

Using classification trees to predict forest structure types from LiDAR data

C. Torresan, P. Corona, G. Scrinzi, J. V. Marsal

Torresan C., Corona P., Scrinzi G., Marsal J.V., 2016. Using classification trees to predict forest structure types from LiDAR data. *Ann. For. Res.* 59(1): 281-298.

Abstract. This study assesses whether metrics extracted from airborne LiDAR (Light Detection and Ranging) raw point cloud can be exploited to predict different forest structure types by means of classification trees. Preliminarily, a bivariate analysis by means of Pearson statistical test was developed to find associations between LiDAR metrics and the proportion of basal area into three stem diameter classes (understory, mid-story, and over-story trees) of 243 random distributed plots surveyed from 2007 to 2012 in Trento Province (Northern Italy). An unsupervised clustering approach was adopted to determine forest structural patterns on the basis of basal area proportion in the three stem diameter classes, using a k-means procedure combined with a previous hierarchical classification algorithm. A comparison among the identified clusters centroids was performed by the Kruskal-Wallis test. A classification tree model to predict forest structural patterns originating from the cluster analysis was developed and validated. Between 18 potential LiDAR metrics, 11 were significantly correlated with the proportion of basal area of understory, mid-story, and overstory trees. The results coming from the agglomerative hierarchical clustering allowed identification of 5 clusters of forest structure: pole-stage (70% of the considered cases), young (15%), adult (24.3%), mature (24.3%), and old forests (30%). Five LiDAR metrics were selected by the classification tree to predict the forest structural types: standard deviation and mode of canopy heights, height at which 95% and 99% of canopy heights fall below, difference between height at which 90% and 10% of canopy heights fall below. The validation tree model process showed a misclassification error of 45.9% and a level of user's accuracy ranging between 100% and 33.3% in the validation data set. The highest level of user's accuracy was reached in the classification of pole-stage forests (100%), in which more than 82% of basal area is due to the understory-trees, follow by the classification of old forests types (63.5% of basal area due to the overstory-trees) achieved 76.5% of user's accuracy. The model has provided moderately satisfactory results in term of classification performance: substantial room for improvement might be established by multi- or hyperspectral imaging that allow detailed characterization of the spectral behaviour of the forest structure types.

Keywords airborne laser scanning, discrete return laser scanner data, stem diameter classes, basal area, bivariate analysis, unsupervised clustering, classification tree model, forest inventory, forest management

Authors. Chiara Torresan (c.torresan@ibimet.cnr.it) - National Research Council - Institute for Biometeorology, via Giovanni Caproni, 8 (50145) Firenze, Italia. FOXLAB Joint CNR-FEM Initiative, Via E. Mach, 1 38010 San Michele all'Adige (TN), Italy; Piermaria Corona - Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria, Forestry Research Centre (CREA-SEL), Viale Santa Margherita, 80 (52100), Arezzo, Italy; Gianfranco Scrinzi - Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria, Forest Monitoring and Management Research Unit (CREA-MPF), Piazza Nicolini, 6 (38123), Trento, Italy; Joan Valls Marsal - Biostatistics Unit, Biomedical Research Institute of Lleida (IRBLLEIDA) Hospital Universitari Arnau de Vilanova de Lleida (HUAU) C/ Rovira Roure, 80 Laboratoris d'investigació HUAU (25198)

Manuscript received August 05, 2015; revised April 20, 2016; accepted April 27, 2016; online first May 06, 2016.

Introduction

Structure is one of the basic forest characteristics. Even though the definition of forest stand structure is not fixed in the same manner as the traditionally considered forest inventory attributes (Maltamo et al. 2005), we can say that the structure represents a mutual arrangement of components next to each other or above each other (Šebeň & Bošela 2010).

For forest management purposes, the stand structure is typically described at various scales in terms of number of trees or percent cover of a given species composition or basal area per hectare within certain size classes. Sometimes these descriptions are reduced to average stand statistics: the problem with this kind of statistics is that the same average can be produced from significantly different distributions of tree diameters (stem diameter measured at breast height, i.e. at 1.30 m from the ground, dbh). Hence if a more detailed description of the dbh distributions were available, then probably more reliable estimates could be made and used for responding to species habitat uses and requirements, for developing silviculture and harvesting systems, and so to forecast timber supply of different quality (Moss 2012).

To classify stand structure is important because the classification enables consistent, precise and verifiable information about the expected dbh distribution of trees within a stand.

Silviculturists and forest managers typically use structural classification for management purposes. Differences in stand structure are important for applying different silvicultural interventions to manipulate stand dynamics in order to direct the stand development toward a specific direction. Moreover, stand structure classification is important for forest inventory purposes, because it can be used to pre-stratify forest areas to reduce the costs of sampling.

Remotely sensed data is increasingly utilized to assist traditional field-based methods in the estimation of forest structural attributes over extensive areas. Remote sensing is a valuable source of information in mapping and monitoring forest features (Corona 2010), and machine learning techniques, such as classification and regression trees, artificial neural networks and support vector machines, have been used to exploit remotely sensed data for forest stand type classification. In the use of remote sensing data for forestry applications, a typical machine learning task is to develop a predictive model which employs a set of remote sensing observations to predict forest conditions. The remotely sensed data most used for this purpose are satellite imagery (e.g. Ghose et al. 2010, Gomez et al. 2012, Fallah et al. 2013, Li et al. 2013, Griffiths et al. 2014). Although satellite data have been widely explored for this forest application, passive remote sensing techniques are limited in their ability to capture forest structural complexity,

particularly in uneven-aged, mixed species forests with multiple canopy layers. Generally, these techniques are only able to provide information on horizontal (two-dimensional) forest extent. The vertical forest structure (or the interior of the canopy and the understory vegetation) cannot be mapped using these passive remote sensing techniques. On the other hand, it has been shown that the LiDAR (Light Detection and Ranging) technique, which has great capability of canopy penetration (Reutebuch et al. 2003, Isenburg et al. 2014), yields such high density sampling that detailed description of the forest structure in three-dimensions can be obtained.

Consequently, the LiDAR has become the primary source for characterizing forest structure (Lefsky et al. 1999, Drake et al. 2002, Lim et al. 2003a, Corona & Fattorini 2008, Hyypää et al. 2008, Vierling et al. 2008, Hudak et al. 2009, Corona et al. 2012, Wulder et al. 2012, Corona et al. 2014). Accordingly, much interest is attached to exploring the use of LiDAR data for modelling structural forest parameters. The LiDAR literature has extensively looked at forest structure characterization, following the approach of deriving LiDAR metrics (descriptive statistics) from point clouds within the specific analysis unit (e.g. plot, or stand) and run linear regression analysis (Næsset 2002, Lim et al. 2003b, Holmgren & Jonsson 2004, Maltamo et al. 2005), multivariate linear regression analysis (Lefsky et al. 1998, Lefsky et al. 1999, Sherrill et al. 2008, Ediriweera et al. 2014), stepwise regression analysis (Means et al. 2000), Weibull probability density function (Coops et al. 2007) to relate these variables to field measurements forest structure attributes, such as canopy height, fractional vegetation cover, basal area, etc.

As Zhao et al. (2008) highlighted, in the case of using statistical analysis tools for inferring relationship and making prediction, of particular importance are the selection of appropriate statistical procedures for model building and, more important, the choice of metrics

that should enter the models, and hence the statistical procedures to determine those predictors. Between hundreds of metrics that can be extracted from LiDAR data, conventional rules are used to select metrics for estimating a canopy variable, such as the physical linkage of metrics to the variable in question, the avoidance of multicollinearity, and the optimization of certain statistical measures like *R*-squared, the Akaike information criterion and the Bayesian information criterion (Zhao et al. 2008).

As confirmed by remote sensing studies conducted in the past years (Durbha et al. 2007, Zhao et al. 2008), supervised learning techniques, e.g. machine learning, may be chosen because of their usefulness in tackling high-dimensional problems in assessing forest biophysical parameters. This outcome has also been demonstrated in recent studies. Zhao et al. (2011) verified that models developed using support vector machine and Gaussian processes as machine learning in conjunction with the LiDAR composite metrics to characterize forest canopy structure outperformed traditional approaches such as the maximum likelihood classifier and linear regression models. Gleason and Im (2012) compared the effectiveness of four machine learning approaches for estimating biomass in moderately dense forest at both tree and plot levels: their results indicated that biomass estimation accuracy improves when modelled at the plot level and that support vector regression produced the most accurate biomass model. Garcia-Gutierrez et al. (2014) presented a comparison between the classic multiple linear regression-based methodology and regression techniques in machine learning (neural networks, support vector machines, nearest neighbour, random forests) to estimate crown biomass, stem biomass, aboveground biomass, basal area, dominant height, mean height, and volume at stand level by LiDAR metrics. The results of Garcia-Gutierrez et al. (2014) showed that classic multiple linear regression is outperformed by

machine learning technique. Gørgens et al. (2015a) compared the performance of three machine learning tools (neural network, random forest and support vector regression) for predicting stand volume of fast-growing forest plantations by metrics extracted from an ALS survey with the conventional method that uses regression models through variables reduction and selection: LiDAR metrics were proven to be more effective for the estimation of stand volume.

Various studies considered the use of machine learning techniques applied to LiDAR data integrated with satellite or airborne imagery for the identification of forest canopy parameters. Dalponte et al. (2008) in their analysis on the joint effect of hyperspectral and LiDAR data for the classification of complex forest areas using support vector machines and Gaussian maximum likelihood proved that support vector machine classifier was robust and accurate in the exploitation of the considered multisource data.

Ke et al. (2010) used synergistically multispectral imagery and LiDAR data for object-based forest species classification utilizing machine learning decision trees: they obtained high classification accuracy using both spectral- and LiDAR-derived metrics based on objects segmented from both spectral and LiDAR layers.

On the basis of the those results, we asserted that machine learning techniques perform better than conventional regression approaches for forest structure classification. In our case study, we applied machine learning to classify the structure of mountain forests with complex structures (multi-layered forests), an issue that has hitherto remained untested due to the fact that the large part of previous studies was conducted mostly in temperate and boreal forests of Europe and North America relatively homogeneous from the structural point of view.

Moreover, with this study we wanted to use machine learning techniques not to predict single field measured attributes which, individually considered, provide information about an

aspect of the forest structure, but not about the forest stand structure itself. Conversely, we wanted to predict forest structure types.

Distinctively, the goal of this research was to explore the potential of LiDAR data to classify forest structure types of mountain forests in Italian Alps (Trento province) by means of classification trees. We examined the potential of LiDAR point cloud-derived metrics to model the proportion of basal area in three dbh classes at stand level (as direct estimator of type of structure). Specific objectives of the study are: (i) to identify different types (patterns) of forest structures on the basis of the proportion of basal area in three dbh classes which describe understory, mid-story and overstory trees via an unsupervised cluster analysis, (ii) to develop a classification tree model based on a set of variables derived from LiDAR point cloud to predict the forest structure types identified through the cluster analysis and (iii) to test and verify the ability of the developed classification tree model as statistical techniques for classifying forest structure type.

Materials and methods

Study area

The study area is the province of Trento (6212 km²) located in the North-East of Italy on the Southern side of the Alps chain. The territory is almost entirely mountainous: around 60% of the surface is located at an altitude higher than 1000 m a.s.l., 19% between 600 and 1000 m a.s.l., and 21% between 200 and 600 m a.s.l.

The climate of the province is cool, temperate and mild continental. The mean annual temperature is 11.5 °C, while the annual rainfall average is 883 mm with two main peak periods, in spring (May rainfall averages 94 mm) and autumn (October rainfall averages 110 mm).

The area of Trento province, according to its geographical and climatic gradients, can be

divided into three zones: endalpica, mesalpica and esalpica (Odasso 2002), to which different forest vegetation correspond. The endalpica zone includes upland areas with higher elevation and landlocked valleys. There are environments with harsh continental climate, particularly favourable to forest communities dominated by boreal conifers like Norway spruce (*Picea abies* (L.) H. Karst), Scots and black pines (*Pinus sylvestris* L. and *Pinus nigra* Arnold), Swiss pine (*Pinus cembra* L.), and European larch (*Larix deciduas* Mill.). The mesalpica zone include mountains with relatively lower elevations, generally found on plateaus and in valleys, typically with East-West orientation and with average elevation around 1000 m a.s.l.. Cool climate, from sub-continental to sub-oceanic, characterizes this zone, where forests are dominated by mesophile tree species like silver fir (*Abies alba* Mill.) and beech (*Fagus sylvatica* L.). The esalpica zone is concentrated in a central strip orientated North-South in the Trento province territory, with elevation below 1000 m a.s.l., characterized by incursions of species with sub-Mediterranean or steppe character, dominated by the forests composed of thermophile broad leaved trees (*Ostrya carpinifolia* Scop., *Carpinus betulus* L., *Fraxinus ornus* L., *Quercus pubescens* Willd., *Quercus petrae* (Mattuschka) Liebl.).

Of the Trento province surface, 55.7% is covered by forests (PAT 2008): 59.2% of the forest are predominantly composed by Norway spruce, 17.3% by European larch, 10.6% by silver fir, 6.1% by Scots and black pines, 4.3% by beech, 1.4% by Swiss pine and 1.1% by other broad-leaved species (PAT 2008).

Public institutions own 76.3% of the forest surface (PAT 2008) which is subject to a forest management plan with broad objectives, such as maintaining productive function of the forest, management of the services provided by the forest (protection, tourism and recreation, carbon dioxide fixation, etc), and improvement and conservation of biodiversity in terms of species, habitat and landscape (Torresan et al.

2014). In Trento province, 79% of the forest surface is managed by the system of seed tree and shelterwood, while the remaining 21% by the coppice system (PAT 2008).

Field data collection and response variable

The dendrometric data used in this study comprise a collection of data coming from three different sites located in Trento province. We used data available from previous experiments to have a sample of the most representative of different cases of basal area distribution into dbh classes because the percentage of basal area in three classes of trees, defined as small trees or understory (17.5 cm to 27.4 cm dbh), medium trees or mid-story (27.5 to 47.4 cm dbh), and large trees or overstory (above 47.5 cm dbh), is the criterion used by the Forest Service of the Trento Province to classify the structure of forest stands.

The first data set includes 92 circular plots with different sizes (531, 1257, 1964 and 2827 m²) randomly distributed in forest compartments of public lands. The field surveys were carried out from June to September 2007. The center of each plot was georeferenced with a Thales MobileMapper CE Global Positioning System (GPS) receiver recording at minimum 200 observations on the ground surface. Afterwards the geographical coordinates of plot centers were differentially corrected. The species was recorded and dbh of trees greater than or equal to 7.5 cm was measured from two orthogonal axes by timber caliper and then averaged.

The second data set refers to data collected during research activities whose aim was to investigate stand structure indices as tools to support forest management (Pastorella and Paletto 2013). Stratified random sampling was applied to place 24 plots in Norway spruce forests pure and mixed with silver fir, 12 plots in beech forests, 12 plots in Scots pine forests, and 6 plots in European larch forests for a total of 54 plots surveyed. Field surveys were carried out from June to September 2012. The

center of each circular plot was georeferenced by a Garmin eMap GPS receiver recording at minimum 200 satellite positions on the ground surface. The plot area was 531 m². For all trees with dbh greater than or equal to 4.5 cm, the specie was recorded and the dbh was measured from two orthogonal axes by timber caliper and then averaged.

The third data set is made up by data collected in three specific forest areas of Trento province: Foresta Demaniale di Paneveggio (North East of Trento province), Padergnone municipality (central part of Trento province), and Val di Sella (South East of Trento province). In total 110 plots were randomly distributed over those three areas. The field surveys were carried out during the 2007 summer season. The plot areas were 531 m² and 1257 m². The nominal coordinates of plot center were reached using a navigation assisted by a GPS receiver according to the protocol adopted in the National Forest Inventory (Scrinzi and Floris 2003). The center of each plot was georeferenced with Thales MobileMapper CE and Trimble Geo-XT GPS receivers recording at minimum 200 satellite positions on the ground surface. Afterwards the geographical coordinates of plot centers were differentially corrected and averaged. In each plot, tree species was recorded and dbh was measured as average of two orthogonal axes from a threshold of 2.5 cm using timber calipers. Figure 1 illustrates the location of 256 plots.

The dbh was used to calculate the basal area



Figure 1 Plot locations over the territory of Trento province

of each tree and successively the field-based proportion of basal area in the three dbh classes (understory, mid-story and over-story) on each plot needed in the cluster analysis “for constructing a sensible and informative classification of the initially unclassified set of data” (Everitt et al. 2001).

LiDAR data acquisition, LiDAR point clouds processing and metrics computation

During the 2006-2007 autumn and winter seasons, the Province of Trento commissioned Blom CGR S.p.A. to conduct an airborne laser scanning (ALS) campaign over its entire territory. Discrete return LiDAR data for the whole surface of Trento province were acquired by an Optech ALTM3100C laser system mounted in Partenavia P68. The LiDAR system recorded range and intensity of 2 returns per pulse, and achieved a nominal density of 1.28 points per m² (Table 1).

In 2013, the Province of Trento made available the raw LiDAR data throughout all its territory. Raw LiDAR data were delivered in tiles of 2 x 2 km² without overlap.

TerraScan (Terrasolid) and LAsTools (rapidlasso GmbH) software were used to process the LiDAR point cloud data. The parameters of the “Classify ground” function, based on the Axelsson (1999) algorithm’s, were set to 5 m in the case of the max building size, to 80° in the case of the terrain angle, to 8° in the case of the iteration angle and to 2 m in the

case of the iteration distance. LAsTools was used to normalized the height of each point with respect to the ground surface. Once the LiDAR data were normalized, the point data were processed to extract LiDAR metrics at the plot level: the plot level variables are statistics calculated on the LiDAR points falling within the spatial

Table 1 Acquisition parameters of the 2006-2007 autumn and winter airborne laser campaign

Sensor	Optech ALTM3100C
Acquisition date	Autumn and winter 2006, 2007
Flight altitude	1500 m above ground
Flight line sidelap	50%
Maximum off-nadir scan angle	25°
Returns/pulse	2
Density	1.28 pulses m ⁻²
Pulse repetition	33 kHz
Laser wavelength	800 nm

Table 2 LiDAR metrics used in the analysis

Name of metric	Description
hmean	Mean of LiDAR heights intersecting plot
hmode	Mode of LiDAR heights intersecting plot
hstd	Standard deviation of LiDAR heights intersecting plot
hcv	Coefficient of variation of LiDAR heights intersecting plot
hiqdist	Interquartile distance of LiDAR heights intersecting plot
hske	Skewness of LiDAR heights intersecting plot
hkur	Kurtosis of LiDAR heights intersecting plot
haad	Average absolute deviation of LiDAR heights intersecting plot (= the average distance between the height of each LiDAR point and the mean height of LiDAR point in the plot)
hp01	Height at which 1% of LiDAR heights intersecting plot fall below
hp05	Height at which 5% of LiDAR heights intersecting plot fall below
hp50	Height at which 50% of LiDAR heights intersecting plot fall below (median of LiDAR height intersecting plot)
hp95	Height at which 95% of LiDAR heights intersecting plot fall below
hp99	Height at which 99% of LiDAR heights intersecting plot fall below
hp90_hp10	Difference between height at which 90% of LiDAR heights intersecting plot fall below and height at which 10% of LiDAR heights intersecting plot fall below
crr	Canopy-relief ratio (ratio of the difference between the mean and the minimum of LiDAR heights intersecting plot over the difference between the maximum and the minimum of LiDAR heights intersecting plot)
vegcover	Canopy cover (ratio of the number of returns intersecting plot with height ≥ 2 m over the total number of returns intersecting plot)
percreturnmean	Percentage of all returns above the mean of LiDAR heights intersecting plot
percreturnmode	Percentage of all returns above the mode of LiDAR heights intersecting plot

border corresponding to the boundary of each field plot. The statistics were computed on all points LiDAR heights (i.e. both first and last return) above 2 m using freely-available FUSION software (McGaughey 2012). For our study, we decided to use only the plots with a density of at least 0.5 points/m²: in this way

243 plots (out of the 256 available) were selected and used for the purposes of this work. LiDAR metrics considered in this study are described in Table 2. Canopy relief ratio is a quantitative descriptor of the relative shape of the canopy from altimetry observation (Pike & Wilson 1971, Parker & Russ 2004) which

describes the degree to which canopy surfaces are in the upper ($crr > 0.5$) or in the lower ($crr < 0.5$) portions of the height range.

Methods

To achieve the goal of the study, as preliminary investigation a bivariate analysis was performed to assess the relationship between LiDAR variables, and the proportion of basal area of understory, mid-story, and overstory trees. For this purpose, a Pearson correlation test was applied using the Bonferroni method to correct p -values for multiple testing.

To further investigate differences in forest structures and hence identify the number of clusters of forest types, an unsupervised cluster analysis was performed to identify the different types of forests based on the proportion of basal area in three classes, using a hierarchical classification algorithm based on a Ward (1963) criterion. The hierarchical clustering was then improved by applying a k-means algorithm. A comparison among the identified clusters centroids was performed by the Kruskal-Wallis test.

Classification trees were used to assess the capability of LiDAR metrics as predictors to explain the forest structures identified through the cluster analysis. The results of this supervised technique were then internally validated with the misclassification error to evaluate its global performance. An external validation was accomplished splitting the general sample into two sub-samples: the training set, composed by the 75% of observations, and the validation set, with the remaining 25% of observations. The results of validation were categorized in a contingency table, reporting errors of commission, sometimes also called “false positives”,

which refer to the percentage of observations incorrectly classified, and errors of omissions, also called “false negative”, which refer to percentage of observations that are missed in classification in the specific group.

All statistical analyses were performed using *stats* and *rpart* R packages.

Results

Basic statistics

Table 3 reports the statistical distribution of response variables across the 243 plots under investigation.

The results from the bivariate analysis between LiDAR metrics and the proportion of basal area of understory, mid-story, and overstory trees are reported in Table 4.

The mean LiDAR height intersecting plots is negatively correlated with the percentage of basal area of small trees: this means that large values of mean height are associated with small values of the proportion of basal area of understory trees. Conversely large values of mean height are associated with large values of the proportion of basal area of overstory trees. The highest percentile (95th and 99th) and the difference between the 90th and the 10th percentile are significantly correlated with the proportion of basal area of understory, mid-story, and overstory trees. No variable related to canopy cover (vegetation cover, percentage of all returns above the mean or the mode of LiDAR heights intersecting plots) was significantly correlated with the proportion of basal area of understory, mid-story, and overstory trees.

Table 3 Statistical distribution of percentage of basal area of understorey, mid-story and overstorey trees

Canopy position of trees	Minimum	Maximum	Mean	Standard deviation
Understory	0.00	100	21.57	21.87
Mid-story	0.00	100	46.04	20.41
Overstory	0.00	97.53	32.39	25.41

Table 4 Pearson correlation coefficient (rP) with the associated Bonferroni corrected p -values (p -value-BC) between LiDAR metrics and proportion of basal area of understory (%BA understory trees), mid-story (%BA mid-story trees), and overstory trees (%BA overstory trees). Values in bold indicate p -values <0.05. The abbreviations follow the Table 2.

Metric name	%BA understory trees	%BA mid-story trees	%BA overstory trees
hmean	-0.54 (0.0000)	-0.13 (0.7205)	0.57 (0.0000)
hmode	-0.49 (0.0000)	-0.09 (1.0000)	0.49 (0.0000)
hstd	-0.68 (0.0000)	-0.21 (0.0164)	0.76 (0.0000)
hcv	-0.28 (0.0002)	-0.06 (1.0000)	0.29(0.0001)
hiqdist	-0.60 (0.0000)	-0.20 (0.0364)	0.67 (0.0000)
hske	-0.02 (1.0000)	-0.06 (1.0000)	0.07 (1.0000)
hkur	0.30 (0.0000)	-0.02 (1.0000)	-0.24 (0.0034)
haad	-0.66 (0.0000)	-0.21 (0.0205)	0.73 (0.0000)
hp01	0.04 (1.0000)	-0.06 (1.0000)	0.02 (1.0000)
hp05	-0.15 (0.2981)	-0.04 (1.0000)	0.17 (0.1559)
hp50	-0.54 (0.0000)	-0.12 (1.0000)	0.56 (0.0000)
hp95	-0.66 (0.0000)	-0.20 (0.0387)	0.73 (0.0000)
hp99	-0.66 (0.0000)	-0.20 (0.0280)	0.73 (0.0000)
hp90_hp10	-0.66 (0.0000)	-0.19 (0.0453)	0.72 (0.0000)
crr	0.00 (1.0000)	0.03 (1.0000)	-0.02 (1.0000)
vegcover	-0.10 (1.0000)	-0.07 (1.0000)	0.14 (0.5272)
percretabmean	-0.11 (1.0000)	-0.03 (1.0000)	0.12 (1.0000)
percretabmode	-0.02 (1.0000)	-0.05 (1.0000)	0.05 (1.0000)

Forest structure patterns

The results coming from the agglomerative hierarchical clustering allowed identification of 5 clusters of forest structure. Consequently the application of k-means algorithm with $k = 5$ in the partitive clustering provides an optimization of this classification.

According to the results, we classified our plots in pole-stage, young, adult, mature, and old forests. Figure 2 shows the dendrogram of the forest structure patterns identified by the cluster analyses after applying the k-means algorithm. Most parts of the forests in Trento province are old forests (30%), while the lower part is represented by pole-stage forests (7%). Young forests represent almost 15% of all forest patterns, and adult and mature forests are equally distributed (24.3%).

Table 5 reports the results from the non-parametric Kruskal-Wallis tests together with the

average values of LiDAR metrics for the five forest structure patterns. These results suggest that different patterns are detected with the hierarchical classification method. Out of 18, 11 LiDAR variables (hmean, hmode, hstd, hcv, hiqdist, hkur, haad, hp50, hp95, hp99, hp90_hp10), with p -values lower than 0.05, resulted in statistically distinguishable differences between the five patterns.

From the interpretation of the results, as indicated by LiDAR metrics values, pole-stage forests have mean height of 13 m, standard deviation of height around 2.6 m, height at the 99th percentile of 17.8 m.

As indicated by LiDAR metrics mean values, young forests are taller than pole-stage forests (mean height of 14 m), standard deviation of height is around 4 m, and height at the 99th percentile around 22 m. According to the average values of LiDAR metrics, in adult forests the mean height is around 16.3 m, stand-

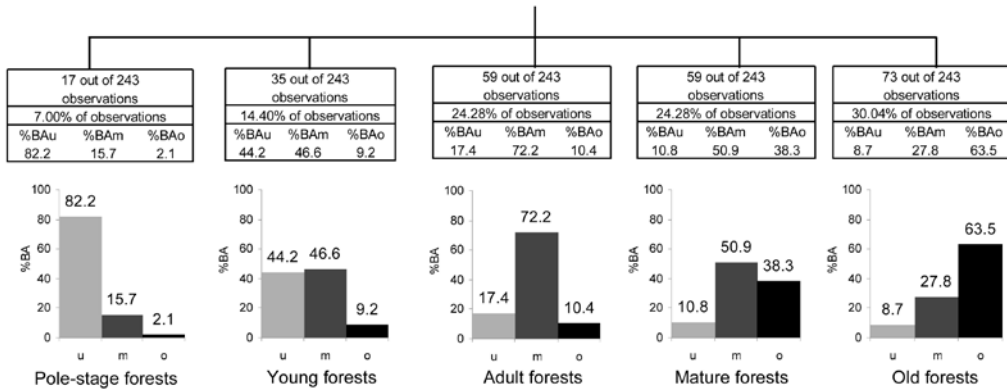


Figure 2 Characteristics of the five forest structure patterns resulting from the cluster analysis. Abbrev.: BAu - basal area of understory, BAm - basal area of mid-story, BAo - basal area of overstory.

Table 5 Average values of LiDAR metrics for the five forest structure patterns and *p*-values from the Kruskal-Wallis test

Metric name	%BA understory trees	%BA mid-story trees	%BA overstory trees
hmean	-0.54 (0.0000)	-0.13 (0.7205)	0.57 (0.0000)
hmode	-0.49 (0.0000)	-0.09 (1.0000)	0.49 (0.0000)
hstd	-0.68 (0.0000)	-0.21 (0.0164)	0.76 (0.0000)
hcv	-0.28 (0.0002)	-0.06 (1.0000)	0.29(0.0001)
hiqdist	-0.60 (0.0000)	-0.20 (0.0364)	0.67 (0.0000)
hske	-0.02 (1.0000)	-0.06 (1.0000)	0.07 (1.0000)
hkur	0.30 (0.0000)	-0.02 (1.0000)	-0.24 (0.0034)
haad	-0.66 (0.0000)	-0.21 (0.0205)	0.73 (0.0000)
hp01	0.04 (1.0000)	-0.06 (1.0000)	0.02 (1.0000)
hp05	-0.15 (0.2981)	-0.04 (1.0000)	0.17 (0.1559)
hp50	-0.54 (0.0000)	-0.12 (1.0000)	0.56 (0.0000)
hp95	-0.66 (0.0000)	-0.20 (0.0387)	0.73 (0.0000)
hp99	-0.66 (0.0000)	-0.20 (0.0280)	0.73 (0.0000)
hp90_hp10	-0.66 (0.0000)	-0.19 (0.0453)	0.72 (0.0000)
crr	0.00 (1.0000)	0.03 (1.0000)	-0.02 (1.0000)
vegcover	-0.10 (1.000)	-0.07 (1.0000)	0.14 (0.5272)
percretabmean	-0.11 (1.0000)	-0.03 (1.0000)	0.12 (1.000)
percretabmode	-0.02 (1.0000)	-0.05 (1.0000)	0.05 (1.0000)

ard deviation is 4.6 m and height at the 99th percentile is around 25 m.

Considering the resulting analysis, mature forests have mean height of 16 m, with standard deviation of height around 6 m, height at the 99th percentile of 30 m.

As indicated by LiDAR metrics mean values, old forests have mean height of 21 m, standard

deviation of height around 7 m, height at the 99th percentile of 33.5 m.

Classification trees to predict forest structure and model validation

The classification tree selected five ALS metrics to predict the forest structural types: stand-

ard deviation and mode of canopy heights, height at which 95% and 99% of canopy heights fall below, difference between height at which 90% and 10% of canopy heights fall below.

The structure of classification tree to predict forest patterns is showed in Figure 3: the values in the rectangles show the number and percentage of observations in the node. The tree model provided classification rules to predict the five forest patterns as in the following lines.

A stand is a pole-stage forest if the standard deviation of LiDAR heights intersecting the woodland is lower than 2.3 m. We can classify a stand as young forest if the standard deviation of LiDAR heights is between 2.3 m and 4.3 m and the 95th percentile of LiDAR heights is lower than 15 m or the standard deviation is in the same range but the 95th percentile is bigger than 15 m and the 99th percentile of the

LiDAR heights is bigger or equal than 22.4 m. We can distinguish adult forests from mature forests because the former have a difference between 90th percentile and 10th percentile of LiDAR heights bigger or equal to 12.2 m while in the mature forest this difference is lower than 12.2 m. A forest stand can be classify as old forest when the standard deviation of LiDAR heights is bigger or equal than 5.1 m and the 95th percentile of LiDAR heights is bigger than 29 m or when the standard deviation is bigger or equal than 7.0 m but the 95% of the LiDAR heights is smaller than 29 m.

Table 6 reports the results of the internal validation of the tree model. The basal error, or the error for the null model, calculated as difference between 1 and the forest structure pattern most represented, reached 70.0%. The misclassification error reached 36.2%. Omission error ranged between 19.2% (old forests) and 52.5% (mature forests), while commission

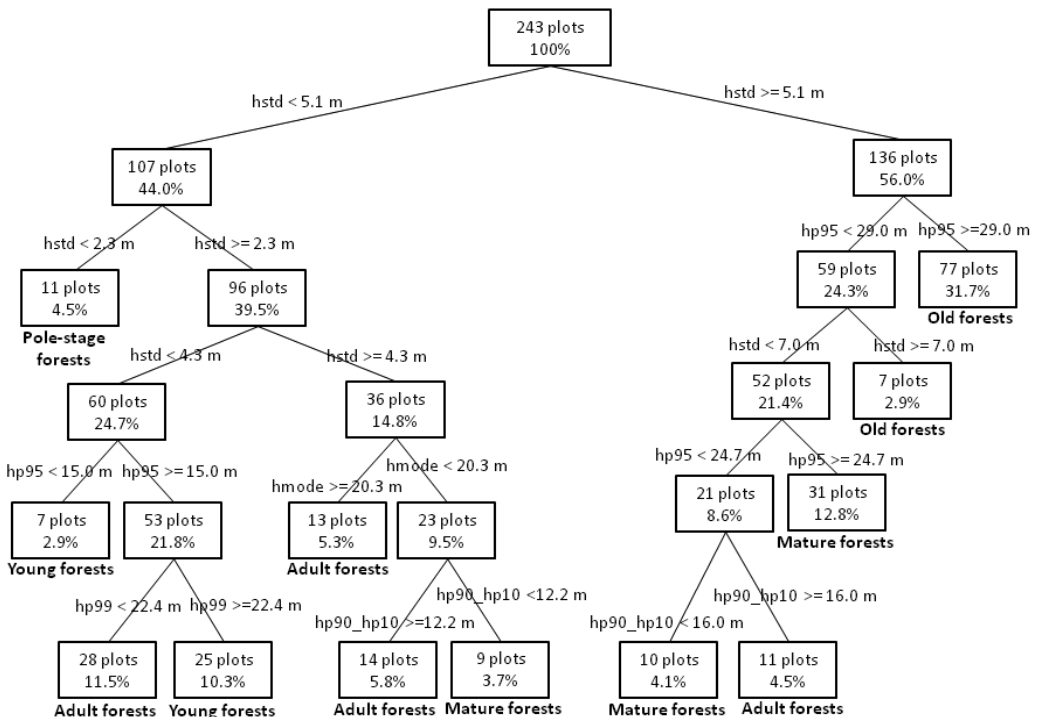


Figure 2 Classification tree predicting the five forest patterns

Table 6 Contingency table resulting from the classification of five forest structural pattern by the classification tree (internal validation on all observations)

Stage/accuracy	Predicted					Total	User accuracy (%)	Commission error (%)	
	Pole-stage forests	Young forests	Adult forests	Mature forests	Old forests				
Observed	Pole-stage forests	9	2	0	0	0	11	81.8	18.2
	Young forests	2	19	8	3	0	32	59.4	40.6
	Adult forests	5	9	40	7	5	66	60.6	39.4
	Mature forests	1	5	7	28	9	50	56.0	44.0
	Old forests	0	0	4	21	59	84	70.2	29.8
Total	17	35	59	59	73	243			
Producer accuracy (%)	53.0	54.3	67.8	47.5	80.8	Total accuracy (%)	63.8		
Omission error (%)	47.0	45.7	32.2	52.5	19.2	Misclassification error (%)	36.2		

Table 7 Contingency table of external validation resulting from the classification of five forest structural patterns by the classification tree (training set made up with 75% of observations)

Stage/accuracy	Predicted					Total	User accuracy (%)	Commission error (%)	
	Pole-stage forests	Young forests	Adult forests	Mature forests	Old forests				
Observed	Pole-stage forests	7	1	0	0	0	8	87.5	12.5
	Young forests	1	20	7	1	1	30	66.7	33.3
	Adult forests	6	4	24	10	2	46	52.2	47.8
	Mature forests	0	3	5	29	11	48	60.4	39.6
	Old forests	0	0	6	7	37	50	74.0	26.0
Total	14	28	42	47	51	182			
Producer accuracy (%)	50.0	71.4	57.1	61.7	72.5	Total accuracy (%)	64.3		
Omission error (%)	50.0	28.6	42.9	38.3	27.5	Misclassification error (%)	35.7		

Table 8 Contingency table of external validation resulting from the classification of five forest structural patterns by the classification tree (validation set made up with 25% of observations)

Stage/accuracy	Predicted					Total	User accuracy (%)	Commission error (%)	
	Pole-stage forests	Young forests	Adult forests	Mature forests	Old forests				
Observed	Pole-stage forests	1	0	0	0	0	1	100	0.0
	Young forests	2	4	1	0	1	8	50.0	50.0
	Adult forests	0	2	7	2	0	11	63.6	36.4
	Mature forests	0	1	7	8	8	24	33.3	66.7
	Old forests	0	0	2	2	13	17	76.5	23.5
Total	3	7	17	12	22	61			
Producer accuracy (%)	33.3	57.1	41.2	66.7	59.1	Total accuracy (%)	54.1		
Omission error (%)	66.7	42.9	58.8	33.3	40.9	Misclassification error (%)	45.9		

error was between 18.2% (pole-stage forests) and 44.0% (mature forests). There are no cases in which old forests have been classified as pole-stage or young forests and vice versa.

Table 7 and Table 8 report the results of the validation of tree model in the training and validation data set respectively. The misclassification error reached 45.9% in the validation data set. No pole-stage forests were classified as old forests, while 2 out of 8 cases of young forests were classified as old forest (25.0%), no adult forests were classified as old forests, and 8 out of 24 cases of mature forests were classified as old forests (33.3%).

Discussion

This study examined the performance of classification trees to predict structure types of mountain forests in the Italian Alps by metrics extracted from ALS data.

This study put together observations of dendrometric measurements from different data sets. Our observations were first of all widely distributed over the area of the territory of Trento province. Second, they included cases with very different basal area distributions in the three dbh classes in order to compose a proper sample size to generate reliable decision trees.

In all data sets, species were recorded but this information was not necessary for the aims of the study. Moreover, diameter of trees was measured starting from different thresholds (7.5, 4.5 and 2.5 cm), but this was not a problem considering that the trees are subdivided in the three diameter classes from the limit of 17.5 cm. Plots have different size because the field surveys were made with various aims, but many studies aimed at analysing the influence of plot size (and LiDAR density in some cases) on forest structure attribute estimates (i.e. Strunk et al. 2012, Watt et al. 2013, Ruiz et al. 2014) have demonstrated that minimum plot areas of 500-600 m² are needed for ba-

sal area estimates and that larger plot sizes do not significantly increase the accuracy of the estimates, but they do, however, increase the cost of fieldwork. We want point out that the assessment of the influence of plot size on forest structure attribute by means of LiDAR data was not a goal of this study.

In consideration of all of the above, the results of this study has evidenced that the proportion of basal area in dbh classes is a valuable parameter to distinguish different forest structural types and confirm that ALS data are a useful information source to predict forest structure types.

In previous work in which the use of LiDAR-derived CHM metrics were used to predict twelve forest structures types defined by the forest structural classification system used by the Forest Service of Trento Province, Torresan et al. (2014) indicated the possibility of investigating the prediction using a simplified number of structures (considering for example mono-layered, bi-layered and the multi-layered forest stands). In this study, five patterns of forest structure have been identified by the unsupervised cluster analysis based on the distribution of basal area in understory, mid-story and overstory trees: pole-stage, young, adult, mature and old forests. The analyses have drawn attention to the fact that the stands of the study area here considered are comprised of mature forests with a trend toward ageing. In fact, most forest stands in Trento province are old forests, with a large percentage of overstory-trees (63.5%). Considering that old trees do not act simply as senescent carbon reservoirs but actively fix large amounts of carbon compared to smaller trees (Stephensons et al. 2014), this is a good news, as the Trento forests are major components of the global carbon cycle, providing substantial feedback to atmospheric greenhouse gas concentrations. In any case, the results from our study can be used by forest managers to address the development of the forests in specific directions with the appropriate silvicultural interventions.

The level of user's accuracy, which refers to the probability that a forest classified as a certain structure type is really this type, of the classification tree model developed in this study ranged between 100% and 33% in the validation data set made up with 25% of observations. The highest level of user's accuracy was reached in the classification of pole-stage forests (100%) in which more than 82% of basal area is due to the understory-trees. This could be due to the fact that the pole-stage forests are dense and homogeneous forests in term of structure, so in this cases LiDAR pulses are detected by a sort of "blanket". When the complexity of the structure increases (multi-layered structures), the probability that LiDAR pulses penetrated below the canopy decreased by interference of middle and understory strata. This caused considerable impact on density of LiDAR points below the canopy, which could affect the accuracy of laser metrics (Ediriweera et al. 2014). This is confirmed also by the fact that the prediction of old forests types (63.5% of basal area due to the overstory-trees) from LiDAR data achieved high level of user's accuracy (76.5%).

Due to the lack of references that used classification trees as machine learning techniques to predict types of forests - instead of single forest biophysical variables - and hence due to the difficulty of finding values of classification accuracy, the discussion of our results relative to other findings in the literature is based on the most similar studies.

The classification results of forest types, presented in Zhao et al. (2011), showed that LiDAR data discriminate between pines and hardwood/mixed work better through support vector machine classifiers than through multivariate linear regression. To train and test the support vector machine classifier, Zhao et al. (2011) randomly select two samples of 5000 pixels at a 20 m resolution, being the class labels of these pixels indirectly obtained from the Quickbird-derived classification map by aggregating it from a resolution of 2.5 m to 20

m. Over 50 runs, after the fifth, support vector machine consistently outperformed multivariate linear regression with an averaged overall accuracy of 80.68% for multivariate linear regression versus 82.27 % for support vector machine. These values are clearly higher than those obtained in the study discussed here, but Zhao et al. (2011) discriminated just between young pine stands and forest stands characterized by relatively complex structure.

Chirici et al. (2013) attempted to classify nine forest fuel types in the Mediterranean province of Palermo and Catania (Italy) previously observed and identified by photo interpretation. They applied the random forest technique using thirty-one ALS-based metrics calculated from the normalized height returns in conjunction with an IRS LISS-III image. The overall accuracy obtained through this technique was 45%, lower than those obtained in our study.

Torresan et al. (2014) evaluated two approaches to predict twelve forest structure classes according to the amount of basal area present in understory, midstory, and overstory trees: in the first, supervised classification with both linear discriminant analysis and random forest were attempted; in the second, prediction of basal areas of lower, mid, and upper canopy trees from CHM-derived variables by k-nearest neighbour imputation and parametric regression were conducted to then classified the observations based on their predicted basal areas. The level of user's accuracy of the random forest (a learning method) ranged between 31% and 62%, which are levels lower than those obtained by the classification tree model of this study.

With regard to the LiDAR variables selected by learning machine approaches, as Görgens et al. (2015a) reported, when the process of selecting and reducing variables is automated, the metrics chosen might not rely on biological assumptions and often differ completely between studies.

For example, referring to the two studies

mentioned above, Zhao et al. (2011) used the canopy density and quantile height to produce a hybrid form of LiDAR composite metric as predictors, while Chirici et al. (2013) established canopy cover (defined as the percent of non-ground returns) as the most relevant metric and other relevant metrics included the spectral information from IRS and several other ALS metrics such as percentiles of the height distribution, the mean height of all returns, and the number of returns.

Conversely, the classification tree model developed in our study, among the original eighteen variables, it selected five variables: standard deviation and mode of canopy heights, height at which 95% and 99% of canopy heights fall below, difference between height at which 90% and 10% of canopy heights fall below. This confirms Gørgens et al. (2015b) findings about the specificity of LiDAR metrics as predictors and the need of identifying stable metrics derived from ALS data to be used as independent variable in specific models.

Classification trees are useful tools with a simple structure, that is with low number of rules and final nodes. In these cases the classification tree constitutes a practical and parsimonious tool for classification of stands for management or planning. The obtained classification tree model is clearly and easily interpretable, so that forest technicians can feasibly use it under operational contexts. The model established here provides moderately satisfactory results in terms of classification performance. We are aware that a misclassification error around 46% in the validation set is still not trusted for practical decision making purposes, but it is reasonably acceptable when the model is used for preliminary analysis in the context of forest inventory processes that may require the recognition of forest structural types for stratification.

Conclusions

A proper understanding of forest structure is

one of the keys to the sustainable management of forests. The description and mapping of forest structure are important aspects for forest inventory purposes (e.g. in the stratification process of a forest estate), for individualization of silvicultural interventions needed to achieve a diversified landscape, to drive a specific forest structure toward desired conditions, for addressing habitat management according to wildlife animals species needs, etc.

Forest technicians often deal with processes of discrimination between different types of forest structures for forests management purposes. Activities, such as photo interpretation or fieldwork, traditionally used for forest structure interpretation are time-consuming and expensive. Although remote sensing is not expected to replace completely field measurement any time in the near future (McRoberts & Tomppo 2007) it would facilitate planning and management with realistic goals.

Our study demonstrated that in the considerably mountainous forests LiDAR data and classification trees modelling would be useful and affordable for assisting in the classification of forest structure types, though caution is required to deal with inherent modelling uncertainties.

Our study represents the first contribution in the context of forest management in Italian Alps to the process of forest structure types discrimination using airborne laser scanner data. Substantial room for improvement is expected by coupling multi- or hyperspectral imagery, which this offers extensive spatial and spectral coverage of measurements and represents an ideal tool for forest mapping at provincial scale. As Dalponte et al. (2009) pointed out, remote sensing hyperspectral sensors are important and powerful instruments for addressing classification problems in complex forests, due to the obtainable detailed characterization of the spectral behaviour of the forest structure types. Therefore, future studies could investigate the potential of the integration of hyperspectral and LiDAR data on the accuracy of classification trees and other machine learning

techniques in the forest structure classification.

References

- Axelsson P., 1999. Processing of laser scanner data - algorithms and applications. *ISPRS Journal of Photogrammetry and Remote Sensing* 54(2): 138-147. DOI: 10.1016/S0924-2716(99)00008-8.
- Chirici G., Scotti R., Montagni A., Barbati A., Cartisano R., Lopez G., Marchetti M., McRoberts R.E., Olson H., Corona P., 2013. Stochastic gradient boosting classification trees for forest fuel types mapping through airborne laser scanning and IRS LISS-III imagery. *International Journal of Applied Earth Observation and Geoinformation* 25: 87-97. DOI: <http://dx.doi.org/10.1016/j.jag.2013.04.006>.
- Coops N.C., Hilker T., Wulder M.A., St-Onge B.A., Newnham G.J., Siggins A., Trofymow J.A., 2007. Estimating canopy structure of Douglas-fir forest stands from discrete-return LiDAR. *Trees - Structure and Function* 21: 295-310. DOI: 10.1007/s00468-006-0119-6.
- Corona P., 2010. Integration of forest mapping and inventory to support forest management. *iForest - Biogeosciences and Forestry* 3: 59-64. DOI: 10.3832/ifor0531-003.
- Corona P., Fattorini L., 2008. Area-based lidar-assisted estimation of forest standing volume. *Canadian Journal of Forest Research* 38: 2911-2916. DOI: 10.1139/X08-122.
- Corona P., Fattorini L., Franceschi S., Scrinzi G., Torresan C., 2014. Estimation of standing wood volume in forest compartments by exploiting airborne laser scanning information: model-based, design-based, and hybrid perspectives. *Canadian Journal of Forest Research* 44: 1303-1311. DOI: 10.1139/cjfr-2014-0203.
- Corona P., Cartisano R., Salvati R., Chirici G., Floris A., Di Martino P., Marchetti M., Scrinzi G., Clementel F., Travaglini D., Torresan C., 2012. Airborne Laser Scanning to support forest resource management under alpine, temperate and Mediterranean environments in Italy. *European Journal of Remote Sensing* 45: 27-37. DOI: 10.5721/EurJRS20124503.
- Dalponte M., Bruzzone L., Gianelle D., 2008. Fusion of Hyperspectral and LiDAR Remote Sensing Data for Classification of Complex Forest Areas. *Geosciences and Remote Sensing, IEEE Transactions on* 46(5): 1416-1427. DOI: 10.1109/TGRS.2008.916480.
- Dalponte M., Bruzzone L., Vescovo L., Gianelle D., 2009. The role of spectral resolution and classifier complexity in the analysis of hyperspectral images of forest areas. *Remote Sensing of Environment* 113 (2009) 2345-2355.
- Drake J.B., Dubayah R.O., Clark R.O., Knox D.B., Blair J.B., Fofon M.A., Chazdon R.L., Weishampel J.F., Prince S., 2002. Estimation of tropical forest structural characteristics using large-footprint LiDAR. *Remote Sensing of Environment* 79(2-3): 305-319.
- Durbha S.S., King R.L., Younan N.H., 2007. Support vector machines regression for retrieval of leaf area index from multiangle imaging spectroradiometer. *Remote Sensing of Environment* 107, 348-361.
- Ediriweera S., Pathirana S., Danaher T., Nichols D., 2014. LiDAR remote sensing of structural properties of subtropical rainforest and eucalypt forest in complex terrain in north-eastern Australia. *Journal of Tropical Forest Science* 26(3): 397-408.
- Everitt B.S., Landau S., Leese M., 2001. *Cluster Analysis*. 4th Edition, Oxford University Press, Inc., New York; Arnold, London.
- Fallah A., Kalbi S., Shataee S., 2013. Forest stand types classification using tree-based algorithms and SPOT-HRG Data. *The International Journal of Environmental Resources Research* 1(3): 31-45.
- Garcia-Gutierrez J., Martinez-Álvarez F., Troncoso A., Riquelme J.C., 2014. A comparative study of machine learning regression methods on LiDAR data: A case study. *International Joint Conference SOCO'13-CIS-IS'13-ICEUTE'13, Advances in Intelligent Systems and Computing* 239: 249-258.
- Ghose M.K., Pradhan R., Ghose S.S., 2010. Decision tree classification of remotely sensed satellite data using spectral separability matrix. *International Journal of Advanced Computer Science and Applications* 1(5): 93-101.
- Gleason C.J., Im J., 2012. Forest biomass from airborne LiDAR data using machine learning approaches. *Remote Sensing of Environment* 125: 80-91.
- Gómez C., Wulder M.A., Montes F., Delgado J.A., 2012. Modeling forest structural parameters in the Mediterranean pines of Central Spain using QuickBird-2 imagery and classification and regression tree analysis (CART). *Remote Sensing* 4: 135-159. DOI: 10.3390/rs4010135.
- Görgens E.B., Montagni A., Estraviz Rodriguez L.C., 2015a. A performance comparison of machine learning methods to estimate the fast-growing forest plantation yield based on laser scanning metrics. *Computers and Electronics in Agriculture* 116: 221-227. DOI: 10.1016/j.compag.2015.07.004.
- Görgens E.B., Packalen P., da Silva A.G.P., Alvares C.A., Campoe O.C., Stape J.L., Rodriguez L.C.E., 2015b. Stand volume models based on stable metrics as from multiple ALS acquisitions in *Eucalyptus* plantations. *Annals of Forest Science* 72(4): 489-498. DOI: 10.1007/s13595-015-0457-x.
- Griffiths P., Kuemmerle T., Baumann M., Radeloff V.C., Abrudan I.V., Lieskovsky J., Munteanu C., Ostapowicz K., Hostert P., 2012. Forest disturbances, forest recovery, and changes in forest types across the Carpathian ecoregion from 1985 to 2010 based on Landsat image composites. *Remote Sensing of Environment* 151: 72-88. DOI: 10.1016/j.rse.2013.04.022.
- Holmgren J., Jonssohn, T. 2004. Large scale airborne laser-scanning of forest resources in Sweden. In: Thies, M., Koch, B., Spiecker, H., Weinacker, H. (eds.), *Laser*

- scanners for forest and landscape assessment. Proceedings of the ISPRS Working Group VIII/2. Freiburg, Germany, October 3-6, 2004. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences. vol. XXXVI, Part 8/W2, pp. 157-160.
- Hudak A.T., Evans J.S., Smith A.M.S., 2009. LiDAR utility for natural resource Managers. *Remote Sensing* 1(4): 934-951. DOI: 10.3390/rs1040934.
- Hyypää J., Hyypää H., Leckie D., Gougeon F., Yua X., Maltamo M., 2008. Review of methods of small footprint airborne laser scanning for extracting forest inventory data in boreal forests. *International Journal of Remote Sensing* 29(5):1339-1366. DOI: 10.1080/01431160701736489.
- Isenburg M., Trunzer H., Malmer F., 2014. Experiences with LiDAR canopy penetration in a dense tropical rainforest. 35th Asian Conference on Remote Sensing 2014, ACRS 2014: Sensing for Reintegration of Societies.
- Ke Y., Quackenbush L.J., Im J., 2010. Synergistic use of QuickBird multispectral imagery and LiDAR data for object-based forest species classification. *Remote Sensing of Environment* 114(2010): 1141-1154.
- Lefsky M.A., Cohen W.B., Acker S.A., Parker G.G., Spies T.A., Harding D., 1999. Lidar remote sensing of the canopy structure and biophysical properties of Douglas-fir western hemlock forests. *Remote Sensing of Environment* 70(3): 339-361.
- Lefsky M.A., Cohen W.B., Acker S.A., Parker G.G., Spies T.A., Parker G.G., Harding, D., 1998. Lidar remote sensing of forest canopy structure and related biophysical parameters at H.J. Andrews experimental forest, Oregon, USA. *Remote Sensing of Environment* 70(3): 339-361.
- Li M., Im J., Beier C., 2013. Machine learning approaches for forest classification and change analysis using multi-temporal Landsat TM images over Huntingdon Wildlife Forest. *GIScience & Remote Sensing* 50(4): 361-384, DOI: 10.1080/15481603.2013.819161.
- Lim K., Treitz P., Baldwin K., Morrison I., Green J., 2003b. LiDAR remote sensing of biophysical properties of tolerant northern hardwood forests. *Canadian Journal of Remote Sensing* 29: 648-678.
- Lim K., Treitz P., Wulder M., St-Onge B., Flood M., 2003a. LiDAR remote sensing of forest structure. *Progress in Physical Geography* 27(1): 88-106.
- Maltamo M., Packalén P., Yu X., Eerikäinen K., Hyypää J., Pitkänen J., 2005. Identifying and quantifying structural characteristics of heterogeneous boreal forests using laser scanner data. *Forest Ecology and Management* 216: 41-50. DOI: 10.1016/j.foreco.2005.05.034.
- McGaughey R.J., 2012. FUSION/LDV: Software for LiDAR data analysis and visualization, Version 3.01. USFS.
- McRoberts R.E., Tomppo E.O., 2007. Remote sensing support for national forests inventories. *Remote Sensing of Environment* 110: 412-419.
- 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. *Photogrammetric Engineering & Remote Sensing* 66(11): 1367-1371.
- Moss I., 2012. Stand structure classification, succession, and mapping using LiDAR. A dissertation submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in the Faculty of Graduate Studies (Forestry), The University of British Columbia (Vancouver), April 2012.
- Næsset E., 2002. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment* 80(1): 88-99.
- Odasso M., 2002. I tipi forestali del Trentino [Forest types of Trentino]. Report del Centro di Ecologia Alpina, 25, 192 p.
- Parker G.G., Russ E.M., 2004. The canopy surface and stand development assessing forest canopy structure and complexity with near-surface altimetry. *Forest Ecology and Management* 189: 307-315. DOI: 10.1016/j.foreco.2003.09.001.
- Pastorella F., Paletto A., 2013. Stand structure indices as tools to support forest management: an application in Trentino forests (Italy). *Journal of Forest Science* 59(4): 159-168.
- PAT (Provincia Autonoma di Trento) e Servizio Foreste e Fauna, 2008. Rapporto sullo stato delle foreste e della fauna [State of the forests and fauna report]. Testo a cura di Stefano Dellantonio e Cristina Gandolfo.
- Pike R.J., Wilson S.E., 1971. Elevation-relief ratio, hypsometric integral, and geomorphic area-altitude analysis. *Bulletin of the Geological Society of America* 82: 1079-1084. DOI: 10.1130/0016-7606(1971)82[1079:ER-HIAG]2.0.CO;2.
- Reutebuch S.E., McGaughey R.J., Andersen H.-E., Carson W.W., 2003. Accuracy of a high-resolution lidar terrain model under a conifer forest canopy. *Canadian Journal of Remote Sensing* 29(5): 527-535.
- Ruiz L.A., Hermosilla T., Mauro F., Godino M., 2014. Analysis of the influence of pot size and LiDAR density on forest structure attribute estimates. *Forests* 5: 936-951. DOI: 10.3390/f5050936.
- Scrinzi G., Floris A., 2003. Procedure di individuazione, materializzazione e ritrovamento dei punti campionamento [Procedures for identifying, materialization and discovery of sampling points]. *Inventario Nazionale delle Foreste e dei Serbatoi forestali del Carbonio – INFC, Corpo Forestale dello Stato. Isafa, Trento, 48 p.*
- Šebeň V., Bošefa M., 2010. Different approaches to the classification of vertical structure in homogeneous and heterogeneous forests. *Journal of Forest Science* 56(4): 171-176.
- Sherrill K.R., Lefsky M.A., Bradford J.B., Ryan M.G., 2008. Forest structure estimation and pattern exploration from discrete-return lidar in subalpine forests of the central Rockies. *Canadian Journal of Forest Research* 38: 2081-2096. DOI: 10.1139/X08-059.
- Stephenson N.L., Das A.J., Condit R., Russo S.E., Baker

- P.J., Beckman N.G., Coomes D.A., Lines E.R., Morris W.K., Ruger N., Alvarez E., Blundo C., Bunyavejchewin S., Chuyong G., Davies S.J., Duque A., Ewago C.N., Flores O., Franklin J.F., Grau H.R., Hao Z., Harmon M.E., Hubbell S.P., Kenfack D., Lin Y., Makana J.-R., Malizia A., Malizia L.R., Pabst R.J., Pongpattananurak N., Su S.-H., Sun I-F., Tan S., Thomas D., van Mantgem P.J., Wang X., Wiser S.K., Zavala M.A., 2014. Rate of tree carbon accumulation increases continuously with tree size. *Nature* 507: 90-93. DOI: 10.1038/nature12914.
- Strunk J., Temesgen H., Andersen H.E., Flewelling J.P., Madsen L., 2012. Effects of LiDAR pulse density and sample size on a model-assisted approach to estimate forest inventory variables. *Canadian Journal of Remote Sensing* 38(5): 644-654.
- Torresan C., Strunk J., Zald H. S. J., Zhiqiang Y., Cohen W. B., 2014. Comparing statistical techniques to classify the structure of mountain forest stands using CHM-derived metrics in Trento province (Italy). *European Journal of Remote Sensing* 47: 75-94. DOI: 10.5721/EuJRS20144706.
- Vierling K.T., Vierling L.A., Gould W.A., Martinuzzi S., Clawges R.M. 2008. Lidar: shedding new light on habitat characterization and modeling. *Frontiers in Ecology and the Environment* 6(2): 90-98.
- Ward J.H., 1963. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association* 58(301): 236-244. DOI: 10.1080/01621459.1963.10500845.
- Watt M.S., Adams T., Gozalez Aracil S., Marshall H., Watt P., 2013. The influence of LiDAR pulse density and plot size on the accuracy of New Zealand plantation stand volume equations. *New Zealand Journal of Forestry Science* 43:15.
- Wulder M.A., White J.C., Nelson R.F., Næsset E., Ørka H.O., Coops N.C., Hilker T., Bater C.W., Gobakken C., 2012. Lidar sampling for large-area forest characterization: A review. *Remote Sensing of Environment* 121: 196-209.
- Zhao K., Popescu S., Meng X., Pang Y., Agca M., 2011. Characterizing forest canopy structure with lidar composite metrics and machine learning. *Remote Sensing of Environment* 115(8): 1978-1996. DOI: 10.1016/j.rse.2011.04.001.
- Zhao K., Popescu S.C., Zhang X., 2008. Bayesian learning with Gaussian processes for supervised classification of hyperspectral data. *Photogrammetric Engineering & Remote Sensing* 74(10): 1223-1234.