## Title

Canopy structural attributes derived from AVIRIS imaging spectroscopy data in a mixed broadleaf/conifer forest



## Authors

Huesca, Margarita
García, Mariano
Roth, Keely L
et al.

## Publication Date

2016-09-01

## DOI

10.1016/j.rse.2016.04.020

Peer reviewed

Remote Sensing of Environment 182 (2016) 208–226

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

# Canopy structural attributes derived from AVIRIS imaging spectroscopy data in a mixed broadleaf/conifer forest

Margarita Huesca [a,*], Mariano García [b,c], Keely L. Roth [a], Angeles Casas [a], Susan L. Ustin [a]

[a] Land, Air and Water Resources Department, Center for Spatial Technologies And Remote Sensing (CSTARS), University of California Davis, USA
[b] University of Leicester, Centre for Landscape and Climate Research, Leicester LE1 7RH, United Kingdom
[c] Jet Propulsion Laboratory (JPL), California Institute of Technology, Pasadena, CA 91109, USA

ARTICLE INFO

*Article history:*
Received 23 July 2015
Received in revised form 15 April 2016
Accepted 28 April 2016
Available online 31 May 2016

*Keywords:*
Canopy structure
AVIRIS
LiDAR
Random forest
Structural types

ABSTRACT

There is a well-established need within the remote sensing community for improved estimation and understanding of canopy structure and its influence on the retrieval of leaf biochemical properties. The main goal of this research was to assess the potential of optical spectral information from NASA's Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) to discriminate different canopy structural types. In the first phase, we assessed the relationships between optical metrics and canopy structural parameters obtained from LiDAR in terms of different canopy structural attributes (biomass (i.e., area under Vegetation Vertical Profile, $VVP_{int}$), canopy height and vegetation complexity). Secondly, we identified and classified different "canopy structural types" by integrating several structural traits using Random Forests (RF). The study area is a heterogeneous forest in Sierra National Forest in California (USA). AVIRIS optical properties were analyzed by means of several sets of variables, including single narrow band reflectance and 1st derivative, sub-pixel cover fractions, narrow-band indices, spectral absorption features, optimized normalized difference indices and Principal Component Analysis (PCA) components. Our results demonstrate that optical data contain structural information that can be retrieved. The first principal component, used as a proxy for albedo, was the most strongly correlated optical metric with vegetation complexity, and it also correlated well with biomass ($VVP_{int}$) and height. In conifer forests, the shade fraction was especially correlated to vegetation complexity, while water-sensitive optical metrics had high correlations with biomass ($VVP_{int}$). Single spectral band analysis results showed that correlations differ in magnitude and in direction, across the spectrum and by vegetation type and structural variable. This research illustrates the potential of AVIRIS to analyze canopy structure and to distinguish several structural types in a heterogeneous forest. Furthermore, RF using optical metrics derived from AVIRIS proved to be a powerful technique to generate maps of structural attributes. The results emphasize the importance of using the whole optical spectrum, since all spectral regions contributed to canopy structure assessment.



## 1. Introduction

Using remote sensing data to characterize ecosystem traits and processes is important to understanding functionality and ecosystem services. At leaf level, we have a strong understanding of the relationship between spectral reflectance, biochemistry and structure (Gates, Keegan, Schleter, & Weidner, 1965; Gausman & Allen, 1973; Gausman, 1984; Asner, Martin, Ford, Metcalfe, & Liddell, 2009; Sánchez-Azofeifa, Castro, Wright, Gamon, et al., 2009; Féret, François, Gitelson, Asner, et al., 2011). However, these relationships become substantially more complex when scaling to the canopy. It is well known that canopy reflectance at the pixel level is a function of the vegetation component properties (leaf and woody components), canopy structure (leaf and stem area and 3-dimensional orientation), understory vegetation and soil, illumination and view geometry (Asner, 1998; Asner & Martin, 2009; Asner, Martin, Anderson, & Knapp, 2015). Considering only within pixel elements, the impact of canopy structure, understory vegetation and soil produces a more complex relationship, making accurate estimates of leaf properties challenging to impossible (Asner, Martin, Anderson and Knapp 2015; Ollinger, 2011).

Some studies have related leaf optical properties to canopy reflectance assuming some approaches to simplify the problem, for instance, Allen and Richardson (1968) modeled this relationship considering the canopy as horizontal layers of stacked leaves, a clearly unrealistic assumption. Other studies have retrieved foliar and canopy chemistry from imaging spectroscopy data (Peterson, Aber, Matson, Card, et al., 1988; Townsend, Foster, Chastain, & Currie, 2003; Ustin, Gitelson, Jacquemoud, Schaepman, et al., 2009; Doughty, Asner, & Martin, 2011) without a canopy model. Many approaches to estimate forest canopy parameters are based on empirical models (Coops, Smith, Martin, & Ollinger, 2003; Kokaly, Asner, Ollinger, Martin, et al. 2009; Clevers & Kooistra, 2012), but their models are constrained by site and sensor dependence. Another approach is to develop physical models, for example, canopy radiative transfer models (forward and inverse

* Corresponding author at: Center for Spatial Technologies and Remote Sensing (CSTARS), University of California Davis, USA.
*E-mail address:* mhuescamartinez@ucdavis.edu (M. Huesca).

http://dx.doi.org/10.1016/j.rse.2016.04.020
0034-4257/

model simulations) that relate leaf optical properties with biophysical and biochemical variables (Verhoef, 1984; Rosema, Verhoef, Noorbergen, & Borgesius, 1992; Jacquemoud, Bacour, Poilvé, & Frangi, 2000; Jacquemoud, Verhoef, Baret, Bacour, et al., 2009; Ceccato, Flasse, Tarantola, Jacquemoud, et al., 2001). Omari, White, Staenz, and King (2013) used PROFLAIR leaf-canopy reflectance model to simulate a broad-leaf forest canopy using Hyperion Imaging spectroscopy data to test retrieval of leaf area index (LAI) and canopy chlorophyll content. Cheng, Zarco-Tejada, Riaño, Rueda, et al., (2006) used three leaf and canopy radiative transfer models to estimate vegetation water content for canopies with three different architectures. Zarco-Tejada, Rueda, and Ustin (2003) estimated leaf water index with MODerate resolution Imaging Spectroradiometer (MODIS) data by inversion of a leaf-canopy model (PROSPECT + SAILH) in chaparral vegetation in California (USA). The main problem they found was the large number of unknown parameters to fix and the constraints they needed to include for accurate retrievals.

More recent research has sought to better quantify the influence of canopy structure on canopy reflectance, (Knyazikhin, Schull, Stenberg, Mõttus, et al., 2013) especially under differing vegetation densities and vegetation gap distributions. Variables such as leaf area index (LAI), leaf angle distribution and canopy gaps play a significant role in canopy reflectance (Asner, 1998; Ollinger, 2011). If the impact of solar and view geometry is added, scaling from leaf to canopy becomes still more challenging. Consequently, it is essential to account for the influence of canopy structure together with solar and view geometry, isolating leaf spectra to retrieve qualified foliar properties from canopy spectra (Asner, Martin, Anderson, and Knapp 2015). Nevertheless, before analyzing the influence of each component, it is necessary to estimate them as accurately as possible. Solar and view geometry are easily measured; however, canopy structure assessment involves greater difficulty, but is critical in the analysis of imaging spectroscopy data. For example, White, Gómez, Wulder, and Coops (2010) found that the spectral variability found in a coastal temperate forest is more related to structural complexity than to species diversity.

The 3-D information provided by Light Detection and Ranging (LiDAR) data enables a characterization of canopy structure unachievable by any other passive remote sensing technique. The ability of LiDAR data to accurately provide structural information has been widely proved over different scenarios. It has proved its capability to estimate height, fractional cover, LAI, clumping index, biomass or volume at different levels of detail, from single trees to stand level (Hilker, Wulder, & Coops, 2008; Hopkinson & Chasmer, 2009; Tang, Brolly, Zhao, Strahler, et al., 2014; Riaño, Valladares, Condes, & Chuvieco, 2004; García, Gajardo, Riaño, Zhao, et al., 2015; Bouvier, Durrieu, Fournier, & Renaud, 2015; Næsset, Gobakken, Bollandsås, Gregoire, et al., 2013; García, Riaño, Chuvieco, & Danson, 2010). This makes LiDAR as a valuable technology to provide reference data for other remote sensors, enabling the development of models using LiDAR as calibration/validation data. The GLAS (Geoscience Laser Altimeter System — *http://attic.gsfc.nasa.gov/glas/*) sensor, on-board the ICESat (Ice Cloud and land Elevation Satellite — *http://icesat.gsfc.nasa.gov/icesat/*) satellite, proved also to be useful to estimate forest structural attributes (Suna, Ransonb, Kimesb, Blairb, et al., 2008; Duncansona, Niemanna, & Wulderb, 2010; Popescu, Zhao, Neuenschwander, & Lin, 2011), although its accuracy was reduced over steep terrain because of the large footprint size (Hilbert & Schmullius, 2012), which caused the mixing of ground and vegetation reflections.

Acquiring large spatial coverages with airborne LiDAR is not feasible (i.e., limited access and expensive) and satellite borne sensors provide a sample of the terrain with footprints spaced 170 m along-track and 15 km across-track at the equator; therefore, integration with passive remote sensing may offer an alternative to overcome this limitation (Lefsky, Cohen, & Spies, 2001). Much previous research in this area used multispectral data, more specifically broadband spectral indices. Cohen and Spies (1992) found a good relationship between spectral variables derived from Landsat TM and several forest structural attributes, such as standard deviation of tree size, structural complexity, and stand age, in a conifer forest in the Cascade Mountains, Oregon, USA. Lu, Batistella, and Moran (2005) explored the influence of forest structure in heterogeneous Amazon forests to estimate aboveground biomass using spectral mixture analysis of Landsat Thematic Mapper data. They found TM4 and green vegetation fractions were the most suitable variables for estimating succession forest biomass. Hall, Shimabukuro, and Huemmrich (1995) found positive relationships between shade fraction and LAI, biomass density, diameter at breast height (DBH), and aboveground net primary productivity (NPP) in a black spruce stand in the Superior National Forest near Ely, Minnesota, USA. Hyde, Dubayah, Walker, Blair, et al. (2006) estimated canopy height and biomass in conifer dominated forests in the Sierra Nevada Mountains, California, USA using Principal Component Analysis (PCA) and Normalized Difference Vegetation Index (NDVI) derived from Landsat ETM+. They found ETM+ accurately predicted canopy height but poorly predicted biomass, probably due to sensor saturation problems at high biomass levels. Ahmed, Franklin, and Wulder (2013) estimated canopy cover and height in a temperate deciduous and coniferous forest integrating Landsat spectral indices, spectral mixture analysis, and disturbance information. Ahmed, Franklin, Wulder, and White (2015) found that forest disturbance derived from Landsat time series improved accuracy of canopy structure estimation information in a coastal temperate forest in Vancouver, British Columbia, Canada, as well as, aiding a pre-stratification between young and mature forests. Along the same lines, Pflugmacher, Cohen, and Kennedy (2012) used forest disturbance and recovery information from Landsat time series to estimate aboveground biomass, basal area, and height in mixed-conifer forests in the Blue Mountains in Oregon, USA.

Nevertheless, important physical and physiological processes of vegetation have not been measured using broadband spectral indices (Blackburn, 1998). To address these data gaps, imaging spectroscopy data has promise since surface reflectance is sampled at high spatial resolution over hundreds of contiguous narrow spectral bands in the optical range (Ollinger, 2011). Imaging spectroscopy data can be used to calculate a wider range of vegetation spectral indices than Landsat data as well as specific absorption features, which combined, allows identification of physiological traits that minimize soil background effects (Asner, 1998; Ustin, Roberts, Gamon, Asner, et al., 2004). For instance, Ogunjemiyo, Parker, and Roberts (2005) found a negative relationship between AVIRIS albedo and canopy rugosity in a conifer forest. Roberts, Ustin, Ogunjemiyo, Greenberg, et al. (2004) evaluated how structural attributes identified with AVIRIS data (albedo, spectral fractions, NDVI and Equivalent Water Thickness (EWT)) varied with stand age in conifer forest. Swatantran, Dubayah, Roberts, Hofton, et al., (2011) found strong correlations between AVIRIS water-sensitive indices and shade fraction with canopy height derived from Laser Vegetation Imaging Sensor (LVIS) in the montane forests of the Sierra Nevada, California, USA. Although, imaging spectroscopy data appears to be promising for canopy structure estimation, it has not yet been widely used for this purpose and its usefulness has not been completely exploited. The proposed Hyperspectral Infrared Imager (HyspIRI) mission (National Research Council (NRC), 2007) will provide frequent global high spectral resolution information that offers an opportunity for globally assessing canopy structure from optical data. Thus, it is necessary to develop and test effective and efficient tools and techniques in preparation for these data. In order to further develop and refine these methods, airborne LiDAR data provides most useful validation and reference data source (Hudak, Lefsky, Cohen, & Berterretche, 2002; Wulder & Seemann, 2003).

The main objective of this research was to assess the relationship among structural metrics from LiDAR and from optical spectral information from AVIRIS, to better understand the effects of canopy structure on canopy spectral reflectance. Improving this understanding is a necessary step in developing methods to accurately estimate leaf properties from image data.

Specifically, this study sought to 1) analyze the effect of stratification by forest type (Conifer versus Hardwood) on the relationships between optical metrics and vegetation structure; 2) identify the spectral regions and optical metrics most significantly correlated with canopy structure; and 3) identify and map what we call "canopy structural types", by integrating several structural traits and evaluating their distributions across a heterogeneous forest. The final goal of this work is to discriminate these canopy structure types rather than to quantitatively predict the structural variables themselves.

Although a body of research exists around the study of canopy structure with optical properties, most of the research has been focused on a single forest type (Cohen & Spies, 1992; Hall, Shimabukuro, and Huemmrich, 1995; Ogunjemiyo, Parker, and Roberts, 2005; Roberts, Ustin, Ogunjemiyo, Greenberg, et al., 2004; Franklin, Hall, Smith, & Gerylo, 2003) or on a single structural variable, or several independently analyzed variables (Ahmed, Franklin, Wulder, and White, 2015; Ogunjemiyo, Parker, and Roberts, 2005; Swatantran, Dubayah, Roberts, Hofton, et al., 2011). In this paper, we focus on a heterogeneous landscape and the combination of structural traits that distinguish different canopy structure types.

## 2. Study site

The study site is Soaproot Saddle (Fig. 1) located in the mid-elevation range of the Sierra National Forest in central California, USA, approximately 1100 m above sea level. The area is composed of a mixed conifer and evergreen and deciduous broadleaf forests, clearly dominated by conifer species mainly *Pinus ponderosa* (ponderosa pine) and *Calocedrus decurrens* (incense cedar). The main broadleaf species are the deciduous *Quercus kelloggii* (black oak) and the evergreen *Quercus chrysolepis* (canyon oak). The forest canopy density varies from dense to relatively open with a dense understory of shrubs and grasses. The topography is complex with steep slopes in some areas. The area has a strong Mediterranean climate characterized by hot and dry summers and mild and wet winters. The mean temperature ranges from 5.5 °C to 18.0 °C, with 805 mm of precipitation per year (http://criticalzone.org/sierra/infrastructure/field-area/flux-tower-at-soaproot-saddle/). The study region covers an area of 2000 ha located within the Soaproot Saddle NEON (National Science Foundation's National Ecological Observatory Network) site.

## 3. Data and methods

### 3.1. AVIRIS preprocessing

AVIRIS data were acquired by the Jet Propulsion Laboratory (JPL) (California Institute of Technology Center) onboard a NASA ER-2 during the HyspIRI flight summer campaign in June 2013. AVIRIS provides surface radiance measurements in 224 spectral bands between 370 and 2500 nm with a bandwidth of 10 nm and a spatial resolution of 18 m (Green, Eastwood, Sarture, Chrien, et al., 1998). The noisy atmospheric absorption bands from 1342 to 1482 nm and from 1800 to 1966 nm were removed; as a result 193 bands were used.

The AVIRIS image was radiometrically calibrated, geometrically orthorectified and atmospherically corrected by JPL. In addition, a topographic correction was applied to avoid the effect of the slope and aspect on reflectance values. We used the model proposed by Soenen, Peddle, and Coburn (2005) which is based on the Sun-Canopy-Sensor (SCS) correction developed by Gu and Gillespie (1998). This correction takes into account the orientation that trees do not grow perpendicular to the inclined terrain. The formulation is as follows (Eq. (1)):

$$\rho_n = \rho \frac{\cos\alpha \cos\theta + C}{\cos i + C} \quad (1)$$

where $\rho_n$ is the normalized reflectance, $\rho$ is the uncorrected reflectance, $\alpha$ is the slope of the terrain, $\theta$ is the solar zenith angle, i is the illumination angle, and C is an empirical parameter introduced by Soenen, Peddle, and Coburn, (2005) to moderate the overcorrection of the SCS correction at large incidence angles. The empirical parameter

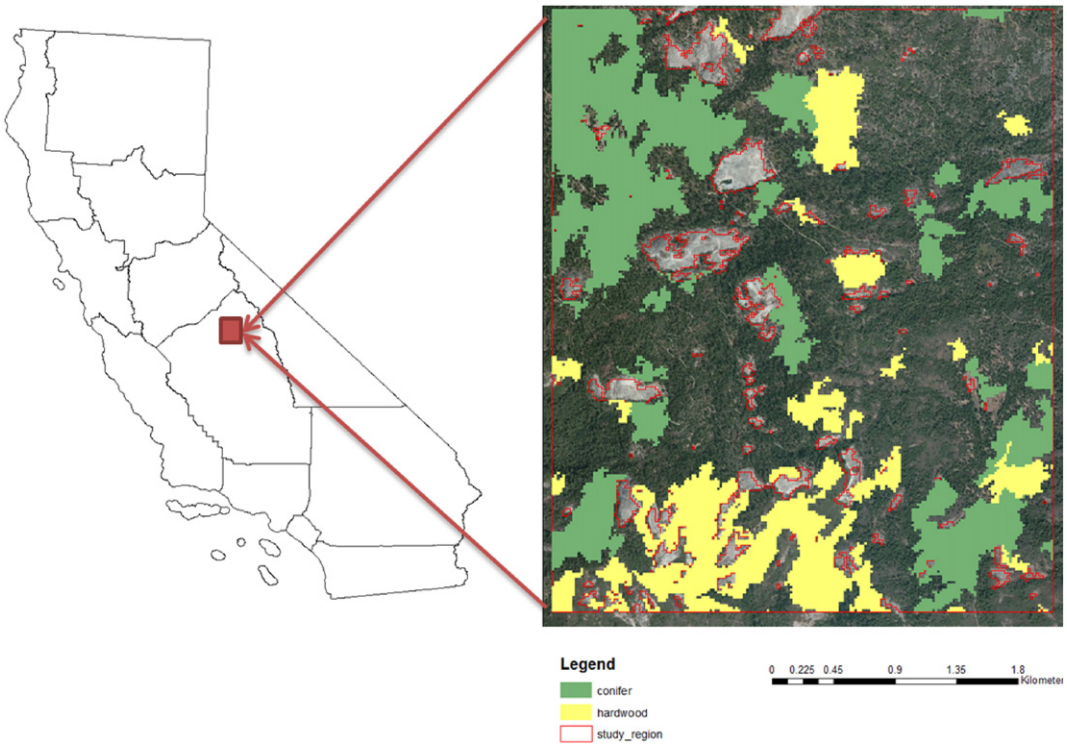


**Fig. 1.** Location and map of the Soaproot Saddle study region with a high resolution aerial image and two major forest types delineated (conifer (green) versus hardwood (yellow)), dark green represents mixed forest and grey rocks.

(C) (Eq. (2)) is a function of the slope (b) and the intercept (a) of the regression line derived from the relationship between the reflectance and the cosine of the illumination angle.

$$\rho = a + b\cos i \Rightarrow C = \frac{a}{b} \quad (2)$$

### 3.2. LiDAR preprocessing

LiDAR data were collected over the same region covered by the AVIRIS data by an Optech Gemini small footprint waveform-recording LiDAR flown by the National Ecological Observatory Network (NEON) (NEON Technical Memo 005, 2013), and provided as discrete returns (up to 4) in LAS Specification 1.4 format. In addition, a Digital Elevation Model (DEM) with 1 m spatial resolution was provided along with the point cloud (19 p/m$^2$), which was used to derive the height of each return above the ground, calculated as the difference between the Z coordinate of the point, and the Z value of the DEM at the same X, Y position.

### 3.3. LiDAR and AVIRIS dataset preparation

LiDAR and AVIRIS data were both acquired in June 2013. LiDAR variables were calculated at 18-m spatial resolution to match that of the AVIRIS data. Pixels with a very low vegetation cover were removed from the analysis; such pixels were identified by NDVI values less than 0.4. Additionally, pure conifer and hardwood pixels were identified using the vegetation map provided by the Forest Service (CALVEG www.fs.fed.us). AVIRIS pixels were labeled as pure conifer or hardwood only when the entire 18 m-pixel was completely within an area of conifer or hardwood according to CALVEG data. In total we analyzed six datasets: a LiDAR and an AVIRIS dataset for three scenarios: all vegetated pixels, conifer forest pixels only and hardwood forest pixels only.

### 3.4. Optical metrics

The optical metrics calculated and evaluated included (Appendix Table A.1): (1) single narrow-band reflectance and 1st derivative, (2) sub-pixel cover fractions, (3) narrow-band indices (4) spectral absorption features, (5) optimized normalized difference indices, and (6) Principal Component Analysis (PCA) components.

#### 3.4.1. Single spectral narrow bands

Reflectance for the 193 narrow-band wavelengths and 192 first order derivative transformed narrow-bands were individually considered. The first derivative is approximated as finite differences.

#### 3.4.2. Sub-pixel fractions

Multiple Endmember Spectral Mixture Analysis (MESMA) (Roberts, Gardner, Church, Ustin, et al., 1998; Dennison, Halligan, & Roberts, 2004) was used to obtain sub-pixel fractions of Green Vegetation (GV), non-photosynthetic vegetation (NPV), soil, and shade for each pixel. MESMA models each pixel's spectrum as a linear combination of pure spectral components or 'endmembers'. The main difference between MESMA and simple Spectral Mixture Analysis (SMA) is that, in MESMA, the number and types of the endmember can vary at the pixel level (Roberts, Gardner, Church, Ustin, et al., 1998). The image is unmixed using all possible combinations of two, three and four endmembers, selecting the best-fit model for each pixel. MESMA provides endmember abundances within the pixel together with a root mean square error (RMSE) model fit. In this analysis, we constrained the maximum allowable RMSE to 2.5%. The minimum and maximum allowable endmember fractions were constrained to fall between −0.05 and 1.05. MESMA was implemented using the ENVI/IDL VIPER Tools add on package (Roberts, Halligan, & Dennison, 2007) which includes a detailed description of the tools.

Due to the lack of pure NPV and soil pixels in this ecosystem at 18 m spatial resolution, a reference spectral library of NPV and soil endmembers was created using spectra collected during fieldwork in 2014 using an ASD full-range spectrometer sensor FieldSpec3 Spectroradiometer (Analytical Spectral Devices Inc., Boulder, CO, USA). These endmembers included several soil types, a range of herbaceous NPV, downed branches, and leaf litter and bark from the dominant tree species. Each spectrum was an average of 50 scans measured approximately 1 m above the surface (~1 cm from the surface for bark) under completely clear sky conditions within 2 h of solar noon. GV endmembers were selected from the AVIRIS image, and photometric shade was used in modeling. Twenty five, ten and twelve spectra were preliminarily selected as NPV, soil, and GV endmembers, respectively. We used three metrics to select the final representative endmembers for each class: (1) Count-based Endmember Selection (COB: Roberts, Dennison, Gardner, Hetzel, et al., 2003) where endmembers are selected based on the maximum number of endmembers that are modeled within their own class. (2) Endmember Average RMSE (EAR: Dennison & Roberts, 2003) where endmembers are selected based on the minimum RMSE within a class and (3) Minimum Average Spectral Angle (MASA: Dennison, Halligan, and Roberts, 2004), where endmembers are selected based on the lowest average spectral angle. The final endmembers selected were the ones high inCOB values and low MASA and EAR values.

#### 3.4.3. Narrow-band indices and absorption features

A set of narrow-band spectral indices (Table 1) and absorption features, which might be related with canopy structure, were calculated. These metrics were selected from the literature and included those indices found to be related to structural parameters such as biomass, LAI or vegetation height. Among the seven spectral narrow-band indices used, some are sensitive to the presence of green foliar biomass, such as Normalized Difference Vegetation Index (NDVI) (Rouse, Haas, Schell, & Deering, 1973), Red Edge Normalized Difference Vegetation Index (NDVI705) (Gitelson & Merzlyak, 1994; Sims & Gamon, 2002), Modified Red Edge Normalized Difference Vegetation Index (mNDVI705) (Sims & Gamon, 2002; Datt, 1999), and Enhanced Vegetation Index (EVI) (Huete, Didan, Miura, Rodriguez, et al., 2002). Others are sensitive to water content, such as Normalized Difference Water Index (NDWI) (Gao, 1996), Normalized Difference Infrared Index (NDII) (Hunt & Rock, 1989) or dry matter content, like the Cellulose Absorption Index (CAI) (Daughtry, 2001).

We additionally calculated several absorption-based derivatives (Wtr1EdgeWvl, Wtr1EdgeWvl: NIR1 water absorption feature edge wavelength and magnitude respectively) and fit water measurements (Wtr1AbAr, Wtr2AbAr: NIR1 and NIR2 water absorption feature areas respectively), and physically-derived values (EWT). The Near-Infrared (NIR) water absorption features are found beginning at approximately

**Table 1**
Narrow-band indices used in this research with their formulas and references.

| Index | Formula | Reference |
|---|---|---|
| NDVI | $\frac{\rho_{800}-\rho_{660}}{\rho_{800}+\rho_{660}}$ | Rouse et al. (1973) |
| NDVI705 | $\frac{\rho_{750}-\rho_{705}}{\rho_{750}+\rho_{705}}$ | Gitelson and Merzlyak (1994), Sims and Gamon (2002) |
| mNDVI705 | $\frac{\rho_{750}-\rho_{705}}{(\rho_{750}+\rho_{705}-2\cdot\rho_{445})}$ | Sims and Gamon (2002), Datt (1999) |
| EVI[a] | $G\cdot\frac{\rho_{800}-\rho_{660}}{\rho_{800}+C_1\rho_{660}-C_2\cdot\rho_{480}+L}$ | Huete et al. (2002) |
| NDWI | $\frac{\rho_{860}-\rho_{1240}}{\rho_{860}+\rho_{1240}}$ | Gao (1996) |
| NDII | $\frac{\rho_{819}-\rho_{1649}}{\rho_{819}+\rho_{1649}}$ | Hunt and Rock (1989) |
| CAI | $0.5\cdot\frac{\rho_{2005}-\rho_{2203}}{\rho_{2106}}$ | Daughtry (2001) |

[a] L = canopy background adjustment for correcting nonlinear, differential NIR and red radiant transfer through a canopy; $C_1$ and $C_2$ = coefficients of the aerosol resistance term (which uses the blue band to correct for aerosol influences in the red band); and G = a gain or scaling factor. The coefficients adopted in the EVI algorithm are, L = 1, $C_1$ = 6, $C_2$ = 7.5, and G = 2.5 (Huete, Justice, & Liu, 1994; Huete, Liu, Batchily, & van Leeuwen, 1997).

958–1073 nm (i.e., NIR1) and 1105–1168 nm (i.e., NIR2), respectively. A detailed explanation of these metrics can be found in Clark and Roberts (2012), Ustin, Riaño, and Hunt (2012) and Hunt, Ustin, and Riaño (2013). EWT was computed using a Beer-Lambert light-extinction model, and additional details can be found in Roberts, Green, and Adams (1997).

#### 3.4.4. Optimized normalized difference indices

We also constructed optimized normalized indices related to canopy structure using the AVIRIS data. Normalized indices are the most commonly used indices to relate remote sensing data and canopy properties (Hansen & Schjoerring, 2003; Ferwerda, Skidmore, & Mutanga, 2005; Zhao, Reddy, Kakani, Read, et al., 2005). Normalized difference indices were calculated with all possible combinations of two AVIRIS bands to explore the band combination indices that had the strongest relationships with canopy structure, as compared to the LiDAR variables. Two types of normalized difference index were computed, one type based on untransformed narrow-bands (Eq. (3)), where $\rho_{\lambda a}$ and $\rho_{\lambda b}$ are the reflectance at bands $\lambda_a$ and $\lambda_b$ respectively. The other used the first derivative transformed narrow-bands (Eq. (4)), where $d\rho_{\lambda a}$ and $d\rho_{\lambda b}$ are the first derivative of reflectance between bands ($\lambda_a$ and $\lambda_{a-1}$) and ($\lambda_b$ and $\lambda_{b-1}$) respectively. The best indices were selected based on the maximum $R^2$ value.

$$ND = \frac{\rho_{\lambda a} - \rho_{\lambda b}}{\rho_{\lambda a} + \rho_{\lambda b}} \quad (3)$$

$$DND = \frac{d\rho_{\lambda a} - d\rho_{\lambda b}}{d\rho_{\lambda a} + d\rho_{\lambda b}} \quad (4)$$

#### 3.4.5. Principal Component Analysis

Principal Component Analysis (PCA) (Richards, 1999) was performed both using the full spectral range as well as specific regions of the spectrum that presented the strongest relationships with canopy structure based on the results from our single band analysis and index optimization. PCA finds a set of orthogonal axes that maximize the variance of the data and produces a set of uncorrelated output variables that are linear combinations of the original values; therefore, the first PC contains the largest percentage of data variance. The information within each selected spectral region is thus summarized by a reduced number of bands. We selected the PCs that together represented 90% of the data variance for use in our analysis. We used the first principal component of reflectance as a proxy for the albedo (Kauth & Thomas, 1976; Small, 2003). The albedo is the fraction of radiation incident that is reflected and this measure has been used to characterize the scattering, thus is expected to be related to canopy structure.

### 3.5. LiDAR metrics

We derived a set of variables from the LiDAR data (Appendix Table A.2) to describe the vegetation structure of the study sites. Since our study area was multilayered with shrub vegetation, which also affected the signal received at the sensor, those returns with a height greater than or equal to 0.5 m was assumed to be tree returns.

The height of the vegetation and its variability were described by means of the maximum (H_max), mean (H_mean), median (H_median) and standard deviation (H_std) of the height, derived from the height distribution of the LiDAR returns. To avoid possible outliers, the 99th percentile was used to describe the maximum height. Canopy surface roughness or complexity was described by the standard deviation of the pixel heights of a canopy height model (CHM_std) created at 1 m spatial resolution. Fractional cover (FC) represents the vertical projection of the vegetation canopy onto the ground, and was used as a proxy to describe the horizontal distribution of the vegetation in the study area. Using LiDAR data, FC can be estimated as the proportion of vegetation returns relative to the sum of vegetation and ground returns. This ratio can be computed based on the first return ratio or based on an all returns ratio (Hopkinson & Chasmer, 2009). Lovell, Jupp, Culvenor, and Coops (2003) showed that FC estimated from first returns (FC_1ret) overestimate the real FC; nevertheless, since first returns resemble the upper canopy more closely we estimated FC using both approaches. Leaf area index (LAI) can be estimated from the canopy cover (Morsdorf, Kotz, Meier, Itten, et al., 2006); however, since no consideration of the spatial distribution of the canopy elements was considered, this estimate represents the effective LAI (LAIe), also since leaf and woody components were no separated, the actual measurement is plant area index (PAI). To describe the vertical distribution of the vegetation for the study area, the Vegetation Vertical Profile (VVP) at a 0.5 m vertical resolution was derived for each pixel as the proportion of canopy returns for each height bin. The area under the profile ($VVP_{int}$) was computed as a proxy for the amount of biomass present in the pixel.

### 3.6. LiDAR and optical metrics relationship

We first assessed correlations among LiDAR variables to evaluate whether these could be grouped in some way to simplify the following analysis. After selecting the LiDAR variables with unique information, the strength and the direction of the relationship between these and the untransformed and the first derivative transformed narrow-band reflectance values were evaluated by Pearson's correlation coefficient. Relationships between all LiDAR variables and the remaining optical metrics were assessed using the coefficient of determination ($R^2$).

### 3.7. Canopy structure assessment

In a first phase, each LiDAR structure variable (e.g. height) was used independently to test whether optical metrics can discriminate different classes (i.e., low, medium, high), and in a second phase we evaluated the ability of optical metrics to discriminate canopy structure types that incorporate multiple structural variables; these are subsequently defined in Section 3.7.1 below. In both phases, we use classification to test the ability of optical metrics to define canopy structure. Our hypothesis is that if we can discriminate different canopy structure types using

**Table 2**
Relationship among all LiDAR variables as measured by the coefficient of determination.

| | LAI | FC | FC_1ret | VVP | H_max | H_mean | H_median | H_std | CHM_std |
|---|---|---|---|---|---|---|---|---|---|
| LAI | 1.00 | 0.94 | 0.92 | 0.87 | 0.21 | 0.04 | 0.03 | 0.40 | 0.29 |
| FC | | 1.00 | 0.98 | 0.94 | 0.24 | 0.06 | 0.05 | 0.41 | 0.32 |
| FC_1ret | | | 1.00 | 0.94 | 0.22 | 0.05 | 0.05 | 0.38 | 0.29 |
| VVP | | | | 1.00 | 0.38 | 0.17 | 0.16 | 0.48 | 0.38 |
| H_max | | | | | 1.00 | 0.82 | 0.79 | 0.77 | 0.74 |
| H_mean | | | | | | 1.00 | 0.99 | 0.38 | 0.38 |
| H_median | | | | | | | 1.00 | 0.35 | 0.34 |
| H_std | | | | | | | | 1.00 | 0.95 |
| CHM_std | | | | | | | | | 1.00 |

optical metrics, optical data can provide critical information on structure for improved retrieval of canopy chemistry. We considered LiDAR variables the reference data for validation purposes.

*3.7.1. Canopy structure type definition using LiDAR*

LiDAR variables (i.e., biomass ($VVP_{int}$), height and complexity) were classified into five ordinal classes (very low, low, medium, high and very high) using an unsupervised IsoData classification. In the first phase each LiDAR variable was used independently and in the second, the three LiDAR structural variables that were binned into five LiDAR classes were combined into a unique classification of canopy structure types. The number of possible combinations among the three variables is 125; these classes were analyzed in terms of their representativeness (i.e., number of pixels) and the proximity among classes with the intent to reduce the number of combinations among classes and variables to an interpretable and ecologically meaningful set. The following analysis was done at the pixel level using each of the different datasets described in Section 3.3 (all vegetated pixels, conifer forest, and hardwood forest pixels).

*3.7.2. Canopy structure types discrimination with optical metrics using Random Forests*

We used the non-parametric regression tree-based classifier Random Forests (RF) (Breiman, 2001; Pal, 2005) to discriminate canopy structural types from the optical metrics. Several studies have shown the utility of non-parametric regression tree methods for ecological applications (Hansen, DeFries, Townshend, Sohlberg, et al., 2002; Moisen & Frescino, 2002; Cutler, Edwards, Beard, Cutler, et al., 2007; Blackard, Finco, Helmer, Holden, et al., 2008). Among these methods, RF is well-suited for analyzing complex non-linear relationships, and it has been satisfactorily used with forest remote sensing data (Ahmed, Franklin, Wulder, and White, 2015; Chan & Paelinckx, 2008; Falkowski, Evans, Martinuzzi, Gessler, et al., 2009; Powell, Cohen, Healey, Kennedy, et al., 2010; Kantola, Vastaranta, Yu, Lyytikainen-Saarenmaa, et al., 2010; Pierce, Farris, & Taylor, 2012; Naidoo, Cho, Mathieu, & Asner, 2012). The main advantages of this method are that (1) it does not require any assumptions between independent and dependent variables (Olden, Lawler, and Poff, 2008); (2) it allows selection of the best variables without any a priori selection; (3) although the whole data set is used, each run has an independent validation; and (4) it provides a ranking of the most important variables (Clark & Roberts, 2012).

RF generates a user-defined number of decision trees. Each tree is created with a randomly selected set of training samples (the reference data) and independently tested with the remaining observations. After the maximum number of trees has been reached, each observation's class is determined by the majority chosen class of all decision trees. The number of decision trees was set at 500 to ensure that every observation was predicted at least several times. This number was selected by running RF with increasing numbers of trees from 10 to 1000. Classification error decreased with increasing number of trees until it reached 500, at which point, the error stabilized. The number of variables randomly sampled at each split was sqrt(p), where p is the number of variables. A third of the observations were randomly selected with replacement from each class for training, and the remaining two thirds retained for validation. The error rate for each classification was evaluated with a confusion matrix, and classification accuracy was estimated by the percentage of Out Of Bag (OOB) data (Breiman, 2001) and the Kappa index.

OOB error is the proportion of misclassified data. Each tree is built using a different bootstrap sample of the original data, and the remaining data is used to test the classification. The proportion of times that an observation is misclassified averaged over all bootstrap samples is the OOB error. Kappa measures the accuracy based on the difference between the actual agreement in the error matrix (values in the major diagonal) and the chance agreement which is indicated by values of rows and columns across the major diagonal (Congalton, 1991) (Eq. (5)), where $k$ is the number of classes.

$$K = \frac{n \cdot \sum_{i-1}^{k} n_{ii} - \sum_{i-1}^{k} n_{i+} \cdot n_{+i}}{n^2 - \sum_{i-1}^{k} n_{i+} \cdot n_{+i}} \quad (5)$$

It is assumed that n samples are distributed into $k^2$ cells. Each sample is assigned to one class ($k$) in the classification. $n_{ij}$ indicates the number

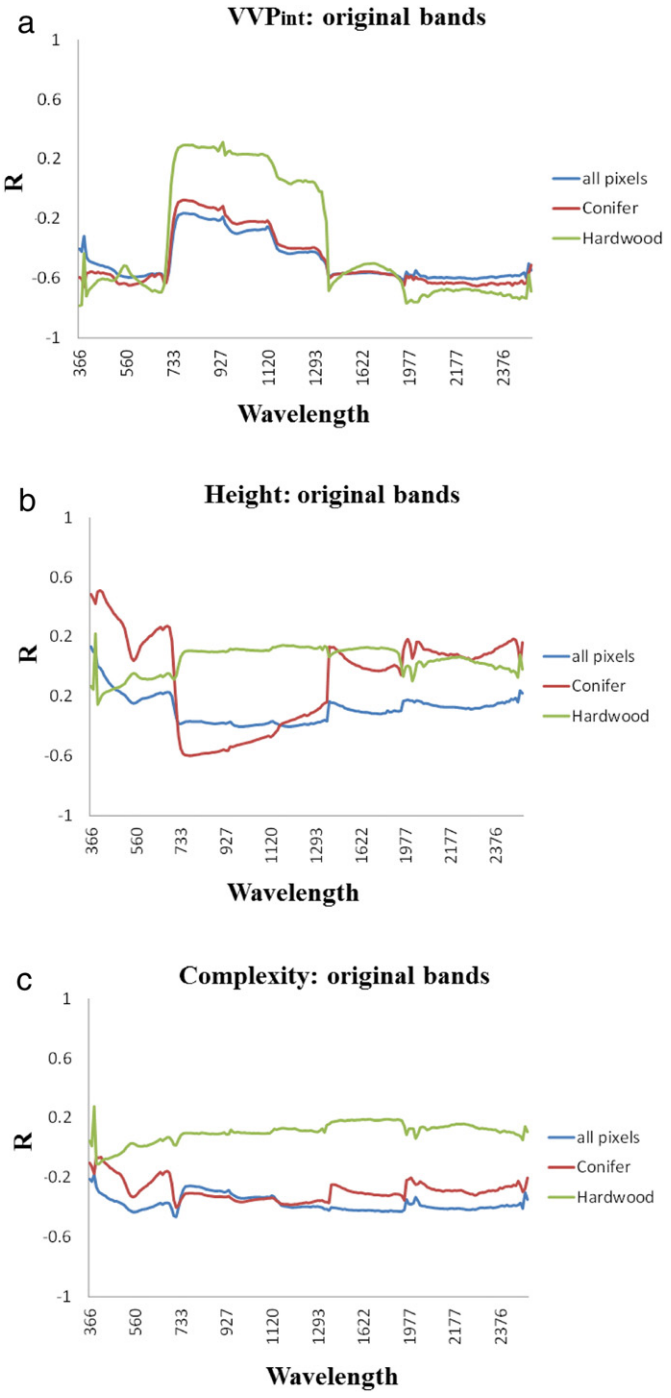


**Fig. 2.** The strength and the direction of Pearson's correlation coefficient between LiDAR variables and original reflectance narrow-bands for a) biomass ($VVP_{int}$), b) height and c) complexity related LiDAR variables as defined in Section 4.

of samples of class j that are classified into class i. $n_{i+}$ and $n_{+i}$ are represented by the following formulas (Eqs. (6) and (7))

$$n_{i+} = \sum_{j-1}^{k} n_{ij} \quad (6)$$

$$n_{+j} = \sum_{i-1}^{k} n_{ij} \quad (7)$$

RF was also used to rank the importance of each optical metric for differentiating the structural classes. RF provides two measures of variable importance: 1) the decrease in root mean square error that quantifies the observed decrease in error when the appropriate variable is included in the model, and 2) the Gini index that quantifies node impurity, that is, the degree to which a variable generates terminal nodes. The RF algorithm (Liaw & Wiener, 2002) was implemented in the R statistical program (3.0.1) (R Development Core Team, 2014).

We ran the RF analysis two times. The first time vegetated pixels from across the entire study region were included. In the second run, the study area was split up into two regions. The RF analysis was run on one subset and then the classifier obtained from this region was applied to the second region. We did this to assess how portable the relationships we observed were when applied to an area not used in model building or assessment. Within each analysis, we further sub-divided our data to evaluate whether accounting for vegetation type could improve the results (i.e., conifer forest and hardwood forest). Lastly, because our optimized normalized difference indices were data dependent, they could introduce a positive bias in the modeling process. To address this, we ran each analysis with and without these variables.

## 4. Results

The results are divided into four subsections. Section 4.1 contains analysis of the relationship among LiDAR structural variables. Section 4.2 presents the relationships between LiDAR structure and optical metrics. In Section 4.3, we present the results of index optimization, and Section 4.4 contains the results of the canopy structural type classifications. Section 4.4 is further divided into two sections. Section 4.4.1 contains the results of each the structural variables independently (Section 4.4.1.1 biomass ($VVP_{int}$), Section 4.4.1.2 height and Section 4.4.1.3 complexity) and Section 4.4.2 presents the results of the combination of all structural variables, defining the so-called structural types.

### 4.1. LiDAR variable relationship

We observed strong inter-correlations (i.e., ≥0.85) among LiDAR variables with respect to three vegetation structural attributes (Table 2): variables related with biomass (LAI, FC, FC_1ret and $VVP_{int}$), variables related with height (H_mean, H_median and H_max), and variables related with vegetation complexity or canopy roughness (H_std and CHM_std). Because of this, we elected to use $VVP_{int}$ as a proxy of the amount of canopy material, i.e., qualitative measure of biomass, based on previous studies that showed that the vertical profile can be related to biomass (Harding, Lefsky, Parker, & Blair, 2001; Zhao, Popescu, & Nelson, 2009). In this study, we used the integral of VVP. H_mean was selected to represent height, and CHM_std to represent vegetation complexity. Subsequent analyses used these three attributes.

### 4.2. LiDAR-optical metrics relationship

Fig. 2 shows the Pearson correlations between the original narrowband reflectance and biomass ($VVP_{int}$), height, and complexity. Table 3 shows the relationships between the structural properties as described by LiDAR and AVIRIS-derived optical metrics considering the three scenarios studied (i.e., all vegetated pixels, conifer and hardwood forest).

Results showed that Pearson correlations between biomass ($VVP_{int}$) and reflectance differed across the spectrum and by vegetation type (Fig. 2). Biomass ($VVP_{int}$) in both hardwood and conifer forest types were strongly negatively correlated (i.e., $R \leq -0.6$ for most spectral bands) in the visible and shortwave-infrared (SWIR) parts of the spectrum, and in hardwood forest, there was also a weak positive correlation, with Pearson coefficients between 0.3 and 0.4, in the NIR region, up to 1150 nm. Conifer height was positively correlated with the visible part of the spectrum and negatively with the NIR. On the other hand, hardwood forests lacked any clear part of the spectrum highly correlated with height. Correlations with vegetation complexity had similar, generally low, values across the spectrum. When using the first derivative of reflectance, we observed generally higher correlations across the

**Table 3**
The strength and the direction of the relationship between LiDAR variables, as defined in Section 4.1, and optical metrics as measured by the coefficient of determination considering all vegetated pixels, conifer pixels and hardwood forest pixels.

| | ALL | | | | | | CONIFER | | | | | | HARDWOOD | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $VVP_{int}$ | | Height | | Complexity | | $VVP_{int}$ | | Height | | Complexity | | $VVP_{int}$ | | Height | | Complexity | |
| GV | 0.01 | + | 0.06 | − | 0.01 | − | 0.03 | + | **0.38** | − | 0.04 | − | 0.32 | + | 0.00 | + | 0.00 | − |
| SHADE | 0.38 | + | **0.29** | + | **0.36** | + | 0.31 | + | 0.26 | + | **0.34** | + | 0.10 | + | 0.02 | + | 0.02 | + |
| NDVI | 0.31 | + | 0.01 | + | 0.13 | + | 0.27 | + | 0.13 | − | 0.02 | + | 0.54 | + | 0.02 | + | 0.00 | + |
| NDVI705 | 0.34 | + | 0.01 | + | 0.14 | + | 0.36 | + | 0.14 | − | 0.03 | + | 0.50 | + | 0.05 | + | 0.02 | + |
| mNDVI705 | 0.35 | + | 0.02 | + | 0.15 | + | 0.38 | + | 0.07 | − | 0.06 | + | 0.45 | + | 0.05 | + | 0.02 | + |
| EVI | 0.04 | + | 0.04 | − | 0.00 | + | 0.05 | + | **0.35** | − | 0.02 | − | 0.39 | + | 0.01 | + | 0.00 | + |
| NDWI | 0.19 | + | 0.01 | + | 0.06 | + | 0.22 | + | 0.11 | − | 0.02 | + | 0.13 | + | 0.00 | − | 0.00 | + |
| NDII | 0.31 | + | 0.02 | + | 0.13 | + | 0.29 | + | 0.12 | − | 0.03 | + | 0.50 | + | 0.00 | + | 0.00 | − |
| CAI | 0.25 | − | 0.00 | − | 0.10 | − | 0.23 | − | 0.08 | + | 0.02 | − | **0.67** | − | 0.01 | − | 0.00 | + |
| EWT | 0.39 | + | 0.06 | + | 0.12 | + | 0.42 | + | 0.03 | + | 0.16 | + | 0.09 | + | 0.00 | − | 0.00 | + |
| Wtr1EdgeWvl | 0.10 | + | 0.00 | + | 0.02 | + | 0.15 | + | 0.02 | − | 0.01 | + | 0.20 | + | 0.04 | − | 0.04 | − |
| Wtr1EdgeMag | 0.10 | + | 0.00 | + | 0.02 | + | 0.15 | + | 0.02 | − | 0.01 | + | 0.19 | + | 0.04 | − | 0.04 | − |
| Wtr1AbAr | 0.35 | + | 0.04 | + | 0.12 | + | 0.42 | + | 0.00 | + | 0.12 | + | 0.25 | + | 0.01 | − | 0.01 | − |
| Wtr2AbAr | **0.46** | + | 0.09 | + | 0.20 | + | **0.47** | + | 0.04 | + | 0.21 | + | 0.30 | + | 0.00 | + | 0.00 | + |
| PC1 | 0.39 | − | **0.25** | − | **0.34** | − | 0.39 | − | 0.21 | − | **0.36** | − | 0.22 | − | 0.00 | + | 0.02 | + |
| PC2 | 0.07 | + | 0.02 | − | 0.00 | + | 0.10 | + | 0.35 | − | 0.02 | − | 0.42 | + | 0.01 | − | 0.04 | − |
| VIS-PC1 | **0.43** | + | 0.04 | + | 0.22 | + | **0.51** | + | 0.05 | − | 0.09 | + | **0.59** | + | 0.00 | − | 0.02 | − |
| VIS-PC2 | 0.30 | − | 0.14 | − | 0.19 | − | 0.42 | − | 0.01 | + | 0.10 | − | 0.05 | + | 0.10 | − | 0.12 | − |
| NIR-PC1 | 0.26 | − | **0.27** | − | **0.29** | − | 0.19 | − | **0.41** | − | **0.35** | − | 0.00 | + | 0.00 | − | 0.00 | − |
| NIR-PC2 | 0.17 | + | 0.01 | + | 0.06 | + | 0.21 | + | 0.15 | − | 0.01 | + | 0.11 | + | 0.01 | − | 0.01 | − |
| SWIR1-PC1 | 0.42 | − | 0.13 | − | 0.27 | − | 0.46 | − | 0.00 | − | 0.17 | − | 0.45 | − | 0.02 | + | 0.06 | + |
| SWIR2-PC1 | **0.44** | − | 0.10 | − | 0.24 | − | **0.51** | − | 0.00 | + | 0.13 | − | **0.62** | − | 0.00 | + | 0.04 | + |

spectrum, but no one wavelength region stood out as being especially strongly related to the LiDAR variables. For height, the improvement was the greatest for conifer pixels.

The relationship between the structural properties optical metrics varied in direction and magnitude (Table 3). Wtr2AbAr, SWIR2-PC1 and VIS-PC1 had the highest correlations with biomass ($VVP_{int}$) when all vegetated pixels were considered and for conifer forest alone. For hardwood forest, the best correlations were found with CAI, SWIR2-PC1 and VIS-PC1. Lower correlations occurred between height and optical metrics. The highest correlations were found with shade, NIR-PC1 and PC1 when all vegetated pixels were considered. Within conifer forest NIR-PC1, GV and EVI had the highest correlations. However, none of the optical metrics consistently showed significant correlation (i.e., $R^2 \leq 0.10$ for all optical metrics) with height in hardwood forests. The weakest correlations were found with vegetation complexity. Shade, PC1 and NIR1-PC1 had the highest correlations, when all vegetated pixels were considered. Relatively low to moderate correlations, with $R^2 \geq 0.30$, occurred for conifer forests, for PC1, NIR-PC1 and shade. No optical metrics showed significant correlation in hardwood forests.

### 4.3. Index optimization

The index optimization analysis is summarized in Fig. 3 and Table 4. The best band combination optimized index and the spectral regions most related to canopy structure are presented in Fig. 3. Table 4 shows the best band combination normalized difference index using original (ND) and the first derivative narrow-bands (DND) compared with the best correlated optical metric found in the previous section for the three LiDAR-derived metrics selected.

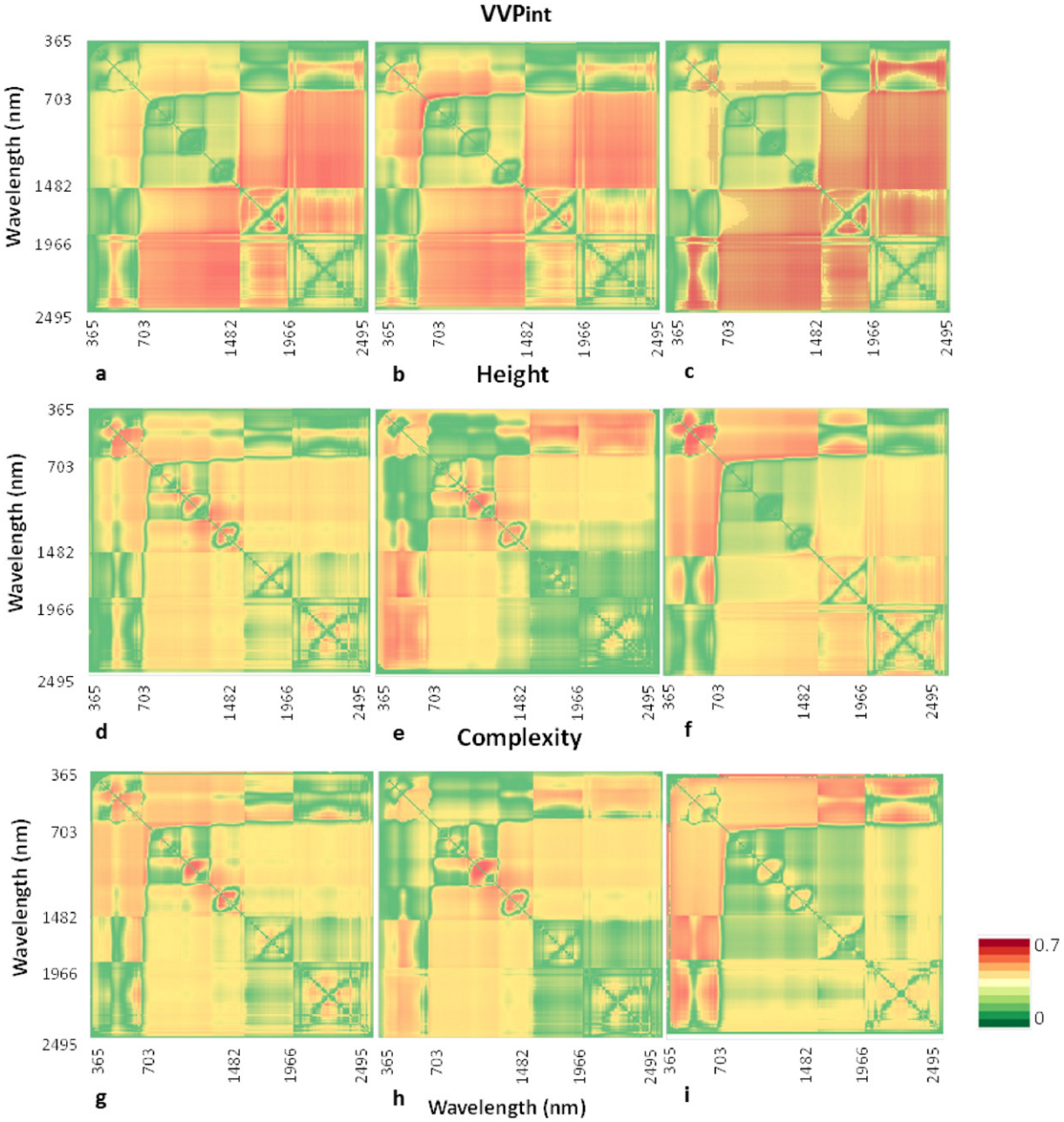


**Fig. 3.** Coefficient of determination from the all possible band combinations of normalized difference indices using untransformed narrow-bands related to biomass ($VVP_{int}$) LiDAR variable for a) all pixels b) conifer pixels and c) hardwood pixels; and those related to LiDAR height for: d) all pixels e) conifer pixels and f) hardwood pixels; and those related to LiDAR complexity for g) all pixels h) conifer pixels and i) hardwood pixels.

**Table 4**
The coefficient of determination of the optical metrics with the highest correlations and the best band combination normalized difference index using original (ND) and the first derivative narrow-bands (DND), for biomass ($VVP_{int}$), height and complexity LiDAR variables considering all pixels, conifer pixels or hardwood forest pixels.

| | All | | Conifer | | Hardwood | |
|---|---|---|---|---|---|---|
| | Index | $R^2$ | Index | $R^2$ | Index | $R^2$ |
| $VVP_{int}$ | Wtr2AbAr | 0.50 | VIS-PC1 | 0.61 | CAI | 0.67 |
| | ND (1148–1800) | 0.50 | ND (733–791) | 0.48 | ND (560–2425) | 0.68 |
| | DND (937–927/2186–2176) | 0.45 | DND (752–742/2186–2176) | 0.40 | DND (667–657/2256–2246) | 0.64 |
| Height | shade | 0.40 | NIR-PC1 | 0.50 | VIS-PC2 | 0.10 |
| | ND (1138–1148) | 0.33 | ND (956–946) | 0.34 | ND (560–628) | 0.46 |
| | DND (521–511/849–840) | 0.40 | DND (927–917/1043–1033) | 0.41 | DND (540–530/684–674) | 0.43 |
| Complexity | shade | 0.36 | PC1 | 0.42 | VIS-PC2 | 0.12 |
| | ND (1205–1215) | 0.30 | ND (1014–1043) | 0.33 | ND (375–713) | 0.36 |
| | DND (840–830/1214–1205) | 0.37 | DND (849–840/1043–1033) | 0.36 | DND (723–713/713–703) | 0.33 |

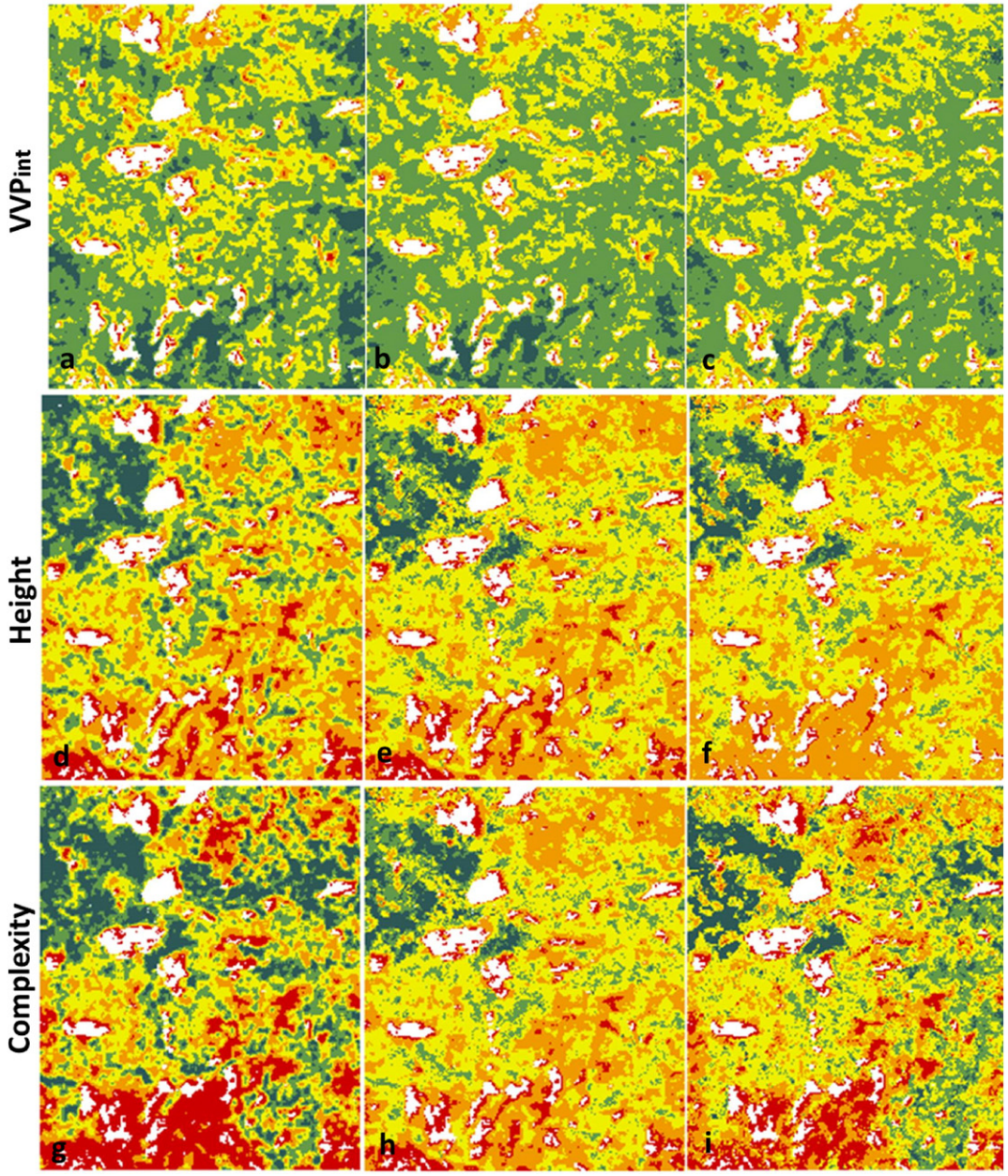


**Fig. 4.** Frames in column 1 show biomass ($VVP_{int}$) (a), height (d) and complexity (g) as classified from the LiDAR data. Frames in column 2 show biomass ($VVP_{int}$) (b), height (e) and complexity (h) as predicted from AVIRIS using training data acquired across the entire study region. Frames in column 3 show biomass ($VVP_{int}$) (c), height (f) and complexity (i) as predicted by AVIRIS but using training data from only half of the study region and applied to the entire region. Color code to indicate: Very low (red), low (orange), medium (yellow), high (light green) and very high (dark green). No data (white).

Overall, we did not find an optimized index that greatly improved $R^2$ values with biomass ($VVP_{int}$), height and vegetation complexity (Fig. 3) when considering all vegetation pixels or conifer forest. However, for hardwood we found stronger relationships with height and vegetation complexity, with an increase from 0.10 to 0.46 and from 0.12 to 0.36, respectively. Among optimized indices, the ones based on the first derivative transformed narrow-bands were slightly more strongly correlated.

The optimized narrow-band index analysis also provided information about the spectral regions more related to canopy structure in terms of biomass ($VVP_{int}$), height and vegetation complexity. When all vegetation pixels were considered, the highest correlations with biomass ($VVP_{int}$) occurred when bands in the NIR and SWIR regions of the spectrum were combined. When separated into conifer and hardwood pixels, we also observed higher correlations using these bands, but the best bands for hardwood were a combination of visible and SWIR bands, and a combination of two NIR bands (~733 nm and 791 nm) in conifer forests. For height, the best band indices for all vegetation pixels were found using wavelengths around 1100 nm, coincident with a water absorption feature, and with slightly weaker correlations around 600 nm. In conifer forests, we found the strongest correlations with indices using pairs of wavelengths in the NIR region, but also found good correlations using bands combining a visible and SWIR wavelength pair. In hardwoods the highest correlations were found using pairs of bands in the visible or in the NIR parts of the spectrum, particularly around 600 nm. For complexity, indices using the bands around 1100 nm were highly correlated for all vegetated pixels and conifer forest only. However for hardwood forest, good correlations were found using bands from the visible to NIR-SWIR infrared portion of the spectrum with most other bands, and the best band combinations were between bands 375 and 713 nm, and between 1482 and 1966 nm.

When all vegetation pixels were considered, the highest correlations between first derivative optimized narrow-band index with biomass ($VVP_{int}$) occurred when bands in the NIR (937–927 nm) and SWIR (2186–2176 nm) regions of the spectrum were combined. When separated into conifer and hardwood pixels, we observed the highest correlations using the combination of visible (752–742 nm and 667–657 nm for conifer and hardwood respectively) and SWIR (2186–2176 nm and 2256–2246 nm for conifer and hardwood respectively) bands. For height, the best band indices using first derivatives bands for all vegetation pixels were found using visible (521–511 nm) and NIR (849–840 nm) parts of the spectrum. In conifer forest, the strongest correlations were combining bands in the NIR (927–917 nm and 1043–1033 nm) region. In hardwood the highest correlations occurred using pairs of bands in the visible (540–530 nm and 684–674 nm) parts of the spectrum. For complexity, the highest correlations were found using bands in the NIR (840–830 nm and 1214–1205 nm for all pixels, and 849–840 nm and 1043–1033 nm for conifer forest). In hardwood forest the highest correlations were found using bands around 700 nm (723–713 nm and 713–703 nm). The highest correlated optical metrics with biomass ($VVP_{int}$), height and vegetation complexity and the best band combination index using untransformed and the first derivative transformed narrow-bands are summarized in Table 4.

### 4.4. Canopy structural type discrimination with optical metrics using random forests

#### 4.4.1. Individual variables

Fig. 4 shows results from RF classifications for the three structural variables and Table 5 summarizes the overall accuracy and Kappa coefficient of biomass ($VVP_{int}$), height and complexity classifications for the three scenarios (i.e., all vegetated pixels, conifer and hardwood forest). There were no differences in accuracy when optimized indices were included or removed from the classifications, thus results will be explained considering the 24 variables in the classifications. By examining the spatial error distribution (Fig. 5), we observed most errors occurred between the most structurally similar classes (Table 6 and Fig. 5). Since the class thresholds were not defined following ecological criteria, we considered the impact of small overlapping zones between adjacent classes in order to evaluate the magnitude of the error and to evaluate the border effect. We greatly improved the classification accuracy for the three variables (i.e., biomass ($VVP_{int}$), height and complexity) in all the studied scenarios when considering these overlapping zones (Tables 5 & 6). Table 7 shows the percentage error for the prediction and the extrapolation of biomass ($VVP_{int}$), height, and complexity classes. The similarity between prediction and extrapolation errors in most cases demonstrates the portability of the method developed. Finally, Table 8 shows the top five optical metrics for structural class prediction based on the decrease in root square mean error and Gini index for biomass ($VVP_{int}$), height, and vegetation complexity. The following subsection explains in more detail the results found for each LiDAR-derived metric (i.e., biomass ($VVP_{int}$), height and vegetation complexity).

*4.4.1.1. Biomass ($VVP_{int}$).* In terms of vegetation stratification, the highest overall accuracy for biomass ($VVP_{int}$) classifications was achieved when all vegetated pixels were included. The overall accuracy for conifer and hardwood forest was 58.40% and 61.80%, respectively. The lowest errors were found in classes 3 (39.10%) and 4 (17.86%) for all vegetated pixels (Table 6). These classes make up the greatest number of pixels within the study area. Within conifer forest, the lowest errors were found in classes 1, 3 and 4. Within hardwood forest, class errors were similar among classes.

Fig. 4a, b and c show a classification comparison among LiDAR-derived biomass ($VVP_{int}$) and RF results, with and without an independent validation dataset. In general terms, results were similar among

**Table 5**
Classification accuracy and Kappa coefficient for three structural variables (biomass ($VVP_{int}$), height and complexity as defined in Section 4.1.) and the three scenarios studied (all vegetated pixels, conifer pixels and hardwood forest pixels) considering all variables (24 variables: 24v) or removing the optimized indices (22 variables: 22v), and with and without considering overlap between consecutive classes.

| | | | $VVP_{int}$ | | Height | | Complexity | |
|---|---|---|---|---|---|---|---|---|
| | | | Accuracy | Kappa | Accuracy | Kappa | Accuracy | Kappa |
| All | 24v | Without | 68.87 | 0.48 | 62.09 | 0.47 | 52.98 | 0.40 |
| | | With | 89.35 | 0.83 | 88.54 | 0.85 | 84.74 | 0.82 |
| | 22v | Without | 68.05 | 0.47 | 60.83 | 0.45 | 53.20 | 0.40 |
| | | With | 89.27 | 0.83 | 88.33 | 0.84 | 86.66 | 0.84 |
| Conifer | 24v | Without | 58.40 | 0.56 | 59.05 | 0.57 | 49.20 | 0.47 |
| | | With | 78.63 | 0.77 | 78.06 | 0.77 | 67.30 | 0.66 |
| | 22v | Without | 57.24 | 0.55 | 57.41 | 0.55 | 48.35 | 0.46 |
| | | With | 78.08 | 0.77 | 76.72 | 0.76 | 66.40 | 0.65 |
| Hardwood | 24v | Without | 61.80 | 0.61 | 61.04 | 0.60 | 59.39 | 0.58 |
| | | With | 81.29 | 0.81 | 80.43 | 0.80 | 78.43 | 0.78 |
| | 22v | Without | 59.59 | 0.59 | 60.31 | 0.59 | 57.82 | 0.57 |
| | | With | 85.18 | 0.85 | 79.39 | 0.79 | 76.84 | 0.76 |

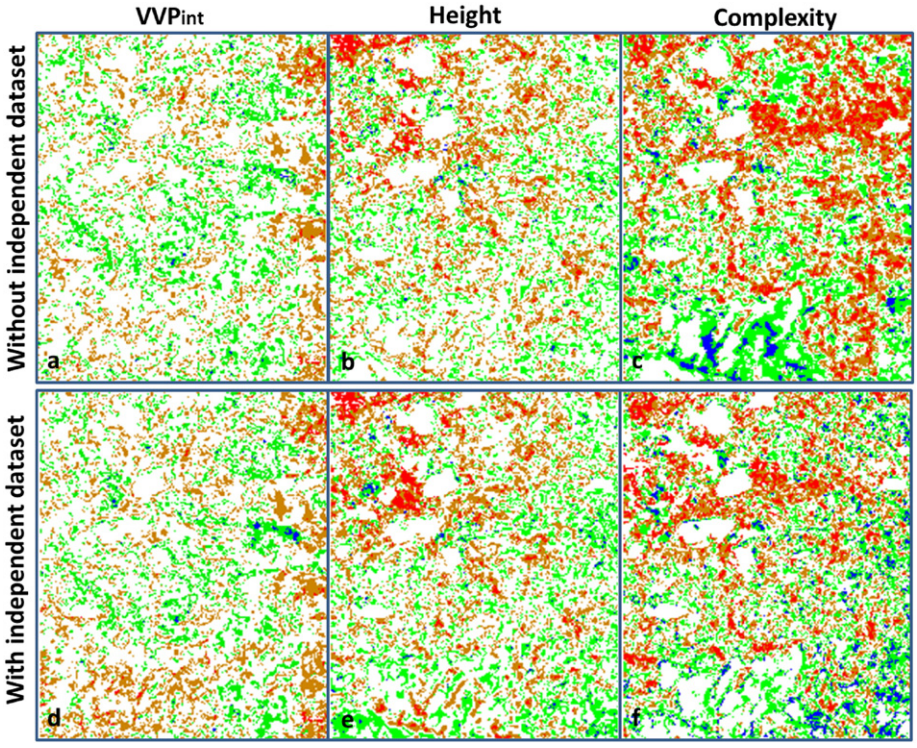


**Fig. 5.** Spatial distribution of classification error for biomass ($VVP_{int}$) (a), height (b) and complexity (c) considering the entire study region (row 1) and classification error for biomass ($VVP_{int}$) (d), height (e) and complexity (f) considering just half of the study region but applied to the full area (row 2). Color code: Green: one class overestimated, blue: more than one class overestimated, yellow: one class underestimated, and red: more than one class underestimated.

classifications, although high biomass ($VVP_{int}$) (Class 5) tended to be underestimated in the eastern and southern parts of the image, especially when an independent dataset was used (Fig. 4c). Fig. 5a and d show the spatial distribution of class error, and the error percentage by vegetation type can be found in Table 7. The lowest accuracy was found in mixed pixels, with a similar percentage using either the entire image or just half for training. The best overall accuracy was achieved for the conifer forest, where 73.58% and 71.11% of the pixels were correctly classified by RF with non-independent and independent datasets respectively. The greatest differences between the two RF classifications were found in hardwood forest with overall accuracies of 70.67% when the whole dataset was considered and 58.65% when an independent validation dataset was used. Errors were similar for mixed pixels and conifer forest, while within hardwood forest, underestimation was more common, especially when an independent dataset was used.

Based on mean decrease in RF overall accuracy ranking (Table 8), the optimized index using the first derivative transformed narrow-bands was the most important optical metric for biomass ($VVP_{int}$) when all metrics were considered, while CAI was the most important without including optimized indices. In addition, metrics sensitive to photosynthetic pigments (EVI, GV) together with more water-sensitive metrics (EWT) or principal components (VIS-PC2, PC2) were within the top-five most important optical metrics. Based on Gini index ranking (Table 8), the optimized indices using the first derivative transformed narrow-bands were the most important optical metrics for biomass ($VVP_{int}$). Within the top-five important metrics we also found the first principal components of the SWIR2 region together with indices sensitive to photosynthetic pigment, such as, NDVI and mNDVI705. When optimized indices were removed, CAI and NDII entered the top five.

*4.4.1.2. Height.* The highest accuracy for height classifications (Table 5) was found without vegetation stratification showing similar overall accuracy for hardwood forest (i.e., 62.09% vs 61.04%). The overall accuracy for conifer forest was 59.05%. Table 6 shows the percentage of class error for height classifications in the three scenarios studied. Considering all vegetated pixels, the highest errors were found in class 4 with a percentage of 70.80%. In conifer and hardwood forests the highest errors were found in class 5 (high biomass ($VVP_{int}$)), but this class only represented 6% and 3% of the conifer and hardwood pixels, respectively. In the hardwood forest, class 2 also showed high errors with an accuracy of 60.43%.

Results comparing LiDAR-derived height classes and RF classifications showed the high similarity (Fig. 4d, e and f). Height tended to be underestimated in the northwestern portion of the image, especially in the high height class (Class 5), while height tended to be overestimated in the northeastern portion and in the south, especially when considering the independent dataset. Fig. 5b and e show the

**Table 6**
Percentage of class error for the biomass ($VVP_{int}$), height and complexity classifications in the three scenarios studied (all vegetated pixels, conifer pixels and hardwood forest pixels) considering all variables (24v), and with and without overlapping between consecutive classes.

| | | $VVP_{int}$ | | Height | | Complexity | |
|---|---|---|---|---|---|---|---|
| | | Without | With | Without | With | Without | With |
| All | Class1 | 54.75 | 20.86 | 36.77 | 1.09 | 18.18 | 3.40 |
| | Class2 | 53.04 | 20.73 | 31.75 | 3.70 | 59.92 | 10.84 |
| | Class3 | 39.10 | 13.03 | 26.92 | 7.20 | 41.08 | 9.84 |
| | Class4 | 17.86 | 6.33 | 70.80 | 29.29 | 65.85 | 22.63 |
| | Class5 | 62.03 | 17.34 | 56.86 | 40.60 | 51.58 | 26.80 |
| Conifer | Class1 | 37.11 | 12.67 | 41.05 | 16.36 | 41.08 | 27.99 |
| | Class2 | 56.72 | 32.61 | 30.59 | 14.98 | 53.95 | 32.81 |
| | Class3 | 32.72 | 17.02 | 45.82 | 24.88 | 45.60 | 29.08 |
| | Class4 | 40.17 | 20.38 | 50.03 | 31.78 | 57.52 | 36.44 |
| | Class5 | 73.42 | 38.62 | 70.04 | 51.75 | 67.08 | 50.57 |
| Hardwood | Class1 | 37.01 | 12.75 | 23.80 | 12.32 | 38.41 | 20.88 |
| | Class2 | 38.95 | 14.22 | 61.17 | 34.64 | 45.81 | 23.68 |
| | Class3 | 50.95 | 20.02 | 30.32 | 15.24 | 39.18 | 20.87 |
| | Class4 | 37.20 | 13.51 | 41.11 | 18.12 | 40.27 | 21.95 |
| | Class5 | 34.25 | 11.50 | 82.72 | 52.47 | 48.55 | 32.20 |

**Table 7**
Percentage of error for the prediction and the extrapolation of biomass ($VVP_{int}$), height and complexity classifications for conifers, hardwoods and mixed forest types.

| | | | $VVP_{int}$ | | Height | | Complexity | |
|---|---|---|---|---|---|---|---|---|
| | | | Prediction | Extrapolation | Prediction | Extrapolation | Prediction | Extrapolation |
| Conifer | Overestimated | One class | 14.13 | 13.99 | 16.12 | 17.47 | 15.34 | 18.49 |
| | | More than one class | 0.24 | 0.35 | 0.87 | 1.06 | 1.83 | 4.43 |
| | Underestimated | One class | 11.92 | 14.29 | 20.30 | 20.84 | 29.78 | 21.55 |
| | | More than One class | 0.14 | 0.26 | 5.92 | 7.32 | 10.92 | 10.95 |
| | Reference | | 73.58 | 71.11 | 56.78 | 53.30 | 42.12 | 44.58 |
| Hardwood | Overestimated | One class | 10.69 | 6.73 | 11.65 | 20.16 | 46.75 | 31.39 |
| | | More than one class | 0.24 | 0.25 | 0.08 | 0.27 | 9.75 | 5.02 |
| | Underestimated | One class | 18.25 | 32.94 | 14.67 | 15.49 | 6.57 | 7.59 |
| | | More than one class | 0.16 | 1.43 | 0.51 | 1.59 | 2.37 | 1.78 |
| | Reference | | 70.67 | 58.65 | 73.10 | 62.49 | 34.57 | 54.22 |
| Mixed pixels | Overestimated | One class | 17.12 | 15.95 | 14.76 | 18.66 | 19.08 | 19.65 |
| | | More than one class | 0.32 | 0.37 | 0.55 | 0.82 | 1.52 | 4.11 |
| | Underestimated | One class | 14.69 | 17.25 | 20.47 | 19.99 | 26.52 | 21.86 |
| | | More than one class | 0.70 | 0.87 | 2.53 | 2.90 | 9.98 | 6.71 |
| | Reference | | 67.16 | 65.56 | 61.70 | 57.64 | 42.90 | 47.67 |

spatial distribution of class error, and the error percentage by vegetation type is found in Table 7. Although most of the errors correspond to adjacent classes, some relatively continuous areas with more than one class underestimated errors were found especially in northwest part of the study area. In general, heights were underestimated. In terms of vegetation stratification, the conifer forest produced the lowest overall accuracies with a similar percentage using either the entire image or with just half for training. The hardwood forest showed the lowest error percentage with overall accuracies of 73.10% and 62.49% with non-independent and independent datasets, respectively. The highest differences between RF classifiers were found in mixed pixels with overall accuracies of 61.70% and 57.64% with and without non-independent dataset.

Optimized indices using the first derivative transformed narrow-bands were the most important optical metric to classify height based on mean decrease in RF overall accuracy ranking (Table 8), while CAI was the most important without the optimized indices. The top-five most important metrics also included water-sensitive metrics such as Wtr1Edge Wvl, NDWI, NDII and the second principal component of the NIR part of the spectrum when the optimized indices were removed. Based on Gini index ranking (Table 8), the optimized index using the first derivative transformed narrow-bands was also the most important optical metric for height together with NDVI705, mNDVI705, Wtr2AbAr and VIS-PIC1. The most important metrics without the optimized indices in the classification were similar except the EWT entered and DND-height dropped out of the top-five metrics.

*4.4.1.3. Complexity.* In terms of vegetation stratification, the highest overall accuracy for complexity classifications (Table 5) was achieved in hardwood forest (i.e., 59.39%) showing similar results when all pixels are considered (i.e., 52.98%). The lowest overall accuracies were found in the conifer forest (i.e., 49.20%). By examining the class error distribution (Table 6) the highest errors for all vegetated pixels were found in classes 2 and 4 (i.e., 59.92% and 65.85%, respectively). Conifer and hardwood forests showed similar error percentages in all classes.

Fig. 4g, h and i show a classification comparison among LiDAR-derived biomass ($VVP_{int}$) and RF results, with and without an independent validation dataset. High vegetation complexity tended to be underestimated in the northwestern and eastern portions of the image, while low vegetation complexity tended to be overestimated in the southern part, especially when using a non-independent dataset. The percentages of error (Fig. 5c and f and Table 7) of more than one class were higher than in biomass ($VVP_{int}$) and height classifications. The greatest differences between RF using a non-independent or independent dataset were found in hardwood forest with overall accuracies of 34.57% and 54.22%, respectively. Conifer forest and mixed pixels presented similar levels of error that, in general terms, tended to underestimate the vegetation complexity. Within hardwood forest, vegetation complexity was clearly overestimated.

Based on mean decrease in RF overall accuracy ranking (Table 8), within the top-five important metrics for classifying vegetation complexity, we found metrics sensitive to photosynthetic pigments, such as NDVI and mNDVI705, water-sensitive metrics, such as, NDII, Wtr1EdgeWvl, or dry matter and principal component metrics, such as, SWIR2-PC1 and VIS-PC2. The optimized index using the first derivative transformed narrow-bands was one of the most important optical metrics for complexity based on Gini index ranking (Table 8). The top-five important optical metrics also included mNDVI705, EWT, PC1 and VIS-PC1. When the optimized indices were removed the most important metrics were similar except the Wtr2AbAr entered and optimized index dropped out of the top-five metrics.

**Table 8**
The top five optical metrics for prediction based on the decrease in root square mean error and Gini index for biomass ($VVP_{int}$), height and vegetation complexity.

| | $VVP_{int}$ | | Height | | Complexity | |
|---|---|---|---|---|---|---|
| | all_24v | all_22v | all_24v | all_22v | all_24v | all_22v |
| RMSE | DND-biomass | CAI | DND-height | CAI | mNDVI705 | CAI |
| | EWT | EWT | CAI | Wtr1EdgeWvl | CAI | VIS-PC2 |
| | CAI | PC2 | NDII | NDII | SWIR2-PC1 | mNDVI705 |
| | VIS-PC2 | GV | NDWI | NIR-PC2 | Wtr1EdgeWvl | Wtr1EdgeWvl |
| | EVI | EVI | Wtr1EdgeWvl | NDWI | NDII | NDVI705 |
| GINI | ND-biomass | SWIR2-PC1 | DND-height | Wtr2AbAr | DND-complex | mNDVI705 |
| | DND-biomass | mNDVI705 | mNDVI705 | mNDVI705 | mNDVI705 | EWT |
| | SWIR2-PC1 | CAI | Wtr2AbAr | VIS-PC1 | EWT | VIS-PC1 |
| | mNDVI705 | NDII | VIS-PC1 | EWT | VIS-PC1 | Wtr2AbAr |
| | NDVI | NDVI | NDVI705 | NDVI705 | PC1 | PC1 |

**Table 9**
Canopy structural type definitions based on the selected combination of the three classes (low, medium, high) for each LiDAR variable (biomass ($VVP_{int}$), height and complexity). The selection criterion is explained in detail in Section 4.4.2.

| Structural type | $VVP_{int}$ | Height | Complexity | % Study area |
|---|---|---|---|---|
| 1 | Low | Low | Low | 4.89 |
| | Medium | Low | Low | 8.84 |
| | Medium | Low | Medium | 3.15 |
| 2 | Medium | Medium | Medium | 7.27 |
| | Medium | Medium | High | 4.83 |
| | Medium | High | High | 5.55 |
| 3 | High | Low | Low | 17.20 |
| | High | Low | Medium | 3.11 |
| | High | Medium | Low | 4.96 |
| 4 | High | Medium | Medium | 8.92 |
| | High | Medium | High | 7.38 |
| | High | High | High | 14.64 |

*4.4.2. Canopy structural types*

Section 4.4.2.1 details the definition of canopy structural types in the study area based on combinations of the three LiDAR-derived structural variables (Table 9). Section 4.4.2.2 contains the results of Random Forests classification of these canopy structural types using optical metrics. Fig. 6 shows these RF classification results, and Table 10 summarizes the overall accuracy, class error and Kappa coefficient. Classification error is mapped in Fig. 7, and Tables 11 and 12 contain the top five optical metrics for canopy structural type classification.

*4.4.2.1. Canopy structural type definition.* The analysis of the 125 possible combinations from the five classes of LiDAR variables showed that 31 combinations were not present, 69 occupied less that the 1% of the study area each, and the percentage of the dominant classes was just 8%. In a second step, the LiDAR structure variables (i.e., biomass ($VVP_{int}$), height, and complexity) were classified in three classes in order to reduce the number of possible combinations from 125 to 27. Table 9 shows the description of the 12 LiDAR variable combinations that represented more than 1% of the study region. However, among the 12 classes, there were still some that were not well-represented in the study area. In addition, the RF results using these 12 classes showed high overlap between consecutive classes. Consequently, the 12 classes were grouped into 4 structural types with varying levels of biomass ($VVP_{int}$), height and complexity (Table 9).

*4.4.2.2. Canopy structural type discrimination.* When classifying these structural types, we observed no significant differences in accuracy when optimized indices were included or removed from the classifications (70.16% and 67.95 respectively) (Table 10). The overall accuracy did not decrease significantly with or without the independent dataset (66.23% and 70.16% with and without independent dataset considering all optical metrics and 63.53% and 67.95 without optimized indices). Vegetation stratification was not considered due to the low representativeness of the 4 classes within each scenario. The lowest and highest errors were found in classes 4 (high biomass ($VVP_{int}$), height and complexity) and 2 (medium biomass ($VVP_{int}$), height and complexity), respectively.

Fig. 6 shows a classification comparison among LiDAR-derived canopy structural types and RF results, with and without an independent validation dataset, and Fig. 7 shows the spatial distribution of class error, and the error percentage by vegetation type can be found in Table 11. In general terms, results were similar among classifications and there was no clear spatial trend in errors. Errors appeared mainly along the borders between classes. The lowest accuracy was found in mixed pixels, with a similar percentage using either the entire image or just half for training. The best overall accuracy was achieved for hardwood forest, where 83.88% and 77.84% of the pixels were correctly classified by RF with non-independent and independent datasets, respectively. The accuracy for conifer forest was also good, with 73.24% correctly classified pixels when the whole dataset was considered and 69.57% when an independent validation dataset was used.

Based on mean decrease in RF overall accuracy ranking (Table 12), within the top-five most important metrics for classifying structural types, we found the optimized indices using the first derivative transformed narrow-bands for biomass ($VVP_{int}$) and height together with the second PC of the VIS region of the spectrum and CAI with and without an independent dataset. The absorption water feature Wtr1EdgeWvl is replaced by the second principal component when an independent dataset is used. Some differences were also found in the top-five most important metrics in the classification without optimized indices with and without an independent dataset. The optimized indices using the first derivative transformed narrow-bands for height and complexity were found within the top-five metrics based on Gini index ranking (Table 12) together with some principal components considering different regions of the spectrum. When the optimized indices were removed, results from the Principal Component Analysis were also found within the most important metrics together with water absorption features.

## 5. Discussion

*5.1. Relationship between single optical metric and each structural*

Our initial analysis of single spectral bands showed that correlations differ, in magnitude and in direction, across the spectrum and by vegetation type for each structural variable. Biomass ($VVP_{int}$) was most similar between the two forest types, producing negative correlations in

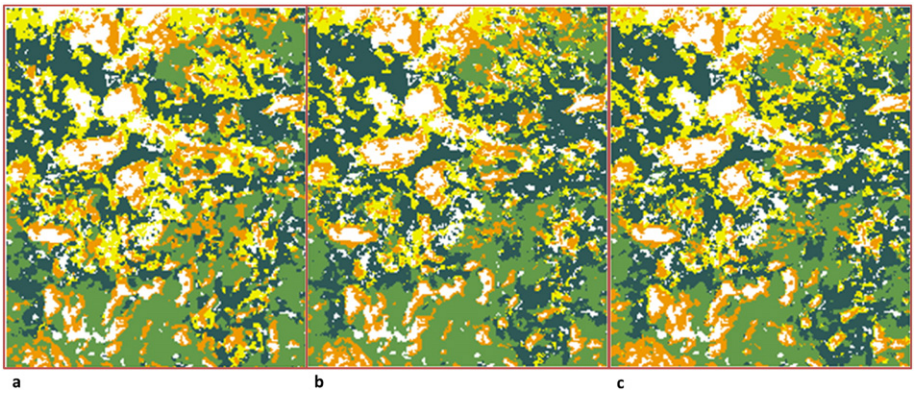


**Fig. 6.** Canopy structural types classified by LiDAR data (a), predicted with Random Forests classification considering the entire study region (b) and predicted from RF classification considering just half of the study region (c). Structural Type 1 (orange), Structural Type 2 (yellow), Structural Type 3 (light green) and Structural Type 4 (dark green) (structural types are defined in Table 9). No data (white).

**Table 10**
Canopy structural types classification percentage of class error, accuracy and Kappa coefficient considering all variables (24 variables: 24v) or removing the optimized indexes (22 variables: 22v) for the prediction and the extrapolation.

| | | | Type 1 | Type 2 | Type 3 | Type 4 | Accuracy | Kappa |
|---|---|---|---|---|---|---|---|---|
| All | 24v | Prediction | 36.13 | 51.02 | 24.44 | 18.74 | 70.16 | 0.59 |
| | | Extrapolation | 40.51 | 48.27 | 30.31 | 21.66 | 66.23 | 0.54 |
| | 22v | Prediction | 37.46 | 56.11 | 28.04 | 18.65 | 67.95 | 0.56 |
| | | Extrapolation | 40.65 | 52.37 | 36.74 | 22.25 | 63.53 | 0.50 |

visible and SWIR part of the spectrum and positive correlation in the NIR. Similar results were found in Marshall and Thenkabail (2014). The highest correlations in the visible and SWIR part of the spectrum for biomass ($VVP_{int}$) could be due to the fact that radiation absorption for the pigments in the visible, and for the lignin, cellulose and protein-N in the SWIR, are less affected by the scattering caused by the canopy structure that dominates the NIR (Kokaly Asner, Ollinger, Martin, et al., 2009). On the other hand, the estimation of height and vegetation complexity seems to be more challenging since only the conifer forest showed different correlations across the spectrum. The best band combination indices method allowed us to distinguish clear spectral regions highly related to biomass ($VVP_{int}$). However, based on height and complexity variables, the best optical band combination is around 1100 nm, coincident with a water absorption feature. Although the water-sensitive optical metrics do not show very high correlations with height and complexity, this analysis demonstrates the importance of these metrics to assess structure. The reason may be because taller trees in this forest often have more complex canopies and higher shade fractions. Our results agree with other researchers who have shown strong correlations between water-sensitive optical metrics and canopy height (Cohen & Spies, 1992; Swatantran, Dubayah, Roberts, Hofton, et al., 2011).

The relationships found in our study between optical indices/absorption features and vegetation structural metrics measured by LiDAR are consistent with the findings of other studies. For example, the shade fraction was correlated with biomass and height (i.e., $R^2 \geq 0.30$) (Hall, Shimabukuro, and Huemmrich, 1995; Swatantran, Dubayah, Roberts, Hofton, et al., 2011), but a stronger correlation was found with LiDAR-derived complexity in conifer forest, which agrees with Lu, Batistella, and Moran, (2005); Ogunjemiyo, Parker, and Roberts, (2005) and Sabol, Gillespie, Adams, Smith, et al., (2002). Similar results were found with the first principal component, used as a proxy for the albedo. This variable seems to be a valuable optical metric to explain biomass (Roberts, Ustin, Ogunjemiyo, Greenberg, et al., 2004) and height (Hyde, Dubayah, Walker, Blair, et al., 2006), but its highest correlations were found with vegetation complexity, specifically in conifer forests, it was the most strongly correlated optical metric. This result agrees with findings by Ogunjemiyo, Parker, and Roberts, (2005). Water-sensitive optical metrics (water indices, absorption features and PCA with SWIR region) had high correlations with biomass ($VVP_{int}$), especially in the conifer forests. The higher correlations found with water-sensitive optical metrics as opposed to pigment-sensitive metrics could be due to the saturation problem of the latter in conifer forests with high biomass density. This did not occur in the hardwood forest, possibly due to the lower biomass density associated with this type of forest. These results are in agreement with other authors in mixed forest types (Roberts, Ustin, Ogunjemiyo, Greenberg, et al., 2004; Swatantran, Dubayah, Roberts, Hofton, et al., 2011; Freitas, Mello, & Cruz, 2005).

We discovered several optimized indices that yielded higher correlations (i.e., $R^2 \geq 0.40$) with the three structural attributes, consistent with studies by Gong, Pu, Biging, and Larrieu (2003) that estimated LAI using normalized indexes derived from Hyperion data, or with studies by Peña, Brenning, and Sagredo (2012) that constructed empirical normalized indexes using Hyperion data to estimate canopy structure variables, such as, tree diameter at breast height or tree height using either original band or its first derivatives.

Optimized indices were especially effective in hardwood forests, resulting in stronger correlations with all three structural variables. Casas, Riaño, Ustin, Dennison, et al., (2014) also showed improvement using a best band combination technique among other techniques to retrieve water vegetation information. We found similar correlations between optimized indices and structural variables using both untransformed and transformed narrow-band indices, which unlike Peña, Brenning, and Sagredo, (2012), showed stronger correlations with the first derivative-based normalized indices than with untransformed band-based indices. Although the correlations are similar, the information provided by these two types of indices is different. Optimized indices using untransformed reflectance bands are based on reflectance values at different wavelengths, while first derivative band indices exploit the change between bands. However, these indices

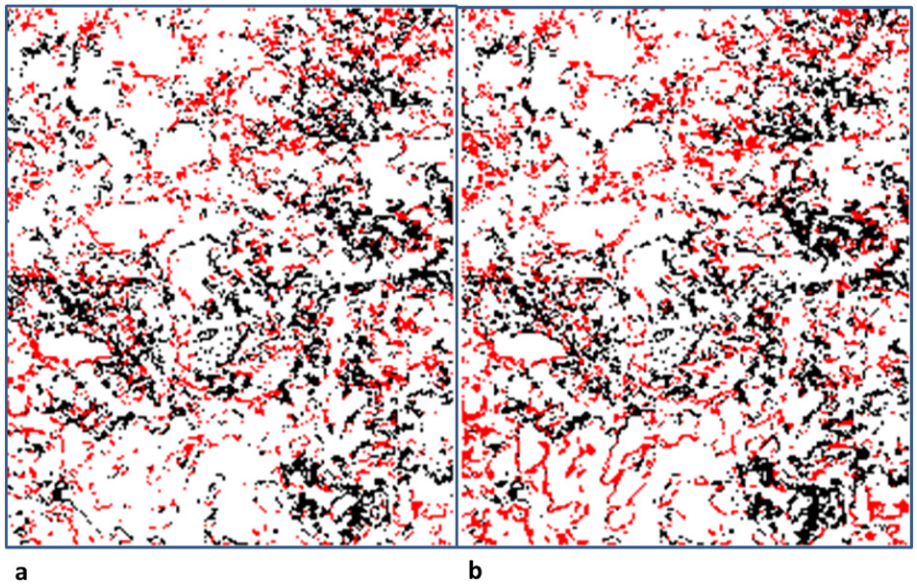


**Fig. 7.** Spatial distribution of classification error for canopy structural types (structural types are defined in Table 9) considering the entire study region (a) and considering just half of the study (b). Black: one or more classes above the correct estimate and red: one or more classes below the correct estimate.

**Table 11**
Percentage of error for the prediction and the extrapolation of canopy structural types classifications for conifer, hardwood and mixed forest.

| | Prediction | Extrapolation |
|---|---|---|
| Conifer | 73.24 | 69.57 |
| Hardwood | 83.88 | 77.84 |
| Mixed pixels | 66.64 | 64.86 |

maybe site and image dependent, and thus, further analysis should be done to assess their usefulness in other study regions. Still, they are promising given that other studies over a range of ecosystems have observed similar patterns in correlation between these wavelength regions and canopy structure.

In general terms, stronger relationships were found between single spectral bands and biomass ($VVP_{int}$) when stratifying by forest type. However, for the other two variables (i.e., height and complexity), increased correlations were only found for the conifer forest. On the other hand, the best band combination indices methods had higher correlations with the three structural variables in hardwood forests. These differences found among vegetation types make sense, due to the different architecture of conifer and hardwood forests. Although conifer and broad-leaf forests can have similar leaf area, the leaf distribution is significantly different. In conifers, needles are packed in dense shoots while hardwood trees have broader, flatter leaves. Crown shape is also different, with conifers having conical crowns whereas hardwood trees have spherical or ellipsoidal crowns. Needles produce self-shading within the shoot, so they tend to reflect less light than broad-leaf canopies (Ollinger, 2011). Needles cause more scatter than hardwood leaves because of lower transmittance and higher absorbance by needles than broad leaves (Roberts, Ustin, Ogunjemiyo, Greenberg, et al., 2004; Allen, Gausman, Richardson, & Thomas, 1969; Knapp & Carter, 1998). Needles often have higher dry matter content, thus absorption could be stronger in the SWIR. Furthermore, the recollision probability, which is sensitive to the canopy structural organization complexity, is larger for conifers than for hardwood (Knyazikhin, Schull, Stenberg, Mõttus, et al., 2013). These dissimilarities result in different relationships between canopy structure and optical properties between forest types.

### 5.2. Relationship between all optical metric and each structural variable

Although we found acceptable correlations (i.e., higher than 0.30) between LiDAR and optical metrics, they were not especially strong. This could be due to several sources of error. For instance the relationship was measured using Pearson's correlation, which assumes a linear relationship but this assumption may not always be true. In addition the information from just one set of optical metrics is not sufficient to describe all the spectral variability in the structure variable. The correlation values indicate whether a relationship exists between variables, but the low values mean that is necessary to have auxiliary information to explain and estimate the structural variable with confidence. Also the mix of understory vegetation, litter, and soil within the pixel may introduce bias in the empirical relationship between pixel reflectance and the canopy structure variable measured by LiDAR. Besides structural-optical relationships are age dependent, for instance, Ogunjemiyo, Parker, and Roberts, (2005) found that the relationship between canopy rugosity and albedo varied nonlinearly with tree age. All these factors make it difficult to estimate canopy structure solely from a single optical metric, especially within heterogeneous forests.

To address the first two points, we used the combinations of optical metrics to classify general patterns in each structural attribute using the non-parametric RF technique. We found the most important metrics were not necessarily the same ones identified in the literature, and theses varied for the three structural attributes studied. However, it is critical to note that optical metrics associated with greenness, water and dry matter, derived from the VIS, NIR and SWIR portions of the spectrum, were always present in the top five RF variable rankings. This underlines the importance of considering the entire spectrum for canopy structure assessment, and moves beyond the findings of previous studies by considering multiple optical metrics. The inclusion of several absorption features (rather than indices) in the top variable ranking also emphasizes the importance of using imaging spectroscopy data, because the absorption features occur in discrete portions of the spectrum and cannot be estimated with a few bands in broadband sensors (Ollinger, 2011). Although other studies have demonstrated the importance of the entire spectrum for other applications, such as species discrimination (Asner & Martin, 2009; Asner, Martin, & Suhaili, 2012; Clark & Roberts, 2012) or for estimating foliar traits from imaging spectroscopy data (Asner, Martin, Anderson, and Knapp, 2015), this issue has not been well-studied for canopy structure assessment. Additionally, it is important to highlight the importance of the SWIR spectral region and the absorption features, specifically those related to water content (Hunt, Ustin, and Riaño, 2013). Further research is needed to more deeply explore the SWIR spectral region and its role in canopy structure assessment. Lastly, stratifying by forest type did not improve RF accuracy, and Swatantran, Dubayah, Roberts, Hofton, et al. (2011) had similar result in similar forest. This indicates our RF approach can be applied to multiple forest types simultaneously.

### 5.3. Individual structural classes' assessment

Of the three structural metrics we considered, biomass is the most frequently estimated with optical remote sensing data. Canopy height and complexity have been less commonly estimated because their relationship with optical data is not well-understood (Hyyppä, Hyyppä, Inkinen, Engdahl, et al., 2000). Our result improves this understanding and demonstrates that optical metrics derived from AVIRIS data and used in the RF classifier can be used to discriminate different tree height and complexity classes and produce continuous maps of these classes with moderately high accuracy (from 60% to 89%). Though several researchers have found good correlations between canopy rugosity (i.e., complexity) and optical data using textural analysis (Ogunjemiyo, Parker, and Roberts, 2005; Sabol, Gillespie, Adams, Smith, 2002), we did not find these relationships in a preliminary assessment of several textural variables, including entropy and homogeneity. In future research, textural analysis could be addressed using higher spatial resolution data or taking into account other textural variables not considered here.

The greatest source of error in the maps was due to the discretization of continuous structural variables into classes. The consideration of

**Table 12**
The top five optical metrics for prediction based on the decrease in root square mean error and Gini index for canopy structural type classifications.

| RMSE | | | | GINI | | | |
|---|---|---|---|---|---|---|---|
| Prediction | | Extrapolation | | Prediction | | Extrapolation | |
| all_24v | all_22v | all_24v | all_22v | all_24v | all_22v | all_24v | all_22v |
| DND-biomass | VIS-PC2 | DND-biomass | VIS-PC2 | VIS-PC1 | SWIR2-PC1 | SWIR2-PC1 | SWIR2-PC1 |
| DND-height | CAI | DND-height | PC2 | SWIR2-PC1 | VIS-PC1 | VIS-PC1 | VIS-PC1 |
| VIS-PC2 | PC2 | VIS-PC2 | SWIR2-PC1 | DND-height | PC1 | DND-height | Wtr2AbAr |
| Wtr1EdgeWvl | Wtr2AbAr | PC2 | GV | DND-complex | SWIR1-PC1 | DND-complex | VIS-PC2 |
| CAI | Wtr1EdgeWvl | CAI | Wtr1EdgeWvl | PC1 | Wtr2AbAr | SWIR1-PC1 | PC1 |

small, overlapping values between adjacent classes greatly improved the classification accuracy in all the studied scenarios, but this highlights the fact that it can be challenging to define discrete classes within each structural attribute. We also demonstrated the portability of these models by assessing a spatially independent classification. While error increased slightly for biomass ($VVP_{int}$) and height, we observed a slight decrease in error for complexity. This may be because different classes were better represented in the dataset used to develop the classifier, resulting in higher accuracy for the independent dataset. As in all classification studies, the training data selected will affect the accuracy. There are established tradeoffs between realistic and equal class representation within training data. Some spatial patterns in the error were observed for height and complexity classifications. Heights tended to be underestimated within the pixels having high biomass and high height. These errors are similar to those found by others (Peña, Brenning, and Sagredo, 2012; Hudak, Lefsky, Cohen, and Berterretche, 2002). Complexity was generally underestimated in the areas of highest complexity areas and overestimated in the areas of lowest complexity. This is a typical issue when using regression analysis. From an ecological perspective, overestimation of complexity in homogeneous areas could be due to spectral variability more related to biochemical/biophysical conditions rather than structural properties; however further analysis is required to confirm this.

### 5.4. Defining and mapping canopy structural types

Because the three structure variables studied (biomass ($VVP_{int}$), height and complexity) covary in an ecologically meaningful way, differences in canopy structure may be better represented by taking into account all three variables simultaneously. This motivated our evaluation of structural type identification, and our results support this idea. We identified four major canopy structural types: The first structural type is represented by low biomass, height and complexity; these pixels are dominated by shrubs. The second structural type is characterized by medium biomass and medium height and complexity; this structural type represents mixed growth form pixels. The third structural type found shows high biomass, low-medium height and complexity; this type is mainly associated with pixels dominated by hardwood species. And the fourth structural type is characterized by high biomass, height and complexity; it is mainly associated with pixels dominated by conifer species.

Although land cover maps, such as CALVEG can be used to identify the species composition of an area, it is also vital to know the canopy structure and how it varies spatio-temporally within and across a mixed species forest. Our research provides a methodology to discriminate such vegetation structure scenarios, providing valuable information for decision making in several disciplines such as fire ecology, pest management or wildlife ecology among others. Easily interpretable maps of canopy structural types are critical for fire or pest behavior modeling since canopy continuity and homogeneity are a key factors in terms of fire spread and suppression, monitoring pest expansion or wildlife habitat (Van Wagner, 1977; Jia, Burke, Goetz, Kaufmann, et al., 2006; Lecina-Diaz, Alvarez, & Retana, 2014). Maps of structural complexity can contribute to increased accuracy of biomass and carbon sequestration. Furthermore, these structural types may facilitate fuel type discrimination and fuel availability, which are both determinants for fire risk estimation, the establishment of appropriate treatment to reduce fire risk (fuel thinning) or to improve carbon storage. Moreover, the approach detailed here can also be used to asses fire impact and severity (e.g., two fires with the same intensity may result in different fire severity depending on canopy structure), as well as to track forest recovery after a disturbance (Keeley, 2009; Ryan, 2002).

### 5.5. Future work

Future research may address some of the assumptions and limitations of our study. For instance, canopy reflectance is impacted by illumination and view angles (Hall, Shimabukuro, and Huemmrich, 1995; Sharma, Kajiwara, & Honda, 2013), atmospheric effects, or misregistration problems that might introduce bias and reduce classification accuracy. For this research, we used LiDAR variables as the reference data. LiDAR data may be affected by acquisition factors, such as altitude and scan angle (and their impact on point density), terrain, and vegetation characteristics, such as species composition or canopy closure (Morsdorf, Frey, Meier, Itten, et al., 2008; Næsset, 2009; Ørka, Næsset, & Bollandsås, 2010). These potential sources of LiDAR inaccuracy may influence AVIRIS canopy structure classification accuracy. However, even considering those LiDAR uncertainties, LiDAR has been proven as an accurate tool for characterizing vegetation structure (Hilker, Wulder, and Coops, 2008; Hopkinson & Chasmer, 2009; Tang, Brolly, Zhao, Strahler, et al., 2014). Existing spaceborne LiDAR systems, such as GLAS, and future missions, such as GEDI, will continue to expand the use of LiDAR for canopy structure assessment. Although optical data will never yield estimates of structural attributes as accurately as LiDAR or radar, results from this research have demonstrated the usefulness of imaging spectroscopy data to obtain this information that will enable extending capability to map structural types into areas where LiDAR data are unavailable. Our future research will more specifically focus on developing models to estimate the structural variables extending the approach demonstrated to a broader study area to test its applicability to other forest types with differing environmental conditions. This is critical given that imaging spectroscopy data will become more widely available when several spaceborne missions are launched, such as the German Aerospace Center (DLR) EnMap mission (http://www.enmap.org/, Kaufmann, Segl, Kuester, Rogass, et al., 2012), the Italian Space Agency (ASI) PRISMA mission (http://www.asi.it/en/activity/observation-earth/prisma, Pignatti, Acito, Amato, Casa, et al., 2012), the Japanese Space Agency HySui mission (Ohgi, Iwasaki, Kawashima, & Inada, 2010; Staenz & Held, 2012), the Indian Space Agency planned hyperspectral imager (Kumar, Thapa, and Kuriakose (2015) on the geostationary imaging satellite (GISAT)), and NASA's planned Hyperspectral Infrared Imager (HyspIRI) mission (National Research Council (NRC), 2007; Lee et al., 2015). These sensors may provide frequent, global, high spectral resolution information, offering a wider opportunity for its use in canopy structure assessment.

## 6. Conclusions

Remote sensing assessments of canopy structure have become critical to many research and management agendas, including biomass estimation, carbon cycle assessment and climate modeling. While it has been well-established that LiDAR can be used to very accurately estimate canopy structure, our research evaluated an alternative approach, using imaging spectroscopy data. Though these data may not be as accurate in direct prediction of canopy structure, our results demonstrate these data can characterize the overall gradients of canopy biomass ($VVP_{int}$), height, and complexity, and to map meaningful canopy structural types using self-learning classifiers like RF. Our method is portable, producing reliable results when applied to an area not used in model-building. Furthermore, this approach moves beyond simply estimating structural attributes to analyzing the combinations of structural characteristics that distinguish different canopy structure types in a heterogeneous forested landscape. We stress both the use of multiple optical metrics and the importance of the entire spectrum when classifying structural types. Our approach can be used in areas where LiDAR data are unavailable and is robust to differences in vegetation type. The growing availability of imaging spectroscopy data will enable more widespread use of these data for future canopy structure assessment.

## Acknowledgements

This research was conducted within the framework of the HyspIRI Planning Mission (NASA Grant # NNX12AP87G). Mariano Garcia is supported by the Marie Curie IOF (ForeStMap — 3D Forest Structure

Monitoring and Mapping, Project Reference: 629376). The contents on this paper reflect only the authors' views and not the views of the European Commission. I would like to thank NEON for providing the LiDAR data and the NASA JPL AVIRIS team for collecting and pre-processing the hyperspectral data. I would also thank the anonymous reviewers for their help in improving the manuscript.

## Appendix A

Tables A-1 and A-2 show a summary of the LiDAR variables and optical metrics used in this research.

**Table A.1**
Summary of the optical metric variables used.

| Optical metrics | |
|---|---|
| CAI | Cellulose Absorption Index |
| EVI | Enhanced Vegetation Index |
| EWT | Equivalent Water Thickness |
| GV | Green Vegetation |
| mNDVI705 | Modified NDVI using 705 nm spectral band |
| NDII | Normalized Difference Infrared Index |
| NIR-PC1 | First Principal Component of NIR region |
| NIR-PC2 | Second Principal Component of NIR region |
| NDVI | Normalized Difference Vegetation Index |
| NDVI705 | NDVI using 705 nm spectral band |
| NDWI | Normalized Difference Water Index |
| PC1 | First Principal Component |
| PC2 | Second Principal Component |
| SHADE | Shade fraction |
| SWIR1-PC1 | First Principal Component of SWIR region |
| SWIR2-PC1 | Second Principal Component of SWIR region |
| VIS-PC1 | First Principal Component of VIS region |
| VIS-PC2 | Second Principal Component of VIS region |
| Wtr1AbAr | NIR1 water absorption feature area |
| Wtr1EdgeWvl | NIR1 water absorption feature edge wavelength |
| Wtr2AbAr | NIR2 water absorption feature area |
| Wtr1EdgeMag | NIR1 water absorption feature edge magnitude |

**Table A.2**
Summary of the LiDAR variables used.

| LiDAR variables | |
|---|---|
| CHM_std | Canopy height model standard deviation |
| FC | Fractional cover |
| FC_1ret | Fractional cover estimated from first returns |
| H_max | Maximum height |
| H_mean | Mean height |
| H_median | Median height |
| H_std | Standard deviation of height |
| LAI | Leaf area index |
| VVP | Vegetation Vertical Profile |

## References

Ahmed, O. S., Franklin, S. E., & Wulder, A. (2013). Interpretation of forest disturbance using a time series of Landsat imagery and canopy structure from airborne lidar. *Canadian Journal of Remote Sensing*, *39*, 521–542.

Ahmed, O. S., Franklin, S. E., Wulder, M. A., & White, J. C. (2015). Characterizing stand-level forest canopy cover and height using Landsat time series, samples of airborne LiDAR, and the Random Forest algorithm. *ISPRS Journal of Photogrammetry and Remote Sensing*, *101*, 89–101.

Allen, W. A., & Richardson, A. J. (1968). Interaction of light with a plant canopy. *Journal of the Optical Society of America*, *58*, 1023–1028.

Allen, W. A., Gausman, H. W., Richardson, A. J., & Thomas, J. R. (1969). Interaction of isotropic light with a compact plant leaf. *Journal of the Optical Society of America*, *59*, 1376–1379.

Asner, G. P. (1998). Biophysical and biochemical sources of variability in canopy reflectance. *Remote Sensing of Environment*, *64*, 234–253.

Asner, G. P., & Martin, R. E. (2009). Airborne spectranomics: Mapping canopy chemical and taxonomic diversity in tropical forests. *Frontiers in Ecology and the Environment*, *7*, 269–276.

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. *Ecological Applications*, *19*, 236–253.

Asner, G., Martin, R., & Suhaili, A. (2012). Sources of canopy chemical and spectral diversity in lowland Bornean Forest. *Ecosystems*, *15*, 504–517.

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

Blackard, J. A., Finco, M. V., Helmer, E. H., Holden, G. R., Hoppus, M. L., Jacobs, D. M., et al. (2008). Mapping U.S. forest biomass using nationwide forest inventory data and moderate resolution information. *Remote Sensing of Environment*, *112*, 1658–1677.

Blackburn, G. A. (1998). Quantifying chlorophylls and caroteniods at leaf and canopy scales: An evaluation of some hyperspectral approaches. *Remote Sensing of Environment*, *66*, 273–285.

Bouvier, M., Durrieu, S., Fournier, R. A., & Renaud, J. (2015). Generalizing predictive models of forest inventory attributes using an area-based approach with airborne LiDAR data. *Remote Sensing of Environment*, *156*, 322–334.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32.

Casas, A., Riaño, D., Ustin, S. L., Dennison, P., & Salas, J. (2014). Estimation of water-related biochemical and biophysical vegetation properties using multitemporal airborne hyperspectral data and its comparison to MODIS spectral response. *Remote Sensing of Environment*, *148*, 28–41.

Ceccato, P., Flasse, S., Tarantola, S., Jacquemoud, S., & Grégoire, J. (2001). Detecting vegetation leaf water content using reflectance in the optical domain. *Remote Sensing of Environment*, *77*, 22–33.

Chan, J. C., & Paelinckx, D. (2008). Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sensing of Environment*, *112*, 2999–3011.

Cheng, Y., Zarco-Tejada, P. J., Riaño, D., Rueda, C. A., & Ustin, S. L. (2006). Estimating vegetation water content with hyperspectral data for different canopy scenarios: Relationships between AVIRIS and MODIS indices. *Remote Sensing of Environment*, *105*, 354–366.

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 Sensing*, *4*, 1820.

Clevers, J. G. P. W., & Kooistra, L. (2012). Using hyperspectral remote sensing data for retrieving canopy chlorophyll and nitrogen content. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *5*, 574–583.

Cohen, W. B., & Spies, T. A. (1992). Estimating structural attributes of Douglas-fir/western hemlock forest stands from landsat and SPOT imagery. *Remote Sensing of Environment*, *41*, 1–17.

Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, *37*, 35–46.

Coops, N. C., Smith, M., Martin, M. E., & Ollinger, S. V. (2003). Prediction of eucalypt foliage nitrogen content from satellite-derived hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, *41*, 1338–1346.

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., et al. (2007). Random Forest for classification in ecology. *Ecology*, *88*, 2783–2792.

Datt, B. (1999). A new reflectance index for remote sensing of chlorophyll content in higher plants: Tests using eucalyptus leaves. *Journal of Plant Physiology*, *154*, 30–36.

Daughtry, C. S. T. (2001). Discriminating crop residues from soil by shortwave infrared reflectance. *Agronomy Journal*, *93*, 125–131.

Dennison, P. E., & Roberts, D. A. (2003). Endmember selection for multiple endmember spectral mixture analysis using endmember average RSME. *Remote Sensing of Environment*, *87*, 123–135.

Dennison, P. E., Halligan, K. Q., & Roberts, D. A. (2004). A comparison of error metrics and constraints for multiple endmember spectral mixture analysis and spectral angle mapper. *Remote Sensing of Environment*, *93*, 359–367.

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

Duncansona, L. I., Niemanna, K. O., & Wulderb, M. A. (2010). Estimating forest canopy height and terrain relief from GLAS waveform metrics. *Remote Sensing of Environment*, *114*, 138–154.

Falkowski, M. J., Evans, J. S., Martinuzzi, S., Gessler, P. E., & Hudak, A. T. (2009). Characterizing forest succession with LiDAR data: An evaluation for the inland northwest, USA. *Remote Sensing of Environment*, *113*, 946–956.

Féret, J., François, C., Gitelson, A., Asner, G. P., Barry, K. M., Panigada, C., et al. (2011). Optimizing spectral indices and chemometric analysis of leaf chemical properties using radiative transfer modeling. *Remote Sensing of Environment*, *115*, 2742–2750.

Ferwerda, J. G., Skidmore, A. K., & Mutanga, O. (2005). Nitrogen detection with hyperspectral normalized ratio indices across multiple plant species. *International Journal of Remote Sensing*, *26*, 4083–4095.

Franklin, S. E., Hall, R. J., Smith, L., & Gerylo, G. R. (2003). Discrimination of conifer height, age and crown closure classes using Landsat-5 TM imagery in the Canadian Northwest Territories. *International Journal of Remote Sensing*, *24*, 1823–1834.

Freitas, S. R., Mello, M. C. S., & Cruz, C. B. M. (2005). Relationships between forest structure and vegetation indices in Atlantic Rainforest. *Forest Ecology and Management*, *218*, 353–362.

Gao, B. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, *58*, 257–266.

García, M., Riaño, D., Chuvieco, E., & Danson, F. M. (2010). Estimating biomass carbon stocks for a Mediterranean forest in Spain using height and intensity LiDAR data. *Remote Sensing of Environment*, *114*, 816–830.

García, M., Gajardo, J., Riaño, D., Zhao, K., Martín, P., & Ustin, S. (2015). Canopy clumping appraisal using terrestrial and airborne laser scanning. *Remote Sensing of Environment*, *161*, 78–88.

Gates, D. M., Keegan, H. J., Schleter, J. C., & Weidner, V. R. (1965). Spectral properties of plants. *Applied Optics*, *4*, 11–20.

Gausman, H. W. (1984). Evaluation of factors causing reflectance differences between sun and shade leaves. *Remote Sensing of Environment*, *15*, 177–181.

Gausman, H. W., & Allen, W. A. (1973). Optical parameters of leaves of 30 plant species. *Plant Physiology*, *52*, 57–62.

Gitelson, A., & Merzlyak, M. N. (1994). Spectral reflectance changes associated with autumn senescence of *Aesculus hippocastanum* L. and *Acer platanoides* L. Leaves. Spectral features and relation to chlorophyll estimation. *Journal of Plant Physiology*, *143*, 286–292.

Gong, P., Pu, R., Biging, G. S., & Larrieu, M. R. (2003). Estimation of forest leaf area index using vegetation indices derived from Hyperion hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, *41*, 1355–1362.

Green, R. O., Eastwood, M. L., Sarture, C. M., Chrien, T. G., Aronsson, M., Chippendale, B. J., ... Williams, O. (1998). Imaging spectroscopy and the Airborne Visible Infrared Imaging Spectrometer (AVIRIS). *Remote Sensing of Environment*, *65*, 227–248.

Gu, D., & Gillespie, A. (1998). Topographic normalization of Landsat TM images of Forest based on subpixel sun–canopy–sensor geometry. *Remote Sensing of Environment*, *64*, 166–175.

Hall, F. G., Shimabukuro, Y. E., & Huemmrich, K. F. (1995). Remote sensing of forest biophysical structure using mixture decomposition and geometric reflectance models. *Ecological Applications*, *5*, 993–1013.

Hansen, P. M., & Schjoerring, J. K. (2003). Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. *Remote Sensing of Environment*, *86*, 542–553.

Hansen, M. C., DeFries, R. S., Townshend, J. R. G., Sohlberg, R., Dimiceli, C., & Carroll, M. (2002). Towards an operational MODIS continuous field of percent tree cover algorithm: Examples using AVHRR and MODIS data. *Remote Sensing of Environment*, *83*, 303–319.

Harding, D. J., Lefsky, M. A., Parker, G. G., & Blair, J. B. (2001). Laser altimeter canopy height profiles: Methods and validation for closed-canopy, broadleaf forests. *Remote Sensing of Environment*, *76*, 283–297.

Hilbert, C., & Schmullius, C. (2012). Influence of surface topography on ICESat/GLAS Forest height estimation and waveform shape. *Remote Sensing*, *4*, 2210–2235.

Hilker, T., Wulder, M. A., & Coops, N. C. (2008). Update of forest inventory data with LiDAR and high spatial resolution satellite imagery. *Canadian Journal of Remote Sensing*, *34*, 5–12.

Hopkinson, C., & Chasmer, L. (2009). Testing LiDAR models of fractional cover across multiple forest ecozones. *Remote Sensing of Environment*, *113*, 275–288.

Hudak, A. T., Lefsky, M. A., Cohen, W. B., & Berterretche, M. (2002). Integration of LiDAR and Landsat ETM+ data for estimating and mapping forest canopy height. *Remote Sensing of Environment*, *82*, 397–416.

Huete, A., Justice, C., & Liu, H. (1994). Development of vegetation and soil indices for MODIS-EOS. *Remote Sensing of Environment*, *49*, 224–234.

Huete, A. R., Liu, H. Q., Batchily, K., & van Leeuwen, W. (1997). A comparison of vegetation indices over a global set of TM images for EOS-MODIS. *Remote Sensing of Environment*, *59*, 440–451.

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, *83*, 195–213.

Hunt, E. R., Jr., & Rock, B. N. (1989). Detection of changes in leaf water content using near- and middle-infrared reflectances. *Remote Sensing of Environment*, *30*, 43–54.

Hunt, E. R., Jr., Ustin, S., & Riaño, D. (2013). Remote sensing of leaf, canopy, and vegetation water contents for satellite environmental data records. *Satellite-based applications on climate change* (pp. 335–357). New York: Springer.

Hyde, P., Dubayah, R., Walker, W., Blair, J. B., Hofton, M., & Hunsaker, C. (2006). Mapping forest structure for wildlife habitat analysis using multi-sensor (LiDAR, SAR/InSAR, ETM+, Quickbird) synergy. *Remote Sensing of Environment*, *102*, 63–73.

Hyyppä, J., Hyyppä, H., Inkinen, M., Engdahl, M., Linko, S., & Zhu, Y. (2000). Accuracy comparison of various remote sensing data sources in the retrieval of forest stand attributes. *Forest Ecology and Management*, *128*, 109–120.

Jacquemoud, S., Bacour, C., Poilvé, H., & Frangi, J. (2000). Comparison of four radiative transfer models to simulate plant canopies reflectance: Direct and inverse mode. *Remote Sensing of Environment*, *74*, 471–481.

Jacquemoud, S., Verhoef, W., Baret, F., Bacour, C., Zarco-Tejada, P. J., Asner, G. P., et al. (2009). PROSPECT + SAIL models: A review of use for vegetation characterization. *Remote Sensing of Environment*, *113*(Suppl. 1), S56–S66.

Jia, G. J., Burke, I. C., Goetz, A. F. H., Kaufmann, M. R., & Kindel, B. C. (2006). Assessing spatial patterns of forest fuel using AVIRIS data. *Remote Sensing of Environment*, *102*, 318–327.

Kantola, T., Vastaranta, M., Yu, X., Lyytikainen-Saarenmaa, P., Holopainen, M., Talvitie, M., et al. (2010). Classification of defoliated trees using tree-level airborne laser scanning data combined with aerial images. *Remote Sensing*, *2*, 2665.

Kaufmann, H., Segl, K., Kuester, T., Rogass, C., Hofer, S., Mueller, A., & Chlebek, C. (2012). The Environmental Mapping and Analysis Programme (EnMAP) — Present status of preparatory phase. *Proceedings of the international geoscience and remote sensing symposium (IAGARSS'12)*. Germany: Munich (4 pp.).

Kauth, R. J., & Thomas, G. S. (1976). The tasselled cap — A graphic description of the spectral-temporal development of agricultural crops as seen by LANDSAT. *Proceedings of the symposium on machine processing of remotely sensed data*. Indiana: Purdue University of West Lafayette (4B-41-4B-51).

Keeley, J. E. (2009). Fire intensity, fire severity and burn severity: A brief review and suggested usage. *International Journal of Wildland Fire*, *18*, 116–126.

Knapp, A. K., & Carter, G. A. (1998). Variability in Leaf Optical properties among 26 species from a broad range of habitats. *American Journal of Botany*, *85*, 940–946.

Knyazikhin, Y., Schull, M. A., Stenberg, P., Mõttus, M., Rautiainen, M., Yang, Y., et al. (2013). Hyperspectral remote sensing of foliar nitrogen content. *Proceedings of the National Academy of Sciences*, *110*, E185–E192.

Kokaly, R. F., Asner, G. P., Ollinger, S. V., Martin, M. E., & Wessman, C. A. (2009). Characterizing canopy biochemistry from imaging spectroscopy and its application to ecosystem studies. *Remote Sensing of Environment*, *113*(Suppl. 1), S78–S91.

Kumar, K. A., Thapa, N., & Kuriakose, S. A. (2015). Advances in spaceborne hyperspectral imaging systems. *Current Science, Special issue: Hyperspectral Remote Sensing*, *108*, 826–832.

Lecina-Diaz, J., Alvarez, A., & Retana, J. (2014). Extreme fire severity patterns in topographic, convective and wind-driven historical wildfires of Mediterranean pine forests. *PLoS ONE*, *9*(1), e85127.

Lee, C. M., Cable, M. L., Hook, S. J., Greena, R. O., Ustin, S. L., Mandl, D. J., & Middleton, E. M. (2015). An introduction to the NASA Hyperspectral InfraRed Imager (HyspIRI) mission and preparatory activities. *Remote Sensing of Environment*, *167*, 6–19.

Lefsky, M. A., Cohen, W. B., & Spies, T. A. (2001). An evaluation of alternate remote sensing products for forest inventory, monitoring, and mapping of Douglas-fir forests in western Oregon. *Canadian Journal of Forest Research*, *31*, 78–87.

Liaw, A., & Wiener, M. (2002). Classification and regression by Random forest. *R News*, *2*, 18–22.

Lovell, J. L., Jupp, D. L. B., Culvenor, D. S., & Coops, N. C. (2003). Using airborne and ground-based ranging LiDAR to measure canopy structure in Australian forests. *Canadian Journal of Remote Sensing*, *29*, 607–622.

Lu, D., Batistella, M., & Moran, E. (2005). Satellite estimation of aboveground biomass and impacts of forest stand structure. *Photogrammetric Engineering and Remote Sensing*, *71*, 967–974.

Marshall, M., & Thenkabail, P. (2014). Biomass modeling of four leading world crops using hyperspectral Narrowbands in support of HyspIRI mission. *Photogrammetric Engineering and Remote Sensing*, *80*, 757–772.

Moisen, G. G., & Frescino, T. S. (2002). Comparing five modelling techniques for predicting forest characteristics. *Ecological Modelling*, *157*, 209–225.

Morsdorf, F., Kotz, B., Meier, E., Itten, K. I., & Allgower, B. (2006). Estimation of LAI and fractional cover from small footprint airborne laser scanning data based on gap fraction. *Remote Sensing of Environment*, *104*, 50–61.

Morsdorf, F., Frey, O., Meier, E., Itten, K. I., & Allgower, B. (2008). Assessment of the influence of flying altitude and scan angle on biophysical vegetation products derived from airborne laser scanning. *International Journal of Remote Sensing*, *29*, 1387–1406.

Næsset, E. (2009). Effects of different sensors, flying altitudes, and pulse repetition frequencies on forest canopy metrics and biophysical stand properties derived from small-footprint airborne laser data. *Remote Sensing of Environment*, *113*, 148–159.

Næsset, E., Gobakken, T., Bollandsås, O. M., Gregoire, T. G., Nelson, R., & Ståhl, G. (2013). Comparison of precision of biomass estimates in regional field sample surveys and airborne LiDAR-assisted surveys in Hedmark County, Norway. *Remote Sensing of Environment*, *130*, 108–120.

Naidoo, L., Cho, M. A., Mathieu, R., & Asner, G. (2012). Classification of savanna tree species, in the Greater Kruger National Park region, by integrating hyperspectral and LiDAR data in a Random Forest data mining environment. *ISPRS Journal of Photogrammetry and Remote Sensing*, *69*, 167–179.

National Research Council (NRC) (2007). *Earth science and applications from space: National imperatives for the next decade and beyond*. Washington, DC, USA: National Academy Press.

Ogunjemiyo, S., Parker, G., & Roberts, D. (2005). Reflections in bumpy terrain: Implications of canopy surface variations for the radiation balance of vegetation. *IEEE Geoscience and Remote Sensing Letters*, *2*, 90–93.

Ohgi, N., Iwasaki, A., Kawashima, T., & Inada, H. (2010). Japanese hyper-multi spectral mission. *Geoscience and remote sensing symposium (IGARSS)*. IEEE.

Olden, J. D., Lawler, J. J., & Poff, N. L. (2008). Machine learning methods without tears: A primer for ecologists. *The Quarterly Review of Biology*, *83*, 171–193.

Ollinger, S. V. (2011). Sources of variability in canopy reflectance and the convergent properties of plants. *New Phytologist*, *189*, 375–394.

Omari, K., White, H. P., Staenz, K., & King, D. J. (2013). Retrieval of Forest canopy parameters by inversion of the PROFLAIR leaf-canopy reflectance model using the LUT approach. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *6*, 715–723.

Ørka, H. O., Næsset, E., & Bollandsås, O. M. (2010). Effects of different sensors and leaf-on and leaf-off canopy conditions on echo distributions and individual tree properties derived from airborne laser scanning. *Remote Sensing of Environment*, *114*, 1445–1461.

Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, *26*, 217–222.

Peña, M. A., Brenning, A., & Sagredo, A. (2012). Constructing satellite-derived hyperspectral indices sensitive to canopy structure variables of a Cordilleran cypress (*Austrocedrus chilensis*) forest. *ISPRS Journal of Photogrammetry and Remote Sensing*, *74*, 1–10.

Peterson, D. L., Aber, J. D., Matson, P. A., Card, D. H., Swanberg, N., Wessman, C., et al. (1988). Remote sensing of forest canopy and leaf biochemical contents. *Remote Sensing of Environment*, *24*, 85–108.

Pflugmacher, D., Cohen, W. B., & Kennedy, R. E. (2012). Using Landsat-derived disturbance history (1972–2010) to predict current forest structure. *Remote Sensing of Environment*, *122*, 146–165.

Pierce, A. D., Farris, C. A., & Taylor, A. H. (2012). Use of random forests for modeling and mapping forest canopy fuels for fire behavior analysis in Lassen Volcanic National Park, California, USA. *Forest Ecology and Management*, *279*, 77–89.

Pignatti, S., Acito, N., Amato, U., Casa, R., de Bonis, R., Diani, M., ... Cuomo, V. (2012). Development of algorithms and products for supporting the Italian hyperspectral PRISMA mission: The SAP4PRISMA project. *Proceedings of the International Geoscience and Remote sensing Symposium (IGARSS'12), Munich, Germany* (4 pp.).

Popescu, S. C., Zhao, K., Neuenschwander, A., & Lin, C. (2011). Satellite lidar vs. small footprint airborne LiDAR: Comparing the accuracy of aboveground biomass estimates and forest structure metrics at footprint level. *Remote Sensing of Environment*, *115*, 2786–2797.

Powell, S. L., Cohen, W. B., Healey, S. P., Kennedy, R. E., Moisen, G. G., Pierce, K. B., et al. (2010). Quantification of live aboveground forest biomass dynamics with Landsat time-series and field inventory data: A comparison of empirical modeling approaches. *Remote Sensing of Environment*, *114*, 1053–1068.

R Development Core Team (2014). *R: A language and environment for statistical computing, Vienna, Austria.* (http://www.R-project.org).

Riaño, D., Valladares, F., Condes, S., & Chuvieco, E. (2004). Estimation of leaf area index and covered ground from airborne laser scanner (LiDAR) in two contrasting forests. *Agricultural and Forest Meteorology*, *124*, 269–275.

Richards, J. A. (1999). *Remote sensing digital image analysis: An introduction.* Berlin, Germany: Springer-Verlag, 240.

Roberts, D. A., Green, R. O., & Adams, J. B. (1997). Temporal and spatial patterns in vegetation and atmospheric properties from AVIRIS. *Remote Sensing of Environment*, *62*, 223–240.

Roberts, D. A., Gardner, M., Church, R., Ustin, S., Scheer, G., & Green, R. O. (1998). Mapping chaparral in the Santa Monica Mountains using multiple endmember spectral mixture models. *Remote Sensing of Environment*, *65*, 267–279.

Roberts, D. A., Dennison, P. E., Gardner, M. E., Hetzel, Y., Ustin, S. L., & Lee, C. T. (2003). Evaluation of the potential of Hyperion for fire danger assessment by comparison to the airborne visible/infrared imaging spectrometer. *IEEE Transactions on Geoscience and Remote Sensing*, *41*, 1297–1310.

Roberts, D. A., Ustin, S. L., Ogunjemiyo, S., Greenberg, J., Dobrowski, S. Z., Chen, J., et al. (2004). Spectral and structural measures of northwest forest vegetation at leaf to landscape scales. *Ecosystems*, *7*, 545–562.

Roberts, D., Halligan, K., & Dennison, P. (2007). *ViperTools.*http://www.vipertools.org

Rosema, A., Verhoef, W., Noorbergen, H., & Borgesius, J. J. (1992). A new forest light interaction model in support of forest modeling. *Remote Sensing of Environment*, *42*, 23–41.

Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1973). Monitoring vegetation systems in the Great Plains with ERTS. *Third ERTS Symposium, NASA*, *1*. (pp. 309–317).

Ryan, K. C. (2002). Dynamic interactions between forest structure and fire behavior in boreal ecosystems. *Silva Fennica*, *36*, 13–39.

Sabol, D. E., Jr., Gillespie, A. R., Adams, J. B., Smith, M. O., & Tucker, C. J. (2002). Structural stage in Pacific northwest forests estimated using simple mixing models of multispectral images. *Remote Sensing of Environment*, *80*, 1–16.

Sánchez-Azofeifa, G. A., Castro, K., Wright, S. J., Gamon, J., Kalacska, M., Rivard, B., et al. (2009). Differences in leaf traits, leaf internal structure, and spectral reflectance between two communities of lianas and trees: Implications for remote sensing in tropical environments. *Remote Sensing of Environment*, *113*, 2076–2088.

Sharma, R. C., Kajiwara, K., & Honda, Y. (2013). Estimation of forest canopy structural parameters using kernel-driven bi-directional reflectance model based multi-angular vegetation indices. *ISPRS Journal of Photogrammetry and Remote Sensing*, *78*, 50–57.

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 Sensing of Environment*, *81*, 337–354.

Small, C. (2003). High spatial resolution spectral mixture analysis of urban reflectance. *Remote Sensing of Environment*, *88*, 170–186.

Soenen, S. A., Peddle, D. R., & Coburn, C. A. (2005). A modified sun-canopy-sensor topographic correction in forested terrain. *IEEE Transactions on Geoscience and Remote Sensing*, *43*, 2148–2159.

Staenz, K., & Held, A. (2012). Summary of current and future terrestrial civilian hyperspectral spaceborne systems. *International GeoScience and remote sensing symposium (IGARSS), IEEE, 22–27 July 2012, Munich, Germany #2607* (4 pp.).

Suna, G., Ransonb, K. J., Kimesb, D. S., Blairb, J. B., & Kovacsc, K. (2008). Forest vertical structure from GLAS: An evaluation using LVIS and SRTM data. *Remote Sensing of Environment*, *112*, 107–117.

Swatantran, A., Dubayah, R., Roberts, D., Hofton, M., & Blair, J. B. (2011). Mapping biomass and stress in the Sierra Nevada using LiDAR and hyperspectral data fusion. *Remote Sensing of Environment*, *115*, 2917–2930.

Tang, H., Brolly, M., Zhao, F., Strahler, A. H., Schaaf, C. L., Ganguly, S., et al. (2014). Deriving and validating Leaf Area Index (LAI) at multiple spatial scales through LiDAR remote sensing: A case study in Sierra National Forest, CA. *Remote Sensing of Environment*, *143*, 131–141.

Townsend, P. A., Foster, J. R., Chastain, R. A., Jr., & Currie, W. S. (2003). Application of imaging spectroscopy to mapping canopy nitrogen in the forests of the central Appalachian Mountains using Hyperion and AVIRIS. *IEEE Transactions on Geoscience and Remote Sensing*, *41*, 1347–1354.

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*, 523–534.

Ustin, S. L., Gitelson, A. A., Jacquemoud, S., Schaepman, M., Asner, G. P., Gamon, J. A., et al. (2009). Retrieval of foliar information about plant pigment systems from high resolution spectroscopy. *Remote Sensing of Environment*, *113*(Suppl. 1), S67–S77.

Ustin, S. L., Riaño, D., & Hunt, E. R. (2012). Estimating canopy water content from spectroscopy. *Israel Journal of Plant Sciences*, *60*, 9–23.

Van Wagner, C. E. (1977). Conditions for the start and spread of crown fire. *Canadian Journal of Forest Research*, *7*, 23–24.

Verhoef, W. (1984). Light scattering by leaf layers with application to canopy reflectance modeling: The SAIL model. *Remote Sensing of Environment*, *16*, 125–141.

White, J. C., Gómez, C., Wulder, M. A., & Coops, N. C. (2010). Characterizing temperate forest structural and spectral diversity with Hyperion EO-1 data. *Remote Sensing of Environment*, *114*, 1576–1589.

Wulder, M. A., & Seemann, D. (2003). Forest inventory height update through the integration of LiDAR data with segmented Landsat imagery. *Canadian Journal of Remote Sensing*, *29*, 536–543.

Zarco-Tejada, P. J., Rueda, C. A., & Ustin, S. L. (2003). Water content estimation in vegetation with MODIS reflectance data and model inversion methods. *Remote Sensing of Environment*, *85*, 109–124.

Zhao, D., Reddy, K. R., Kakani, V. G., Read, J. J., & Koti, S. (2005). Selection of optimum reflectance ratios for estimating leaf nitrogen and chlorophyll concentrations of field-grown cotton. *Agronomy Journal*, *97*, 89–98.

Zhao, K. G., Popescu, S., & Nelson, R. (2009). LiDAR remote sensing of Forest biomass: A scale-invariant estimation approach using airborne lasers. *Remote Sensing of Environment*, *113*, 182–196.