

Generality of leaf spectroscopic models for predicting key foliar functional traits across continents: A comparison between physically- and empirically-based approaches

Zhihui Wang^{a,*}, Jean-Baptiste Féret^b, Nanfeng Liu^c, Zhongyu Sun^a, Long Yang^a, Shoubao Geng^a, Hui Zhang^{d,e}, Adam Chlus^f, Eric L. Kruger^f, Philip A. Townsend^f

^a Guangdong Provincial Key Laboratory of Remote Sensing and Geographical Information System, Guangdong Open Laboratory of Geospatial Information Technology and Application, Guangzhou Institute of Geography, Guangdong Academy of Sciences, Guangzhou 510070, China

^b TETIS, INRAE, AgroParisTech, CIRAD, CNRS, Université Montpellier, Montpellier, France

^c Guangdong Provincial Key Laboratory of Urbanization and Geo-simulation, School of Geography and Planning, Sun Yat-Sen University, Guangzhou 510275, China

^d College of Forestry/Wuzhishan National Long-Term Forest Ecosystem Monitoring Research Station, Hainan University, Haikou 570228, China

^e Key Laboratory of Genetics and Germplasm Innovation of Tropical Special Forest Trees and Ornamental Plants (Hainan University), Ministry of Education, College of Forestry, Hainan University, Haikou 570228, China

^f Department of Forest and Wildlife Ecology, University of Wisconsin-Madison, 1630 Linden Drive, Madison, WI 53706, USA

ARTICLE INFO

Keywords:

Foliar functional traits
Leaf spectroscopy
PROSPECT
COSINE
Physically-based model
Empirical approach
PLSR
Model transferability

ABSTRACT

Leaf spectroscopy provides an efficient way of predicting foliar functional traits, commonly using physically- and empirically-based models. However, the generality of both models has not been fully investigated, and it is not clear if inversion strategies of physically-based models can be transferred across datasets. In this study, we evaluated the generality of leaf spectroscopic models for predicting key foliar functional traits and compared the performance of physically- and empirically-based approaches. Two extensive datasets compiling a total of 3861 foliar samples were collected from 24 field sites in eastern United States and south China. The leaf radiative transfer model PROSPECT was coupled with COSINE (PROCOSINE) to retrieve foliar traits from leaf bidirectional reflectance factor (BRF). A commonly used empirically-based model, partial least squares regression (PLSR) was performed as a comparison. Results showed that both PROSPECT and PROCOSINE can accurately estimate leaf mass per area (LMA) and equivalent water thickness (EWT). Inversion strategies including the selection of optimal spectral domains and the use of prior information (*IS3*) greatly improved the estimation accuracy of leaf nitrogen, leaf chlorophyll *a + b* and carotenoids. The estimation accuracies were similar when transferring inversion strategies across datasets, indicating a high level of transferability of physically-based models. PLSR and interval PLSR (iPLSR, via feature selection) could predict foliar traits with high accuracies when cross-validation was performed, and iPLSR achieved higher accuracies. But both the empirical approaches demonstrated low transferability when applied to an independent dataset. Our findings highlight the importance of generalized traits models with respect to development and calibration of leaf radiative transfer model, as well as incorporating representative samples in training empirical models. This study can help us to better understand the variation of foliar traits among and within species, their response to environmental change, as well as plant biodiversity.

1. Introduction

Plant functional traits refer to the morphological, biochemical and physiological properties of plants which determine the establishment, growth, reproduction and survival of plants, and reflect the adaptation

and acclimation of plants to environment (Pérez-Harguindeguy et al., 2013; Reich, 2014; Wright et al., 2004). Foliar functional traits play an important role in ecosystem processes and functions such as nutrient cycling and gross primary productivity (Cornwell et al., 2008; Schimel and Schneider, 2019; van Bodegom et al., 2014), and therefore are

* Corresponding author at: Guangzhou Institute of Geography, Guangdong Academy of Sciences, Guangzhou 510070, China.

E-mail address: zwang896@wisc.edu (Z. Wang).

<https://doi.org/10.1016/j.rse.2023.113614>

Received 21 March 2023; Received in revised form 24 April 2023; Accepted 30 April 2023

Available online 8 May 2023

0034-4257/© 2023 Elsevier Inc. All rights reserved.

considered as essential biodiversity variables and have been used to parameterize Earth system models (Pereira et al., 2013; Rogers et al., 2017; Skidmore et al., 2021). The most studied foliar functional traits include leaf mass per area (LMA), leaf water content (also known as equivalent water thickness, EWT), leaf nitrogen (N), phosphorus (P) and potassium (K), and leaf pigments (chlorophyll *a* + *b*, carotenoids) (Díaz et al., 2016; Ustin et al., 2009; Wright et al., 2004). LMA and N are often used to define the “leaf economics spectrum” which reflects the trade-off between resource acquisition and allocation (Reich et al., 1997; Wright et al., 2004). Leaf water transports nutrients within plants and is of great importance for photosynthesis, respiration and transpiration (Ustin et al., 2012). P and K are key macro nutrients for plant growth and metabolism (Taiz and Zeiger, 2010). Leaf pigments play an essential role in harvesting light for photosynthesis and dissipating excess light to provide photoprotection under high illumination (Croft and Chen, 2018).

Traditionally, foliar traits are collected through field sampling and laboratory chemistry analysis, which is time-consuming and expensive. In contrast, leaf spectroscopy provides an efficient way of determining foliar traits (Féret et al., 2021; Serbin et al., 2019; Wang et al., 2020; Yang et al., 2016). In leaf spectroscopy, physically- and empirically-based approaches are two categories of methods used to link foliar traits to leaf spectra (Féret et al., 2019; Li et al., 2018; Shiklomanov et al., 2016). The physically-based approach relies on the inversion of radiative transfer models (RTMs). Based on physical laws, RTMs describe the absorption, scattering and reflection processes of light within leaves (Féret et al., 2019; Jacquemoud and Baret, 1990). Of all radiative transfer models, PROSPECT is the most widely used. It simulates the leaf directional-hemispherical reflectance (DHR) and transmittance (DHT) spectra within 400–2500 nm based on the content per surface unit of a set of chemical constituents characterized by a specific absorption coefficient, and a structure parameter to account for scattering (Féret et al., 2008; Jacquemoud and Baret, 1990).

The physically-based approach generally shows a stronger robustness and transferability than the empirically-based approach (Darvishzadeh et al., 2008; Féret et al., 2019). However, limitations still exist in this approach. First, the foliar traits that can be estimated by the physically-based approach are limited to the input parameters of RTMs. For instance, the leaf constituents included in PROSPECT are EWT, LMA, chlorophyll *a* + *b* content (C_{ab}), carotenoids (C_{xc}), anthocyanins, nitrogen-based proteins and carbon-based constituents (Féret et al., 2017, 2021; Féret et al., 2008; Jacquemoud and Baret, 1990). Second, the estimation accuracy of physically-based approach varies among constituents and several inversion strategies have been proposed to improve the performance of model inversion, such as the incorporation of prior information, the selection of optimal spectral domains, and the application of ecological constraints (Combal et al., 2003; Darvishzadeh et al., 2008; Féret et al., 2021; Spafford et al., 2021; Yebra and Chuvieco, 2009). However, the prior information, optimal spectral domains and ecological constraints found or used in these studies often vary with datasets (Darvishzadeh et al., 2008; Jurdao et al., 2013; Yebra and Chuvieco, 2009). In other words, there is still no consensus on what and how inversion strategies should be applied, and it is unclear if these inversion strategies are transferable across datasets.

Finally, intensive field measurements of leaf DHR and DHT spectra are usually inaccessible due to the high cost of the integrating sphere and long measurement time. A new double integrating sphere provides a more efficient way of measuring DHR and DHT, but there is a compromise between measurement time and data accuracy (Hovi et al., 2018; Möttus et al., 2017). Leaf bi-directional reflectance factor (BRF) measured using a leaf contact probe can be an alternative with lower cost, higher signal-to-noise ratio and easier portability during field data collection (Li et al., 2018, 2019; Sims and Gamon, 2002; Yang et al., 2016). BRF may diverge from DHR (Bousquet et al., 2005; Jay et al., 2016; Li et al., 2018, 2019; Maccioni et al., 2001), because leaf surface may include waxes or trichomes interacting with light and resulting in

directional effects such as specular reflectance or enhanced scattering. Several studies have shown that a replacement of DHR with BRF in the PROSPECT model inversion can lead to a poor accuracy in trait estimations (Li and Wang, 2011; Ma et al., 2012). To solve this problem, Jay et al. (2016) developed the COSINE (CLOSe-range Spectral Imaging of Leaves) model to relate leaf DHR to BRF. By coupling PROSPECT and COSINE (the coupled model is called ‘PROCOSINE’), previous studies have showed that foliar traits (C_{ab} , EWT and LMA) in crops (e.g., wheat and rice) can be estimated from leaf BRF with a high accuracy (Fu et al., 2020; Li et al., 2018, 2019; Wang et al., 2021). However, it is unclear if PROCOSINE can estimate foliar traits across more diverse plant species.

The abovementioned facts make it challenging for ecologists to apply the physically-based approach to estimate foliar traits. In this sense, empirical approaches represent interesting alternative methods for trait estimation due to the ease of implementation and no limitation for foliar traits (Burnett et al., 2021; Serbin et al., 2019; Wang et al., 2019). The empirically-based approach aims to establish statistical models between foliar traits and leaf spectra, which include ordinary least squares regression (OLSR) (Féret et al., 2011; Li et al., 2019; Wang et al., 2016), stepwise linear regression (Grossman et al., 1996; Wang et al., 2015b), partial least squares regression (PLSR) (Martin et al., 2008; Nakaji et al., 2019; Wang et al., 2020), and machine learning algorithms such as supporting vector machine (SVM), gaussian process regression (GPR), and convolutional neural network (CNN) (Féret et al., 2019; Pullanagari et al., 2021; Verrelst et al., 2012). However, the empirically-based approach is often criticized for its poor transferability across sites, species and dates (Nakaji et al., 2019; Yang et al., 2016), because the models are driven by the training data. Serbin et al. (2019) showed that LMA could be predicted across different biomes with a high accuracy when building statistical models with the data collected from a wide range of sites, species and dates. However, whether such an approach applies to other foliar traits is still unclear.

Therefore, a comprehensive comparison of physically- and empirically-based approaches is needed to evaluate the model performance and transferability in predicting foliar traits, which will provide helpful guidance for selecting the optimal approach. In this study, we aim to assess the generality of leaf spectroscopic models for predicting key foliar functional traits from leaf BRF measurements with both physically- and empirically-based approaches. We collected two extensive datasets from 20 temperate and subtropical sites in eastern United States (US) and four subtropical and tropical sites in south China (CN). US and CN datasets represented foliar samples from different geographic areas (North America vs. East Asia), plant functional types (deciduous vs. evergreen) and plant species. These samples covered a wide range of geographic areas, species, light conditions, and growth periods. Thus, the two datasets provided a great opportunity to evaluate the generality of leaf spectroscopic models, i.e., the model performance when applied to an independent dataset (Martin et al., 2008; Serbin et al., 2019). Our specific objectives are:

- (1) to test the applicability of the physically-based approach, i.e., coupled PROSPECT and COSINE, in predicting foliar traits using leaf bidirectional reflectance;
- (2) to evaluate the ability of various inversion strategies in improving leaf trait estimations and their transferability across datasets;
- (3) to assess the model performance and transferability of the empirically-based approach.

2. Materials and methods

2.1. Study sites and field sampling

Two datasets of foliar samples were collected from a wide range of geographic areas, species, light conditions and growth periods. One was from the subtropical, temperate forest and grass ecosystems in eastern United States (hereafter referred to as “US”), and the other from the

subtropical and tropical forests in south China (hereafter referred to as “CN”). Details of the sample sizes, plant species, sampling dates and field sites are summarized in Tables 1 and S1 – S2. The sampling locations of each dataset are shown in Fig. 1.

The US dataset is composed of foliar samples from 19 field sites across seven NEON domains and foliar samples from Madison, Wisconsin (Chlus, 2020; Wang et al., 2020) (Fig. 1). NEON (National Ecological Observation Network) is a continental-scale observation facility collecting open access ecological data for monitoring ecosystem changes and responses to environment (Kampe et al., 2010; Schimel et al., 2007). The NEON data were collected during the peak growth season of 2017. Foliar samples included mature leaves from both the sunlit top and the lowest reachable shade branches of each individual tree or shrub, as well as leaves from grasses and forbs. The Madison subset was collected from May to November in 2016 to capture the seasonal variations in foliar traits. Sunlit and shade foliar samples were collected from over 100 species such as broadleaf trees, graminoids, forbs and vine species.

The CN dataset was collected from four subtropical and tropical forests sites located within three provinces in south China, including Guangxi, Guangdong and Hainan (Fig. 1). Field work was conducted from July to October in 2020, and from May to July in 2021 (Table 1). Samples were mainly from broadleaf trees with a small number of shrubs (21 out of the 360 samples). Leaf samples were collected from sunlit top of canopy branches of trees for all sites except the Guangxi site. At this site, samples were collected from the top, middle and bottom of the canopy to capture the vertical profile of foliar traits.

Fresh leaf spectra were measured on foliar samples using an ASD FieldSpec 3 spectrometer (ASD Inc., Boulder, USA) coupled with a plant contact probe with an external light source. The instrument was first optimized, and dark current was automatically corrected. A spectralon 99% white reference was then measured to obtain the leaf bidirectional reflectance factor (BRF). Ten readings were averaged per measurement (of sample and white reference). The integrating time was set to 1 s. One spectrum was made on one place of the adaxial side of each leaf with a black background by avoiding the main veins. The leaf samples were measured onsite within three hours upon collection. The measurement protocol was consistent for all field campaigns. After leaf BRF measurements, each fresh leaf was weighted for fresh weight (FW, g) using a digital scale (precision 0.001 g) and was scanned for leaf area (LA, cm²) using a flatbed scanner (Epson, Nagano, Japan). Then, the foliar samples were dried at 65 °C for at least 72 h and were measured for dry weight (DW, g). Equivalent water thickness (EWT, mg/cm²) was calculated as (FW-DW)/LA. Leaf mass per area (LMA, mg/cm²) was calculated as DW/LA.

A subset of foliar samples was flash frozen in the field in liquid

Table 1

Summary of the field sites, sampling dates, and number of samples and species. CN: south China; US: eastern United States. The values in brackets indicate the number of samples and species measured for equivalent water thickness (EWT) and leaf mass per area (LMA) in the US extensive dataset. The NEON site codes and names are listed in Fig. 1. The coordinate, climate, vegetation type of each site are listed in Table S1. The sampled species of each dataset are listed in Table S2.

Dataset	Field site	Sampling dates	Number of samples	Number of species
CN	Heshan	Sep.-Oct. 2020; May 2021	196	34
	Liuzhou	Jul. 2020	39	2
	Shenzhen	Nov. 2021	24	18
	Wanning	Jul. 2021	101	53
	Total		360	97
US	19 NEON sites	May-Oct. 2017	111 (3498)	76 (178)
	Madison	May-Nov. 2016	273	121
Total			384 (3498)	186 (178)

nitrogen, and was later analyzed for pigments including leaf chlorophyll *a* + *b* content per surface unit (C_{ab} , $\mu\text{g}/\text{cm}^2$) and carotenoid content per surface unit (C_{xc} , $\mu\text{g}/\text{cm}^2$) using the high-performance liquid chromatography (HPLC, Agilent 1200 Series; Agilent Technologies) (Kothari et al., 2018; Schweiger et al., 2018). Another subset of foliar samples was oven-dried, ground, and then sent for chemistry analysis of nitrogen concentration (N_{mass} , mg/g). Nitrogen content (N_{area} , g/m²) was calculated using LMA ($N_{\text{area}} = N_{\text{mass}} \times \text{LMA}$). It should be noted that the pigment analysis of the CN samples was different from that of the US samples. In the case of CN, five leaf discs ($\sim 1.414 \text{ cm}^2$) were obtained from a leaf sample and flash frozen in the field by liquid nitrogen. In laboratory, the leaf discs were ground in liquid nitrogen and incubated in 95% ethanol (v/v) to determine C_{ab} and C_{xc} using the equations in Lichtenthaler (1987).

As summarized in Table 1, the NEON dataset includes 3498 leaves measured for EWT and LMA, 111 samples for foliar pigments and nitrogen. In the Madison dataset, 273 leaves were collected for EWT, LMA, foliar pigments and nitrogen. In the CN dataset, 360 foliar samples were collected for LMA, EWT, nitrogen and foliar pigments.

2.2. The physically-based approach

2.2.1. The PROCOSINE model inversion

The PROCOSINE model results from the coupling of PROSPECT and COSINE models. Here, we used two versions of the PROSPECT model. The PROSPECT-D model (Féret et al., 2017) simulates DHR and DHT from a set of leaf biophysical descriptors, including a refractive index, the leaf structure parameter (N_{struct}), C_{ab} , C_{xc} , EWT, LMA, the leaf anthocyanin content (C_{ant}), and their specific absorption coefficients. The refractive index and specific absorption coefficients are fixed values in the model. The PROSPECT-PRO model (Féret et al., 2021) simulates DHR and DHT from the same set of leaf biophysical descriptors, except that LMA is divided into two distinct constituents, the leaf protein content (PROT) and leaf carbon-based constituents (CBC), which combine cellulose, lignin, and other structural and non-structural carbohydrates. Then the DHR simulated with PROSPECT-D and PROSPECT-PRO over the visible to shortwave infrared (VSWIR) domain from 400 nm to 2500 nm is expressed as in Eqs. (1) and (2).

$$\text{DHR}_{\text{sim,D}} = \text{PROSPECT} - \text{D}(N_{\text{struct}}, C_{ab}, C_{xc}, \text{EWT}, \text{LMA}) \quad (1)$$

$$\text{DHR}_{\text{sim,PRO}} = \text{PROSPECT} - \text{PRO}(N_{\text{struct}}, C_{ab}, C_{xc}, \text{EWT}, \text{PROT}, \text{CBC}) \quad (2)$$

where $\text{DHR}_{\text{sim,D}}$ is the DHR simulated with PROSPECT-D, $\text{DHR}_{\text{sim,PRO}}$ is the DHR simulated with PROSPECT-PRO, N_{struct} , C_{ab} , C_{xc} , EWT, LMA, PROT and CBC are the leaf structure parameter (unitless), leaf chlorophyll *a* + *b* content ($\mu\text{g}/\text{cm}^2$), leaf carotenoids content ($\mu\text{g}/\text{cm}^2$), leaf dry matter per area (mg/cm^2), leaf protein content (mg/cm^2) and leaf carbon-based constituents content (mg/cm^2). Specifically, PROSPECT-D was used to estimate C_{ab} , C_{xc} , EWT and LMA, and PROSPECT-PRO was used for predicting PROT.

PROCOSINE simulated BRF from DHR simulated with any version of PROSPECT, and three additional parameters corresponding to leaf orientation and a specular term (Jay et al., 2016; Li et al., 2018) (Eq. (3)):

$$\text{BRF}_{\text{sim}} = \frac{\cos\theta_i}{\cos\theta_s} (\text{DHR}_{\text{sim}} + b_{\text{spec}}) \quad (3)$$

where θ_i and θ_s are the light incident angle (angle between the light source and the normal to the leaf), and illumination zenith angle, respectively, and b_{spec} is a wavelength-independent specular term (unitless) corresponding to the difference between leaf BRF and DHR (Bousquet et al., 2005; Jay et al., 2016; Li et al., 2018).

The inversion of PROCOSINE involves the optimization of model parameter vector θ by minimizing the difference between the measured and modeled leaf BRF (Eq. (4)):

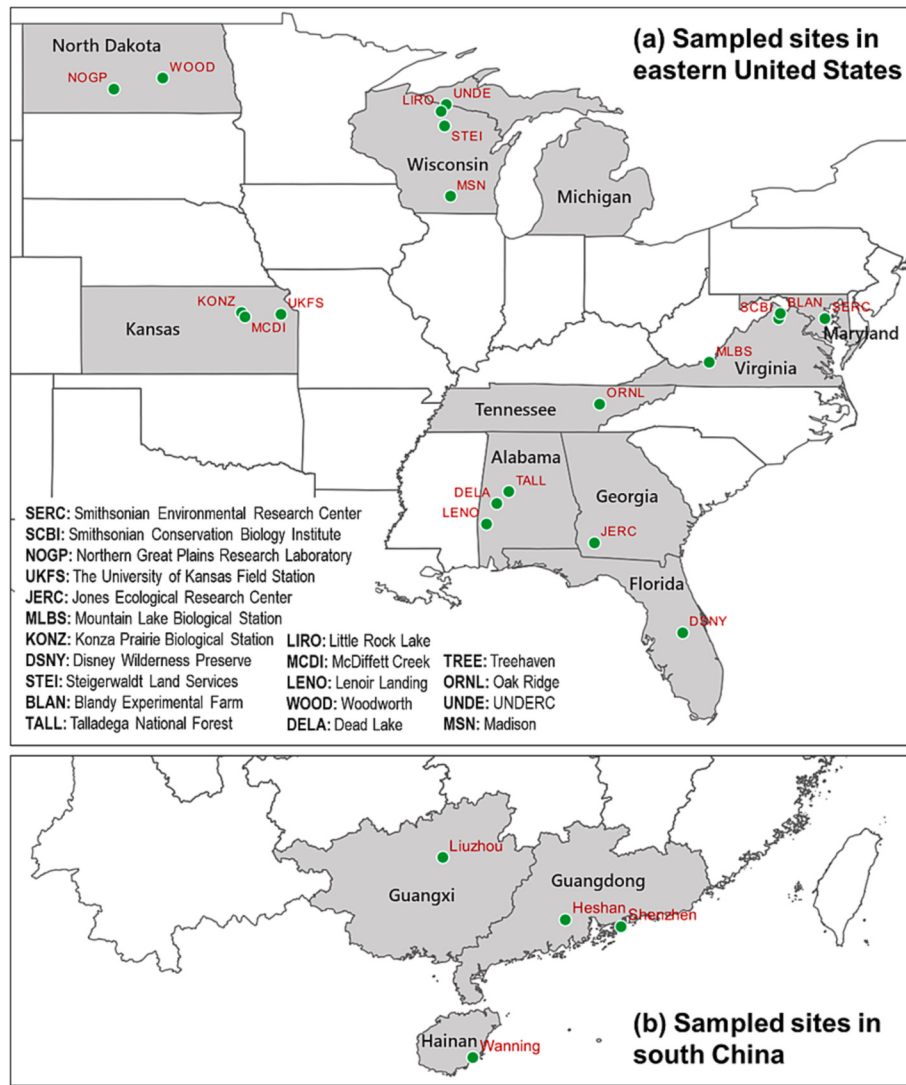


Fig. 1. Location of field sites with sample collection: (a) eastern United States; (b) south China.

$$J(\theta) = \sum_{\lambda=\lambda_1}^{\lambda_n} (BRF_{mes}(\lambda) - BRF_{sim}(\lambda, \theta))^2 \quad (4)$$

where the model parameter vector θ combines the set of leaf biophysical descriptors corresponding to the PROSPECT version in use and b_{spec} . λ_1 and λ_n represent the wavelength range used for model inversions; $BRF_{mes}(\lambda)$ is the bidirectional reflectance factor measured at wavelength λ ; $BRF_{sim}(\lambda, \theta)$ is the leaf bidirectional reflectance factor at wavelength λ , which is simulated by PROCOSINE with model parameters θ (Eqs. (1)–(3)). Here we used a simplified version of PROCOSINE, as the geometry

of acquisition defined by θ_i and θ_s was set to 15° for all leaves.

The Matlab code of PROSPECT-PRO was downloaded from https://gitlab.com/jbferet/prospect_pro_matlab. The optimization was performed using the function *fminsearchbnd* in Matlab (The MathWorks, Inc.). The optimization range of foliar traits (C_{ab} , C_{xc} , EWT, LMA, PROT, and CBC) was determined by in situ measurements (see Table 2). To estimate leaf nitrogen content (N_{area} , mg/cm²), a scale factor of 4.43 was applied to leaf protein content (PROT): $N_{area} = PROT/4.43$ (Yeoh and Wee, 1994).

Table 2

The range of each leaf parameter used in the coupled PROSPECT (-D or -PRO) and COSINE.

Parameter	Abbreviation	Unit	Initial value	Minimum	Maximum (US)	Maximum (CN)
Leaf structure	Nstruct	–	1.5	0.5	3.5	3.5
Chlorophyll $a + b$	C_{ab}	μg/cm ²	40	0.5	120	90
Carotenoids	C_{xc}	μg/cm ²	10	0.5	25	20
Anthocyanin	C_{anth}	μg/cm ²	0.1	0	20	20
Equivalent water thickness	EWT	mg/cm ²	10	1	60	30
Leaf mass per area	LMA	mg/cm ²	10	1	30	30
Protein	PROT	mg/cm ²	1	0.1	3	3
Carbon-based constituents	CBC	mg/cm ²	9	0.9	27	27
Specular term	b_{spec}	–	0.2	–0.2	0.6	0.6

2.2.2. Iterative optimization: inversion strategies

Various inversion strategies have been proposed to reduce the uncertainty in parameter estimations such as prior information, the selection of optimal spectral domain, and ecological constraints using empirical relationships to eliminate unrealistic combinations of parameters (Combal et al., 2003; Darvishzadeh et al., 2008; Jurdao et al., 2013; Yebra and Chuvieco, 2009). Here, we assessed the performance of three inversion strategies (hereafter referred to *IS1*–*3*):

- (1) using the full spectrum without any prior information;
- (2) using the optimal spectral domains reported by previous studies (Féret et al., 2019, 2021; Spafford et al., 2021) with prior information on N_{struct} and b_{spec} ;
- (3) selecting optimal non-contiguous spectral domains from US and CN datasets following the procedure described in Féret et al. (2021) with prior information on N_{struct} and b_{spec} .

It should be noted that we did not identify optimal spectral domains by using the whole US dataset when testing inversion strategy *IS3*, due to the high computation time. Instead, we applied inversion strategy *IS3* on a subset of US foliar samples ($n = 384$) with concurrent measurements of ten foliar traits. Then, we applied the selected optimal spectral regions to predict EWT and LMA for the US extensive dataset with 3498 measurements of EWT and LMA. In this way, we could assess if the optimal spectral domains obtained from a subset of the dataset can be applied to the whole dataset.

In *IS1*, the full spectrum was simply applied to the cost function to estimate the foliar traits, that is, $\lambda_1 = 400$ nm and $\lambda_n = 2500$ nm in Eq. (4). In *IS2*, the optimal spectral domains identified by Féret et al. (2021, 2019) and Spafford et al. (2021) were applied to estimate the foliar traits. They were 1700–2400 nm for EWT and LMA (i.e., $\lambda_1 = 1700$ nm, $\lambda_n = 2400$ nm), 2100–2139 nm and 2160–2179 nm for leaf protein, 700–720 nm for C_{ab} , and 520–560 nm for C_{xc} .

In *IS3*, we used a sequential forward feature selection (SFS) technique to identify the non-contiguous optimum spectral domains (Kudo and Sklansky, 2000; Marcano-Cedeno et al., 2010). This method started with an empty feature set and sequentially added features which generated the minimum RMSE between the trait estimations and the measurements (Féret et al., 2021). We created 20 evenly-sized spectral features of 50 nm from 1400 nm to 2399 nm for EWT, LMA and N_{area} , and 17 spectral features of 20 nm from 460 nm to 799 nm for C_{ab} and C_{xc} . The root mean squared error (RMSE) between the predicted and measured traits was calculated to assess the model performance, and the minimum RMSE was used to identify the optimal spectral domains. Take EWT as an example, we firstly performed the model inversion using each of 20 features, and searched for the one which generated the minimum RMSE between the measured and the predicted EWT. Then we identified the next spectral feature among the remaining 19 features by combining the previously identified feature, which led to the minimum RMSE with two features. The procedure continued until all features were added and the full spectrum of 1400–2399 nm was used for model inversion. The RMSE for each set of features was calculated. The order of spectral feature added to the optimal non-contiguous spectral domains was recorded.

We performed inversion of PROSPECT and PROCOSINE to predict the five foliar traits (EWT, LMA, N_{area} , C_{ab} , C_{xc}), and compared the retrieval performance of the two models in order to test the contribution of additional specular term and leaf orientation in PROCOSINE. In PROSPECT inversion, the prior information on the leaf structure was obtained using an empirical relationship built on the DHR at 1131 nm following Spafford et al. (2021). For PROCOSINE, the approach may not be applicable to BRF due to the differences between DHR and BRF. We found that EWT and LMA could be estimated with moderate accuracies using the full spectrum without prior information (*IS1*). Thus, we assumed that other parameters such as leaf structure and b_{spec} were also well retrieved. In PROCOSINE, we used the N_{struct} and b_{spec} estimated

from *IS1* as prior information for ease of operation.

In addition, we applied the optimal non-contiguous spectral domains (*IS3*) obtained from one dataset to the other dataset to evaluate the model performance and transferability of inversion strategies across datasets.

2.3. Partial least squares regression

Partial least squares regression (PLSR) transforms the original data to a smaller number of orthogonal latent vectors and maximizes the correlation between the response variables and the predictor variables (Wold et al., 2001). It deals with the problem of multicollinearity inherent in hyperspectral data and has been widely used to estimate foliar functional traits using leaf spectroscopy or imaging spectroscopy (Asner et al., 2015; Serbin et al., 2019; Wang et al., 2020; Yang et al., 2016). To select the most informative spectral bands, we performed one of the feature selection methods in PLSR, the interval PLSR (iPLSR) (Mehmood et al., 2012; Nørgaard et al., 2000). iPLSR is one of the wrapper methods in PLSR, which is similar to the sequential forward feature selection approach used in RTM inversion.

Similar to Section 2.2.2, we created evenly-sized spectral features for EWT, LMA, N_{area} , C_{ab} and C_{xc} . For each trait, we firstly built iPLSR models using each of the features, and searched for the one which generated the minimum RMSE between the trait measurements and predictions. Then we identified the next spectral feature among the remaining features by combining the previously identified one, which led to the minimum RMSE with two features. The procedure continued until all features were added. The RMSE for each set of features was calculated, and the minimum RMSE was used to identify the optimal spectral domains. For comparison, we also built PLSR models with the full spectrum (400–2400 nm).

To avoid model overfitting, the number of latent vectors was determined by minimizing the prediction residual sum of squares (PRESS) statistic through 200 70/30 jack-knifed splits of the calibration dataset (Chen et al., 2004). In this study, we built iPLSR and PLSR models between foliar traits and leaf spectra to explore the generality of empirically-based approaches.

First, we evaluated the performance of models within each dataset. That is, iPLSR and PLSR models were built on the CN or US dataset. The original dataset was randomly split with 70% for model calibration and 30% for model validation. Within the calibration subset, 70% of the samples were randomly selected to generate a model. To minimize the effect of random sampling on model calibration, we repeated the random sampling procedure for 200 times and thus generated 200 models.

Second, we evaluated the transferability of iPLSR and PLSR models across different datasets. That is, models were calibrated on the US (or CN) dataset and validated on the CN (or US) dataset. In this scenario, the calibration dataset was one of the full datasets (US or CN). Similarly, 200 models were generated by randomly sampling 70% of the calibration dataset for 200 times.

Finally, the 200 models were applied to the validation subset. The validation subset was the 30% withheld data of each dataset in the first scenario, and one of the datasets (CN or US) in the second scenario. The average of the resultant 200 predictions was used as an estimate of foliar traits. For iPLSR, the selected optimal spectral domains were used to evaluate the informative spectral bands. For PLSR, the variable importance of projection (VIP) was calculated to evaluate the contribution of each wavelength to the trait prediction (Wold et al., 2001).

2.4. Model evaluation

Four statistics, including the coefficient of determination (R^2), the root mean squared error (RMSE), the normalized RMSE (NRMSE = RMSE/mean), and the bias (BIAS) between model predictions and field measurements, were calculated to evaluate the performance of the

physically- and empirically-based approaches. BIAS was calculated as the difference between the averages of predictions and measurements. A positive value of BIAS often indicates over-estimation of the model predictions, and a negative value of BIAS means under-estimation.

To evaluate the model transferability, we calculated the relative RMSE difference,

$$RMSE_{diff} = \frac{(RMSE_{across} - RMSE_{within})}{RMSE_{within}} * 100\% \quad (5)$$

where $RMSE_{within}$ is the RMSE when cross-validation is performed within a dataset, $RMSE_{across}$ is the RMSE when models are applied across datasets (e.g., models developed on US were applied to CN).

3. Results

3.1. Statistics of the foliar functional traits

The statistics and distribution patterns of the foliar functional traits from CN and US datasets are shown in Table 3, as well as Figs. 2 and S1 - S2. All foliar traits in US dataset showed similar to broader range in terms of content than in CN dataset (Table 3). The range measured for LMA was similar across CN and US datasets, but the mean LMA in CN was higher, which may be explained by ecological factors and strategies, as CN dataset was mainly composed of evergreen broadleaf species. The EWT range in US was broader than that in CN due to the presence of species such as cattail (*Typha angustifolia*) and yellow iris (*Iris pseudacorus*) with particularly high EWT ($> 30 \text{ mg/cm}^2$). The mean of N_{area} was higher in CN (0.18 mg/cm^2) than in US (0.16 mg/cm^2). The distribution and mean values of C_{ab} in CN and US were similar. The mean value of C_{xc} in CN ($6.30 \text{ } \mu\text{g/cm}^2$) was lower than in US ($8.17 \text{ } \mu\text{g/cm}^2$).

3.2. The variability of leaf bidirectional reflectance factor

The mean and standard deviation of the leaf BRF for two datasets are shown in Figs. 3 and S3. The mean reflectance of US in the visible spectral region was much higher than that of CN, which was attributed to a lower chlorophyll content in young and senesced leaves in the Madison subset. The mean reflectance in the near-infrared region was higher in CN than US due to more scattering caused by thicker leaves of evergreen broadleaf species in CN. The lower mean reflectance of CN in the shortwave infrared region can be attributed to the absorption by higher EWT and LMA (Table 3). The standard deviation of leaf reflectance was similar for the two datasets.

3.3. The performance of PROSPECT and PROCOSINE

3.3.1. The optimal non-contiguous spectral domains

3.3.1.1. PROSPECT. The optimal non-contiguous spectral domains for predicting foliar traits (IS3) in CN and US datasets are listed in Fig. 4 and Table S3. The optimal spectral domains to estimate EWT and LMA used

in PROSPECT model inversion were mostly overlapped for CN and US (Fig. 4). The optimal spectral domains to predict N_{area} were the same for CN and US, which was 2100–2149 nm. For pigments, the optimal spectral domains for the two datasets were similar sharing the red-edge spectral segments of 740–759 nm and 760–779 nm.

3.3.1.2. PROCOSINE. The optimal non-contiguous spectral domains for the estimation of EWT and LMA with PROCOSINE largely differed from those obtained with PROSPECT (-D or -PRO), but were quite consistent for N_{area} and pigments (Fig. 4). The optimal spectral regions to predict EWT for CN and US datasets only shared one spectral segment of 2050–2099 nm. For LMA, the optimal spectral regions for the two datasets had the common spectral segments of 2300–2349, 2250–2299, and 2350–2399 nm. The most accurate prediction of N_{area} for US was obtained by six segments of 50 nm starting with 2050, 2150, 2100 and 2250 nm. For CN, the optimal subdomains only included one segments starting with 2100 nm. The optimal spectral domains to predict C_{ab} were similar for CN and US, both including the red-edge spectral region of 520–559 nm. When predicting C_{xc} , the optimal spectral regions were also similar for the two datasets sharing the spectral segment of 700–779 nm.

3.3.2. Model performance within each individual dataset

3.3.2.1. PROSPECT. Both EWT and LMA were accurately estimated by inverting PROSPECT with the full spectrum ($RMSE = 2.37\text{--}4.50 \text{ mg/cm}^2$, Table 4). IS2 led to improved results compared to IS1 when using BRF. The accuracy was further improved when using the optimal non-contiguous spectral domains identified in this study (Fig. 5).

EWT was more accurately predicted in CN than in US. This could be largely explained by the presence of samples corresponding to forb species in the US dataset, with EWT higher than 25.0 mg/cm^2 (e.g., cattail, *Typha angustifolia*; rattlesnake master, *Eryngium yuccifolium*). With these samples being removed, the estimation accuracy of EWT in US was greatly improved with RMSE decreasing from 3.62 mg/cm^2 to 2.69 mg/cm^2 . For LMA, an underestimation was found for both CN and US when performing IS1 (Table S4). Selecting optimal non-contiguous spectral regions (IS3) was the best inversion strategy for LMA (Table 4).

Both the selection of optimal spectral domains and the use of prior information significantly improved the estimation of N_{area} . The RMSE of N_{area} reduced from 0.350 to 0.058 mg/cm^2 for CN and from 0.320 to 0.049 mg/cm^2 for US, respectively (Table 4).

C_{ab} was most accurately estimated with the IS3 (Table 4). C_{ab} in US was more accurately predicted than that in CN. C_{xc} was poorly estimated for both datasets with any of the three inversion strategies.

3.3.2.2. PROCOSINE. The estimation accuracies of foliar traits using PROCOSINE were lower than those from PROSPECT (Table 4). Notably, the RMSE of LMA increased from 1.16 to 1.36 mg/cm^2 with PROSPECT to 1.58–2.24 mg/cm^2 with PROCOSINE. The inversion strategies yielding the most accurate estimations were similar for the two models.

Table 3

Statistics of the foliar functional traits in CN and US datasets used in this study. EWT: equivalent water thickness; LMA: leaf mass per area; C_{ab} : leaf chlorophyll $a + b$; C_{xc} : leaf carotenoids. SD: standard deviation; CV: coefficient of variation.

Dataset	Trait	Unit	Number of samples	Min	Max	Range	Mean	SD	CV
CN	EWT	mg/cm^2	359	5.50	26.43	20.93	13.91	4.06	0.29
	LMA	mg/cm^2	360	2.79	23.06	20.27	8.66	3.17	0.37
	N_{area}	mg/cm^2	357	0.07	0.38	0.31	0.18	0.06	0.36
	C_{ab}	$\mu\text{g/cm}^2$	172	6.94	80.70	73.76	36.54	14.13	0.39
	C_{xc}	$\mu\text{g/cm}^2$	172	1.11	13.20	12.09	6.30	2.30	0.37
US	EWT	mg/cm^2	3769	0.38	68.03	67.65	10.32	5.31	0.51
	LMA	mg/cm^2	3791	0.21	24.36	24.15	7.41	3.01	0.41
	N_{area}	mg/cm^2	294	0.03	0.39	0.36	0.16	0.07	0.43
	C_{ab}	$\mu\text{g/cm}^2$	371	0.78	109.11	108.33	39.97	18.95	0.47
	C_{xc}	$\mu\text{g/cm}^2$	372	0.89	20.95	20.95	8.17	3.15	0.39

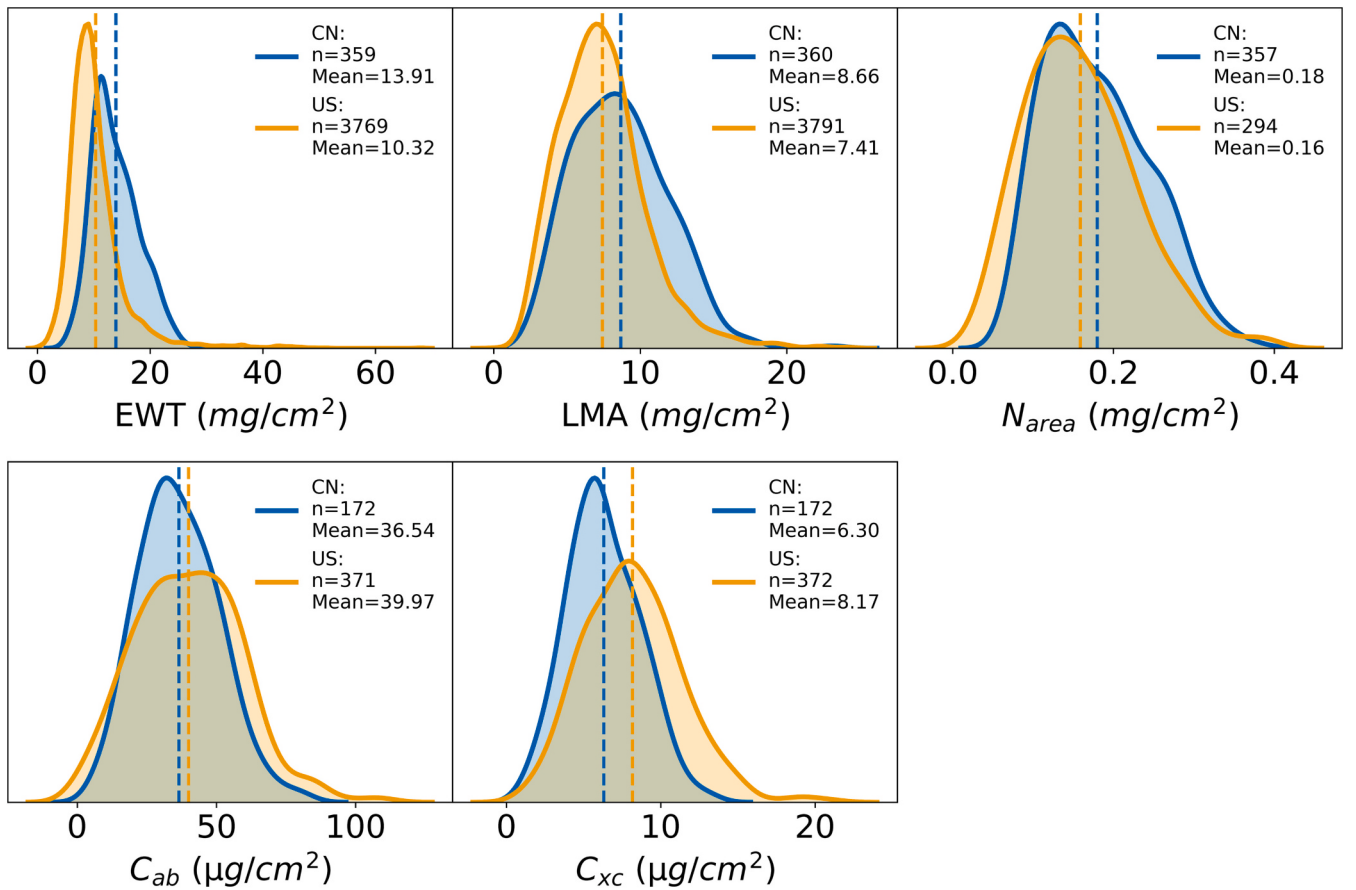


Fig. 2. Histogram and distribution density plot of the foliar traits in CN and US datasets. EWT: equivalent water thickness; LMA: leaf mass per area; C_{ab} : leaf chlorophyll $a + b$; C_{xc} : leaf carotenoids.

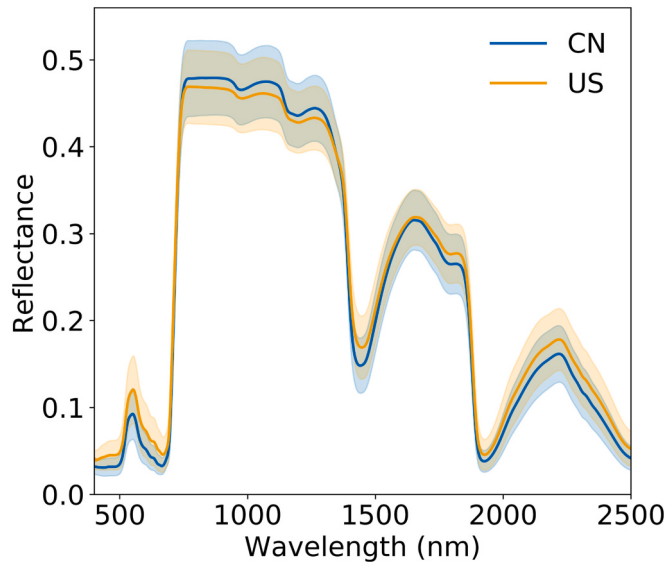


Fig. 3. The mean of leaf bi-directional reflectance factor (with standard deviation) for CN and US datasets.

The results of each dataset by light conditions (sun and shade leaves), and by growth periods (mature, young, and old leaves) were shown in Figs. S4 – S7. Foliar traits of the shade leaves were more accurately estimated than those of sun leaves in both CN and US (Figs. S4 – S5). With respect to growth periods, the estimation accuracies of foliar traits

were quite similar for young and mature leaves in CN (Fig. S6). In US, no clear patterns were found in the estimation accuracies of foliar traits for mature, young, and old leaves (Fig. S7).

3.3.3. Transferability of inversion strategies across datasets

3.3.3.1. PROSPECT. EWT and LMA in CN could be accurately estimated using the optimal spectral domains obtained from the US dataset (Table 5). The estimation accuracy was high for LMA (RMSE = 1.39 mg/cm²) but lower for EWT (RMSE = 3.62 mg/cm²) in US when applying the optimal spectral domains identified for the CN dataset.

N_{area} was accurately predicted both in CN and US (RMSE = 0.058 mg/cm² and 0.049 mg/cm², respectively) with the optimal spectral domains obtained from the counterpart, indicating a high transferability of the inversion strategy (Table 5). C_{ab} was less accurately estimated when applying the optimal spectral domains identified in one dataset to the other (Table 5, Figs. 6, S8 – S9). C_{xc} was poorly predicted for both datasets when using the optimal spectral domains identified by the counterpart.

We also applied the selected optimal spectral regions from the US dataset (Table S3) to the US extensive dataset with 3498 measurements of EWT and LMA (Table 6). The most accurate estimation of LMA was achieved using the optimum non-contiguous spectral regions (RMSE = 1.30 mg/cm²). However, EWT was most accurately predicted by *IS2* using the optimal spectral regions reported in previous studies (RMSE = 2.96 mg/cm²).

3.3.3.2. PROCOSINE. PROCOSINE resulted in less accurate estimations of foliar traits than PROSPECT when applying the optimal spectral domains identified in one dataset to the other (Table 5). LMA in CN was

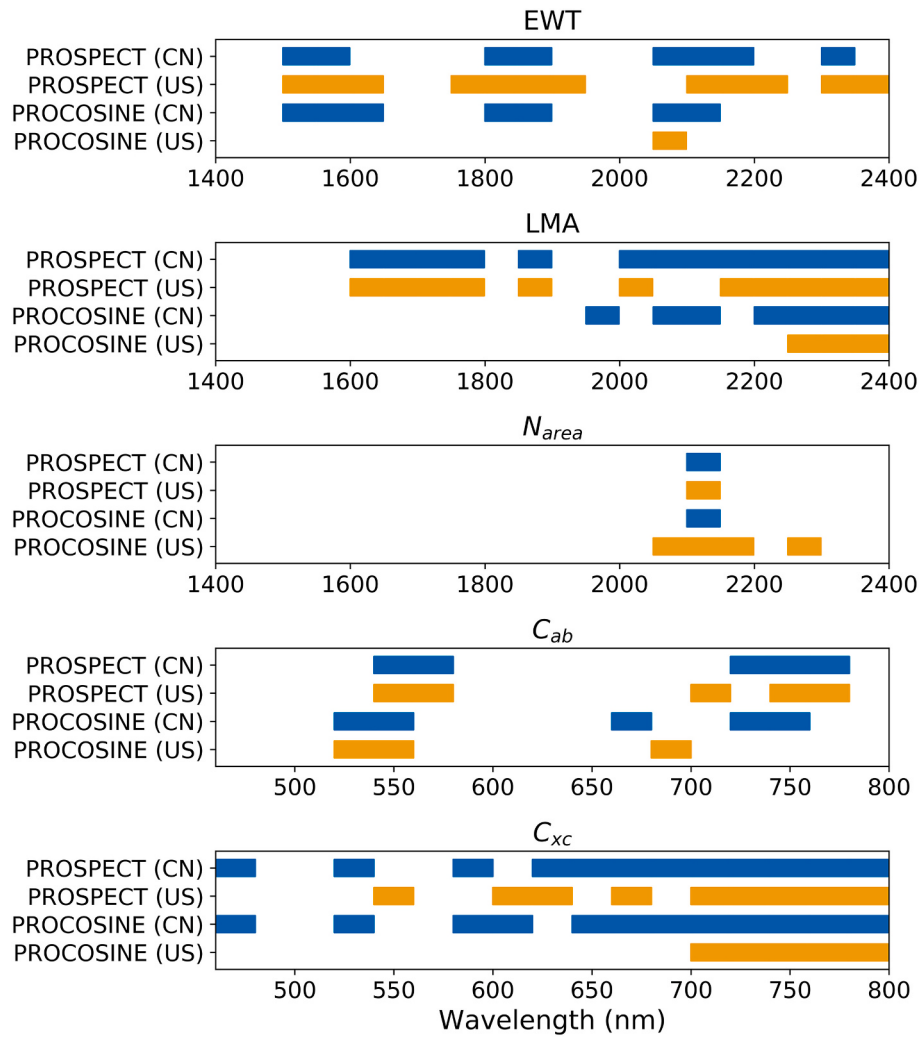


Fig. 4. The optimal non-contiguous spectral domains for predicting foliar traits in CN and US datasets used in model inversion of PROSPECT and PROCOSINE. EWT: equivalent water thickness; LMA: leaf mass per area; N_{area} : leaf nitrogen; C_{ab} : leaf chlorophyll $a + b$; C_{xc} : leaf carotenoids.

Table 4

The model performance of PROSPECT and PROCOSINE to predict foliar traits in CN and US datasets with three inversion strategies (*IS1–3*). *IS1*: uses the full spectrum; *IS2*: uses the optimal spectral regions as suggested in previous studies (Féret et al., 2019, 2021; Spafford et al., 2021); *IS3*: uses the optimal non-contiguous spectral regions identified in this study. Bold values are the most accurate estimation among the three inversion strategies. The values of NRMSE and BIAS are listed in Tables S4 and S5. EWT: equivalent water thickness; LMA: leaf mass per area; N_{area} : leaf nitrogen; C_{ab} : leaf chlorophyll $a + b$; C_{xc} : leaf carotenoids.

Trait	Inversion strategy	PROSPECT				PROCOSINE			
		CN dataset		US dataset		CN dataset		US dataset	
		R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
EWT (mg/cm^2)	<i>IS1</i>	0.56	4.02	0.64	4.50	0.62	3.89	0.68	4.58
	<i>IS2</i>	0.53	3.17	0.65	4.72	0.60	3.52	0.65	4.60
	<i>IS3</i>	0.74	2.18	0.77	3.62	0.66	2.82	0.69	4.19
LMA (mg/cm^2)	<i>IS1</i>	0.77	2.37	0.72	2.78	0.80	2.27	0.72	2.89
	<i>IS2</i>	0.85	1.24	0.82	1.44	0.81	1.96	0.74	2.51
	<i>IS3</i>	0.87	1.16	0.84	1.36	0.81	1.58	0.74	2.24
N_{area} (mg/cm^2)	<i>IS1</i>	0.29	0.35	0.43	0.32	0.19	0.34	0.33	0.30
	<i>IS2</i>	0.35	0.070	0.61	0.060	0.35	0.077	0.57	0.067
	<i>IS3</i>	0.50	0.058	0.63	0.049	0.47	0.067	0.48	0.069
C_{ab} ($\mu\text{g}/\text{cm}^2$)	<i>IS1</i>	0.21	25.13	0.60	13.12	0.25	25.15	0.64	13.15
	<i>IS2</i>	0.36	22.92	0.71	11.74	0.27	25.71	0.61	15.09
	<i>IS3</i>	0.33	13.13	0.69	10.92	0.14	18.49	0.48	16.13
C_{xc} ($\mu\text{g}/\text{cm}^2$)	<i>IS1</i>	0.14	8.81	0.13	6.05	0.15	9.29	0.19	7.15
	<i>IS2</i>	0.17	11.81	0.18	6.97	0.29	12.31	0.27	8.50
	<i>IS3</i>	0.07	3.47	0.14	5.31	0.04	4.52	0.21	5.72

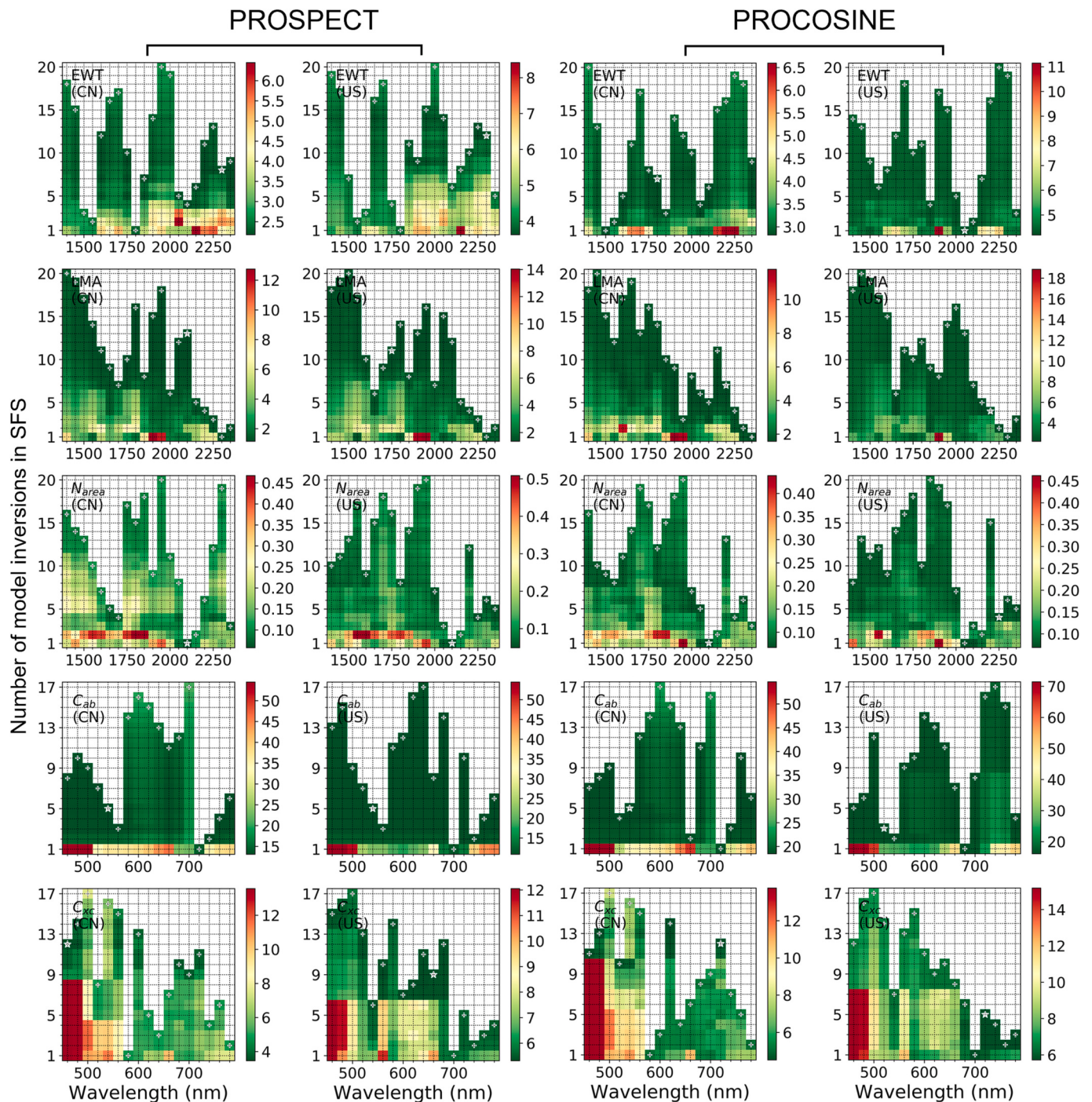


Fig. 5. RMSE between the measured and predicted foliar traits from leaf bidirectional reflectance factor with PROSPECT and PROCOSINE. The X axis is the wavelength added to the feature set in the selection of optimal non-contiguous spectral regions (IS3). The Y axis is the number of model inversions in the sequential forward feature selection (SFS). The white “+” symbol indicates the wavelength added to the feature set which leads to minimum RMSE in each run. The white star symbol indicates the wavelength which leads to minimum RMSE among all runs. EWT: equivalent water thickness; LMA: leaf mass per area; N_{area} : leaf nitrogen; C_{ab} : leaf chlorophyll $a + b$; C_{xc} : leaf carotenoids.

predicted with moderate accuracy ($RMSE = 1.56 \text{ mg/cm}^2$) with the optimal spectral domains obtained from the US dataset, while LMA in US was poorly estimated ($RMSE = 2.76 \text{ mg/cm}^2$) with the optimal spectral domains identified by the CN dataset. Pigments were predicted with poor accuracies for both datasets when using the optimal spectral domains identified by their counterpart.

EWT and LMA were retrieved with lower accuracies, when applying the optimal spectral regions identified for the US dataset to the US extensive dataset using PROCOSINE compared to PROSPECT (Table 6).

3.4. The performance of the empirically-based approach

3.4.1. Model performance within each individual dataset

Both the iPLSR and PLSR model accurately estimated the majority of foliar traits when it was cross-validated within the same dataset (Tables 7 and S8), and the accuracies were higher than those of PROSPECT and PROCOSINE (Table 4). iPLSR achieved higher accuracies than PLSR (Tables 7 and S8). EWT and LMA were most accurately predicted, followed by N_{area} , C_{ab} and C_{xc} .

Table 5

The cross-dataset validation of PROSPECT and PROCOSINE to predict foliar traits using the optimal non-contiguous spectral regions (IS3) identified in each dataset (Table S3). US $\rightarrow$ CN denotes the prediction accuracy statistics for the CN dataset using the optimal spectral domains obtained from US. CN $\rightarrow$ US denotes the prediction accuracy statistics for the US dataset using the optimal spectral domains obtained from CN. The values of NRMSE and BIAS are listed in Tables S6 and S7. EWT: equivalent water thickness; LMA: leaf mass per area; C_{ab} : leaf chlorophyll $a + b$; C_{xc} : leaf carotenoids; N_{area} : leaf nitrogen.

Trait	PROSPECT				PROCOSINE			
	US $\rightarrow$ CN		CN $\rightarrow$ US		US $\rightarrow$ CN		CN $\rightarrow$ US	
	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
EWT (mg/cm ²)	0.73	2.27	0.78	3.62	0.62	3.06	0.66	4.40
LMA (mg/cm ²)	0.88	1.22	0.83	1.39	0.83	1.56	0.67	2.76
N_{area} (mg/cm ²)	0.50	0.058	0.63	0.049	0.33	0.071	0.55	0.074
C_{ab} (μ g/cm ²)	0.35	19.42	0.60	15.21	0.19	24.70	0.38	19.24
C_{xc} (μ g/cm ²)	0.14	3.71	0.10	5.12	0.24	5.47	0.02	6.53

3.4.2. Model performance across datasets

When the iPLSR and PLSR model was validated across datasets, the estimation accuracies decreased (Tables 7–9, S9). The model built on US generally performed better than that built on CN (Table 8 and Figs. 7, S10–S15). The trait model built on US worked well for predicting LMA in CN, resulted in systematic negative bias for the estimation EWT, estimated C_{ab} and C_{xc} with moderate accuracy, but accurately predicted N_{area} .

The trait model built on CN could only predict EWT and LMA in US with high accuracies (Table 8 and Fig. 7). The LMA in US was estimated with $R^2 = 0.82$ and RMSE = 1.60 mg/cm². Since the EWT in CN ranged from 5.50 to 26.43 mg/cm², the EWT in US was underestimated for the measurements higher than 25.0 mg/cm². N_{area} was moderately estimated, and pigments were poorly estimated when transferring the models from CN to US. This is likely due to the fact that the trait range in CN is smaller than that in US (Table 3).

3.4.3. The feature importance of iPLSR and PLSR

For iPLSR, the optimal spectral domains for predicting foliar traits in CN and US datasets were largely inconsistent (Fig. 8). For N_{area} , the optimal spectral domains for the two datasets were quite similar sharing the spectral segments of 1650–1800 nm, 2050–2100 nm, 2200–2250

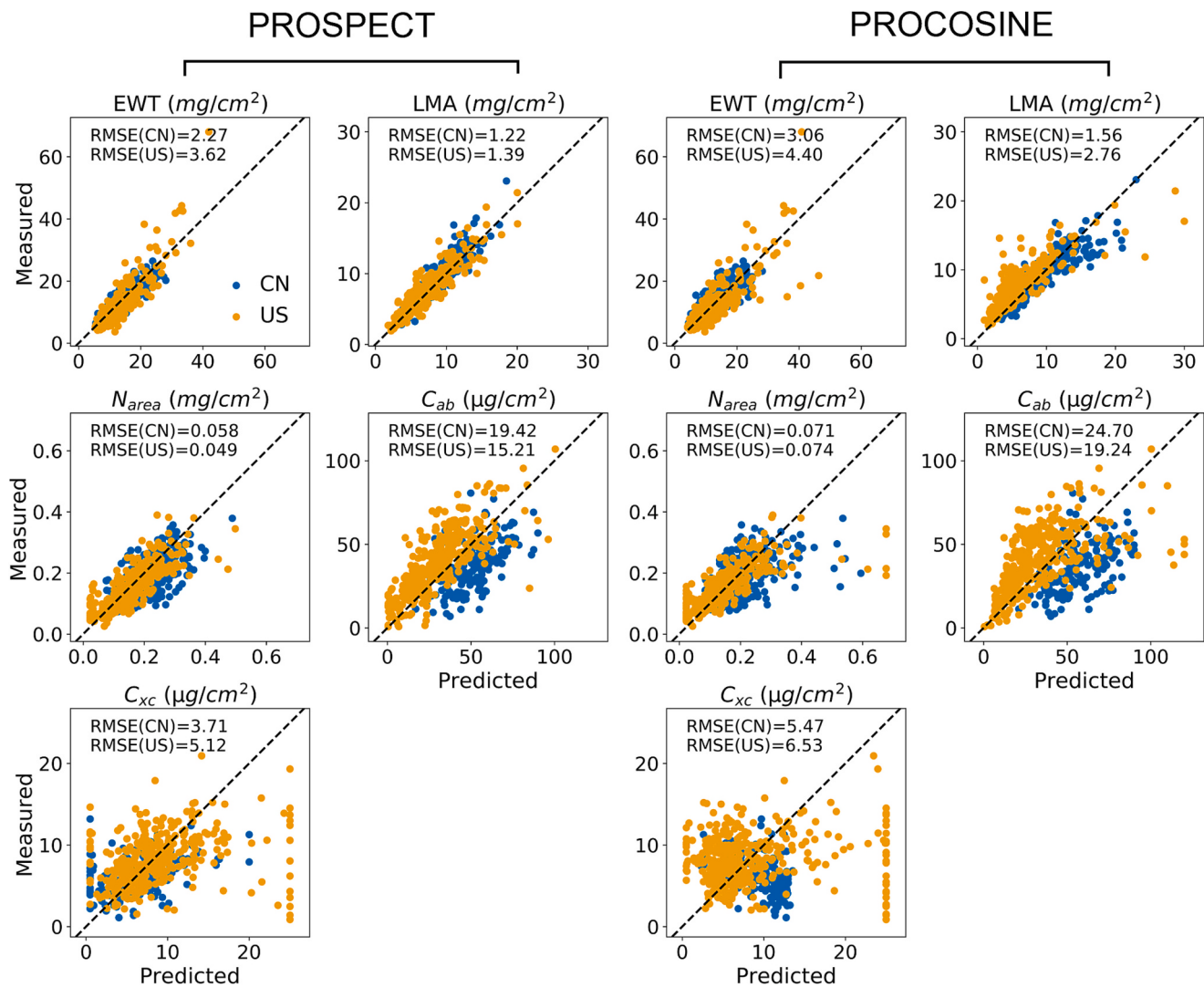


Fig. 6. Measured versus predicted foliar traits from the leaf bidirectional reflectance factor with PROSPECT and PROCOSINE. Predictions for CN were obtained using the optimal non-contiguous spectral regions obtained from US, and *vice versa* (Table S3). EWT: equivalent water thickness; LMA: leaf mass per area; C_{ab} : leaf chlorophyll $a + b$; C_{xc} : leaf carotenoids.

Table 6

The model performance of PROSPECT and PROCOSINE to predict equivalent water thickness (EWT) and leaf mass per area (LMA) in the US extensive dataset of 3498 foliar samples with three inversion strategies (IS1–3). IS1: uses the full spectrum; IS2: uses the optimal spectral regions as suggested in previous studies (Féret et al., 2019, 2021; Spafford et al., 2021); IS3 uses the optimal non-contiguous spectral regions identified for the US dataset for predicting EWT and LMA, respectively (Table S3). Bold values are the most accurate estimation among the three inversion strategies.

Model	Trait	Inversion strategy	US extensive dataset			
			R ²	RMSE	NRMSE	BIAS
PROSPECT	EWT (mg/cm ²)	IS1	0.67	4.00	7.87	1.88
		IS2	0.67	2.96	5.83	2.65
		IS3	0.78	2.99	5.90	1.76
	LMA (mg/cm ²)	IS1	0.72	2.50	10.92	−0.98
		IS2	0.81	1.34	5.85	−0.14
		IS3	0.83	1.30	5.69	−0.33
PROCOSINE	EWT (mg/cm ²)	IS1	0.71	4.53	8.92	2.59
		IS2	0.69	4.31	8.50	2.06
		IS3	0.48	3.80	7.49	−0.05
	LMA (mg/cm ²)	IS1	0.73	2.52	10.97	−0.82
		IS2	0.76	2.27	9.91	−0.32
		IS3	0.79	1.98	8.61	−0.31

nm, and 2300–2400 nm. For other traits, the optimal spectral domains for CN and US were mostly different.

For PLSR, the variable importance of projection (VIP) values of EWT and LMA models built on CN and US were similar (Fig. S16), which

resulted in good model transferability of those two foliar traits. For EWT, VIP had higher values between 1350 nm and 1450 nm. The VIP values of LMA models were higher at 1200–1340 nm, 1660–1750 nm and 2250–2320 nm. N_{area} models built on CN and US had similar VIP values in the spectral region of 1400–2400 nm. The VIP values of pigment models were high in the visible spectral region, with peaks at 700–740 nm shared by both models built on CN and US.

4. Discussion

4.1. The feasibility of the physically-based approach in predicting foliar traits

This study confirms the applicability of physical model inversion on leaf bidirectional reflectance factor for the estimation of LMA, EWT and N_{area}. Here, both PROSPECT and PROCOSINE model inversion resulted in satisfactory accuracy across geographic areas, plant species, light conditions and growth periods (Table 4, Figs. S4–S7). The estimation accuracies of foliar traits using PROSPECT were generally higher than those from PROCOSINE. We also confirmed the capability of various inversion strategies in improving leaf trait estimations and their transferability across datasets of different continents (Table 5, Fig. 6). Similarly, PROSPECT resulted in more accurate estimations of foliar traits than PROCOSINE when applying the non-contiguous optimal spectral domains identified in one dataset to the other (Table 5). The optimal spectral domains for predicting EWT and LMA with PROSPECT largely differed from those obtained with PROCOSINE, but were quite

Table 7

The model performance of interval partial least squares regression (iPLSR) by selecting the optimal spectral domains. iPLSR model was built on CN dataset or US dataset. Cross-validation was performed on 30% of each dataset. *h* is the number of latent vectors used to build iPLSR models. EWT: equivalent water thickness; LMA: leaf mass per area; C_{ab}: leaf chlorophyll *a* + *b*; C_{xc}: leaf carotenoids.

Trait	Independent Validation									
	CN dataset					US dataset				
	<i>h</i>	R ²	RMSE	NRMSE	BIAS	<i>h</i>	R ²	RMSE	NRMSE	BIAS
EWT (mg/cm ²)	11	0.92	1.13	6.50	0.08	16	0.92	1.75	4.56	0.07
LMA (mg/cm ²)	17	0.97	0.61	4.09	0.14	14	0.93	0.69	6.28	0.01
N _{area} (mg/cm ²)	16	0.88	0.022	8.30	0.0002	14	0.87	0.021	7.94	0.005
C _{ab} (μg/cm ²)	9	0.61	8.68	14.30	1.83	8	0.81	8.02	8.74	−0.08
C _{xc} (μg/cm ²)	7	0.61	1.34	15.43	0.06	4	0.60	1.86	14.04	0.08

Table 8

The cross-dataset validation of interval partial least squares regression (iPLSR) to predict foliar traits. iPLSR model was built on one of the two datasets (CN and US) and applied to the other for validation. *h* is the number of latent vectors used to build iPLSR models. EWT: equivalent water thickness; LMA: leaf mass per area; C_{ab}: leaf chlorophyll *a* + *b*; C_{xc}: leaf carotenoids.

Trait	US → CN				CN → US			
	R ²	RMSE	NRMSE	BIAS	R ²	RMSE	NRMSE	BIAS
EWT (mg/cm ²)	0.68	2.61	12.48	−0.68	0.71	4.67	7.26	1.70
LMA (mg/cm ²)	0.84	2.13	10.49	−1.70	0.82	1.60	8.20	0.85
N _{area} (mg/cm ²)	0.74	0.033	10.72	−0.0009	0.68	0.044	12.11	−0.02
C _{ab} (μg/cm ²)	0.38	17.73	24.04	13.72	0.29	22.21	20.50	−9.25
C _{xc} (μg/cm ²)	0.42	3.47	28.67	2.99	0.05	4.39	21.91	−1.76

Table 9

The relative RMSE difference between the RMSE when cross-validation was performed within a dataset and the RMSE when the models were applied across datasets. Leaf trait models included PROSPECT, PROCOSINE, PLSR and interval PLSR (iPLSR).

Trait	PROSPECT		PROCOSINE		PLSR		iPLSR	
	US→CN	CN→US	US→CN	CN→US	US→CN	CN→US	US→CN	CN→US
EWT	4%	0%	9%	5%	218%	61%	131%	167%
LMA	5%	2%	−1%	23%	37%	30%	249%	132%
N _{area}	0%	0%	6%	10%	−3%	136%	50%	110%
C _{ab}	48%	39%	34%	46%	46%	98%	104%	177%
C _{xc}	7%	−4%	21%	14%	91%	87%	159%	136%

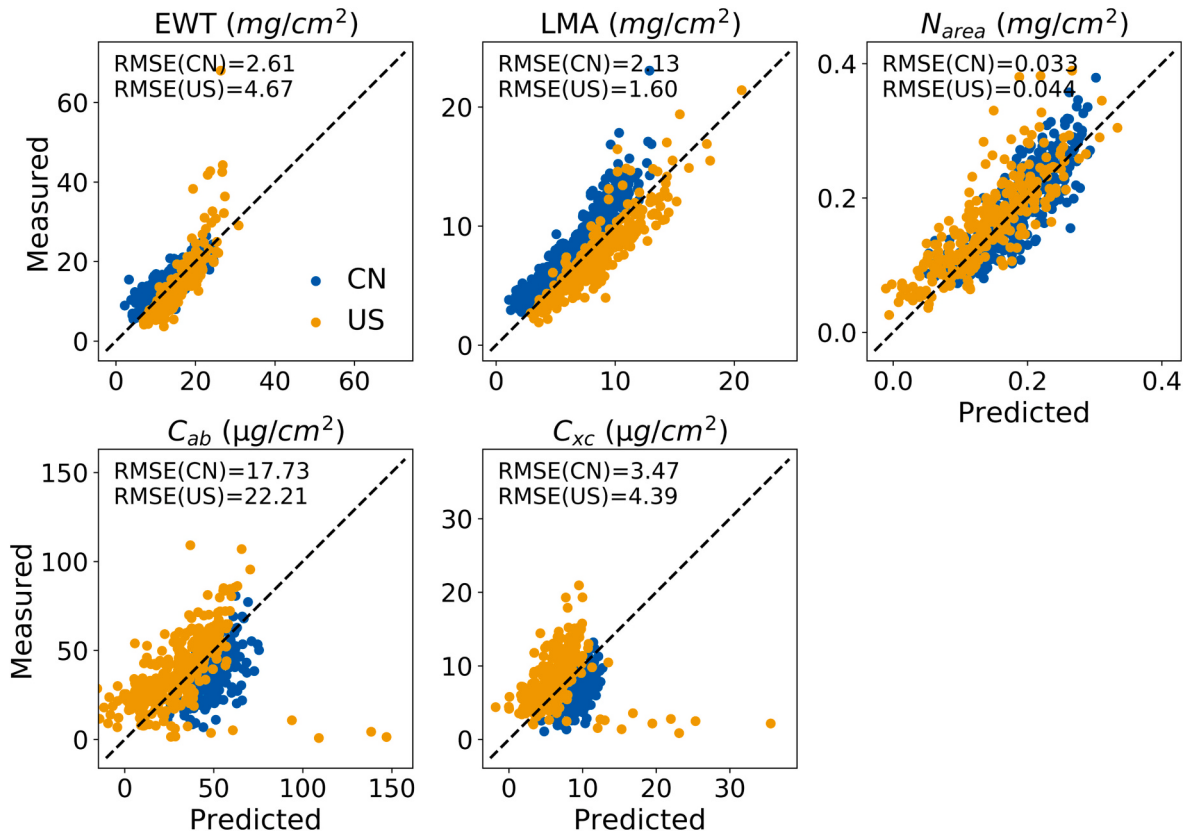


Fig. 7. Measured versus predicted foliar traits from the leaf bidirectional reflectance factor using interval partial least squares regression (iPLSR). Foliar traits in CN dataset were predicted with the US model, while foliar traits in US dataset were predicted with the CN model. EWT: equivalent water thickness; LMA: leaf mass per area; C_{ab} : leaf chlorophyll $a + b$; C_{xc} : leaf carotenoids.

consistent for N_{area} and pigments (Fig. 4, Table S3).

Our estimation accuracy of EWT and LMA was comparable with previous studies (Féret et al., 2019; Feret et al., 2008; Li et al., 2018; Spafford et al., 2021). This can be attributed to the fact that water and dry matter are the dominant drivers of the variations in leaf reflectance and have strong absorption features (Feret et al., 2008; Jacquemoud et al., 2009). LMA was often less accurately estimated than EWT due to the overlapping effect of water absorption features (Ali et al., 2016; Colombo et al., 2008; Feret et al., 2008; Riano et al., 2005). However, this study found that LMA can be more accurately predicted than EWT by using the physically-based approach with properly selected spectral information (Table 4). The optimal spectral domains for LMA mainly located in 1600–1800 nm and 2050–2400 nm (Fig. 4 and Table S3). EWT has weaker absorption in these spectral domains (Wang et al., 2015a; Féret et al., 2019), which thereby improved the prediction of LMA. Our results also confirmed that applying various inversion strategies could further improve the estimation accuracy of both LMA and EWT.

More importantly, our study confirmed the feasibility of retrieving foliar N from leaf bidirectional reflectance factor by inverting the PROSPECT or PROCOSINE model on a wider range of dataset (Table 4). However, inversion strategies are needed in order to achieve a satisfactory estimation accuracy. The selection of optimal non-contiguous spectral region can significantly improve the prediction (Table 4), and the use of leaf structure and b_{spec} as prior information can further increase the accuracy. The optimal spectral subdomains were similar for the two datasets (Fig. 4, Table S3), and had the common feature of 2100 nm related to the absorption features of protein (Curran, 1989; Fourty et al., 1996).

Compared to EWT and LMA, N_{area} had a lower estimation accuracy because of the weaker absorption of proteins which was overlapped by

the absorption features of water and other constituents of dry matter (Curran, 1989). Féret et al. (2021) reported lower NRMSE for the estimation of proteins using PROSPECT-PRO (NRMSE = 15.1%), possibly because the calibration and validation samples were from the same dataset LOPEX. This would lead to consistent experimental uncertainty during data collection (including protocol, equipment, operator and error in measurement). Although RTM model was developed based on physical laws, some important parameters need to be empirically calibrated using in-situ data. For instance, the specific absorption coefficients (SAC) of dry matter, chlorophyll, carotenoids, and protein were calibrated from leaf directional-hemispherical reflectance and transmittance (Féret et al., 2017, 2021; Feret et al., 2008). As shown in Table S10, the SAC calibration of PROSPECT-PRO was performed using a small number of samples ($n = 33$) from LOPEX which is the only publicly available dataset. We expect that the estimation accuracy of N_{area} can be further improved if datasets with the concurrent measurements of leaf spectra and traits from more diverse samples can be used for the calibration of PROSPECT-PRO (Féret et al., 2021).

Leaf chlorophyll in US dataset was accurately estimated through PROSPECT and PROCOSINE inversion (Table 4), and the model performance was similar to previous studies (Feret et al., 2008; Féret et al., 2017; Spafford et al., 2021). Predicting leaf carotenoids is still challenging due to its low concentration and overlapping absorption features with chlorophyll (Feret et al., 2008). We obtained less accurate estimation for carotenoids than chlorophyll (Table 4), which is in agreement with previous studies (Feret et al., 2008; Spafford et al., 2021). Leaf chlorophyll in CN dataset was less accurately estimated (Table 4), possibly due to the discrepancy in the plant functional type between the SAC calibration dataset and our CN dataset, as well as the uncertainty in pigment analysis. US samples were determined using the high-performance liquid chromatography (HPLC) techniques, while CN

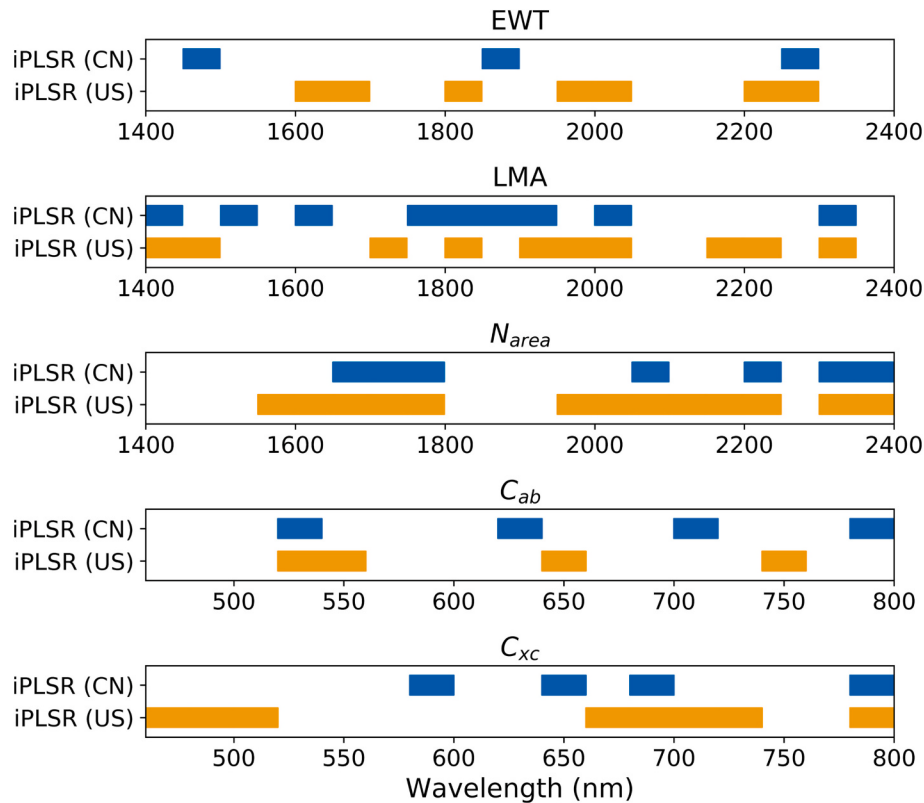


Fig. 8. The optimal spectral domains for predicting foliar traits in CN and US datasets used in interval PLSR (iPLSR). EWT: equivalent water thickness; LMA: leaf mass per area; N_{area} : leaf nitrogen; C_{ab} : leaf chlorophyll $a + b$; C_{xc} : leaf carotenoids.

samples were measured using the spectrophotometric analysis. The plant functional types used for the SAC calibration of PROSPECT are mainly European deciduous species (Table S10) while the plant functional types of our CN dataset used for testing PROSPECT are mainly evergreen broadleaf species. Again, this emphasized the importance of incorporating more diverse leaf samples for the SAC calibration of PROSPECT.

Other inversion strategies such as ecological constraints or wavelet transform can potentially improve the estimation by reducing the unrealistic combination of traits (Banskota et al., 2013; Jurdao et al., 2013; Li et al., 2018; Wang et al., 2018; Yebra and Chuvieco, 2009). Our study did not include the use of ecological constraints in model inversion for the following reasons. The ecological rules are often built on empirical relationships which may be species- or site- specific and not be applicable when transferring to another dataset. For instance, N_{area} and EWT was found to be highly correlated (Homolová et al., 2013; Sullivan et al., 2013; Wang et al., 2015a), but not in the CN and US datasets used in this study ($r = 0.30$ and 0.32 , respectively). The relationship between N_{area} and LMA was often more correlated within species and become much weaker at high taxonomic levels (Anderegg et al., 2018). Therefore, it is difficult to apply a universal relationship in model inversion, and the use of such ecological rules may decrease the generality of RTM. For local studies with similar species, the use of ecological constraints is recommended which may significantly improve the estimation accuracy of N_{area} .

A previous study performed the model inversion by coupling PROSPECT and continuous wavelet transform in the merit function, and achieved more accurate estimation of LMA than PROCOSINE for wheat and rice (Li et al., 2018). This was because wavelet transform could both alleviate the effect of specular reflectance and enhance the absorption of dry matter absorption. In this study, we assessed the performance of wavelet transform but did not achieve satisfactory results (results not shown). Therefore, this study chose to couple PROSPECT and COSINE,

which has a stronger physical basis to consider the specular reflectance effect than wavelet transform (Jay et al., 2016; Li et al., 2018). We also performed model inversion simply using PROSPECT without accounting for the specular reflectance, and obtained trait estimations with similar or higher accuracies than PROCOSINE (Tables 4–5, Figs. 5–6). The good performance of PROSPECT (designed for DHR) on BRF could be attributed to an overestimation of leaf structure using Spafford et al. (2021), which may offset the influence of specular reflectance. Meanwhile, the lower estimation accuracies by PROCOSINE may be caused by the uncertainties in estimating leaf structure and b_{spec} .

4.2. Towards a generalized foliar trait model using PLSR

A generalized empirical model (PLSR) was newly developed to predict LMA across plant species, leaf ages and biomes (Serbin et al., 2019). Our results further confirmed the feasibility of developing a generalized LMA model by extending the dataset to south China which was rarely sampled. Furthermore, the interval PLSR via feature selection further improved the estimation accuracies of foliar traits than PLSR using the full spectrum. The model developed on US can accurately predict the LMA in CN but with some bias, which may be due to the fact that the dominant plant species in US and CN were deciduous and evergreen, respectively.

We also demonstrated that a generalized PLSR model for EWT can be developed across continents (Table 8) due to the dominant role of water in driving the variation of leaf spectra. There was a notable underestimation in EWT predictions when transferring the models from US to CN (Fig. 7), which was likely due to the fact that the CN dataset mainly consisted of evergreen broadleaf species with higher EWT (Fig. 2). The range of EWT limited the model transferability by generating predictions with a systematic bias but a high R^2 (Fig. 7, Table 8). When transferring the EWT model from CN to US, a non-linear relationship between the predictions and the measurements was observed, which

was overestimation for samples with low values and underestimation for those with high values (Fig. 7). Therefore, the development of robust and transferable models across biomes needs the calibration data covering diverse plant functional types.

Nitrogen can be accurately predicted when the models were transferred across datasets (Table 8), which can be due to the common optimal spectral domains shared by CN and US (Fig. 8). The empirically-based approach poorly estimated leaf pigments when transferring from one dataset to the other (Table 8). Leaf chlorophyll $a + b$ was more accurately estimated than carotenoids. Although empirical approaches are often criticized by the low generality, including more representative samples in model development can potentially overcome the challenges (Nakaji et al., 2019; Serbin et al., 2019; Wang et al., 2020; Yang et al., 2016).

4.3. Implication for foliar trait prediction and ecological applications

Physically- and empirically-based approaches are complementary in predicting foliar traits. The selection of approaches depends on the traits of interest, estimation accuracy, model transferability and ease of operation. In general, the physically-based approach can be used for predicting EWT, LMA, N_{area} , C_{ab} and C_{xc} , while the empirically-based approach can be used for any foliar trait of interest with or without obvious absorption features (Burnett et al., 2021). Foliar traits such as P and K are not incorporated in RTM and can only rely on empirical approaches for prediction (Asner et al., 2015; Wang et al., 2020).

Both approaches could be used to predict EWT and LMA with high accuracies over a wide range of plant species (Tables 4 and 7). The physically-based approach could accurately predict C_{ab} for deciduous plant species but not for evergreen species, estimate N_{area} with moderate accuracy, and only poorly estimate C_{xc} . For those cases, empirical approaches such as PLSR are suggested to achieve higher estimation accuracy (Féret et al., 2019; Wang et al., 2018).

RTM offers robustness and transferability, but inversion strategies may affect the model transferability. For instance, the optimal spectral regions and ecological constraints used for model inversion may vary depending on the dataset. The optimal spectral regions suggested in previous studies (IS2) did not yield satisfactory results when using BRF (Table 4), which suggests that these optimal domains defined using DHR and DHT should be refined. With the two extensive datasets collected in eastern United States and south China, we found that the estimation accuracies were quite similar when transferring inversion strategies across datasets, indicating a high level of transferability of the physically-based models (Tables 4, 5 and 9). RTM was most transferable for LMA and EWT when using BRF, which agrees with Féret et al. (2019) who used directional-hemispherical leaf optical properties. The model transferability was slight lower for N_{area} , and least for C_{ab} and C_{xc} . The possible reasons include the trait and spectral differences between the two datasets (Figs. 2 and S17), the experimental uncertainty or bias in the two datasets, as well as the suboptimal performance of BRF compared to DHR. For empirically-based approaches such as PLSR and iPLSR, the estimation accuracies were high when cross-validation was performed, but greatly decreased if models were applied to an independent dataset (Table 9). This demonstrated lower transferability of empirically-based approaches across datasets. It is capable to build generalized trait models for EWT and LMA irrespective of plant functional types. To develop generalized trait models for leaf pigments and nitrogen, an extensive dataset or representative samples from diverse sites, biomes and species are needed (Serbin et al., 2019). Before applying empirical models built on an existing dataset to a new dataset, the similarity of two datasets in terms of foliar traits and spectral characteristics should be evaluated. Transfer learning and model updating can serve as promising techniques to improve the transferability of empirical models such as PLSR (Wan et al., 2022).

The ease of operation of empirically-based approach makes it more popular for ecologists without high skills in remote sensing (Asner et al.,

2015; Burnett et al., 2021; Streher et al., 2020). The operation of RTM model is more challenging for most researchers, and improper use of the RTM model often led to poor estimation accuracies. The software package named Automated Radiative Transfer Models Operator (ARTMO) provides a user-friendly tool of retrieving foliar traits using RTM models and standard inversion (Verrelst et al., 2015, 2019). An R package for PROSPECT (<https://jbferet.gitlab.io/prospect/>) has implemented inversion strategies assessed in this study (i.e., prior information and optimal spectral domains), which can help achieve satisfactory estimation of foliar traits.

Most of the previous studies used leaf hemispherical-directional reflectance (DHR) as model input to PROSPECT to predict foliar traits such as EWT and LMA (Féret et al., 2019; Spafford et al., 2021). Our results confirmed the applicability of leaf bidirectional reflectance factor (BRF) in predicting five key foliar traits. More importantly, the estimation accuracy was comparable to that obtained from DHR (Féret et al., 2019; Feret et al., 2008; Spafford et al., 2021). Therefore, BRF measured using a leaf contact probe can be used to estimate foliar traits from RTM inversion when an integrating sphere is not available (Comar et al., 2012; Li et al., 2018, 2019; Sims and Gamon, 2002). More extensive datasets of BRF with concurrent trait measurements are shared on the Ecological Spectral Information System (EcoSIS; <https://www.ecosis.org/>). Such datasets will advance the development of generalized spectroscopic models for predicting foliar traits, and help the identification of current gaps in characterizing the spectral space of plants living on the planet (Serbin et al., 2019).

Our study also suggested the possibility of integrating leaf PROSPECT or PROCOSINE model (e.g., by transforming BRF to hemispherical-directional reflectance and transmittance) with canopy radiative transfer models to predict foliar traits from canopy BRF. In this coupled model, canopy radiative transfer models such as SAIL (Jacquemoud et al., 2009; Verhoef, 1984), INFORM (Atzberger, 2000) and DART (Gastellu-Etchegorry, 1996) can be used to account for the effects of canopy structure, background, illumination and viewing geometry on canopy BRF. With the already operational and planned satellite missions such as PRISMA (Loizzo et al., 2019), HiSui (Iwasaki et al., 2011), GaoFen-5 (Liu et al., 2019), EnMAP (Guanter et al., 2015), CHIME (Rast et al., 2019) and Surface Biology and Geology (SBG; National Academies of Sciences, 2018), global monitoring of foliar traits regularly will become possible (Berger et al., 2020). This will help us to better understand foliar trait variation at broad scales, links with foliar traits and ecosystem function, as well as assess global functional biodiversity (Rogers et al., 2017; Schimel and Schneider, 2019; Skidmore et al., 2021; Wang et al., 2020).

5. Conclusions

This study compared the generality of RTM and empirical approaches for predicting key foliar functional traits using leaf bidirectional reflectance factor (BRF). Two extensive datasets were collected in eastern United States and south China which covered a large number of species, leaf age and growth condition. By coupling PROSPECT and COSINE, leaf BRF was used as model input to estimate foliar traits. We found that EWT and LMA can be accurately estimated from RTM, while inversion strategies were needed to improve accuracies in predicting N_{area} and pigments. Moreover, the estimation accuracies were similar when transferring inversion strategies across datasets of different continents, indicating high transferability of physically-based models. The empirical approaches, PLSR and interval PLSR, demonstrated lower transferability, e.g., by yielding accurate estimations when cross-validation was performed, but lower accuracies if models were applied to a new dataset. Generalized models can be developed for EWT and LMA by RTM or empirical approaches such as PLSR. In terms of leaf pigments and nitrogen, calibration of PROSPECT by incorporating more diverse leaf samples across biomes is recommended to further improve the estimation accuracy. An extensive dataset or representative samples

from diverse sites, biomes and species are needed to build generalized models with empirical approaches. In addition, transfer learning and model updating can serve as promising techniques to improve the model transferability. With such generalized spectroscopic models for predicting foliar traits, we can better understand the variation of foliar traits among and within species, their response to environmental change, as well as plant biodiversity.

CRedit authorship contribution statement

Zhihui Wang: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Resources, Software, Visualization, Writing – original draft, Writing – review & editing. **Jean-Baptiste Féret:** Conceptualization, Formal analysis, Investigation, Methodology, Writing – review & editing. **Nanfeng Liu:** Conceptualization, Formal analysis, Investigation, Methodology, Visualization, Writing – review & editing. **Zhongyu Sun:** Data curation, Resources, Writing – review & editing. **Long Yang:** Data curation, Resources, Writing – review & editing. **Shoubao Geng:** Data curation, Writing – review & editing. **Hui Zhang:** Data curation, Resources, Writing – review & editing. **Adam Chlus:** Data curation, Writing – review & editing. **Eric L. Kruger:** Data curation, Funding acquisition, Writing – review & editing. **Philip A. Townsend:** Conceptualization, Data curation, Funding acquisition, Resources, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Leaf spectra and trait data will be made available in the EcoSIS spectral repository (<https://ecosis.org/>), and the Matlab codes will be made available at the Ecological Spectral Model Library (<https://ecosml.org/>) on publication of the manuscript.

Acknowledgements

This work is funded by National Natural Science Foundation of China (42001305), Guangdong Basic and Applied Basic Research Foundation (2022A1515011459) and GDAS' Special Project of Science and Technology Development (2020GDASYL-20200102001) to ZW, NSF Macro-systems Biology and NEON-Enabled Science (MSB-NES) award DEB 1638720 to PAT and ELK, and NSF ASCEND Biology Integration Institute (BII) DBI 2021898 to PAT. NL is in part supported by Guangdong Basic and Applied Basic Research Foundation (2022A1515110575, 2023A1515011406). Jean-Baptiste Féret acknowledges financial support from Agence Nationale de la Recherche (BioCop project ANR-17-CE32-0001). We acknowledge the help from NEON site managers with logistics in fieldwork. We thank Ryan Geygan, Hannah Manninen, John Joutras, Andrew Kluck, Tyler Crass, James Fang Gui, Ben Townsend, Ben Spaier, Ittai Herrmann, John Clare, Andrew Jablonski, Tedward Erker, Aidan Mazur, Brendan Heberlein, Abigail Ernst, Haley Knight, Dewi Atikah Radin Umar, Ashley Seufzer, Diana Barrera, Alex Horvath, Angad Dhariwal, Josephine Mayhew, Jacob Gold, Abigail Walther, Raina Eddy, Xu, Alex Brito, Jeannine Cavender-Bares, Anna Schweiger, Cathleen Nguyen, Sarah Hobbie, Richard Lindroth, Kennedy Rubert-Nason, Xizhen Zhu, Zhipeng Chen, Meijie Liu, Meimei Zhang, Shangjie Tang, Jialing Dai, Kai Jiang, Yihui Xu, Jianbin Chen, Wenwen Liu, and Chentao Zhang for their help in field data collection, sample processing and chemical analyses. The authors thank the five anonymous reviewers for their valuable and constructive comments which helped greatly improve the quality of the paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.rse.2023.113614>.

References

- Ali, A.M., Darvishzadeh, R., Skidmore, A.K., van Duren, I., Heiden, U., Heurich, M., 2016. Estimating leaf functional traits by inversion of PROSPECT: assessing leaf dry matter content and specific leaf area in mixed mountainous forest. *Int. J. Appl. Earth Obs. Geoinf.* 45, 66–76. <https://doi.org/10.1016/j.jag.2015.11.004>.
- Anderegg, L.D.L., Berner, L.T., Badgley, G., Sethi, M.L., Law, B.E., HilleRisLambers, J., 2018. Within-species patterns challenge our understanding of the leaf economics spectrum. *Ecol. Lett.* 21, 734–744. <https://doi.org/10.1111/ele.12945>.
- Asner, G.P., Martin, R.E., Anderson, C.B., Knapp, D.E., 2015. Quantifying forest canopy traits: imaging spectroscopy versus field survey. *Remote Sens. Environ.* 158, 15–27. <https://doi.org/10.1016/j.rse.2014.11.011>.
- Atzberger, C., 2000. Development of an invertible forest reflectance model: The INFOR-model. *Decade of Trans-European Remote Sensing Cooperation*.
- Banskota, A., Wynne, R., Thomas, V., Serbin, S., Kayastha, N., Gastellu-Etchegorry, J., et al., 2013. Investigating the utility of wavelet transforms for inverting a 3-D radiative transfer model using hyperspectral data to retrieve Forest LAI. *Remote Sens.* 5, 2639–2659. <https://doi.org/10.3390/rs5062639>.
- Berger, K., Verrelst, J., Féret, J.-B., Wang, Z., Woche, M., Strathmann, M., et al., 2020. Crop nitrogen monitoring: recent progress and principal developments in the context of imaging spectroscopy missions. *Remote Sens. Environ.* 242, 111758. <https://doi.org/10.1016/j.rse.2020.111758>.
- Bousquet, L., Laché, S., Jacquemoud, S., Moya, I., 2005. Leaf BRDF measurements and model for specular and diffuse components differentiation. *Remote Sens. Environ.* 98, 201–211. <https://doi.org/10.1016/j.rse.2005.07.005>.
- Burnett, A.C., Anderson, J., Davidson, K.J., Ely, K.S., Lamour, J., Li, Q., et al., 2021. A best-practice guide to predicting plant traits from leaf-level hyperspectral data using partial least squares regression. *J. Exp. Bot.* <https://doi.org/10.1093/jxb/erab295>.
- Chen, S., Hong, X., Harris, C.J., Sharkey, P.M., 2004. Sparse modeling using orthogonal forward regression with PRESS statistic and regularization. *IEEE Trans. Syst. Man, Cybern. Part B Cybern.* 34, 898–911. <https://doi.org/10.1109/TSMCB.2003.817107>.
- Chlus, A., 2020. Spectroscopic characterization of spatial and temporal patterns in foliar biochemistry. *University of Wisconsin-Madison*.
- Colombo, R., Meroni, M., Marchesi, A., Busetto, L., Rossini, M., Giardino, C., et al., 2008. Estimation of leaf and canopy water content in poplar plantations by means of hyperspectral indices and inverse modeling. *Remote Sens. Environ.* 112, 1820–1834. <https://doi.org/10.1016/j.rse.2007.09.005>.
- Comar, A., Baret, F., Viénot, F., Yan, L., de Solan, B., 2012. Wheat leaf bidirectional reflectance measurements: description and quantification of the volume, specular and hot-spot scattering features. *Remote Sens. Environ.* 121, 26–35. <https://doi.org/10.1016/j.rse.2011.01.028>.
- Combal, B., Baret, F., Weiss, M., Trubuil, A., Macé, D., Pragnère, A., et al., 2003. Retrieval of canopy biophysical variables from bidirectional reflectance using prior information to solve the ill-posed inverse problem. *Remote Sens. Environ.* 84, 1–15. [https://doi.org/10.1016/S0034-4257\(02\)00035-4](https://doi.org/10.1016/S0034-4257(02)00035-4).
- Cornwell, W.K., Cornelissen, J.H.C., Amatangelo, K., Dorrepaal, E., Eviner, V.T., Godoy, O., et al., 2008. Plant species traits are the predominant control on litter decomposition rates within biomes worldwide. *Ecol. Lett.* 11, 1065–1071. <https://doi.org/10.1111/j.1461-0248.2008.01219.x>.
- Croft, H., Chen, J.M., 2018. Leaf pigment content. In: *Comprehensive Remote Sensing*. Elsevier, pp. 117–142. <https://doi.org/10.1016/B978-0-12-409548-9.10547-0>.
- Curran, P.J., 1989. Remote sensing of foliar chemistry. *Remote Sens. Environ.* 30, 271–278. [https://doi.org/10.1016/0034-4257\(89\)90069-2](https://doi.org/10.1016/0034-4257(89)90069-2).
- Darvishzadeh, R., Skidmore, A., Schlerf, M., Atzberger, C., 2008. Inversion of a radiative transfer model for estimating vegetation LAI and chlorophyll in a heterogeneous grassland. *Remote Sens. Environ.* 112, 2592–2604. <https://doi.org/10.1016/j.rse.2007.12.003>.
- Díaz, S., Kattge, J., Cornelissen, J.H.C., Wright, I.J., Lavorel, S., Dray, S., et al., 2016. The global spectrum of plant form and function. *Nature* 529, 167–171. <https://doi.org/10.1038/nature16489>.
- Féret, J.-B., Berger, K., de Boissieu, F., Malenovsky, Z., 2021. PROSPECT-PRO for estimating content of nitrogen-containing leaf proteins and other carbon-based constituents. *Remote Sens. Environ.* 252, 112173. <https://doi.org/10.1016/j.rse.2020.112173>.
- Féret, J.-B., François, C., Gitelson, A., Asner, G.P., Barry, K.M., Panigada, C., et al., 2011. Optimizing spectral indices and chemometric analysis of leaf chemical properties using radiative transfer modeling. *Remote Sens. Environ.* 115, 2742–2750. <https://doi.org/10.1016/j.rse.2011.06.016>.
- Féret, J.-B., le Maire, G., Jay, S., Berveiller, D., Bendoula, R., Hmimina, G., et al., 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 Sens. Environ.* 231, 110959. <https://doi.org/10.1016/j.rse.2018.11.002>.
- Feret, J.B., François, C., Asner, G.P., Gitelson, A.A., Martin, R.E., Bidet, L.P.R., et al., 2008. PROSPECT-4 and 5: advances in the leaf optical properties model separating photosynthetic pigments. *Remote Sens. Environ.* 112, 3030–3043. <https://doi.org/10.1016/j.rse.2008.02.012>.

- Féret, J.B., Gitelson, A.A., Noble, S.D., Jacquemoud, S., 2017. PROSPECT-D: towards modeling leaf optical properties through a complete lifecycle. *Remote Sens. Environ.* 193, 204–215. <https://doi.org/10.1016/j.rse.2017.03.004>.
- Fourty, T., Baret, F., Jacquemoud, S., Schmuck, G., Verdebout, J., 1996. Leaf optical properties with explicit description of its biochemical composition: direct and inverse problems. *Remote Sens. Environ.* 56, 104–117. [https://doi.org/10.1016/0034-4257\(95\)00234-0](https://doi.org/10.1016/0034-4257(95)00234-0).
- Fu, P., Meacham-Hensold, K., Guan, K., Wu, J., Bernacchi, C., 2020. Estimating photosynthetic traits from reflectance spectra: a synthesis of spectral indices, numerical inversion, and partial least square regression. *Plant Cell Environ.* 43, 1241–1258. <https://doi.org/10.1111/pce.13718>.
- Gastellu-Etchegorry, J., 1996. Modeling radiative transfer in heterogeneous 3-D vegetation canopies. *Remote Sens. Environ.* 58, 131–156. [https://doi.org/10.1016/0034-4257\(95\)00253-7](https://doi.org/10.1016/0034-4257(95)00253-7).
- Grossman, Y.L., Ustin, S.L., Jacquemoud, S., Sanderson, E.W., Schmuck, G., Verdebout, J., 1996. Critique of stepwise multiple linear regression for the extraction of leaf biochemistry information from leaf reflectance data. *Remote Sens. Environ.* 56, 182–193. [https://doi.org/10.1016/0034-4257\(95\)00235-9](https://doi.org/10.1016/0034-4257(95)00235-9).
- Guanter, L., Kaufmann, H., Segl, K., Foerster, S., Rogass, C., Chabrillat, S., et al., 2015. The EnMAP spaceborne imaging spectroscopy mission for earth observation. *Remote Sens.* 7, 8830–8857. <https://doi.org/10.3390/rs70708830>.
- Homolová, L., Malenovsky, Z., Clevers, J.G.P.W., García-Santos, G., Schaepman, M.E., 2013. Review of optical-based remote sensing for plant trait mapping. *Ecol. Complex.* 15, 1–16. <https://doi.org/10.1016/j.ecocom.2013.06.003>.
- Hovi, A., Forsström, P., Möttö, M., Rautiainen, M., 2018. Evaluation of accuracy and practical applicability of methods for measuring leaf reflectance and transmittance spectra. *Remote Sens.* 10, 25. <https://doi.org/10.3390/rs10010025>.
- Iwasaki, A., Ohgi, N., Tanii, J., Kawashima, T., Inada, H., 2011. Hyperspectral Imager Suite (HISUI)-Japanese hyper-multi spectral radiometer. In: 2011 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 1025–1028. <https://doi.org/10.1109/IGARSS.2011.6049308>.
- Jacquemoud, S., Baret, F., 1990. PROSPECT: a model of leaf optical properties spectra. *Remote Sens. Environ.* 34, 75–91. [https://doi.org/10.1016/0034-4257\(90\)90100-Z](https://doi.org/10.1016/0034-4257(90)90100-Z).
- Jacquemoud, S., Verhoef, W., Baret, F., Bacour, C., Zarco-Tejada, P.J., Asner, G.P., et al., 2009. PROSPECT + SAIL models: a review of use for vegetation characterization. *Remote Sens. Environ.* 113, S56–S66. <https://doi.org/10.1016/j.rse.2008.01.026>.
- Jay, S., Bendoula, R., Hadoux, X., Féret, J.-B., Gorretta, N., 2016. A physically-based model for retrieving foliar biochemistry and leaf orientation using close-range imaging spectroscopy. *Remote Sens. Environ.* 177, 220–236. <https://doi.org/10.1016/j.rse.2016.02.029>.
- Jurdao, S., Yebra, M., Guerschman, J.P., Chuvieco, E., 2013. Regional estimation of woodland moisture content by inverting radiative transfer models. *Remote Sens. Environ.* 132, 59–70. <https://doi.org/10.1016/j.rse.2013.01.004>.
- Kampe, T.U., Johnson, B.R., Kuester, M.A., Keller, M., 2010. NEON: the first continental-scale ecological observatory with airborne remote sensing of vegetation canopy biochemistry and structure. *J. Appl. Remote Sens.* 4, 043510. <https://doi.org/10.1117/1.3361375>.
- Kothari, S., Cavender-Bares, J., Bitan, K., Verhoeven, A.S., Wang, R., Montgomery, R.A., et al., 2018. Community-wide consequences of variation in photoprotective physiology among prairie plants. *Photosynthetica* 56, 455–467. <https://doi.org/10.1007/s11099-018-0777-9>.
- Kudo, M., Sklansky, J., 2000. Comparison of algorithms that select features for pattern classifiers. *Pattern Recogn.* 33, 25–41. [https://doi.org/10.1016/S0031-3203\(99\)00041-2](https://doi.org/10.1016/S0031-3203(99)00041-2).
- Li, D., Cheng, T., Jia, M., Zhou, K., Lu, N., Yao, X., et al., 2018. PROCWT: coupling PROSPECT with continuous wavelet transform to improve the retrieval of foliar chemistry from leaf bidirectional reflectance spectra. *Remote Sens. Environ.* 206, 1–14. <https://doi.org/10.1016/j.rse.2017.12.013>.
- Li, D., Tian, L., Wan, Z., Jia, M., Yao, X., Tian, Y., et al., 2019. Assessment of unified models for estimating leaf chlorophyll content across directional-hemispherical reflectance and bidirectional reflectance spectra. *Remote Sens. Environ.* 231, 111240. <https://doi.org/10.1016/j.rse.2019.111240>.
- Li, P., Wang, Q., 2011. Retrieval of leaf biochemical parameters using PROSPECT inversion: a new approach for alleviating ill-posed problems. *IEEE Trans. Geosci. Remote Sens.* 49, 2499–2506. <https://doi.org/10.1109/TGRS.2011.2109390>.
- Lichtenthaler, H.K., 1987. Chlorophylls and carotenoids: pigments of photosynthetic biomembranes. *Methods Enzymol.* 148, 350–382. [https://doi.org/10.1016/0076-6879\(87\)48036-1](https://doi.org/10.1016/0076-6879(87)48036-1).
- Liu, Y.-N., Zhang, J., Zhang, Y., Sun, W.-W., Jiao, L.-L., Sun, D.-X., et al., 2019. The advanced hyperspectral imager: aboard China's GaoFen-5 satellite. *IEEE Geosci. Remote Sens. Mag.* 7, 23–32. <https://doi.org/10.1109/MGRS.2019.2927687>.
- Loizzo, R., Daraio, M., Guarini, R., Longo, F., Lorusso, R., Dini, L., 2019. Prisma mission status and perspective. In: IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 4503–4506. <https://doi.org/10.1109/IGARSS.2019.8899272>.
- Ma, Z.G., Chen, X., Wang, Q., Li, P.H., Jiapaer, G., 2012. Retrieval of leaf biochemical properties by inverted PROSPECT model and hyperspectral indices: an application to Populus euphratica polymorphic leaves. *J. Arid Land* 4, 52–62. <https://doi.org/10.3724/SP.J.1227.2012.00052>.
- Maccioni, A., Agati, G., Mazzinghi, P., 2001. New vegetation indices for remote measurement of chlorophylls based on leaf directional reflectance spectra. *J. Photochem. Photobiol. B Biol.* 61, 52–61. [https://doi.org/10.1016/S1011-1344\(01\)00145-2](https://doi.org/10.1016/S1011-1344(01)00145-2).
- Marcano-Cedeno, A., Quintanilla-Dominguez, J., Cortina-Januchs, M.G., Andina, D., 2010. Feature selection using Sequential Forward Selection and classification applying Artificial Metaplasticity Neural Network. In: IECON 2010 - 36th Annual Conference on IEEE Industrial Electronics Society. IEEE, pp. 2845–2850. <https://doi.org/10.1109/IECON.2010.5675075>.
- Martin, M.E., Plourde, L.C., Ollinger, S.V., Smith, M.L., McNeil, B.E., 2008. A generalizable method for remote sensing of canopy nitrogen across a wide range of forest ecosystems. *Remote Sens. Environ.* 112, 3511–3519. <https://doi.org/10.1016/j.rse.2008.04.008>.
- Mehmood, T., Liland, K.H., Snipen, L., Sæbø, S., 2012. A review of variable selection methods in partial least squares regression. *Chemom. Intell. Lab. Syst.* 118, 62–69. <https://doi.org/10.1016/j.chemolab.2012.07.010>.
- Möttus, M., Hovi, A., Rautiainen, M., 2017. Theoretical algorithm and application of a double-integrating sphere system for measuring leaf transmittance and reflectance spectra. *Appl. Opt.* 56, 563. <https://doi.org/10.1364/AO.56.000563>.
- Nakaji, T., Oguma, H., Nakamura, M., Kachina, P., Asanok, L., Marod, D., et al., 2019. Estimation of six leaf traits of east asian forest tree species by leaf spectroscopy and partial least square regression. *Remote Sens. Environ.* 233, 111381. <https://doi.org/10.1016/j.rse.2019.111381>.
- National Academies of Sciences, 2018. Thriving on our changing planet: A decadal strategy for Earth observation from space. National Academies Press, Washington, D. C. <https://doi.org/10.17226/24938>.
- Nørgaard, L., Saudland, A., Wagner, J., Nielsen, J.P., Munck, L., Engelsen, S.B., 2000. Interval partial least-squares regression (i PLS): a comparative chemometric study with an example from near-infrared spectroscopy. *Appl. Spectrosc.* 54, 413–419. <https://doi.org/10.1366/0003702001949500>.
- Pereira, H.M., Ferrier, S., Walters, M., Geller, G.N., Jongman, R.H.G., Scholes, R.J., et al., 2013. Essential biodiversity variables. *Science* 339, 277–278. <https://doi.org/10.1126/science.1229931>.
- Pérez-Harguindeguy, N., Díaz, S., Garnier, E., Lavorel, S., Poorter, H., Jaureguiberry, P., et al., 2013. New handbook for standardised measurement of plant functional traits worldwide. *Aust. J. Bot.* 61, 167. <https://doi.org/10.1071/BT12225>.
- Pullanagari, R.R., Dehghan-Shoar, M., Yule, L.J., Bhatia, N., 2021. Field spectroscopy of canopy nitrogen concentration in temperate grasslands using a convolutional neural network. *Remote Sens. Environ.* 257, 112353. <https://doi.org/10.1016/j.rse.2021.112353>.
- Rast, M., Ananasso, C., Bach, H., Ben-Dor, E., Chabrillat, S., Colombo, R., 2019. Copernicus hyperspectral imaging mission for the environment : Mission requirements document, 2.1. ed. European Space Agency (ESA), Noordwijk.
- Reich, P.B., 2014. The world-wide “fast-slow” plant economics spectrum: a traits manifesto. *J. Ecol.* 102, 275–301. <https://doi.org/10.1111/1365-2745.12211>.
- Reich, P.B., Walters, M.B., Ellsworth, D.S., 1997. From tropics to tundra: global convergence in plant functioning. *Proc. Natl. Acad. Sci. U. S. A.* 94, 13730–13734. <https://doi.org/10.1073/pnas.94.25.13730>.
- Riño, D., Vaughan, P., Chuvieco, E., Zarco-Tejada, P.J., Ustin, S.L., 2005. Estimation of fuel moisture content by inversion of radiative transfer models to simulate equivalent water thickness and dry matter content: analysis at leaf and canopy level. *IEEE Trans. Geosci. Remote Sens.* 43, 819–826. <https://doi.org/10.1109/TGRS.2005.843316>.
- Rogers, A., Medlyn, B.E., Dukes, J.S., Bonan, G., von Caemmerer, S., Dietze, M.C., et al., 2017. A roadmap for improving the representation of photosynthesis in earth system models. *New Phytol.* 213, 22–42. <https://doi.org/10.1111/nph.14283>.
- Schimel, D., Hargrove, W., Hoffman, F., MacMahon, J., 2007. NEON: a hierarchically designed national ecological network. *Front. Ecol. Environ.* 5, 59. [https://doi.org/10.1890/1540-9295\(2007\)5\[59:NAHDNE\]2.0.CO;2](https://doi.org/10.1890/1540-9295(2007)5[59:NAHDNE]2.0.CO;2).
- Schimel, D., Schneider, F.D., 2019. Flux towers in the sky: global ecology from space. *New Phytol.* 224, 570–584. <https://doi.org/10.1111/nph.15934>.
- Schweiger, A.K., Cavender-Bares, J., Townsend, P.A., Hobbie, S.E., Madritch, M.D., Wang, R., et al., 2018. Plant spectral diversity integrates functional and phylogenetic components of biodiversity and predicts ecosystem function. *Nat. Ecol. Evol.* 2, 976–982. <https://doi.org/10.1038/s41559-018-0551-1>.
- Serbin, S.P., Wu, J., Ely, K.S., Kruger, E.L., Townsend, P.A., Meng, R., et al., 2019. From the Arctic to the tropics: multi-biome prediction of leaf mass per area using leaf reflectance. *New Phytol.* 224, 1557–1568. <https://doi.org/10.1111/nph.16123>.
- Shiklomanov, A.N., Dietze, M.C., Viskari, T., Townsend, P.A., Serbin, S.P., 2016. Quantifying the influences of spectral resolution on uncertainty in leaf trait estimates through a bayesian approach to RTM inversion. *Remote Sens. Environ.* 183, 226–238. <https://doi.org/10.1016/j.rse.2016.05.023>.
- Sims, D.A., Gamon, J.A., 2002. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. *Remote Sens. Environ.* 81, 337–354. [https://doi.org/10.1016/S0034-4257\(02\)00010-X](https://doi.org/10.1016/S0034-4257(02)00010-X).
- Skidmore, A.K., Coops, N.C., Neinavaz, E., Ali, A., Schaepman, M.E., Paganini, M., et al., 2021. Priority list of biodiversity metrics to observe from space. *Nat. Ecol. Evol.* 5, 896–906. <https://doi.org/10.1038/s41559-021-01451-x>.
- Spafford, L., le Maire, G., MacDougall, A., de Boissieu, F., Féret, J.-B., 2021. Spectral subdomains and prior estimation of leaf structure improves PROSPECT inversion on reflectance or transmittance alone. *Remote Sens. Environ.* 252, 112176. <https://doi.org/10.1016/j.rse.2020.112176>.
- Streher, A.S., Torres, R.S., Morello, L.P.C., Silva, T.S.F., 2020. Accuracy and limitations for spectroscopic prediction of leaf traits in seasonally dry tropical environments. *Remote Sens. Environ.* 244, 111828. <https://doi.org/10.1016/j.rse.2020.111828>.
- Sullivan, F.B., Ollinger, S.V., Martin, M.E., Ducey, M.J., Lepine, L.C., Wicklein, H.F., 2013. Foliar nitrogen in relation to plant traits and reflectance properties of New Hampshire forests. *Can. J. For. Res.* 43, 18–27. <https://doi.org/10.1139/cjfr-2012-0324>.
- Taiz, L., Zeiger, E., 2010. Plant Physiology, 5th ed. Sinauer Associates, Sunderland, MA.
- Ustin, S.L., Gitelson, A.A., Jacquemoud, S., Schaepman, M., Asner, G.P., Gamon, J.A., et al., 2009. Retrieval of foliar information about plant pigment systems from high

- resolution spectroscopy. *Remote Sens. Environ.* 113, S67–S77. <https://doi.org/10.1016/j.rse.2008.10.019>.
- Ustin, S.L., Riaño, D., Hunt, E.R., 2012. Estimating canopy water content from spectroscopy. *Isr. J. Plant Sci.* 60, 9–23. <https://doi.org/10.1560/IJPS.60.1-2.9>.
- van Bodegom, P.M., Douma, J.C., Verheijen, L.M., 2014. A fully traits-based approach to modeling global vegetation distribution. *Proc. Natl. Acad. Sci. U. S. A.* 111, 13733–13738. <https://doi.org/10.1073/pnas.1304551110>.
- Verhoef, W., 1984. Light scattering by leaf layers with application to canopy reflectance modeling: the SAIL model. *Remote Sens. Environ.* 16, 125–141. [https://doi.org/10.1016/0034-4257\(84\)90057-9](https://doi.org/10.1016/0034-4257(84)90057-9).
- Verrelst, J., Camps-Valls, G., Muñoz-Marí, J., Rivera, J.P., Veroustraete, F., Clevers, J.G. P.W., et al., 2015. Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties - a review. *ISPRS J. Photogramm. Remote Sens.* 108, 273–290. <https://doi.org/10.1016/j.isprsjprs.2015.05.005>.
- Verrelst, J., Malenovský, Z., Van der Tol, C., Camps-Valls, G., Gastellu-Etchegorry, J.-P., Lewis, P., et al., 2019. Quantifying vegetation biophysical variables from imaging spectroscopy data: a review on retrieval methods. *Surv. Geophys.* 40, 589–629. <https://doi.org/10.1007/s10712-018-9478-y>.
- Verrelst, J., Muñoz, J., Alonso, L., Delegido, J., Rivera, J.P., Camps-Valls, G., et al., 2012. Machine learning regression algorithms for biophysical parameter retrieval: opportunities for Sentinel-2 and -3. *Remote Sens. Environ.* 118, 127–139. <https://doi.org/10.1016/j.RSE.2011.11.002>.
- 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 Sens. Environ.* 269, 112826. <https://doi.org/10.1016/j.rse.2021.112826>.
- Wang, S., Guan, K., Wang, Z., Ainsworth, E.A., Zheng, T., Townsend, P.A., et al., 2021. Unique contributions of chlorophyll and nitrogen to predict crop photosynthetic capacity from leaf spectroscopy. *J. Exp. Bot.* 72, 341–354. <https://doi.org/10.1093/jxb/eraa432>.
- Wang, Z., Chlus, A., Geygan, R., Ye, Z., Zheng, T., Singh, A., et al., 2020. Foliar functional traits from imaging spectroscopy across biomes in eastern North America. *New Phytol.* 228, 494–511. <https://doi.org/10.1111/nph.16711>.
- Wang, Z., Skidmore, A.K., Darvishzadeh, R., Heiden, U., Heurich, M., Wang, T., 2015a. Leaf nitrogen content indirectly estimated by leaf traits derived from the PROSPECT model. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 8, 3172–3182. <https://doi.org/10.1109/JSTARS.2015.2422734>.
- Wang, Z., Skidmore, A.K., Darvishzadeh, R., Wang, T., 2018. Mapping forest canopy nitrogen content by inversion of coupled leaf-canopy radiative transfer models from airborne hyperspectral imagery. *Agric. For. Meteorol.* 253–254, 247–260. <https://doi.org/10.1016/j.agrformet.2018.02.010>.
- Wang, Z., Skidmore, A.K., Wang, T., Darvishzadeh, R., Hearne, J., 2015b. Applicability of the PROSPECT model for estimating protein and cellulose + lignin in fresh leaves. *Remote Sens. Environ.* 168, 205–218. <https://doi.org/10.1016/j.rse.2015.07.007>.
- Wang, Z., Townsend, P.A., Schweiger, A.K., Couture, J.J., Singh, A., Hobbie, S.E., et al., 2019. Mapping foliar functional traits and their uncertainties across three years in a grassland experiment. *Remote Sens. Environ.* 221, 405–416. <https://doi.org/10.1016/j.rse.2018.11.016>.
- Wang, Z., Wang, T., Darvishzadeh, R., Skidmore, A.K., Jones, S., Suarez, L., et al., 2016. Vegetation indices for mapping canopy foliar nitrogen in a mixed temperate forest. *Remote Sens.* 8, 491. <https://doi.org/10.3390/rs8060491>.
- Wold, S., Sjöström, M., Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* 58, 109–130. [https://doi.org/10.1016/S0169-7439\(01\)00155-1](https://doi.org/10.1016/S0169-7439(01)00155-1).
- Wright, I.J., Reich, P.B., Westoby, M., Ackerly, D.D., Baruch, Z., Bongers, F., et al., 2004. The worldwide leaf economics spectrum. *Nature* 428, 821–827. <https://doi.org/10.1038/nature02403>.
- Yang, X., Tang, J., Mustard, J.F., Wu, J., Zhao, K., Serbin, S., et al., 2016. Seasonal variability of multiple leaf traits captured by leaf spectroscopy at two temperate deciduous forests. *Remote Sens. Environ.* 179, 1–12. <https://doi.org/10.1016/j.rse.2016.03.026>.
- Yebra, M., Chuvielo, E., 2009. Linking ecological information and radiative transfer models to estimate fuel moisture content in the Mediterranean region of Spain: solving the ill-posed inverse problem. *Remote Sens. Environ.* 113, 2403–2411. <https://doi.org/10.1016/j.rse.2009.07.001>.
- Yeoh, H.H., Wee, Y.C., 1994. Leaf protein contents and nitrogen-to-protein conversion factors for 90 plant species. *Food Chem.* 49, 245–250. [https://doi.org/10.1016/0308-8146\(94\)90167-8](https://doi.org/10.1016/0308-8146(94)90167-8).