Received: 6 December 2024 | Accepted: 23 April 2025

DOI: 10.1111/2041-210X.70100

Methods in Ecology and Evolution

RESEARCH ARTICLE

Spectral wavelength range influences the performance of chemometric models estimating various foliar functional traits

Minjee Park[1,2] | Lorenzo Cotrozzi[3] | Geoffrey M. Williams[1] | Matthew D. Ginzel[1,4] | Michael V. Mickelbart[2,5,6] | Douglass F. Jacobs[1] | John J. Couture[1,2,4]

[1]Department of Forestry and Natural Resources, Purdue University, West Lafayette, Indiana, USA; [2]Center for Plant Biology, Purdue University, West Lafayette, Indiana, USA; [3]Department of Agriculture, Food, and Environment, University of Pisa, Pisa, Italy; [4]Department of Entomology, Purdue University, West Lafayette, Indiana, USA; [5]Department of Botany and Plant Pathology, Purdue University, West Lafayette, Indiana, USA and [6]Department of Horticulture and Landscape Architecture, Purdue University, West Lafayette, Indiana, USA

Correspondence
John J. Couture
Email: couture@purdue.edu

Present address
Geoffrey M. Williams, Department of Forestry, Michigan State University, East Lansing, Michigan, USA

Funding information
Hardwood Tree Improvement and Regeneration Center, Purdue University; National Science Foundation, Grant/Award Number: 1916587; USDA, Grant/Award Number: IND011490

Handling Editor: Jessica Royles

Abstract

1. Hyperspectral reflectance can potentially be used to non-destructively estimate a diverse suite of plant physiochemical functional traits by applying chemometric approaches to leverage absorption features related to chemical compounds and physiological processes associated with these traits. This approach has considerable implications in advancing plant physiological and chemical ecology. For complex functional traits, however, there is a lack of well-defined absorption features and features may be unevenly distributed across the reflectance spectrum, suggesting that the influence of wavelength ranges on the performance of chemometric models is potentially important for accurately estimating foliar functional traits.
2. Here, we investigate the influence of spectral ranges on the performance of models estimating six tree functional traits: CO_2 assimilation rate, specific leaf area, leaf water content and concentrations of foliar nitrogen, sugars and gallic acid. Using data collected from multiple different experiments, we quantified plant functional trait responses using standard reference measurements and paired them with proximal leaf-level hyperspectral reflectance measurements spanning the wavelength range of 400–2400 nm. A total of 100 different wavelength range combinations were evaluated using partial least squares regression to determine the influence of wavelength range on model performance.
3. We found that the influence of starting or ending wavelength on model performance was trait specific and better model outcomes were achieved when the starting and ending wavelengths encompassed absorption features associated with the specific leaf trait modelled. Interestingly, we found that including shortwave-infrared wavelength ranges (1300–2500 nm) improved performance for all trait models.

4. Collectively, our findings underscore the importance of optimal spectral range selection in enhancing the accuracy of chemometric models for specific foliar trait estimates. An emergent outcome of this work is that the approach can be used to (1) identify the important spectral features of traits that currently lack known absorption features or have multiple or weak absorption features, (2) expand the current suite of plant functional traits that can be estimated using spectroscopy and (3) ultimately advance the integration of a spectral biology approach in ecological research.

KEYWORDS
black walnut, hyperspectral, Juglans nigra, leaf functional traits, northern red oak, PLSR, Quercus rubra, wavelength range

1 | INTRODUCTION

Vegetation spectroscopy has emerged as a powerful tool for assessing plant health, especially considering its potential for estimating plant functional trait responses to biotic and abiotic stress (Asner et al., 2015; Lausch et al., 2016; Serbin & Townsend, 2020). This approach has been used to rapidly and non-destructively classify various stress events and assess stress severity in plants, and it can be applied across large spatial and temporal scales using aerial and spaceborne measurement platforms (Hill et al., 2019; Huang et al., 2019; Sapes et al., 2024; Thomas et al., 2018). Vegetation spectroscopy can be used to monitor changes in plants caused by environmental variation and pest and pathogen pressure (Asner et al., 2018; Cotrozzi, 2022; Garcia et al., 2025; Lassalle, 2021) and has shown promise to advance trait-based approaches for understanding plant ecology (Cavender-Bares et al., 2016; Couture et al., 2015; DeLaMater et al., 2021; Struckman et al., 2019; Wang et al., 2022). While relationships among plant spectral data and physiochemical traits have enabled monitoring of specific responses of plants to different stressors (Lausch et al., 2016), it is not yet well understood how spectral range contributes to the accuracy of estimates of many complex physiochemical traits.

Plant functional traits encompass a wide variety of plant characteristics, including chemical composition, physiological processes, morphology, anatomical structure and phenological changes (Violle et al., 2007). Spectral estimation of these traits relies on the absorption and reflectance patterns of vegetation within and across particular wavelength ranges (Curran, 1989). Estimating specific chemical compounds, such as chlorophyll, may rely on fewer, more well-understood absorption features (Curran, 1989), while physiological processes, such as photosynthesis, are a product of a combination of a number of functional traits and may require a more complex set of absorption features related to the underlying mechanisms to estimate the emergent physiological process (Doughty et al., 2011; Sexton et al., 2021; Wu et al., 2024). In addition, when absorption features for traits are neither well-defined nor singular in nature, estimating traits from spectral data can be challenging, and different wavelength or wavelength range combinations may be needed for accurate estimation.

A preferred approach to estimating foliar functional traits in chemometric modelling is partial least squares regression (PLSR; Wold et al., 1984, 2001). This method is an empirical multivariate statistical approach that builds a predictive model by relating variability in predictor variables to variability in response variables (Burnett et al., 2021; Wold et al., 1984, 2001). In contrast to other linear regression approaches, which can generate spurious coefficients when predictor variables are highly correlated, as is the case with spectral data, PLSR reduces the predictor data matrix into few, relatively uncorrelated latent variables (Burnett et al., 2021). Because PLSR decomposes spectral data into latent variables to extract information from spectral profiles, the approach can estimate plant functional traits even in cases where complex traits lack clear, specific absorption features or rely on absorption features from different subcomponent traits that define the functioning of an emergent complex trait using the input spectral data (Kothari & Schweiger, 2022). Additionally, PLSR can indirectly infer traits by leveraging the covariance among other traits that do have clear absorption features (Chadwick & Asner, 2016; Chen et al., 2022; Nunes et al., 2017). Because of these characteristics, however, the estimation accuracy of PLSR is not inherently reliant on known absorption features, and thus it is important to identify optimal input wavelength ranges.

Currently, there is a lack of well-defined absorption features for many complex or less commonly studied plant functional traits. In addition, features for these traits may be unevenly distributed across the reflectance spectrum, and the wavelength ranges used can influence how PLSR decomposes the spectral matrix to produce latent variables, ultimately affecting prediction outcomes. Because of these knowledge gaps, research into identifying optimal wavelength ranges for plant functional trait predictions is crucial for improving chemometric outcomes. Previous studies have employed different approaches for selecting wavelength ranges for PLSR analysis. Some studies have used a 'full wavelength range' (i.e. 350–2500 nm, or a close variation to this range) for all plant functional traits (Dao et al., 2025; Ely

et al., 2019; Kothari, Beauchamp-Rioux, Blanchard, et al., 2023; Nakaji et al., 2019), while others have focused on smaller wavelength regions based on historically known absorption features of specific traits (Calzone et al., 2021; Cotrozzi, Peron, et al., 2020; Couture et al., 2016; Dechant et al., 2017; Nunes et al., 2017; Serbin et al., 2014; Xie et al., 2024). In addition, studies have explored various combinations of the visible (VIS), near infrared (NIR) and shortwave infrared (SWIR) ranges, selecting wavelength ranges that yield the best model outcomes (Chen et al., 2022; Cotrozzi, Peron, et al., 2020; Kothari, Beauchamp-Rioux, Laliberté, et al., 2023; Sexton et al., 2021). Some recent studies have aimed to explore the optimal wavelength range for trait estimations by first dividing spectral features into evenly sized, small intervals and then combining these intervals to identify spectral regions that maximized performance in PLSR models (Wan et al., 2022; Wang et al., 2023). Despite acknowledging the general influence of wavelength selection on the performance of models estimating plant functional traits, the influence of spectral features for individual leaf traits on model performance, particularly in terms of model accuracy over multiple different wavelength ranges, is not fully understood.

In this study, we aim to fill this knowledge gap by exploring the influence of different wavelength ranges on chemometric model performance for predicting foliar functional traits. We first quantified a suite of plant foliar physiological, anatomical and chemical traits that reflect plant growth, defence and stress responses including net CO_2 assimilation rate (*A*), specific leaf area (SLA), leaf water content (LWC) and nitrogen (N), sugars and gallic acid (GA) concentrations. Assimilation rates, SLA and foliar N and sugars concentrations are often implicated in having an impact on multiple ecosystem processes such as plant growth and leaf litter decomposition (Asner et al., 2017; Cui et al., 2020; Kazakou et al., 2009; Wright et al., 2004). Leaf water content is frequently used to assess plant water status (Cheng et al., 2011) and phenolic acids, including GA, contribute to plant defences by acting as antioxidants and exhibiting antimicrobial activity (Kumar et al., 2020). These traits capture essential plant function and together they provide insights into how plant growth and stress responses adapt to varying environmental conditions. We then examined the influence of multiple different combinations of wavelength ranges on the ability to estimate foliar functional traits from leaf spectral data using PLSR.

This study explores how the accuracy of PLSR models varies depending on the wavelength ranges chosen within the full spectrum (400–2400 nm) due to the varying presence of relevant absorption features for the foliar trait of interest in the different subsets of the spectrum. Our specific objectives were to (1) evaluate optimal spectral regions to estimate variation in plant functional traits, (2) investigate the accuracy of these models across different wavelength ranges selected and (3) assess the extent to which changes in model prediction accuracy across wavelength regions are related to known absorption features of specific plant functional traits.

2 | MATERIALS AND METHODS

2.1 | Data collections

Data used in this study were collected from several different experiments over a 2-year period (2018–2019) examining responses of two tree species, black walnut (*Juglans nigra*) and red oak (*Quercus rubra*), to multiple abiotic and biotic stress factors, alone and in combination, in a controlled environment. Abiotic stress environments included water, salt and nutrient stress conditions and biotic stress environments included fungal pathogen inoculations (Table 1). Detailed descriptions of experimental design and conditions can be found in Appendix S1.

2.2 | Foliar spectral and functional trait collections

Spectral data used in this study were collected individually from all living, attached leaves that were selected for physiochemical reference measurements in studies described in Appendix S1. Gas-exchange measurements were conducted on leaves of red oak or leaflets of black walnut from the top fully expanded leaf per tree to determine net CO_2 assimilation rate (*A*). Specific leaf area (SLA), leaf water content (LWC), foliar nitrogen (N), sugars and gallic acid (GA) concentrations were measured from a single leaf or leaflet, for SLA and LWC, or four to six leaves or leaflets, for N, sugars and GA, that were collected and then transported in a cooler on ice to the laboratory.

2.2.1 | Foliar spectral collections

In all experiments, full-range (350–2500 nm) reflectance profiles of all leaves or leaflets sampled for reference measurements were collected using a SVC HR-1024i spectroradiometer (Spectral Vista Corporation, Poughkeepsie NY, USA) including a fibre optic cable connected to a plant probe, which was equipped with a leaf clip containing an internal halogen light source. Leaf spectral data were collected in the middle of the adaxial side of each leaf or leaflet immediately following gas-exchange measurements and before tissue was collected for determining foliar anatomical, water-related or chemical traits. Integration time (i.e. the length of time that the detectors are allowed to collect photons before passing the accumulated charge to the converter for processing) was set at 2 s. The relative reflectance of each leaf was determined from the measurement of radiance collected from the leaf or leaflet divided by the radiance collected from a white reference panel internal to the leaf clip that was measured every six spectral collections. Spectral measurements were resampled to a single nanometre resolution and spectral and physiochemical reference measurements were then paired. In the case of spectral measurements for chemistry (N, sugars and GA), the spectra from the four or six leaves or leaflets collected were combined by calculating the mean

TABLE 1 Four experimental designs.

Experiment	Tree species	Stress	Treatment
2018 EEL	Black walnut	Fungal disease	Control: Deionized water Pathogen infection: *Geosmithia morbida* inoculation
		Soil type	Forest soil, plantation soil, sterile soil
2018 Wright Center	Black walnut	Nutrient availability	Control: Fertigation Nutrient deficiency: Tap water
		Water availability	Control: Full irrigation Drought: 50 mL day^{-1} irrigation
2019 EEL	Black walnut	Fungal disease	Control: Deionized water Fungal infection: *G. morbida* or *Fusarium solani* inoculation
		Water availability	Control: Full irrigation Drought: No irrigation
2019 Wright Center	Black walnut Red oak	Nutrient availability	Control: Fertigation Nutrient deficiency: Tap water
		Salinity	Control: Fertigation Salt deposition: salt solution (NaCl, 50 mM)

Note: Experiments were conducted in 2018 and 2019 to expose one-year-old black walnut (*Juglans nigra* L.) and northern red oak (*Quercus rubra* L.) seedlings to various biotic and abiotic stressors in two greenhouse environments: The Purdue Entomology Environmental Lab (EEL) at Purdue University, West Lafayette, IN, USA (40°25′23″ N, 86°54′52″ W) and the John S. Wright Forestry Center (Wright Center) at the Purdue Martell Experimental Forest, West Lafayette, IN, USA (40°25′56″ N, 87°02′19″ W). Details of the stress treatments are provided in Appendix S1.

at each wavelength to produce an average spectrum. Wavelength regions 350–399, 2401–2500 and 1891–1911 nm were removed because of noise associated with detector limits (i.e. 350–399 and 2401–2500) or noise associated with the SWIR 1 and SWIR 2 detector overlap region (i.e. 1891–1911).

2.2.2 | Foliar functional trait collections

Net CO_2 assimilation rate was determined using an LI-6400XT Portable Photosynthesis System (LI-COR Biosciences, Lincoln, NE, USA) with a light source of a 6400-02B LED, operating at 400 μL L^{-1} CO_2 concentration and saturating light conditions (1700 μmol m^{-2} s^{-1} photosynthetically active radiation). Specific leaf area and LWC were determined by measuring the fresh weight (FW) of a leaf or leaflet and then oven-drying the leaf or leaflet at 60°C until a constant mass to obtain the dry weight (DW). The petiole of leaf or petiolule of leaflet was removed, and the leaf or leaflet was scanned before drying, and the area of leaf or leaflet was determined using the software ImageJ v.1.38 (National Institutes of Health). Specific leaf area was calculated as leaf area divided by DW and LWC was calculated as [(FW − DW)/FW] × 100.

Additional foliar tissue collected for chemical quantification was immediately wrapped in aluminium foil and stored in liquid nitrogen until being transferred to a −20°C freezer. Foliar samples were lyophilized and ground to fine powder using a ball mill. Leaf N concentration was determined using a combustion analyser (Costech Analytical Technologies Inc., Valencia, CA) with atropine (CE Elantech, Lakewood, NJ, USA) serving as a standard. Sugars, determined as the sum of glucose, fructose and sucrose, were quantified according to Pellegrini et al. (2015) with minor modifications. Briefly, 50 mg of freeze-dried and ground leaf material was homogenized in 1 mL of HPLC-grade water and placed in a water bath at 60°C for 1 h. The extracts were then centrifuged at 5000 *g* for 20 min at room temperature. The supernatant was filtered through a 0.22 μm PES filter (Celltreat Scientific Products, USA). Glucose, fructose and sucrose were separated using high-performance liquid chromatography (HPLC, Shimadzu Co., Kyoto, Japan). The HPLC system consisted of a Shimadzu LC-20AB solvent delivery pump, a SIL-20AC HT autosampler, a CTO-20A column oven and a CBM-20A communications bus module. The system was fitted with a Rezex RCM monosaccharide Ca^{+} size exclusion column (300 mm × 7.8 mm diameter, Phenomenex, Torrance, CA, USA) and a Carbo-Ca^{2+} security guard cartridge (4 mm × 3 mm diameter, Phenomenex, Torrance, CA, USA). The separated sugars were then

detected using a refractive index detector (RID-20A, Shimadzu). The injection volume was 5 μL and the mobile phase was HPLC-grade water at a flow rate of 0.6 mL min^{-1}. The total run time was 20 min, and the column temperature was maintained at 65°C. Standards of glucose (MP Biomedicals, Solon, OH, USA), fructose (Acros Organics, Fair Lawn, NJ, USA) and sucrose (Alfa Aesar, Ward Hill, MA, USA) were used to identify the retention time of individual sugars and concentrations of individual sugars were determined by comparing the peak area of individual sugars to an external standard curve developed for each individual sugar.

Gallic acid was analysed according to Nour et al. (2012) with slight modifications. Briefly, 1 mL methanol and 1% butylhydroxytoluene were added to 14.5 mg of dried leaf tissue and sonicated at 25°C for 40 min. Extracts were centrifuged at 1200 *g* and the supernatants were filtered through a 0.2 μm syringe filter (13 mm diameter, Fisherbrand, PTFE) and stored at −20°C. Gallic acid was identified by the same HPLC system that was used for sugar quantification, equipped with a photodiode array detector (SPD-M20A) and Hypersil Gold column (4.6 mm × 250 mm, 5 μm, Thermo Scientific). The mobile phase used was 1% acetic acid in water (A) and methanol (B) in a gradient mode: 0–27 min (10% B), 27–55 min (10%–40% B), 55–60 min (40% B), 60–62 min (40%–44% B), 62–70 min (44% B), 70–71 min (44%–10% B) and 71–75 min (10% B). The flow rate of the mobile phase and the injection volume were 1 mL min^{-1} at 30°C and 5 μL, respectively. The peak representing GA was identified from chromatograms at 280 nm and by comparing their retention time to that of a pure standard (Sigma-Aldrich, St. Louis, MO, USA). Concentrations of GA were determined by comparing peak area of GA to an external standard curve of the GA standard. Summary statistics of observed foliar functional traits from the different stress combinations for all experiments can be found in Table 2. In total, trait data were collected for *A* ($n = 449$), SLA ($n = 340$), LWC ($n = 340$), N ($n = 196$), sugars ($n = 196$) and GA ($n = 197$).

2.2.3 | Chemometric model calibration and validation

Untransformed reflectance profiles were used to generate predictive models using PLSR (Burnett et al., 2021; Wold et al., 2001). Data analyses for PLSR were conducted in *R v. 3.6.1* (R Core Team, 2019) using the package *pls v. 2.7.1* (Mevik et al., 2019). Eighty percent of the full dataset was used for calibration and cross-validation and the remaining 20% of the full dataset was used for external validation to evaluate the final performance of the models on unseen data. Within the 80% of the dataset used for calibration and cross-validation, the data were split into 80% for model fitting (training) and 20% for cross-validation (testing), stratified across the reference data range over 100 randomized permutations. Using a randomization approach allows for the assessment of model stability and the determination of model uncertainty in prediction.

Model calibration and cross-validation were conducted across 100 different wavelength ranges. To generate different intervals and locations of wavelength ranges, we adjusted the starting and

TABLE 2 Variation in plant functional traits.

	2018 EEL			2018 Wright Center			2019 EEL			2019 Wright Center			All experiments		
Leaf trait	No. of samples	Range	Mean ± SD	No. of samples	Range	Mean ± SD	No. of samples	Range	Mean ± SD	No. of samples	Range	Mean ± SD	No. of samples	Range	Mean ± SD
A	118	1.78–19.08	11.72 ± 3.93	77	0.07–18.22	6.85 ± 4.93	134	0.15–16.29	7.4 ± 4.04	120	0.04–12.92	5.04 ± 3.27	449	0.04–19.08	7.81 ± 4.71
SLA	0	NA	NA	78	197.64–530.54	326.39 ± 65.2	135	205.9–453.84	306.54 ± 54.91	127	155.72–428.86	247.24 ± 65.73	340	155.72–530.54	288.94 ± 69.72
LWC	0	NA	NA	78	38.7–84.69	66.55 ± 5.35	135	48.39–76.29	70.64 ± 2.93	127	43.85–81.9	62.16 ± 7.24	340	38.7–84.69	66.54 ± 6.58
N	54	0.95–2.78	2.01 ± 0.41	39	1.34–4.01	2.4 ± 0.73	55	2.4–4.08	3.14 ± 0.32	48	1.66–3.3	2.39 ± 0.41	196	0.95–4.08	2.5 ± 0.64
Sugars	54	8.91–14.57	12.55 ± 1.38	39	4.37–15.55	11.39 ± 2.89	55	8.32–15.47	12.18 ± 1.7	48	4.81–15.31	9.49 ± 3.41	196	4.37–15.55	11.47 ± 2.68
GA	54	13.02–243.75	63.57 ± 48.39	39	50.1–377.5	146.49 ± 84.76	56	51.81–252.79	137.74 ± 46.93	48	11.01–239.83	59.28 ± 39.94	197	11.01–377.5	100.02 ± 68.08

Note: CO_2 assimilation rate (*A*, μmol CO_2 m^{-2} s^{-1}), specific leaf area (SLA cm^2 g^{-1}), leaf water content (LWC, %), nitrogen (N, % dry mass), sugars (% dry mass) and gallic acid (GA, mg 100 g^{-1} dry mass) collected from four different experiments conducted in 2018 and 2019.

Abbreviation: SD, standard deviation.

ending wavelengths across the spectral signature in the process of calibration. We also ensured that each spectral range encompassed a continuous segment of at least 300 nm. This procedure enabled the inclusion or exclusion of specific characteristic patterns within the vegetation spectrum, generating 100 different wavelength ranges (Figure S1). We employed this approach, as opposed to a feature selection algorithm technique for two reasons: the potentially overly context-dependent nature of these feature selection approaches in application of chemometrics and the reductionist nature of these approaches never ensures, or forces, specific combinations of wavelengths together. Because of the context-dependent nature of most feature selection approaches, outcomes that are considered by an algorithm to be influential in trait estimation may be a secondary product of some other change in plant spectral profiles. For example, such changes may involve colour shifts that are correlated with the trait of interest but not directly related to the absorption features of the trait. Moreover, the reductionist approach used by many feature selection algorithms (e.g. LASSO, ReliefF) potentially excludes the opportunity to incorporate specific combinations or sets of wavelength ranges, and remove wavelengths considered not useful. While removing wavelengths unimportant for a singular dataset, they limit the discovery of potentially novel absorption features and focus only on wavelengths important to the dataset analysed.

During model calibration, the number of latent variables retained in individual models was determined based on reduction of the predicted residual sum of square (PRESS) statistic (Chen et al., 2004) using leave-one-out cross-validation. Prediction residuals more than twice the standard deviation were determined to be outliers. These outliers were removed from the calibration dataset because of either faulty spectral or reference measurements, and models were subsequently re-trained. Specifically, *A*: 16/361 (4.4%), SLA: 9/272 (3.3%), LWC: 8/272 (2.9%), N: 5/156 (3.2%), sugars: 7/156 (4.5%) and GA: 9/157 (5.7%) were removed. Model performance was evaluated using goodness-fit (R^2) and normalized root mean square error (NRMSE, RMSE divided by the reference data range), representing precision and accuracy, respectively, across the different permutations. R^2 represents the proportion of variance in the dependent variable that is explained by the independent variables and is useful for assessing the overall strength of the model in capturing the variability in the data. Conversely, NRMSE focuses on the magnitude of the error between predicted and observed values, providing a measure of accuracy.

Based on the results comparing the model performance from calibration across 100 spectral ranges, final PLSR models were built using the optimal range for each trait that demonstrated the absolute highest R^2 and the lowest NRMSE. The same modelling approach described above was followed, except that 500 randomized permutations were used. To determine the importance of specific wavelengths contributing to final models, we calculated the variable importance of projection (VIP) statistic (Chong & Jun, 2005; Wold et al., 2001). External validation was performed by applying the coefficients from all 500 PLSR models to the data withheld from modelling. Relationships between the mean estimated and observed values were tested by regression analysis. Fit statistics (R^2 and NRMSE) were again used to assess the precision and accuracy of model estimates.

2.2.4 | Statistical analyses

Significant differences ($p < 0.05$) in R^2 or NRMSE of the models using 100 different wavelength ranges in the calibration process, over 100 iterations each, were tested using analysis of variance (ANOVA). A post-hoc Tukey HSD test was used to assess statistically significant pairwise differences ($p = 0.05$). The relationships between wavelength starting and ending and model prediction accuracy were analysed by linear and second-order polynomial regression analysis. The performance of the optimal wavelength range and the full wavelength range were compared by applying the multiple model coefficient iterations of each model on the external validation set separately, and differences were assessed using a *t*-test. Pearson correlation was used to assess the relationship between the correlations of foliar traits estimated using spectral data and those measured using traditional standardized methods. Statistical analyses were performed in GraphPad Prism v 9.5.1 (GraphPad Software, Boston, MA, USA) and JMP v. 16.1 (SAS Institute Inc., Cary, NC, USA).

3 | RESULTS

Model performance parameters (R^2 and NRMSE) for cross-validation outcomes varied among the different wavelength ranges used for building the PLSR models (Figures 1 and 2; Tables S1–S12). Changing the spectral range had a significant impact on model performance, resulting in a variation of 1.7 to 3.0 times in R^2 and 1.3 to 2.6 times in NRMSE when all functional traits were considered. The range of R^2 over the 100 wavelength combinations was 0.23–0.68 for *A*, 0.43–0.90 for SLA, 0.54–0.91 for LWC, 0.40–0.91 for N, 0.29–0.61 for sugars and 0.20–0.51 for GA. The absolute range of NRMSE was 14%–22% for *A*, 7%–17% for SLA, 5%–12% for LWC, 6%–17% for N, 15%–21% for sugars and 16%–21% for GA (Tables S1–S12).

When predicting each foliar functional trait, the wavelength range with the highest R^2 and the wavelength range with the lowest NRMSE were the same for *A* and LWC. For other traits, the wavelength range that yielded the best performance consistently ranked either first or second in these evaluations. However, the number of optimal spectral ranges varied depending on whether R^2 or NRMSE was considered. The exception to this was *A*, in which nine out of the 100 evaluated wavelength ranges exhibited R^2 values above 0.65, and the same nine wavelength ranges were identified as optimal when considering NRMSE, with values less than 14.6% (Figures 1a and 2a; Tables S1 and S7). The wavelength ranges included in these optimal *A* models began at either 400 or 500 nm and extended beyond 1700 nm, with the highest R^2 of 0.68 obtained using the range

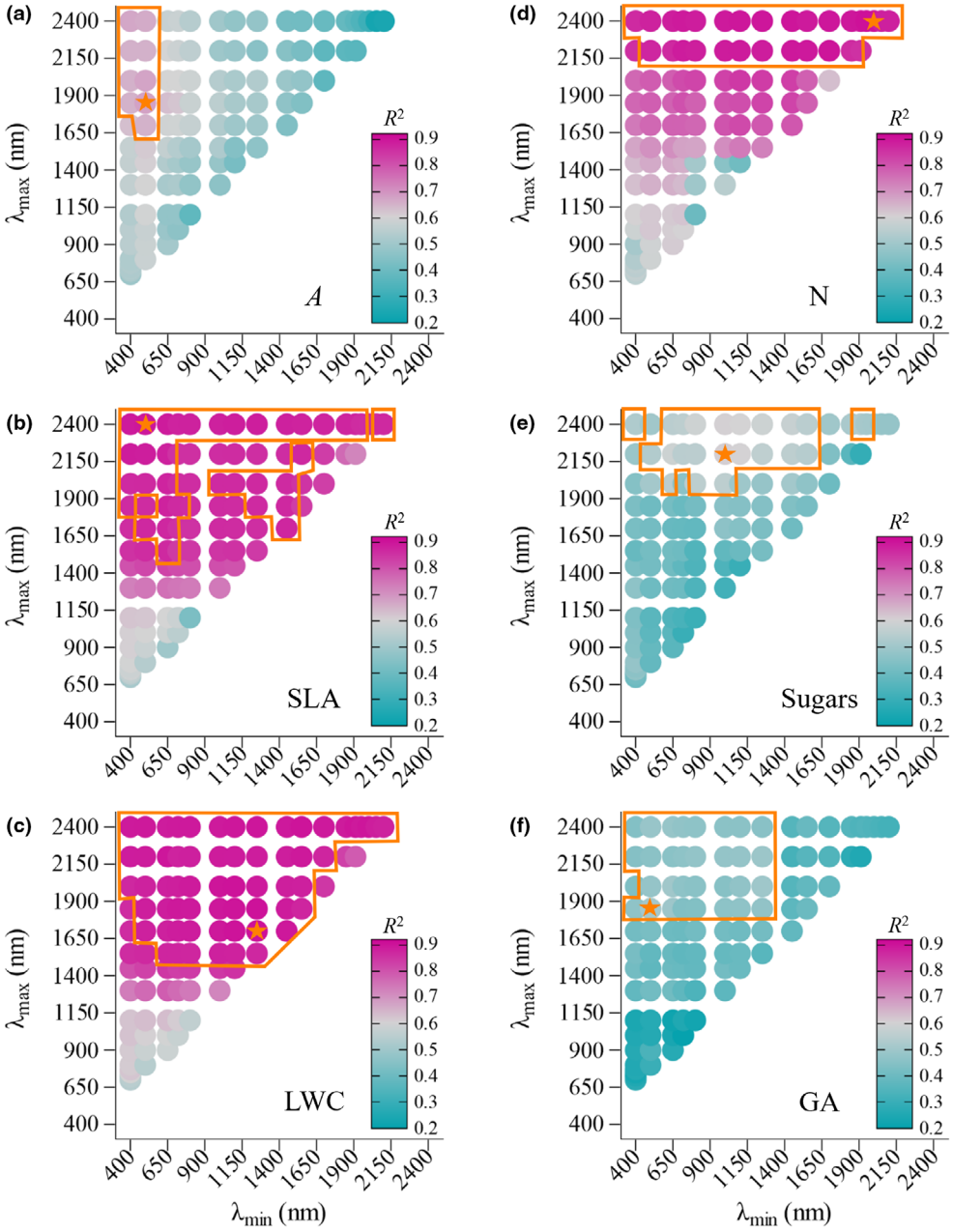

FIGURE 1 Wavelength ranges starting from λ_{min} to λ_{max} influence partial least squares regression (PLSR) model performance for estimating major leaf functional traits, (a) CO_2 assimilation rate (*A*), (b) specific leaf area (SLA), (c) leaf water content (LWC), (d) nitrogen (N), (e) sugars and (f) gallic acid (GA). PLSR model outcomes were made based on the 100 different wavelength ranges, all over 100 iterations 80%/20% calibration and cross-validation approach. Colour maps are contours of the model performance, goodness-of-fit (R^2), on the testing dataset. Data points within the orange-coloured box represent wavelength ranges exhibiting significantly highest R^2 values that did not differ significantly each other and the orange star highlights the spectral region yielding the best result.

of 500–1850 nm (Figure 1a; Table S1). For SLA, 34 optimal models out of the 100 evaluated spectral regions were obtained when wavelengths beyond 1550 nm were included, with the highest performance achieved in the 500–2400 nm range (Figure 1b; Table S2). The number of wavelength ranges that minimized NRMSE for SLA models was reduced to 20, and when the ending wavelength was 2200 nm or less, the starting wavelength needed to include the visible wavelength range (Figure 2b; Table S8). Similarly, for assessing LWC, 61 optimal spectral regions were identified, which could begin at almost any wavelength, but optimal models were obtained when wavelengths beyond 1550 nm were included. For LWC, the best model R^2 of 0.91 was found at 1250–1700 nm (Figure 1c; Table S3). Some of the models with spectral regions that started within the SWIR region or ended at 1550 nm were not identified as optimal spectral regions when using NRMSE as a criterion for evaluation (i.e. below 6%; Figure 2c; Table S9).

Model combinations of foliar N revealed 28 optimal spectral ranges, starting from any wavelength, but optimal models were

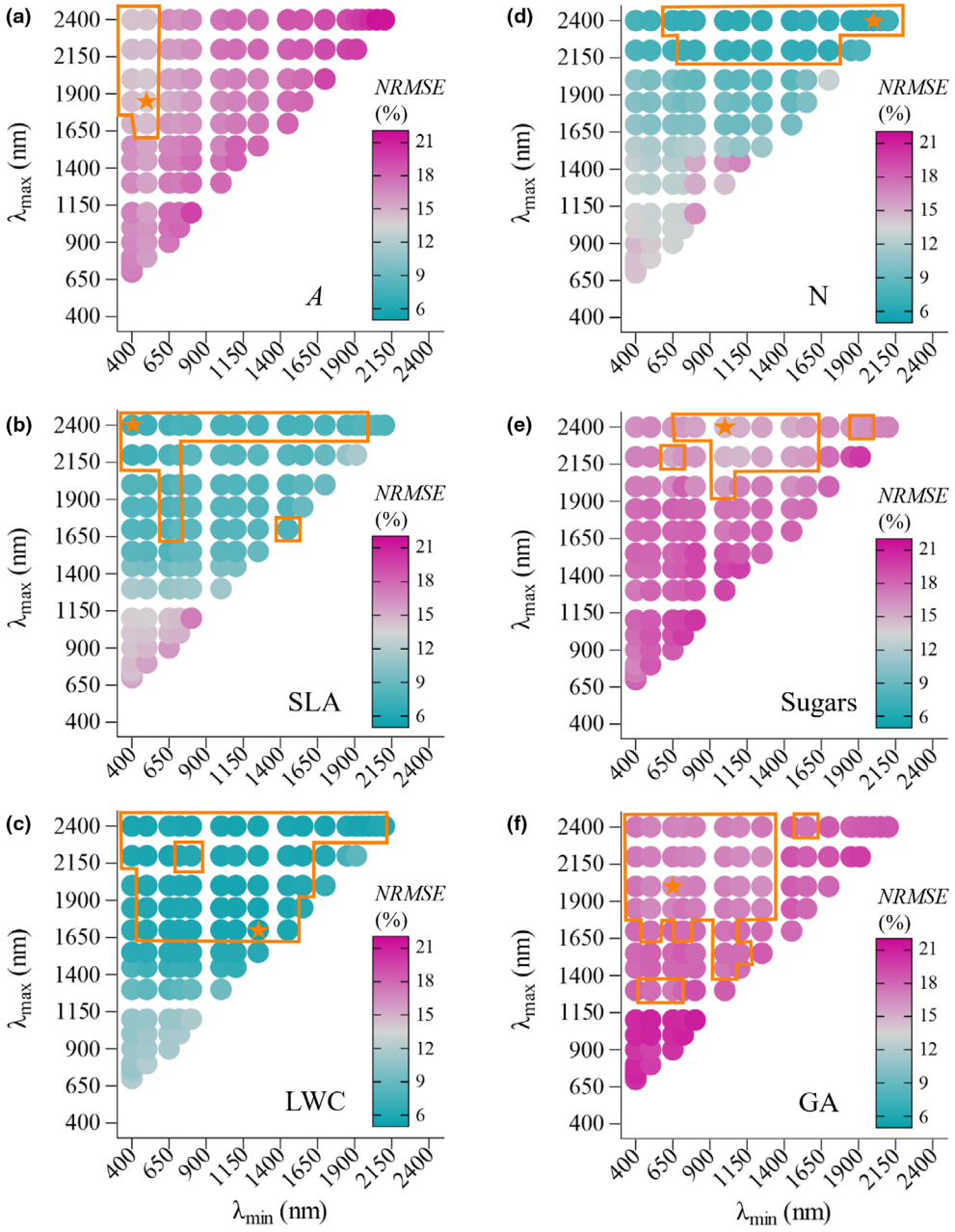

FIGURE 2 Wavelength ranges starting from λ_{min} to λ_{max} influence partial least squares regression (PLSR) model performance for estimating major leaf functional traits, (a) CO_2 assimilation rate (*A*), (b) specific leaf area (SLA), (c) leaf water content (LWC), (d) nitrogen (N), (e) sugars and (f) gallic acid (GA). PLSR model outcomes were made based on the 100 different wavelength ranges, all over 100 iterations 80%/20% calibration and cross-validation approach. Colour maps are contours of the model performance, normalized root mean square error (NRMSE), on the testing dataset. Data points within the orange-coloured box represent wavelength ranges exhibiting significantly lowest NRMSE values that did not differ significantly from each other, and the orange star highlights the spectral region yielding the best result.

obtained when wavelengths beyond 2200 nm were included, with the highest R^2 of 0.91 achieved within the 2000–2400 nm wavelength range (Figure 1d; Table S4). There were 23 optimal spectral ranges for which foliar N estimates were most accurate, with NRMSE values ranging from 6.5% to 7.2%. When the ending wavelength range was 2200 nm, the starting wavelength needed to begin between 720 and 1700 nm and when the wavelength range ended at 2400 nm, the starting wavelength could range from 650 to 2000 nm (Figure 2d; Table S10). For sugars, 22 wavelength ranges were identified and optimal models were obtained when wavelengths beyond 2000 nm were included (Figure 1e; Table S5). The highest R^2 of 0.61 was observed when using the wavelength range 1000–2200 nm. Additionally, 15 wavelength ranges were recognized as optimal based on the lowest NRMSE (14.7%–16.1%). The starting wavelength of these optimal ranges was more narrowly confined to 650 to 1550 nm (Figure 2e; Table S11). For GA, 31 optimal spectral ranges

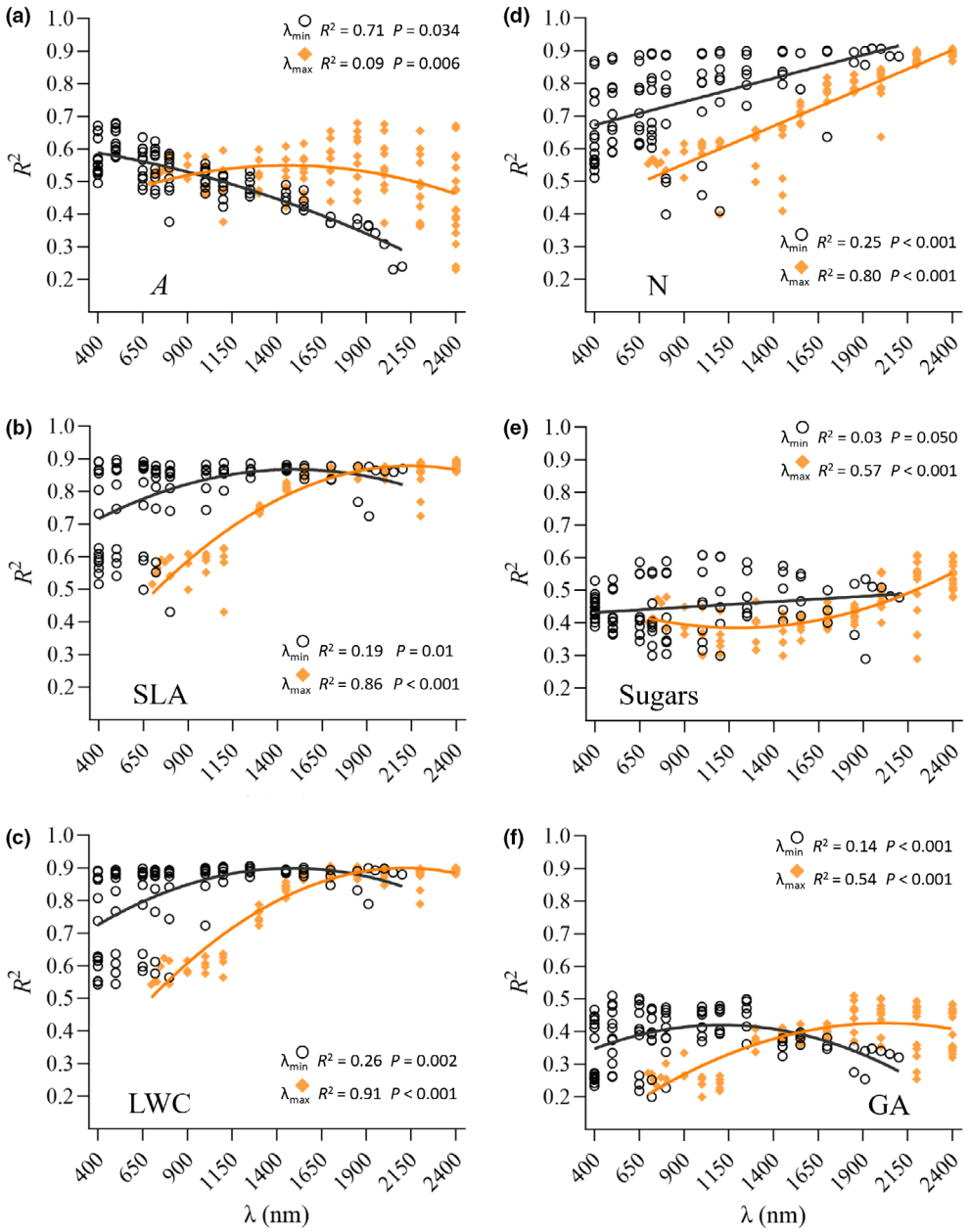

FIGURE 3 Partial least squares regression model performance metric, R^2, as a function of λ_{min} (a starting wavelength, open circle) and λ_{max} (an ending wavelength, orange diamond) for estimating major leaf functional traits, (a) CO_2 assimilation rate (*A*), (b) specific leaf area (SLA), (c) leaf water content (LWC), (d) nitrogen (N), (e) sugars and (f) gallic acid (GA). The lines represent the significant λ–R^2 relationships based on regression analyses ($p < 0.05$).

were identified starting between 400 and 1250 nm and ending between 1850 and 2400 nm. The highest R^2 of 0.51 was found in the wavelength range of 500–1850 nm (Figure 1f; Table S6). When evaluating the model performance based on NRMSE, 41 optimal wavelength ranges were found. These ranges fall within the same optimal spectral regions as those identified by R^2; however, they exhibited an extension in the ending wavelengths, ranging between 1300 and 2400 nm (Figure 2f; Table S12).

We found statistically significant relationships between the starting or ending wavelength and R^2 (Figure 3). For *A*, model R^2 decreased non-linearly with the starting wavelength, whereas it had a weak, concave relationship with the ending wavelength, reaching a maximum of R^2 at 1850 nm (Figure 3a). For SLA, model R^2 increased non-linearly with both the starting and ending wavelength points, but the model performance was less related to the starting wavelength than the ending wavelength (Figure 3b). Similar patterns were observed for LWC, but the difference between R^2 values was smaller than for SLA (Figure 3c). We found a linear increase in R^2 with both the starting and ending wavelengths for estimating N (Figure 3d). However, the relationship between R^2 and the starting wavelength was weaker than the relationship between R^2 and the ending wavelength. For sugars, there was a weak and linear relationship between

R^2 and the starting wavelength, but the relationship between R^2 and the ending wavelength was stronger and demonstrated a convex pattern (Figure 3e). For GA, the R^2 has a weak, concave relationship with the starting wavelength, whereas R^2 increased non-linearly with the ending wavelength (Figure 3f).

The trends between the starting or ending wavelengths and NRMSE were opposite to those between the starting or ending wavelengths and R^2 due to the inverse relationship between R^2 and NRMSE (Figure 4). The effects of starting or ending wavelengths on the NRMSE of each trait were similar to those of the starting or ending wavelengths on the R^2. For *A*, NRMSE was more dependent on the starting wavelength than the ending wavelength (Figure 4a). For all other foliar traits, the ending wavelength was more influential to determine NRMSE than the starting wavelength (Figure 4b–f).

The final models accurately estimated foliar traits in the calibration and cross-validation approach (Table 3). Model performance in the external validation dataset was lower than the calibration outcomes but produced reasonably good metrics (Figure 5). Overall, R^2 values ranged from 0.43 to 0.86 and NRMSE ranged from 9% to 21%. In general, model performance was better for SLA, LWC and N than for *A*, and

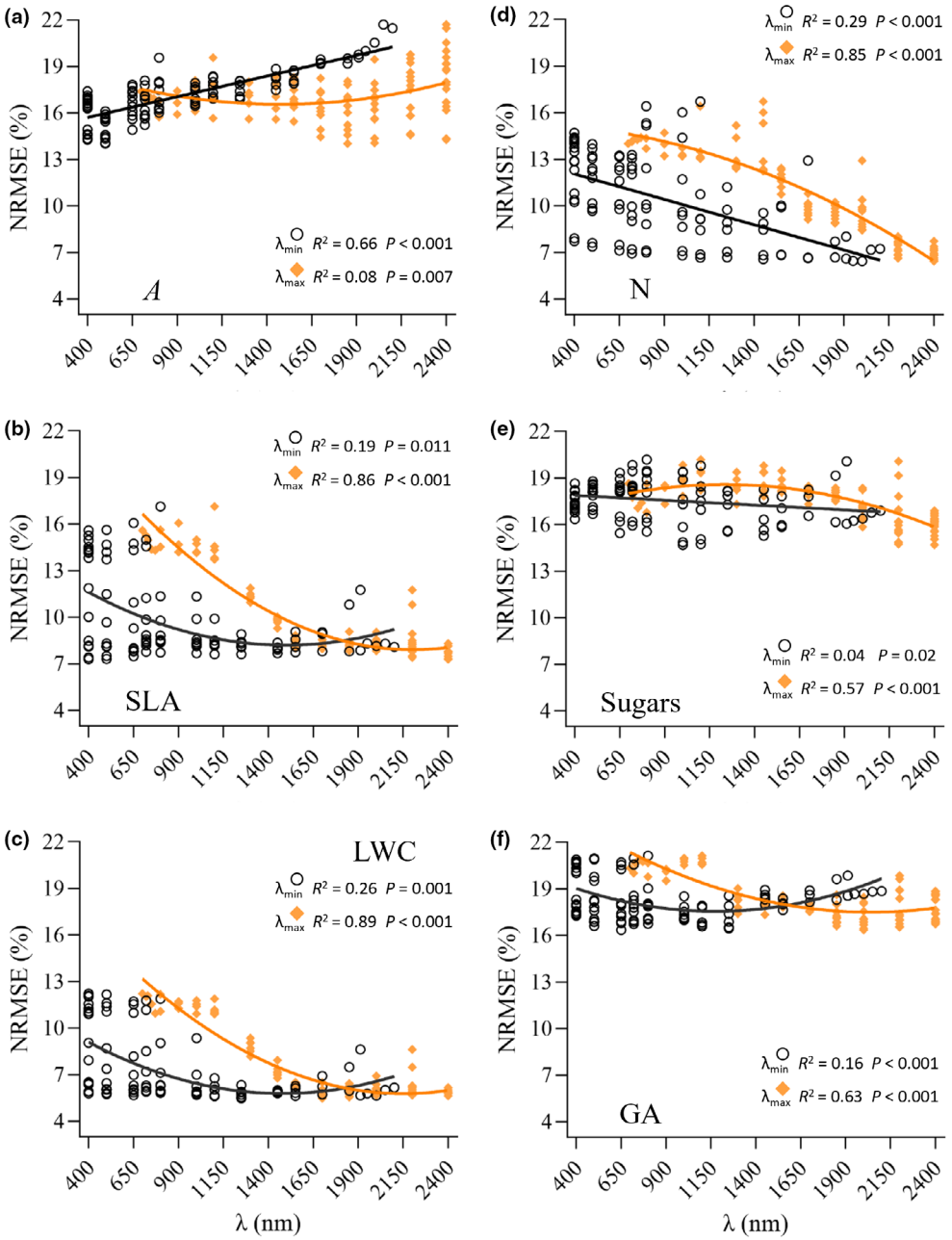

FIGURE 4 Partial least squares regression model performance metric, NRMSE, as a function of λ_{min} (a starting wavelength, open circle) and λ_{max} (an ending wavelength, orange diamond) for estimating major leaf functional traits, (a) CO_2 assimilation rate (*A*), (b) specific leaf area (SLA), (c) leaf water content (LWC), (d) nitrogen (N), (e) sugars and (f) gallic acid (GA). The lines represent the significant λ–NRMSE relationships based on regression analyses ($p < 0.05$).

TABLE 3 Final PLSR model fit (R^2), root mean square error (RMSE), bias and normalized root mean square error (NRMSE) for calibration (C) and cross-validation (CV) data, generated via using 500 random permutations of the data with 80% used for C and 20% for CV for models predicting plant functional traits from black walnut and red oak spectra.

	Range	R^2		RMSE		Bias		NRMSE (%)	
Trait	(nm)	C	CV	C	CV	C	CV	C	CV
A (μmol CO_2 m^{-2} s^{-1})	500–1850	0.84 ± 0.01 (0.82–0.86)	0.68 ± 0.05 (0.46–0.82)	1.88 ± 0.04 (1.73–1.99)	2.65 ± 0.20 (1.96–3.34)	0.00 ± 0.00 (0.00–0.00)	−0.01 ± 0.34 (−1.30–1.09)	9.86 ± 0.23 (9.10–10.46)	13.91 ± 1.03 (10.31–17.53)
SLA (cm^2 g^{-1})	500–2400	0.93 ± 0.00 (0.92–0.94)	0.89 ± 0.02 (0.79–0.95)	17.46 ± 0.42 (16.04–18.60)	21.87 ± 2.12 (15.88–29.79)	0.00 ± 0.00 (0.00–0.00)	0.45 ± 3.44 (−8.55–10.79)	5.88 ± 0.14 (5.40–6.26)	7.36 ± 0.71 (5.34–10.03)
LWC (%)	1250–1700	0.93 ± 0.01 (0.91–0.94)	0.90 ± 0.03 (0.76–0.96)	1.62 ± 0.06 (1.39–1.76)	1.88 ± 0.24 (1.20–2.55)	0.00 ± 0.00 (0.00–0.00)	0.00 ± 0.28 (−0.81–0.86)	4.83 ± 0.18 (4.16–5.25)	5.63 ± 0.71 (3.59–7.63)
N (% dry mass)	2000–2400	0.96 ± 0.00 (0.95–0.97)	0.91 ± 0.02 (0.83–0.97)	0.12 ± 0.00 (0.10–0.13)	0.18 ± 0.02 (0.13–0.24)	0.00 ± 0.00 (0.00–0.00)	0.00 ± 0.04 (−0.12–0.11)	4.06 ± 0.14 (3.46–4.41)	6.37 ± 0.68 (4.50–8.17)
Sugars (% dry mass)	1000–2400	0.93 ± 0.01 (0.90–0.95)	0.62 ± 0.11 (0.16–0.85)	0.68 ± 0.04 (0.56–0.79)	1.57 ± 0.19 (1.07–2.31)	0.00 ± 0.00 (0.00–0.00)	0.03 ± 0.32 (−0.82–1.04)	6.35 ± 0.35 (5.21–7.42)	14.77 ± 1.80 (10.00–21.68)
GA (mg/100 g dry mass)	650–2000	0.70 ± 0.02 (0.65–0.77)	0.48 ± 0.12 (0.10–0.75)	31.16 ± 1.44 (26.04–34.03)	41.54 ± 5.84 (27.55–58.72)	0.00 ± 0.00 (0.00–0.00)	0.03 ± 8.54 (−26.53–25.88)	12.53 ± 0.58 (10.47–13.68)	16.70 ± 2.35 (11.07–23.60)

Note: Final models were built using the optimal wavelength range for each trait, selected based on the highest R^2 and lowest NRMSE. Data are shown as mean ± SD. Values in parentheses are fifth and 95th percentile confidence intervals.

concentrations of sugars and GA. The VIP values generated from models using the selected optimal spectral ranges indicated contributions from more influential spectral wavelengths that aligned with known absorption features for each trait (Figure 6). In the final model, VIP values of *A* highlighted important wavelengths in the 550, 650–750, 970, 1350–1500 and around 1720 nm range (Figure 6a). VIP values of the SLA model highlighted important wavelengths in the visible wavelength range, around 1400–1500 nm and in the 1900–2400 nm range (Figure 6b). LWC was influenced by shortwave infrared absorption features associated with water content (around 1400–1450 nm), but also was influenced by wavelengths around 1690 nm (Figure 6c). VIP values of nitrogen showed multiple peaks in the 2000–2400 nm range, specifically at 2060, 2130, 2180, 2240, 2300 and 2350 nm (Figure 6d). Sharp peaks at 1000, 1370–1450, 1900–2050 nm, along with smaller peaks above 2000 nm, were observed for sugars (Figure 6e). Finally, VIP values of GA showed major peaks at 650–750, 1450, 1720 nm and at around 1900 nm (Figure 6f). Variation in raw spectral data, compared with VIP values and known absorption features, reveal variation along spectral profiles in areas high VIP values and absorption features (Figure 6g).

External validation of models using full-range spectral data performed differently with the optimal-range models for most traits (Figure S2). A comparison of R^2 values between full-range and optimal-range spectral models showed statistically significant differences across all traits. The full range was statistically better for *A* ($t = 17.8$, $p < 0.001$), SLA ($t = 7.5$, $p < 0.001$) and LWC ($t = 50.0$, $p < 0.001$), but the increases in R^2 were modest (≤4%; *A*: R^2 0.533 and 0.512, SLA: R^2 0.842 and 0.838, LWC: R^2 0.692 and 0.672, for full and optimal ranges, respectively). In contrast, the optimal range performed better for N ($t = -3.6$, $p < 0.001$), sugars ($t = -57.6$, $p < 0.001$) and GA ($t = -11.1$, $p < 0.001$), with R^2 values that were 0.4%, 30% and 7% greater, respectively, compared to the full wavelength range (N: R^2 0.857 and 0.861, sugars: R^2 0.455 and 0.590, GA: R^2 0.426 and 0.457, for full and optimal ranges, respectively).

Traits were correlated in both the observed and estimated leaf trait datasets (Figure 7). We found that correlation patterns between observed foliar traits (Figure 7a) were similar to those found in estimated traits (Figure 7b), although we identified stronger relationships in the estimated dataset compared to the observed dataset. While there was a statistically significant and positive relationship between the foliar trait correlations in the observed and estimated datasets (Figure 7c), we did find variation in both the magnitude and direction of trait correlations across individual experiments (Tables S13–S16 and S17–S20 for observed traits and estimated traits, respectively).

When data were split between the two tree species, model performance for most traits (i.e. *A*, LWC, N and sugars) was comparable (Figure S3a,b,d,e). Model performance for SLA and GA was lower in red oak compared to black walnut, but SLA was still well predicted for red oak, with a mean R^2 of 0.63 and mean NRMSE comparable to black walnut (red oak mean NRMSE of 7.3% and black walnut mean NRMSE of 7.6%; Figure S3c). R^2 was approximately 40% lower, and NRMSE was 40% higher for models estimating GA from red oak compared to black walnut (Figure S3f).

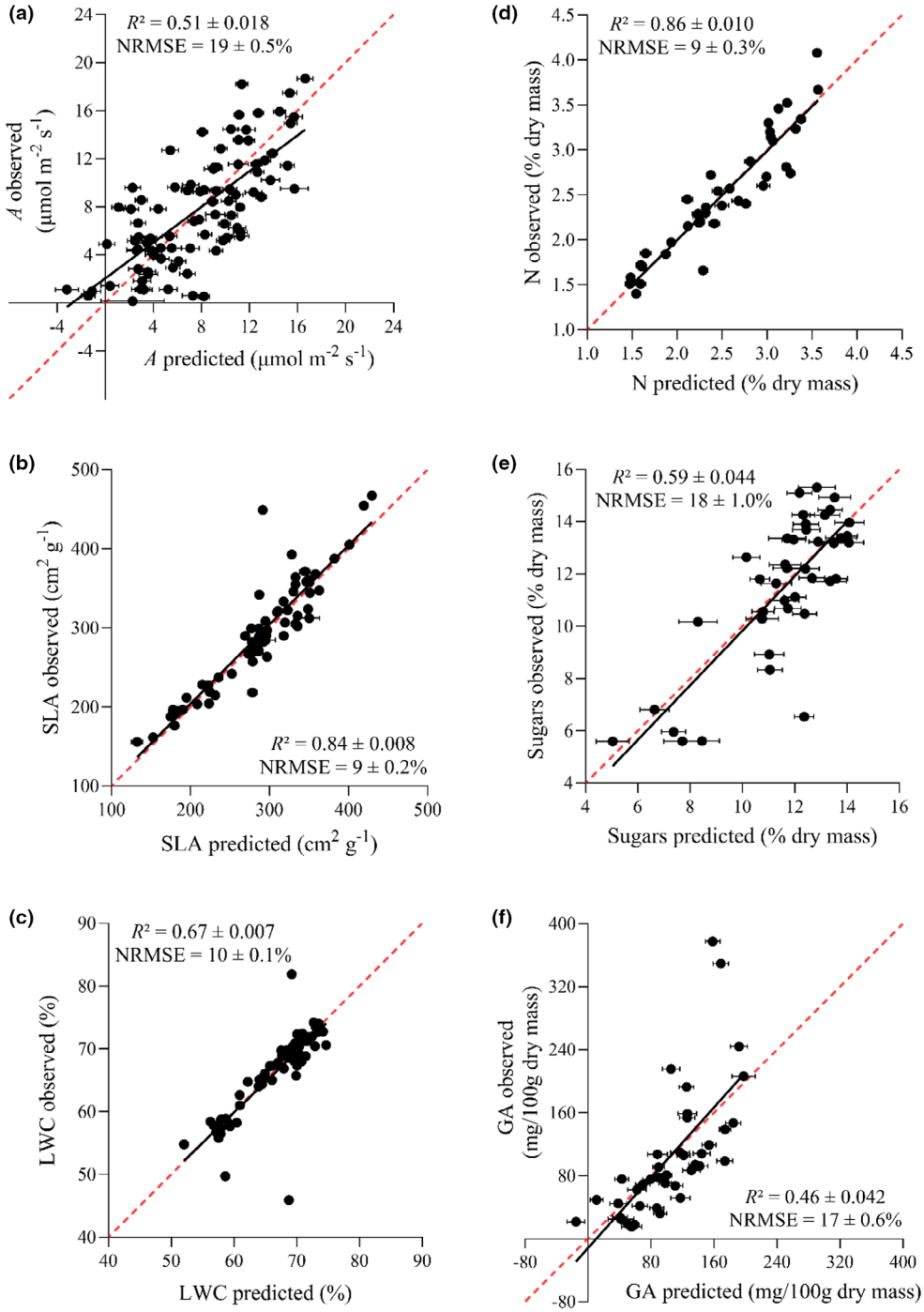

FIGURE 5 External validation of partial least squares regression-estimated traits using final models. (a) CO_2 assimilation rate (*A*), (b) specific leaf area (SLA), (c) leaf water content (LWC), (d) nitrogen (N), (e) sugars and (f) gallic acid (GA) showed precision and accuracy of spectral approach using each optimal wavelength range. Error bars for estimated values represent the standard deviations generated from 500 simulated models. Red line is 1:1 relationship; model goodness-fit (R^2), normalized root mean square-error (NRMSE) for external validation data.

4 | DISCUSSION

In this study, we found that the wavelength range used in chemometric modelling affected model performance of six ecologically relevant foliar functional traits that included physiological, anatomical and chemical parameters. We observed distinct improvements in model performance when incorporating spectral regions characterized by absorption features of specific traits. We also found that for most traits, full-range spectral data performed similarly to a reduced spectral range, a likely outcome of the distribution of absorption features associated with most traits used in this study. We do note, however, that model performance of traits with more specific absorption

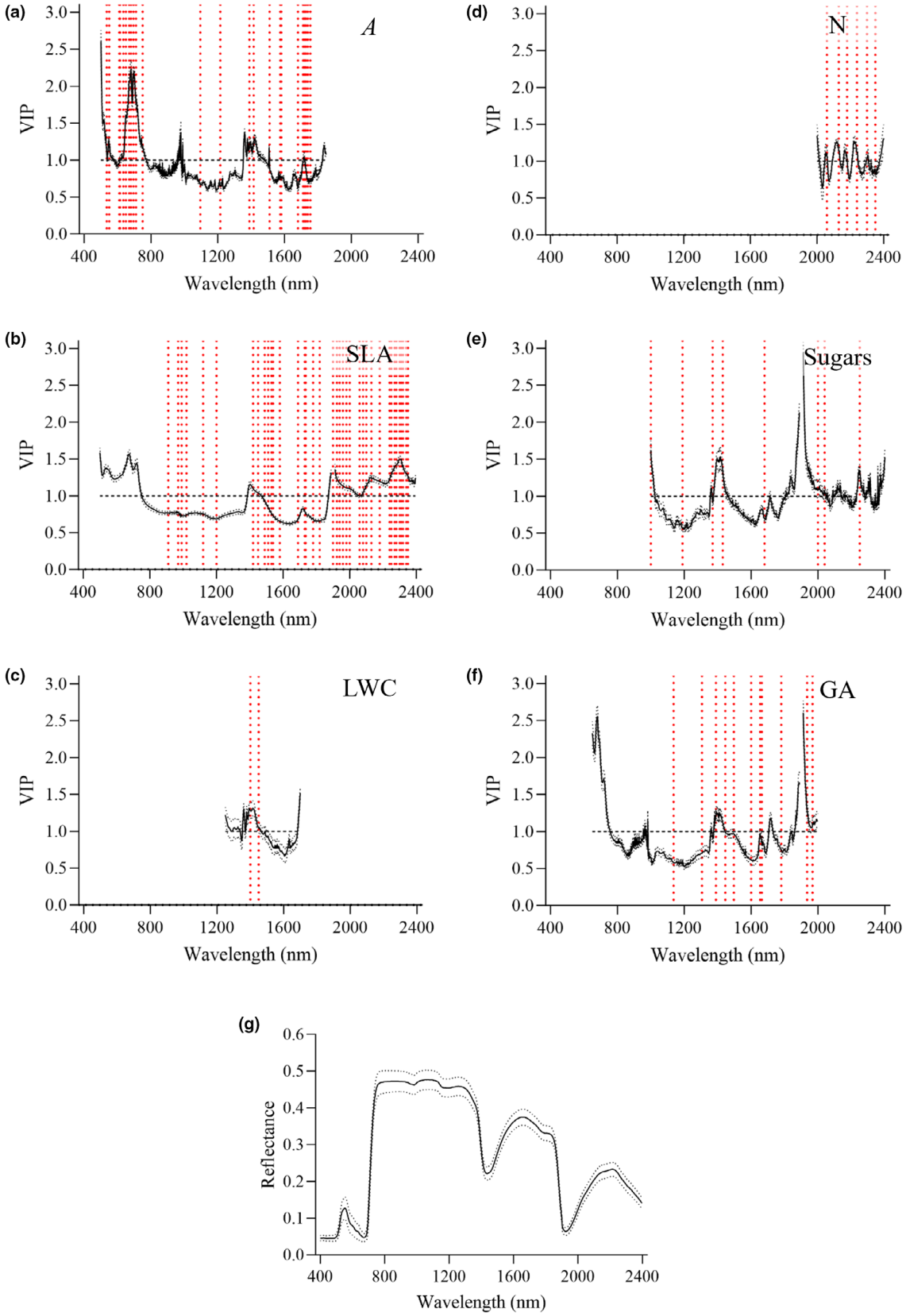

FIGURE 6 Mean (solid) and ±1 standard deviation (dotted) of variable importance of projection (VIP) of partial least squares regression leaf trait models using the best range selected for each trait, (a) CO_2 assimilation rate (*A*), (b) specific leaf area (SLA), (c) leaf water content (LWC), (d) nitrogen (N), (e) sugars and (f) gallic acid (GA) and (g) mean (solid) and ±1 standard deviation (dotted) of spectral profiles of leaves collected for this study. VIP denotes the contribution of each wavelength to the trait estimation and a VIP value of >1 (dashed line) is considered significant with. Wavelengths associated with optical features of each trait are presented as red vertical dotted lines (Cotrozzi et al., 2017; Curran, 1989; Gara et al., 2021; Kumar et al., 2001; Serbin et al., 2012; Figure S4).

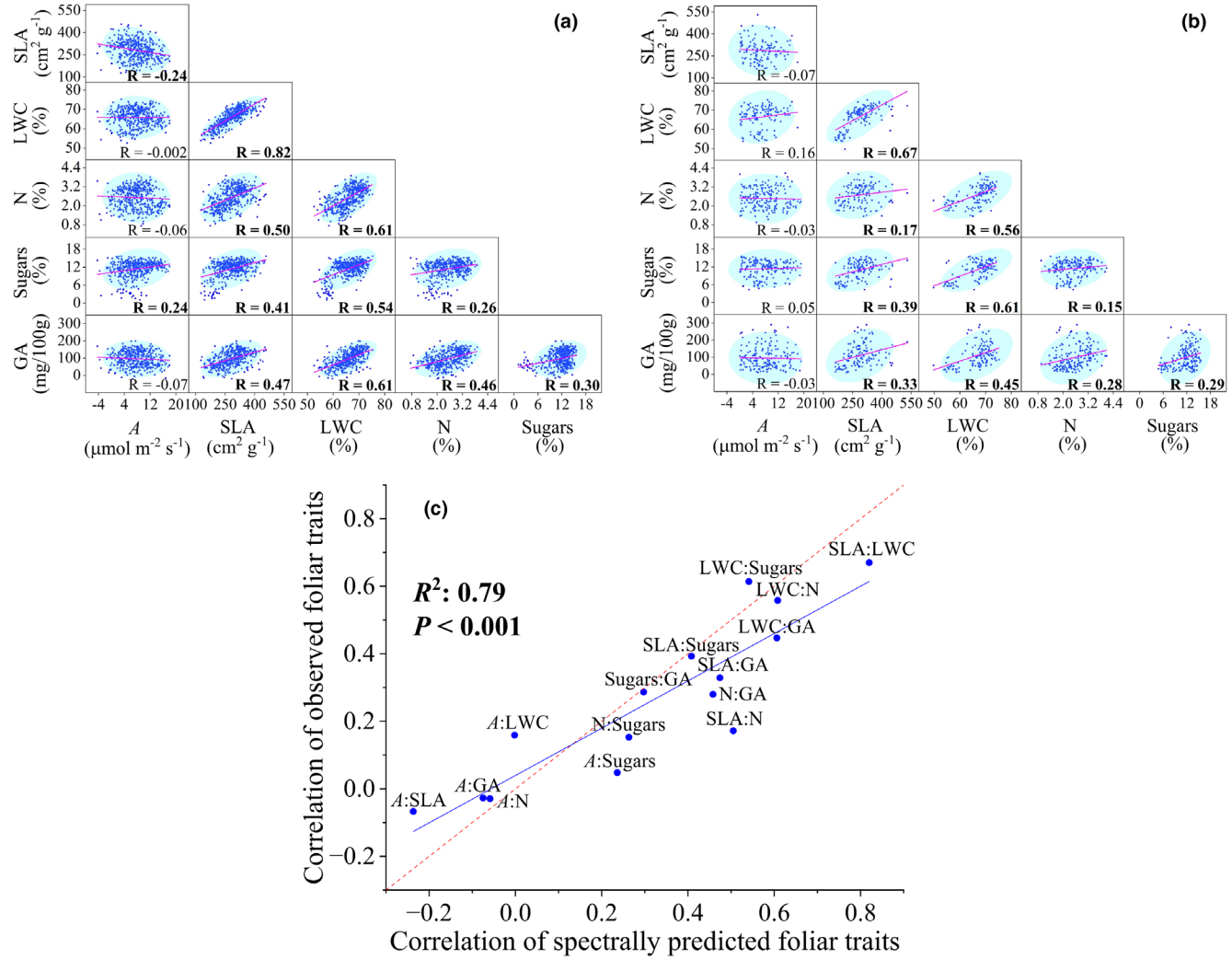

FIGURE 7 Scatterplot of pairwise trait correlations from (a) estimated and (b) observed data sets across all experiments combined and (c) relationship between the correlations between foliar traits predicted using spectral data and those measured using traditional standardized methods. R^2 is calculated on correlation values from (a) and (b).

features in a narrower spectral region was negatively affected when using full-range spectral data, suggesting additional spectral bands can introduce noise and error into chemometric approaches. We suggest that the approach used in this study can (1) assess the relative impact of absorption features specific to individual functional traits on model performance and (2) possibly identify novel spectral regions containing absorption features for complex traits or traits with unknown or potentially multiple non-singular or weak absorption features. An emergent outcome of this study is that including the SWIR wavelength region in modelling improved the accuracy of all trait estimations. Our findings underscore the importance of optimal spectral range selection in enhancing the performance of chemometric models for specific foliar trait estimations, enabling trait estimates to serve as surrogates for standard reference measurements, which are often logistically challenging to collect and analyse. This approach can help broaden the suite of measurable plant functional traits that are influential in ecosystem functioning and highlights the utility of integrating spectral biology into a trait-based ecological framework.

4.1 | The importance of starting or ending wavelength on model performance differs depending on plant functional traits

We found that starting or ending wavelength influenced model performance in a trait-specific manner. Furthermore, we observed trends in model performance that were predominantly non-linear, indicating that the impact of wavelength range variation on model accuracy is not consistent across different wavelengths. This outcome is likely a result of the differential distribution of spectral signals related to absorption features that reflect the distinctive spectral characteristics of each trait.

Of the traits we measured, model performance for estimation of *A* was found to be more dependent on starting wavelengths. The greater influence of starting wavelengths on the prediction accuracy of *A*, relative to other functional traits, was likely because including or excluding the visible range had a more substantial impact on model performance due to the relationship with pigment pools affecting net

CO_2 assimilation rates. Other studies using spectral data to estimate photosynthetic parameters have shown the importance of longer wavelengths (i.e. >1100 nm) for improved model performance (Serbin et al., 2015; Yendrek et al., 2016). While we found that model performance for *A* improved when longer wavelengths were included, the models we generated were less dependent on ending wavelengths. This outcome may be due to the enhanced importance of VNIR wavelengths due to the different stress conditions the plants experienced (Appendix S1) when collecting data in this study that likely affected pigment pools more than Rubisco pools (Calzone et al., 2021; Cotrozzi, Peron, et al., 2020; Cotrozzi, Lorenzini, et al., 2020). Moreover, previous studies (Serbin et al., 2015; Yendrek et al., 2016) focused on more inherently stable photosynthetic parameters (e.g. V_{cmax}) that potentially leverage longer wavelengths associated with nitrogen absorption features related to Rubisco pools preferentially to shifts in pigment pools that might be more related to maximum photosynthetic rate (Sexton et al., 2021). This outcome highlights that starting and ending wavelengths play an important role in determining absorption features that can improve model performance in chemometric modelling. However, the relative importance of wavelength ranges with known absorption features might be trait specific and depend on vegetation stress or health status, especially for complex traits.

4.2 | Convergence of starting and ending wavelength range for models appears to encompass spectral ranges of known trait absorption features

In this study, we found that when we overlay the patterns in the relationship between the starting and ending wavelengths and model performance parameters, the range between the peaks of two lines appears to contain spectral regions of known or speculated absorption features for individual traits. This outcome suggests that the peaks of these regression lines may bookend regions of important absorption features. For example, the two regression lines depicting the starting and ending wavelengths for *A* intersect near 800 nm, which is close to the red edge (700–750 nm), a spectral region strongly influenced by chlorophyll content (Mutanga & Skidmore, 2007; Smith et al., 2003). Additionally, the range between the two peaks of the lines, spanning from 500 to 1850 nm, includes absorption features for both chlorophyll *a* and *b*, as well as proteins in the longer wavelengths (Curran, 1989; Elvidge, 1990). For SLA, the critical range, encompassing from 1400 to 2400 nm, includes absorption features related to the ligno-cellulose and starch (Gara et al., 2021). Moreover, the intersection of the two curves for starting and ending wavelength ranges for LWC falls within a range containing water absorption features, specifically from 1400 to 2000 nm (Féret et al., 2019).

Improved models for both N and sugars were found when the spectral range extended into longer wavelengths. Unlike other traits whose model performance stabilizes beyond ~1850 nm, suggesting that key spectral features have already been captured, N and sugars continue to benefit from an extended wavelength range. Nitrogen and sugars did not demonstrate the same relationship as other traits, but converged or approached to convergence when wavelengths above 2000 nm were included. For N, major absorption features associated with N-containing biochemical constituents, such as proteins, contribute strongly in the longer wavelengths (Wan et al., 2022). Sugar-related absorption features span from 1000 to 2200 nm (Cotrozzi et al., 2017), with improvements in sugar estimation extending to 2200 nm. Meanwhile, GA exhibited a clear convergence near 1600 nm and two peaks between 1100 and 2100 nm, encompassing phenol absorption (~1660 nm; Couture et al., 2016; Kokaly & Skidmore, 2015).

4.3 | Including longer wavelengths improves trait estimations

We found that the most accurate models were achieved by including SWIR wavelength ranges. Our finding that longer wavelength ranges can improve model performance is supported by previous findings highlighting numerous absorption features in the SWIR region related with functional traits used in this study, including those associated with water, cellular structure, proteins, carbohydrates and phenolic compounds (Calzone et al., 2021; Chen et al., 2022; Cotrozzi et al., 2017; Couture et al., 2016; Curran, 1989; Dechant et al., 2017; Féret et al., 2019; Kokaly & Skidmore, 2015; Nunes et al., 2017; Ramirez et al., 2015; Xie et al., 2024; Yan et al., 2021). Because of the information in the SWIR region, models containing this range were identified as among the most optimal, exhibiting stronger predictive performance. In contrast, models based solely on the VIS–NIR range showed reduced accuracy for all traits relative to models containing SWIR regions. These findings are consistent with previous studies demonstrating that although the metrics often highlight the importance of the VIS–NIR region when using the full spectral range, excluding VIS–NIR wavelengths in favour of SWIR alone does not compromise model performance (Kothari et al., 2024; Sexton et al., 2021; Wan et al., 2022). This outcome also highlights the importance of evaluating individual wavelength regions to isolate and validate the contribution of SWIR-specific absorption features to trait prediction (Sexton et al., 2021).

Building on these findings, our analysis further demonstrates that model performance for some traits remains relatively comparable regardless of extending the starting or ending wavelength point, as long as the wavelength ranges that contain relevant spectral information for a specific trait are included. For example, when assessing *A*, we achieved better model performance with spectral ranges starting at 400–500 nm and extending through 1700–2400 nm. In contrast, models estimating LWC were less influenced by the starting wavelength and more dependent on ending wavelength position, resulting in an approximately sevenfold increase in optimal ranges that were not statistically different. This outcome likely arises from how the distribution of important absorption features associated with each trait influences the model performance. The assessment of *A* benefitted from the inclusion of both VIS and SWIR regions in this study, likely because pigments and photosynthetic proteins playing pivotal roles in determining photosynthetic capacity exhibit

absorption features in the VIS and SWIR regions, respectively (Cotrozzi, Peron, et al., 2020; Serbin et al., 2012). This underscores the critical role of combining these spectral regions for accurate estimates of A, emphasizing how longer wavelengths improve leaf-level photosynthesis predictions (Doughty et al., 2011). Conversely, LWC is a measure of the amount of foliar water, and optimal models required the inclusion of major water absorption features which are predominant in the SWIR; thus, the starting wavelength range was less important. The variability observed in the number of optimal wavelength ranges among the different foliar traits suggests that this depends on the amount, magnitude and distribution of absorption features of target traits across the spectrum.

Many studies have employed the full spectral range (400–2400 nm) for estimating different functional traits and reported strong model performance (Ely et al., 2019; Kothari, Beauchamp-Rioux, Blanchard, et al., 2023; Serbin et al., 2014; Yan et al., 2021). In our study, it is worth noting that the full spectral region was also included as one of the optimal wavelength ranges for most traits. External validation of PLSR models revealed comparable model performance between the selected range with the highest model performance and the full spectral range for most models, with minimal differences (i.e. ≤4%). We did find, however, a more pronounced reduction in model performance when using the full spectral range for estimating sugars and GA. The decline in model performance for these specific traits may be attributed to the relatively specific spectral signals associated with sugars and GA across the full spectrum, compared with the other traits investigated in this study. This potentially implies that including wavelength ranges that do not contain absorption features related to a specific trait can add noise and negatively affect spectral estimation of some traits, especially those with narrow absorption features or multiple, potentially weak, absorption features. Characterization of absorption features for ecologically relevant compounds thus becomes imperative for reducing any noise from unnecessary wavelengths (Serbin et al., 2012). These findings suggest that while employing the entire spectral range may be adequate for predicting certain traits, it may not be adequate for all plant functional traits.

4.4 | Reliable trait predictions across tree species

Generalized trait models that cover diverse tree species have been developed using PLSR (Nakaji et al., 2019; Serbin et al., 2019; Wan et al., 2022; Wang et al., 2023). The transferability of PLSR models across different species has been shown to depend on the representativeness of the calibration dataset, particularly whether it encompasses diverse plant species. Additionally, a dataset that includes diverse species is likely to capture variation from multiple absorption features of certain traits when many absorption features are present (Wan et al., 2022). We acknowledge that this study is limited by using only two species, black walnut and red oak, both of which are the same age, to develop trait models. However, after selecting an optimal wavelength range that maximized model performance, we applied it separately to each species, resulting in comparable performance for most traits when each species was assessed independently. The decline in performance of models estimating SLA and GA may be attributed to a narrower range of trait values and a reduced number of samples available for the evaluations. However, for SLA, mean performance was still relatively good given the reduction in trait range and sample size, confirming that single and multispecies spectral models estimating some traits can be reliable even though there is variation among species and functional groups (Kothari, Beauchamp-Rioux, Blanchard, et al., 2023; Kothari, Beauchamp-Rioux, Laliberté, et al., 2023; Serbin et al., 2019).

4.5 | Consistent functional trait relationships across studies and integrating spectral biology and plant ecology

We found that pairwise trait correlations were similar for both observed and estimated traits, as seen in Chadwick and Asner (2016). While these correlations varied somewhat across the different treatments, the overall pattern of trait correlations was consistent between observed and estimated data when considering all treatments collectively. The consistency in this relationship suggests that spectrally estimated functional trait data can act as a surrogate for standard reference-collected trait data and has been shown to produce similar statistical outcomes when compared with analyses using standard reference data (Cotrozzi, Peron, et al., 2020). In addition, studies using only spectrally estimated functional traits have revealed insights into plant responses to environmental change, the influence of functional trait variation on plant–insect interactions and the influence of geographic variation of insect herbivory on plant demographic processes (Cavender-Bares et al., 2016; Couture et al., 2015; DeLaMater et al., 2021; Struckman et al., 2019). We suggest that the use of spectrally estimated functional traits can integrate spectral biology into a plant trait-based ecological framework.

5 | CONCLUSIONS

Outcomes from this work suggest this modelling approach has broader implications beyond the modelling of only known absorption features. It can be extended to identify unknown absorption features in leaves for traits, potentially enhancing our ability to predict and understand these traits and their responses to stress. By leveraging the relationships between starting and ending wavelengths and their convergence at or near significant spectral signals, we offer a novel approach for identifying absorption features of traits that are ecologically relevant but to date have not been focused on in the literature because of unknown or complex absorption features. We also highlight consistencies among trait correlations for observed and estimated traits data, suggesting that estimated data can act as a surrogate for standard reference measurements. This outcome highlights that a spectral biology approach can be integrated into an ecological framework, providing access to a suite of plant functional

traits that can be logistically challenging to collect, ultimately advancing our understanding of relationships among vegetation optical properties, plant functional traits and plant ecology.

AUTHOR CONTRIBUTIONS

Minjee Park collected the data, conducted the data analysis, wrote the first draft of the manuscript and edited the manuscript. Lorenzo Cotrozzi collected the data and edited the manuscript. Geoffrey M. Williams was involved in the fungal inoculation and edited the manuscript. John J. Couture conceptualized the study, designed the methodology, contributed to initial writing and led the review and editing process. John J. Couture secured funding through USDA NIFA Hatch, Matthew D. Ginzel and Douglass F. Jacobs secured funding through CAFS, while John J. Couture, Matthew D. Ginzel and Douglass F. Jacobs secured funding through HTIRC. Matthew D. Ginzel, Michael V. Mickelbart, Geoffery M. Williams and Douglass F. Jacobs contributed to manuscript editing. All authors reviewed and approved the final version of the manuscript for publication.

ACKNOWLEDGEMENTS

We thank Mike Gosney, Daniel Edwards, Kristin Sauder and Megan Haas who helped with the spectral and reference measurements in the field and laboratory. This research was supported by the USDA NIFA Hatch award IND011490 to John J. Couture, the National Science Foundation Center for Advanced Forestry Systems (CAFS) Award #1916587 to Matthew D. Ginzel and Douglass F. Jacobs, and the Hardwood Tree Improvement and Regeneration Center (HTIRC) at Purdue University to John J. Couture, Matthew D. Ginzel and Douglass F. Jacobs. This research was in partial fulfilment of a doctoral degree for Minjee Park from Purdue University.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflict of interest.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/2041-210X.70100.

DATA AVAILABILITY STATEMENT

Data and code used in this study are archived in the Ecological Spectral Information System (EcoSIS) https://doi.org/10.21232/bCRYo6R2 (Park et al., 2025).

ORCID

Minjee Park https://orcid.org/0000-0001-9921-4890
Lorenzo Cotrozzi https://orcid.org/0000-0002-4401-3896
Geoffrey M. Williams https://orcid.org/0000-0001-7585-704X
Matthew D. Ginzel https://orcid.org/0000-0003-2444-491X
Michael V. Mickelbart https://orcid.org/0000-0001-5939-3126
Douglass F. Jacobs https://orcid.org/0000-0002-5580-2516
John J. Couture https://orcid.org/0000-0003-4784-4537

REFERENCES

Asner, G. P., Martin, R. E., Anderson, C. B., & Knapp, D. E. (2015). Quantifying forest canopy traits: Imaging spectroscopy versus field survey. *Remote Sensing of Environment*, *158*, 15–27.

Asner, G. P., Martin, R. E., Anderson, C. B., Kryston, K., Vaughn, N., Knapp, D. E., Bentley, L. P., Shenkin, A., Salinas, N., Sinca, F., Tupayachi, R., Quispe Huaypar, K., Montoya Pillco, M., Ccori Álvarez, F. D., Díaz, S., Enquist, B. J., & Malhi, Y. (2017). Scale dependence of canopy trait distributions along a tropical forest elevation gradient. *New Phytologist*, *214*(3), 973–988.

Asner, G. P., Martin, R. E., Keith, L. M., Heller, W. P., Hughes, M. A., Vaughn, N. R., Asner, G., Martin, R., Keith, L., Heller, W., Hughes, M., Vaughn, N., Hughes, R., & Balzotti, C. (2018). A spectral mapping signature for the rapid ohia death (ROD) pathogen in Hawaiian forests. *Remote Sensing*, *10*, 404.

Burnett, A. C., Anderson, J., Davidson, K. J., Ely, K. S., Lamour, J., Li, Q., Morrison, B. D., Yang, D., Rogers, A., & Serbin, S. P. (2021). A best-practice guide to predicting plant traits from leaf-level hyperspectral data using partial least squares regression. *Journal of Experimental Botany*, *72*, 6175–6189.

Calzone, A., Cotrozzi, L., Remorini, D., Lorenzini, G., Nali, C., & Pellegrini, E. (2021). Oxidative stress assessment by a spectroscopic approach in pomegranate plants under a gradient of ozone concentrations. *Environmental and Experimental Botany*, *182*, 104309.

Cavender-Bares, J., Meireles, J. E., Couture, J. J., Kaproth, M., Singh, A., Kingdon, C. C., Center, A., Zuniga, E., Pilz, G., & Townsend, P. A. (2016). Associations of leaf spectra with genotypic and phylogenetic variation in oaks: Prospects for remote detection of biodiversity. *Remote Sensing*, *8*, 221.

Chadwick, K. D., & Asner, G. P. (2016). Organismic-scale remote sensing of canopy foliar traits in lowland tropical forests. *Remote Sensing*, *8*, 87.

Chen, L., Zhang, Y., Nunes, M. H., Stoddart, J., Khoury, S., Chan, A. H. Y., & Coomes, D. A. (2022). Predicting leaf traits of temperate broadleaf deciduous trees from hyperspectral reflectance: Can a general model be applied across a growing season? *Remote Sensing of Environment*, *269*, 112767.

Chen, S., Hong, X., Harris, C. J., & Sharkey, P. M. (2004). Sparse modeling using orthogonal forward regression with PRESS statistic and regularization. *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, *34*, 898–911.

Cheng, T., Rivard, B., & Sanchez-Azofeifa, A. (2011). Spectroscopic determination of leaf water content using continuous wavelet analysis. *Remote Sensing of Environment*, *115*(2), 659–670.

Chong, I.-G., & Jun, C.-H. (2005). Performance of some variable selection methods when multicollinearity is present. *Chemometrics and Intelligent Laboratory Systems*, *78*, 103–112.

Cotrozzi, L. (2022). Spectroscopic detection of forest diseases: A review (1970–2020). *Journal of Forestry Research*, *33*(1), 21–88.

Cotrozzi, L., Couture, J. J., Cavender-Bares, J., Kingdon, C. C., Fallon, B., Pilz, G., Pellegrini, E., Nali, C., & Townsend, P. A. (2017). Using foliar spectral properties to assess the effects of drought on plant water potential. *Tree Physiology*, *37*, 1582–1591.

Cotrozzi, L., Lorenzini, G., Nali, C., Pellegrini, E., Saponaro, V., Hoshika, Y., Arab, L., Rennenberg, H., & Paoletti, E. (2020). Hyperspectral reflectance of light-adapted leaves can predict both dark- and light-adapted Chl fluorescence parameters, and the effects of chronic ozone exposure on date palm (*Phoenix dactylifera*). *International Journal of Molecular Sciences*, *21*, 6441.

Cotrozzi, L., Peron, R., Tuinstra, M. R., Mickelbart, M. V., & Couture, J. J. (2020). Spectral phenotyping of physiological and anatomical leaf traits related with maize water status. *Plant Physiology*, *184*, 1363–1377.

Couture, J. J., Serbin, S. P., & Townsend, P. A. (2015). Elevated temperature and periodic water stress alter growth and quality of common

milkweed (*Asclepias syriaca*) and monarch (*Danaus plexippus*) larval performance. *Arthropod-Plant Interactions*, *9*, 149–161.

Couture, J. J., Singh, A., Rubert-Nason, K. F., Serbin, S. P., Lindroth, R. L., & Townsend, P. A. (2016). Spectroscopic determination of ecologically relevant plant secondary metabolites. *Methods in Ecology and Evolution*, *7*, 1402–1412.

Cui, E., Weng, E., Yan, E., & Xia, J. (2020). Robust leaf trait relationships across species under global environmental changes. *Nature Communications*, *11*(1), 2999.

Curran, P. J. (1989). Remote sensing of foliar chemistry. *Remote Sensing of Environment*, *30*, 271–278.

Dao, P. D., He, Y., Lu, B., & Axiotis, A. (2025). Imaging spectroscopy reveals topographic variability effects on grassland functional traits and drought responses. *Ecology*, *106*(3), e70006.

Dechant, B., Cuntz, M., Vohland, M., Schulz, E., & Doktor, D. (2017). Estimation of photosynthesis traits from leaf reflectance spectra: Correlation to nitrogen content as the dominant mechanism. *Remote Sensing of Environment*, *196*, 279–292.

DeLaMater, D. S., Couture, J. J., Puzey, J. R., & Dalgleish, H. J. (2021). Range-wide variations in common milkweed traits and their effect on monarch larvae. *American Journal of Botany*, *108*(3), 388–401.

Doughty, C. E., Asner, G. P., & Martin, R. E. (2011). Predicting tropical plant physiology from leaf and canopy spectroscopy. *Oecologia*, *165*, 289–299.

Elvidge, C. D. (1990). Visible and near infrared reflectance characteristics of dry plant materials. *Remote Sensing*, *11*(10), 1775–1795.

Ely, K. S., Burnett, A. C., Lieberman-Cribbin, W., Serbin, S. P., & Rogers, A. (2019). Spectroscopy can predict key leaf traits associated with source–sink balance and carbon–nitrogen status. *Journal of Experimental Botany*, *70*, 1789–1799.

Féret, J. B., le Maire, G., Jay, S., Berveiller, D., Bendoula, R., Hmimina, G., Cheraiet, A., Oliveira, J. C., Ponzoni, F. J., Solanki, T., de Boissieu, F., Chave, J., Nouvellon, Y., Porcar-Castell, A., Proisy, C., Soudani, K., Gastellu-Etchegorry, J.-P., & Lefèvre-Fonollosa, M. J. (2019). Estimating leaf mass per area and equivalent water thickness based on leaf optical properties: Potential and limitations of physical modeling and machine learning. *Remote Sensing of Environment*, *231*, 110959.

Gara, T. W., Rahimzadeh-Bajgiran, P., & Darvishzadeh, R. (2021). Forest leaf mass per area (LMA) through the eye of optical remote sensing: A review and future outlook. *Remote Sensing*, *13*, 3352.

Garcia, M. N., Tameirão, L. B., Schietti, J., Aleixo, I., Domingues, T. F., Huemmrich, K. F., Tameirão, L. B. S., Campell, P. K. E., & Albert, L. P. (2025). Predicting drought vulnerability with leaf reflectance spectra in Amazonian trees. *Remote Sensing of Environment*, *318*, 114562.

Hill, J., Buddenbaum, H., & Townsend, P. A. (2019). Imaging spectroscopy of forest ecosystems: Perspectives for the use of space-borne hyperspectral earth observation systems. *Surveys in Geophysics*, *40*, 553–588.

Huang, C.-y., Anderegg, W. R. L., & Asner, G. P. (2019). Remote sensing of forest die-off in the Anthropocene: From plant ecophysiology to canopy structure. *Remote Sensing of Environment*, *231*, 111233.

Kazakou, E., Violle, C., Roumet, C., Pintor, C., Gimenez, O., & Garnier, E. (2009). Litter quality and decomposability of species from a Mediterranean succession depend on leaf traits but not on nitrogen supply. *Annals of Botany*, *104*(6), 1151–1161.

Kokaly, R. F., & Skidmore, A. K. (2015). Plant phenolics and absorption features in vegetation reflectance spectra near 1.66 μm. *International Journal of Applied Earth Observation and Geoinformation*, *43*, 55–83.

Kothari, S., Beauchamp-Rioux, R., Blanchard, F., Crofts, A. L., Girard, A., Guilbeault-Mayers, X., Hacker, P. W., Pardo, J., Schweiger, A. K., Demers-Thibeault, S., Bruneau, A., Coops, N. C., Kalacska, M., Vellend, M., & Laliberté, E. (2023). Predicting leaf traits across functional groups using reflectance spectroscopy. *New Phytologist*, *238*, 549–566.

Kothari, S., Beauchamp-Rioux, R., Laliberté, E., & Cavender-Bares, J. (2023). Reflectance spectroscopy allows rapid, accurate and non-destructive estimates of functional traits from pressed leaves. *Methods in Ecology and Evolution*, *14*, 385–401.

Kothari, S., Hobbie, S. E., & Cavender-Bares, J. (2024). Rapid estimates of leaf litter chemistry using reflectance spectroscopy. *Canadian Journal of Forest Research*, *54*(9), 978–991.

Kothari, S., & Schweiger, A. K. (2022). Plant spectra as integrative measures of plant phenotypes. *Journal of Ecology*, *110*, 2536–2554.

Kumar, L., Schmidt, K., Dury, S., & Skidmore, A. (2001). Imaging spectrometry and vegetation science. In F. D. van der Meer & S. M. De Jong (Eds.), *Imaging spectrometry: Basic principles and prospective applications* (pp. 111–155). Springer.

Kumar, S., Abedin, M. M., Singh, A. K., & Das, S. (2020). Role of phenolic compounds in plant-defensive mechanisms. In R. Lone, R. Shuab, & A. Kamili (Eds.), *Plant phenolics in sustainable agriculture: Volume 1* (pp. 517–532). Springer.

Lassalle, G. (2021). Monitoring natural and anthropogenic plant stressors by hyperspectral remote sensing: Recommendations and guidelines based on a meta-review. *Science of the Total Environment*, *788*, 147758.

Lausch, A., Erasmi, S., King, D. J., Magdon, P., & Heurich, M. (2016). Understanding forest health with remote sensing -part I—A review of spectral traits, processes and remote-sensing characteristics. *Remote Sensing*, *8*, 1029.

Mevik, B.-H., Wehrens, R., & Liland, K. H. (2019). *Pls: Partial least squares and principal component regression*. R package version 2.7–2. https://CRAN.R-project.org/package=pls

Mutanga, O., & Skidmore, A. K. (2007). Red edge shift and biochemical content in grass canopies. *ISPRS Journal of Photogrammetry and Remote Sensing*, *62*, 34–42.

Nakaji, T., Oguma, H., Nakamura, M., Kachina, P., Asanok, L., Marod, D., Aiba, M., Kurokawa, H., Kosugi, Y., Kassim, A. R., & Hiura, T. (2019). Estimation of six leaf traits of east Asian forest tree species by leaf spectroscopy and partial least square regression. *Remote Sensing of Environment*, *233*, 111381.

Nour, V., Trandafir, I., & Cosmulescu, S. (2012). HPLC determination of phenolic acids, flavonoids and juglone in walnut leaves. *Journal of Chromatographic Science*, *51*, 883–890.

Nunes, M. H., Davey, M. P., & Coomes, D. A. (2017). On the challenges of using field spectroscopy to measure the impact of soil type on leaf traits. *Biogeosciences*, *14*, 3371–3385.

Park, M., Cotrozzi, L., Williams, G. M., Ginzel, M. D., Mickelbart, M. V., Jacobs, D. F., Couture, J. J., (2025). Purdue LeafSpectral and Functional Trait Data used in PLSR modeling v2. *Data set*.

Pellegrini, E., Campanella, A., Paolocci, M., Trivellini, A., Gennai, C., Muganu, M., Nali, C., & Lorenzini, G. (2015). Functional leaf traits and diurnal dynamics of photosynthetic parameters predict the behavior of grapevine varieties towards ozone. *PLoS One*, *10*, e0135056.

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ramirez, J. A., Posada, J. M., Handa, I. T., Hoch, G., Vohland, M., Messier, C., & Reu, B. (2015). Near-infrared spectroscopy (NIRS) predicts non-structural carbohydrate concentrations in different tissue types of a broad range of tree species. *Methods in Ecology and Evolution*, *6*, 1018–1025.

Sapes, G., Schroeder, L., Scott, A., Clark, I., Juzwik, J., Montgomery, R. A., Guzmán Q, J. A., & Cavender-Bares, J. (2024). Mechanistic links between physiology and spectral reflectance enable previsual detection of oak wilt and drought stress. *Proceedings of the National Academy of Sciences of the United States of America*, *121*(7), e2316164121.

Serbin, S. P., Dillaway, D. N., Kruger, E. L., & Townsend, P. A. (2012). Leaf optical properties reflect variation in photosynthetic metabolism and its sensitivity to temperature. *Journal of Experimental Botany*, *63*, 489–502.

Serbin, S. P., Singh, A., Desai, A. R., Dubois, S. G., Jablonski, A. D., Kingdon, C. C., Kruger, E. L., & Townsend, P. A. (2015). Remotely

estimating photosynthetic capacity, and its response to temperature, in vegetation canopies using imaging spectroscopy. *Remote Sensing of Environment*, *167*, 78–87.

Serbin, S. P., Singh, A., McNeil, B. E., Kingdon, C. C., & Townsend, P. A. (2014). Spectroscopic determination of leaf morphological and biochemical traits for northern temperate and boreal tree species. *Ecological Applications*, *24*, 1651–1669.

Serbin, S. P., & Townsend, P. A. (2020). Scaling functional traits from leaves to canopies. In *Remote sensing of plant biodiversity* (pp. 43–82). Springer.

Serbin, S. P., Wu, J., Ely, K. S., Kruger, E. L., Townsend, P. A., Meng, R., Wolfe, B. T., Chlus, A., Wang, Z., & Rogers, A. (2019). From the Arctic to the tropics: Multibiome prediction of leaf mass per area using leaf reflectance. *New Phytologist*, *224*, 1557–1568.

Sexton, T., Sankaran, S., & Cousins, A. B. (2021). Predicting photosynthetic capacity in tobacco using shortwave infrared spectral reflectance. *Journal of Experimental Botany*, *72*, 4373–4383.

Smith, M.-L., Martin, M. E., Plourde, L., & Ollinger, S. V. (2003). Analysis of hyperspectral data for estimation of temperate forest canopy nitrogen concentration: Comparison between an airborne (AVIRIS) and a spaceborne (Hyperion) sensor. *IEEE Transactions on Geoscience and Remote Sensing*, *41*, 1332–1337.

Struckman, S., Couture, J. J., LaMar, M. D., & Dalgleish, H. J. (2019). The demographic effects of functional traits: An integral projection model approach reveals population-level consequences of reproduction-defense tradeoffs. *Ecology Letters*, *22*, 1396–1406.

Thomas, S., Kuska, M. T., Bohnenkamp, D., Brugger, A., Alisaac, E., Wahabzada, M., Behmann, J., & Mahlein, A.-K. (2018). Benefits of hyperspectral imaging for plant disease detection and plant protection: A technical perspective. *Journal of Plant Diseases and Protection*, *125*, 5–20.

Violle, C., Navas, M.-L., Vile, D., Kazakou, E., Fortunel, C., Hummel, I., & Garnier, E. (2007). Let the concept of trait be functional! *Oikos*, *116*, 882–892.

Wan, L., Zhou, W., He, Y., Wanger, T. C., & Cen, H. (2022). Combining transfer learning and hyperspectral reflectance analysis to assess leaf nitrogen concentration across different plant species datasets. *Remote Sensing of Environment*, *269*, 112826.

Wang, Z., Féret, J.-B., Liu, N., Sun, Z., Yang, L., Geng, S., Zhang, H., Chlus, A., Kruger, E. L., & Townsend, P. A. (2023). Generality of leaf spectroscopic models for predicting key foliar functional traits across continents: A comparison between physically- and empirically-based approaches. *Remote Sensing of Environment*, *293*, 113614.

Wang, Z., Townsend, P. A., & Kruger, E. L. (2022). Leaf spectroscopy reveals divergent inter-and intra-species foliar trait covariation and trait–environment relationships across NEON domains. *New Phytologist*, *235*(3), 923–938.

Wold, S., Ruhe, A., Wold, H., & Dunn, W. J., III. (1984). The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. *SIAM Journal on Scientific and Statistical Computing*, *5*, 735–743.

Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, *58*, 109–130.

Wright, I. J., Reich, P. B., Westoby, M., Ackerly, D. D., Baruch, Z., Bongers, F., Cavender-Bares, J., Chapin, T., Cornelissen, J. H., Diemer, M., Flexas, J., Garnier, E., Groom, P. K., Gulias, J., Hikosaka, K., Lamont, B. B., Lee, T., Lee, W., Lusk, C., … Villar, R. (2004). The worldwide leaf economics spectrum. *Nature*, *428*(6985), 821–827.

Wu, F., Liu, S., Lamour, J., Atkin, O. K., Yang, N., Dong, T., Xu, W., Smith, N. G., Wang, Z., Wang, H., Su, Y., Liu, X., Shi, Y., Xing, A., Dai, G., Dong, J., Swenson, N. G., Kattge, J., Reich, P. B., … Yan, Z. (2024). Linking leaf dark respiration to leaf traits and reflectance spectroscopy across diverse forest types. *New Phytologist*, *246*, 481–497.

Xie, R., Darvishzadeh, R., Skidmore, A., & van der Meer, F. (2024). Characterizing foliar phenolic compounds and their absorption features in temperate forests using leaf spectroscopy. *ISPRS Journal of Photogrammetry and Remote Sensing*, *212*, 338–356.

Yan, Z., Guo, Z., Serbin, S. P., Song, G., Zhao, Y., Chen, Y., Wu, S., Wang, J., Wang, X., Li, J., Wang, B., Wu, Y., Su, Y., Wang, H., Rogers, A., Liu, L., & Wu, J. (2021). Spectroscopy outperforms leaf trait relationships for predicting photosynthetic capacity across different forest types. *New Phytologist*, *232*(1), 134–147.

Yendrek, C. R., Tomaz, T., Montes, C. M., Cao, Y., Morse, A. M., Brown, P. J., McIntyre, L. M., Leakey, A. D. B., & Ainsworth, E. A. (2016). High-throughput phenotyping of maize leaf physiological and biochemical traits using hyperspectral reflectance. *Plant Physiology*, *173*, 614–626.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Appendix S1: Plant material and experimental designs.

Table S1: List of 100 wavelength ranges used for building PLSR models estimating CO_2 assimilation rate (*A*), ranked in descending order by the mean goodness of fit, R^2, across model evaluations by conducting 100 permutations.

Table S2: List of 100 wavelength ranges used for building PLSR models estimating specific leaf area, ranked in descending order by the mean goodness of fit, R^2, across model evaluations by conducting 100 permutations.

Table S3: List of 100 wavelength ranges used for building PLSR models estimating leaf water content, ranked in descending order by the mean goodness of fit, R^2, across model evaluations by conducting 100 permutations.

Table S4: List of 100 wavelength ranges used for building PLSR models estimating nitrogen, ranked in descending order by the mean goodness of fit, R^2, across model evaluations by conducting 100 permutations.

Table S5: List of 100 wavelength ranges used for building PLSR models estimating sugars, ranked in descending order by the mean goodness of fit, R^2, across model evaluations by conducting 100 permutations.

Table S6: List of 100 wavelength ranges used for building PLSR models estimating gallic acid, ranked in descending order by the mean goodness of fit, R^2, across model evaluations by conducting 100 permutations.

Table S7: List of 100 wavelength ranges used for building PLSR models estimating CO_2 assimilation rate (*A*), ranked in ascending order by the mean normalized root mean square error, NRMSE, across model evaluations by conducting 100 permutations.

Table S8: List of 100 wavelength ranges used for building PLSR models estimating specific leaf area, ranked in ascending order by the mean normalized root mean square error, NRMSE, across model evaluations by conducting 100 permutations.

Table S9: List of 100 wavelength ranges used for building PLSR models estimating leaf water content, ranked in ascending order by the mean normalized root mean square error, NRMSE, across model evaluations by conducting 100 permutations.

Table S10: List of 100 wavelength ranges used for building PLSR models estimating nitrogen, ranked in ascending order by the

mean normalized root mean square error, NRMSE, across model evaluations by conducting 100 permutations.

Table S11: List of 100 wavelength ranges used for building PLSR models estimating sugars, ranked in ascending order by the mean normalized root mean square error, NRMSE, across model evaluations by conducting 100 permutations.

Table S12: List of 100 wavelength ranges used for building PLSR models estimating gallic acid, ranked in ascending order by the mean normalized root mean square error, NRMSE, across model evaluations by conducting 100 permutations.

Table S13: Pearson's correlation matrix describing relationships among observed leaf traits collected in EEL in 2018.

Table S14: Pearson's correlation matrix describing relationships among observed leaf traits collected in the Wright Center in 2018.

Table S15: Pearson's correlation matrix describing relationships among observed leaf traits collected in EEL in 2019.

Table S16: Pearson's correlation matrix describing relationships among observed leaf traits collected in the Wright Center in 2019.

Table S17: Pearson's correlation matrix describing relationships among estimated leaf traits collected in EEL in 2018.

Table S18: Pearson's correlation matrix describing relationships among estimated leaf traits collected in the Wright Center in 2018. Statistically significant ($p < 0.05$) correlations are bolded.

Table S19: Pearson's correlation matrix describing relationships among estimated leaf traits collected in EEL in 2019.

Table S20: Pearson's correlation matrix describing relationships among estimated leaf traits collected in the Wright Center in 2019.

Figure S1: 100 wavelength ranges selected for building PLSR model in this study.

Figure S2: External validation of partial least squares regression-predicted traits using the full wavelength range (400–2400 nm) model.

Figure S3: Correlation scatterplot between observed values and partial least squares regression-estimated trait values for black walnut (black line) and red oak (red line), derived from the calibration dataset.

Figure S4: First-derivative reflectance profiles of the gallic acid standard within the optimal range for gallic acid (650–2000 nm).

How to cite this article: Park, M., Cotrozzi, L., Williams, G. M., Ginzel, M. D., Mickelbart, M. V., Jacobs, D. F., & Couture, J. J. (2025). Spectral wavelength range influences the performance of chemometric models estimating various foliar functional traits. *Methods in Ecology and Evolution*, *16*, 1703–1722. https://doi.org/10.1111/2041-210X.70100