Int J Appl Earth Obs Geoinformation 65 (2018) 12–23

Contents lists available at ScienceDirect

Int J Appl Earth Obs Geoinformation

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

Ensemble classification of individual *Pinus* crowns from multispectral satellite imagery and airborne LiDAR

Collins B. Kukunda[a,*], Joaquín Duque-Lazo[b], Eduardo González-Ferreiro[a,c], Hauke Thaden[d], Christoph Kleinn[a]

[a] University of Goettingen, Forest Inventory and Remote Sensing, Faculty of Forest Sciences, Büsgenweg 5, D-37077 Göttingen, Germany
[b] Department of Forestry, School of Agriculture and Forestry, University of Córdoba, Laboratory of Dendrochronology, DendrodatLab-ERSAF Edf Leonardo da Vinci, Campus de Rabanales s/n, 14071 Córdoba, Spain
[c] Unidade de Xestión Forestal Sostible (UXFS) – Departamento de Enxeñería Agroforestal, Escola Politécnica Superior, R/Benigno Ledo, Campus Terra, E-27002 Lugo, Spain
[d] University of Goettingen, Chair for Statistics, Humboldallee 3, D-37073 Göttingen, Germany

ARTICLE INFO

Keywords:
Individual tree crown segmentation
Ensemble regression and classification
Machine learning
Data integration
Spectrally and structurally similar tree species

ABSTRACT

Distinguishing tree species is relevant in many contexts of remote sensing assisted forest inventory. Accurate tree species maps support management and conservation planning, pest and disease control and biomass estimation. This study evaluated the performance of applying ensemble techniques with the goal of automatically distinguishing *Pinus sylvestris* L. and *Pinus uncinata* Mill. Ex Mirb within a 1.3 km^2 mountainous area in Barcelonnette (France). Three modelling schemes were examined, based on: (1) high-density LiDAR data (160 returns m^{-2}), (2) Worldview-2 multispectral imagery, and (3) Worldview-2 and LiDAR in combination. Variables related to the crown structure and height of individual trees were extracted from the normalized LiDAR point cloud at individual-tree level, after performing individual tree crown (ITC) delineation. Vegetation indices and the Haralick texture indices were derived from Worldview-2 images and served as independent spectral variables. Selection of the best predictor subset was done after a comparison of three variable selection procedures: (1) Random Forests with cross validation (AUCRFcv), (2) Akaike Information Criterion (AIC) and (3) Bayesian Information Criterion (BIC). To classify the species, 9 regression techniques were combined using ensemble models. Predictions were evaluated using cross validation and an independent dataset. Integration of datasets and models improved individual tree species classification (True Skills Statistic, TSS; from 0.67 to 0.81) over individual techniques and maintained strong predictive power (Relative Operating Characteristic, ROC = 0.91). Assemblage of regression models and integration of the datasets provided more reliable species distribution maps and associated tree-scale mapping uncertainties. Our study highlights the potential of model and data assemblage at improving species classifications needed in present-day forest planning and management.

1. Introduction

Physically and spectrally similar species such as *P. sylvestris* (*Pinus sylvestris* L.) and *P. uncinata* (*Pinus uncinata* Mill. Ex Mirb) are hard to distinguish in the field and in remote sensing imagery. In some cases microbiological analyses (Alvarez et al., 2009; Boratynska and Boratynski, 2007) are required to separate their identity. The microbiological methods provide high taxonomic precision but are inapplicable when, for example, species identification is required in forest inventories over large areas. Remotely sensed hyperspectral and/or multispectral data show potential to address this challenge. They can accurately recognize and map continuous stochastic distributions of vegetation communities, species groups, land-use and land-cover types and individual species across different genera. However, optical remote sensing data can fail when discriminating individual tree species that have similar appearances. An early attempt by Coleman et al. (1990) using Landsat TM data failed to discriminate between *Pinus* species stands because of similarity in spectral responses. Later, Goodwin et al. (2005) using high spatial resolution airborne CASI-2 data found individual *Eucalyptus* species – "spectrally complex" – and opted for genera groups. In the time since the above mentioned studies were published, spatial and spectral resolutions of datasets have increased

* Corresponding author.
E-mail addresses: ckukund@gwdg.de (C.B. Kukunda), jduquelazo@gmail.com (J. Duque-Lazo), edu.g.ferreiro@gmail.com (E. González-Ferreiro), hauke.thaden@wiwi.uni-goettingen.de (H. Thaden), ckleinn@gwdg.de (C. Kleinn).

http://dx.doi.org/10.1016/j.jag.2017.09.016
Received 18 July 2017; Received in revised form 22 September 2017; Accepted 29 September 2017

making available less correlated sub-genus reflectance data. However, spectrally and structurally related species continue to present a challenge. Studies reported that discrimination of closely related tree species was hampered by: (1) high structural similarity undermining the utility of structural characteristics like crown shape, size and leaf area index (Goodwin et al., 2005), (2) a higher intra-species than inter-species leaf spectral variability implying a fuzzy spectral signal (Youngentob et al., 2011) and (3) mixed spectral signals resulting from the influence of canopy scale structure (e.g. vegetation to background ratio in a pixel) (Aberle, 2016), crown aspect (i.e. shaded or non-shaded crowns) (Gerard and North, 1997) and noise from tree age and phenology (Clark et al., 2005; Peerbhay et al., 2014). Such challenges explain the limited usefulness of optical structural and spectral attributes in distinguishing similar tree species.

Complementary to spectral features recorded by means of optical remote sensing are structural features of the tree crowns or forest canopies that can be derived from LiDAR (Light Detection and Ranging) point clouds (Vauhkonen et al., 2014). LiDAR data offers the opportunity of describing some differences in properties of species crowns (e.g. the amount and allocation of biomass to branches and leaves) by recording differences in point height distributions (Ørka et al., 2009; Vauhkonen et al., 2014), especially when high point densities are available. LiDAR derived height and density distributions can be isolated for individual trees by means of individual tree detection techniques (e.g. Wulder et al. (2000)) and crown segmentation methods. Ultimately, the individual crown metrics can be used as predictors of species identities in object oriented approaches (Gougeon and Leckie, 2006; Ke et al., 2010). Studies have also shown that LiDAR intensity data i.e. strength of the back scattered energy, is useful in distinguishing between tree species, particularly when used in conjunction with other structural LiDAR-derived variables (Kim et al., 2009; Suratno et al., 2009; Zhang and Lui, 2013). For example, combined intensity and structural features to distinguish Norway Spruce and Birch trees which resulted in an overall accuracy of 88%. Similarly, Zhang and Lui (2013) demonstrated the applicability of LiDAR-derived structure and intensity variables to distinguish *Nothofagus cunninghamii* (Hook.) Oerst. and *Acacia dealbata* Link using Support Vector Machines (SVM) attaining overall accuracy up to 88.6%. Suratno et al. (2009) also used both structural and intensity predictors for identifying four species of individual trees in a mixed coniferous forest and reported kappa of 56% compared to kappa of 93% based on stands. Nonetheless, some researchers argue that the lack of a spectral signal remains an important limitation of LiDAR data in identifying tree species (Deng et al., 2007; Leckie et al., 2003; Swatantran et al., 2011) and they propose integration of LiDAR and spectral datasets as a more effective method.

Irrespective of the dataset in question and the similarity of the target objects, automatic species differentiation requires a classification model to link field observations to predictors. In many cases, a convenient multivariate form of the classification model is not known. In their review article, Fassnacht et al. (2016) show that different studies have explored various classification techniques including parametric approaches like: Discriminant analysis (FDA), Maximum likelihood (ML), Spectral Angle Mapper (SAM), Bayesian regression, Generalized Linear Models (GLM), Spectral Mixture Analysis (SMA), Logistic regression, Fuzzy logic, and thresholding and non-parametric approaches like; SVM, Random Forests (RF) and Artificial Neural Networks (ANN). We argue here that for any given choice of a model (parametric, semi- or non-parametric) or any choice of a variable set (dependent on the chosen variable selection procedure and datasets), the accuracy in prediction may vary since: (1) different model families fit the data to varying degrees (Appendix 1 in supplementary material shows differences in model fit along the data range given the model type) and, (2) a predictor effect is partially influenced/confounded by effects of other covariates. Recently, some studies proposed ensemble classification approaches to address the issues related to uncertainty in prediction across classifiers (Engler et al., 2013; Ko et al., 2014; Duque-Lazo and Navaro-Cerrillo, 2017). These techniques essentially combine decisions from several statistical classifiers with the aim of minimizing generalization error (Banfield et al., 2007; Ko et al., 2014; Engler et al., 2013). In the same way, their research demonstrates a consistent improvement in classification accuracies when employing ensemble modelling methods. However, these methods have not been tested in distinction of "spectrally and structurally similar" species and from that originates the motivation for this study.

The study contributes to improved mapping of structurally and spectrally similar tree species based on two remote sensing data sources at individual tree scale. We compared their predictive performance and later integrated high density airborne LiDAR data and high resolution Worldview-2 optical satellite data and regression classifiers. We additionally evaluated the performance of three variable selection procedures in the selection of the best predictor subset across classifiers.

2. Materials and methods

2.1. Study area

The study site is located in the South Eastern alpine part of France, district of Barcelonnette (latitude 44° 25′ 22.87″ N and longitude 6° 40′22.43″ E; Fig. 1). The area is about 1.3 km^2 and is covered by mainly unmanaged forests of *P. sylvestris* and *P.uncinata* (≈95% of total area). Norway spruce (*Picea abies* (L.) H.Karst.), European Larch (*Larix decidua* Mill.) and some other broadleaved species cover the remaining ≈5% of the study area. Altitude ranges between 1400 and 2020 m.a.s.l. Topography is irregular (mountainous) with slope gradients ranging between 10° and 70°. This type of terrain posses some challenges especially during the generation of elevation models and topographic normalization of spectral responses.

2.2. Characteristics of studied species

To demonstrate the potential of ensemble techniques, our study focused on the two dominant species in the study area. *P. sylvestris* and *P. uncinata* have chromatic and morphological differences that can be leveraged for their distinction (Farjon, 2010). These are: (1) the crown of *P. uncinata* is conical with narrow spreading lateral branches whereas that of *P. sylvestris* is conical-ovoid to ovoid with widely spreading to ascending lateral branches when mature but conical when young; (2) The density of the branches varies with the growth of the tree but is generally denser and grows to a lower base height for *P. uncinata* than *P. sylvestris*; (3) Visually, *P. uncinata* has a greyish-black trunk whereas *P. sylvestris*'s bark is reddish gray at the base and orange at the thin bark of the upper trunk and major branches; (4) *P. uncinata* is generally shorter (12–20 m) than *P. sylvestris* (15–35 m) at maturity and (5) the barks of both species are scaly plated or fissured with varying degrees of surface roughness but generally with the bark of *P. sylvestris* rougher than that of *P. uncinata* (Farjon, 2010). Despite these differences in crown appearance and architecture, a visual distinction of *P. sylvestris* and *P. uncinata* via needles can be difficult (Fauvart et al., 2012). This property exacerbates the challenge of their characterization from optical remote sensing as needles contribute the largest portion of reflectance in the imagery. We expected that combining LiDAR and Worldview-2 datasets as well as regression classifiers would help capture various biophysical traits and therefore increase the chance of their automatic distinction.

2.3. Data

LiDAR data, a 15 × 15 cm RGB aerial orthophoto and Worldview-2 (2 × 2 m, 8 bands and 0.5 × 0.5 panchromatic band) datasets were acquired during leaf-on and snow free conditions in June of 2009 and September of 2010, respectively. LiDAR data and the aerial orthophoto

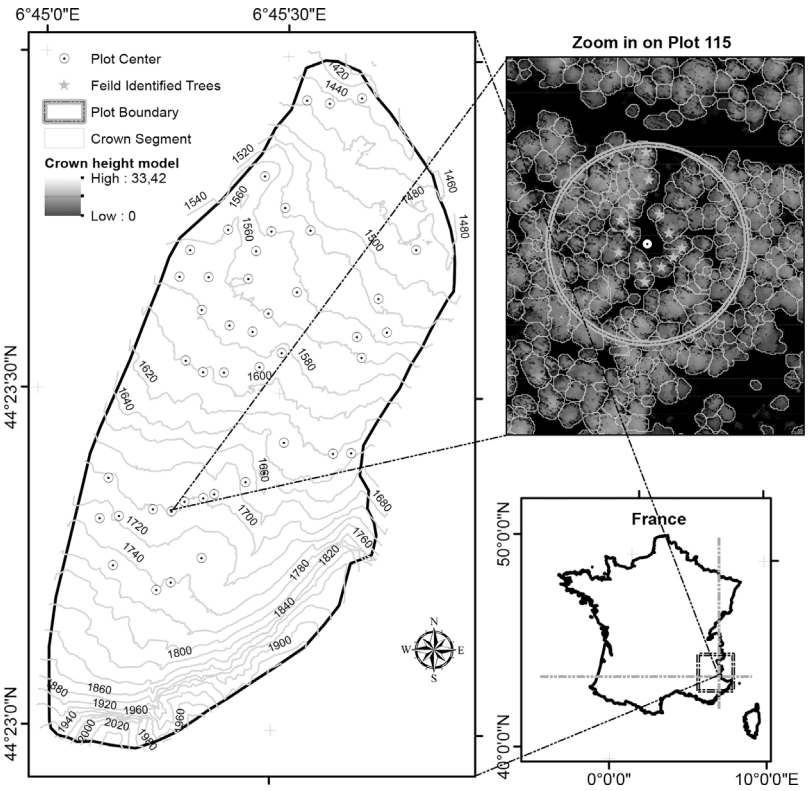

Fig. 1. Map of the study area located within the following X and Y UTM 32N coordinates (WGS84 reference system): 320,797.6; 4,916,782.5 (lower left) and 321,992.5; 4,918,787.1 (upper right).

were collected from an aircraft flying at an average 300 m above ground. A Riegl VQ480i system with a pulse repetition rate of up to 300 kHz and a rotating mirror scanning method was used. A very high density of the point cloud (160 points m^{-2}, seven discrete returns per pulse) was achieved from seven overlapping flight lines.

For the Worldview-2 imagery, level 2 post processing had been done by the vendor. The average sun elevation and azimuth angles were 48.1° and 161.7° and the average satellite elevation and azimuth angles were 74.8° and 55.0° respectively. The sun elevation and azimuth angles were used as input parameters during topographic normalization of the images to eliminate potential effects of shadows resulting from the rugged terrain.

2.4. Fieldwork

Field work was done in autumn 2012, 2- and 3-years after collection of LiDAR and Worldview-2 imagery respectively. A total of 48 circular field plots of 500 m^2 (with slope correction) were measured. Within each plot, all trees with diameter at breast height (DBH, 1.3 m above the ground) larger than 7 cm were recorded and callipered in two orthogonal directions. However, of all the trees encountered on the plots, only a subset of 544 trees were identified and their location recorded: 273 individuals of *P. sylvestris* and 271 individuals of *P. uncinata*. Note that the 544 trees are those whose crowns were visible in the dominant canopy and whose treetops could be determined on the canopy height model (CHM). We were able to verify locations of the identified dominant trees in the field using a CHM (Fig. 2), whereby the ground horizontal distance and orientation from a landmark were compared to the CHM distance and orientation. The landmarks included other isolated trees and canopy gaps.

2.5. Canopy height model generation and individual tree crown (ITC) delineation

The LiDAR Digital Elevation Model (DEM) and Digital Surface Model (DSM) were generated by griding Delaunay triangulated ground and first returns into 15 cm resolution rasters respectively. Tree position, height, and crown width was retrieved from the CHM by using a modified version of the mixed-pixel and region-based algorithm designed by González-Ferreiro et al. (2013). This algorithm is proposed as a sequence of routines programmed in the integrated development environment (IDE) (eCognition Developer 8.7 (®Trimble GmbH, Munich, Germany)). All logical procedures in the construction of the canopy delineation algorithm were arranged into five groups: CHM smoothing, segmentation, classification of canopy areas, iterative process, and data export. For this study, one parameter of the algorithm has been tuned, to adapt to the analyzed species. Concretely, we have changed the break value for the roundness index from 0.5 to 0.25 in the shape decision criteria (see the Fig. 3 in González-Ferreiro et al. (2013)) in order to discard and control for unnatural crown shapes.

Finally, the ITC result (Fig. 2) and crown attributes were then exported as vector polygons in an ESRI™ shapefile with the associated database. Tree tops and height attributes were also exported as a point vector shapefile for subsequent analysis.

2.6. Assessment of accuracy of delineated crowns

A good ITC segmentation was prerequisite in order to accurately isolate individual tree predictors. To evaluate the accuracy of the crown delineation algorithm, we closely followed two area-based methods presented by Clinton et al. (2010). The reference polygons (assumed truth), relative to which the performance of the segmentation algorithm was judged, were manually digitized from a CHM overlain on a high resolution (0.15 × 0.15 m) aerial orthophoto collected together with the LiDAR data. Three student assistants performed this task, stratawise – i.e. the study area was divided into three parts – and each independently, on desktop screens of 1600 × 900 pixel resolution and at a common map scale of 1:50. The assistants were trained before the digitization according to a defined protocol detailing ways to identify crown boundaries based on changes in grey-color contrasts. For all study participants, 1615 individual crown polygons spread across the entire study area, were manually digitized and used in the assessment

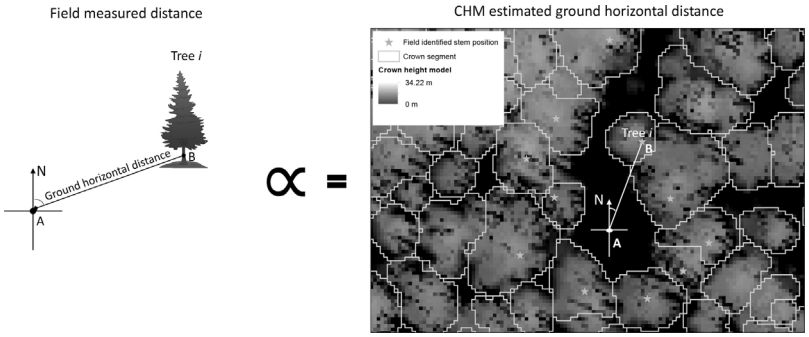

Fig. 2. Identification of individual tree stem locations in the field, whereby; A is the plot center or landmark, B is the stem position, *i* is the tree identification number ranging from 1 to 544 and N signifies the North direction. The identification of tree stem positions on the Canopy Height Model is shown on the right whereby, ground horizontal distance (A to B) ≈ CHM distance (A to B) × CHM image resolution. The crown segments are an output of automatic individual tree crown delineation using the Canopy Height Model as input data.

of the segmentation accuracy.

We computed the segmentation goodness metric (*Eq. (1)*) for each reference polygon based on relative overlap with the largest corresponding automatically generated segment (Clinton et al., 2010).

$$\text{segmentation goodness}_{ij} = 1 - \frac{A_{Ii}}{A_{Rj}} \quad (1)$$

A_{Ii} is the area in intersection ($A_{Ai} \cap A_{Rj}$) between an automatically generated segment and a corresponding manually delineated segment, A_{Ai} is the area of segment i and A_{Rj} is the area of the corresponding reference segment.

To compare pairwise sizes and context of automatically generated crowns relative to the reference polygons, we computed the area fit index (*AFI*) (Clinton et al., 2010) as shown in *Eq. (2)*.

$$\text{AFI}_j = \frac{A_{Rj} - A_{maxAi}}{A_{Rj}} \quad (2)$$

AFI_j is the Area Fit Index for segment j and A_{maxAi} is the area of the largest automatically generated segment intersecting with A_{Rj}.

Following Clinton et al. (2010), we classified over-segmented reference polygons (O) as reference objects with less than 100% overlap with automatic polygons and $AFI > 0$ and under-segmented reference polygons (U) as those of 100% overlap with automatic polygons and $AFI < 0$. By averaging over all over-segmented and under-segmented polygons we computed the closeness index D shown in *Eq. (3)*.

$$D = \sqrt{\frac{O_{\in k}^2 + U_{\in m}^2}{2}}. \quad (3)$$

O is the mean segmentation goodness across the k over-segmented reference polygons and the term U is the mean segmentation goodness across m under-segmented reference polygons. We judged our segmentation result based on Clinton et al. (2010) that identified closeness index values of 0.3 as good segmentations.

Lastly, an overall rate of correct detection was computed as the ratio of correctly detected reference polygons and all reference polygons. Correct detection was defined by the following limits: overlap ⩾70% and $-3 \leqslant AFI \geqslant 2$. The *AFI* thresholds were determined based on visual inspection.

2.7. Preparation of predictors

The individual crown segments delineated in the previous step were used to isolate individual crown variables from both remote sensing datasets. From the height normalized point cloud, metrics describing per-crown return intensity and height distributions and variables describing sizes and shapes of individual crowns were extracted using the FUSION LiDAR Toolkit (McGaughey, 2014). As no flight trajectory file was available for this study, radiometric normalization of intensity data was done following González-Ferreiro et al. (2014) and based on a user defined standard range. Here note that Korpela et al. (2010) report a marginal gain in accuracy (2–3%) of species classification after intensity normalization while distinguishing conifers in boreal conditions. Fassnacht et al. (2016) mention that *"this task can be quasi-impossible in mountainous terrain"*.

Standard crown size and shape variables were extracted from the points and polygon shapefiles generated in the ITC delineation. LiDAR metrics were derived from the normalized LiDAR height and intensity distributions within the limits of the delineated individual tree crowns. The minimum height threshold (MHT), which is commonly specified as the lower boundary for calculating height metrics (central tendency, dispersion, shape and percentile statistics), was established at 1 m. The height break threshold (HBT), which is the limit for separating the point cloud data into two sets to separate canopy returns from the under canopy returns, in order to estimate canopy cover metrics, was established as 5 m (based on field observation). In total 108 metrics were derived from the LiDAR data (see Table 1 for a complete description of LiDAR-derived metrics).

Before extracting per-crown optical metrics, topographic normalization based on local parameter estimation of spectral differences (Mo et al., 2015) was done using the LiDAR DEM. The topographic normalization aimed at reducing radiometric distortions brought about by shadow effects in rugged terrain especially enhanced by a low scanning and sun elevation angle at the time of image acquisition. Optical images were resampled using a nearest neighbor interpolation from 2 m to 0.5 m before computing the per crown statistics (Table 1). The following broadband greeness and leaf pigmentation products were derived: Normalized Difference Vegetation Index (NDVI), Simple Ratio Index (SRI), Enhanced Vegetation Index (EVI), Atmospherically Resistant Vegetation Index (ARVI) and Anthocyanin Reflectance Index (ARI). Additionally, per crown statistics were computed from texture indices derived from the panchromatic band including; energy, entropy, correlation, inverse distance moment, inertia, cluster shade, cluster prominence and Haralick correlation. The same statistics were calculated for each of the eight multispectral bands (Table 1). In total 230 spectral predictors were prepared.

2.8. Statistical models and calibration

The predictors described above were used to distinguish between the two tree species via multiple regression approaches. We used the default settings of the *biomod2* R-package version 3.3-7 (Thuiller et al., 2016) to ensemble results of 9 regression techniques into a final prediction. The models are described in detail in . The included modelling techniques were; Classification and Regression Trees (CTA), Generalized Linear Models (GLM), Generalized Boosting Models (GBM), Generalized Additive Models (GAM), Artificial Neural Networks (ANN), Flexible Discriminant Analysis (FDA), Multiple Adaptive Regression Splines (MARS), Random Forests (RF) and MAXENT.Phillips. The models were calibrated with species data based on the field observations.

Multi-collinearity analysis was done prior to the fitting process to assess whether two or more explanatory variables were significantly correlated. Collinearity among all explanatory variables was evaluated by the Variance Inflation Factor (VIF) (Fahrmeir et al., 2013, P. 156),

Table 1
Potential LiDAR and optical predictors. Note that descriptions of metrics are listed in the order of appearance of the acronyms and not repeated for each variable group. Metrics are grouped into: Intensity (*I*), Crown cover, shape and size (*C*), Height (*H*), Individual bands (*B*1, …, *B*8), Texture indices (*T*) and Vegetation indices (*V*). "…" refers to "in the same sequence ".

Abbreviation	Description
LiDAR metrics on cover, size and shape of crowns (C)	
C_{cr}, C_{cwr}, C_{len}, C_{cls}, C_{cbh}, C_{per},C_{per1}, …, C_{per4}, C_{crr}, C_{r1}, …, C_{r7} .	Crown radius, Crown-width ratio, Crown perimeter length, Crown closure, Crown base height, Percentage first returns above 5 m, Percentage first returns above mean, Percentage first returns above mode, Percentage all returns above 5 m, Percentage all returns above mean, Percentage all returns above mode, Canopy relief ratio, Count of returns by return number (1-7).
LiDAR metrics on height (H) and intensity (I) distributions – dispersion statistics	
H_{IQD}, H_{SD}, H_{var}, H_{CV}, H_{skew}, H_{kur}, H_{AAD}, H_{MeAD}, H_{MoAD}, H_{L2}, …, H_{L4}, H_{LSkew}, H_{LKur}, I_{IQD}, …, I_{LKur}.	Inter-quartile distance, Standard deviation, Variance, Coefficient of variation, Skewness, Kurtosis, Average absolute deviation, Median of the absolute deviations from median, Median of the absolute deviations from mode, L-Moments (2:4), L-Moment of Skewness, L-Moment of Kurtosis.
LiDAR metrics on height (H) and intensity (I) distributions – descriptive statistics	
H_{max}, H_{min}, H_{mean}, H_{med}, H_{mode}, H_{P01}, …, H_{P99}, I_{max}, …, I_{P99}.	Maximum, Minimum, Mean, Median, Mode, Percentiles (1,5,10,20,25,30, …, 70,75,80,90,95,99).
Spectral metrics from individual bands (B) and vegetation indices	
$V_{minNDVI}$, $V_{maxNDVI}$, $V_{sumNDVI}$, $V_{cntNDVI}$, $V_{meanNDVI}$, V_{sdNDVI}, $V_{uqeNDVI}$, $V_{rangeNDVI}$, $V_{varNDVI}$, $V_{medNDVI}$, $V_{modeNDVI}$, V_{minSRI}, …, $V_{modeSRI}$, V_{minEVI}, …, $V_{modeEVI}$, $V_{minARVI}$, …, $V_{modeARVI}$, V_{minARI}, …, $V_{modeARI}$, V_{minB1}, …, V_{modeB1}, …, V_{minB8}, …, V_{modeB8}.	Minimum, Maximum, Sum, Count, Mean, Standard deviation, Unique, Range, Variance, Median, Mode of: vegetation indices (NDVI, EVI, ARVI, ARI), and individual bands (*B*1, …, *B*8).
Spectral metrics from texture indices (T)	
$T_{minInertia}$, $T_{maxInertia}$, $T_{sumInertia}$, $T_{cntInertia}$, $T_{meanInertia}$, $T_{sdInertia}$, $T_{uqeInertia}$, $T_{rangeInertia}$, $T_{varInertia}$, $T_{medInertia}$, $T_{modeInertia}$, $T_{minEner}$, …, $T_{modeEner}$, T_{minEnt}, …, $T_{modeEnt}$, T_{minCor}, …, $T_{modeCor}$, T_{minIDM}, …, $T_{modeIDM}$, T_{minCP}, …, T_{modeCP}, $T_{minHCor}$, …, $T_{modeHCor}$.	Minimum, Maximum, Sum, Count, Mean, Standard deviation, Unique, Range, Variance, Median, Mode of: Inertia, Energy (Ener), Entropy (Ent), Correlation (Cor), Inverse distance moment (IDM), Cluster prominence (CP), Haralick correlation (HCor).

considering VIF > =10 as the step-wise elimination threshold (Duque-Lazo et al., 2016).

To overcome dimensionality issues commonly faced in multiple regressions of small sample sizes to many predictor variables, and so as to select for the most robust predictor set, three variable selection procedures were considered, namely: (1) AUCRFcv (Calle et al., 2011), (2) AIC and (3) BIC (Fahrmeir et al., 2013).

A variable importance measure (VI) was used to assess the relative contributions of individual variables in a given model. To allow for comparability of variable importance across models, VI was computed from model predictions (for all model types). After obtaining a prediction (P_1), the variable of interest was shuffled while holding all the other variables in the model at their median value and a new model prediction made (P_2). A simple correlation between P_1 and P_2 was then calculated to obtain VI according to Eq. (4):

$$VI = 1 - \text{cor}(P_1, P_2) \quad (4)$$

A zero *VI* value meant no importance whereas the importance of a variable in a predictor set increased with VI scores (Thuiller et al., 2016).

2.9. Model and map evaluation

Model fit was evaluated using the relative operating characteristic (ROC) and map quality was assessed with the true skills statistic (TSS). Both ROC and TSS were computed from cross validation and split sample procedures. During cross validation, all observational data (544 field identified trees) was utilized where models were fit with 80% of the data and evaluated with the remaining 20% of the data. An almost 1:1 ratio, 271 training and 273 validation, was used in the split sample approach. The same training and validation set was kept constant across all runs and variable selection procedures. Models were run 50 times in order to monitor variation in variable importance and prediction accuracies. Robust models were defined by a combination of both ROC and TSS.

2.10. Ensemble prediction

A stacked assemblage of predictions across individual models was based on mean, median, inferior confidence interval, superior confidence interval, committee average and a weighted mean (Thuiller et al., 2016). Not all individual models computed per run were included in the ensembles. Different ROC quality thresholds were tested to select models into the ensemble and minimize loss in the resultant ensemble model and map accuracy brought about by inclusion of weak or less informative individual models. Eventually, ROC = 0.8 was used. The model and map evaluation described in Section 2.9 applies to ensemble models.

Finally, three modelling schemes were considered based on: (1) high-density LiDAR data, (2) spectral image layers derived from multispectral imagery (Worldview-2), and (3) both the spectral and LiDAR data. For each scheme, all variable selection and model types were evaluated.

3. Results

3.1. Individual crown detection

More than 70% of dominant crowns in the canopy were correctly delineated according to the defined criteria. Closeness index values averaged around 0.33 as shown strata-wise in (Table 2).

Table 2
Segmentation goodness across study participants.

Participant	Segmentation measures		No. of Crowns
	Correct detections (%)	Closeness index (D)	
I	73.36	0.31	517
II	60.65	0.41	310
III	77.48	0.33	826
Total	73.18	0.33	1615

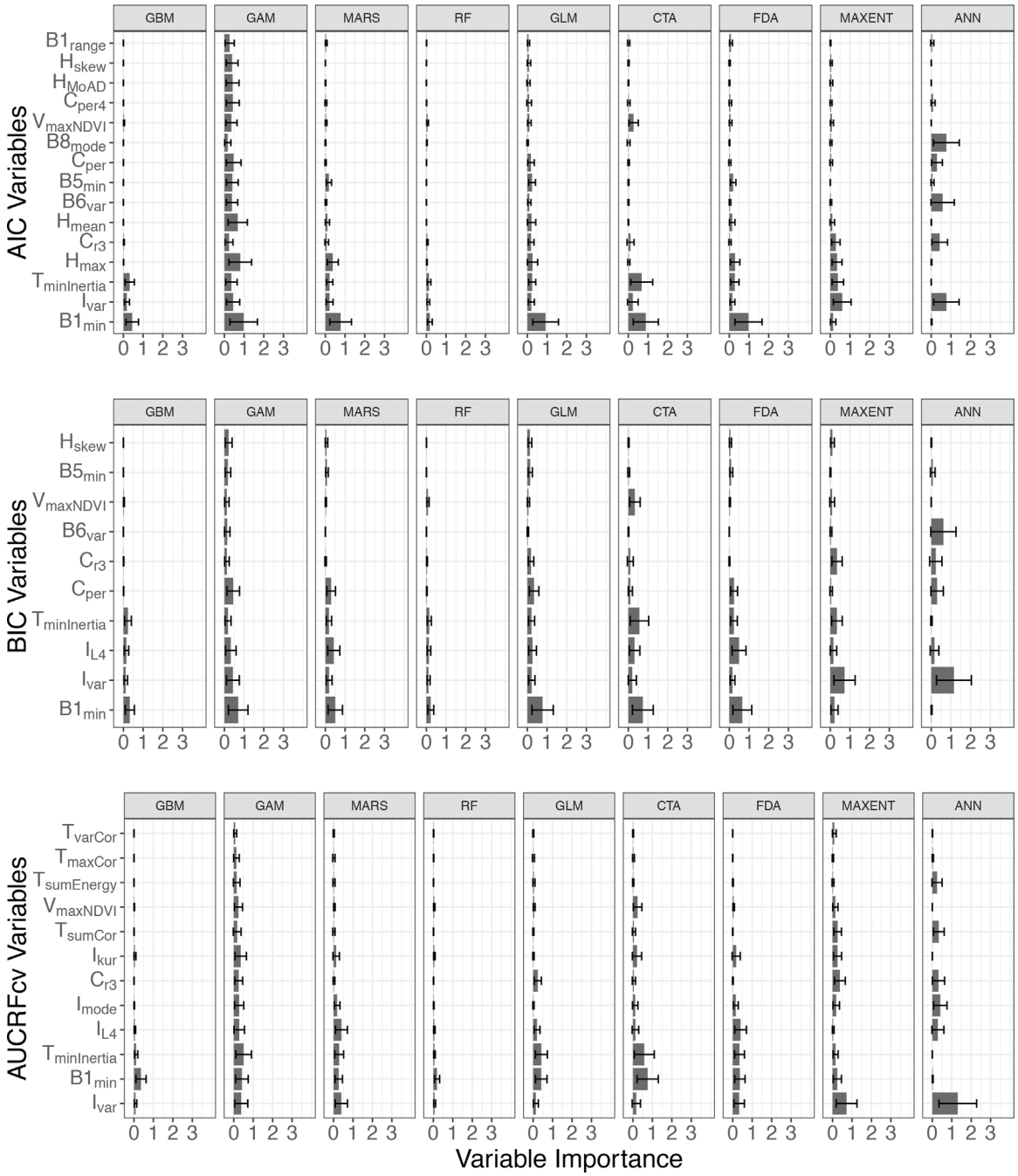

Fig. 3. Mean variable importance of spectral and LiDAR predictors across 50 model runs. Variable importance is sorted across models in an ascending order. Sorting of variable importances was global so as to identify variables of highest importance across models. Groups of metrics are: Intensity (I), Height (H), Crown cover, size and shape (C), Texture index (T), Vegetation index (V) and Individual band (B1, …, B8). For detailed predictor variable names including explanation of the subscript in each acronym, refer to Table 1. Models: Artificial Neural Networks (ANN), Classification and Regression Trees (CTA), Flexible Discriminant Analysis (FDA), Generalized Additive Models (GAM), Generalized Boosting Models (GBM), Generalized Linear Model (GLM), Multiplicative Regression Splines (MARS), Maximum Entropy (MAXENT.Phillips), Random Forests (RF).

3.2. Importances of predictor variables

Both LiDAR and optical metrics were of high importance in the computed models (Fig. 3). We suggest that identification of the most important predictors considers both the magnitude of variable importance (as shown in Fig. 3) and the consistence of selection across modelling schemes (Table 3). The reader is referred to supplementary material for importances of variables selected when the datasets were employed in isolation (individually). Amongst the LiDAR variables, the following were of high importance: I_{var}, I_{L4}, C_{per}, C_{r3}, and H_{max}, and in Worldview-2 imagery: $B1_{min}$, $B5_{min}$, $T_{minInertia}$, and $V_{maxNDVI}$.

Judged by the variable importance scores alone (i.e. without considering consistency of selection across modelling schemes), LiDAR intensity, optical imagery texture indices, and individual band reflectance characteristics received higher importance in the models compared to height, canopy cover metrics and vegetation indices (Fig. 3). Similar trends were observed when the datasets were employed in isolation (Appendices 2 and 3). There was generally agreement across individual models and variable selection procedures (Fig. 3). However, GAM and ANN assigned higher importance to variables less important in other models. Additionally, some models variable importances were more precise across runs compared to others (based on

Table 3
Selected predictors across modelling schemes and variable selection procedures. Metrics are grouped into: Intensity (*I*), Crown cover, shape and size (*C*), Height (*H*), Individual bands (*B*1, …, *B*8), Texture indices (*T*) and Vegetation indices (*V*). ☑ Refers to selected into the final predictor set given a variable selection procedure. For detailed predictor variable names including explanation of the subscript in each acronym, refer to Table 1.

		Predictor	Modelling scheme & variable selection procedure					
			Both			LiDAR		
			AUCRFcv	BIC	AIC	AUCRFcv	BIC	AIC
1.		I_{var}	☑	☑	☑	☑	☑	☑
2.		I_{L4}	☑	☑		☑	☑	☑
3.	Intensity	I_{kur}	☑			☑		☑
4.		I_{mode}	☑			☑		
5.		I_{L3}				☑		
6.		I_{P01}						☑
7.		C_{r3}	☑	☑	☑	☑		☑
8.		C_{r2}				☑		
9.		C_{r1}				☑		
10.	Crown cover,	C_{per}		☑	☑	☑	☑	☑
11.	size & shape	C_{per4}			☑			
12.		C_{length}				☑		
13.		C_{per3}						☑
14.		C_{cr}						☑
15.		H_{skew}		☑	☑	☑		
16.		H_{max}			☑	☑	☑	☑
17.		H_{mean}			☑	☑	☑	☑
18.	Height	H_{MoAD}			☑		☑	☑
19.		H_{L3}				☑		
20.		H_{var}				☑		
21.		H_{kur}					☑	
22.		H_{CV}						☑

			Both			Spectral		
23.		$B1_{min}$	☑	☑	☑	☑	☑	☑
24.		$B5_{min}$		☑	☑	☑	☑	☑
25.		$B6_{var}$		☑	☑			
26.	Individual	$B1_{range}$			☑		☑	☑
27.	Band	$B8_{mode}$			☑	☑		
28.		$B4_{med}$					☑	☑
29.		$B1_{var}$						☑
30.		$B2_{mode}$						☑
31.		$B8_{var}$						☑
32.		$T_{minInertia}$	☑	☑	☑		☑	☑
33.		$T_{sumEnergy}$	☑			☑		
34.		T_{sumCor}	☑			☑		
35.		T_{maxCor}	☑			☑		
36.	Texture	$T_{varEnergy}$	☑			☑		
37.	index	T_{varCor}	☑			☑		
38.		T_{sumCP}						☑
39.		$T_{varInertia}$						☑
40.		T_{minCS}						☑
41.		T_{modeCP}						☑
42.		$V_{maxNDVI}$	☑	☑	☑			
43.	Vegetation	V_{minSRI}				☑	☑	☑
44.	index	V_{sumARI}				☑		

standard error bars in Fig. 3). There was a higher precision in *VI* from RF, GBM and GLM compared to the other model categories.

3.3. Individual model and map quality

Two clusters in performance of individual models were observed when comparing individual models (Fig. 4). In general, RF, GBM, GLM, FDA, MARS formed a cluster of the better performing models compared to MAXENT.Phillips, CTA, ANN and GAM in the cluster of lower performing models. The former cluster had considerably lower variance in both quality of model fit and map accuracies across model runs. There was preponderance of linear models when spectral predictors were employed in isolation and a majority of classification tree models when LiDAR predictors were employed (Fig. 4). Additionally, spectral models were more stable than models using LiDAR only and Spectral and LiDAR. Generally, individual model and map quality across variable selection procedures was similar (Fig. 4).

All individual models and datasets failed to differentiate the species when employed independently (Fig. 4). Individual performance among the best five models was on average TSS = 0.55 for LiDAR and TSS = 0.64 for Spectral. However, the average individual performance of the same models increased slightly to about TSS = 0.67 when both datasets were integrated. In isolation, spectral variables exhibited higher predictive power compared to LiDAR variables.

3.4. Ensemble model and map quality

In general, all ensemble models performed equally for both

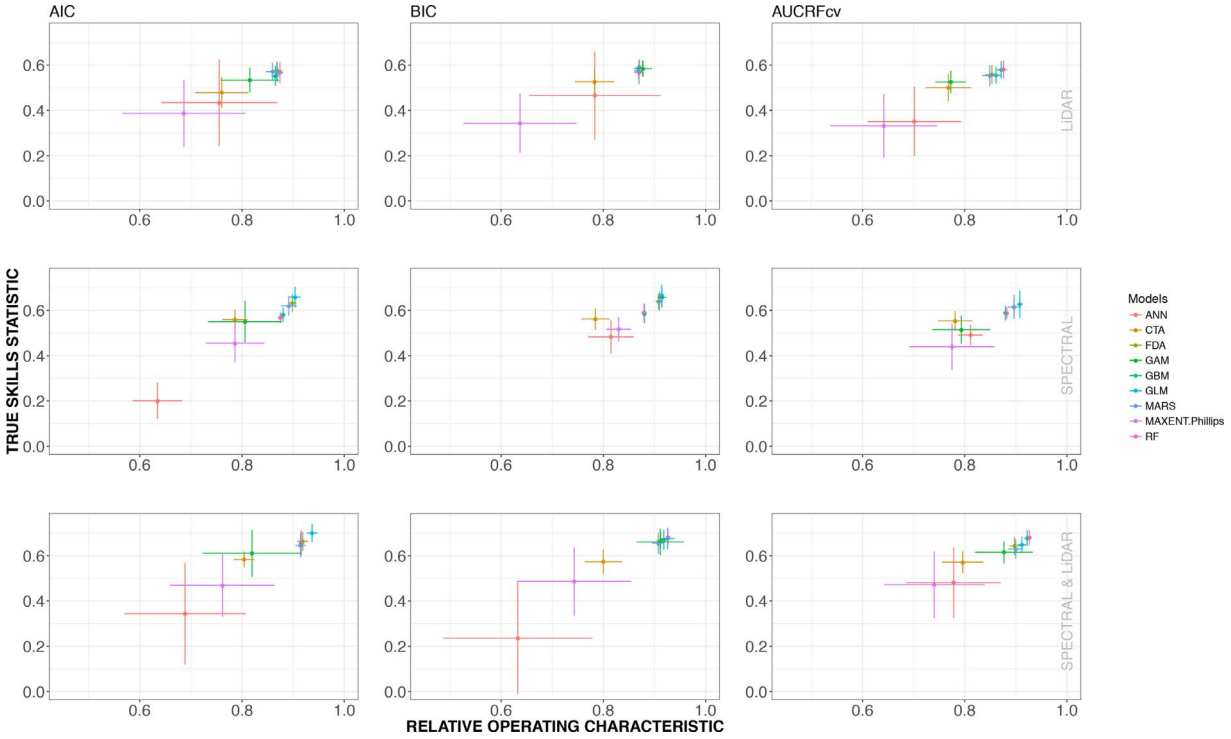

Fig. 4. Performance by model type (Artificial Neural Networks – ANN, Classification and Regression Trees – CTA, Flexible Discriminant Analysis – FDA, Generalized Additive Models – GAM, Generalized Boosting Models – GBM, Generalized Linear Model – GLM, Multiplicative Regression Splines – MARS, Maximum Entropy – MAXENT.Phillips and Random Forests – RF) and variable selection procedure (Top row = LiDAR, Middle row = Spectral, Bottom row = Spectral & LiDAR). The dots represent the mean and the lines represent the associated standard errors across 50 model runs.

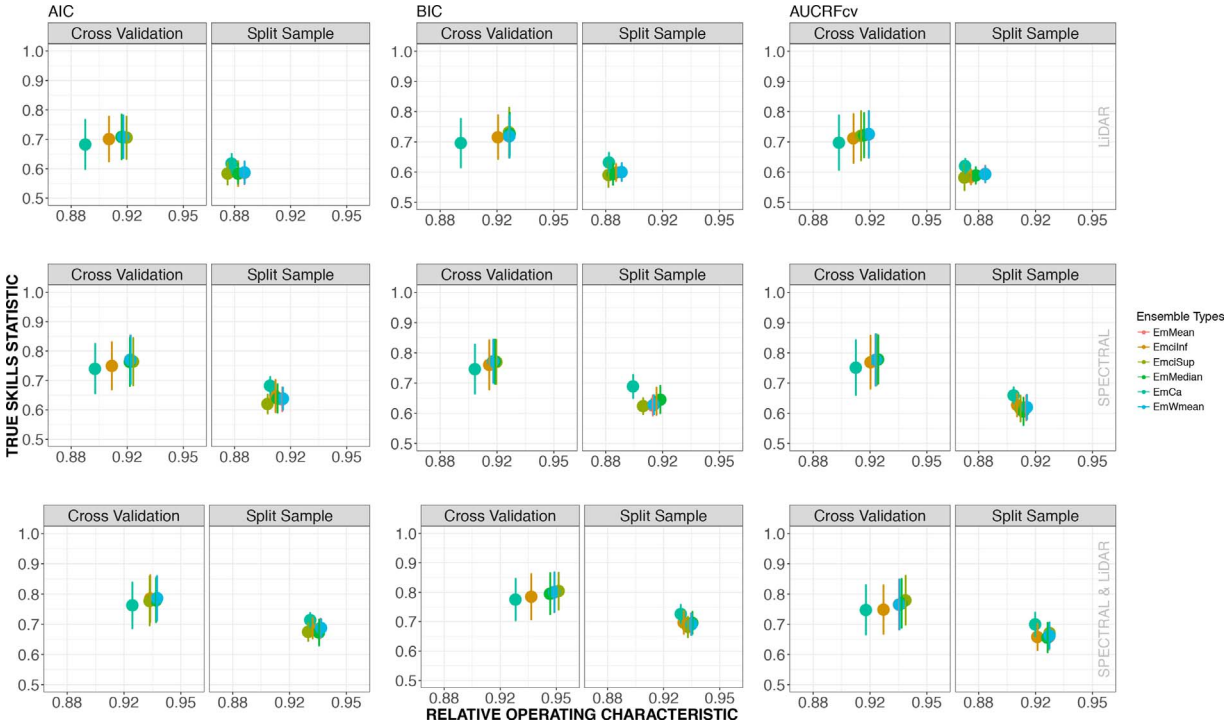

Fig. 5. Performance by ensemble model type (Mean – EmMean, Inferior confidence interval – EmciInf, Superior confidence interval – EmciSup, Median – EmMedian, Committee average – EmCa and Weighted mean EmWmean) and variable selection procedure (Top row = LiDAR, Middle row = Spectral, Bottom row = Both). The dots represent the mean and the lines, the associated standard errors across 50 model runs.

validation procedures. Assemblage of individual models increased map quality with across-model TSS averages to 0.63 for LiDAR models, to 0.68 for spectral models and to 0.73 for both spectral and LiDAR powered models (Fig. 5). Higher model and map quality values were reported by the cross validation procedure (Fig. 5). The Committee Average ensemble (EMca) was consistently the highest performing ensemble when evaluated using a split sample approach. EMca was the worst performing ensemble model based on cross validation.

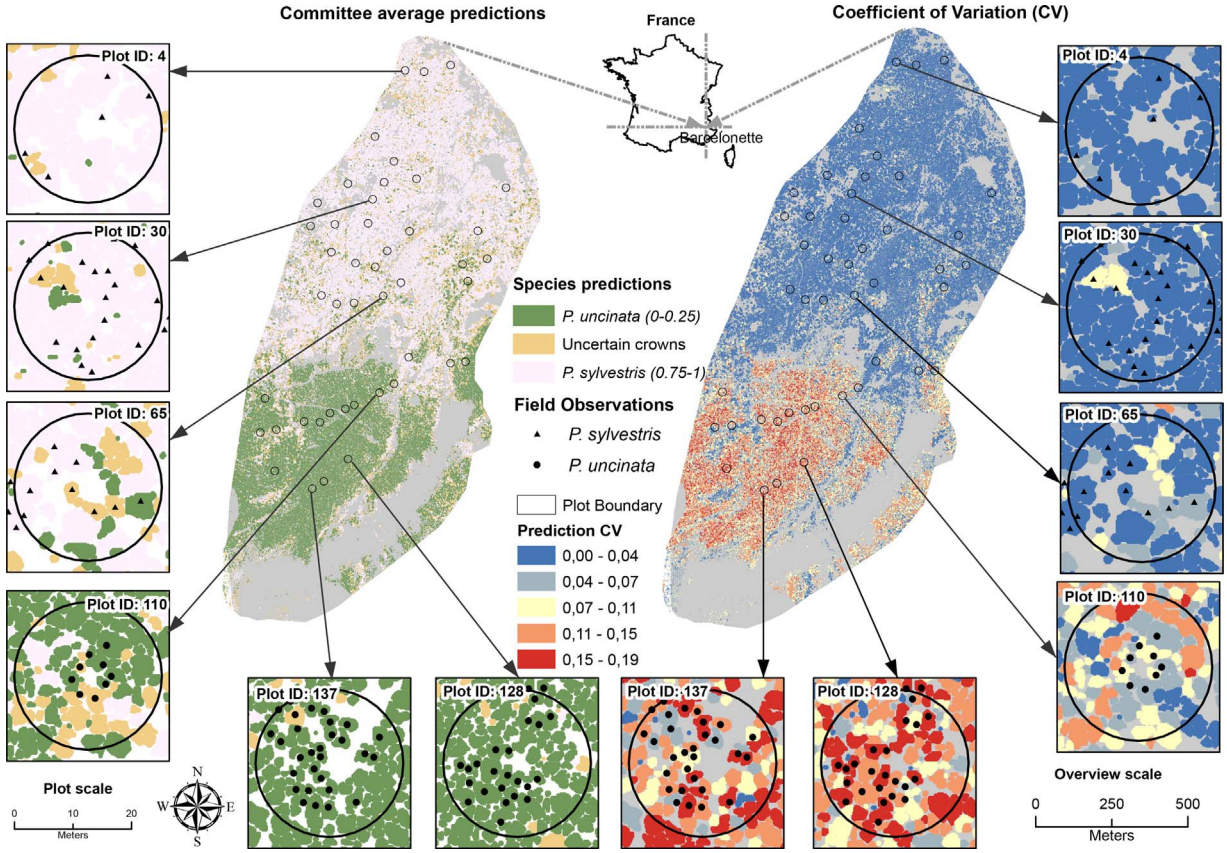

Fig. 6. An individual tree species map and associated across model uncertainty computed via the coefficient of variation.

Fig. 6 shows example maps of tree species at individual crown level. The maps show a single committee average prediction and the respective uncertainty estimate i.e. Coefficient of Variation (CV) across modelling techniques. There was higher uncertainty across models for predicting *P. uncinata* than was for predicting *P. sylvestris*. However, the uncertainty was generally low (CV ≤19%) indicating general agreement among individual models in the ensemble.

4. Discussion

Our result shows that it is possible to perform an accurate automatic tree crown segmentation (Closeness index = 0.33) of two very similar *Pinus* species in complex terrain, and that unlike the variable selection procedure, a combination of LiDAR and multispectral variables improved model prediction. More concretely, we have demonstrated that the ensemble model approach is superior to single model prediction. However, the key challenge is a high intra-species variance in classifications of individual trees.

4.1. Effect of a high intra- vs inter-species variation

A common recommendation when variance is greater within species is to employ object-oriented techniques (Gougeon and Leckie, 2006) with a higher spatial or spectral resolution, either in isolation or after some form of fusion (Leckie et al., 2003; Ke et al., 2010) in order to summarize the within class (tree) variation at a lower scale of observation. At the scale of single trees (crowns) our results have shown that the high within-tree variation can persist (Fig. 4). However, classification accuracies can improve with ensemble models (Fig. 5). The very high intra-species variation is attributed to differences in individual tree morphology and differences in local site conditions. No management planning information was available to this study, and neither was tree age data collected. We however did observe that there were differences in stand characteristics (for example, the size of trees) suggesting differences in either site specific rates of growth or stand ages. The difference has influenced within species variation in particular related to structure of crowns. *P. sylvetris* crowns are known to change with stages of maturity (Farjon, 2010; Ross et al., 1986). Similarly, the rugged terrain introduced variations in site relative to local slope, aspect and soil conditions and thus increased tree-scale within species variation. The terrain effect is further seen in the spatial pattern of model uncertainty (Fig. 6) as dissimilarity among models increased with elevation and ruggedness (Fig. 1) in areas dominated by *P. uncinata*. In other words, errors in normalization of remote sensing datasets induced higher model uncertainty in areas of rugged terrain. It should be noted however, that we utilized very high point density LiDAR which supported generation of very high quality elevation models uncommonly possible in complex terrain. This enhanced robustness of ITC delineation and precision of within crown statistics. In common practice, a good ITC delineation is possible with densities from between 4 and 10 points m^{-2} (Hamraz et al., 2017). However, at such lower densities, it will be probable that fewer points inside of each segment (delineated crown) affect precision of computed statistics and metrics. Therefore, additional within species variation could arise from errors in ITC delineation and metrics calculation. However, the latter two sources of variation are here considered of low importance based on the high segmentation accuracy achieved and the employment of advanced topographic and intensity normalization procedures. The higher intra-species variation triggered the need to utilize individual models with several predictor variables.

4.2. Important species predictors

As expected, a combination of both LiDAR and spectral variables resulted in increased differentiation capabilities as either datasets captured varying characteristics of the species crowns. Similarly, a

comparison between prediction power of multispectral and LiDAR variables in isolation suggested higher structural than spectral similarity between the species; resulting in a higher prediction power for spectral predictors. Combined, the most important spectral predictors: $B1_{min}$, $B5_{min}$, $T_{minInertia}$, and $V_{maxNDVI}$, highlight differences in condition, pigmentation and internal structure of the needles useful for their characterization. Similar conclusions were drawn by Alvarez et al. (2009) using microscopy in analysis of epidermal characteristics of needles of both species. Additionally, we attribute the high importance of the texture index $T_{minInertia}$ to the combined effect of chromatic differences in the upper trunk and branches of the two species and their differences in branch density (refer to Section 2.2 for details). $T_{minInertia}$ measures inter-pixel contrast within each crown area. The 50 cm resolution of the Worldview-2 imagery captured inter-pixel variation in brightness of the leaves and trunks that was useful in the distinction of the species. On the other hand, LiDAR variables highlighted the importance of bark roughness (via intensity metrics, I_{var}, I_{L4}), branch density (via canopy cover metrics, C_{per}, C_{r3}) and height (H_{max}) differences in characterization of the species.

4.3. Evaluation of individual models

The peculiar clustering in individual model performance showed; MAXENT.Phillips, ANN, CTA and GAM to have been consistently poorly performing and unstable across multiple runs. This was unsurprising for GAM and ANN since they simultaneously assigned higher importance to variables less important in other models (Fig. 3). In a similar ensemble setting, akin results regarding variable usage in ANN models were reported by Marmion et al. (2009). They related this unique behavior of the ANN model to its inherent non-selective nature, given that it builds intermediate relationships between predictors and therefore risks assigning higher importance to less effective variables based on their indirect contribution to the prediction process. On the other hand, GAMs fit local splines along the data range of each available variable, estimate a single smooth curve per variable, and then additively combine the results. Both structures of the GAM and ANN models are prone to becoming exceedingly complex and may therefore face over-fitting (Marmion et al., 2009). The challenge therefore was finding the optimal threshold between individual model complexity and overall ensemble mapping accuracy. Here, more complex individual models resulted in higher map accuracies attributed to their ability to minimize model bias. However, based on the comparison between the split-sample and cross-validation results (Fig. 5), we posit that the observed marginal discrepancy between map accuracies reported by both validation procedures is testimony that over-fitting was not a significant problem since ensemble models generalized well on new data.

Further, the poor performance of the CTA model is linked to the lack of bagging and bootstrapping capabilities, disposing the model type to bias and high across run variances (Briem et al., 2007). One notes that despite their similarity in structure, the RF model consistently outperformed CTA. This can be explained by regularization, bagging and bootstrapping algorithms in the RF model affording better performance at generalization on validation datasets. On the other hand, it was unique to this study that the performance of MAXENT.Phillips was not comparable with other models, such as RF and GBM, that are known to be consistently highest performing (Elith et al., 2011). We found out that the poor performance of MAXENT.Phillips – in this specific case – is linked to its calibration with presence-absence data rather than presence only data and to a sub-optimal specification of the prevalence parameter (set here to 0.5) from which a logistic output is generated after combining presence (here *P. sylvestris*) and background (here *P. uncinata*) data (Guillera-Arroita et al., 2014). Based on several modelling techniques, Fig. 6 shows that *P. sylvestris* potentially has a higher prevalence in the study area compared to *P. uncinata*. Guillera-Arroita et al. (2014) recommend to arrive at a prevalence estimate when parameterizing MAXENT.Phillips via the sample data, however, our individual tree selection approach did not afford the option to estimate unbiased prevalence before hand. The ideal practice would have been modelling the distribution of each species separately and employing the respective prevalence parameters independently. Nonetheless, we emphasize that the individual performance of MAXENT.Phillips had marginal to no effect on the presented ensemble results, given that an ROC threshold of 0.8 was used in selection of models into the ensemble which MAXENT.Phillips rarely passed. Remember that even within the cluster of better performing models (i.e. GLM, MARS, FDA, GBM, RF), the best individual model only achieved up to 40% better-than-random map accuracy (TSS = 0.69).

4.4. The contribution of an ensemble approach

Ensemble modelling improved individual models by increasing both map accuracy and minimizing prediction variance. The improvement was achieved through stacked fusion of predictions as well as due to the fact that the ensemble approach gives more weight to models with both good fit and effective bias-reduction. The bias minimization properties stemmed from tree-based and gradient boosting algorithms – such as RF and GBM – that sequentially arrived at a prediction by building several classifiers in a complementary tandem (Banfield et al., 2007; Briem et al., 2007). This way, our result is in agreement with Engler et al. (2013) that assembled per-pixel predictions of six species obtaining a cross validated Cohen's Kappa, $\kappa = 0.65$. For comparison with their study, the best mean cross-validation accuracy obtained in this study is TSS = 0.81 based on a combination of LiDAR and spectral datasets and BIC variable selection procedure (Fig. 5). A mean TSS of 0.81 means that the ensemble models were on average approximately 60% better-than-random at distinguishing *Pinus sylvestris* and *Pinus uncinata* at an individual crown scale. Note that TSS and κ are equivalent and that Engler et al. (2013) had to distinguish up to six species per pixel. Therefore, their expected error rate is higher. If the number of target species would have been more than two, we would have separate predictions of each species and spatially combine the result as done by Engler et al. (2013). However, we linked one species to the other and assigned class probabilities to either class in a single step. This approach had the advantage of computational efficiency but fell short for particular models (e.g. MAXENT.Phillips) as previously discussed. Similarly, it is worth explaining the discrepancy in behavior of the committee average ensemble (EmCa) given that it is the highest performing ensemble accuracy when evaluated in a split sample approach and the worst when evaluated by cross validation. This issue is related to the EmCa calculation which transforms all predictive probabilities into binaries (0 or 1) according to the maximum ROC/TSS threshold (the same as used in selection of individual models into the ensemble). Later, the new class probability is calculated as the average of the combined votes. During cross validation, a change in the training sample induces differences in model fit and therefore affects individual model votes – especially when encountering a high intra-class variation.

Further advantage of the ensemble approach rests in reporting across run variances of map accuracy scores. This is important in order to communicate the underlying precision of the estimated map accuracy. We show that with a single model, map accuracies can fluctuate across multiple runs with higher variances in cross validation than the split sample. At this point, it is important to understand that the two validation procedures convey fundamentally different assessments. While the cross validation approach estimates the expected prediction map accuracy, the split sample approach estimates both the conditional prediction map accuracy and model generalization capabilities (Hastie et al., 2009, p. 242). Expectedly, as with any model, our ensemble model generalization capabilities were lower (from TSS = 0.81 to TSS = 0.73) although better than individual models in isolation. The higher variances seen in the cross validation compared to the split sample approach are determined by the similarity of across fold training samples in the cross validation procedure (Hastie et al., 2009,

pp. 242–243). In other words, the more similar the training sample, the lower the across fold variation and therefore relates directly to the problem of intra-species variation. Certainly, the validation and test sample sizes as well affect the prediction precision when comparing between cross validation and split sample results.

Lastly, a spatial context to the precision of the estimated map accuracy is conveyed by across model coefficient of variation (Fig. 6) which communicates the level of agreement among individual models. This is a valuable resource when making inference towards the spatial variability of mapping errors. Therefore, such an output can in operational settings, for example, support selection of sites to focus map ground truthing exercises. Fig. 6 shows that there was a high agreement across models conveyed by a < 20% coefficient of variation. When such an estimate is used in tandem with the mean and variance estimates of map quality (e.g. TSS), it can enhance interpretation of the quality of the mapping product.

5. Conclusion

We have shown that assembly of regression models and integration of the datasets can provide a more reliable species distribution map with associated tree-scale mapping uncertainties. We have also shown that the approach can provide more transparent assessments of errors around modelled species distributions. Given that all the tried ensemble approaches performed equally, we do not recommend any in particular. We recommend leveraging of models and data assemblages in order to provide improved and transparent species classifications for forest planning, management and science at local to landscape scales. We speculate that the techniques and approaches used here lend themselves to other important areas such as classification of forest health conditions and forest degradation, among other classification challenges where inter-class overlap is pronounced. Similarly, the ensemble approaches presented here may perform better than conventional methods for species that are more clearly distinguishable and therefore should be studied further.

Conflict of interests

All authors declare no competing interests.

Acknowledgments

This research was conducted with financial support from the Deutsche Foschungsgemeinschaft (DFG) through the Research Training Group 1644 ‘scaling problems in statistics’, University of Goettingen, run by Prof. Thomas Kneib. Field work was supported by the European Union through the Erasmus Mundus Global Environment Modelling programme at the University of Twente, Faculty of Geoinformation Science and Earth Observation. Remote sensing data was collected under the auspices of the University of Twente, Faculty of Geoinformation Science and Earth Observation. We appreciate the support of MSc. Sunil Thapa during field work in the mountains of Barcelonette and thank Dr. Anahita Khosravipour, Dr. Hein van Gils and Dr. Yousif Hussin for their helpful comments during the project design phase and also for their support during field work. We are grateful to the student assistants; MSc. Volkin Nekondo, MSc. Susan Keuber Rweyemamu and MSc. Mireille Biraro, for their support in manual digitization of individual crowns. Further acknowledgement to the Galician Government and European Social Fund (Official Journal of Galicia DOG no. 52, 17/03/2014 p. 11343, exp: POS-A/2013/049) for funding the postdoctoral research stays of E. González-Ferreiro.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jag.2017.09.016.

References

Aberle, H., 2016. Hyperspectral remote sensing and field measurements for forest characteristics – a case study in the Hainich National Park, Central Germany. University of Goettingen (Ph.D. thesis).
Alvarez, S.G., Juaristi, C.M., Gutierrez, J.S., Garcia-Amorena, I., 2009. Taxonomic differences between *Pinus sylvestris* and *P. uncinata* revealed in the stomata and cuticle characters for use in the study of fossil material. Rev. Palaeobot. Palynol 155, 61–68.
Banfield, R., Hall, L., Bowyer, K., Kegelmeyer, W., 2007. A comparison of decision tree ensemble creation techniques. IEEE Trans. Pattern Anal. Mach. Intell. 29, 173–180.
Boratynska, K., Boratynski, A., 2007. Taxonomic differences among closely related pines *Pinus sylvestris*, *P. mugo*, *P. uncinata* and *P. uliginosa* as revealed in needle sclerenchyma cells. Flora 202, 555–569.
Briem, G., Benediktsson, J., Sveinsson, J., 2007. Multiple classifier applied to multisource remote sensing data. IEEE Trans. Geosci. Remote Sens. 29, 173–180.
Calle, L., Urrea, V., Boulesteix, A., Malats, N., 2011. AUC-RF: a new strategy for genomic profiling with Random Forest. Hum. Heredity 72, 121–132.
Clark, M., Roberts, D., Clark, D., 2005. Hyperspectral discrimination of tropical rain forest tree species at leaf to crown scales. Remote Sens. Environ. 96, 375–398.
Clinton, N., Holt, A., Scarborough, J., Yan, L., Gong, P., 2010. Accuracy assessment measures for object-based image segmentation goodness. Photogramm. Eng. Remote Sens. 76, 289–299.
Coleman, T., Gudapati, L., Derrington, J., 1990. Monitoring forest plantations using Landsat Thematic Mapper data. Remote Sens. Environ. 33, 211–221.
Deng, F., Li, S., Su, G., 2007. A methodology for fusion LIDAR and digital images. Proc. SPIE 6795, Second International Conference on Space Information Technology, 67952C.
Duque-Lazo, J., Navaro-Cerrillo, R., 2017. What to save, the host or the pest? The spatial distribution of xylophage insects within the Mediterranean oak woodlands SouthWestern Spain. Forest Ecol. Manag. 392, 90–104.
Duque-Lazo, J., van Gils, H., Groen, T., Navarro-Cerrillo, R., 2016. Transferability of species distribution models: the case of *Phytophthora cinnamomi* in Southwest Spain and Southwest Australia. Ecol. Model. 320, 62–70.
Elith, J., Phillips, S., Hastie, T., Dudík, M., chee, Y., Yates, C., 2011. A statistical explanation of MaxEnt for ecologists. Divers. Distrib. 17, 43–57.
Engler, R., Waser, L., Zimmermann, N., Schaub, M., Berdos, S., Ginzler, C., Psomas, A., 2013. Combining ensemble modeling and remote sensing for mapping individual tree species at high spatial resolution. Forest Ecol. Manag. 310, 64–73.
Fahrmeir, L., Kneib, T., Lang, S., Marx, B., 2013. Regression: Models, Methods and Applications. Springer, Berlin, Heidelberg ISBN:978-3-642-34332-2.
Farjon, A., 2010. A Handbook of the World's Conifers. Koninklijke Brill, Leiden E-ISBN: 9789047430629.
Fassnacht, F., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L., Straub, C., Ghosh, A., 2016. Review of studies on tree species classification from remotely sensed data. Remote Sens. Environ. 186, 64–87.
Fauvart, N., Ali, A., Terral, J., Roiron, P., Blarquez, O., Carcaillet, C., 2012. Holocene upper tree-limits of *Pinus* section *sylvestris* in the western Alps as evidenced from travertine archives. Rev. Palaeobot. Palynol. 169, 96–102.
Gerard, F., North, P., 1997. Analyzing the effect of structural variability and canopy gaps on Forest BRDF using a geometric-optical model. Remote Sens. Environ. 62, 46–62.
González-Ferreiro, E., Diéguez-Aranda, U., Barreiro-Fernández, L., Buján, S., Barbosa, M., Suárez, J., Bye, I., Miranda, D., 2013. A mixed pixel- and region-based approach for using airborne laser scanning data for individual tree crown delineation in *Pinus radiata D. Don* plantations. Int. J. Remote Sens. 34, 7671–7690.
González-Ferreiro, E., Diéguez-Aranda, U., Crecente-Campo, F., Barreiro-Fernández, L., Miranda, D., Castedo-Dorado, F., 2014. Modelling canopy fuel variables for *Pinus radiata D. Don* in NW Spain with low-density LiDAR data. Int. J. Wildland Fire 23, 350–362.
Goodwin, N., Turner, R., Merton, R., 2005. Classifying eucalyptus forests with high spatial and spectral resolution imagery: an investigation of individual species and vegetation communities. Aust. J. Bot. 53, 337–345.
Gougeon, F., Leckie, D., 2006. The individual tree crown approach applied to ikonos images of a coniferous plantation area. Photogramm. Eng. Remote Sens. 72, 1287–1297.
Guillera-Arroita, G., Lahoz-Monfort, J., Elith, J., 2014. Maxent is not a presence–abscence method: a comment on Thibuad et al. Methods Ecol. Evol. 5, 1192–1197.
Hamraz, H., Contreras, M., Zhang, J., 2017. Forest understory trees can be segmented accurately within sufficiently dense airborne laser scanning point clouds. Sci. Rep. 7 (1), 2045–2322.
Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning, 2nd ed. Springer ISBN: 978-0-387-84857-0.
Ke, Y., Quackenbush, L., Im, J., 2010. Synergistic use of QuickBird multispectral imagery and LIDAR data for object-based forest species classification. Remote Sens. Environ. 114, 1141–1154.
Kim, M., McGaughey, R., Andersen, H., Schreuder, G., 2009. Tree species differentiation using intensity data derived from leaf-on and leaf-off airborne laser scanner data. Remote Sens. Environ. 113, 1575–1586.
Ko, C., Sohn, G., Remmel, T., Miller, J., 2014. Hybrid ensemble classification of tree genera using airborne LIDAR data. Remote Sens ISSN 2072-4292.
Korpela, I., Ørka, H., Hyyppä, J., Heikkinen, V., Tokola, T., 2010. Range and AGC normalization in airborne discrete-return LiDAR intensity data for forest canopies. ISPRS J. Photogramm. Remote Sens. 65, 369–379.
Leckie, D., Gougeon, F., Hill, D., Quinn, R., Armstrong, L., Shreenan, R., 2003. Combined high-density lidar and multispectral imagery for individual tree crown analysis. Can. J. Remote Sens. 29, 633–649.

Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R., Thuiller, W., 2009. Evaluation of consensus methods in predictive species distribution modelling. Divers. Distrib. 15, 59–69.
McGaughey, R., 2014. FUSION/LDV: software for lidar data analysis and visualization, v. 3.42. USDA Forest Service, Pacific Northwest Research Station, Seattle, WA.
Mo, D., Fuchs, H., Fehrmann, L., Yang, H., Lu, Y., Kleinn, C., 2015. Local parameter estimation of topographic normalization for forest type classification. IEEE Geosci. Remote Sens. Lett. 12, 1998–2002.
Ørka, H., Næsset, E., Bollandsås, O., 2009. Classifying species of individual trees by intensity and structure features derived from airborne laser scanner data. Remote Sens. Environ. 113, 1163–1174.
Peerbhay, K., Mutanga, O., Ismail, R., 2014. Does simultaneous variable selection and dimension reduction improve the classification of Pinus forest species? J. Appl. Remote Sens. 8, 085194-L.
Ross, J., Kellomäki, S., Oler-Blom, P., Ross, V., Vilikainen, L., 1986. Architecture of scots pine crown: phytometrical characteristics of needles and shoots. Silva Fennica 20, 91–105.
Suratno, A., Seielstad, C., Queen, L., 2009. Tree species identification in mixed coniferous forest using airborne laser scanning. ISPRS J. Photogramm. Remote Sens. 64, 683–693.
Swatantran, A., Dubayah, R., Roberts, D., Hofton, M., Blair, J., 2011. Mapping biomass and stress in the sierra nevada using lidar and hyperspectral data fusion. Remote Sens. Environ. 115, 2917–2930.
Thuiller, W., Georges, D., Engler, R., Breiner, F., 2016. biomod2: Ensemble Platform for Species Distribution Modeling. R package version 3.3-7. https://CRAN.R-project.org/package=biomod2.
Vauhkonen, J., Maltamo, M., McRoberts, R., Næsset, E., 2014. Introduction to forestry applications of airborne laser scanning. Springer, Dord ISSN: 1568-1319.
Wulder, M., Niemann, K., Goodenough, D., 2000. Local maximum filtering for the extraction of tree locations and basal area from high spatial resolution imagery. Remote Sens. Environ. 73, 103–114.
Youngentob, K., Roberts, D., Held, A., Dennison, P., Jia, X., Lindenmayer, D., 2011. Mapping two eucalyptus subgenera using multiple end member spectral mixture analysis and continuum-removed imaging spectrometry data. Remote Sens. Environ. 115, 1115–1128.
Zhang, Z., Lui, X., 2013. Support vector machines for tree species identification using LiDAR-derived structure and intensity variables. Geocarto Int. 28, 364–378.