltoides*, D×M = *P. deltoides* × *P. maximowiczii*, D×N = *P. deltoides* × *P. nigra*, (D×N)×M = *P. deltoides* × *P. nigra* × *P. maximowiczii*, D×T = *P. deltoides* × *P. trichocarpa*, T×D = *P. trichocarpa* × *P. deltoides*, and T×M = *P. trichocarpa* × *P. maximowiczii*.

https://doi.org/10.1371/journal.pone.0264780.g002

and PC2 loadings indicated that the wavelengths near 935 nm were significant (Fig 4), and the same wavelength was also selected by the LASSO models for estimating V_{cmax} and J_{max} (Table 2).

The PCA models had comparable RMSPE and R^2 as the LASSO models for V_{cmax}, J_{max}, and TPU. Only the PCA model for N_{area} had substantially lower R^2 than the LASSO model for N_{area} (PCA = 0.15 vs. LASSO = 0.29) but RMSPE was comparable (PCA = 23% vs. LASSO = 21%) (Table 2). Among the PCA models, the PCA model that estimated N_{area} had the lowest RMSPE (RMSPE = 24%), followed by V_{cmax} (RMSPE = 31%), J_{max} (RMSPE = 32%), and TPU (RMSPE = 33%) (Table 2 and Fig 3). The PCA model for J_{max} had the highest R^2 (R^2 = 0.53), followed by models for V_{cmax} (R^2 = 0.51), TPU (R^2 = 0.50), and N_{area} (R^2 = 0.15).

The model validation using repeated 10-fold cross validation indicated that there was a very slight difference in RMSEs among the training data, test data, and entire dataset for both

Table 1. Pearson correlation coefficient (*r*) matrix among tree biomass, V_{cmax}, J_{max}, TPU, and N_{area}.

	Tree biomass	V_{cmax}	J_{max}	TPU	N_{area}
Tree biomass	1.00				
V_{cmax}	**0.33**	1.00			
J_{max}	**0.35**	**0.97**	1.00		
TPU	**0.36**	**0.96**	**0.99**	1.00	
N_{area}	0.12 ^{ns}	**0.39**	**0.38**	**0.36**	1.00

Non-significant correlations are denoted by "ns" superscript, and the rest with "bold" font are significant at *P*-value < 0.05.

https://doi.org/10.1371/journal.pone.0264780.t001

Table 2. Selected predictor variables in the models and averaged 10-fold cross validation results.

Model	Model form	R^2	RMSPE	RMSE (Entire data)	Cross validated RMSE	
					Training data	Test data
LASSO: V_{cmax}	-30.2 + (601.1 × ρ758.29) + (302.7 × ρ935.71)	0.51	31%	57.20	59.07	57.37
LASSO: J_{max}	-44.6 + (796.2 × ρ758.29) + (200.57 × ρ935.71)	0.54	32%	60.98	61.68	60.64
LASSO: TPU	-6.4 + (-62.0 × ρ687.03) + (35.2 × ρ745.99) + (59.9 × ρ756.76)	0.56	31%	4.02	4.02	4.37
LASSO: N_{area}	2.0 + (28.3 × ρ303.51) + (-24.4 × ρ711.97) + (0.7 × ρ920.68) + (5.9 × ρ1021.44)	0.29	21%	0.51	0.53	0.51
PCA: V_{cmax}	184.4 + (-1.9 × PC1) + (-5.6 × PC2) + (-0.03 × PC1 × PC2)	0.52	31%	56.99	56.68	59.54
PCA: J_{max}	193.5 + (-2.4 × PC1) + (-5.3 × PC2) + (-0.01 × PC1 × PC2)	0.53	32%	61.92	62.46	64.72
PCA: TPU	13.2 + (-0.2 × PC1) + (-0.3 × PC2) + (-0.0007 × PC1 × PC2)	0.50	33%	4.30	4.20	4.65
PCA: N_{area}	2.5 + (0.0008 × PC1) + (-0.04 × PC2) + (-0.0005 × PC1 × PC2)	0.15	23%	0.56	0.60	0.55

Training and testing data contained 70% and 30% of the entire data, respectively. ρ is spectral reflectance at the corresponding wavelengths.

https://doi.org/10.1371/journal.pone.0264780.t002

LASSO and PCA models (Table 2). Therefore, both LASSO and PCA models were not over-fit. Cross-validation also indicated that RMSEs were generally higher, albeit modestly, for the PCA models compared with the LASSO models.

Discussion

Model performance

Our results showed promise in estimating photosynthetic capacity parameters (V_{cmax}, J_{max}, and TPU) using hyperspectral leaf reflectance (Table 2). In both LASSO and PCA, the models for estimating N_{area} had the lowest R^2, but RMSPE was smaller for N_{area} than that of the estimates of V_{cmax}, J_{max}, and TPU (Table 2). The N_{area} model likely had the smallest RMSPE

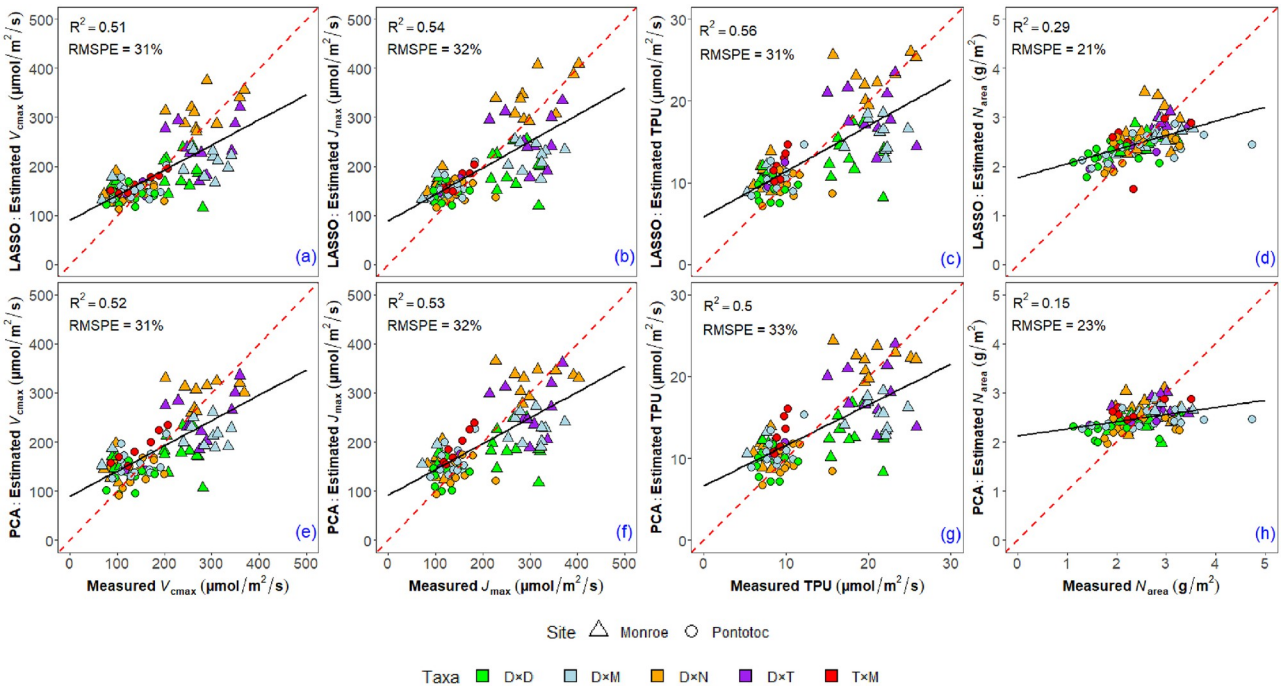

Fig 3. Scatterplots showing the linear relationship between measured and estimated V_{cmax}, J_{max}, TPU and N_{area} for the LASSO models: (a), (b), (c), and (d) and the PCA models: (e), (f), (g), and (h). The solid black line is the regression line, and the dotted red line is the 1:1 line.

https://doi.org/10.1371/journal.pone.0264780.g003

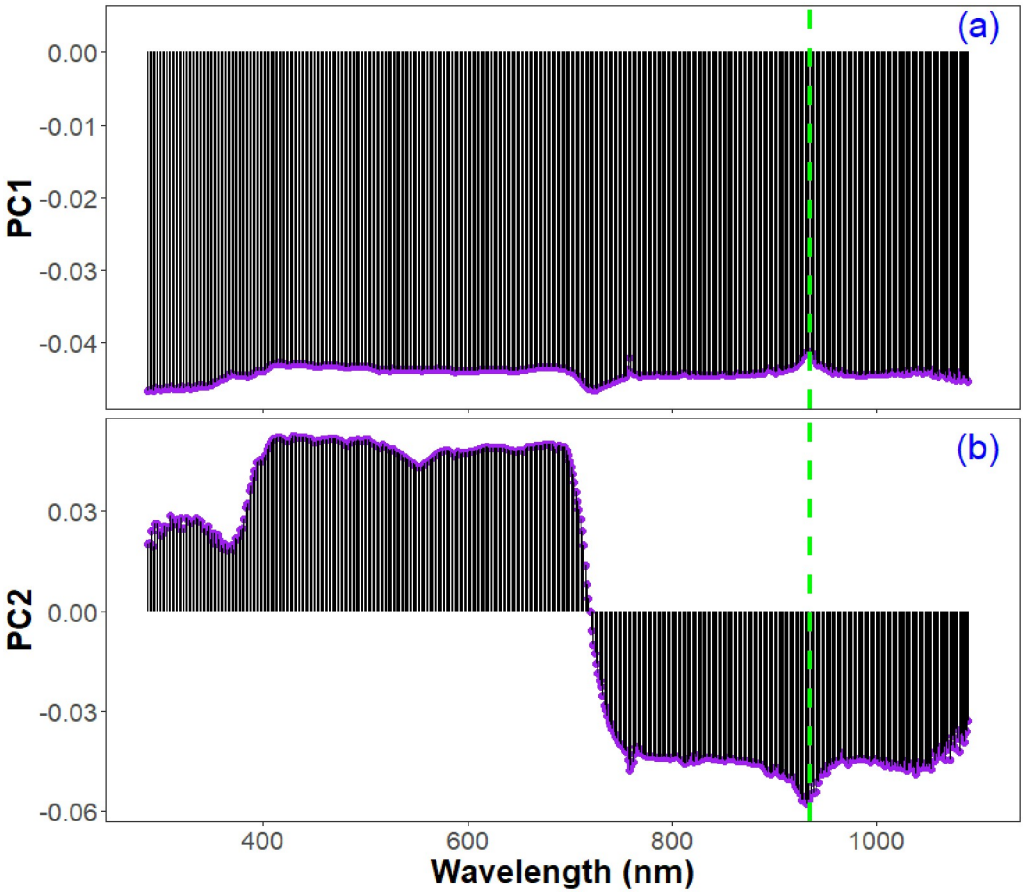

Fig 4. Lollipop plot showing the loadings of all spectral wavelengths in (a) PC1 and (b) PC2. As shown by vertical green dashed line, a significant peak and trough occurred at or near 935 nm, which was also selected by the LASSO models for estimating V_{cmax} and J_{max}.

https://doi.org/10.1371/journal.pone.0264780.g004

because of having smaller variance compared to that of V_{cmax}, J_{max}, and TPU (Fig 1). The N_{area} model had the smallest R^2 probably because N_{area} was weakly correlated with photosynthetic capacity parameters (Table 1) suggesting that a large portion of N in leaves was not involved in photosynthesis, and consequently, might not generate a strong spectral signal. In addition to N_{area}, we also developed a model for leaf N content per unit mass (i.e., N_{mass} (%); calculated by dividing leaf N mass with leaf sample mass and then, multiplying with 100). Nonetheless, the N_{area} model (R^2 = 0.29 for LASSO and 0.15 for PCA) fit the data better than that of N_{mass} (%) (R^2 = 0.07 for LASSO and 0.02 for PCA). N_{area} resulted in a better model likely because N_{area} incorporated the information of both N_{mass} and leaf mass per area in its calculation.

For screening, new genotypes are developed by breeding within and potentially between taxa. Therefore, creating models from taxa- or genotype-specific data may not be robust enough because new genotypes that need to be screened for productivity potential are being developed. In this study, we used all leaf sample data (n = 105) measured across 7 taxa and 62 genotypes while developing LASSO and PCA models, with many of these genotypes commonly selected for establishing short rotation plantations across different geographic regions. Our measurements at both study plantations represented a single growing season, but we measured a wide variety of taxa and genotypes to generate large variability in physicochemical

parameters (V_{cmax}, J_{max}, TPU, and N_{area}) and to develop robust models for predicting *Populus* parameters. Therefore, our hyperspectral leaf reflectance captured variability in physiological functioning across a diverse range of *Populus* taxa and genotypes.

LASSO and PCA model development techniques had similar capacity in estimating V_{cmax}, J_{max}, and TPU. However, the LASSO models used only 2 to 4 wavelengths to estimate all response variables, while the PCA models used 2 PCs, including all 512 wavelengths. Consequently, it is easier to interpret the results of the LASSO models compared to the PCA models. On the other hand, the LASSO method requires more sophisticated analysis. Since LASSO is a supervised feature extraction method, it needs to define an optimal value for the tuning parameter (λ) using cross-validation procedures to adjust the balance between sparsity (i.e., the model containing few non-zero coefficients) and model prediction accuracy [59, 71]. Furthermore, LASSO selects only one variable from the group if predictor variables within a group are highly correlated [67, 95]. Despite these setbacks, our findings suggested that sparse statistical modeling with LASSO could simplify models by selecting a small subset of predictor variables, which was especially helpful for analyzing data with high dimensionality and complexity in which small and critical differences in data could not be easily detected [72]. Therefore, we found that the LASSO method was able to develop acceptable models for the hyperspectral data in which the number of predictor variables was significantly larger than the sample size.

In contrast to the LASSO method, the PCA method was an unsupervised dimension reduction technique. Consequently, it is easier and quicker to generate results with minimum input from the analyst. The PCA method transforms the high-dimensional space of correlated variables into a new feature space of high variance, and consequently can achieve the tasks of both feature and noise reduction [96]. However, the LASSO method does not transform the data, and consequently, the data might still contain noise [97]. The disadvantage of the PCA method is that, while constructing the principal components to minimize the correlations among them, the principal components do not necessarily tie in with the interest of specific model outcome [67]. Because all the variables exist in each principal component, it is harder to interpret the direct relationship between individual spectral wavelengths and leaf physicochemical parameters. The model performance between LASSO and PCA methods was equivalent in estimating photosynthetic capacity parameters, however, the LASSO model outperformed the PCA model in estimating N_{area} (Table 2). Therefore, the LASSO method was able to provide a simple and robust model for all response variables even though it required more time and effort during model development.

We also analyzed previously published vegetation indices in linear regressions by using our measured data in order to evaluate whether these vegetation indices were comparable with our LASSO and PCA models in estimating V_{cmax}, J_{max}, TPU, and N_{area}. We calculated R^2 and RMSE of vegetation indices and compared these parameters with our best-fit LASSO and PCA models. We found that published vegetation indices did not perform better than our models because our models, particularly the LASSO models, had the highest R^2 and the lowest RMSE for estimating photosynthetic capacity parameters and N_{area} in our *Populus* trees (Table 3). This was likely because these vegetation indices were developed using other species and to estimate other physiological and leaf- and canopy-level parameters than we estimated. After LASSO and PCA models, the hyperspectral vegetation indices developed for estimating canopy-level growth parameters in cotton (*Gossypium hirsutum* L.) and leaf-level chlorophyll content in a wide range of plants [98–102] showed the next best performance for estimating V_{cmax}, J_{max}, and TPU (Table 3 and S2 Table). Regarding N_{area}, the LASSO model performed the best. However, hyperspectral red edge index (ρ738/ρ720) [103] and linear red edge index [98] previously used for estimating leaf area index in wheat and canopy N in cotton, and

Table 3. Comparison of model performance among the LASSO model, the PCA model, and published vegetation indices.

Source	Formula	V_{cmax}		J_{max}		TPU		N_{area}	
		R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
LASSO	This study	0.51	**57.20**	0.54	**60.98**	0.56	**4.02**	0.29	**0.51**
PCA	This study	0.52	**56.99**	0.53	**61.92**	0.50	**4.30**	0.15	0.56
Carter et al. [104]	ρ695/ρ420	0.02^{ns}	81.21	0.03^{ns}	88.67	0.03^{ns}	5.97	0.01^{ns}	0.60
Carter et al. [104]	ρ695/ρ760	0.00^{ns}	81.85	0.00^{ns}	89.88	0.00^{ns}	6.05	0.03^{ns}	0.60
Datt [99]	ρ672/(ρ550×ρ708)	0.08	**78.62**	0.09	**85.74**	0.11	**5.72**	0.03^{ns}	0.60
Gamon (Photochemical Reflectance Index) [105]	(ρ531-ρ570)/(ρ531+ρ570)	0.00^{ns}	81.80	0.00^{ns}	89.70	0.01^{ns}	6.04	0.04^{ns}	0.60
Gitelson and Merzlyak [106]	ρ750/ρ705	0.05	79.80	0.04	88.22	0.03^{ns}	5.97	0.15	0.56
Gitelson and Merzlyak [107]	(ρ750-ρ705)/ (ρ750+ρ705)	0.03^{ns}	80.53	0.02^{ns}	88.88	0.02^{ns}	6.00	0.13	0.57
Gitelson and Merzlyak [108]	ρ750/ρ550	0.06	79.28	0.04	87.94	0.03^{ns}	5.96	0.11	0.57
Gitelson and Merzlyak [108]	ρ750/ρ700	0.02^{ns}	80.88	0.01^{ns}	89.31	0.01^{ns}	6.03	0.09	0.58
Gitelson et al. (Green Normalized Difference Vegetation Index) [109]	(ρ750-ρ550)/(ρ750+ρ550)	0.04	80.23	0.03^{ns}	88.76	0.02^{ns}	6.00	0.10	0.58
Gitelson et al. [100]	1/ρ700	0.16	**74.85**	0.20	**80.34**	0.21	**5.40**	0.00^{ns}	0.61
Gupta et al. (Hyperspectral Red Edge Index) [103]	ρ735/ρ720	0.06	79.54	0.05	87.59	0.04	5.93	0.17	0.55
Gupta et al. (Hyperspectral Red Edge Index) [103]	ρ741/ρ717	0.01^{ns}	81.26	0.01^{ns}	89.48	0.01^{ns}	6.03	0.16	0.56
Gupta et al. (Hyperspectral Red Edge Index) [103]	ρ747/ρ708	0.06	79.49	0.05	87.84	0.04^{ns}	5.95	0.17	0.55
Gupta et al. (Hyperspectral Red Edge Index) [103]	ρ738/ρ720	0.07	78.78	0.07	86.75	0.06	5.88	0.19	0.55
Le Maire et al. [101]	(ρ749-ρ720)-(ρ701-ρ672)	0.29	**69.06**	0.30	**75.48**	0.27	**5.18**	0.21	**0.54**
Maccioni et al. [102]	(ρ780-ρ710)/(ρ780-ρ680)	0.12	**76.57**	0.12	**84.25**	0.10	**5.73**	0.25	**0.52**
Penuelas et al. (Structure Insensitive Pigment Index) [110]	(ρ800-ρ445)/(ρ800-ρ680)	0.03^{ns}	80.47	0.04	88.09	0.04	5.94	0.00^{ns}	0.61
Raper and Varco (Linear Red Edge Index) [98]	700+40[(ρ670+ρ780)/2-ρ700]/ (ρ740-ρ700)	0.12	**76.95**	0.11	**84.92**	0.09	**5.78**	0.24	**0.53**
Sims and Gamon [111]	(ρ750-ρ445)/(ρ705-ρ445)	0.06	79.26	0.06	87.36	0.04	5.92	0.17	0.55
Sims and Gamon [111]	(ρ800–680)/(ρ800+ρ680–2ρ445)	0.03^{ns}	80.64	0.04^{ns}	88.29	0.03^{ns}	5.95	0.00^{ns}	0.61
Sims and Gamon [111]	(ρ800-ρ445)/(ρ680-ρ445)	0.01^{ns}	81.36	0.01^{ns}	89.38	0.02^{ns}	6.00	0.03^{ns}	0.60
Stimson et al. (Hyperspectral Normalized Difference Vegetation Index) [112]	(ρ860-ρ690) /(ρ860 +ρ690)	0.00^{ns}	81.75	0.01^{ns}	89.56	0.01^{ns}	6.03	0.02^{ns}	0.60
Vogelmann et al. [113]	ρ740/ρ720	0.07	78.85	0.07	86.82	0.06	5.88	0.19	0.55
Wen et al. [114]	ρ764/ρ716	0.00^{ns}	81.71	0.00^{ns}	89.88	0.00^{ns}	6.06	0.15	0.56
Zarco-Tejada et al. (Red Edge Optical Index) [115]	ρ750/ρ710	0.06	79.28	0.05	87.63	0.04	5.93	0.18	0.55

Spectral reflectance is denoted by ρ, and non-significant correlations are indicated by the "ns" superscript. The models and vegetation indices having the lowest RMSE for estimating each parameter are bolded. The sources of the alphabetically ordered vegetation indices and their study objectives are described in S2 Table.

https://doi.org/10.1371/journal.pone.0264780.t003

vegetation indices developed by Maccioni et al. [102] and Le Maire et al. [101] for estimating leaf chlorophyll content in trees outperformed the PCA model (Table 3 and S2 Table).

Interpretation of selected wavelengths

As with previous studies [32, 116], we found a strong positive correlation between V_{cmax} and J_{max} ($r = 0.97$) (Table 1) likely due to their coordinated processes in CO_2 assimilation. The LASSO models selected the same two wavelengths: 758 and 935 nm for estimating both V_{cmax} and J_{max} (Table 2 and S1 Fig). PCA also indicated 935 nm as a significant wavelength in both PC1 and PC2 (Fig 4). In our leaf reflectance curves (Fig 2 and S1 Fig), peaks and troughs of spectral reflectance mainly occurred at 758 nm and 935 nm. The wavelength 758 nm is an important wavelength for photosynthesis as it corresponds with the sun-stimulated fluorescence emission spectrum in plants [117, 118]. The spike in reflectance near 758 nm (Fig 2 and

S1 Fig) is the spectral signature caused by chlorophyll molecules that re-emit light to dissipate energy and return from excited to non-excited states [118]. In addition, chlorophyll fluorescence emission can be a sign of plant water stress, and steady-state chlorophyll fluorescence emission can reduce net CO_2 assimilation [119–122]. According to Frankenberg et al. [123], Fraunhofer lines at 758.8 and 770.1 nm can be used for retrieving chlorophyll fluorescence. Satellite-derived solar-induced chlorophyll fluorescence data have also been suggested as a proxy for estimating and mapping V_{cmax} for global climate change studies [124–126]. Moreover, chlorophyll fluorescence emission can be an indicator of increased photorespiration, which can reduce the rate of photosynthetic capacity [127]. The near infrared wavelength 935 nm is commonly identified as a key wavelength in several vegetation indices. For example, this wavelength was found in the normalized difference vegetation index (NDVI = (ρ935 –ρ661)/(ρ935 + ρ661), in which ρ = spectral reflectance) used to evaluate the visual quality and the drought stress responses of turfgrasses [128–130] as well as to explore the photosynthetic responses under varying salinity stresses in seashore paspalum (*Paspalum vaginatum* Swartz) ecotypes [131]. Furthermore, according to Wang et al. [132], 935 nm was sensitive to leaf greenness and chlorophyll content [133]. Moreover, 935 nm was associated with thermal dissipation of absorbed energy under stress conditions. For example, Zhang et al. [134] observed that non-photochemical quenching (i.e., a mechanism used by plants to protect from adverse effects of high light intensity by dissipating the excessively absorbed radiant energy into heat in the photosystem II antenna complexes) was negatively correlated with the spectral reflectance range 935–945 nm under salinity stress. There was also a significant inverse relationship between non-photochemical quenching and light-adapted fluorescence yield [135, 136] because the fluorescence signal could originate from photosystem II [137]. In addition to its correlation with the leaf chlorophyll fluorescence signal under different abiotic stresses [66, 134], reflectance at 935 nm is related with the water content in plant leaves [138] and could provide potential for understanding leaf water status [139].

The TPU model selected remote sensing reflectance at 687, 746, and 757 nm (Table 2 and S1 Fig). Due to its capacity to detect chlorophyll fluorescence signal, the wavelength 757 nm has been widely used for modeling terrestrial gross primary productivity from space [140–145]. In addition, 687 nm is also used for sun-induced fluorescence retrieval for estimating leaf- and canopy-level net photosynthesis of vegetation [117, 146–149]. The absorption maximum positions of the excitation spectra of fluorescence also occurred at 746 nm in pea plants [150]. The wavelength 746 nm is also associated with photosynthetic processes. Traditionally, the wavelength range of 400–700 nm is considered photosynthetically active radiation (PAR). However, recent studies suggested wavelengths 711 nm, 723 nm, and 746 nm as extended PAR wavelengths as well [151, 152] because these far-red wavelengths also manifested comparable efficiency at driving canopy photosynthesis and yield of both C_3 and C_4 species and their cultivars [152]. The reflectance at 746 nm may distinguish levels of accessary pigments with Sonobe et al. [153] finding that reflectance at 746 nm had a strong negative correlation with carotenoid concentration in shade-grown tea leaves. Therefore, the predictors included in our TPU model (687, 746, and 757 nm) are in agreement with formerly proposed wavelengths for estimating photosynthetic capacity.

Our hyperspectral leaf reflectance model for estimating N_{area} had lower R^2 (0.29 for the LASSO model and 0.15 for the PCA model) than models for estimating V_{cmax}, J_{max}, and TPU (Table 2). Nevertheless, all four wavelengths selected in our N_{area} LASSO model have been reported elsewhere as critical wavelengths, which were largely associated with, and also useful for, estimating leaf N content in different plant species. Our best model for estimating N_{area} contained four wavelengths– 304, 712, 921, and 1021 nm (Table 2 and S1 Fig). The first wavelength, 304 nm, is in the ultraviolet-B (UV-B) region, which extends from 280 to 320 nm

[154–156]. Leaf protective chemicals have been shown to absorb UV-B radiation at 305 nm [157]. Leaves with more protective chemicals to absorb UV-B radiation might have less N content because protective compounds (e.g., flavonoid content) in plants were found to have a negative relationship with leaf N content [158]. The wavelength near 712 nm was also highly correlated with total N content in cotton plants (*Gossypium hirsutum* L.) [159]. In Chinese cabbage (*Brassica campestris* L. ssp. *Pekinensis* 'Norgangbom') leaves, 710 nm was correlated with chlorophyll content and was also a significant wavelength for predicting leaf N [160]. According to Zhao et al. [41], changes in reflectance at 710 nm could detect levels of N deficiency because of its negative relationship with leaf N in corn (*Zea mays* L.). Besides agricultural crops, 711 nm was one of the correlated spectral bands with leaf N in balsam fir (*Abies balsamea* L. Mill) [161]. The wavelength near 921 nm can be a leaf N predictor as well. Zhao et al. [162] found that the reflectance at 920 nm was associated with leaf N concentration and canopy N density in winter wheat (*Triticum aestivum* L.), and the reflectance at this wavelength was controlled by leaf structural features. Moreover, Tarpley et al. [163] used 920 nm for estimating leaf N content in cotton, and Yu et al. [164] used 921 nm for estimating whole-plant total (including leaves, stems, and roots) N content in pepper (*Capsicum annuum* L.) plants. The wavelength near 1021 nm also showed a relationship with leaf N content. Due to the N-H stretch from protein (i.e., the stretching vibrations of the strong molecular bonds between hydrogen atoms and nitrogen atoms), the N absorption feature centers at 1020 nm in leaves [165]. Moreover, 1020 nm was used as a N absorption feature for estimating canopy N in sagebrush (*Artemisia* spp.) [166] and as a protein absorption feature for estimating leaf N concentrations in grass [167]. Therefore, previously documented studies support the wavelengths selected in our LASSO models for estimating V_{cmax}, J_{max}, TPU, and N_{area}.

Conclusion

LASSO and PCA model development methods showed similar capacity for estimating photosynthetic capacity parameters, but the LASSO model outperformed the PCA model in estimating N_{area}. For analyzing hyperspectral data, the LASSO models resulted in superior model interpretability compared to the PCA models. With only a few spectral wavelengths, narrow-band hyperspectral reflectance can estimate photosynthetic capacity and leaf N content of highly variable eastern cottonwood and hybrid poplar genotypes. This is particularly critical when the most important spectral bands are adjacent to each other (within about 10 nm in TPU model; Table 2) suggesting that high spectral resolution is needed for spectral precision and modeling biophysical traits. Previous physiological studies also found that wavelengths selected by our LASSO models were strongly correlated with photosynthetic capacity and leaf N content. The results of our single-leaf studies corroborated that sun-induced chlorophyll fluorescence retrieval bands (687 and 758 nm) [117, 146–149] could predict V_{cmax}, J_{max}, and TPU. This suggests that sun-induced fluorescence techniques can be successful in estimating V_{cmax} and J_{max} of *Populus* plantations at the ecosystem level. In total, hyperspectral leaf reflectance can be used as a cost-effective means for high-throughput phenotyping and rapid clonal screening due to its ability to model photosynthetic capacity and leaf N content of a wide variety of *Populus* genotypes.

Supporting information

S1 Fig. Locations of wavelengths selected by the LASSO models for estimating V_{cmax}, J_{max}, TPU, and N_{area}. D×T and T×D taxa were combined and shown as D×T, and (D×N)×M taxa was included in D×N (S1 Table). D×D = *P. deltoides* × *P. deltoides*, D×M = *P. deltoides* × *P. maximowiczii*, D×N = *P. deltoides* × *P. nigra*, (D×N)×M = *P. deltoides* × *P. nigra* × *P.*

maximowiczii, D×T = *P. deltoides* × *P. trichocarpa*, T×D = *P. trichocarpa* × *P. deltoides*, and T×M = *P. trichocarpa* × *P. maximowiczii*.
(DOCX)

S1 Table. List of taxa and genotypes measured at the Monroe and Pontotoc sites. Monroe and Pontotoc columns describe the number of leaves on which measurements were made and measurement dates. There were a total of seven taxa and 62 genotypes measured at both study sites.
(DOCX)

S2 Table. Study objectives of selected vegetation indices.
(DOCX)

Acknowledgments

This publication is a contribution of the Forest and Wildlife Research Center, Mississippi State University. We thank Dr. Randall J. Rousseau for supervising the *Populus* plantations. We also thank two anonymous reviewers for their constructive suggestions.

Author Contributions

Conceptualization: Thu Ya Kyaw, Courtney M. Siegert, Heidi J. Renninger.

Data curation: Thu Ya Kyaw, Justin J. Pitts, Heidi J. Renninger.

Formal analysis: Thu Ya Kyaw, Courtney M. Siegert.

Funding acquisition: Courtney M. Siegert, Heidi J. Renninger.

Investigation: Thu Ya Kyaw, Courtney M. Siegert, Heidi J. Renninger.

Methodology: Thu Ya Kyaw, Courtney M. Siegert, Padmanava Dash, Krishna P. Poudel, Heidi J. Renninger.

Project administration: Thu Ya Kyaw, Courtney M. Siegert, Heidi J. Renninger.

Resources: Thu Ya Kyaw, Courtney M. Siegert, Padmanava Dash, Heidi J. Renninger.

Supervision: Thu Ya Kyaw, Courtney M. Siegert, Heidi J. Renninger.

Validation: Thu Ya Kyaw, Courtney M. Siegert.

Writing – original draft: Thu Ya Kyaw.

Writing – review & editing: Courtney M. Siegert, Padmanava Dash, Krishna P. Poudel, Justin J. Pitts, Heidi J. Renninger.

References

1. Stanton B, Eaton J, Johnson J, Rice D, Schuette B, Moser B. Hybrid poplar in the Pacific Northwest: The effects of market-driven management. J For. 2002; 100(4):28–33.
2. Vogt KA, Andreu MG, Vogt DJ, Sigurdardottir R, Edmonds RL, Schiess P, et al. Societal values and economic return added for forest owners by linking forests to bioenergy production. J For. 2005; 103 (1):21–27.
3. Volk TA, Abrahamson LP, Nowak CA, Smart LB, Tharakan PJ, White EH. The development of short-rotation willow in the northeastern United States for bioenergy and bioproducts, agroforestry and phytoremediation. Biomass Bioenergy. 2006; 30(8–9):715–727.
4. Shi E, Hanson S. EIA projects U.S. biofuel production to slowly increase through 2050. U.S. Energy Information Administration (EIA); 2020 [cited 19 March 2020]; https://www.eia.gov/todayinenergy/detail.php?id=43096.

5. Hinchee M, Rottmann W, Mullinax L, Zhang C, Chang S, Cunningham M, et al. Short-rotation woody crops for bioenergy and biofuels applications. In Vitro Cell Dev Biol Plant. 2009; 45(6):619–629. https://doi.org/10.1007/s11627-009-9235-5 PMID: 19936031

6. Kelley S, Gustafson R, Puettmann M, Oneil E, Lippke B, Morales-Vera R, et al. Carbon cycling, environmental & rural economic impacts of collecting & processing specific woody feedstocks in biofuels. CORRIM-Consortium for Research on Renewable Industrial Materials, 2020.

7. Oliveira N, Sixto H, Cañellas I, Rodríguez-Soalleiro R, Pérez-Cruzado C. Productivity model and reference diagram for short rotation biomass crops of poplar grown in Mediterranean environments. Biomass Bioenergy. 2015; 72:309–320.

8. Dowell RC, Gibbins D, Rhoads JL, Pallardy SG. Biomass production physiology and soil carbon dynamics in short-rotation-grown *Populus deltoides* and *P. deltoides* × *P. nigra* hybrids. For Ecol Manage. 2009; 257(1):134–142.

9. Harper RJ, Sochacki SJ, Smettem KRJ, Robinson N. Bioenergy feedstock potential from short-rotation woody crops in a dryland environment. Energy Fuels. 2010; 24(1):225–231.

10. Kauter D, Lewandowski I, Claupein W. Quantity and quality of harvestable biomass from *Populus* short rotation coppice for solid fuel use—a review of the physiological basis and management influences. Biomass Bioenergy. 2003; 24(6):411–427.

11. Ye X, Busov V, Zhao N, Meilan R, McDonnell LM, Coleman HD, et al. Transgenic *Populus* trees for forest products, bioenergy, and functional genomics. Crit Rev Plant Sci. 2011; 30(5):415–434.

12. Zalesny RS, Wiese AH, Bauer EO, Riemenschneider DE. *Ex situ* growth and biomass of *Populus* bioenergy crops irrigated and fertilized with landfill leachate. Biomass Bioenergy. 2009; 33(1):62–69.

13. Volk TA, Berguson B, Daly C, Halbleib MD, Miller R, Rials TG, et al. Poplar and shrub willow energy crops in the United States: Field trial results from the multiyear regional feedstock partnership and yield potential maps based on the PRISM-ELM model. Glob Change Biol Bioenergy. 2018; 10 (10):735–751.

14. Kszos LA, Downing ME, Wright LL, Cushman JH, McLaughlin SB, Tolbert VR, et al. Bioenergy feedstock development program status report. ORNL/TM-2000/292. Oak Ridge, TN, USA: Oak Ridge National Laboratory, 2000.

15. Nielsen UB, Madsen P, Hansen JK, Nord-Larsen T, Nielsen AT. Production potential of 36 poplar clones grown at medium length rotation in Denmark. Biomass Bioenergy. 2014; 64:99–109.

16. Orság M, Fischer M, Tripathi AM, Žalud Z, Trnka M. Sensitivity of short rotation poplar coppice biomass productivity to the throughfall reduction—Estimating future drought impacts. Biomass Bioenergy. 2018; 109:182–189.

17. González-García S, Bacenetti J, Murphy RJ, Fiala M. Present and future environmental impact of poplar cultivation in the Po Valley (Italy) under different crop management systems. Journal of Cleaner Production. 2012; 26(56–66).

18. Oliveira N, Pérez-Cruzado C, Cañellas I, Rodríguez-Soalleiro R, Sixto H. Poplar short rotation coppice plantations under Mediterranean conditions: The case of Spain. Forests. 2020; 11:1352.

19. Verlinden MS, Broeckx LS, Ceulemans R. First vs. second rotation of a poplar short rotation coppice: Above-ground biomass productivity and shoot dynamics. Biomass Bioenergy. 2015; 73:174–185.

20. Dillen SY, Djomo SN, Al Afas N, Vanbeveren S, Ceulemans R. Biomass yield and energy balance of a short-rotation poplar coppice with multiple clones on degraded land during 16 years. Biomass Bioenergy. 2013; 56:157–165.

21. Fahrenkrog AM, Neves LG, Resende MF Jr, Dervinis C, Davenport R, Barbazuk WB, et al. Population genomics of the eastern cottonwood (*Populus deltoides*). Ecol Evol. 2017; 7(22):9426–9440. https://doi.org/10.1002/ece3.3466 PMID: 29187979

22. Barrett JW, Rajora OP, Yeh FH, Dancik BP, Strobeck C. Mitochondrial DNA variation and genetic relationships of *Populus* species. Genome. 1993; 36(1):87–93. https://doi.org/10.1139/g93-012 PMID: 18469973

23. Davis JM. Genetic improvement of poplar (*Populus* spp.) as a bioenergy crop. In: Vermerris W, editor. Genetic improvement of bioenergy crops. New York, USA: Springer; 2008. p. 377–396.

24. Stanton BJ, Neale DB, Li S. *Populus* breeding: from the classical to the genomic approach. In: Jansson S, Bhalerao R, Groover A, editors. Genetics and genomics of *Populus*. New York, USA: Springer; 2010. p. 309–348.

25. Herr JR, Carlson JE. Traditional breeding, genomics-assisted breeding, and biotechnological modification of forest trees and short rotation woody crops. In: Jacobson M, Ciolkosz D, editors. Wood-based energy in the northern forests: Springer; 2013. p. 69–99.

26. Sannigrahi P, Ragauskas AJ, Tuskan GA. Poplar as a feedstock for biofuels: A review of compositional characteristics. Biofuel Bioprod Biorefin. 2010; 4(2):209–226.

27. Stanton BJ, Haiby K, Gantz C, Espinoza J, Shuren RA. The economics of rapid multiplication of hybrid poplar biomass varieties. Forests. 2019; 10(5):446.

28. Riemenschneider DE, Berguson WE, Dickmann DI, Hall RB, Isebrands JG, Mohn CA, et al. Poplar breeding and testing strategies in the north-central U.S.: Demonstration of potential yield and consideration of future research needs. For Chron. 2001; 77(2):245–253.

29. Alves FC, Balmant KM, Resende MFR Jr, Kirst M, Campos Gdl. Accelerating forest tree breeding by integrating genomic selection and greenhouse phenotyping. Plant Genome. 2020; 13(3):1–13. https://doi.org/10.1002/tpg2.20048 PMID: 33217213

30. El-Sharkawy MA, Cock JH, Lynam JK, del Pilar Hernàndez A, Cadavid LFL. Relationships between biomass, root-yield and single-leaf photosynthesis in field-grown cassava. Field Crops Res. 1990; 25 (3–4):183–201.

31. Farquhar GD, Caemmerer Sv, Berry JA. A biochemical model of photosynthetic CO_2 assimilation in leaves of C_3 species. Planta. 1980; 149(1):78–90. https://doi.org/10.1007/BF00386231 PMID: 24306196

32. Wullschleger SD. Biochemical limitations to carbon assimilation in C_3 plants—A retrospective analysis of the A/C_i curves from 109 species. J Exp Bot. 1993; 44(5):907–920.

33. Ericsson T. Nutrient dynamics and requirements of forest crops. N Z J For Sci. 1994; 24(2/3):133–168.

34. Garten CT. Correlations between concentrations of elements in plants. Nature. 1976; 261:686–688.

35. Vitousek PM, Howarth RW. Nitrogen limitation on land and in the sea: how can it occur? Biogeochemistry. 1991; 13(2):87–115.

36. Saidana D, Braham M, Boujnah D, Mariem FB, Ammari S, El Hadj SB. Nutrient stress, ecophysiological, and metabolic aspects of olive tree cultivars. J Plant Nutr. 2009; 32:129–145.

37. Grassi G, Meir P, Cromer R, Tompkins D, Jarvis PG. Photosynthetic parameters in seedlings of *Eucalyptus grandis* as affected by rate of nitrogen supply. Plant Cell Environ. 2002; 25(12):1677–1688.

38. Reich PB, Walters MB, Ellsworth DS. Leaf age and season influence the relationships between leaf nitrogen, leaf mass per area and photosynthesis in maple and oak trees. Plant Cell Environ. 1991; 14 (3):251–259.

39. Zhao D, Reddy KR, Kakani VG, Reddy VR. Nitrogen deficiency effects on plant growth, leaf photosynthesis, and hyperspectral reflectance properties of sorghum. Eur J Agron. 2005; 22(4):391–403.

40. Mu X, Chen Y. The physiological response of photosynthesis to nitrogen deficiency. Plant Physiol Biochem. 2021; 158:76–82. https://doi.org/10.1016/j.plaphy.2020.11.019 PMID: 33296848

41. Zhao D, Reddy KR, Kakani VG, Read JJ, Carter GA. Corn (*Zea mays* L.) growth, leaf pigment concentration, photosynthesis and leaf hyperspectral reflectance properties as affected by nitrogen supply. Plant Soil. 2003; 257:205–218.

42. Knipling EB. Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation. Remote Sens Environ. 1970; 1(3):155–159.

43. Barnes ML, Breshears DD, Law DJ, Van Leeuwen WJ, Monson RK, Fojtik AC, et al. Beyond greenness: Detecting temporal changes in photosynthetic capacity with hyperspectral reflectance data. PLoS One. 2017; 12(12):e0189539. https://doi.org/10.1371/journal.pone.0189539 PMID: 29281709

44. Qian X, Zhang Y, Liu L, Du S. Exploring the potential of leaf reflectance spectra for retrieving the leaf maximum carboxylation rate. Int J Remote Sens. 2019; 40(14):5411–5428.

45. Vilfan N, van der Tol C, Verhoef W. Estimating photosynthetic capacity from leaf reflectance and Chl fluorescence by coupling radiative transfer to a model for photosynthesis. New Phytol. 2019; 223 (1):487–500. https://doi.org/10.1111/nph.15782 PMID: 30861144

46. Yoder BJ, Pettigrew-Crosby RE. Predicting nitrogen and chlorophyll content and concentrations from reflectance spectra (400–2500 nm) at leaf and canopy scales. Remote Sens Environ. 1995; 53 (3):199–211.

47. Ye X, Abe S, Zhang S. Estimation and mapping of nitrogen content in apple trees at leaf and canopy levels using hyperspectral imaging. Precis Agric. 2020; 21:198–225.

48. Wang S, Guan K, Wang Z, Ainsworth EA, Zheng T, Townsend PA, et al. Unique contributions of chlorophyll and nitrogen to predict crop photosynthetic capacity from leaf spectroscopy. J Exp Bot. 2021; 72(2):341–354. https://doi.org/10.1093/jxb/eraa432 PMID: 32937655

49. Jin J, Pratama BA, Wang Q. Tracing leaf photosynthetic parameters using hyperspectral indices in an alpine deciduous forest. Remote Sens. 2020; 12(7):1124.

50. Meacham-Hensold K, Montes CM, Wu J, Guan K, Fu P, Ainsworth EA, et al. High-throughput field phenotyping using hyperspectral reflectance and partial least squares regression (PLSR) reveals genetic modifications to photosynthetic capacity. Remote Sens Environ. 2019; 231(111176). https://doi.org/10.1016/j.rse.2019.04.029 PMID: 31534277

51. Wei F, Yan Z, Yongchao T, Weixing C, Xia Y, Yingxue L. Monitoring leaf nitrogen accumulation in wheat with hyper-spectral remote sensing. Acta Ecol Sin. 2008; 28(1):23–32.

52. Ramachandiran K, Pazhanivelan S. Determination of nitrogen and water stress with hyper spectral reflectance on maize using classification tree (CT) analysis. J Agrometeorol. 2015; 17(2):213–218.

53. Yao X, Huang Y, Shang G, Zhou C, Cheng T, Tian Y, et al. Evaluation of six algorithms to monitor wheat leaf nitrogen concentration. Remote Sens. 2015; 7(11):14939–14966.

54. He J, Zhang X, Guo W, Pan Y, Yao X, Cheng T, et al. Estimation of vertical leaf nitrogen distribution within a rice canopy based on hyperspectral data. Front Plant Sci. 2020; 10(1802). https://doi.org/10.3389/fpls.2019.01802 PMID: 32117352

55. Zhou Y, Ju W, Sun X, Hu Z, Han S, Black TA, et al. Close relationship between spectral vegetation indices and Vcmax in deciduous and mixed forests. Tellus B Chem Phys Meteorol. 2014; 66(23279).

56. Camino C, González-Dugo V, Hernández P, Sillero JC, Zarco-Tejada PJ. Improved nitrogen retrievals with airborne-derived fluorescence and plant traits quantified from VNIR-SWIR hyperspectral imagery in the context of precision agriculture. Int J Appl Earth Obs Geoinf. 2018; 70:105–117.

57. Sun J, Yang J, Shi S, Chen B, Du L, Gong W, et al. Estimating rice leaf nitrogen concentration: influence of regression algorithms based on passive and active leaf reflectance. Remote Sens. 2017; 9 (951).

58. Yi Q-X, Huang J-F, Wang F-M, Wang X-Z, Liu Z-Y. Monitoring rice nitrogen status using hyperspectral reflectance and artificial neural network. Environ Sci Technol. 2007; 41(19):6770–6775. https://doi.org/10.1021/es070144e PMID: 17969693

59. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning. Vol. 112. New York: Springer; 2013. p. 18.

60. Smith G. Step away from stepwise. J Big Data. 2018; 5(32).

61. Bratsch S, Epstein H, Buchhorn M, Walker D, Landes H. Relationships between hyperspectral data and components of vegetation biomass in Low Arctic tundra communities at Ivotuk, Alaska. Environ Res Lett. 2017; 12(025003):1–14.

62. Vergara-Diaz O, Vatter T, Kefauver SC, Obata T, Fernie AR, Araus JL. Assessing durum wheat ear and leaf metabolomes in the field through hyperspectral data. Plant J. 2020; 103(2). https://doi.org/10.1111/tpj.14636 PMID: 31808224

63. Galán RJ, Bernal-Vasquez A, Jebsen C, Piepho H, Thorwarth P, Steffan P, et al. Integration of genotypic, hyperspectral, and phenotypic data to improve biomass yield prediction in hybrid rye. Theor Appl Genet. 2020; 133(11):3001–3015. https://doi.org/10.1007/s00122-020-03651-8 PMID: 32681289

64. Farrugia J, Griffin S, Valdramidis VP, Camilleri K, Falzon O. Principal component analysis of hyperspectral data for early detection of mould in cheeselets. Curr Res Food Sci. 2021; 4:18–27. https://doi.org/10.1016/j.crfs.2020.12.003 PMID: 33554131

65. Rodarmel C, Shan J. Principal component analysis for hyperspectral image classification. Surv Land Inf Sci. 2002; 62(2):115–122.

66. Zhang H, Zhu L, Hu H, Zheng K, Jin Q. Monitoring leaf chlorophyll fluorescence with spectral reflectance in rice (*Oryza sativa* L.). Procedia Eng. 2011; 15:4403–4408.

67. Bunea F, She Y, Ombao H, Gongvatana A, Devlin K, Cohen R. Penalized least squares regression methods and applications to neuroimaging. Neuroimage. 2011; 55(4):1519–1527. https://doi.org/10.1016/j.neuroimage.2010.12.028 PMID: 21167288

68. Pope J, Powers D, Connell J, Jasemi M, Taylor D, Fafoutis X, editors. Supervised machine learning and feature selection for a document analysis application. Proceedings of the 9th International Conference on Pattern Recognition Applications and Methods; 2020: SCITEPRESS Digital Library.

69. Kotlar AM, Iversen BV, de Jong van Lier Q. Evaluation of parametric and nonparametric machine-learning techniques for prediction of saturated and near-saturated hydraulic conductivity. Vadose Zone J. 2019; 18(18014):1–13.

70. Singh K, Miska S, Ozbayoglu E, Alp Aydin B, editors. Using supervised machine learning algorithms to predict pressure drop in narrow annulus. Proceedings of SPE/AAPG Eastern Regional Meeting; 2018; Pittsburgh, PA, USA.

71. Mwangi B, Tian TS, Soares JC. A review of feature reduction techniques in neuroimaging. Neuroinformatics. 2014; 12(2):229–244. https://doi.org/10.1007/s12021-013-9204-3 PMID: 24013948

72. Rich DC, Livingston KM, Morgan SL. Evaluating performance of Lasso relative to PCA and LDA to classify dyes on fibers. Forensic Chem. 2020; 18(100213):1–7.

73. Faurina R, Winduratna B, Nugroho P, editors. Predicting stock movement using unidirectional LSTM and feature reduction: The case of an Indonesia stock. International Conference on Electrical Engineering and Computer Science (ICEECS); 2018; Bali, Indonesia.

74. Li Z, Ma X, Shen F, Lu H, Xia Y, Lu J. Evaluating treatment response to neoadjuvant chemoradiotherapy in rectal cancer using various MRI-based radiomics models. BMC Med Imaging. 2021; 21(30):1–10. https://doi.org/10.1186/s12880-021-00560-0 PMID: 33593304

75. Yendrek CR, Tomaz T, Montes CM, Cao Y, Morse AM, Brown PJ, et al. High-throughput phenotyping of maize leaf physiological and biochemical traits using hyperspectral reflectance. Plant Physiol. 2017; 173(1):614–626. https://doi.org/10.1104/pp.16.01447 PMID: 28049858

76. Dietze MC. Gaps in knowledge and data driving uncertainty in models of photosynthesis. Photosynth Res. 2014; 119(1–2):3–14. https://doi.org/10.1007/s11120-013-9836-z PMID: 23645396

77. Zaehle S, Sitch S, Smith B, Hatterman F. Effects of parameter uncertainties on the modeling of terrestrial biosphere dynamics. Global Biogeochem Cycles. 2005; 19(3):1–16.

78. Verheijen LM, Brovkin V, Aerts R, Bönisch G, Cornelissen JHC, Kattge J, et al. Impacts of trait variation through observed trait–climate relationships on performance of an Earth system model: a conceptual analysis. Biogeosciences. 2013; 10(8):5497–5515.

79. Bonan GB, Lawrence PJ, Oleson KW, Levis S, Jung M, Reichstein M, et al. Improving canopy processes in the Community Land Model version 4 (CLM4) using global flux fields empirically inferred from FLUXNET data. J Geophys Res Biogeosci. 2011; 116(G2).

80. Croft H, Chen JM, Luo X, Bartlett P, Chen B, Staebler RM. Leaf chlorophyll content as a proxy for leaf photosynthetic capacity. Glob Chang Biol. 2017; 23(9):3513–3524. https://doi.org/10.1111/gcb.13599 PMID: 27976452

81. Rogers A. The use and misuse of $V_{c,max}$ in Earth System Models. Photosynth Res. 2014; 119:15–29. https://doi.org/10.1007/s11120-013-9818-1 PMID: 23564478

82. Schaefer K, Schwalm CR, Williams C, Arain MA, Barr A, Chen JM, et al. A model-data comparison of gross primary productivity: Results from the North American Carbon Program site synthesis. J Geophys Res Biogeosci. 2012; 117(G03010).

83. Natural Resources Conservation Service, United States Department of Agriculture [database on the Internet]. Web Soil Survey. [cited 28 September 2020]. https://websoilsurvey.nrcs.usda.gov/app/WebSoilSurvey.aspx.

84. Arguez A, Durre I, Applequist S, Vose RS, Squires MF, Yin X, et al. NOAA's 1981–2010 U.S. climate normals: An overview. Bull Am Meteorol Soc. 2012; 93(11):1687–1697. Available from https://www.ncdc.noaa.gov/cdo-web/datatools/normals.

85. Hornslein NJ, Siegert C, Renninger HJ. Physiological response of mid-canopy sweetgum trees to overstory loblolly pine mortality. Trees. 2019; 33:139–151.

86. Kassahun Z, Yow JN, Renninger HJ. Diversity or redundancy in leaf physiological and anatomical parameters in a species diverse, bottomland hardwood forest? Forests. 2020; 11(5):519.

87. Renninger H, Durbin T, Gentry A, Kassahun Z. Relationships between leaf anatomy and physiological functioning of Southern US oak species differing in flood tolerance. Forests. 2020; 11(1):73.

88. Sharkey TD, Bernacchi CJ, Farquhar GD, Singsaas EL. Fitting photosynthetic carbon dioxide response curves for C_3 leaves. Plant Cell Environ. 2007; 30(9):1035–1040. https://doi.org/10.1111/j.1365-3040.2007.01710.x PMID: 17661745

89. R Core Team. R: A language and environment for statistical computing (Version 4.0.4). R Foundation for Statistical Computing, Vienna, Australia. https://www.R-project.org.

90. Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc Series B Stat Methodol. 1996; 58(1):267–288.

91. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw. 2010; 33(1):1–22. PMID: 20808728

92. Kuhn M. caret: Classification and Regression Training. R package version 6.0–86. 2020; http://CRAN.R-project.org/package=caret.

93. Moreno-Torres JG, Sáez JA, Herrera F. Study on the impact of partition-induced dataset shift on *k*-fold cross-validation. IEEE Trans Neural Netw Learn Syst. 2012; 23(8):1304–1312. https://doi.org/10.1109/TNNLS.2012.2199516 PMID: 24807526

94. Kaiser HF. The application of electronic computers to factor analysis. Educ Psychol Meas. 1960; 20(1):141–151.

95. Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Stat Soc Series B Stat Methodol. 2005; 67(2):301–320.

96. Kumar B, Dikshit O, Gupta A, Singh MK. Feature extraction for hyperspectral image classification: A review. Int J Remote Sens. 2020; 41(16):6248–6287.

97. Wong M, Hung C, editors. Dimensionality reduction with weighted K-Means for hyperspectral image classification. Proceedings of IEEE International Geoscience and Remote Sensing Symposium; 2020: IEEE.

98. Raper TB, Varco JJ. Canopy-scale wavelength and vegetative index sensitivities to cotton growth parameters and nitrogen status. Precis Agric. 2015; 16(1):62–76.

99. Datt B. Remote sensing of chlorophyll a, chlorophyll b, chlorophyll a+ b, and total carotenoid content in eucalyptus leaves. Remote Sens Environ. 1998; 66(2):111–121.

100. Gitelson AA, Buschmann C, Lichtenthaler HK. The chlorophyll fluorescence ratio F735/F700 as an accurate measure of the chlorophyll content in plants. Remote Sens Environ. 1999; 69(3):296–302.

101. Le Maire G, Francois C, Dufrene E. Towards universal broad leaf chlorophyll indices using PROSPECT simulated database and hyperspectral reflectance measurements. Remote Sens Environ. 2004; 89(1):1–28.

102. Maccioni A, Agati G, Mazzinghi P. New vegetation indices for remote measurement of chlorophylls based on leaf directional reflectance spectra. J Photochem Photobiol B. 2001; 61(1–2):52–61. https://doi.org/10.1016/s1011-1344(01)00145-2 PMID: 11485848

103. Gupta RK, Vijayan D, Prasad TS. Comparative analysis of red-edge hyperspectral indices. Adv Space Res. 2003; 32(11):2217–2222.

104. Carter GA. Ratios of leaf reflectances in narrow wavebands as indicators of plant stress. Int J Remote Sens. 1994; 15(3):697–703.

105. Gamon JA, Serrano L, Surfus JS. The photochemical reflectance index: an optical indicator of photosynthetic radiation use efficiency across species, functional types, and nutrient levels. Oecologia. 1997; 112(4):492–501. https://doi.org/10.1007/s004420050337 PMID: 28307626

106. Gitelson A, Merzlyak MN. Quantitative estimation of chlorophyll-a using reflectance spectra: Experiments with autumn chestnut and maple leaves. J Photochem Photobiol B. 1994a; 22(3):247–252.

107. Gitelson A, Merzlyak MN. Spectral reflectance changes associated with autumn senescence of *Aesculus hippocastanum* L. and *Acer platanoides* L. leaves. Spectral features and relation to chlorophyll estimation. J Plant Physiol. 1994b; 143(3):286–292.

108. Gitelson AA, Merzlyak MN. Remote estimation of chlorophyll content in higher plant leaves. Int J Remote Sens. 1997; 18(12):2691–2697.

109. Gitelson AA, Kaufman YJ, Merzlyak MN. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. Remote Sens Environ. 1996; 58(3):289–298.

110. Penuelas J, Baret F, Filella I. Semi-empirical indices to assess carotenoids/chlorophyll a ratio from leaf spectral reflectance. Photosynthetica. 1995; 31(2):221–230.

111. Sims DA, Gamon JA. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. Remote Sens Environ. 2002; 81(2–3):337–354.

112. Stimson HC, Breshears DD, Ustin SL, Kefauver SC. Spectral sensing of foliar water conditions in two co-occurring conifer species: *Pinus edulis* and *Juniperus monosperma*. Remote Sens Environ. 2005; 96(1):108–118.

113. Vogelmann J, Rock B, Moss D. Red edge spectral measurements from sugar maple leaves. Int J Remote Sens. 1993; 14(8):1563–1575.

114. Wen P-F, Wang R, Shi Z-J, Ning F, Wang S-L, Zhang Y-J, et al. Effects of N application rate on N remobilization and accumulation in maize (*Zea mays* L.) and estimating of vegetative N remobilization using hyperspectral measurements. Comput Electron Agric. 2018; 152:166–181.

115. Zarco-Tejada PJ, Miller JR, Noland TL, Mohammed GH, Sampson PH. Scaling-up and model inversion methods with narrowband optical indices for chlorophyll content estimation in closed forest canopies with hyperspectral data. IEEE Trans Geosci Remote Sens. 2001; 39(7):1491–1507.

116. Leuning R. Scaling to a common temperature improves the correlation between the photosynthesis parameters J_{max} and V_{cmax}. J Exp Bot. 1997; 48(2):345–347.

117. Rascher U, Agati G, Alonso L, Cecchi G, Champagne S, Colombo R, et al. CEFLES2: the remote sensing component to quantify photosynthetic efficiency from the leaf to the region by measuring sun-induced fluorescence in the oxygen absorption bands. Biogeosciences. 2009; 6(7):1181–1198.

118. RayChaudhuri B. Remote sensing of solar-induced chlorophyll fluorescence at atmospheric oxygen absorption band around 760 nm and simulation of that absorption in laboratory. IEEE Trans Geosci Remote Sens. 2012; 50(10):3908–3914.

119. Calderón R, Navas-Cortés JA, Lucena C, Zarco-Tejada PJ. High-resolution airborne hyperspectral and thermal imagery for early detection of *Verticillium* wilt of olive using fluorescence, temperature and narrow-band spectral indices. Remote Sens Environ. 2013; 139:231–245.

120. Flexas J, Briantais J, Cerovic Z, Medrano H, Moya I. Steady-state and maximum chlorophyll fluorescence responses to water stress in grapevine leaves: A new remote sensing system. Remote Sens Environ. 2000; 73(3):283–297.

121. Flexas J, Escalona J, Medrano H. Water stress induces different levels of photosynthesis and electron transport rate regulation in grapevines. Plant Cell Environ. 1999; 22:39–48.

122. Wang H, Gu M, Cui J, Shi K, Zhou Y, Yu J. Effects of light quality on CO_2 assimilation, chlorophyll-fluorescence quenching, expression of Calvin cycle genes and carbohydrate accumulation in *Cucumis sativus*. J Photochem Photobiol B. 2009; 96(1):30–37. https://doi.org/10.1016/j.jphotobiol.2009.03.010 PMID: 19410482

123. Frankenberg C, O'Dell C, Berry J, Guanter L, Joiner J, Köhler P, et al. Prospects for chlorophyll fluorescence remote sensing from the Orbiting Carbon Observatory-2. Remote Sens Environ. 2014; 147:1–12.

124. Zhang Y, Guanter L, Berry JA, Joiner J, van der Tol C, Huete A, et al. Estimation of vegetation photosynthetic capacity from space-based measurements of chlorophyll fluorescence for terrestrial biosphere models. Glob Chang Biol. 2014; 20(12):3727–3742. https://doi.org/10.1111/gcb.12664 PMID: 24953485

125. He L, Chen JM, Liu J, Zheng T, Wang R, Joiner J, et al. Diverse photosynthetic capacity of global ecosystems mapped by satellite chlorophyll fluorescence measurements. Remote Sens Environ. 2019; 232:111344. https://doi.org/10.1016/j.rse.2019.111344 PMID: 33149371

126. Koffi EN, Rayner PJ, Norton AJ, Frankenberg C, Scholze M. Investigating the usefulness of satellite-derived fluorescence data in inferring gross primary productivity within the carbon cycle data assimilation system. Biogeosciences. 2015; 12(13):4067–4084.

127. Wingler A, Quick WP, Bungard RA, Bailey KJ, Lea PJ, Leegood RC. The role of photorespiration during drought stress: an analysis utilizing barley mutants with reduced activities of photorespiratory enzymes. Plant Cell Environ. 1999; 22(4):361–373.

128. Bremer DJ, Lee H, Su K, Keeley SJ. Relationships between normalized difference vegetation index and visual quality in cool-season turfgrass: II. Factors affecting NDVI and its component reflectances. Crop Sci. 2011; 51(5):2219–2227.

129. Guertal EA, Shaw JN. Multispectral radiometer signatures for stress evaluation in compacted bermudagrass turf. HortScience. 2004; 39(2):403–407.

130. Merewitz E, Meyer W, Bonos S, Huang B. Drought stress responses and recovery of Texas× Kentucky hybrids and Kentucky bluegrass genotypes in temperate climate conditions. Agron J. 2010; 102 (1):258–268.

131. Lee G, Carrow RN, Duncan RR. Photosynthetic responses to salinity stress of halophytic seashore paspalum ecotypes. Plant Sci. 2004; 166(6):1417–1425.

132. Wang C, Feng M, Yang W, Ding G, Xiao L, Li G, et al. Extraction of sensitive bands for monitoring the winter wheat (*Triticum aestivum*) growth status and yields based on the spectral reflectance. PLoS One. 2017; 12(1):e0167679. https://doi.org/10.1371/journal.pone.0167679 PMID: 28060827

133. Lombard K, O'Neill M, Mexal J, Ulery A, Onken B, Bettmann G, et al. Can soil plant analysis development values predict chlorophyll and total Fe in hybrid poplar? Agrofor Syst. 2010; 78:1–11.

134. Zhang H, Hu H, Zhang X, Wang K, Song T, Zeng F. Detecting *Suaeda salsa* L. chlorophyll fluorescence response to salinity stress by using hyperspectral reflectance. Acta Physiol Plant. 2012; 34 (2):581–588.

135. Naumann JC, Young DR, Anderson JE. Linking leaf chlorophyll fluorescence properties to physiological responses for detection of salt and drought stress in coastal plant species. Physiol Plant. 2007; 131 (3):422–433. https://doi.org/10.1111/j.1399-3054.2007.00973.x PMID: 18251881

136. Naumann JC, Young DR, Anderson JE. Leaf chlorophyll fluorescence, reflectance, and physiological response to freshwater and saltwater flooding in the evergreen shrub, *Myrica cerifera*. Environ Exp Bot. 2008; 63(1–3):402–409.

137. Krause GH, Weis E. Chlorophyll fluorescence and photosynthesis: the basics. Annu Rev Plant Physiol Plant Mol Biol. 1991; 42:313–349.

138. Lisker IS, Radchenko NS, Radchenko SS. Dependence of the absorption coefficient of 935-nm radiation on water content of plant leaves. Russ Agric Sci. 2008; 34(3):160–161.

139. Huang B, Fry J, Wang B. Water relations and canopy characteristics of tall fescue cultivars during and after drought stress. HortScience. 1998; 33(5):837–840.

140. Guanter L, Frankenberg C, Dudhia A, Lewis PE, Gómez-Dans J, Kuze A, et al. Retrieval and global assessment of terrestrial chlorophyll fluorescence from GOSAT space measurements. Remote Sens Environ. 2012; 121:236–251.

141. Li X, Xiao J, He B. Chlorophyll fluorescence observed by OCO-2 is strongly related to gross primary productivity estimated from flux towers in temperate forests. Remote Sens Environ. 2018; 204:659–671.

142. Wang X, Chen JM, Ju W. Photochemical reflectance index (PRI) can be used to improve the relationship between gross primary productivity (GPP) and sun-induced chlorophyll fluorescence (SIF). Remote Sens Environ. 2020; 246(111888):1–17.

143. Li X, Xiao J, He B, Altaf Arain M, Beringer J, Desai AR, et al. Solar-induced chlorophyll fluorescence is strongly correlated with terrestrial photosynthesis for a wide variety of biomes: First global analysis based on OCO-2 and flux tower observations. Glob Chang Biol. 2018; 24(9):3990–4008. https://doi.org/10.1111/gcb.14297 PMID: 29733483

144. Sun Y, Frankenberg C, Jung M, Joiner J, Guanter L, Köhler P, et al. Overview of Solar-Induced chlorophyll Fluorescence (SIF) from the Orbiting Carbon Observatory-2: Retrieval, cross-mission comparison, and global monitoring for GPP. Remote Sens Environ. 2018; 209:808–823.

145. Wood JD, Griffis TJ, Baker JM, Frankenberg C, Verma M, Yuen K. Multiscale analyses of solar-induced florescence and gross primary production. Geophys Res Lett. 2017; 44(1):533–541.

146. Zhang YJ, Hou MY, Xue HY, Liu LT, Sun HC, Li CD, et al. Photochemical reflectance index and solar-induced fluorescence for assessing cotton photosynthesis under water-deficit stress. Biol Plant. 2018; 62(4):817–825.

147. Amoros-Lopez J, Gomez-Chova L, Vila-Frances J, Alonso L, Calpe J, Moreno J, et al. Evaluation of remote sensing of vegetation fluorescence by the analysis of diurnal cycles. Int J Remote Sens. 2008; 29(17–18):5423–5436.

148. Verrelst J, van der Tol C, Magnani F, Sabater N, Rivera JP, Mohammed G, et al. Evaluating the predictive power of sun-induced chlorophyll fluorescence to estimate net photosynthesis of vegetation canopies: A SCOPE modeling study. Remote Sens Environ. 2016; 176:139–151.

149. Amir M, Chen J, Chen B, Wang S, Zhu K, Li Y, et al. Reflectance and chlorophyll fluorescence-based retrieval of photosynthetic parameters improves the estimation of subtropical forest productivity. Ecol Indic. 2021; 131(108133):1–12.

150. Kochubey SM, Samokhval EG. Long-wavelength chlorophyll forms in Photosystem I from pea thylakoids. Photosynth Res. 2000; 63(3):281–290. https://doi.org/10.1023/A:1006482618292 PMID: 16228438

151. Zhen S, van Iersel M, Bugbee B. Why far-red photons should be included in the definition of photosynthetic photons and the measurement of horticultural fixture efficacy. Front Plant Sci. 2021; 12 (693445):1–4. https://doi.org/10.3389/fpls.2021.693445 PMID: 34239530

152. Zhen S, Bugbee B. Far-red photons have equivalent efficiency to traditional photosynthetic photons: Implications for redefining photosynthetically active radiation. Plant Cell Environ. 2020; 43(5):1259–1272. https://doi.org/10.1111/pce.13730 PMID: 31990071

153. Sonobe R, Miura Y, Sano T, Horie H. Monitoring photosynthetic pigments of shade-grown tea from hyperspectral reflectance. Can J Remote Sens. 2018; 44(2):104–112.

154. Jiao Y, Lau OS, Deng XW. Light-regulated transcriptional networks in higher plants. Nat Rev Genet. 2007; 8(3):217–230. https://doi.org/10.1038/nrg2049 PMID: 17304247

155. Rozema J, van de Staaij J, Björn LO, Caldwell M. UV-B as an environmental factor in plant life: stress and regulation. Trends Ecol Evol. 1997; 12(1):22–28. https://doi.org/10.1016/s0169-5347(96)10062-8 PMID: 21237957

156. Brosché M, Strid Å. Molecular events following perception of ultraviolet-B radiation by plants. Physiol Plant. 2003; 117:1–10.

157. Krause GH, Gallé A, Gademann R, Winter K. Capacity of protection against ultraviolet radiation in sun and shade leaves of tropical forest plants. Funct Plant Biol. 2003; 30(5):533–542. https://doi.org/10.1071/FP03047 PMID: 32689038

158. Li JW, Zhang JX, Zhao Z, Lei XD, Xu XL, Lu XX, et al. Use of fluorescence-based sensors to determine the nitrogen status of paddy rice. J Agric Sci. 2013; 151(6):862–871.

159. Abulaiti Y, Sawut M, Maimaitiaili B, Chunyue M. A possible fractional order derivative and optimized spectral indices for assessing total nitrogen content in cotton. Comput Electron Agric. 2020; 171:105275.

160. Min M, Lee WS, Kim YH, Bucklin RA. Nondestructive detection of nitrogen in Chinese cabbage leaves using VIS–NIR spectroscopy. HortScience. 2006; 41(1):162–166.

161. Luther JE, Carroll AL. Development of an index of balsam fir vigor by foliar spectral reflectance. Remote Sens Environ. 1999; 69(3):241–252.

162. Zhao C, Wang Z, Wang J, Huang W. Relationships of leaf nitrogen concentration and canopy nitrogen density with spectral features parameters and narrow-band spectral indices calculated from field winter wheat (*Triticum aestivum* L.) spectra. Int J Remote Sens. 2012; 33(11):3472–3491.

163. Tarpley L, Reddy KR, Sassenrath-Cole GF. Reflectance indices with precision and accuracy in predicting cotton leaf nitrogen concentration. Crop Sci. 2000; 40(6):1814–1819.

164. Yu K-Q, Zhao Y-R, Li X-L, Shao Y-N, Liu F, He Y. Hyperspectral imaging for mapping of total nitrogen spatial distribution in pepper plant. PLoS One. 2014; 9(12):e116205. https://doi.org/10.1371/journal.pone.0116205 PMID: 25549353

165. Curran PJ. Remote sensing of foliar chemistry. Remote Sens Environ. 1989; 30(3):271–278.

166. Mitchell JJ, Glenn NF, Sankey TT, Derryberry DR, Germino MJ. Remote sensing of sagebrush canopy nitrogen. Remote Sens Environ. 2012; 124:217–223.

167. Ramoelo A, Skidmore AK, Cho MA, Mathieu R, Heitkönig IMA, Dudeni-Tlhone N, et al. Non-linear partial least square regression increases the estimation accuracy of grass nitrogen and phosphorus using *in situ* hyperspectral and environmental data. ISPRS J Photogramm Remote Sens. 2013; 82:27–40.