Forest Ecology and Management 433 (2019) 111–121

Contents lists available at ScienceDirect

Forest Ecology and Management

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

A simple approach to forest structure classification using airborne laser scanning that can be adopted across bioregions

Syed Adnan[a,b,c,*], Matti Maltamo[a], David A. Coomes[b], Antonio García-Abril[d], Yadvinder Malhi[e], José Antonio Manzanera[d], Nathalie Butt[e,f], Mike Morecroft[g], Rubén Valbuena[b,*]

[a] University of Eastern Finland, Faculty of Forest Sciences, PO Box 111, FI-80101 Joensuu, Finland
[b] University of Cambridge, Department of Plant Sciences, Forest Ecology and Conservation, Downing Street, CB2 3EA Cambridge, UK
[c] National University of Sciences and Technology, Institute of Geographical Information Systems, 44000 Islamabad, Pakistan
[d] Universidad Politecnica de Madrid, College of Forestry and Natural Environment, Research Group SILVANET, Ciudad Universitaria, 28040 Madrid, Spain
[e] University of Oxford, School of Geography and the Environment, Environmental Change Institute, OX1 3QY Oxford, UK
[f] The University of Queensland, School of Biological Sciences, St. Lucia, Queensland 4072, Australia
[g] Natural England, Cromwell House, 15 Andover Road, SO23 7BT Winchester, UK

ARTICLE INFO

Keywords:
Structural heterogeneity
LiDAR
Nearest neighbour imputation
Classification and regression trees
Forest structural types

ABSTRACT

Reliable assessment of forest structural types (FSTs) aids sustainable forest management. We developed a methodology for the identification of FSTs using airborne laser scanning (ALS), and demonstrate its generality by applying it to forests from Boreal, Mediterranean and Atlantic biogeographical regions. First, hierarchal clustering analysis (HCA) was applied and clusters (FSTs) were determined in coniferous and deciduous forests using four forest structural variables obtained from forest inventory data – quadratic mean diameter (*QMD*), Gini coefficient (*GC*), basal area larger than mean (*BALM*) and density of stems (*N*) –. Then, classification and regression tree analysis (CART) were used to extract the empirical threshold values for discriminating those clusters. Based on the classification trees, *GC* and *BALM* were the most important variables in the identification of FSTs. Lower, medium and high values of *GC* and *BALM* characterize single storey FSTs, multi-layered FSTs and exponentially decreasing size distributions (reversed J), respectively. Within each of these main FST groups, we also identified young/mature and sparse/dense subtypes using *QMD* and *N*. Then we used similar structural predictors derived from ALS – maximum height (*Max*), L-coefficient of variation (*Lcv*), L-skewness (*Lskew*), and percentage of penetration (*cover*), – and a nearest neighbour method to predict the FSTs. We obtained a greater overall accuracy in deciduous forest (0.87) as compared to the coniferous forest (0.72). Our methodology proves the usefulness of ALS data for structural heterogeneity assessment of forests across biogeographical regions. Our simple two-tier approach to FST classification paves the way toward transnational assessments of forest structure across bioregions.

1. Introduction

The structural complexity of forest affects the growth rate of individual trees and the dynamics of tree communities (Donato et al., 2012). Knowledge of this structural variations is key to understand ecosystem functioning (Coomes and Allen, 2007a) and sustainable forest management planning (Bergeron et al., 2002). Accurate structural heterogeneity assessment and stand development categorization is important for long-term prediction of biomass production (Gove, 2004; Bourdier et al., 2016) and turnover (Marvin et al., 2014), biodiversity (Gove et al., 1995; Pommerening, 2002), and for identifying important habitats for wildlife (Vihervaara et al., 2015). It can also assist the planning and monitoring of different silvicultural regimes and forest management strategies (McElhinny et al., 2005; Valbuena et al., 2016a). Forest structure information may also be helpful to reduce sampling efforts and costs (Maltamo et al., 2010; Moss, 2012).

From an ecological point of view, forest structure is an important attribute at community level and consists of three major components:

* Corresponding authors at: University of Eastern Finland, Faculty of Forest Sciences, PO Box 111, FI-80101 Joensuu, Finland (S. Adnan). University of Cambridge, Department of Plant Sciences, Forest Ecology and Conservation, Downing Street, CB2 3EA Cambridge, UK (R. Valbuena).

E-mail addresses: adnan@uef.fi (S. Adnan), matti.maltamo@uef.fi (M. Maltamo), dac18@cam.ac.uk (D.A. Coomes), antonio.garcia.abril@upm.es (A. García-Abril), yadvinder.malhi@ouce.ox.ac.uk (Y. Malhi), joseantonio.manzanera@upm.es (J.A. Manzanera), n.butt@uq.edu.au (N. Butt), mike.morecroft@naturalengland.org.uk (M. Morecroft), rv314@cam.ac.uk (R. Valbuena).

https://doi.org/10.1016/j.foreco.2018.10.057
Received 18 July 2018; Received in revised form 24 October 2018; Accepted 25 October 2018
Available online 03 November 2018
0378-1127/

horizontal structure (spatial pattern, gaps and tree groups), vertical structure (number of tree layers) and species richness (O'Hara et al., 1996; Zimble et al., 2003; Pascual et al., 2008). However, unlike other forest attributes, forest structure lacks a clear and fixed definition, which thus varies from one application to another (Maltamo et al., 2005). Various approaches are found in the literature for identifying forest structural types (FSTs), such as stand developments classes (Valbuena et al., 2016a), patterns of growth and mortality (Coomes and Allen, 2007b), ecology of tree populations (O'Hara et al., 1996), stand age (O'Hara and Gersonde, 2004) or tree diameter distributions (Linder et al., 1997). There is also no consensus on the relevant classes to identify as FSTs, and thus a disparate number of them can be found, for example including understorey vegetation/regeneration (Gougeon et al., 2001), single storey to multi-storey structures (Zimble et al., 2003; O'Hara and Gersonde, 2004; Maltamo et al., 2005), suppressed tree storey (Hyyppä et al., 2008), young and mature stands (Means et al., 2000; Næsset, 2002), sparse and dense stands (Maltamo et al., 2004; Hyyppä et al., 2008) and reversed J-types of forest structures (Linder et al., 1997; Valbuena et al., 2013). There is also great disparity on the forest variables and indicators employed for quantitative assessment of structural heterogeneity (Lexerød and Eid, 2006; Valbuena et al., 2014) and FST categorization (Valbuena et al., 2013). Overall, FST definition and description may be dependent on the observer and thus there is a need to develop more objective quantitative approaches (e.g., Moss, 2012; Valbuena et al., 2013) that can be useful across biomes and bioregions. Here we propose a region-independent FST characterization by a combination of attributes describing tree diameter distribution – location, spread, skewness and density – using the following forest structural attributes: quadratic mean diameter (*QMD*), Gini coefficient (*GC*), basal area larger than mean (*BALM*) and density of stems (*N*).

The most common descriptors used to categorize forest dynamics and development are the *QMD* and *N* (Gove, 2004). The *QMD* can be described as the diameter of a tree having an average basal area and *N* is the number of stems per hectare (Curtis, 1982). These two parameters (*QMD* and *N*) are key to determine the need for planting or thinning in forest stands. Combinations of *QMD* and *N* are typically employed in the determination of forest development classes (e.g., Valbuena et al., 2016a), maximum stand density limits and occurrences of mortality in forest stands, impacts of habitat fragmentation on forest structure (Echeverría et al., 2007) and development of stand density diagrams (Newton, 1997; Gove, 2004).

The *GC*, an index of inequality widely used in econometrics has become popular in forest science due to its robust statistical properties and capacity to rank FSTs based on tree size variability (Lexerød and Eid, 2006; Duduman, 2011; Valbuena et al., 2012). It has been used to evaluate size inequality (Weiner, 1985), structural heterogeneity (Lexerød and Eid, 2006), successional stages (Duduman, 2011; Valbuena et al., 2013), relationship of relative dominance in forest stands (Valbuena et al., 2012) and to discriminate among differently-shaped diameter distributions (Bollandsås and Næsset, 2007; Valbuena et al., 2016b). Valbuena (2015) postulated that values of *GC* and *BALM* describe the spread and skewness of the tree size distribution, respectively, and that together they provide the best means of categorising FSTs (Gove, 2004; Valbuena et al., 2014). These FSTs can be analysed further to indicate whether trees interaction are dominated by symmetric competition associated with resource depletion, or asymmetric competition associated with resource pre-emption (Weiner, 1985). Although some theoretical values have been postulated discriminating FSTs from *GC* and *BALM* (Valbuena et al., 2013, 2014), there is a need to empirically investigate threshold values of *GC* and *BALM* in such categorization.

Airborne laser scanning provides an excellent means for forest structural heterogeneity assessment as the ALS data produce accurate canopy information (Maltamo et al., 2005; Valbuena et al., 2016b). Metrics derived from ALS height distribution describe the key characteristics of forest structure and could be used to monitor various aspects of forest dynamics (Jaskierniak et al., 2011; Valbuena et al., 2013). Numerous studies have used ALS data and demonstrated that it is a useful tool to characterize variation in forest structure (Maltamo et al., 2005; Pascual et al., 2008; Valbuena et al., 2017; Fedrigo et al., 2018). For this reason, it is important to find methodologies for prediction of FSTs from ALS which can be robust across ecoregions.

The objective of this research was to carry out a classification of FSTs using a combination of these four forest attributes – *QMD*, *GC*, *BALM* and *N* – postulating that together they can achieve a full description for forest structure where each FST contains a range of all possible horizontal and vertical structures. Using data from three different biogeographical regions – Boreal, Mediterranean and Atlantic –, we aimed at developing a region-independent methodology for FST characterization. We also evaluated the capacity of using ALS to achieve a reliable classification of those FSTs.

2. Material and methods

2.1. Study sites and data collection

Forest and ALS data from three biogeographical regions (Fig. 1) were used to identify, classify and predict FSTs:

(a) *Boreal: Kiihtelysvaara Forest, Finland*

Kiihtelysvaara forest is a common boreal managed forest located in the Eastern Finland (62°31′N, 30°10′E). The area is dominated by Scots pine with the presence of Norway spruce and deciduous species as minor tree species. The field data consisted of 79 squared plots collected during May-June 2010 (Maltamo et al., 2012). Plot size was 20 × 20 m, after some of them were subsampled from larger plots (Valbuena et al., 2014) with the intention to analyse a homogeneous dataset consistent with the other two regional sites involved in this study. The data included diameters and breast height (*dbh*) for all trees with a height greater than 4 m or *dbh* > 5 cm. A high resolution ALS dataset was acquired on June 26, 2009 using ATM Gemini sensor (Optech, Canada), Its scan density was 11.9 pulses·m^{-2} obtained from 600 to 700 m above ground level at a pulse rate of 125 kHz. Field of view (FOV) was 26° and scan swath was 320 m wide with a 55% side overlap between the strips.

(b) *Mediterranean: Valsaín Forest, Spain*

Valsain forest is a shelterwood managed (Valbuena et al., 2013) Scots pine area located in Segovia province, Spain (40°48′N 4°01′W), at 300–1500 m above sea level. The field data consisted of 37 circular plots with 20 m radius measured during summer 2006. All seedlings and saplings were measured within an inner 10 m radius subplot, whereas in the outer annulus only trees with *dbh* > 10cm were measured. ALS data were captured on September 2006 using an ALS50-II from 1500 m above ground level with a pulse rate of 55 kHz from Leica Geosystems (Switzerland). A FOV of 25° rendered a 665 m ground bi-directional scan width with 40% side lap. The average scan density of ALS data was 1.15 pulses·m^{-2}.

(c) *Atlantic: Wytham Woods, United Kingdom*

Wytham Woods is a managed lowland ancient woodland located in Oxfordshire, UK (51°46′N, 1°20′W). The dominant species are ash, sycamore as well as oak, hazel and maple trees (Savill et al., 2011). We used data from a permanent plot with a total area of 18 ha measured in 2010. The area of the permanent plot is further subdivided into 450 subplots sizing 20 × 20 m each. Field data included *dbh* of all stems greater than 1 cm. Leica ALS50-II LiDAR system with a 96.8 kHz pulse rate and 35° FOV was used from 2500 m above sea level for ALS data

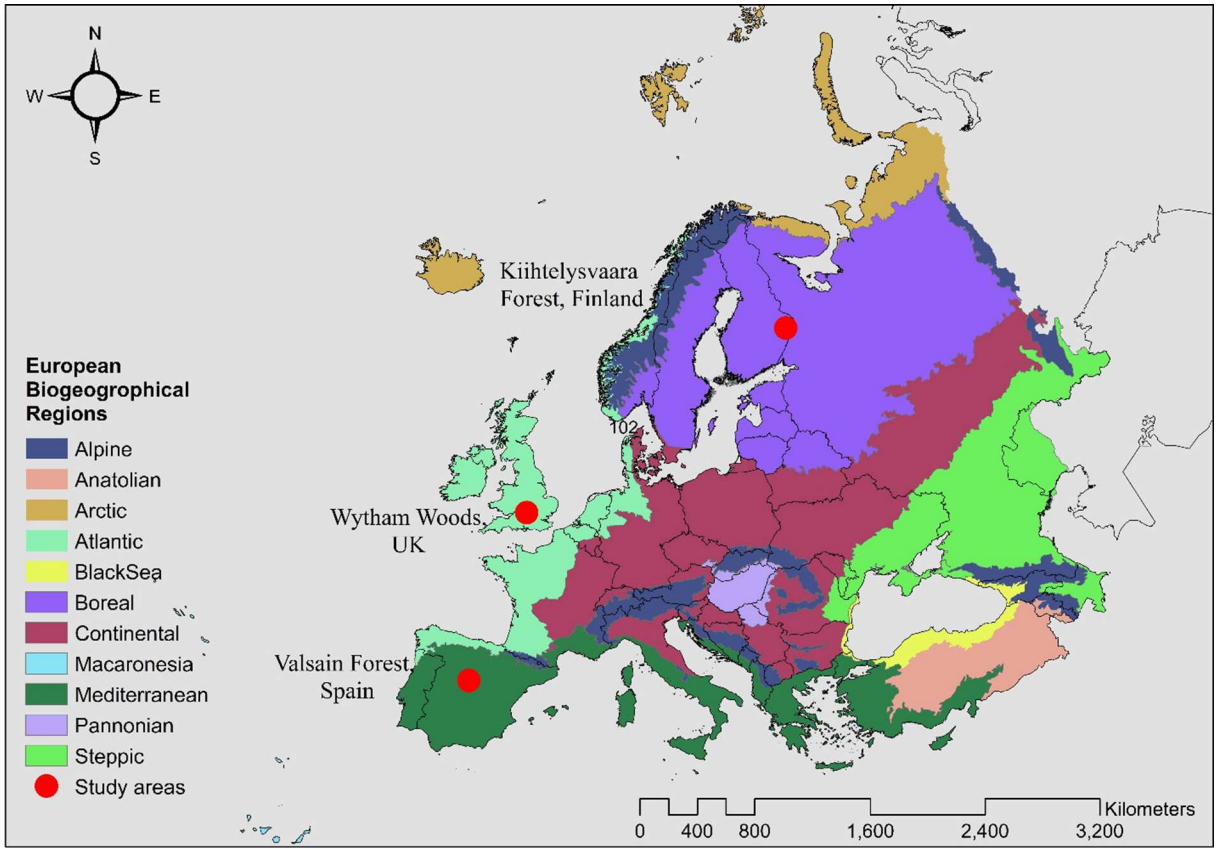

Fig. 1. Map of the European biogeographical regions and the study sites (Boreal/Finland, Mediterranean/Spain and Atlantic/UK) (European Environmental Agency, 2018).

acquisition and a low resolution ALS data of 0.918 pulses·m^{-2} density were acquired on June 24, 2014. Since growth is low in ancient woodlands and FST dynamics change slowly, the time differences between field and remote sensing acquisition can be assumed to have little effect in the classifications.

2.2. Data analyses

Forest stand attributes and characteristics were calculated by aggregating the tree-level information into per-hectare totals at plot-level (Tables 1a and 1b): we calculated quadratic mean diameter (*QMD*, cm), the Gini coefficient (*GC*) (Weiner, 1985), the proportion of basal area larger than the *QMD* (*BALM*) (Gove, 2004), and stem density (*N*, stems·ha^{-1}). The first task was to identify the potential clusters that could be rendered when using these four descriptors (*QMD*, *GC*, *BALM* and *N*). We grouped the data into coniferous (Boreal plus Mediterranean combined) and deciduous forests (Atlantic), after preliminary results showed that it was more convenient to carry out separate analyses for these two groups. The total number of field plots in the coniferous group was 116, and thus we randomly subsampled 116 out of 450 field plots from the deciduous group, to make further analysis consistent and obtain directly comparable results. Then, we applied hierarchical clustering analysis (HCA) to both coniferous and deciduous forest to optimize the clusters that can be rendered from the chosen forest attributes. The second task was to find the

Table 1b
Summary of study area properties.

Number of plots	Global 232			
Parameter	Min	Mean	Max	SD
QMD	10.23	21.49	48.3	8.76
GC	0.15	0.57	0.89	0.19
BALM	0.52	0.78	0.95	0.09
N	75	1146	3500	629

QMD: quadratic mean diameter (cm); *GC*: Gini coefficient; *BALM*:Basal area larger than mean; *N*: stand density (stems·ha^{-1}); SD: standard deviation.

Table 1a
Summary of study area properties.

Number of plots	Finland 79				Spain 37				UK 116			
Parameter	Min	Mean	Max	SD	Min	Mean	Max	SD	Min	Mean	Max	SD
QMD	10.33	16.87	29.26	4.09	14.5	33.13	48.3	12.3	10.23	20.92	46.3	6.06
GC	0.21	0.45	0.81	0.15	0.15	0.43	0.87	0.25	0.33	0.69	0.89	0.1
BALM	0.52	0.77	0.95	0.08	0.55	0.72	0.93	0.12	0.58	0.81	0.95	0.06
N	425	1288	2025	612	167	732	1918	559	75	1181	3500	609

QMD: quadratic mean diameter (cm); *GC*: Gini coefficient; *BALM*:Basal area larger than mean; *N*: stand density (stems·ha^{-1}); SD: standard deviation.

threshold values in both coniferous and deciduous forests which, when applied to *QMD*, *GC*, *BALM* and *N*, were best able to determine FSTs. This task was carried out using classification and regression trees (CART), which in this case were employed to classify the forest data into the clusters identified by the HCA analysis. The last task was to investigate the reliability of the FST classification obtained from ALS. The ALS classification was carried out using nearest neighbour (kNN) imputation method. The FSTs identified as a result of the HCA were employed as response variable in the kNN. All analyses were carried out using the R environment (R Core Team, 2018).

2.2.1. Hierarchical clustering analysis

HCA consists of a series of successive merging (agglomerative method) or splitting (divisive method) steps of individual observations based on proximity measures (similarity, dissimilarity or distance) and is used to determine meaningful clusters in a large group of data. We calculated the most widely used proximity measure, which is the Euclidian distance:

$$d_{kl} = \sqrt{\sum_{m=1}^{p} (X_{km} - X_{lm})^2}, \quad (1)$$

where d_{kl} is the Euclidian distance between two individual cases k and l in a m-dimensional space (of $m = 1, 2 \cdots p$ variables), and X_{km} and X_{lm} are their values of the m^{th} variable. Since *QMD*, *GC*, *BALM* and *N* were measured in different units, calculating the proximity measure d_{kl} directly on their original scales would unfairly weight some variables over others. To deal with this contingency, we applied a standardization of the raw variables prior to Euclidian distance calculation. We chose a range-equalization method. Thus, each variable value X was normalized to a scale 0–1, according to their empirical minimum (X_{min}) and maximum (X_{max}) values (Table 1):

$$Z = (X - X_{min})/(X_{max} - X_{min}), \quad (2)$$

Then, one of the most challenging stages in clustering analysis is the need to determine an optimal number c of clusters because the HCA may run until a single cluster containing all observations (agglomerative method) or c number of clusters each containing one observation (divisive method) are produced (Everitt et al., 2011). We used a distortion curve to choose the optimum number c of clusters (Sugar and James, 2003), since it shows the evolution of within-cluster sum of squares for increasing number of clusters. Thereafter, we used function *hclust* included in package *fastcluster* for HCA (Müllner, 2013), applied the agglomerative procedure included in the function and divided the data into the required optimum number c of clusters (FSTs).

2.2.2. Classification and regression tree (CART) analysis

After obtaining the HCA results and defining the FSTs that can be identified, we were interested to find out empirical threshold values for the chosen forest attributes (*QMD*, *GC*, *BALM* and *N*) that can be used to separate different FSTs. To answer this question, we used CART analysis which is a commonly used statistical modelling to identify important ecological patterns (Breiman et al., 1984; Lawrence and Wright, 2001). For the CART analysis, we employed the package recursive partitioning and regression trees (*rpart*; Breiman et al., 1984), where the HCA results (clusters) were the response variable and *QMD*, *GC*, *BALM* and *N* the explanatory variables. The *QMD* and *N* were log-normalized to avoid the high skewness of their distributions and make them approximately normal. CART resolved values among the explanatory variables that minimize the unexplained variance in response variable, the HCA clusters in this case, recursively splitting the data into those clusters/FSTs. Since the process is recursive, the result resembles a tree where each split is a node with a classification decision between two branches. A large tree was first produced, which was later pruned back to a desired size using function *prune* included in the package *rpart*, and we made it coincide with the optimal c decided upon at the HCA stage.

2.2.3. Classification of forest structural types from ALS datasets

Supervised machine learning methods use a set of features to generalize phenomena observed from a sample or describe the relationships among indicators. Examples of these methods include maximum likelihood classification, nearest neighbour imputation, artificial neural networks, random forest, support vector machine and naïve Bayes classifier (see e.g. Hastie et al., 2009). In the case of ALS, metrics that describe the distribution of ALS return heights over the forest plots are used to predict a FST that corresponds to each of them (Valbuena et al., 2017; Adnan et al., 2017). These ALS metrics are then employed as auxiliary variables to make a prediction throughout the scanned area (Næsset, 2002; Maltamo et al., 2006). In this study, kNN method of package *class* (Venables and Ripley, 2002) was used for prediction because of its simplicity and capacity to model complex covariance structures. This method has been successfully employed for predicting stand density, volume and cover types (Franco-Lopez et al., 2001). kNN classification is based on dissimilarity measures that are computed as a statistical distance to a reference sample plot in a feature space (Kilkki and Päivinen, 1987), just like those explained for HCA (Eq. (1)). The ALS metrics used in the kNN were the maximum of (*Max*) of ALS return heights over an area, the L-coefficient of variation (*Lcv*), L-skewness (*Lske*), and the percentage of all returns above 0.1 m (*Cover*), because of their high correlation with the chosen forest attributes (Lefsky et al., 2005; Valbuena et al., 2017). The *Max* could be related to *QMD* because of a strong tree diameter-height relationship (Enquist and Niklas, 2001; Sumida et al., 2013) and *Cover* is useful to characterize the stand density (Lefsky et al., 2005; Görgens et al., 2015). Similarly, *Lcv* and *Lske* are related to tree dominance (*GC* and *BALM*) and can be used to detect tree size inequality and light availability (Valbuena et al., 2017; Moran et al., 2018). Description of these ALS metrics and their related proxy forest characteristics are also described in Table 2.

For accuracy assessment we used a leave-one-out cross-validation, which consisted in eliminating each sample plot from the training data before fitting a separate nearest neighbour model for predicting it. *CrossTable* function of package *gmodels* (Warnes, 2013) was used to elaborate a detailed accuracy assessment of the cross-validated contingency metrics and infer their statistical significance. Bias towards each given FST was assessed as the difference between producer's and user's accuracies. Producer's accuracy for a given FST was calculated as the proportion of the observed field plots for that FST which were correctly classified, whereas its user's accuracy was the proportion of field plots being classified as that FST which were correct (Story and Congalton, 1986). To evaluate the degree of misclassification, we calculated the overall accuracy (OA) and kappa coefficient (κ) included in the package *vcd* (Meyer et al., 2014).

3. Results

3.1. Classification of field data into homogeneous clusters

The first step was to determine a statistical optimal number of clusters for the HCA, which was found to be $c = 5$ for both the coniferous and deciduous groups, when the decrease in within-cluster variation stabilized after high decreases along the range $c = 1$-5. In the next step, we identified the threshold values for each explanatory variable – *QMD*, *GC*, *BALM* and *N* – using CART. Each node

Table 2
Description of ALS metrics and their related forest characteristics.

Symbol	Description	Proxy forest characteristic
Max	Maximum height of ALS metrics	Dominant tree height
Lcv	L-coefficient of variation	Tree size inequality
Lske	L-skewness of ALS metrics	Competitive dominance
Cover	Percentage of all returns above 0.1 m	Canopy cover

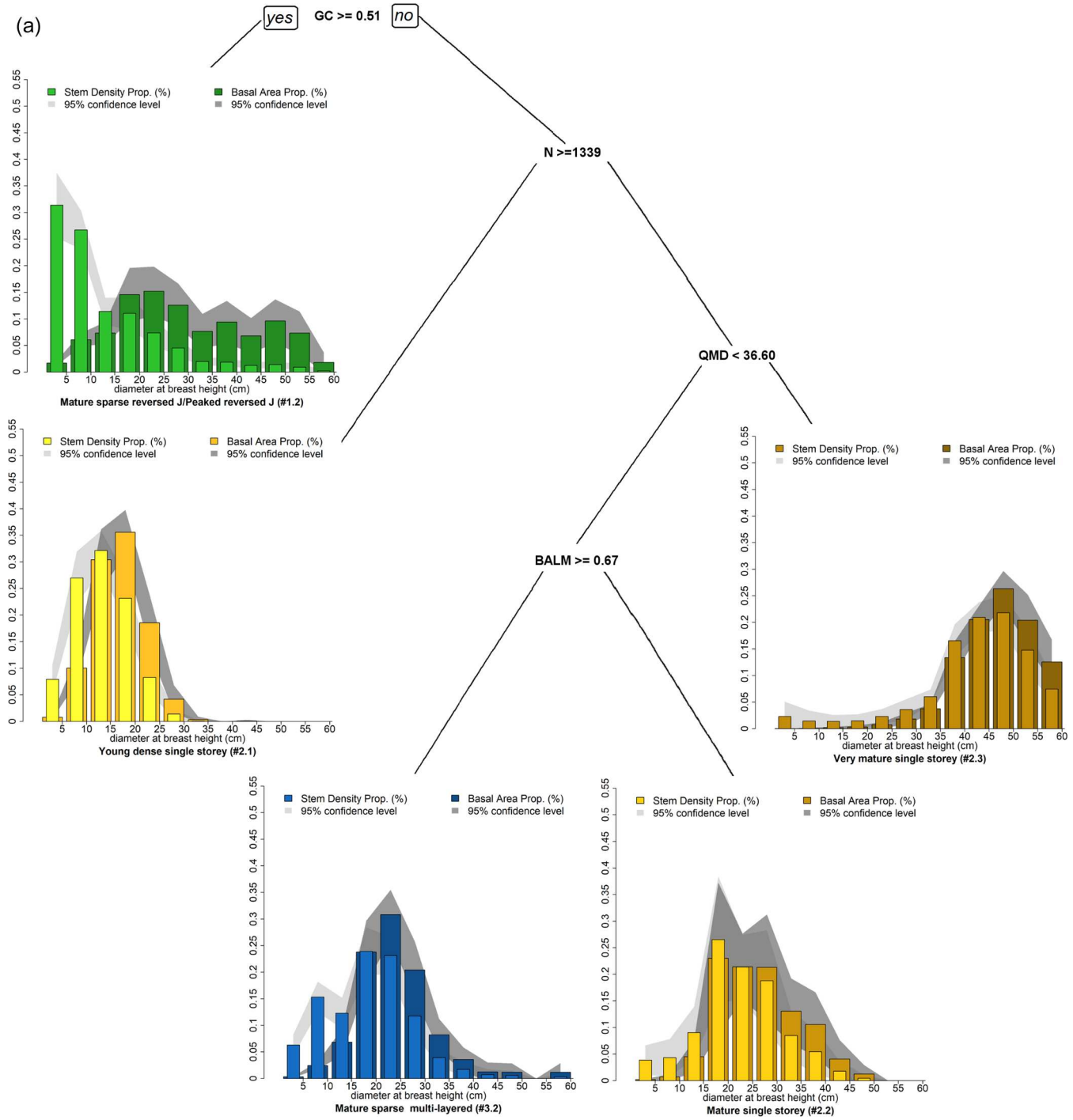

Fig. 2. Classification tree based on the field data from (a) coniferous forest and (b) deciduous forest as a result classification and regression tree (CART) analysis. Threshold values of the explanatory variables recursively divide the data into homogeneous clusters at each node, according to whether they meet the criterion (positive to the left and negative to the right). Each cluster is then classified as a forest structural type (FST) according to criteria in Table 3 and their diameter distributions (stem density and basal area proportions, and their 95% confidence intervals are shown). *QMD*: quadratic mean diameter (cm); *GC*: Gini coefficient; *BALM*: basal areal larger than mean; and *N*: stand density (stems·ha^{-1}).

maximized the between-cluster explained variability, and thus their order shows the importance of each variable in determining the FSTs (Fig. 2a and b). In coniferous forest, the first cluster (having lowest within-group variability) was produced by $GC \geq 0.51$ (Fig. 2a) and, in deciduous forest, it was produced by $BALM > 0.87$ (Fig. 2b). This iterative procedure was applied on either sides of the classification trees and at the end, clusters with lowest within-cluster variability were produced.

3.2. Identification of forest structural types

The threshold values obtained from each classification tree (Fig. 2a and b) were used in the identification of the FSTs, which we assigned after inspecting simultaneously their diameter and basal area-weighted distributions (these are the proportions per diameter class of the total number of stems and basal area, respectively). Table 3 summarizes the characteristics of each FST, and Fig. 3 shows the scatterplots which were also useful for the identification of relevant FSTs.

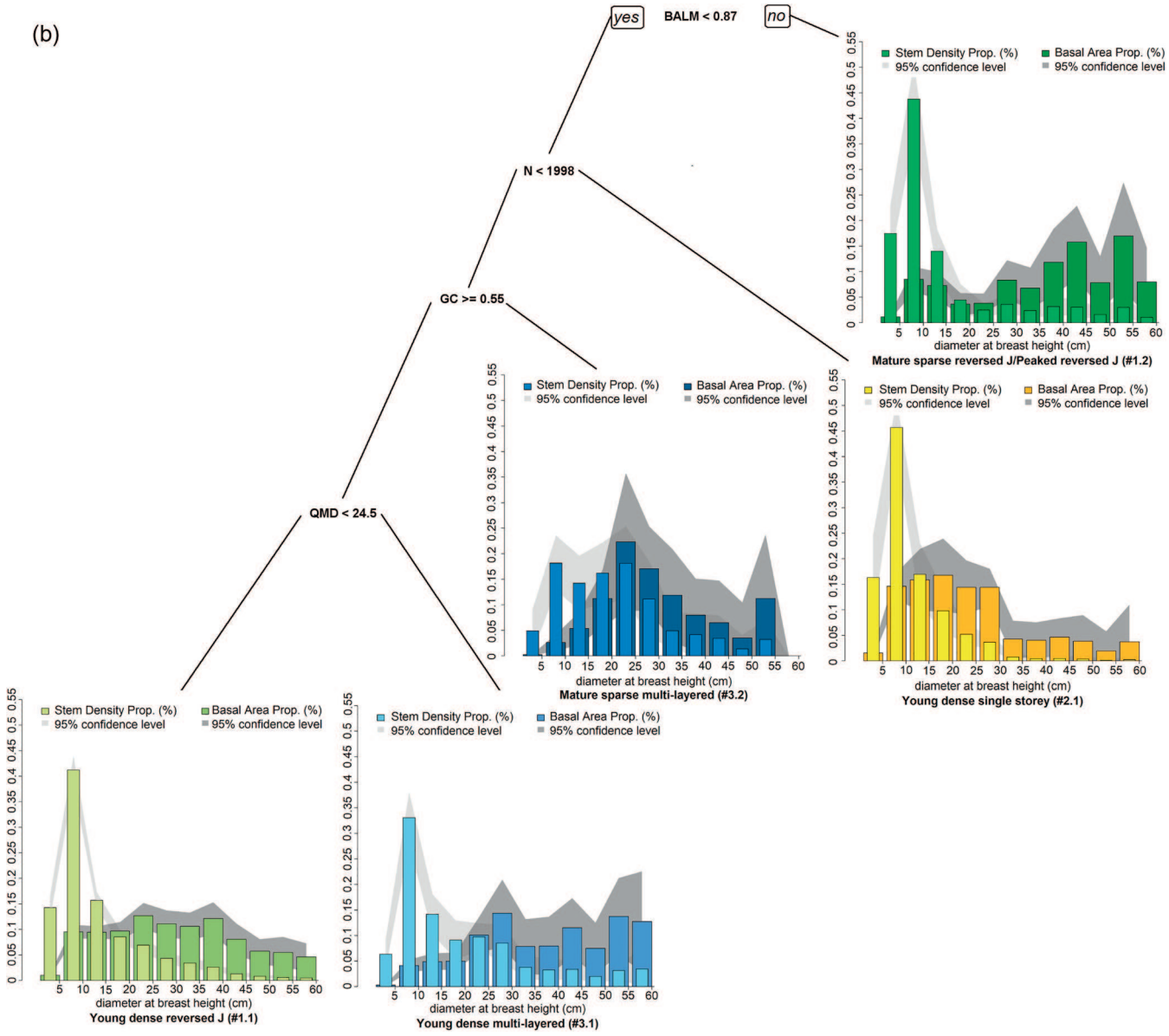

Fig. 2. (*continued*)

Fig. 2a shows the classification tree and diameter distribution of each FSTs found in the coniferous forest group. Higher *GC* values (≥0.51) produced a neat segregation of reversed J distributions from single storey and multi-layered types. This first cluster corresponded to mature sparse reversed J, commonly called peaked reversed J (FST #1.2) because they are characterized by a peak at the right end of their distribution where very big trees take a large proportion of the total basal area (which is best appreciated from the basal area-weighed distributions in Fig. 2a and b) The next node identified young forests by their high density of stems ($N > 1339$ trees·ha^{-1}), which in this case was a young dense single storey FST (*#2.1*). Then the threshold regarded the distinction of very mature single storey (*#2.3*) identified by a high *QMD* > 36.6cm. The last node separated mature sparse multi-layered FST (*#3.2*) areas from mature single storey FST (*#2.2*) by *BALM* ≥ 0.67.

Fig. 2b shows the classification tree and diameter distribution of

Table 3
Denomination of FSTs based on the four forest structural variables and their characteristics.

FST#	Denomination	Characteristics
#1.1	Young dense reversed J	High *GC*, medium/high *BALM*, high *N*, and low *QMD*
#1.2	Mature sparse reversed J (Peaked reversed J)	High *GC*, high *BALM*, medium/low *N* and high *QMD*
#2.1	Young dense single storey	Medium *GC*, medium *BALM*, high *N* and low *QMD*
#2.2	Mature single storey	Low *GC*, low *BALM*, medium *N* and medium *QMD*
#2.3	Very mature single storey	Low *GC*, medium/low *BALM*, low *N* and high *QMD*
#3.1	Young dense multi-layered	Medium *GC*, medium *BALM*, low *N* and high *QMD*
#3.2	Mature sparse multi-layered	Medium *GC*, medium *BALM*, medium *N* and medium *QMD*

QMD: quadratic mean diameter (cm); *GC*: Gini coefficient; *BALM*: Basal area larger than mean; *N*: stand density (stems·ha^{-1}).

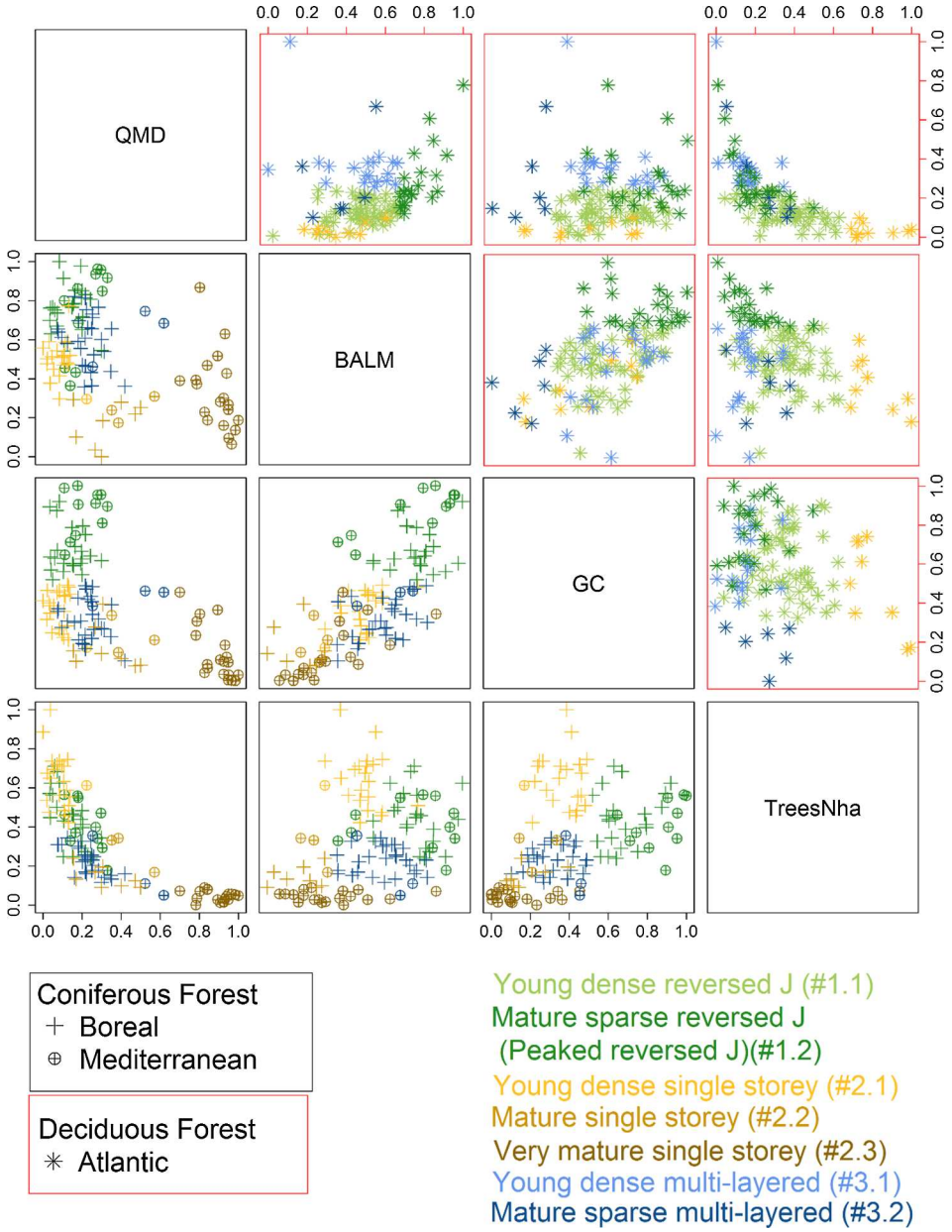

Fig. 3. Scatterplot showing five clusters/FSTs in each coniferous and deciduous forest. Axes show the normalized variable values (Eq. (2)). (*QMD*: quadratic mean diameter (cm);*BALM*:Basal area larger*QMD*; *GC*: Gini coefficient; *N*: stand density (stems·ha^{-1})).

each FSTs found in deciduous forest. High values of *BALM* (> 0.87) separated mature sparse reversed J (*#1.2*). As in the coniferous group, the next node identified young dense single storey (*#2.1*) forests by their high stand density, $N > 1998$ trees·ha^{-1} in this case. The next node found the threshold value of $GC < 0.55$ to identify mature sparse multi-layered (*#3.2*) areas. Higher values of *GC* were found for the remaining FSTs, young dense multi-layered (*#3.1*) and young dense reversed J (*#1.1*), the latter identified by their lower *QMD* > 24.5cm.

The scatterplot distribution of all FSTs in the feature space of *QMD*, *GC*, *BALM* and *N* (Fig. 3) showed that some FSTs are clearly distinct while others present some degree of overlap. The most relevant relationships were found in the cluster disaggregation observed on the $GC - BALM$ feature space, whereas the more traditional $QMD - N$ comparison can be useful to identify young/dense and mature/sparse sub-types.

Tables 4a and 4b describes the statistical properties of each FST, where young dense reversed J FST (*#1.1*) and mature sparse reversed J (*#1.2*) were found to be the most frequent FSTs with 29.3% and 51.7% observations in deciduous and coniferous forests, respectively. Fig. 4 shows the thematic map at the permanent plots in Wytham forest, illustrating the natural spatial distributions of the resulting FSTs.

3.3. Prediction of forest structural types from ALS datasets

Table 5 shows the cross-validated results of the kNN predictions of FSTs from ALS datasets of coniferous forest. Mature sparse reversed J/peaked reversed J (*#1.2*) was accurately predicted. Young dense single storey (*#2.1*) and mature single storey (*#2.2*) were slightly underestimated due to a high confusion with mature sparse multi-layered (*#3.2*), which was in turn slightly overestimated. Very mature single storey (*#2.3*) was also slightly overestimated. The overall accuracy of the classification was OA = 0.73 and $\kappa = 0.64$.

The results for kNN classification in deciduous forest are shown in Table 6. All reversed J diameter distributions were very accurately estimated, both the young dense (*#1.1*) and mature sparse (*#1.2*) reversed J subtypes. The remaining also obtained unbiased predictions,

Table 4a
Total number of observations (field plots) and statistical properties of each forest structural type in coniferous group.

Forest Structural Types (FST)		#1.1	#2.1	#2.2	#2.3	#3.2
Number of Observations		34	22	11	19	30
QMD	Minimum	11.62	10.33	16.22	36.97	13.22
	Maximum	22.84	18.81	32.02	48.3	33.86
	Mean	16.60	13.60	23.72	44.14	19.86
	SD	3.20	1.98	4.98	3.08	4.43
GC	Minimum	0.52	0.27	0.21	0.15	0.22
	Maximum	0.86	0.5	0.43	0.47	0.5
	Mean	0.68	0.41	0.28	0.25	0.38
	SD	0.10	0.07	0.08	0.10	0.08
BALM	Minimum	0.67	0.64	0.52	0.54	0.67
	Maximum	0.95	0.85	0.65	0.89	0.87
	Mean	0.84	0.73	0.60	0.66	0.78
	SD	0.07	0.04	0.04	0.09	0.06
N	Minimum	676	1375	425	167	310
	Maximum	2200	3025	1146	421	1185
	Mean	1402	2000	682	293	805
	SD	383	403	257	74	229

QMD: quadratic mean diameter (cm); *GC*: Gini coefficient;*BALM*: basal areal larger than mean; *N*: stand density (stems·ha^{-1}); SD: standard deviation.
#1.2: mature sparse reversed J/peaked reversed J; *#2.1*: young dense single storey; *#2.2*: mature single storey; *#2.3*: very mature single storey; *#3.2*: mature sparse multi-layered.

Table 4b
Total number of observations (field plots) and statistical properties of each forest structural type in deciduous group.

Forest Structural Types (FST)		*#1.1*	*#1.2*	*#2.1*	*#3.1*	*#3.2*
Number of Observations		60	22	9	19	6
QMD	Minimum	12.2	17.96	11.93	24.95	17.00
	Maximum	24.17	51.27	16.89	62.46	45.70
	Mean	18.52	26.81	13.80	30.50	25.60
	SD	3.02	8.19	1.54	8.09	10.88
GC	Minimum	0.56	0.64	0.48	0.60	0.40
	Maximum	0.86	0.91	0.78	0.85	0.54
	Mean	0.71	0.80	0.65	0.72	0.49
	SD	0.08	0.09	0.12	0.08	0.06
BALM	Minimum	0.64	0.87	0.70	0.63	0.69
	Maximum	0.87	0.99	0.85	0.87	0.83
	Mean	0.80	0.91	0.76	0.80	0.76
	SD	0.05	0.03	0.05	0.07	0.05
N	Minimum	275	175	2150	150	300
	Maximum	1975	1600	3050	1150	1250
	Mean	1181	699	2494	598	871
	SD	389	332	350	246	363

QMD: quadratic mean diameter (cm); *GC*: Gini coefficient;*BALM*: basal areal larger than mean; *N*: stand density (stems·ha^{-1}); SD: standard deviation.
#1.1: young dense reversed J; *#1.2*: Mature sparse reversed J (Peaked reversed J); *#2.1*: young dense single storey; *#3.1*: young dense multi-layered; *#3.2*: mature sparse multi-layered.

although with lesser accuracy in the estimation following this order: young dense single storey (*#2.1*) and multi-layered (*#3.1*), and mature sparse multi-layered (*#3.2*) being the least accurately estimated because it was the least frequent FST. The prediction was overall fairly unbiased, with OA = 0.87 and κ = 0.81.

4. Discussion

In this article we present a two-tier methodology for forest structure classification. The higher tier consists in using values of *GC* and *BALM* to characterize reversed J (exponentially decreasing size distributions), single storey and multi-layered. In a lower tier, *QMD* and *N* were used

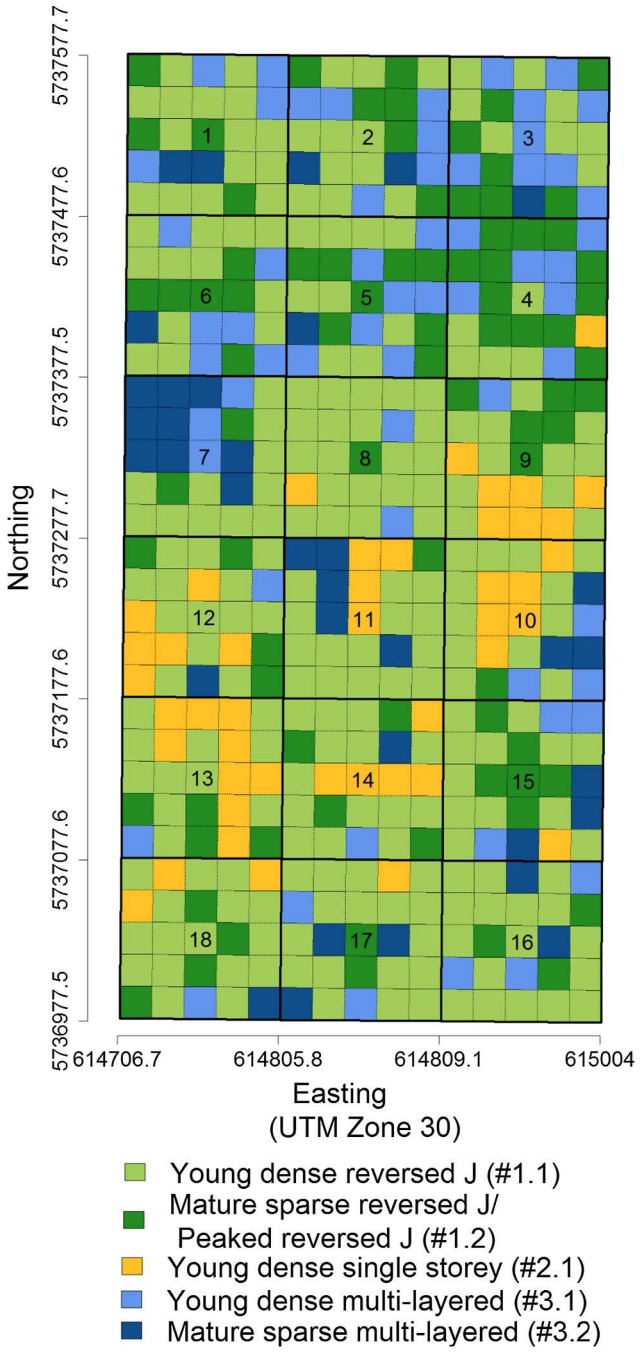

Fig. 4. Thematic map showing the natural spatial distribution of the forest structural types based on classification tree and field data from deciduous forest (Atlantic biogeographical region) at Wytham Woods (UK) permanent experimental plots.

to discriminate young/mature and sparse/dense subtypes for each of those described for the higher tier. These FSTs can provide important ecological information about natural dynamics – competitive (self) thinning, mature thinning, and disturbances – (Coomes and Allen, 2007a), or help in identifying where these dynamics have been artificially modified (Valbuena et al., 2016b). In that same order, they also show a degree in tree community development between those ecosystems following metabolic scaling (Enquist and Niklas, 2001) to those regulated by demographic equilibrium (Muller-Landau et al., 2006). The simplicity of this two-tier approach to FSTs makes it feasible for its adoption across ecoregions.

The proposed FST classification method has purposely been

Table 5
Nearest Neighbour imputation contingency table (coniferous forest: Boreal/Kiihtelysvaara Forest, Finland and Mediterranean/Valsaín Forest, Spain). User's and producer's accuracies are calculated over row and column totals, respectively.

Predicted	Observed					
	#1.2	*#2.1*	*#2.2*	*#2.3*	*#3.2*	User's Accuracy
#1.2	**26**	7	0	0	1	0.76
#2.1	4	**11**	0	0	1	0.69
#2.2	0	0	**3**	0	3	0.50
#2.3	4	0	0	**19**	0	0.83
#3.2	0	4	8	0	**25**	0.68
Producer's Accuracy	0.76	0.50	0.27	1.00	0.83	

#1.2: mature sparse reversed J/peaked reversed J; *#2.1*: young dense single storey; *#2.2*: mature single storey; *#2.3*: very mature single storey; *#3.2*: mature sparse multi-layered.

Table 6
Nearest Neighbour imputation contingency table (deciduous forest: Atlantic biogeographic region/Wytham woods, UK). User's and producer's accuracies are calculated over row and column totals, respectively.

Predicted	Observed					
	#1.1	*#1.2*	*#2.1*	*#3.1*	*#3.2*	User's Accuracy
#1.1	**40**	0	2	1	0	0.93
#1.2	0	**41**	0	2	2	0.91
#2.1	0	0	**5**	1	0	0.83
#3.1	1	1	0	**10**	2	0.71
#3.2	0	1	0	2	**5**	0.62
Producer's Accuracy	0.98	0.95	0.71	0.62	0.56	

#1.1: young dense reversed J; *#1.2*: mature sparse reversed J (peaked reversed J); *#2.1*: young dense single storey; *#3.1*: young dense multi-layered; *#3.2*: mature sparse multi-layered.

designed to allow its general application for FSTs other than those present in the case studies shown hereby. The higher tier was proposed by Valbuena (2015) as a comprehensive bivariate description of forest structure, more meaningful than recovering parameters of diameter distributions (Gove, 2004; Lexerød and Eid, 2006). The addition proposed in this article is to include a lower tier of classification, using *QMD* and *N* to attain a greater span of possibilities with FST subtypes according to the stage of development and density of forests. The Valsain site was designed to cover a wide range of plausible FSTs (Valbuena et al., 2012), some occurring by natural dynamics and others driven by management (Valbuena et al., 2013), while those in Finland are highly managed forests (Valbuena et al., 2014, 2016a). With the inclusion of results from Whytham Woods, we have also extended the empirical evidence previously shown for conifers. The two-tier method should also be largely independent of the sampling design employed. Any effects due to changes in plot size, sampling design, minimum *dbh*, etc., would be unimportant whenever the field data can be employed as good estimators of the variables in hand: *QMD*, *GC*, *BALM* and *N*. The estimation of variables like *QMD* and *N* is well studied, and known unbiased when plots are allocated by simple random sampling. Adnan et al. (2017) showed that the effects of plot size on *GC* estimation are negligible for the plot sizes involved in this study. To the best of our knowledge no studies have tackled with *BALM* estimators, but similar assumptions may be presumed as per its relationship to the basal area and *QMD* (Gove, 2004). Moreover, these effects on variable estimators lessen when propagated toward FST classification, because only values trespassing thresholds have a practical effect. For the purpose of our study we shall assume that the plots are good estimators of the population values for these variables, and thus changes in the CART thresholds among classes due to these effects are only marginal.

We used HCA and CART to identify different FSTs using the four forest variables (*QMD*, *GC*, *BALM* and *N*). HCA is a widely used unsupervised statistical method to classify a large group of observations into several clusters according to similarity, dissimilarity or distance among individual observations (Bien and Tibshirani, 2011). On the other hand, CART is a statistical technique for selecting those variables and their interactions that are most important in determining an outcome or dependent variable (Breiman et al., 1984). We also used the kNN method (Venables and Ripley, 2002) to predict those FSTs obtained from ALS datasets (Kim et al., 2009). All these were applied to data from three biogeographical regions: Boreal, Mediterranean and Atlantic.

There was an interest in exploring empirical threshold values of the four forest variables (*QMD*, *GC*, *BALM* and *N*), and we used the CART analysis for this purpose (Breiman et al., 1984). Fig. 2a and b show these threshold values at each node for classifying into FSTs. The first nodes were based on *GC* and *BALM* (Gove, 2004; Lexerød and Eid, 2006; Valbuena, 2015), which indicates the importance of these two parameters in the disaggregation of the higher tier in FSTs classification (Fig. 3; $BALM - GC$ feature space). The empirical results yielded values of $GC = 0.51$ and $GC = 0.55$ (Fig. 2a and b), which were both very close to the theoretical value at $GC = 0.5$ envisaged by Valbuena et al. (2012) as a beacon for maximum entropy. Multi-layered FSTs are thus signalled around this value, while values below/above must necessarily denote diameter distributions close to Gaussian/negative exponential, respectively. These values were roughly consistent with previous results obtained by Duduman (2011) and Valbuena et al. (2013). Our results from deciduous forest are also similar to those obtained by Simpson et al. (2017) from the same area, however, they used vertical gap probability (proportion ALS returns at specific heights) for structural classification. On the other hand, there was a lack of previous studies analysing empirical values for *BALM* at different FSTs (Valbuena, 2015). One very relevant result was the peaked reversed J diameter distributions (*#1.2*) which can be identified by large values of $BALM > 0.87$ (Fig. 2b). This FST was characterized by two distinctive storeys – one mature and spare trees accompanied by dense young ingrowth in the understorey –. Conversely, low values of $BALM < 0.67$ (Fig. 2a) may indicate the presence of forest ecosystems with very closed canopies and competitive conditions dominated by mature thinning, hence denoting single storey FSTs. Thus, *BALM* was chosen by the CART algorithm to separate single storey (with lower *BALM*) and multi-layered (with medium/higher *BALM*) (Gove, 2004).

The more traditionally used forest variables, *QMD* and *N*, were useful to identify lower-tier sub-types: young/mature and dense/sparse FSTs, respectively (Dodson et al., 2012). CART analysis effectively separated the very mature single storey FST (*#2.3*) in coniferous forest (Fig. 2a) which contained very mature trees (above 100 years old) from Valsaín forest (Spain) as a result of group shelterwood forest management based on long rotation periods (Valbuena et al., 2013). The statistical properties of these FSTs are given in Tables 4a and 4b, wherein, young dense reversed J FST (*#1.1*) and mature sparse reversed J/peaked reversed J (*#1.2*) had the largest number of individual observations in deciduous and coniferous forests, respectively. The performance of the clustering analysis can also be appreciated in the scatterplot distribution in the feature space of *QMD*, *GC*, *BALM* and *N* (Fig. 3). The widest separation among FSTs was found in the $GC - BALM$ feature space (Gove, 2004) which showed that the *GC* and *BALM* are the best indicators in FSTs classifications, as postulated by Valbuena (2015).

ALS is a useful tool for the structural heterogeneity assessment (Zimble et al., 2003; Lefsky et al., 2005; Marvin et al., 2014) and mapping of broad forest areas (Asner and Mascaro, 2014). Our results for predicting FSTs from ALS dataset are shown in Tables 5 and 6. Generally, unbiased estimations were found in both groups and the observed errors were mostly between FSTs that were, structurally speaking, close to one-another. The highest confusion was found in

misclassifying mature single storey (*#2.2*) as mature sparse multi-layered (*#3.2*). These two classes were the most loosely discriminated ones from the forest variables themselves (Fig. 2a), and thus it was not surprising that they showed worse results in their ALS prediction. Such narrow differences and misclassifications are less important because classifying a mature single storey (*#2.2*) as mature sparse multi-layered (*#3.2*) would have a lesser impact in terms of forest management and practical decision-making than a misclassification as a young dense reversed J (*#1.1*). We obtained a greater overall accuracy and kappa coefficient in the deciduous forest (0.87 and 0.81) as compared to the coniferous forest (0.73 and 0.64), which can be simply due to the differences in the ALS datasets employed in the coniferous group. These accuracies obtained, however, show that the methodology may reliably be applied to disparate ALS datasets surveyed at diverse ecoregions and forest types.

The analysis and classification of forest structural types proposed here is of interest for the conservation and promotion of biodiversity, prevention of natural disasters and other ecosystem services. Therefore, forest and natural area managers, nature conservation bodies, landscape planning and ecotourism stakeholders are among the activities and professionals potentially interested in the application of our methodology. Furthermore, this methodology is well adapted to monitor changes over space and time, as it is based on remote sensors such as LiDAR, which is nowadays used for great extensions and even for nation-wide area coverage. The approach presented in this article could, thanks to its simplicity, be adopted at many different forest types across all geographical zones. It could thus be beneficial for international efforts for harmonizing national forest inventories, initialized by the COST Action E43 (COST, 2006; McRoberts et al., 2008, 2012). At pan-European level it could, for instance, contribute to further developments in the ICP Forests, which is International Co-operative Programme on Assessment and Monitoring of Air Pollution Effects on Forests (JRC, 2011; Giannetti et al., 2018). More globally, it could assist the development of essential biodiversity variables from ALS (Pereira et al., 2013; Proença et al., 2017), and contribute to the use of remote sensing to inform policy-makers on progress towards sustainable development goals and biodiversity targets (O'Connor et al., 2015; Vihervaara et al., 2017).

5. Conclusions

In this research, we developed a region-independent methodology for forest structural types assessment, and demonstrated its utility by using disparate datasets from three biogeographical regions – Boreal, Mediterranean and Atlantic –. The methodology is a simple two-tier approach, feasible for its adoption across ecoregions. We separated FSTs at coniferous (Boreal plus Mediterranean combined) and deciduous (Atlantic) forests, using four forest variables – *QMD*, *GC*, *BALM* and *N* – and found empirical threshold values for using them in the identification of different FSTs. We found that the *GC* and *BALM* are the most important variables in the identification of a higher tier of FSTs: reversed J, single storey and multi-layered. Furthermore, a lower tier young/mature and sparse/dense sub-types can be further identified using *QMD* and *N*. We also used nearest neighbour imputation method and the FSTs identified from field data were predicted from ALS data. In spite of using very disparate ALS surveys, the results yielded reliable FST classification. The simplicity of this approach paves the way toward transnational assessments of FSTs across bioregions.

Acknowledgements

This article summarizes the main results of project LORENZLIDAR: Classification of Forest Structural Types with LiDAR remote sensing applied to study tree size-density scaling theories, funded by a Marie S. Curie Individual Fellowship H202-MSCA-IF-2014 (project 658180). Syed Adnan's PhD is funded by National University of Sciences and Technology (NUST), Pakistan under FDP 2014-15. The authors are grateful for the constructive comments from two anonymous reviewers.

References

Adnan, S., Maltamo, M., Coomes, D.A., Valbuena, R., 2017. Effects of plot size, stand density, and scan density on the relationship between airborne laser scanning metrics and the Gini coefficient of tree size inequality. Can. J. For. Res. 47 (12), 1590–1602.
Asner, G.P., Mascaro, J., 2014. Mapping tropical forest carbon: calibrating plot estimates to a simple LiDAR metric. Remote Sens. Environ. 140, 614–624.
Bergeron, Y., Leduc, A., Harvey, B.D., Gauthier, S., 2002. Natural fire regime: a guide for sustainable management of the Canadian boreal forest. Silva Fennica 36 (1), 81–95.
Bien, J., Tibshirani, R., 2011. Hierarchical clustering with prototypes via minimax linkage. J. Am. Stat. Assoc. 106 (495), 1075–1084.
Bollandsås, O.M., Næsset, E., 2007. Estimating percentile-based diameter distributions in uneven-sized Norway spruce stands using airborne laser scanner data. Scand. J. For. Res. 22 (1), 33–47.
Bourdier, T., Cordonnier, T., Kunstler, G., Piedallu, C., Lagarrigues, G., Courbaud, B., 2016. Tree size inequality reduces forest productivity: an analysis combining inventory data for ten European species and a light competition model. PloS One 11 (3), e0151852.
Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Wadsworth, Monterey, Calif., USA.
Coomes, D.A., Allen, R.B., 2007a. Effects of size, competition and altitude on tree growth. J. Ecol. 95 (5), 1084–1097.
Coomes, D.A., Allen, R.B., 2007b. Mortality and tree-size distributions in natural mixed-age forests. J. Ecol. 95 (1), 27–40.
COST: cooperation in the field of science and technology, 2006. Harmonization of national forest inventories in Europe: techniques for common reporting. COST Action E43. Finnish Forest Research Institute.
Curtis, R.O., 1982. A simple index of stand density for Douglas-fir. For. Sci. 28 (1), 92–94.
Dodson, E.K., Ares, A., Puettmann, K.J., 2012. Early responses to thinning treatments designed to accelerate late successional forest structure in young coniferous stands of western Oregon, USA. Can. J. For. Res. 42 (2), 345–355.
Donato, D.C., Campbell, J.L., Franklin, J.F., 2012. Multiple successional pathways and precocity in forest development: can some forests be born complex? J. Veg. Sci. 23 (3), 576–584.
Duduman, G., 2011. A forest management planning tool to create highly diverse uneven-aged stands. Forestry 84 (3), 301–314.
Echeverría, C., Newton, A.C., Lara, A., Benayas, J.M.R., Coomes, D.A., 2007. Impacts of forest fragmentation on species composition and forest structure in the temperate landscape of southern Chile. Glob. Ecol. Biogeogr. 16 (4), 426–439.
Enquist, B.J., Niklas, K.J., 2001. Invariant scaling relations across tree-dominated communities. Nature 410 (6829), 655.
European Environmental Agency, 2018. Biogeographical regions. <https://www.eea.europa.eu/data-and-maps/data/biogeographical-regions-europe-3> (accessed May 14, 2018).
Everitt, B.S., Landau, S., Leese, M., Stahl, D., 2011. Cluster analysis: Wiley series in probability and statistics.
Fedrigo, M., Newnham, G.J., Coops, N.C., Culvenor, D.S., Bolton, D.K., Nitschke, C.R., 2018. Predicting temperate forest stand types using only structural profiles from discrete return airborne lidar. ISPRS J. Photogr. Remote Sens. 136, 106–119.
Franco-Lopez, H., Ek, A.R., Bauer, M.E., 2001. Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method. Remote Sens. Environ. 77 (3), 251–274.
Giannetti, F., Barbati, A., Mancini, L.D., Travaglini, D., Bastrup-Birk, A., Canullo, R., Nocentini, S., Chirici, G., 2018. European Forest Types: toward an automated classification. Ann. For. Sci. 75 (1), 6.
Görgens, E.B., Packalen, P., Da Silva, A.G.P., Alvares, C.A., Campoe, O.C., Stape, J.L., Rodriguez, L.C.E., 2015. Stand volume models based on stable metrics as from multiple ALS acquisitions in Eucalyptus plantations. Ann. For. Sci. 72 (4), 489–498.
Gougeon, F.A., St-Onge, B.A., Wulder, M., Leckie, D.G., 2001. Synergy of airborne laser altimetry and digital videography for individual tree crown delineation. In: Proceedings of the 23rd Canadian Symposium on Remote Sensing, August, pp. 21–24.
Gove, J.H., 2004. Structural stocking guides: a new look at an old friend. Can. J. For. Res. 34 (5), 1044–1056.
Gove, J.H., Patil, G.P., Taillie, C., 1995. A mathematical programming model for maintaining structural diversity in uneven-aged forest stands with implications to other formulations. Ecol. Model. 79 (1–3), 11–19.
Hastie, T., Tibshirani, R., Friedman, J., 2009. Unsupervised learning. In: The Elements of Statistical Learning. Springer, New York, NY, pp. 485–585.
Hyyppä, J., Hyyppä, H., Leckie, D., Gougeon, F., Yu, X., Maltamo, M., 2008. Review of methods of small-footprint airborne laser scanning for extracting forest inventory data in boreal forests. Int. J. Remote Sens. 29 (5), 1339–1366.
Jaskierniak, D., Lane, P.N., Robinson, A., Lucieer, A., 2011. Extracting LiDAR indices to characterise multilayered forest structure using mixture distribution functions. Remote Sens. Environ. 115 (2), 573–585.
JRC: joint research centre, 2011. Biosoil Biodiversity Executive Report. Report Number: Joint Research Centre Report 64509. Publications Office of the European Union ISSN1018–5593.
Kilkki, P., Päivinen, R., 1987. Reference sample plots to combine field measurements and satellite data in forest inventory. Department of Forest Mensuration and Management, University of Helsinki, Research Notes 19, 210–215.
Kim, S., McGaughey, R.J., Andersen, H.E., Schreuder, G., 2009. Tree species

differentiation using intensity data derived from leaf-on and leaf-off airborne laser scanner data. Remote Sens. Environ. 113 (8), 1575–1586.
Lawrence, R.L., Wright, A., 2001. Rule-based classification systems using classification and regression tree (CART) analysis. Photogramm. Eng. Remote Sens. 67 (10), 1137–1142.
Lefsky, M.A., Hudak, A.T., Cohen, W.B., Acker, S.A., 2005. Geographic variability in lidar predictions of forest stand structure in the Pacific Northwest. Remote Sens. Environ. 95 (4), 532–548.
Lexerød, N., Eid, T., 2006. An evaluation of different diameter diversity indices based on criteria related to forest management planning. For. Ecol. Manage. 222 (1–3), 17–28.
Linder, P., Elfving, B., Zackrisson, O., 1997. Stand structure and successional trends in virgin boreal forest reserves in Sweden. For. Ecol. Manage. 98 (1), 17–33.
Maltamo, M., Bollandsås, O., Næsset, E., Gobakken, T., Packalén, P., 2010. Different plot selection strategies for field training data in ALS-assisted forest inventory. Forestry 84 (1), 23–31.
Maltamo, M., Eerikäinen, K., Packalén, P., Hyyppä, J., 2006. Estimation of stem volume using laser scanning-based canopy height metrics. Forestry 79 (2), 217–229.
Maltamo, M., Eerikäinen, K., Pitkänen, J., Hyyppä, J., Vehmas, M., 2004. Estimation of timber volume and stem density based on scanning laser altimetry and expected tree size distribution functions. Remote Sens. Environ. 90 (3), 319–330.
Maltamo, M., Mehtätalo, L., Vauhkonen, J., Packalén, P., 2012. Predicting and calibrating tree size and quality attributes by means of airborne laser scanning and field measurements. Can. J. For. Res. 42 (11), 1896–1907.
Maltamo, M., Packalén, P., Yu, X., Eerikäinen, K., Hyyppä, J., Pitkänen, J., 2005. Identifying and quantifying structural characteristics of heterogeneous boreal forests using laser scanner data. For. Ecol. Manage. 216 (1–3), 41–50.
Marvin, D.C., Asner, G.P., Knapp, D.E., Anderson, C.B., Martin, R.E., Sinca, F., Tupayachi, R., 2014. Amazonian landscapes and the bias in field studies of forest structure and biomass. Proc. Natl. Acad. Sci. 111 (48), E5224–E5232.
McElhinny, C., Gibbons, P., Brack, C., Bauhus, J., 2005. Forest and woodland stand structural complexity: its definition and measurement. For. Ecol. Manage. 218 (1–3), 1–24.
McRoberts, R.E., Winter, S., Chirici, G., Hauk, E., Pelz, D.R., Moser, W.K., Hatfield, M.A., 2008. Large-scale spatial patterns of forest structural diversity. Can. J. For. Res. 38 (3), 429–438.
McRoberts, R.E., Winter, S., Chirici, G., LaPoint, E., 2012. Assessing forest naturalness. For. Sci. 58 (3), 294–309.
Means, J.E., Acker, S.A., Fitt, B.J., Renslow, M., Emerson, L., Hendrix, C.J., 2000. Predicting forest stand characteristics with airborne scanning lidar. Photogramm. Eng. Remote Sens. 66 (11), 1367–1372.
Meyer, D., Zeileis, A., Hornik, K., 2014. vcd: Visualizing Categorical Data. R package version 1.3-2.
Moran, C.J., Rowell, E.M., Seielstad, C.A., 2018. A data-driven framework to identify and compare forest structure classes using lidar. Remote Sens. Environ. 211, 13–166.
Moss, I., 2012. Stand structure classification, succession, and mapping using LiDAR (Doctoral dissertation). University of British Columbia.
Muller-Landau, H.C., Condit, R.S., Harms, K.E., Marks, C.O., Thomas, S.C., Bunyavejchewin, S., Chuyong, G., Co, L., Davies, S., Foster, R., Gunatilleke, S., 2006. Comparing tropical forest tree size distributions with the predictions of metabolic ecology and equilibrium models. Ecol. Lett. 9 (5), 589–602.
Müllner, D., 2013. fastcluster: fast hierarchical, agglomerative clustering routines for R and Python. J. Stat. Softw. 53 (9), 1–18.
Næsset, E., 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. Remote Sens. Environ. 80 (1), 88–99.
Newton, P., 1997. Stand density management diagrams: review of their development and utility in stand-level management planning. For. Ecol. Manage. 98 (3), 251–265.
O'Connor, B., Secades, C., Penner, J., Sonnenschein, R., Skidmore, A., Burgess, N.D., Hutton, J.M., 2015. Earth observation as a tool for tracking progress towards the Aichi Biodiversity Targets. Remote Sens. Ecol. Conserv. 1 (1), 19–28.
O'Hara, K.L., Gersonde, R.F., 2004. Stocking control concepts in uneven-aged silviculture. Forestry 77 (2), 131–143.
O'Hara, K.L., Latham, P.A., Hessburg, P., Smith, B.G., 1996. A structural classification for Inland Northwest forest vegetation. Western J. Appl. Forest. 11 (3), 97–102.
Pascual, C., García-Abril, A., García-Montero, L.G., Martín-Fernández, S., Cohen, W.B., 2008. Object-based semi-automatic approach for forest structure characterization using lidar data in heterogeneous *Pinus sylvestris* stands. For. Ecol. Manage. 255 (11), 3677–3685.
Pereira, H.M., Ferrier, S., Walters, M., Geller, G.N., Jongman, R.H.G., Scholes, R.J., Bruford, M.W., Brummitt, N., Butchart, S.H.M., Cardoso, A.C., Coops, N.C., 2013. Essential biodiversity variables. Science 339 (6117), 277–278.
Pommerening, A., 2002. Approaches to quantifying forest structures. For.: Int. J. For. Res. 75 (3), 305–324.
Proença, V., Martin, L.J., Pereira, H.M., Fernandez, M., McRae, L., Belnap, J., Böhm, M., Brummitt, N., García-Moreno, J., Gregory, R.D., Honrado, J.P., 2017. Global biodiversity monitoring: from data sources to essential biodiversity variables. Biol. Conserv. 213, 256–263.
R Core Team, 2018. R: a language and environment for statistical computing [online]. Available from. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.
Savill, P., Perrins, C., Fisher, N., Kirby, K. (Eds.), 2011. Wytham Woods: Oxford's Ecological Laboratory. Oxford University Press.
Simpson, J.E., Smith, T.E., Wooster, M.J., 2017. Assessment of Errors caused by forest vegetation structure in airborne LiDAR-derived DTMs. Remote Sens. 9 (11), 1101.
Story, M., Congalton, R.G., 1986. Accuracy assessment: a user's perspective. Photogramm. Eng. Remote Sens. 52 (3), 397–399.
Sumida, A., Miyaura, T., Torii, H., 2013. Relationships of tree height and diameter at breast height revisited: analyses of stem growth using 20-year data of an even-aged *Chamaecyparis obtusa* stand. Tree Physiol. 33 (1), 106–118.
Sugar, C.A., James, G.M., 2003. Finding the number of clusters in a dataset: an information-theoretic approach. J. Am. Stat. Assoc. 98 (463), 750–763.
Valbuena, R., 2015. Forest structure indicators based on tree size inequality and their relationships to airborne laser scanning. Dissertation Forestales 1, 55–59.
Valbuena, R., Packalén, P., Martín-Fernández, S., Maltamo, M., 2012. Diversity and equitability ordering profiles applied to study forest structure. For. Ecol. Manage. 276, 185–195.
Valbuena, R., Packalen, P., Mehtätalo, L., García-Abril, A., Maltamo, M., 2013. Characterizing forest structural types and shelterwood dynamics from Lorenz-based indicators predicted by airborne laser scanning. Can. J. For. Res. 43 (11), 1063–1074.
Valbuena, R., Vauhkonen, J., Packalen, P., Pitkänen, J., Maltamo, M., 2014. Comparison of airborne laser scanning methods for estimating forest structure indicators based on Lorenz curves. ISPRS J. Photogramm. Remote Sens. 95, 23–33.
Valbuena, R., Maltamo, M., Packalen, P., 2016a. Classification of multilayered forest development classes from low-density national airborne lidar datasets. For.: Int. J. For. Res. 89 (4), 392–401.
Valbuena, R., Eerikäinen, K., Packalen, P., Maltamo, M., 2016b. Gini coefficient predictions from airborne lidar remote sensing display the effect of management intensity on forest structure. Ecol. Ind. 60, 574–585.
Valbuena, R., Maltamo, M., Mehtätalo, L., Packalen, P., 2017. Key structural features of Boreal forests may be detected directly using L-moments from airborne lidar data. Remote Sens. Environ. 194, 437–446.
Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S, fourth ed. Springer.
Vihervaara, P., Mononen, L., Auvinen, A., Virkkala, R., Lü, Y., Pippuri, I., Packalen, P., Valbuena, R., Valkama, J., 2015. How to integrate remotely sensed data and biodiversity for ecosystem assessments at landscape scale. Landscape Ecol. 30 (3), 501–516.
Vihervaara, P., Auvinen, A.P., Mononen, L., Törmä, M., Ahlroth, P., Anttila, S., Böttcher, K., Forsius, M., Heino, J., Heliölä, J., Koskelainen, M., 2017. How essential biodiversity variables and remote sensing can help national biodiversity monitoring. Global Ecol. Conserv. 10, 43–59.
Warnes, G., 2013. gmodels: Various R programming tools for model fitting. R package version 2.16.2.
Weiner, J., 1985. Size hierarchies in experimental populations of annual plants. Ecology 66 (3), 743–752.
Zimble, D.A., Evans, D.L., Carlson, G.C., Parker, R.C., Grado, S.C., Gerard, P.D., 2003. Characterizing vertical forest structure using small-footprint airborne LiDAR. Remote Sens. Environ. 87 (2–3), 171–182.