Title

Leaf spectral clusters as potential optical leaf functional types within California ecosystems

Authors

Roth, Keely L
Casas, Angeles
Huesca, Margarita
et al.

Publication Date

2016-10-01

DOI

10.1016/j.rse.2016.07.014

Peer reviewed

Remote Sensing of Environment 184 (2016) 229–246

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Leaf spectral clusters as potential optical leaf functional types within California ecosystems

Keely L. Roth [a,*], Angeles Casas [a], Margarita Huesca [a], Susan L. Ustin [a], Maria Mar Alsina [a,b], Spencer A. Mathews [c], Michael L. Whiting [a]

[a] Department of Land, Air and Water Resources, University of California, Davis, One Shields Ave, Davis, CA 95616, United States
[b] E & J Gallo Winery, Modesto, CA, United States
[c] Department of Computer Sciences, University of California, Davis, One Shields Ave, Davis, CA 95616, United States

ARTICLE INFO

Article history:
Received 30 December 2015
Received in revised form 11 June 2016
Accepted 8 July 2016
Available online xxxx

Keywords:
Leaf spectra
Plant functional types
Optical functional types
Hierarchical clustering
Spectroscopy

ABSTRACT

Our ability to measure and map plant function at multiple ecological scales is critical for understanding current and future changes in Earth's ecosystems and the global carbon budget. Conventional plant functional types (cPFTs) based on a few productivity-related traits have been previously used to simplify and represent major differences in global plant functions, but more recent research has directly focused on the use of functional trait information. Still, sampling limitations have constrained efforts to truly understand the variance and covariance of functional traits globally. Reflectance spectra offer a fast, repeatable, simultaneous measurement of a wide variety of leaf functional traits and could be used to optically define leaf functional types. To evaluate this concept, we measured leaf reflectance from a wide range of species in a diverse set of ecosystems across central and northern California, including observations from multiple individuals, sites, and seasons. Using principal components analysis, we analyzed spectral variation in relation to categorical attributes such as species and cPFTs, as well as to a set of functional trait metrics calculated from the spectra. We found the first three principal components (PCs) to be weakly related to categorical attributes and more strongly related to spectrally-derived functional metrics. Each PC was more strongly associated with different portions of the spectrum and contained different functional information. We applied a hybrid clustering algorithm to the PC coordinates of the observations to define potential optical leaf functional types. Twelve spectral clusters were identified, and these did not correspond directly to either single cPFTs or species. However, each cluster had a unique functional metric profile. Clusters represented both inter- and intra-species and cPFT functional differences driven by taxonomy, trait evolution and environmental responses, demonstrating their value as optical leaf functional types and the value of the clustering approach used here for defining optical types from leaf spectra. Our findings support the notion that cPFTs do not adequately capture differences in leaf function. They demonstrate that spectral measurements can be used to improve both the definition of PFTs as well as our knowledge regarding the covariance of functional traits within these classes.

1. Introduction

Plant functioning plays a central role in biosphere-atmosphere interactions (e.g., gas and energy exchange), community assembly and biodiversity, and ecosystem services and resilience (Reichstein et al., 2014). The current rates of environmental change place even greater importance on improving our ability to measure, map, and model plant function across Earth's surface (Schimel et al., 2013). Within ecology, scientists have long sought to represent and understand observed patterns by grouping the diversity of plant life based on major differences in biological or ecological functioning, termed Plant Functional Types (PFTs) (Smith et al., 1997; Woodward & Cramer, 1996). While the literature on how to define PFTs and how to assign species to them is extensive (see Duckworth et al., 2000; Semenova and Van Der Maarel, 2000; Wilson, 1999), the climate and terrestrial biosphere modeling community has been particularly reliant on relatively simple conventional PFTs (cPFTs), based on a few, easily observed physiological and morphological attributes (i.e., photosynthetic pathway, leaf type, leaf phenology and life form) (Bonan et al., 2002; Moorcroft, 2006). Using this concept, individual plant species are assigned to a single cPFT based on these attributes. However, cPFTs have important limitations. There are many potential sets of cPFTs, and the categorical attributes they rely upon may not be those most strongly related to functioning (Wright et al., 2006). Assigning constant attributes to cPFTs makes predicting continuous variation in function impossible

* Corresponding author.
E-mail address: klroth@ucdavis.edu (K.L. Roth).

http://dx.doi.org/10.1016/j.rse.2016.07.014
0034-4257/

and does not allow for changes in function in response to environmental changes (Van Bodegom et al., 2012; Yang et al., 2015). These categorizations also ignore within-cPFT functional variability driven by differences among individual species, as well as by interactions with climate and weather, and changes throughout the growing season (Caldararu et al., 2015). In fact, biosphere-modeling studies are finding more and more frequently cPFTs are insufficient for capturing variation in plant function (Alton, 2011; Groenendijk et al., 2011; Pavlick et al., 2012).

To address these shortcomings, among others, ecologists have shifted focus from cPFTs toward functional traits to quantify processes and define more appropriate PFTs (Bernhardt-Römermann et al., 2008; Kattge et al., 2011; Violle et al., 2007). The characterization of plants based on measureable attributes, or traits, has a long history within ecology. *Functional traits* are considered to be those 1) measured at the level of the individual and 2) having the potential to influence fitness (Pérez-Harguindeguy et al., 2013). Examples include leaf mass area (LMA), water content, and pigment content. A wealth of studies have examined how functional traits covary and relate to biospheric functioning and model parameterization (Pappas et al., 2014; Reichstein et al., 2014; Sakschewski et al., 2015; Van Bodegom et al., 2014; Westoby & Wright, 2006). Researchers have explored the major axes of functional trait spectra with relation to evolution, life history strategies, and climate, demonstrating functional tradeoffs between productivity and stress tolerance (i.e., "quick" vs. "slow" return on investment of dry mass and nutrients) (Adler et al., 2014; Reich et al., 2003; Shipley et al., 2006; Wright et al., 2004). Functional traits are also frequently used to investigate patterns of community assembly, biodiversity and ecosystem function (Ackerly & Cornwell, 2007; Cadotte et al., 2009; Laughlin & Laughlin, 2013; Lavorel & Garnier, 2002; Violle & Jiang, 2009).

Despite widespread success, several challenges remain for using functional traits to further our understanding of functional patterns. Two stand out among these: 1) incorporating functional trait variation across biological and ecological scales; and 2) adequate sampling of functional traits across the globe (necessary for #1). Most often, mean trait values for single species are used in large-scale analyses of functional trait spectra, ignoring sometimes significant intraspecific variation (Albert et al., 2012), although there is plenty of evidence to demonstrate this variation is an important consideration when using traits to estimate function, quantify diversity, and understand community composition (Albert et al., 2010a; Albert et al., 2010b; Wright & Sutton-Grier, 2012). Substantial functional trait variation can also occur at different ecological scales (i.e., within individuals, populations and sites) (Messier et al., 2010). In order to address this variation (and understand its drivers), we need spatially and temporally extensive measurements of functional traits across many individuals, species, and ecosystems. Even though we have larger databases of these measurements than ever before (e.g., TRY; Kattge et al., 2011), the sampling of Earth's biomes has been uneven and bias (Sandel et al., 2015; Schimel et al., 2015).

Given the impossibility of sampling the entire globe, remote sensing is necessary to address these shortcomings and move toward the previously-described goals by providing spatially explicit, repeatable measurements of many of these traits (Asner, 2013; Schimel et al., 2013). Historically, remote sensing data have been used to map PFT composition (Bonan et al., 2002; Sun et al., 2008), but due to the sensitivity of multispectral sensors, only generalized cPFTs could be mapped. However, the development of high fidelity imaging spectrometers has enabled scientists to differentiate among more specific taxonomic classes and move toward the direct estimation of specific leaf functional traits (Schaepman et al., 2009). Because the interactions of light with leaves are mediated by leaf traits (Asner, 1998; Féret et al., 2008), spectroscopy has been successfully used to estimate and predict a wide range of leaf traits from many ecosystems at both leaf and canopy scales (Asner et al., 2015; Homolová et al., 2013; Ustin et al., 2004, 2009). Thus leaf spectral measurements can serve as a proxy for some leaf functional traits, and spectra can be rapidly measured for many leaves on many individuals in the field, providing a standardized, quantitative measurement that can be compared across locations and through time.

Beyond directly estimating traits and separating pre-defined classes (e.g., cPFTs or species), it has been proposed that spectra could be used to directly define "optical functional types" (i.e., PFTs defined from optical remote sensing data) (Ustin & Gamon, 2010). Because the spectrum is sensitive to changes in functional traits, classes defined based on spectral information should capture key functional trait patterns and thus account for inter- and intra- species and cPFT variation. Unsupervised clustering analysis is one approach that can be used to define such optical functional types. It has been applied in aquatic studies to define optical water types (Shi et al., 2013), benthic habitat types (Garcia et al., 2015), and phytoplankton pigment assemblages (Torrecilla et al., 2011). Although clustering has been previously used to generally assess patterns of spectral and functional similarity among terrestrial plant species (Asner & Martin, 2009; Asner et al., 2009; Hesketh & Sánchez-Azofeifa, 2012), it has not, to our knowledge been used to define discrete optical functional classes at leaf-level.

Our main objective in this research was to summarize and interpret the information contained in a diverse set of leaf spectra and assess our ability to define novel optical leaf functional types with these data. More specifically, we sought to:

1) evaluate the relationship of observed leaf spectral variance with commonly used categorical attributes (e.g., cPFTs and species), and with functional metrics calculated from the spectra
2) define distinct optical groups (clusters) using an unsupervised hybrid clustering approach
3) investigate the correspondence of these spectrally-defined clusters with species, cPFTs, and functional metrics

This study aims to improve our understanding of how leaf spectra can be used to define unique functional groups. Because of the large amount of functional trait variance observed within cPFTs, and even within individual species, in prior studies, we hypothesized that classes defined based on spectral measurements alone (optical functional types), can better capture functional trait differences, and provide greater insight into patterns of functional trait covariance within and among species and cPFTs, and across ecosystem types and multiple seasons.

2. Methods

2.1. Study sites and sample collection

Our nine field sites were located across northern and central California (Fig. 1). These sites were selected in support of NASA's Hyperspectral Infrared Imager (HyspIRI) Preparatory Science initiative, which collected hyperspectral imagery over large sections of the state in spring, summer, and fall of 2013, 2014, and 2015. The sites cover several of California's most prominent ecosystem types as well as a range of environmental conditions (e.g., the same ecosystem type but located much further north or south). Ecosystem types visited included agriculture (Russell Ranch and Gallo Vineyard) and pasture (Sherman Island-Pasture), oak savanna/oak woodland (Tonzi Ranch and San Joaquin Experimental Range), mixed conifer and broadleaf forest (Blodgett Forest and Soaproot Saddle), montane conifer forest (Teakettle) and wetlands (Sherman Island-Wetland). At each study site, we identified and selected the set of most dominant plant species (Table 1). While more species are present at each site, we faced a tradeoff among the number of sites, species, and individuals that could be reasonably collected and properly processed while coinciding with the HyspIRI aerial data acquisitions. As such, the cPFTs included in our study are a result of the species selected and were defined based on commonly used combinations of life form, leaf form, and leaf duration, similar to those found in Bonan et al. (2002). The species measured on a particular visit were determined based on the expected seasonal change in traits of interest. For example,

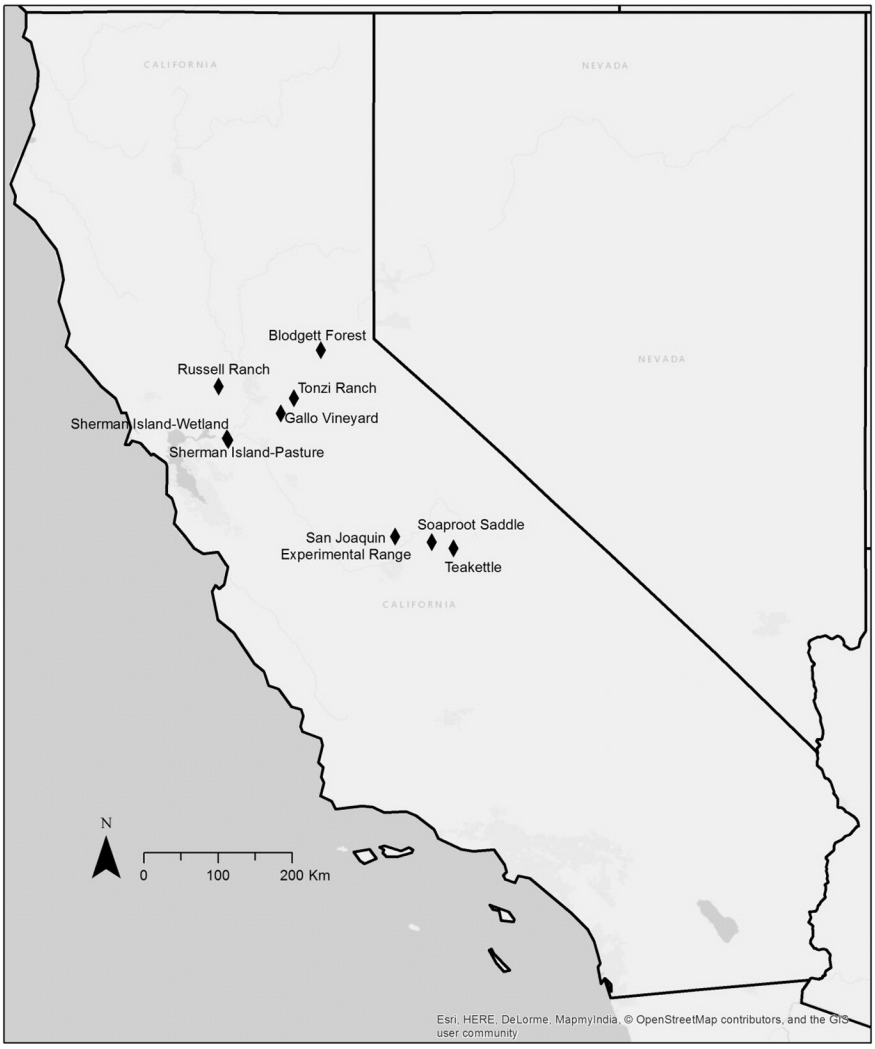

Fig. 1. Map of California showing the nine study site locations.

we expected greater seasonal change for broadleaf deciduous trees than for evergreen needleleaf trees. Although not all species were measured on every date, every effort was made to capture the full seasonal profile of each species. In 2013, samples were collected from large patches dominated by one or two species to facilitate later single-species analysis at image level. In 2014, we expanded our sampling locations to cover a wider spatial distribution and greater range of site conditions to test future image-level modeling more rigorously. No image analysis is presented here.

On each field visit, we collected sets of fully expanded leaves from at least five individuals per species (one sample per individual). For all trees, leaves were collected from the upper, sunlit portion of the canopy. For conifers, collection was restricted to second and third year needles under the assumption that these represent the majority of mature canopy needles. Leaves were immediately placed in foil packets, to ensure no light reached them, and stored on blue ice (or in a lab refrigerator) until lab measurements were made (<48 h.). We also recorded categorical attributes for each sample, including collection information (e.g., species, site) and some of the functional attributes used to define the cPFTs used in this study (Table 2). This information can provide context for observed spectral differences among samples.

2.2. Leaf spectral data collection and processing

In the lab, leaf directional-hemispherical reflectance from 350 to 2500 nm was measured for three to five replicates from each sample using an ASD FieldSpec Pro (1–2 nm sampling interval, with 3–10 nm resolution) attached to a Licor 1800-12 integrating sphere with a 6 V 10 W 3100 K illumination source (LI-COR, 1983). For broadleaf samples, five leaves were selected at random as sample replicates. For needles and other species whose leaves did not cover the entire sample port of the LI-1800, three sample replicates were created by arranging leaves as close together as possible with no overlap (i.e., minimizing gaps between leaves) and sealing the ends of the leaves using either black rubber gaskets or black electric tape. All spectra, including a LI-1800 compacted $BaSO_4$ powder standard and a sphere stray light measurement, were collected in radiance. Absolute reflectance (R_{sample}) was calculated for each leaf spectrum using its measured radiance ($I_{reflected}$), corresponding standard measurement ($I_{reference}$), correcting for stray light (I_{stray}) and multiplying by the efficiency of the $BaSO_4$ standard ($R_{reference}$) (Eq. (1)).

$$R_{sample} = \frac{(I_{reflected} - I_{stray}) * R_{reference}}{(I_{reference} - I_{stray})} \quad (1)$$

For replicates with potential gaps between leaves, a high resolution (12 MP+) digital photo was collected while the replicate was mounted in the integrating sphere port. A Matlab thresholding algorithm was developed to estimate the fraction of gap within the area illuminated in the sphere. Using this gap fraction estimate, we corrected the reflectance spectrum following the approach used in Mesarch et al. (1999) and further tested in Yáñez-Rausell et al. (2014a, 2014b). In this

Table 1
The table shows the conventional plant functional types (cPFTs), species, ecosystem types, field sites, and seasons for which leaf spectra were collected. RURA, Russell Ranch; GALL, Gallo Vineyard; SHIP, Sherman Island-Pasture; SHIW, Sherman Island-Wetland; TEAK, Teakettle; BLOF, Blodgett Forest; SOAP, Soaproot Saddle; TONZ, Tonzi Ranch; SJER, San Joaquin Experimental Range. Sp, Sm, and F refer to the season (Spring, Summer, and Fall, respectively) in which data were collected at a particular site for each species.

cPFT and species	Ecosystem type	Field sites (and seasons)
Broadleaf annual crop		
Zea mays	Agriculture	$RURA_{Sm}$
Deciduous broadleaf shrub		
Vitis vinifera	Agriculture	$GALL_{Sm/F}$
Perennial herb		
Lepidium latifolium	Pasture	$SHIP_{Sp/Sm/F}$
Wetland emergent perennial		
Schoenoplectus acutus	Wetland	$SHIW_{Sp/Sm/F}$
Typha spp.	Wetland	$SHIW_{Sp/Sm/F}$
Evergreen needleleaf tree		
Abies magnifica	Montane conifer forest	$TEAK_{Sm/F}$
Abies concolor	Mixed conifer-broadleaf forest; montane conifer forest	$BLOF_{Sp/Sm/F}$, $SOAP_F$, $TEAK_{Sm/F}$
Calocedrus decurrens	Mixed conifer-broadleaf forest	$BLOF_F$, $SOAP_F$
Pinus jeffreyi	Montane conifer forest	$TEAK_{Sm/F}$
Pinus ponderosa	Mixed conifer-broadleaf forest	$BLOF_{Sp/Sm/F}$, $SOAP_{Sm/F}$
Pinus lambertiana	Mixed conifer-broadleaf forest	$SOAP_F$
Pinus sabiniana	Oak woodland-savanna	$TONZ_{Sp/Sm/F}$, $SJER_{Sm/F}$
Evergreen broadleaf tree		
Quercus chrysolepsis	Mixed conifer-broadleaf forest	$SOAP_{Sp/Sm/F}$
Quercus wislizeni	Oak woodland-savanna	$SJER_{Sp/Sm/F}$
Deciduous broadleaf tree		
Quercus douglasii	Oak woodland-savanna	$TONZ_{Sp/Sm/F}$, $SJER_{Sp/Sm/F}$
Quercus kelloggii	Mixed conifer-broadleaf forest	$BLOF_{Sp/Sm/F}$, $SOAP_{Sp/Sm/F}$
Evergreen broadleaf shrub		
Arctostaphylos viscida	Mixed conifer-broadleaf forest	$SOAP_{Sm/F}$
Ceanothus cordulatus	Montane conifer forest	$TEAK_{Sm/F}$

correction, reflectance is linearly scaled by the fraction of leaf cover within the sample port using the equation below in which ρ is leaf reflectance, ρ_{total} is the measured reflectance, and GF is the gap fraction (Eq. (2)).

$$\rho = \frac{\rho_{total}}{(1-GF)} \quad (2)$$

All processed spectra were averaged by sample. The hemispheric scattering of light in the integrating sphere did reduce the spectrometer signal and increased the noise, even with substantial increases in the integrating period (number of spectra averaged for each measurement), particularly for longer wavelengths. Furthermore, the BaSO4 spectrum is somewhat noisy in this region (>2000 nm). The SWIR portion of the spectrum (1600–2500 nm) was smoothed using a Savitzky-Golay filter (Savitzky & Golay, 1964). This filter type was used because we could control the amount of smoothing by fine tuning the parameters, and it is known to remove noise while preserving absorption features (Ruffin & King, 1999). For our data, we first smoothed the 1600–2500 nm wavelengths using a second degree polynomial and a kernel size of ±25 nm. After examining the resulting spectra, we applied a secondary smoothing to the 1960–2500 nm wavelengths, again using a second degree polynomial with a kernel size of ±50 nm. Similar smoothing approaches have been used in a number of leaf spectra studies to reduce noise (e.g., Stimson et al., 2005; Atzberger et al., 2010; Rautiainen et al., 2012). Visual assessment of smoothed spectra confirmed that absorption features remained intact. The resulting spectral library contained 505 individual plant spectra, representing observations from eight cPFTs, eighteen species, six ecosystem types and nine sites (with multiple seasons for most species and sites).

Using the leaf spectra, we calculated a suite of nearly one hundred spectral functional metrics. These metrics approximate, or are sensitive to, a range of leaf functional traits (e.g., water, pigments, dry matter, leaf structure). They included reflectance at a given wavelength, simple ratios, normalized differences indices, optimized indices, spectral features and trait estimates based on published empirical equations. We examined the cross-correlations among this large group of metrics and selected a subset of eleven to represent a range of ecologically relevant leaf functional traits in our study (Table 3). We recognize the resulting values are not directly measured traits, but are quantitative characterizations of the leaf traits in our data set, and are interpreted in the published literature (Table 3) to represent the traits we describe.

Table 2
List and description of the categorical attributes recorded for each leaf sample, including the potential values for a given attribute or reference to another table containing this information.

Attribute	Description
Species	Species from which sample was collected
cPFT	Conventional plant functional type (see Table 1)
Leaf form	Broadleaf, needleleaf, or other
Leaf duration	Annual, deciduous, evergreen, or perennial
Life form	Graminoid, herbaceous, shrub, or tree
Campaign	Combination of month and year sample was collected
Season	Spring, summer, or fall
Ecosystem type	Ecosystem type from which sample was collected (see Table 1)
Site	Site at which sample was collected (see Table 1)

2.3. Analysis

2.3.1. Overview

We have provided a work flow diagram (Fig. 2) to outline the steps taken in our analysis. First, we summarized the spectral data using Principal Components Analysis (PCA) and evaluated the association of the resulting principal components (PCs) with categorical attributes and spectral functional metrics. We then applied agglomerative hierarchical clustering and Dynamic Hybrid Clustering to the standardized PC coordinate values to define spectral clusters. We tested the validity and stability of these clusters using an iterative subsampling approach, and finally assessed the clusters' membership (using categorical attributes) and functional profiles. All statistical analyses were done using the R statistical software package (R Core Team, 2015), implemented in RStudio (RStudio Team, 2015). We also used the R package 'dynamicTreeCut' for the hybrid clustering (Langfelder et al., 2014).

Table 3
List of the functional metrics calculated from the spectra. These included single band values, simple ratios, optimized indices and empirical estimates. The wavelengths used in each metric and the appropriate references are listed.

Trait	Metric type	Wavelength(s) used (nm)	Reference
Leaf structure	Single band	800	Sims & Gamon (2002)
Surface scattering	Single band	445	Sims & Gamon (2002)
Lignin	Simple ratio	1028, 2101	Almeida & De Souza Filho (2004)
Equivalent water thickness (EWT)	Empirical estimation	1062, 1393	Féret et al., 2011
Leaf mass area (LMA)	Empirical estimation	1368, 1722	Féret et al. (2011)
Pigments	Narrow-band normalized difference index	660, 800	Sims & Gamon (2003)
Total chlorophyll	Optimized index	680–730; 755–780; 780–800	Féret et al. (2011)
Chlorophyll a	Simple ratio	638, 807	Almeida & De Souza Filho (2004)
Chlorophyll b	Simple ratio	648, 807	Almeida & De Souza Filho (2004)
Carotenoids	Empirical estimation	530, 800	Féret et al. (2011)
Pigment ratios	Narrow-band normalized difference index	531, 570	Sims & Gamon (2002)

2.3.2. Summarizing and interpreting information in the leaf spectra

We used PCA both to assess the spectral variability within the data and to summarize the information present in the spectra. PCA transforms the data by maximizing the variance explained by each orthogonal component, and it has been widely used in remote sensing to reduce spectral dimensionality (e.g., Féret & Asner, 2014; Hesketh & Sánchez-Azofeifa, 2012). Reducing the dimensionality of the spectra is also essential for clustering analysis because high dimensional data create sparse distance matrices, leading to poor results (Berkhin, 2006; Steinbach et al., 2004).

In order to determine which PCs to retain in our analysis, we examined the scree plot of variance explained by each component, as well as the cumulative spectral variance explained as new components were added. The first three PCs explained >98% of the cumulative variance, and each of these PCs explained >5% of the spectral variance. Additional PCs explained ≤0.6% of the spectral variance each. Based on these observations, only the first three PCs were retained for use in subsequent analyses. In order to evaluate the spectral information represented by these PCs, we calculated the correlation of PCs 1–3 with each wavelength measured and with the spectral functional metrics previously calculated (see Table 3). We also calculated the variance explained (R^2) by each categorical variable (see Table 2) within each PC.

2.3.3. Clustering

2.3.3.1. Agglomerative hierarchical clustering. We used the standardized spectral PC coordinates of each individual sample in our clustering analysis, rather than an aggregated value (e.g., species' mean). By neither aggregating nor restricting observations from the same cPFT, species, site, and sampling date to be in the same cluster, we allowed for intraspecific and intra-cPFT variance to be incorporated into the clustering. We applied agglomerative hierarchical clustering analysis because it is a very flexible algorithm, making no assumptions about data structure or number of clusters prior to clustering and can use any form of similarity or distance metric (Berkhin, 2006). It allows users to see nested relationships among observations, making the results more interpretable (Chipman & Tibshirani, 2005). Each observation starts as its own cluster, and observations and clusters are subsequently merged based on a user-defined linkage until all observations are in a single group (Kaufman & Rousseeuw, 2005). We calculated the dissimilarity matrix using Euclidean distance, and spectral clusters were merged using the Ward method, which combines the two clusters resulting in the minimum increase in the sum of the within cluster variance at each merge step (Ward, 1963). This method, similar to approaches used in other studies with spectral data (e.g., Asner et al., 2009; Göttlicher et al., 2011), results in fairly compact, homogenous clusters, and is less susceptible to noise and outliers, but can be biased toward spherical clusters (Wishart, 1998).

The resulting clustering solution was visualized using a dendrogram, which represents the clustering structure of the data. The height at which two clusters merge in the dendrogram corresponds to the distance between those two clusters. We used two metrics to assess the quality of the dendrogram, and thus, the strength of the clustering structure in the data: the cophenetic correlation coefficient and the agglomerative coefficient. The cophenetic correlation coefficient is a measure of how well the dendrogram represents the dissimilarities in the

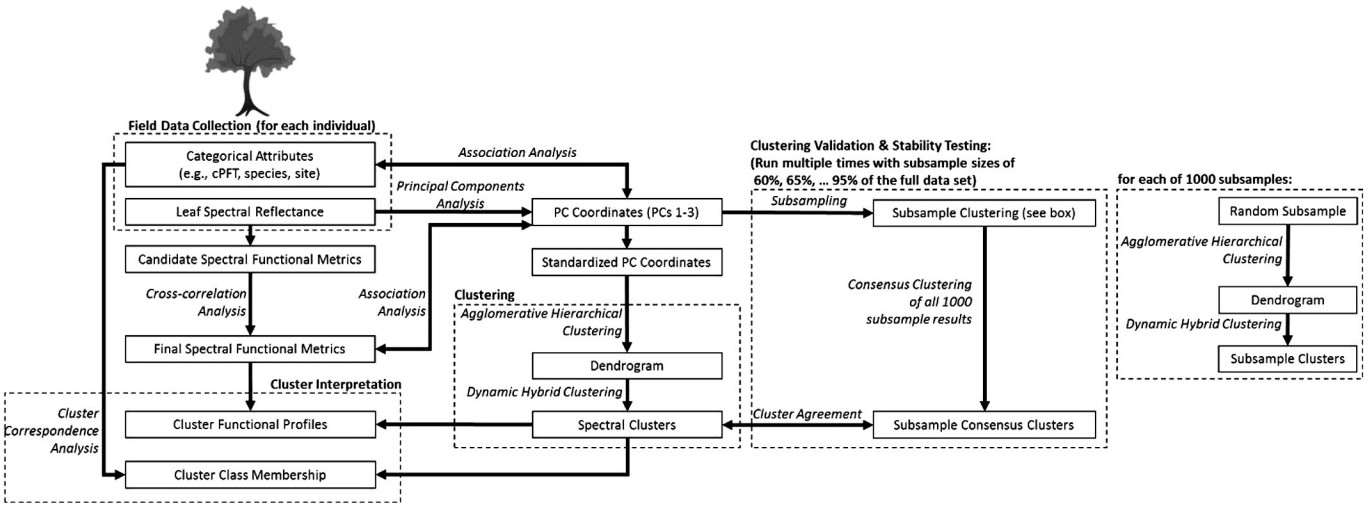

Fig. 2. Conceptual diagram showing the analysis steps for the study.

original set of observations, with values closer to one indicating better representation (Sokal & Rohlf, 1962). The agglomerative coefficient is the average of one minus the distance of each object to its nearest cluster divided by the distance between the final two clusters to merge, and greater structure within the data produces values closer to one (Kaufman & Rousseeuw, 2005).

2.3.3.2. Dynamic Hybrid Clustering. There are some drawbacks to agglomerative hierarchical clustering, particularly for defining hard classes. The final clusters must be determined by the user (who selects either a total number of clusters or a height at which to cut the dendrogram, somewhat subjectively). Furthermore, clusters are not re-visited after merging, meaning once an observation is merged into a cluster, it cannot be placed in a different cluster (Hardle & Simar, 2003; Langfelder et al., 2008). To overcome these drawbacks, hybrid methods have been proposed that combine agglomerative hierarchical clustering with other approaches, such as k-means or divisive hierarchical clustering, to gain the advantages of each (Chipman & Tibshirani, 2005; Langfelder et al., 2008; van der Laan & Pollard, 2003). We selected the Dynamic Hybrid Clustering algorithm proposed by Langfelder et al. (2008) and implemented in the R package 'dynamicTreeCut'. In this method, agglomerative hierarchical clustering is combined with Partitioning Around Medoids (PAM) clustering to balance the tendency of former to create small clusters, with tendency of the latter to create large ones. The final number of clusters is determined by the algorithm. Dynamic Hybrid Clustering begins by using the dendrogram created by agglomerative hierarchical clustering to identify a set of core clusters meeting the following criteria:

1. contains a set minimum number of observations
2. excludes observations too far from the cluster
3. distinct from surrounding clusters
4. the core (tip of dendrogram branch) is tightly connected

After these core clusters are identified, unassigned observations are assigned to the nearest cluster if they are sufficiently close. The algorithm allows the user to more finely tune the four criteria above. In our implementation, we specified the minimum number of observations within a core cluster to be two (the minimum possible given that each leaf on the dendrogram represents one single observation). We explored different settings for the other criteria, and elected to use the default settings for this analysis, as these parameter settings have been successfully applied in other applications (Langfelder et al., 2008).

2.3.3.3. Cluster validation and stability assessment. With any clustering technique, cluster validity and stability are important considerations for both interpretation and replicability. Ideally, the final clusters are robust to both samples containing different observations and different sample sizes. One method for evaluating the validity and stability of clusters is to use subsampling or bootstrapping (Bhattacharjee et al., 2001; Monti et al., 2003). Monti et al. (2003) describe one such approach: consensus clustering. In consensus clustering, a user-defined number of randomly selected subsets of observations are clustered, and the agreement among iterations is used to construct a consensus matrix. This consensus matrix is used as a similarity matrix in agglomerative hierarchical clustering to assign a final clustering. We used the R package ConsensusClusterPlus (Wilkerson & Hayes, 2010), which implements and extends the consensus clustering algorithm by Monti et al. (2003), allowing users to define their own clustering algorithm. The final number of clusters was set to the number found by the Dynamic Tree clustering based on all observations. We tested the algorithm using eight different sample sizes (60%–95% of observations, in 5% intervals). For each sample size, consensus clustering with Dynamic Hybrid Clustering was run for 1000 random subsamples. Consensus matrices and the average item and cluster consensus values across subsamples were used to assess the validity of the clusters found. The kappa coefficient of agreement was calculated between the full sample cluster assignment and each sample size's cluster assignment. Higher kappa values indicate the clustering solution is insensitive random sampling at the given sample size, and similar kappa values across sample sizes would demonstrate the clustering is insensitive to sample size.

To further assess the clustering solution, we calculated 1) the mean within-cluster distance between observations (a measure of cluster compactness), 2) the mean distance between observations in a cluster to those in other clusters (a measure of cluster separation), and 3) the mean cluster silhouette coefficient (a combined metric) (Rousseeuw, 1986). The silhouette coefficient was calculated for each observation (i) using the equation below, where a_i is the average dissimilarity of i with all other observations within the same cluster and b_i is the lowest average dissimilarity of i to any other cluster (Eq. (3)):

$$s_i = \frac{(b_i - a_i)}{max\{a_i, b_i\}} \tag{3}$$

To evaluate the spectral similarity/dissimilarity among the resulting clusters, we calculated the spectral angle (cosine distance) between each pair of cluster mean spectra. This metric allows us to determine if two clusters differ in their mean spectral shape, rather than just in magnitude.

2.3.4. Cluster correspondence with categorical attributes and functional metrics

To interpret cluster membership, we compared cluster correspondence to sample metadata (see Table 2), using the Adjusted Rand index. This index measures the agreement between two sets of partitions, accounting for chance agreement when group size is unequal (Hubert & Arabie, 1985; Milligan & Cooper, 1986). The index ranges from zero to one, with one representing perfect correspondence, and zero representing the expected correspondence between two random partitions. We also calculated the Fowlkes-Mallows Index, or the probability that a pair of observations in the same group of one partition set are found together in the other partition set (Fowlkes & Mallows, 1983). To identify which species and cPFTs were most strongly related with individual clusters, we calculated the point-biserial correlation coefficient between each cluster and class.

Lastly, we evaluated the functional profiles of the clusters by examining and comparing their respective distributions of spectral functional metrics. To test for significant differences in the means and dispersions of these metrics among clusters, we used permutational MANOVA (PERMANOVA) and permutational dispersion analysis (PERMDISP). These two approaches compare cluster data to those of random permutations to determine if the clusters explain significantly more variance than would be expected from random partitions.

3. Results

3.1. Summarizing and interpreting information in the leaf spectra

The correlations between the PCs 1–3 and the wavelengths measured show that each PC carries information from different regions of the spectrum. While the entire spectrum was moderately correlated with the first PC, the strongest correlations were found with the near-infrared (NIR) region, approaching one for most wavelengths between 750 and 1400 nm (Fig. 3). The second PC was most strongly correlated with the shortwave-infrared (SWIR) region, and the third PC with the visible (VIS) region.

Plotting the individual observations in PC space, we visually assessed how they grouped by cPFT (Fig. 4a–b) and species (Fig. 4c–d). While we observed some differentiation among cPFTs and species, there appeared to be substantial overlap among observations from different groups on all three PC axes, indicating that some observations from different species, or even different cPFTs, were very spectrally similar. For example, separation is evident between evergreen needleleaf trees and deciduous

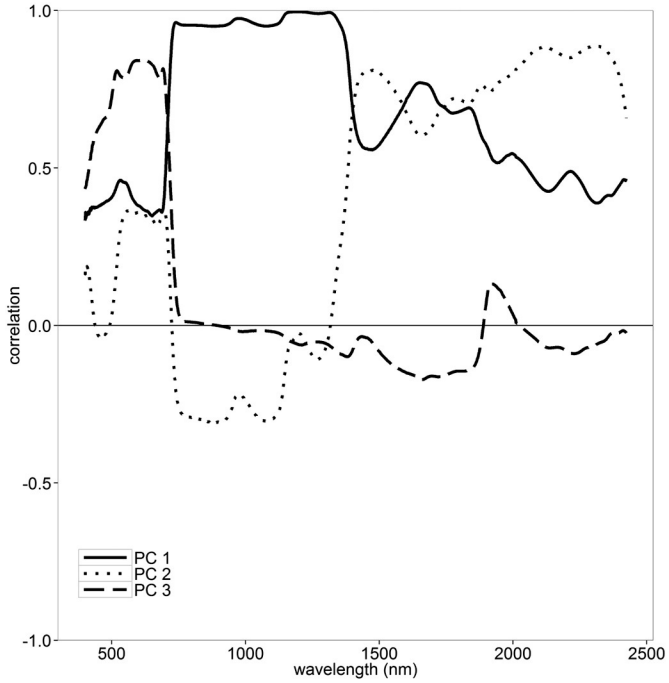

Fig. 3. Graph showing the correlation coefficient between each wavelength measured and the first three PCs: PC1 in solid line; PC2 in dotted line; PC3 in dashed line.

Table 4
Variance explained (R^2) by each categorical variable (defined in Table 2) for each PC.

	PC1	PC2	PC3
Species	0.20	0.80	0.47
cPFT	0.15	0.68	0.24
Leaf form	0.01	0.53	0.04
Leaf duration	0.12	0.55	0.05
Life form	0.10	0.11	0.02
Campaign	0.48	0.23	0.08
Season	0.04	0.02	0.05
Ecosystem type	0.15	0.35	0.04
Site	0.18	0.38	0.09

broadleaf trees, but separation between deciduous broadleaf trees and deciduous broadleaf shrubs is less clear. Looking at species, overlap occurs on all axes for observations from *Zea mays* (corn) and *Quercus kelloggii* (black oak).

We can assess how much information each PC carries about the samples' categorical attributes (defined in Table 2) by looking at the variance explained (R^2) (Table 4). It is important to note that these values do not take into account the inter-relationships among attributes and only provide a measure of how well a particular attribute is associated with each PC. cPFT explains the majority (68%) of the variance in PC 2, as well as 15% and 24% of the variances in PCs 1 and 3, respectively. Species follows a similar pattern, explaining greater proportions of the PC variances overall (20%, 80% and 47% for PCs 1–3). Other categorical attributes also explained substantial proportions of the variance in the three

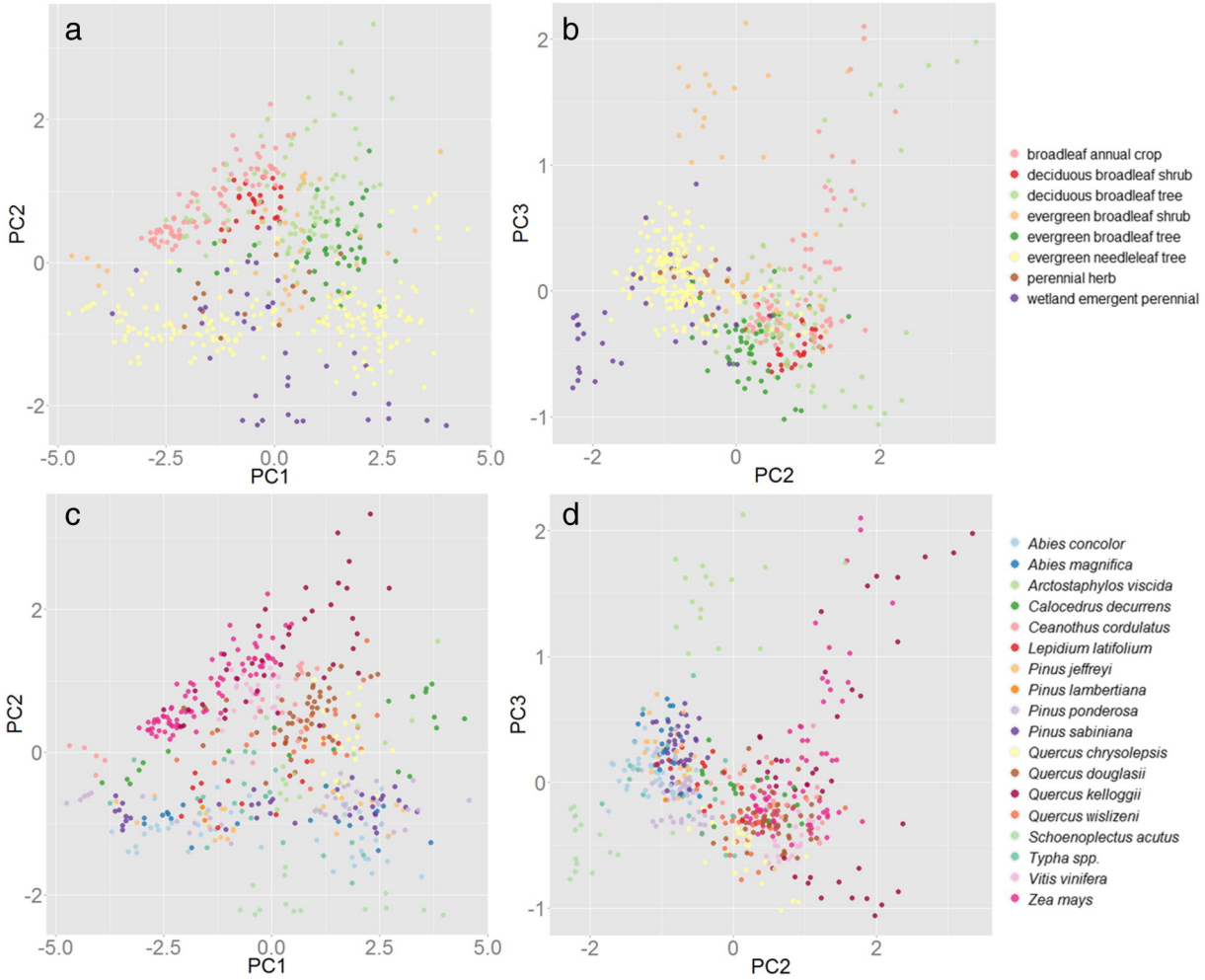

Fig. 4. Scatterplots show the PC coordinates of the observations for PC1 v. PC2 (a & c) and PC2 v. PC3 (b & d). Points are colored by cPFT (a & b) and species (c & d).

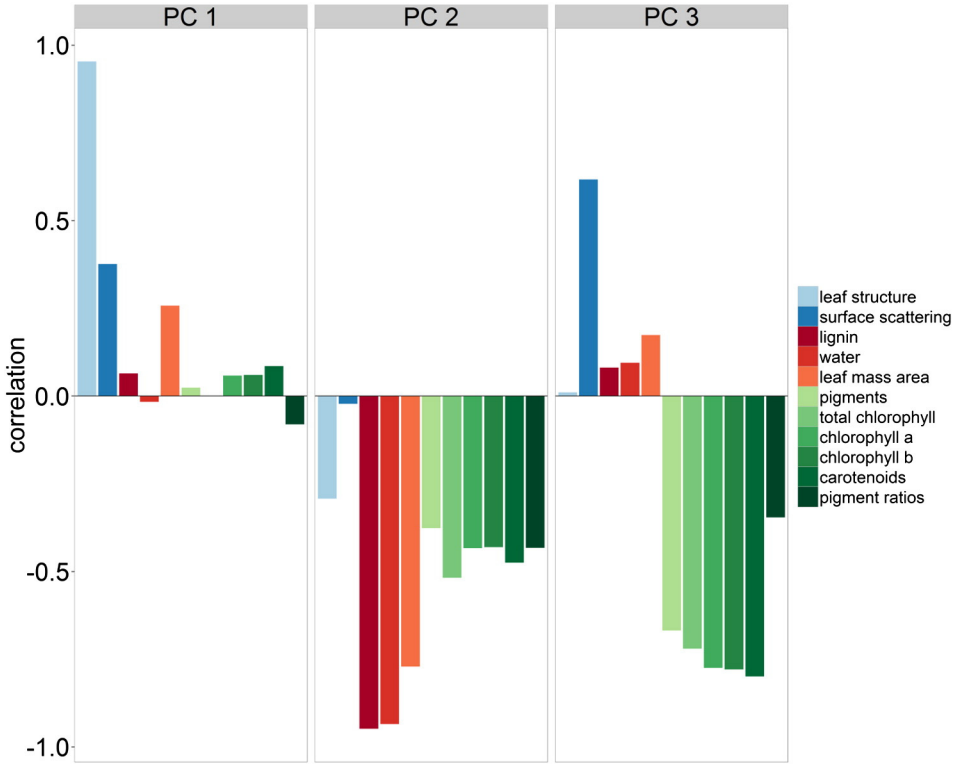

Fig. 5. Fig. shows the correlation coefficient between each spectral functional metric and each PC. Metrics related to surface and structure are in blue, related to water and dry matter are in red, and related to pigments are in green. (For interpretation of the references to in this figure legend, the reader is referred to the web version of this article.)

PCs. For PC 1, Campaign (i.e., month and year) and Site explained 48% and 18% of the variance, respectively. All categorical attributes, with the exceptions of season and life form, explained >35% of the variance in PC 2; however, leaf form and leaf duration explained >50% of the variance. Species and cPFT were the only variables to explain >10% of the variance in PC 3.

We also examined the correlations between spectral functional metrics (see Table 3 for definitions) with each of the PCs to determine what functional information is carried by each. PC 1 was strongly correlated with leaf internal structure, somewhat correlated with surface reflectance, and weakly correlated with LMA (Fig. 5). PC 2 was strongly correlated with lignin, EWT, and LMA, and PC 3 was strongly correlated with pigments. These correlations can be interpreted in light of the previously observed individual wavelength correlations with each of the PCs. They demonstrate that PC 1 mainly contains information about leaf surface and internal structure; PC 2 contains information on water, dry/leaf structure, and PC 3 contains information on pigments.

3.2. Leaf spectral clusters

The dendrogram resulting from hierarchical agglomerative clustering had a cophenetic correlation of 0.6 and an agglomerative coefficient of 0.99, indicating moderate to strong clustering structure in the data. Dynamic Hybrid Clustering resulted in twelve clusters, ranging in size from 12 to 75 observations (Fig. 6). The height at which two clusters merge indicates their similarity (i.e., more similar clusters merge first because they are closer together in data space). Clusters 8 and 9 were the most similar, then 11 and 12, followed by 6 and 7, then 1 and 2. At approximately the same distances, Clusters 3 and 4 merged, Cluster 5 merged with 6 and 7 and Cluster 10 merged with 11 and 12. Next, Clusters 8–9 merged with Clusters 10–12. At the furthest distances, Clusters 5–7 merged with 8–12, followed by 3–4, and finally, Clusters 1–2 merged with the entire set. Average within-cluster distance between observations ranged from 0.12 to 0.27, and average distance from observations within a cluster to members of other clusters ranged from 0.32 to 0.77 (Table 5). Cluster mean silhouette coefficients ranged from 0.05 (Cluster 9) to 0.48 (Cluster 5). These results demonstrate that the values within the majority of clusters are relatively homogeneous, and differences between most cluster pairs are larger than within cluster differences.

Varying subsamples and subsample sizes with consensus clustering demonstrated this solution to be stable and valid, with all subsample proportions having high (and similar) agreement with the clustering of the entire data set (Table 6). Furthermore, the cluster consensus values (i.e., average consensus value between all observation pairs in a given cluster, 0–1) were high for all subsample proportions.

The cluster spectral signatures appeared to differ across most wavelengths (Figs. 6 and 7).The within cluster spectral variance also differed across clusters and wavelengths (Fig. 6). Some clusters, such as Clusters 10–12, had very low variance across most wavelengths, while others, such as Clusters 2, 6 and 7 had higher variance. For most clusters, spectral variance was highest in the NIR region of the spectrum, and the VIS and SWIR regions had similar variance. We calculated the spectral angle (SA; cosine distance) between pairs of cluster mean spectra to compare similarities in spectral shape. An SA value of 0.0 indicates two spectra are the same shape and differ only in amplitude. SA values varied from 0.01 (between Clusters 1 and 4) to 0.335 (between Clusters 2 and 7), with an overall average of 0.145 (Table 7).

3.3. Cluster correspondence with categorical attributes

There was generally low correspondence observed between categorical attributes (defined in Table 2) and spectral clusters (Table 8). The Adjusted Rand Index (1.0 denotes perfect agreement) ranged from 0.041 (season) to 0.274 (cPFT), and the Fowlkes-Mallows Index indicated that for nearly all categorical attributes, two observations from the same category had a 0.25–0.40 probability of being assigned to the same spectral cluster. In particular, the twelve spectral clusters were

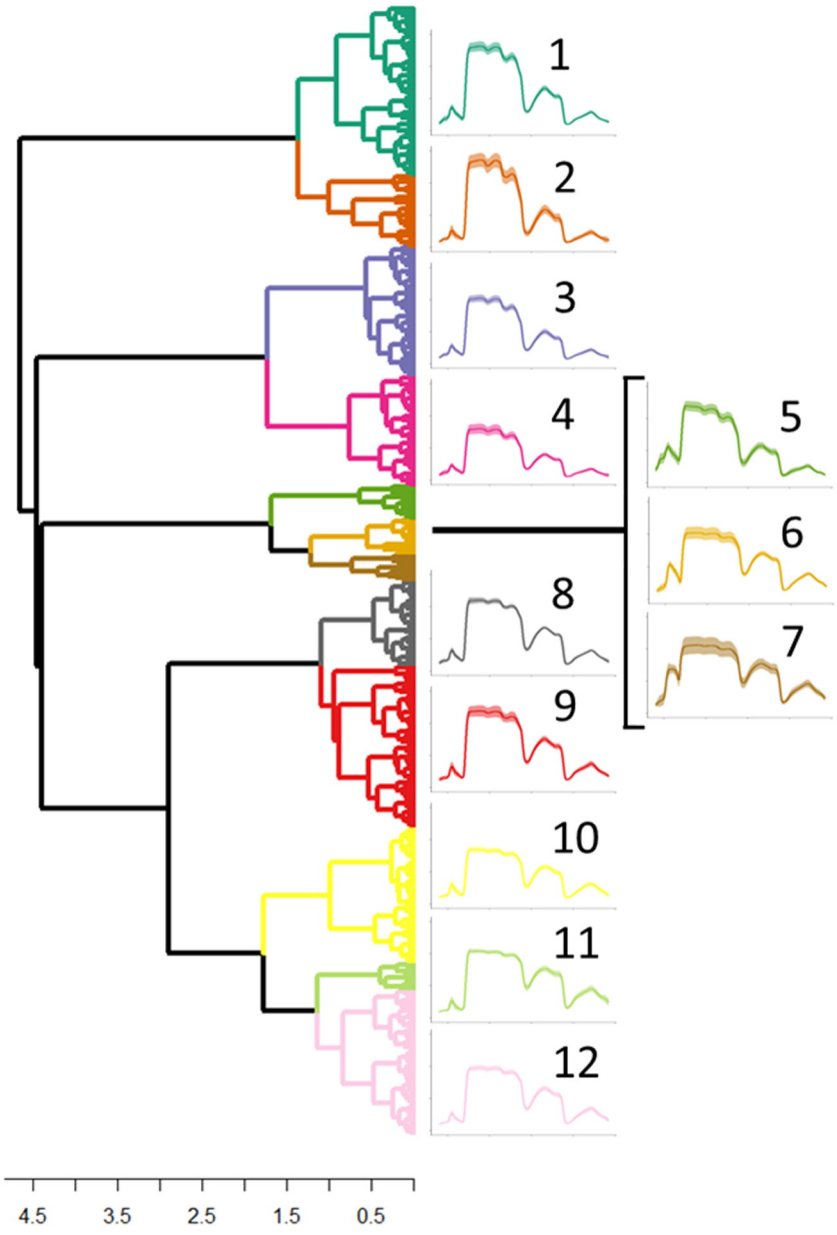

Fig. 6. Dendrogram showing the clustering of the spectral observations into twelve clusters. The dendrogram axis is in Euclidean distance. The solid lines in the spectral plots show the mean spectrum for each cluster, and the shaded areas represent ± one standard deviation.

not consistent with individual cPFTs or species, though differentiation between groups of cPFTs was observed (Fig. 8). For example, Clusters 1–4 were dominated by evergreen needleleaf tree observations, while broadleaf trees observation were found more frequently in Clusters 6–12. Still, observations from the same species had approximately a 33% probability of being clustered together, and observations from the same cPFT were clustered together with a probability of ~39% (Table 8).

Table 5
Individual cluster values for the three clustering assessment metrics.

	Mean within-cluster distance	Mean distance to other clusters	Mean silhouette coefficient
1	0.16	0.43	0.34
2	0.23	0.49	0.08
3	0.14	0.38	0.40
4	0.15	0.50	0.38
5	0.20	0.58	0.48
6	0.18	0.49	0.45
7	0.27	0.77	0.28
8	0.12	0.32	0.36
9	0.19	0.41	0.05
10	0.14	0.41	0.37
11	0.16	0.49	0.36
12	0.14	0.37	0.21

Table 6
Kappa coefficients between the clustering of all observations and for each subsample proportion tested to establish cluster stability and validity.

Subsample proportion	Kappa coefficient with full data set clustering
60%	0.80
65%	0.82
70%	0.82
75%	0.82
80%	0.82
85%	0.82
90%	0.83
95%	0.83

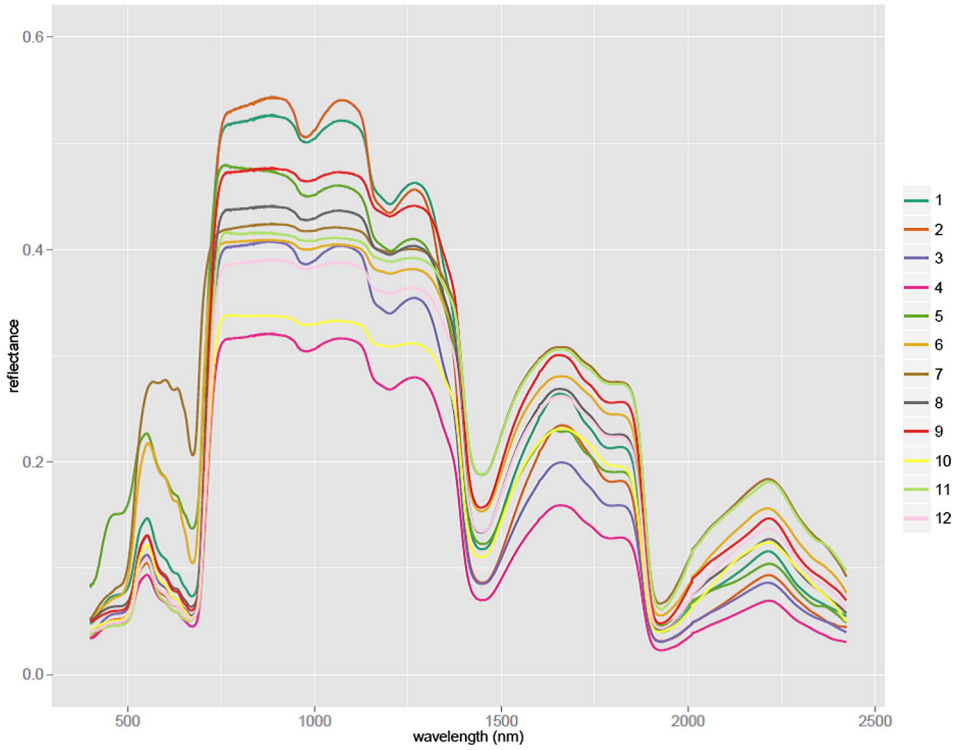

Fig. 7. Mean reflectance for each of the twelve clusters.

Table 7
Spectral angle values between each pair of cluster mean spectra (0–1). Lower values indicate two mean spectra are more similar.

	2	3	4	5	6	7	8	9	10	11	12
1	0.078	0.015	0.010	0.118	0.184	0.262	0.086	0.108	0.143	0.213	0.136
2		0.067	0.078	0.175	0.260	0.335	0.157	0.174	0.212	0.275	0.202
3			0.012	0.121	0.195	0.272	0.099	0.121	0.154	0.225	0.148
4				0.113	0.186	0.263	0.092	0.114	0.147	0.220	0.142
5					0.159	0.211	0.143	0.166	0.174	0.251	0.186
6						0.092	0.123	0.124	0.090	0.140	0.115
7							0.209	0.209	0.175	0.202	0.199
8								0.025	0.062	0.131	0.053
9									0.053	0.109	0.032
10										0.089	0.034
11											0.081

Despite this lack of correspondence to single cPFTs or species, significant correlations between clusters and certain species and cPFTs were observed (Table 9; $p = 0.05^{*}$, 0.01^{**} and 0.005^{***}). The majority of cPFTs were significantly correlated with several clusters, the exceptions being deciduous broadleaf shrub and evergreen broadleaf shrub. The evergreen broadleaf tree, perennial herb and wetland emergent perennial cPFTs were significantly correlated with two clusters each, and broadleaf annual crop, deciduous broadleaf tree and evergreen needleleaf tree cPFTs were significantly correlated with four clusters each. Likewise, the majority of species were significantly correlated with two or three different spectral clusters. *Arctostaphylos visicida*, *Quercus chrysolepsis*, and *Schoenoplectus acutus* were the exception to this observation, having just one significant cluster each (Clusters 5, 9, and 2, respectively).

Table 8
Correspondence metric values between each categorical variable (defined in Table 2) and the twelve spectral clusters.

	Adjusted Rand Index	Fowlkes-Mallows Index
Species	0.255	0.326
cPFT	0.274	0.394
Leaf Form	0.133	0.377
Leaf Duration	0.168	0.358
Life form	0.115	0.336
Campaign	0.168	0.276
Season	0.041	0.255
Ecosystem type	0.150	0.289
Site	0.166	0.267

In general, Clusters 1, 3, and 4 were mainly correlated with evergreen needleleaf trees, Cluster 2 with wetland emergent perennials, and Clusters 5–12 with broadleaf species. The most spectrally similar Clusters, 8 and 9, were dominated by small-leaved deciduous (*Q. douglasii*) and evergreen (*Q. wislizeni, Q. chrysolepsis*) oaks. Cluster 8 was also significantly correlated with *C. cordulatus* (evergreen broadleaf shrub) and *Lepidium latifolium* (perennial herbaceous), and Cluster 9 with *Calocedrus decurrens*. Clusters 11 and 12 were dominated by large-leaved annual and deciduous species, with Cluster 11 being almost entirely *Q. kelloggii*, and Cluster 12 having mainly agricultural species (*Vitis vinifera, Zea mays*). Joining Clusters 11 and 12 later in the clustering process (Fig. 6), Cluster 10 was strongly correlated with *Z. mays*. The next most spectrally similar pair of Clusters, 6 and 7, also

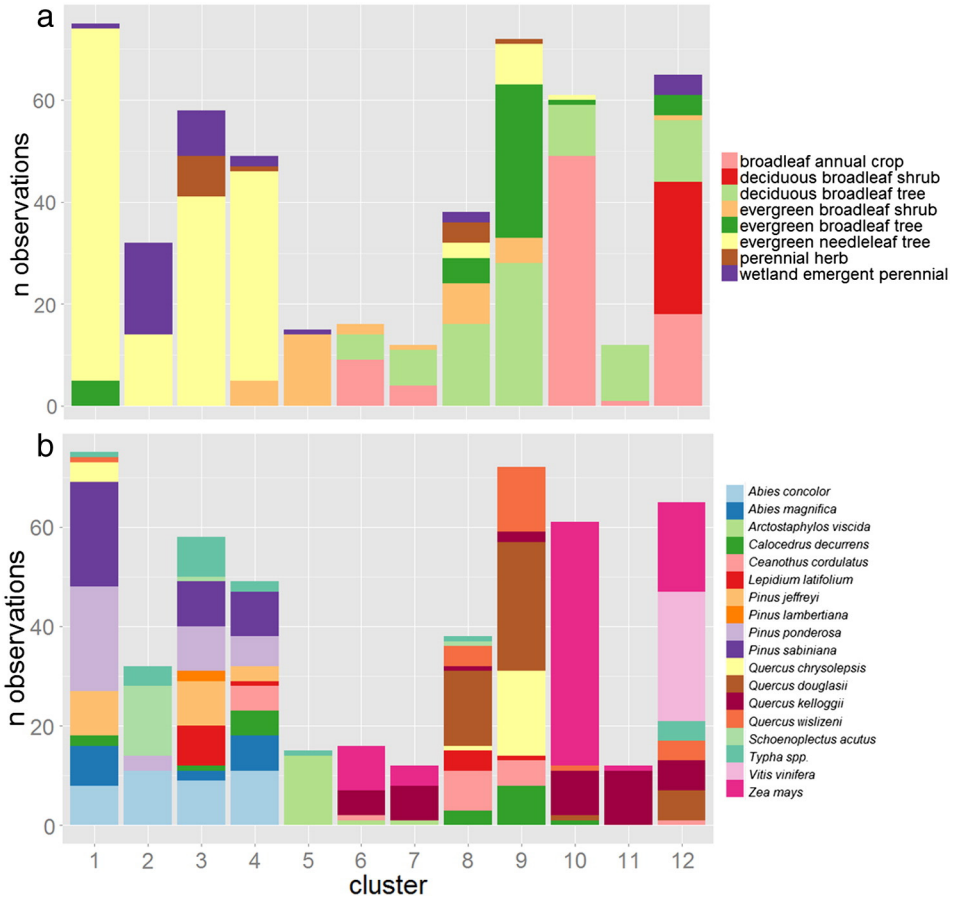

Fig. 8. Figures show the cPFT (a) and species (b) membership of each of the twelve spectral clusters.

both contained *Q. kelloggii* and *Z. mays*. Cluster 5 was the next closest cluster to these and contained almost solely *A. viscida*. Cluster 1, the largest cluster, was dominated by pine species (*Pinus ponderosa, P. sabiniana, P. jeffreyi*) and *Abies magnifica*. The next most similar cluster was Cluster 2, which was most strongly correlated with emergent wetland perennial species (primarily *S. acutus*) and *Abies concolor*. Clusters

Table 9
Correlation coefficient between each group (rows) and each cluster (columns) ($p = 0.05^{*}, 0.01^{**}, 0.005^{***}$). For clarity, correlations not found to be significant at $p = 0.05$ are not shown.

	1	2	3	4	5	6	7	8	9	10	11	12
cPFT												
Broadleaf annual crop	–	–	–	–	–	0.20***	0.07*	–	–	0.65***	–	0.12*
Deciduous broadleaf shrub	–	–	–	–	–	–	–	–	–	–	–	0.61***
Deciduous broadleaf tree	–	–	–	–	–	–	0.17*	0.18*	0.23*	–	0.30***	–
Evergreen broadleaf shrub	–	–	–	–	0.59***	–	–	–	–	–	–	–
Evergreen broadleaf tree	–	–	–	–	–	–	–	0.04*	0.47***	–	–	–
Evergreen needleleaf tree	0.50***	0.05***	0.27***	0.33***	–	–	–	–	–	–	–	–
Perennial herb	–	–	0.24***	–	–	–	–	0.13***	–	–	–	–
Wetland emergent perennial	–	0.49***	0.11*	–	–	–	–	–	–	–	–	–
Species												
Abies concolor	–	0.26***	0.11*	0.18***	–	–	–	–	–	–	–	–
Abies magnifica	0.17***	–	–	0.20***	–	–	–	–	–	–	–	–
Arctostaphylos manzanita	–	–	–	–	0.90***	–	–	–	–	–	–	–
Calocedrus decurrens	–	–	–	0.1*	–	–	–	0.06*	0.15***	–	–	–
Ceanothus cordulatus	–	–	–	0.1*	–	–	–	0.25***	–	–	–	–
Lepidium latifolium	–	–	0.24***	–	–	–	–	0.13**	–	–	–	–
Pinus jeffreyi	0.16***	–	0.20***	0.03*	–	–	–	–	–	–	–	–
Pinus lambertiana	–	–	0.18*	–	–	–	–	–	–	–	–	–
Pinus ponderosa	0.32***	–	0.11**	0.06*	–	–	–	–	–	–	–	–
Pinus sabiniana	0.32***	–	0.11***	0.13***	–	–	–	–	–	–	–	–
Quercus chrysolepsis	–	–	–	–	–	–	–	–	0.38***	–	–	–
Quercus douglasii	–	–	–	–	–	–	–	0.29***	0.37***	–	–	–
Quercus kelloggii	–	–	–	–	–	–	0.29*	–	–	–	0.48***	–
Quercus wislizeni	–	–	–	–	–	–	–	0.08***	0.26***	–	–	–
Schoenoplectus acutus	–	0.60***	–	–	–	–	–	–	–	–	–	–
Typha spp.	–	0.11*	0.17**	–	–	–	–	–	–	–	–	–
Vitis vinifera	–	–	–	–	–	–	–	–	–	–	–	0.61***
Zea mays	–	–	–	–	–	0.20***	–	–	–	0.65***	–	0.12*

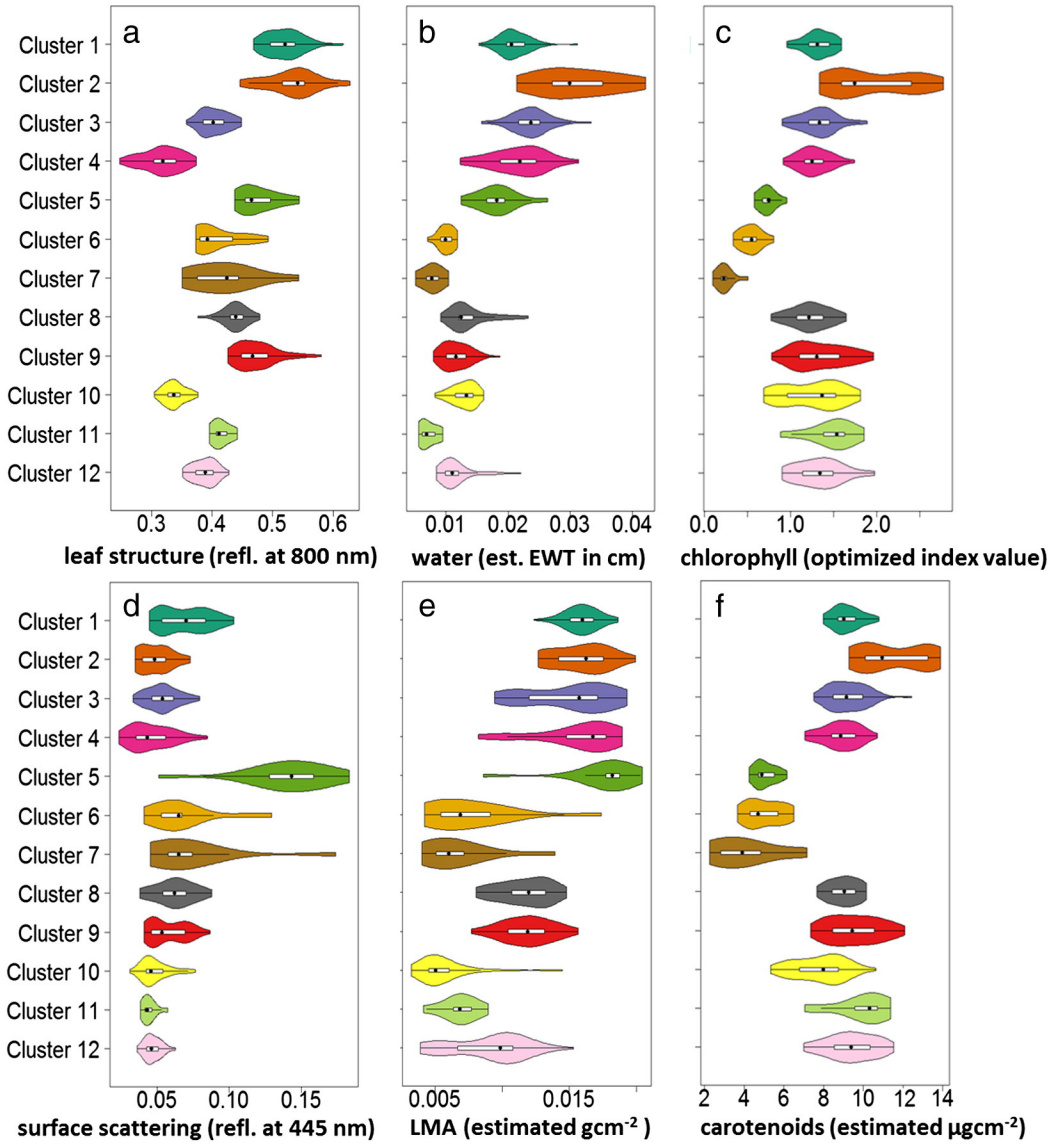

Fig. 9. Panels show the density distributions for the entire range of values for six spectral functional metrics by cluster (a: leaf structure; b: water; c: chlorophyll; d: surface scattering; e: LMA; f: carotenoids). Black dots indicate the median value, and white boxes indicate the 25th to 75th quantile range.

3 and 4 were also both dominated by evergreen needleleaf tree species, particularly *Pinus* spp. in Cluster 3 and *Abies* spp. in Cluster 4. Cluster 3 was also significantly correlated with L. *latifolium* and *Typha spp.*, while Cluster 4 was significantly, but weakly, correlated with *C. decurrens* and *C. cordulatus*.

3.4. Cluster spectral functional profiles

The twelve spectral clusters also differed functionally from one another. Fig. 9 shows the density distributions for a subset of six of the spectral functional metrics: leaf internal structure, leaf surface reflectance LMA, EWT, total chlorophyll, and carotenoids. Fig. 10 shows the distributions for all functional metrics for each cluster.

Considering the three metric types associated with each PC (leaf surface/interior structure, dry matter/water, and pigments), we observed variation in cluster values across the range of each of these axes. Clusters 1 and 2 had very high internal structure values, and high and average surface reflectance values, respectively. Clusters 5, 8 and 9 had high internal structure values, but varied in their surface reflectance values (very high, high and average, respectively). We observed average internal structure values for Clusters 6, 7 and 11, but Clusters 6 and 7 had high surface reflectance values, while Cluster 11 had low values for this metric. Clusters 3 and 12 had low internal structure values, and average and low surface reflectance values, respectively. Lastly, Clusters 4 and 10 had very low internal structure values and average surface reflectance values.

Cluster 5 had the highest LMA values, as well as high lignin and EWT. Clusters 1, 2 and 4 had high LMA, and while Clusters 1 and 4 also had high lignin and EWT, Cluster 2 had very high lignin and EWT values. Clusters 3, 8 and 9 had average LMA values, but differed in their lignin and EWT values. Cluster 3 had high lignin and EWT; Cluster 8 had average lignin and low EWT, and Cluster 9 had low lignin and EWT. Low LMA and very low lignin values were observed for Clusters 6 and 12, and these clusters had very low and low EWT values, respectively. Clusters 7, 10 and 11 had very low LMA values. Additionally, Cluster 10 had low lignin and EWT, while Clusters 7 and 11 had very low lignin and EWT.

High to very high pigment values were observed for Clusters 2 and 11. In particular, Cluster 2 had very high total chlorophyll and

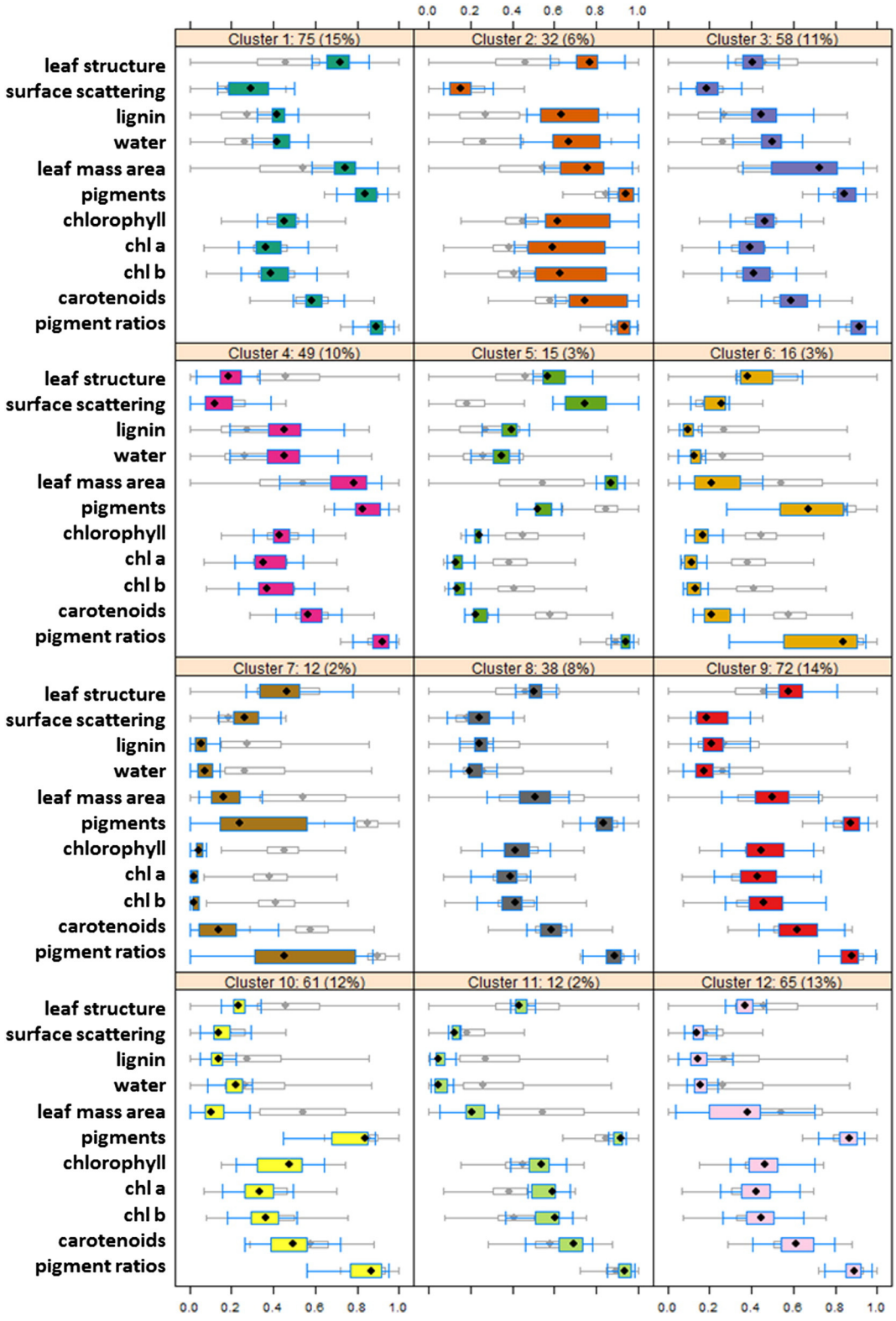

Fig. 10. Boxplots show the distributions of spectral functional metrics for each cluster (in color) with respect to the global distributions (in grey). All metrics were scaled to 0–1 over their range. Black dots show the median value, and boxes show the 25th through 75th quantiles. The number of observations and percent of total observations in each cluster is shown next to the cluster label.

carotenoids, and Cluster 11 had very high chlorophyll a and b values. Both clusters had high pigment ratio values. Half the clusters had average pigment and pigment ratio values (Clusters 1, 3, 4, 8, 9 and 12), as did Cluster 10, with the exception of having low carotenoids. Clusters 5, 6 and 7 had very low pigment values overall, but differed in their pigment ratios. Cluster 5 had high pigment ratio values, and Cluster 6 had average, but widely distributed pigment ratio values. Cluster 7 had very low, widely distributed values.

PERMANOVA indicated that the functional profiles (i.e., the unique combinations of trait metric values) were statistically significantly different for all pairs of clusters (Fig. 10). Spectral clusters explained 69% of the variance in functional metric profiles, whereas species explained 65% and cPFT explained only 48%, demonstrating that the spectral clusters were better able to represent functional differences among observations. PERMDISP analysis showed significant differences in the dispersion of metric profiles for only a few cluster pairs, further supporting the PERMANOVA analysis. Because these results support the hypothesis that spectral clusters are functionally distinct, we can further posit that the dendrogram (Fig. 6) represents the relationship among functional traits within the data set. Each node of the dendrogram should indicate some key functional difference(s) between the observations split by this node.

The first node in the dendrogram (Fig. 6) separated Clusters 1 and 2 (very high leaf internal structure) from the remaining clusters (low -to-average internal structure values). These two clusters both generally had higher LMA, lignin and EWT values, and average-to-high pigment values. Between the two clusters, Cluster 1 had higher leaf surface reflectance values, and Cluster 2 had much higher (and more variable) lignin, EWT, pigment and pigment ratio values. The second node separated Clusters 3 and 4 from the remaining clusters. These two clusters had low-to-very low internal leaf structure values along with generally high LMA, lignin, and EWT, and average pigments/pigment ratios. They differed from one another mainly in internal structure (Cluster 4 having much lower values), as well as in LMA (Cluster 4 having higher values). The third node separated Clusters 5, 6 and 7 from Clusters 8–12. These three clusters were mainly characterized by very low pigment values. Despite this, Cluster 5 was quite unique from Clusters 6 and 7. Its pigment ratio values were high, and its leaf surface, internal structure, lignin, LMA and EWT were also very high. Clusters 6 and 7 differed mainly in that Cluster 7 had lower LMA, pigment and pigment ratio values, as well as broader distributions of pigment ratios and carotenoids. The fourth node separated Clusters 8 and 9 from 10 to 12. These two clusters both had higher LMA values than did 10–12. Cluster 9 had lower lignin values, and Cluster 8 had higher leaf surface reflectance. The fifth node separated Cluster 10 from Clusters 11 and 12. Cluster 10 was characterized by very low internal structure values, average surface reflectance, low lignin and EWT, very low LMA and average pigment and pigment ratio values (but low carotenoids). Its very low LMA and internal structure differentiated it from Clusters 11 and 12. Clusters 11 and 12 both had low surface reflectance values and low-to-very low lignin, EWT and LMA. Lignin, EWT and LMA values were lower for Cluster 11. Cluster 11 also had higher internal structure values and overall high pigment values, as compared to Cluster 12′s average pigment/pigment ratio values.

4. Discussion

4.1. Drivers of leaf spectral variability

The three major axes of spectral variability found within our data are supported by well-established relationships between leaf properties and optical reflectance (Gates et al., 1965; Ollinger, 2011). The majority of the spectral variation occurred in the NIR wavelength regions, captured by PC 1; PC 2 represented the SWIR region, and the lowest amount of variation (PC 3) was in the VIS region. Similar patterns in variance have been found in other studies of diverse leaf spectra (e.g., Asner et al., 2014a). PC 1 carried information about the leaf surface and structure of the cuticle, as well as leaf internal structure. Leaf surface features, such as wax, hairs, and cuticle roughness and thickness, control light reflected specularly from and scattered at the leaf surface (Grant et al., 1993). These surface variations can be related to species or to physiological status, but contain no information from the leaf interior (Vanderbilt et al., 1987), and are a large source of spectral variation (Sims & Gamon, 2002). In our work, this component was most strongly correlated with campaign (i.e., combination of month and year), but also with species and site. Others have found species to have unique signatures based on their surface and structural characteristics (e.g., da Luz, 2006). PC 2 in our study was most strongly correlated with the SWIR region, which contains information related to water, dry matter and internal leaf structure (Cheng et al., 2011, 2014; Kokaly & Clark, 1999). As such, this component was strongly correlated with species and cPFT, and to a lesser extent, with leaf form and leaf duration. Research by Clark and Roberts (2012) found several SWIR features to be critical for differentiating tropical tree species at leaf-level, and Lehmann et al. (2015) found this portion of the spectrum was best able to identify Mediterranean shrub and tree species. PC 3, representing the VIS region, carried information related to pigments and was most strongly correlated with species. Relationships between pigment composition and species have been found in many studies (e.g., Asner & Martin, 2009; Carter & Knapp, 2001; Peñuelas et al., 1995). It is important to note that the samples in our study consisted mainly of green leaves and included only a small proportion of senescing leaves, likely accounting for weaker correlations between PC 3 and season.

4.2. Interpreting leaf spectral clusters

The combination of information contained within the three PCs resulted in twelve spectrally unique clusters in our data. These clusters did not correspond directly to species or cPFTs; however, they did exhibit clear functional differences based on spectral metrics. Observations from a single species or cPFT were often found in multiple clusters, demonstrating functional variance within traditionally defined classes was captured by the spectra, but represented in various clusters. Intra-class functional variance is driven by several factors, including taxonomic identity, the evolution of functional traits, and trait plasticity, or adjustments with respect to season (i.e., phenology) and environmental conditions (e.g., elevation, soil type, weather, climate), (Asner et al., 2014b; Ustin et al., 1993). Our clusters separated observations from species with related taxonomy (e.g., same genus), but that have evolved different traits in response to their respective ecosystem adaptations (i.e., differences in climate, community composition, etc.). For example, despite both being deciduous broadleaf oak trees, *Q. douglasii* and *Q. kelloggii* shared no clusters in common. Instead, *Q. douglasii* observations were clustered with the evergreen *Q. wislizeni* in Clusters 8, 9 and 12. Both these oak species grow in the oak savanna ecosystem, and have developed convergent leaf structural traits (i.e., smaller, somewhat sclerophyllous), perhaps in response to long summer droughts in this environment (Nixon et al., 2002).

Additionally, our clustering was able to detect seasonal differences across ecosystems within the same cPFT. Two evergreen broadleaf oaks, *Q. wislizeni* and *Q. chrysolepsis*, have similar leaf morphology, but *Q. chrysolepsis* grows at higher elevation and gets more water during the growing season (Nixon et al., 2002). Spring and fall observations from both species were present in Cluster 9, whereas summer samples from *Q. wislizeni* were in Cluster 12 while those from *Q. chrysolepsis* were in Cluster 1. This indicates that while these two oaks may function similarly during the spring and fall, they seem to differ functionally during the summer. *Q. wislizeni* is subject to drier summer conditions, and indeed, Cluster 12 had much lower EWT than did Cluster 1. Our clusters also captured seasonal changes in function within a single species. For example, *Z. mays*, was present in 4 clusters (6, 7, 10 and 12). Cluster 10 contained mainly May samples; Cluster 12 contained mainly August samples, and Clusters 6 and 7 contained leaves in two stages of senescence (early and late, respectively). Similarly, fall observations from the deciduous *Q. kelloggii* were grouped into Clusters 6 (from the more northern Blodgett Forest site) and 7 (from the more southern Soaproot Saddle site). For this species, site environmental conditions also appeared to play a role. Cluster 6 contained fall observations from the more northern site, and Cluster 7 contained those from

the more southern site. When within-class functional variance was low, nearly all observations from a species were found within a single cluster. The lower functional variance in these observations could be due to their collection at just one site or due to little difference across seasons (e.g., *S. acutus* and *A. viscida*), or because of very stable environmental conditions (e.g., *V. vinifera*). Our results demonstrate that species and cPFTs, while representing some functional trait differences in general, do not adequately account for differences among individuals or in response to environmental constraints or community interactions.

However, our clusters support the notion that unique functional groups do exist in these data because observations within each cluster shared similar functional metric profiles. Thus, the clusters represent optically-defined leaf functional types. For example, observations from the evergreen broadleaf shrub, *A. visicida*, which has whitish, thick, tough leaves, formed a single cluster (5). This cluster was distinguished by a combination of high surface reflectance, very low pigment values, and high-to-very high EWT, lignin and LMA, traits that are advantageous in a chaparral species. Two other clusters represented thin, broad annual or deciduous, leaves in different stages of senescence (6 and 7). These clusters were characterized by low LMA, lignin and water, and especially low pigment values. The cluster representing late stages of senescence (7) had extremely low pigment values as well as much lower pigment ratio values, indicating these leaves have even lower light use efficiency (Gamon et al., 1997). This ability to account for temporal change (i.e., phenology) demonstrates a key advantage of optically defined functional types over cPFTs. Another cluster (2) combined observations from two very different cPFTs: those from wetland emergent species with those from the evergreen needleleaf tree *A. concolor*. This cluster had very high values for nearly all functional metrics including internal structure values, lignin, EWT, LMA, and pigments. While these species are distinctly different, their leaves share surprisingly similar traits in that they are thick, have rounded surfaces and contain high amounts of water and dry matter. Finally, functional differences among clusters varied across all metrics, signifying the covariance of sets of functional traits are critical in defining cluster function.

4.3. Additional considerations in defining optical functional types with clustering

The clusters found in this study varied in their compactness and separation. This is mainly because both leaf spectra and leaf functional traits are actually continuous variables, existing across gradients. Assigning observations to hard classes will always be a challenge. However, if we accept that spectra represent functional trait space, then some regions of this feature space will be more densely populated due to resource limitations and tradeoffs that drive functional strategies. One potential method for addressing this problem is to use fuzzy clustering to allow multiple class memberships per observations (Almeida & Sousa, 2006). Still, there is value in defining clusters (i.e., functional types). For example, to better model plant functional behavior, we need distributions of trait values, not only mean values. Spectral clusters represent variation in trait values *within* and *across* cPFTs and species, better compartmentalizing actual functionality (i.e., during a given time of year, in a particular location), and yielding more precise trait distributions.

The cPFTs and species measured for this study are clearly a small subset of the diversity of plant species. Thus, the trait variation encompassed in these data will differ from that of other spectral data sets, depending on the previously discussed sources of functional variance. The clusters defined in our analysis do not represent final optical leaf functional types, but rather the approach presented here could be used to identify relevant optical functional types in other studies (likely resulting in different clusters). Ideally, as larger spectral databases are compiled, this approach could be used to define an over-arching set of optical functional types along with their respective trait distributions. Several such database initiatives are underway (e.g., EcoSIS, Specchio).

It is important to note that while we examined cluster differences for a specific set of functional metrics, the leaf spectrum is sensitive to many other leaf functional traits not considered in this analysis, including nitrogen, phenols, and perhaps others (Asner et al., 2015; Kokaly & Skidmore, 2015). Some of these will be important for quantifying additional functional differences among the clusters. However, because these clusters were created using all spectral information, we assume these traits were accounted for more completely than if the clusters were defined by spectral metrics (Asner et al., 2014b). Lastly, spectral reflectance is not sensitive to all functional traits, (e.g., rooting depth), representing an unquantified limitation on its use to define PFTs reliant on such traits. Some of these issues will be addressed in forthcoming research, which will examine how well spectral clusters correspond with measured leaf functional traits, and will also compare them with clusters defined using functional traits. We will also use these data to more precisely evaluate the trait distributions of these optical leaf functional types. Such distributions may be used to constrain both leaf- and canopy-level radiative transfer inversions, potentially leading to more accurate estimations of leaf biochemical properties from remote sensing data.

Lastly, identification of leaf functional traits is only one piece of the PFT puzzle. Other functional trait differences at the canopy scale, such as life form or leaf area index, are clearly important for measuring and mapping plant functions. A similar clustering approach could be taken at the level of the individual plant or canopy to define optical plant or ecosystem functional types, by measuring spectra at these scales and including additional functional metrics in the analysis. Scaling this approach to the canopy will enable us to evaluate the impact of leaf traits, canopy structure, and species and cPFT composition on the spectral properties of entire landscapes.

5. Conclusions

We determined that leaf spectral variance, as captured by the first few PCs, was more closely related to differences in leaf functional traits than to categorical attributes such as cPFT or species. Clustering these observations with the Dynamic Hybrid Clustering approach resulted in valid and stable, ecologically meaningful groups that reflected key functional differences. The clusters did not directly correspond with cPFTs or with species, but instead represented common patterns of trait co-variance related to known plant functional strategies.

It is critical to note that simply knowing the cPFTs, or even the species, present in an environment is not enough to predict their functioning. While species/cPFTs may represent some general trait patterns, they do not adequately account for differences due to individual variability within the species or type, or in response to environmental constraints or community interactions. We must continue to define PFTs that allow us to more directly characterize functional differences. Furthermore, it is not enough define PFTs that rely on single traits or even just a few traits. Instead, we should seek to characterize functional types by the covariance traits and their distributions. The large number of observations required to address these goals will require greater use of remote sensing data, and spectral data in particular. Our research has demonstrated the promise in using such data to define optical functional types at the leaf-level through a novel clustering approach using spectra collected from a broad range of species and cPFTs, and incorporating samples from a diverse range of ecosystem conditions, multiple sites and across seasons. This approach uses no a priori information and groups observations into functionally distinct clusters. Consequently, membership in spectral clusters and functional profiles will provide valuable insight into the drivers of functional trait variance within and among species.

Acknowledgements

We would like to thank all those who contributed to this work, especially those who helped us collect and process field data especially Mui Lay, Kristen Shapiro, Roger Stephens, Chris Preston, Shruti Khanna, Larry Costick, George Scheer, Darren Drewry, Jasmine Shen, and Paul Brower. This research was funded by NASA Grant #NNX12AP87G, Identification of Plant Functional Types by Characterization of Canopy Chemistry, as part of the HyspIRI Preparatory Science Campaign. We also want to thank the three anonymous reviewers who gave their time and provided excellent feedback in improving this manuscript.

References

Ackerly, D.D., Cornwell, W.K., 2007. A trait-based approach to community assembly: partitioning of species trait values into within- and among-community components. Ecol. Lett. 10 (2), 135–145. http://dx.doi.org/10.1111/j.1461-0248.2006.01006.x.

Adler, P.B., Salguero-Gómez, R., Compagnoni, A., Hsu, J.S., Ray-Mukherjee, J., Mbeau-Ache, C., Franco, M., 2014. Functional traits explain variation in plant life history strategies. Proc. Natl. Acad. Sci. U. S. A. 111 (2), 740–745. http://dx.doi.org/10.1073/pnas.1315179111.

Albert, C.H., De Bello, F., Boulangeat, I., Pellet, G., Lavorel, S., Thuiller, W., 2012. On the importance of intraspecific variability for the quantification of functional diversity. Oikos 121 (1), 116–126 Retrieved from http://dx.doi.org/10.1111/j.1600-0706.2011.19672.x.

Albert, C.H., Thuiller, W., Yoccoz, N.G., Douzet, R., Aubert, S., Lavorel, S., 2010a. A multi-trait approach reveals the structure and the relative importance of intra- vs. interspecific variability in plant traits. Funct. Ecol. 24 (6), 1192–1201. http://dx.doi.org/10.1111/j.1365-2435.2010.01727.x.

Albert, C.H., Thuiller, W., Yoccoz, N.G., Soudant, A., Boucher, F., Saccone, P., Lavorel, S., 2010b. Intraspecific functional variability: extent, structure and sources of variation. J. Ecol. 98 (3), 604–613. http://dx.doi.org/10.1111/j.1365-2745.2010.01651.x.

Almeida, T.I.R., De Souza Filho, C.R., 2004. Principal component analysis applied to feature-oriented band ratios of hyperspectral data: a tool for vegetation studies. Int. J. Remote Sens. 25, 5005–5023.

Almeida, R.J., Sousa, J.M.C., 2006. Comparison of fuzzy clustering algorithms for classification. International Symposium on Evolving Fuzzy Systems, pp. 112–117 (IEEE).

Alton, P.B., 2011. How useful are plant functional types in global simulations of the carbon, water, and energy cycles? J. Geophys. Res. Biogeosci. 116, 1–13. http://dx.doi.org/10.1029/2010JG001430.

Asner, G.P., 1998. Biophysical and biochemical sources of variability in canopy reflectance. Remote Sens. Environ. 64 (3), 234–253 (Retrieved from <Go to ISI>://000074765100010).

Asner, G.P., 2013. Biological diversity mapping comes of age. Remote Sens. 5 (1), 374–376. http://dx.doi.org/10.3390/rs5010374.

Asner, G.P., Martin, R.E., 2009. Airborne spectranomics: mapping canopy chemical and taxonomic diversity in tropical forests. Front. Ecol. Environ. 7 (5), 269–276. http://dx.doi.org/10.1890/070152.

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. http://dx.doi.org/10.1016/j.rse.2014.11.011.

Asner, G.P., Martin, R.E., Carranza-Jiménez, L., Sinca, F., Tupayachi, R., Anderson, C.B., Martinez, P., 2014a. Functional and biological diversity of foliar spectra in tree canopies throughout the Andes to Amazon region. New Phytol. http://dx.doi.org/10.1111/nph.12895.

Asner, G. P., Martin, R. E., Ford, A. J., Metcalfe, D. J., & Liddell, M. J. (2009). Leaf chemical and spectral diversity in Australian tropical forests. Ecol. Appl., 19(1), 236–253. Retrieved from (http://www.ncbi.nlm.nih.gov/pubmed/19323186)

Asner, G.P., Martin, R.E., Tupayachi, R., Anderson, C.B., Sinca, F., Carranza-Jiménez, L., Martinez, P., 2014b. Amazonian functional diversity from forest canopy chemical assembly. Proc. Natl. Acad. Sci. U. S. A. 111 (15), 5604–5609. http://dx.doi.org/10.1073/pnas.1401181111.

Atzberger, C., Guérif, M., Baret, F., Werner, W., 2010. Comparative analysis of three chemometric techniques for the spectroradiometric assessment of canopy chlorophyll content in winter wheat. Comput. Electron. Agric. 73, 165–173.

Berkhin, P., 2006. A survey of clustering data mining techniques. In: Kogan, J., Nicholas, C., Teboulle, M. (Eds.), Grouping Multidimensional Data SE - 2. Springer, Berlin Heidelberg, pp. 25–71 http://dx.doi.org/10.1007/3-540-28349-8_2.

Bernhardt-Römermann, M., Römermann, C., Nuske, R., Parth, A., Klotz, S., Schmidt, W., & Stadler, J. (2008). On the identification of the most suitable traits for plant functional trait analyses. Oikos, 117(10), 1533–1541. (Retrieved from) http://dx.doi.org/10.1111/j.0030-1299.2008.16776.x

Bhattacharjee, A., Richards, W.G., Staunton, J., Li, C., Monti, S., Vasa, P., ... Meyerson, M., 2001. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. Proc. Natl. Acad. Sci. 98 (24), 13790–13795. http://dx.doi.org/10.1073/pnas.191502998.

Bonan, G.B., Levis, S., Kergoat, L., Oleson, K.W., 2002. Landscapes as patches of plant functional types: an integrating concept for climate and ecosystem models. Glob. Biogeochem. Cycles 16 (2) (Retrieved from <Go to ISI>://000178887900007).

Cadotte, M.W., Cavender-Bares, J., Tilman, D., Oakley, T.H., 2009. Using phylogenetic, functional and trait diversity to understand patterns of plant community productivity. PLoS One 4 (5), e5695. http://dx.doi.org/10.1371/journal.pone.0005695.

Caldararu, S., Purves, D.W., Smith, M.J., 2015. The effect of using the plant functional type paradigm on a data-constrained global phenology model. Biogeosci. Discuss. 12 (20), 16847–16884. http://dx.doi.org/10.5194/bgd-12-16847-2015.

Carter, G., Knapp, A., 2001. Leaf optical properties in higher plants: linking spectral characteristics to stress and chlorophyll concentration. Am. J. Bot. 88 (4), 677–684.

Cheng, T., Rivard, B., Sánchez-Azofeifa, A., 2011. Spectroscopic determination of leaf water content using continuous wavelet analysis. Remote Sens. Environ. 115 (2), 659–670. http://dx.doi.org/10.1016/j.rse.2010.11.001.

Cheng, T., Rivard, B., Sánchez-Azofeifa, A.G., Féret, J.-B., Jacquemoud, S., Ustin, S.L., 2014. Deriving leaf mass per area (LMA) from foliar reflectance across a variety of plant species using continuous wavelet analysis. ISPRS J. Photogramm. Remote Sens. 87, 28–38. http://dx.doi.org/10.1016/j.isprsjprs.2013.10.009.

Chipman, J., Tibshirani, R., 2005. Hybrid hierarchical clustering with applications to microarray data. Biostatistics 7 (2), 286–301.

Clark, M.L., Roberts, D.A., 2012. Species-level differences in hyperspectral metrics among tropical rainforest trees as determined by a tree-based classifier. Remote Sens. 4 (6), 1820–1855. http://dx.doi.org/10.3390/rs4061820.

da Luz, B.R., 2006. Attenuated total reflectance spectroscopy of plant leaves: a tool for ecological and botanical studies. New Phytol. 172 (2), 305–318. http://dx.doi.org/10.1111/j.1469-8137.2006.01823.x.

Duckworth, J.C., Kent, M., Ramsay, P.M., 2000. Plant functional types: an alternative to taxonomic plant community description in biogeography? Prog. Phys. Geogr. 24 (4), 515–542 (Retrieved from <Go to ISI>://000165678600003).

Féret, J.-B., Asner, G.P., 2014. Mapping tropical forest canopy diversity using high-fidelity imaging spectroscopy. Ecol. Appl. 24, 1289–1296. http://dx.doi.org/10.1890/13-1824.1.

Féret, J.-B., François, C., Gitelson, A.A., Asner, G.P., Barry, K.M., Panigada, C., Richardson, A.D., Jacquemoud, S., 2011. Optimizing spectral indices and chemometric analysis of leaf chemical properties using radiative transfer modeling. Remote Sens. Environ. 115, 2742–2750.

Féret, J.-B., François, C., Asner, G.P., Gitelson, A.A., Martin, R.E., Bidel, L.P.R., ... Jacquemoud, S., 2008. PROSPECT-4 and 5: advances in the leaf optical properties model separating photosynthetic pigments. Remote Sens. Environ. 112 (6), 3030–3043. http://dx.doi.org/10.1016/j.rse.2008.02.012.

Fowlkes, E.B., Mallows, C.L., 1983. A method for comparing two hierarchical clusterings. J. Am. Stat. Assoc. 78 (383), 553. http://dx.doi.org/10.2307/2288117.

Gamon, J.A., Serrano, L., Surfus, J.S., 1997. The photochemical reflectance index: an optical indicator of photosynthetic radiation use efficiency across species, functional types, and nutrient levels. Oecologia 112 (4), 492–501 (Retrieved from <Go to ISI>://A1997YK58000008).

Garcia, R., Hedley, J., Tin, H., Fearns, P., 2015. A method to analyze the potential of optical remote sensing for benthic habitat mapping. Remote Sens. 7 (10), 13157–13189. http://dx.doi.org/10.3390/rs71013157.

Gates, D.M., Keegan, H.J., Weidner, V.R., Schleter, J.C., 1965. Spectral properties of plants. Appl. Opt. 4 (1), 11. http://dx.doi.org/10.1364/AO.4.000011.

Göttlicher, D., Albert, J., Nauss, T., Bendix, J., 2011. Optical properties of selected plants from a tropical mountain ecosystem – traits for plant functional types to parametrize a land surface model. Ecol. Model. 222 (3), 493–502. http://dx.doi.org/10.1016/j.ecolmodel.2010.09.021.

Grant, L., Daughtry, C.S.T., Vanderbilt, V.C., 1993. Polarized and specular reflectance variation with leaf surface-features. Physiol. Plant. 88 (1), 1–9 (Retrieved from <Go to ISI>://A1993LH31400001).

Groenendijk, M., Dolman, A.J., van der Molen, M.K., Leuning, R., Arneth, A., Delpierre, N., Wohlfahrt, G., 2011. Assessing parameter variability in a photosynthesis model within and between plant functional types using global Fluxnet eddy covariance data. Agric. For. Meteorol. 151 (1), 22–38. http://dx.doi.org/10.1016/j.agrformet.2010.08.013.

Hardle, W., Simar, L., 2003. Applied Multivariate Statistical Analysis.

Hesketh, M., Sánchez-Azofeifa, G.A., 2012. The effect of seasonal spectral variation on species classification in the Panamanian tropical forest. Remote Sens. Environ. 118, 73–82. http://dx.doi.org/10.1016/j.rse.2011.11.005.

Homolová, L., Malenovský, 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. http://dx.doi.org/10.1016/j.ecocom.2013.06.003.

Hubert, L., Arabie, P., 1985. Comparing partitions. J. Classif. 2 (1), 193–218. http://dx.doi.org/10.1007/BF01908075.

Kattge, J., Diaz, S., Lavoral, S., Prentice, I.C., Leadley, P.W., BÖNISCH, G., ... WIRTH, C., 2011. TRY - a global database of plant traits. Glob. Chang. Biol. 17 (9). http://dx.doi.org/10.1111/j.1365-2486.2011.02451.x.

Kaufman, L., Rousseeuw, P.J., 2005. Finding Groups in Data: an Introduction to Cluster Anlaysis. second ed. John Wiley & Sons, Inc.

Kokaly, R.F., Clark, R.N., 1999. Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression. Remote Sens. Environ. 67 (3), 267–287 (Retrieved from <Go to ISI>://000078472700002).

Kokaly, R.F., Skidmore, A.K., 2015. Plant phenolics and absorption features in vegetation reflectance spectra near 1.66 μm. Int. J. Appl. Earth Obs. Geoinf. 43, 1–29. http://dx.doi.org/10.1016/j.jag.2015.01.010.

Langfelder, P., Zhang, B., Horvath, S., 2008. Defining clusters from a hierarchical cluster tree: the dynamic tree cut package for R. Bioinformatics 24 (5), 719–720. http://dx.doi.org/10.1093/bioinformatics/btm563.

Langfelder, P., Zhang, B., Horvath, S., 2014. dynamicTreeCut: Methods for Detection of Clusters in Hierarchical Clustering Dendrograms. R Package Version 1.62. https://CRAN.R-project.org/package=dynamicTreeCut.

Laughlin, D.C., Laughlin, D.E., 2013. Advances in modeling trait-based plant community assembly. Trends Plant Sci. 18 (10), 584–593. http://dx.doi.org/10.1016/j.tplants.2013.04.012.

Lavorel, S., Garnier, E., 2002. Predicting changes in community composition and ecosystem functioning from plant traits: revisiting the Holy Grail. Funct. Ecol. 16 (5), 545–556 (Retrieved from <Go to ISI>://000178119300001).

Lehmann, J., Große-Stoltenberg, A., Römer, M., Oldeland, J., 2015. Field spectroscopy in the VNIR-SWIR region to discriminate between Mediterranean native plants and exotic-invasive shrubs based on leaf tannin content. Remote Sens. 7, 1225–1241. http://dx.doi.org/10.3390/rs70201225.

LI-COR, 1983. LI-1800-12 Integrating Sphere Instruction Manual. LI-COR, Inc., Lincoln, NE (Pub. No. 8305-0034).

Mesarch, M.A., Walter-Shea, E.A., Asner, G.P., Middleton, E.M., Chan, S.S., 1999. A revised measurement methodology for conifer needles spectral optical properties: evaluating the influence of gaps between elements. Remote Sens. Environ. 68 (2), 177–192 (Retrieved from <Go to ISI>://000080083800007).

Messier, J., McGill, B.J., Lechowicz, M.J., 2010. How do traits vary across ecological scales? A case for trait-based ecology. Ecol. Lett. 13 (7), 838–848. http://dx.doi.org/10.1111/j.1461-0248.2010.01476.x.

Milligan, G.W., Cooper, M.C., 1986. A study of the comparability of external criteria for hierarchical cluster analysis. Multivar. Behav. Res. 21 (4), 441–458.

Monti, S., Tamayo, P., Mesirov, J., Golub, T., Sebastiani, P., Kohane, I.S., Ramoni, M.F., 2003. Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data. Mach. Learn. 52 (i), 91–118. http://dx.doi.org/10.1023/A:1023949509487.

Moorcroft, P.R., 2006. How close are we to a predictive science of the biosphere? Trends Ecol. Evol. 21 (7), 400–407. http://dx.doi.org/10.1016/j.tree.2006.04.009.

Nixon, K., Standiford, R., McCreary, D., Purcell, K., 2002. The oak (Quercus) biodiversity of California and adjacent regions. In: Standiford, R.B., McCreary, D., Purcell, K.L. (Eds.), Proceedings of the Fifth Symposium on Oak Woodlands: Oaks in California's Challenging Landscape , pp. 3–20 (Albany, CA. Retrieved from) http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:The+Oak+(+Quercus+)+Biodiversity+of+California+and+Adjacent+Regions+1#0.

Ollinger, S.V., 2011. Sources of variability in canopy reflectance and the convergent properties of plants. New Phytol. 189 (2), 375–394. http://dx.doi.org/10.1111/j.1469-8137.2010.03536.x.

Pappas, C., Fatichi, S., Burlando, P., 2014. Terrestrial water and carbon fluxes across climatic gradients: does plant diversity matter? New Phytol. 16 (i), 3663.

Pavlick, R., Drewry, D.T., Bohn, K., Reu, B., Kleidon, A., 2012. The Jena Diversity-Dynamic Global Vegetation Model (JeDi-DGVM): a diverse approach to representing terrestrial biogeography and biogeochemistry based on plant functional trade-offs. Biogeosci. Discuss. 9 (4), 4627–4726. http://dx.doi.org/10.5194/bgd-9-4627-2012.

Peñuelas, J., Baret, F., Filella, I., Penuelas, J., Baret, F., Filella, I., 1995. Semiempirical indexes to assess carotenoids chlorophyll-a ratio from leaf spectral reflectance. Photosynthetica 31 (2), 221–230 (Retrieved from <Go to ISI>://WOS:A1995RK28000008).

Pérez-Harguindeguy, N., Díaz, S., Garnier, E., Lavorel, S., Poorter, H., Jaureguiberry, P., ... Cornelissen, J.H.C., 2013. New handbook for standardised measurement of plant functional traits worldwide. Aust. J. Bot. 61, 167–234.

Core Team, R., 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing (Vienna, Austria. URL https://www.R-project.org/.).

Rautiainen, M., Mõttus, M., Yáñez-Rausell, L., Homolová, L., Malenovský, Z., Schaepman, M.E., 2012. A note on upscaling coniferous needle spectra to shoot spectral albedo. Remote Sens. Environ. 117, 469–474.

Reich, P.B., Wright, I.J., Cavender-Bares, J., Craine, J.M., Oleksyn, J., Westoby, M., Walters, M.B., 2003. The evolution of plant functional variation: traits, spectra, and strategies. Int. J. Plant Sci. 164 (S3), S143–S164.

Reichstein, M., Bahn, M., Mahecha, M. D., Kattge, J., & Baldocchi, D. D. (2014). Linking plant and ecosystem functional biogeography. Proc. Natl. Acad. Sci., (iDiv). doi:http://dx.doi.org/10.1073/pnas.1216065111

Rousseeuw, P.J., 1986. A visual display for hierarchical classification. In: Diday, E., Escoufier, Y., Lebart, L., Pages, J., Schektman, Y., Tomassone, R. (Eds.), Data Analysis and Informatics 4. North-Holland, Amsterdam, pp. 743–748.

RStudio Team, 2015. RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URLhttp://www.rstudio.com.

Ruffin, C., King, R., 1999. The analysis of hyperspectral data using Savitzsky-Golay filtering – theoretical basis (part 1). Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Hamburg, Germany (EEE Catalog Number 99CH36293C).

Sakschewski, B., von Bloh, W., Boit, A., Rammig, A., Kattge, J., Poorter, L., ... Thonicke, K., 2015. Leaf and stem economics spectra drive diversity of functional plant traits in a dynamic global vegetation model. Glob. Chang. Biol. 21, 2711–2725. http://dx.doi.org/10.1111/gcb.12870.

Sandel, B., Gutiérrez, A.G., Reich, P.B., Schrodt, F., Dickie, J., Kattge, J., 2015. Estimating the missing species bias in plant trait measurements. J. Veg. Sci. 26 (5), 828–838. http://dx.doi.org/10.1111/jvs.12292.

Savitzky, A., Golay, M.J.E., 1964. Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem. 36 (8), 1627–1639. http://dx.doi.org/10.1021/ac60214a047.

Schaepman, M.E., Ustin, S.L., Plaza, A.J., Painter, T.H., Verrelst, J., Liang, S., 2009. Earth system science related imaging spectroscopy—an assessment. Remote Sens. Environ. 113 (Supplement 1), S123–S137. http://dx.doi.org/10.1016/j.rse.2009.03.001.

Schimel, D.S., Asner, G.P., Moorcroft, P., 2013. Observing changing ecological diversity in the Anthropocene. Front. Ecol. Environ. 11 (3), 129–137. http://dx.doi.org/10.1890/120111.

Schimel, D.S., Pavlick, R., Fisher, J.B., Asner, G.P., Saatchi, S., Townsend, P., ... Cox, P., 2015. Observing terrestrial ecosystems and the carbon cycle from space. Glob. Chang. Biol. 21, 1762–1776. http://dx.doi.org/10.1111/gcb.12822.

Semenova, G.V., Van Der Maarel, E., 2000. Plant functional types: a strategic perspective. J. Veg. Sci. 11 (6), 917–922. http://dx.doi.org/10.2307/3236562.

Shi, K., Li, Y., Li, L., Lu, H., Song, K., Liu, Z., ... Li, Z., 2013. Remote chlorophyll-a estimates for inland waters based on a cluster-based classification. Sci. Total Environ. 444, 1–15. http://dx.doi.org/10.1016/j.scitotenv.2012.11.058.

Shipley, B., Lechowicz, M. J., Wright, I. J., & Reich, P. B. (2006). Fundamental trade-offs generating the worldwide leaf economics spectrum. Ecology, 87(3), 535–41. Retrieved from (http://www.ncbi.nlm.nih.gov/pubmed/16602282)

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 (2–3), 337–354 (Pii S0034-4257(02)00010-X).

Sims, D.A., Gamon, J.A., 2003. Estimation of vegetation water content and photosynthetic tissue area from spectral reflectance: a comparison of indices based on liquid water and chlorophyll absorption features. Remote Sens. Environ. 84 (4), 526–537.

Smith, T.M., Shugart, H.H., Woodward, F.I., 1997. Plant Functional Types: Their Relevance to Ecosystem Properties and Global Change. International Geosphere-Biosphere Programme Book Series. Cambridge University Press, New York.

Sokal, R.R., Rohlf, F.J., 1962. The comparison of dendrograms by objective methods. Taxon 11, 33–40.

Steinbach, M., Ertoz, L., Kumar, V., 2004. The challenges of clustering high dimensional data. New Directions in Statistical Physics, pp. 273–309 http://dx.doi.org/10.1007/978-3-662-08968-2_16.

Stimson, H.C., Breshears, D.D., Ustin, S.L., Kefauver, S.C., 2005. Spectral sensing of foliar water conditions in two co-occurring conifer species: *Pinus edulis* and *Juniperus monosperma*. Remote Sens. Environ. 96, 108–118.

Sun, W.X., Liang, S.L., Xu, G., Fang, H.L., Dickinson, R., 2008. Mapping plant functional types from MODIS data using multisource evidential reasoning. Remote Sens. Environ. 112 (3), 1010–1024 (Retrieved from <Go to ISI>://000254443700031).

Torrecilla, E., Stramski, D., Reynolds, R.a., Millán-Núñez, E., Piera, J., 2011. Cluster analysis of hyperspectral optical data for discriminating phytoplankton pigment assemblages in the open ocean. Remote Sens. Environ. 115 (10), 2578–2593. http://dx.doi.org/10.1016/j.rse.2011.05.014.

Ustin, S.L., Gamon, J.A., 2010. Remote sensing of plant functional types. New Phytol. 186 (4), 795–816. http://dx.doi.org/10.1111/j.1469-8137.2010.03284.x.

Ustin, S.L., Gitelson, A.A., Jacquemoud, S., Schaepman, M., Asner, G.P., Gamon, J.A., Zarco-Tejada, P., 2009. Retrieval of foliar information about plant pigment systems from high resolution spectroscopy. Remote Sens. Environ. 113, S67–S77. http://dx.doi.org/10.1016/j.rse.2008.10.019.

Ustin, S.L., Roberts, D.A., Gamon, J.A., Asner, G.P., Green, R.O., 2004. Using imaging spectroscopy to study ecosystem processes and properties. Bioscience 54 (6), 523. http://dx.doi.org/10.1641/0006-3568(2004)054[0523:UISTSE]2.0.CO;2.

Ustin, S.L., Sanderson, E.W., Grossman, Y., Hart, Q., Haxo, R., 1993. Relationships between pigment composition variation and reflectance for plant species from a coastal savannah in California. In: Green, R.O. (Ed.), Fourth Annual JPL Airborne Science Workshop. NASA Jet Propulsion Laboratory, Washington, D.C., pp. 181–184.

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 (38), 13733–13738. http://dx.doi.org/10.1073/pnas.1304551110.

Van Bodegom, P.M., Douma, J.C., Witte, J.P.M., Ordoñez, J.C., Bartholomeus, R.P., Aerts, R., 2012. Going beyond limitations of plant functional types when predicting global ecosystem-atmosphere fluxes: exploring the merits of traits-based approaches. Glob. Ecol. Biogeogr. 21 (6), 625–636. http://dx.doi.org/10.1111/j.1466-8238.2011.00717.x.

van der Laan, M.J., Pollard, K.S., 2003. A new algorithm for hybrid hierarchical clustering with visualization and the bootstrap. Journal of Statistical Planning and Inference 117, 275–303. http://dx.doi.org/10.1016/S0378-3758(02)00388-9.

Vanderbilt, V.C., Grant, L., Daughtry, C.S.T., 1987. Polarization of light scattered by clover. Proc. IEEE 73 (6), 1012–1024. http://dx.doi.org/10.1016/0034-4257(87)90011-3.

Violle, C., Jiang, L., 2009. Towards a trait-based quantification of species niche. J. Plant Ecol. 2 (2), 87–93. http://dx.doi.org/10.1093/jpe/rtp007.

Violle, C., Navas, M.-L.L., Vile, D., Kazakou, E., Fortunel, C., Hummel, I., Garnier, E., 2007. Let the concept of trait be functional! Oikos 116 (5), 882–892. http://dx.doi.org/10.1111/j.2007.0030-1299.15559.x.

Ward, J.H., 1963. Hierarchical grouping to optimize an objective function. J. Am. Stat. Assoc. 58 (301), 236–244.

Westoby, M., Wright, I.J., 2006. Land-plant ecology on the basis of functional traits. Trends Ecol. Evol. 21 (5), 261–268. http://dx.doi.org/10.1016/j.tree.2006.02.004.

Wilkerson, M.D., Hayes, D.N., 2010. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics 26 (12), 1572–1573. http://dx.doi.org/10.1093/bioinformatics/btq170.

Wilson, J.B., 1999. Guilds, functional types and ecological groups. Oikos 86 (3), 507–522 (Retrieved from <Go to ISI>://000082248200011).

Wishart, D., 1998. Efficient hierarchical cluster analysis for data mining and knowledge discovery. Computer Science and Statistics, pp. 257–263.

Woodward, F.I., Cramer, W., 1996. Introduction. J. Veg. Sci. 7 (3), 306–308 (Retrieved from) http://www.jstor.org/stable/3236273.

Wright, I.J., Reich, P.B., Westoby, M., Ackerly, D.D., Baruch, Z., Bongers, F., ... Villar, R., 2004. The worldwide leaf economics spectrum. Nature 428 (6985), 821–827. http://dx.doi.org/10.1016/j.tree.2010.11.011.

Wright, J.P., Naeem, S., Hector, A., Lehman, C., Reich, P.B., Schmid, B., Tilman, D., 2006. Conventional functional classification schemes underestimate the relationship with ecosystem functioning. Ecol. Lett. 9 (2), 111–120. http://dx.doi.org/10.1111/j.1461-0248.2005.00850.x.

Wright, J.P., Sutton-Grier, A., 2012. Does the leaf economic spectrum hold within local species pools across varying environmental conditions? Funct. Ecol. 26 (6), 1390–1398. http://dx.doi.org/10.1111/1365-2435.12001.

Yáñez-rausell, L., Schaepman, M.E., Member, S., Clevers, J.G.P.W., 2014a. Minimizing Measurement Uncertainties of Coniferous Needle-Leaf Optical Properties, Part I : Methodological Review. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 7(2) pp. 399–405.

Yáñez-Rausell, L., Malenovský, Z., Clevers, J.G.P.W., Schaepman, M.E., 2014b. Minimizing Measurement Uncertainties of Coniferous Needle-Leaf Optical Properties. Part II : Experimental Setup and Error Analysis. IEEE International Geoscience and Remote Sensing Symposium 7(2) pp. 406–420.

Yang, Y., Zhu, Q., Peng, C., Wang, H., Chen, H., 2015. From plant functional types to plant functional traits: a new paradigm in modelling global vegetation dynamics. Prog. Phys. Geogr. 39 (4), 514–535. http://dx.doi.org/10.1177/0309133315582018.