Comparing statistical techniques to classify the structure of mountain forest stands using CHM-derived metrics in Trento province (Italy)

Chiara Torresan, Jacob Strunk, Harold S.J. Zald, Yang Zhiqiang & Warren B. Cohen

Published online: 17 Feb 2017.

Submit your article to this journal

Article views: 47

View related articles

View Crossmark data

Comparing statistical techniques to classify the structure of mountain forest stands using CHM-derived metrics in Trento province (Italy)

Chiara Torresan[1]*, Jacob Strunk[2], Harold S.J. Zald[3],
Yang Zhiqiang[3] and Warren B. Cohen[3]

[1]Department for Innovation in Biological, Agro-food and Forest Systems,
University of Tuscia, Viterbo, Italy,
[2]Department of Forest Engineering, Resources and Management, College of Forestry,
Oregon State University, 204 Peavy Hall, Corvallis, OR 97331, USA
[3]U.S.D.A. Forest Service, Pacific Northwest Research Station,
3200 SW Jefferson Way, Corvallis, OR 97731, USA
*Corresponding author, e-mail address: chiara.torresan@entecra.it

Abstract
In some cases a canopy height model (CHM) is the only available source of forest height information. For these cases it is important to understand the predictive power of CHM data for forest attributes. In this study we examined the use of lidar-derived CHM metrics to predict forest structure classes according to the amount of basal area present in understory, midstory, and overstory trees. We evaluated two approaches to predict size-based forest classifications: in the first, we attempted supervised classification with both linear discriminant analysis (LDA) and random forest (RF); in the second, we predicted basal areas of lower, mid, and upper canopy trees from CHM-derived variables by k-nearest neighbour imputation (k-NN) and parametric regression, and then classified observations based on their predicted basal areas. We used leave-one-out cross-validation to evaluate our ability to predict forest structure classes from CHM data and in the case of prediction-based classification approach we look at the performances in predicting basal area. The strategies proved moderately successful with a best overall classification accuracy of 41% in the case of LDA. In general, we were most successful in predicting the basal areas of small and large trees (R^2 respectively of 71% and 69% in the case of k-NN imputation).

Keywords: forest structure, lidar, linear discriminant analysis, random forests, k-nearest neighbour imputation, parametric regression.

Introduction
Forest structure analysis and characterization is part of most forest planning processes for sustainable forest management and is of special interest from a silvicultural point of view. Knowledge about current forest structural conditions is crucial to understanding the specific

silvicultural prescriptions necessary to meet management objectives for forest attributes such as biomass, wildlife habitat, and species diversity.

The distribution of basal area in size classes allows forest managers to quantify the current structure and condition of forest stands and predict how this may change over time, and consequently to take management-decision [Ancel et al., 1999; Scrinzi et al., 2011].

Forest structure directly affect the capacity of stands to provide market products (i.e. wood) and no-market products (i.e. ecosystems services, including soil conservation and carbon sequestration), moreover forest structure contribute to the conservation of biological diversity [Staudhammer and LeMay, 2001].

Lidar, as an active remote sensing technology, has been used in many studies to describe and estimate forest structure in three-dimensions: lidar data for forestry applications involves processing the raw point cloud to compute a wide range of vegetation metrics to predict and map many aspects of forest composition and structure, including basal area [Næsset, 1997; Means et al., 1999; Dubayak et al., 2000; Means et al., 2000; Næsset, 2002; Andersen et al., 2003; Zimble et al., 2003; Riaño et al., 2004; Maltamo et al., 2005; Barilotti et al., 2007; Travaglini et al., 2007; Heurich and Thoma, 2008; Ferraz et al., 2012, Alberti et al., 2013]. There are times that the original raw lidar point cloud may be not available, in particular when lidar data are collected for other non-forestry purposes, and only the Digital Terrain Model (DTM) and Digital Surface Model (DSM) are at one's disposal. It has been typically believed that these two elevation models have limited applications for forestry purposes that require characterization of forest structure. Consequently, there has been limited empirical exploration of the potential use of only digital elevation models in forest characterization and classification [Suárez et al., 2012]. Despite declining costs and increasing coverage, lidar data are still not widely available in all locations for forestry applications, so the ability to use lidar-derived DTMs and DSMs has the potential to improve forest classification and mapping in areas where raw lidar data are otherwise not available.

We carried out the present study because we are interested in assessing whether lidar-derived predictor variables extracted from CHMs can support the classification of the forests in the province of Trento (Italy). Within this general purpose, the specific goal of this research was to compare the performance of supervised classification techniques toward specific methods of basal area prediction in the process of structural classification of the mountain forests, a theme in development in the Italian forestry research based on the exploitation of lidar data [Corona et al., 2012].

Methods
Study area
The province of Trento (6212 Km^2) is situated in the North-East of Italy on the Southern side of the Alps chain (Fig. 1).

The territory is almost entirely mountainous, and 60% of its surface is covered by forests [De Natale and Pignatti, 2011]. Approximately, 72% of the forest surface is owned by public institutions and 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.

Figure 1 - The study area: the Trento province.

The area of Trento province, according to its geographical and climatic gradients, can be divided into three zones: *endalpica*, *mesalpica* and *esalpica* [Odasso, 2002]. 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*) or cembro pine (*Pinus cembra)*. 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*) and beech (*Fagus sylvatica*). 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*, *Carpinus betulus*, *Fraxinus ornus*, *Quercus pubescens*, *Quercus petrae*, etc.).

The forest structural classification system used by the Forest Service of Trento Province
The Forest Service of Trento Province is nowadays using a structural classification system based on the distribution of basal area in specific tree size classes both in inventory and forest management decision making phases. The classification in forest structural types

is one of the primary reference (with the species composition) for the forest managers to identify the strata in the process of forest inventory based on a stratified sampling. During the decision making process, forest managers use the forest structural classification to identify the potential evolution of a stand, that is to say the potential transition of a specific structural type toward another one, and consequently decide which kind of silvicultural interventions to implement estimating the potential volume to fell in each tree size class.

The recently forest structural classification system adopted by the Forest Service of Trento Province is based on the percentage of basal area in three classes of trees [Scrinzi et al., 2011], defined as small trees (17.5 cm to 27.4 cm diameter at breast height, dbh), medium trees (27.5 to 47.4 cm dbh), and large trees (above 47.5 cm dbh).

When the percentage of basal area per hectare is greater than 15% in all classes, then the structure is classified as irregular, which could be small trees dominated (I1), medium tree dominated (I2) or large tree dominated (I3). If the percentage of basal area per hectare is greater than 15% in two classes, then the stand is classified as regular dominated by small (R21 and R31), medium (R12 and R32) or large trees (R13 and R23). When the percentage of basal area per hectare is greater than 15% in only one class, then the stand is classified as regular widely dominated by small (R11), medium (R22) or large trees (R33). For example, if in a forest stand the basal area of small trees is 25%, the basal area of medium trees is 30% and the basal area of large trees is 45% then the stand is classified as irregular dominated by large trees, and identified by the code I3 (I stands for irregular and 3 is the number that distinguishes large trees).

Based on these combinations, the structure of forest stands will classified as one of twelve forest structural types, including three irregular types and nine regular types (Fig. 2).
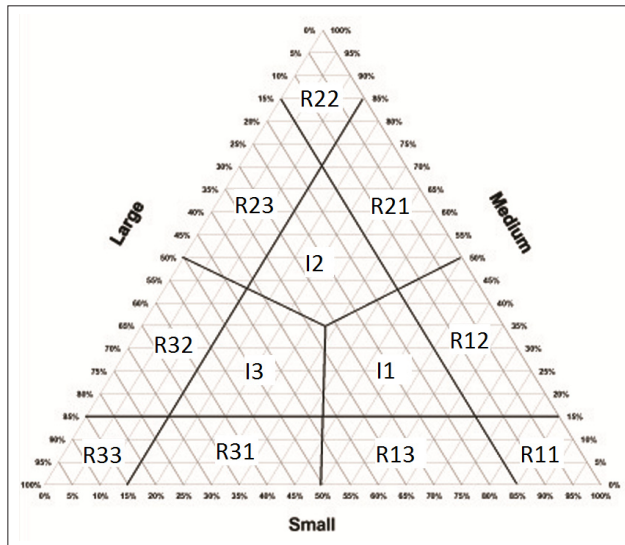

Figure 2 - Forest structural classification system used by the Forest Service of Trento Province where R=regular, I=irregular, 1=small trees, 2=medium trees, 3=large trees.

Field sampling and response variables

The number of field plots was established using a two step sampling design to ensure adequate representation of all structural types present in Trento province.

It is necessary to point out that in that province for management purposes (i.e. falls and loggings) the public forest surface subjected to a management plan is divided on *forest compartments*, a forest unit marked on the ground, 15 ha wide on average.

In the first step of the sampling, all forest compartments callipered for forest planning purposes from 1990 to 2006 and dominated by a single species (80% of species compositions) were selected. In the second step, the 1843 compartments selected through the first phase was pruned to 90, maintaining the proportional representation of different forest structural type. Finally, ground-truth point were randomly placed within the forest compartments ensuring the presence of two points per compartment, one as the point to be surveyed and the other as a reserve to cover those cases in which conditions in the field did not meet the target forest type.

The field surveys were carried out during the Summer (June-September) of 2007. The centre of each circular plot was georeferenced with a Thales MobileMapper CE Global Positioning System (GPS) receiver recording at minimum 200 satellite positions on the ground surface. Afterwards the geographical coordinates of plot centres were differentially corrected and averaged obtaining in this way a plot location accuracy of less than 3 m. In this study, different plot sizes (531, 1257, 1964 and 2827 m²) were employed to ensure the minimum level of tree density in each plot. The dbh and height were measured for all trees in each plot. The diameter of each tree (dbh ≥7.5 cm) was measured from two orthogonal axes using a timber callipers at breast height (1.3 m) and then averaged. The height of each tree was measured by a Vertex-II ultrasonic hypsometer.

The dbh was measured to calculate the field-based proportion of basal area on each plot in the three size classes (small, medium and large) needed to classify the forest structure: both the proportion of basal area, as continuous variable, and forest structure type, as categorical variable, were used as response variable.

The heights and diameters were used to build a height-diameter model: in fact, although we already had field measured tree height, we modelled the tree height to obtain a diameter-height function working for all trees and for all different forest structural type. In this way it was possible to identify specific height thresholds in which the small, medium and large trees could be placed, and consequently threshold the prediction variables.

The model was built according to the Curtis [1967] equation:

$$h = 1.3 + \frac{d^2}{a + bd + cd^2} \quad [1]$$

where *h* is the estimated height of the tree, and *d* is dbh.

The main statistic parameters of the height-diameter function are reported in Table 1.

Table 1 - Main statistic parameters of the height-diameter model according to the Curtis (1967) equation.

Statistic parameter	Model
Dependent variable	h
Measure unit	m
Model equation	$h = 1.3 + (d^2 / a + bd + cd^2)$
a	42.31
b	-0.60
c	0.03
a significance level (p-value)	$p < 0.001$
b significance level (p-value)	$p < 0.05$
c significance level (p-value)	$p < 0.001$
R^2	0.51
Adjusted R^2	0.50

By means of the height-diameter function we estimated that the height of small trees ranged from 8.5 to 15.6 m, the height of medium trees ranged from the 15.6 to 26.5 m and large trees were greater than 26.5 m.

The 90 plots surveyed were classified into nine forest structure types: I2 (6 plots), I3 (3 plots), R11 (9 plots), R12 (7 plots), R21 (14 plots), R23 (22 plots), R31 (1 plot), R32 (25 plots), and R33 (3 plots).

Lidar sampling and lidar predictor variables

During the 2006-2007 autumn and winter seasons, the administrative Province of Trento commissioned Blom CGR S.p.A. an airborne laser scanning (ALS) campaign over its entire territory, aimed to produce a DTM to be used for land planning, hydro geological purposes, topographic measurements, etc. As known, the surveys made for these goals are frequently characterized by a relatively low density of points per square meter and, above all, they are usually made in winter (between November and March in temperate and Mediterranean environments and even under alpine environments, at least in those areas not permanently covered by snow in that period) to minimize the noise by vegetation, since the aim is to achieve a high penetration rate through the vegetation canopy [Corona et al., 2012].

Discrete return lidar data for the whole surface of Trento province were acquired in October, November and December 2006 and then in January, February, March, November and December 2007 using an Optech ALTM3100C laser system mounted in Partenavia P68 (Tab. 2). The lidar system recorded range and intensity of 2 returns per pulse, and achieved a nominal pulses density of 1.28 per m².

Table 2 - Lidar data acquisition parameters.

Sensor	Optech ALTM3100C
Acquisition date	October, November, December 2006, January, February, March, April, May, July, August, November, December 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

The raw data were filtered and classified by the vendor using Terrascan software (Terrasolid Ltd.) which utilizes the Axelsson [1999] filtering algorithm based on Triangulated Irregular Network densification. Data delivered by the vendor included a DSM and a DTM as a regular grid with a spatial resolution of 1 m². We calculated the CHM as the DSM minus the DTM by means of spatial analyst tools in ArcGis 9.x software (ESRI). Canopy height values greater than 2 m were assumed to be vegetation hits, this means that values less than 2 m were set to 0.

From the CHM for each of the 90 plots, grid cells, corresponding in shape and size at those of each field plot, were extracted using ArcGIS. The *maptools* package in R [Lewin-Koh and Bivand, 2012] was used to compute the predictor variables from the height probability distributions (thirteen predictor variables), and from the relative frequency distributions of vegetation heights (four predictor variables) of each plot: in total, seventeen candidate predictor variables were generated for modelling (Tab. 3).

Table 3 - Seventeen candidate predictor variables.

HMAX	Maximum of height pixels intersecting plot
HMEAN	Mean of height pixels intersecting plot
HKURT	Kurtosis of the height pixels distribution intersecting the plot
HSKEW	Skewness of the height pixels distribution intersecting the plot
HVAR	Variance of height pixels intersecting plot
HCV	Coefficient of variation of height pixels intersecting plot
HSTD	Standard deviation of height pixels intersecting plot
H05PCT	Height at which 5% of pixels intersecting plot fall below
H10PCT	Height at which 10% of pixels intersecting plot fall below
H25PCT	Height at which 25% of pixels intersecting plot fall below
H50PCT	Height at which 50% of pixels fall below (median of height pixels intersecting plot)
H75PCT	Height at which 75% of pixels intersecting plot fall below
H90PCT	Height at which 90% of pixels intersecting plot fall below
COVSMALL	Cover of small trees (ratio of the number of pixels intersecting plot with height >=8.5 m and <15.6 m over the total number of pixels intersecting plot)
COVMEDIUM	Cover of medium trees (ratio of the number of pixels intersecting plot with height >=15.6 m and <26.5 m over the total number of pixels intersecting plot)
COVLARGE	Cover of large trees (ratio of the number of pixels intersecting plot with height >=26.5 over the total number of pixels intersecting plot)
COV2	Cover of pre-inventory trees (ratio of the number of pixels intersecting plot with height >=2.0 m and <8.5 m over the total number of pixels intersecting plot)

Statistical analyses
Supervised classification approach
### Linear discriminant analysis
Linear discriminant analysis (LDA) is a supervised classification approach. In LDA, a

linear combination of auxiliary variables is identified which maximizes separation between categorical response groups [Reimann et al., 2008]:

$$w_1 = a_1 x_1 + a_2 x_2 + \dotsc + a_k x_k = \sum_{i=1}^{k} a_i x_i \quad [2]$$

The weights a_i are chosen to maximize the separation between groups. A classification rule is developed in combination with the weighting function in [2]. Using maximum likelihood theory, classifications are assigned to ranges of values. E.g., if $w_1 \in (l_1, l_2]$ for an observation where l_1 and l_2 represent limits for a given class, then the observation is assigned to that class. If $w_1 \notin (l_1, l_2]$ then the observation is assigned to an alternate class (Fig. 3).
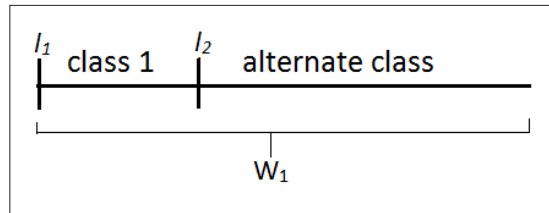
Figure 3 - Schematic describing assignment of LDA values (predictions) to response classes.

LDA was performed in R using the *MASS* package [Venables and Ripley, 2002], which implements a Bayesian decision theory approach. For LDA the covariates are assumed to have a common multivariate normal distribution.

Random forest

RF is a classification technique, based on the use of classification and regression trees, developed by Breiman [2001].

Classification trees build rules for assigning current observations into classes using numerical and/or categorical predictor variables by recursive binary partitioning into regions that are increasingly homogeneous with respect to the class variable. The homogeneous regions are called nodes [Cutler et al., 2007]. RF fits many classification trees to a data set, and then combines the predictions from all the trees. The algorithm begins with the selection of many (e.g., 500) bootstrap samples from the data. Observations in the original data set that do not occur in a bootstrap samples are called out-of-bag observations. A classification tree is fit to each bootstrap sample, and at each node, a small number of randomly selected variables are available for the binary partitioning. The trees are fully grown and each is used to predict the out-of-bag observations. The predicted class of an observation is calculated by majority vote of the out-of-bag predictions for that observation, with ties split randomly [Ok et al., 2012]. Breiman [2001] called this procedure random forest because the base constituents of the ensemble are tree-structured predictors, and because each of these trees is constructed using an injection of randomness.

The RF classification was carried out using the *randomForest* package in R [Liaw and Wiener, 2002]. The RF algorithm in R works as follows:

- grow a specified number of trees (n_{tree}) by the bootstrap method from the original data;
- for each of the bootstrap samples, grow an unpruned classification so that at each node randomly sample a number of variables (m_{try}) of the predictors as candidates at each split;
- predict new data by aggregating the predictions of the n_{tree} trees.

The randomForest package optionally produces the measure of the importance of the predictor variables, and a measure of the internal structure of the data, that give information about the proximity of different data points to one another.

Prediction-based classification approach
k-nearest neighbour (k-NN)

k-NN imputation predicts the value of the variable of interest as a weighted average of values of nearest neighbouring observations [Maltamo, 2003]. The neighbours are defined with some similarity (or distance) measure in the predicting variables, selecting the neighbours from the observations whose variables were previously measured.

For nearest neighbour search and imputation different methods can be used: in this study the most similar neighbour (MSN) and the RF methods were considered. In the MSN method, the nearness is defined using the weighted Euclidean distance [Crookston and Finley, 2008] based on canonical correlation analysis between the independent variables and the dependent variables. In the RF method, the observations are considered similar if they tend to end up in the same terminal nodes in a suitably constructed collection of classification and regression trees. The distance measure is one minus the fraction of trees in which two separate observations fall in the same terminal node.

The *yaImpute* package in R [Crookston and Finley, 2008] was used for *k*-NN imputation by the MSN and RF methods. In this package, any *k* number of reference observations can be selected to impute a target observation. The prediction accuracy usually increases for some number of *k* greater than one (e.g 3 or 5), then gradually declines for larger values of *k* [Muinonen et al., 2001].

Parametric regression

The second strategy used to predict basal areas for small medium and large trees used parametric linear and nonlinear models.

Parametric regression analysis is performed to evaluate the dependence of a response variable on one or several predictors. The function is specified explicitly and can be linear and non linear in the parameters. In this study, a linear regression model with the selected explanatory variables was developed by using ordinary least squares (OLS). There are extensive examples in which OLS has been successfully applied to predict basal area [e.g. Lefsky et al., 1999; Hudak et al., 2006].

A non linear regression model with multiple explanatory variables was also developed by using the non-linear least squares (NLS).

Selection of predictor variables

To avoid the risk of over fitting in case of a high number of predictor variables, and to be able to compare the results obtained with the different statistical techniques for the structure classification and prediction considered in this study, starting from our pool of seventeen candidate predictor variables, the best subsets regression modelling approach was employed to support the choice of the best predictor variables for predicting the total basal area per hectare (small, medium and large trees). We used the *regsubsets* function available in the "leaps" package of R [Lumley, 2004], which selects the best regression subsets through exhaustive search. The *regsubsets* function requires the user to set a maximum number of variables in a subset model (in our case set to 4). The model statistic used to determine best subsets was the Bayesian Information Criterion (BIC) of Schwarz [1978]. The predictor variables selected by this approach were HCV, COVSMALL, COVMEDIUM, COVLARGE.

Performance statistics

Performance statistics were computed both for the classification accuracy of our multiple approaches, and for the prediction accuracy from modelling basal area with CHM derivatives.

We used leave-one-out (LOO) cross-validation to evaluate our ability to predict forest structure classes from CHM data. The results of LOO were categorized in confusion matrices, and evaluated by the overall, producer's, and user's accuracies. The overall model accuracy represents the percentage of plots correctly classified with respect to the total number of sample plots. The producer's accuracy refers to the probability that a certain forest structural type of a stand on the ground is classified as such. The user's accuracy refers to the probability that a stand labelled as a certain forest structural type is really this type.

We looked at the prediction performances of basal area models using the Root Mean Square Error (RMSE). According to Kendall and Buckland [1975], "*in general, the mean square error of a set of values is the arithmetic mean of the squares of their differences from some given value, namely their second moment about that value. When the mean square is regarded as an estimator of certain parental variance components the sum of the squares about the observed mean is usually divided by the number of degrees of freedom, not the number of observation*".

When considering the basal area per hectare of small, medium, large and all trees as the response variables, the RMSE was calculated as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(G_i - \hat{G}_i\right)^2}{n - p}} \quad [3]$$

where G_i is the observed basal area per hectare of sample plot i, $\hat{G}_i$ is the basal area of sample plot i estimated from the predicted distribution, $n - p$ is the degrees of freedom.

To facilitate comparisons between the performances of basal area prediction techniques,

we also report a scaled RMSE obtained by dividing the RMSE by the standard deviation of the training dataset.

For all basal area prediction techniques considered, we also computed the coefficient of determination, which is the proportion of total variation explained by the model.

Results

The results of the analyses within the supervised classification approach are shown with confusion matrices in Table 4 (LDA technique) and Table 5 (RF technique).

Table 4 - Classification result by LDA.

		Truth forest structural type									N° classified plots	User's accuracy
		I2	I3	R11	R12	R21	R23	R31	R32	R33		
Classified forest structural types by LDA	I2	**1**	0	0	1	1	0	0	0	0	3	33%
	I3	0	**0**	0	0	0	0	0	0	0	0	-
	R11	0	0	**6**	3	2	1	0	0	0	12	50%
	R12	0	0	0	**0**	0	0	0	0	0	0	-
	R21	3	1	2	1	**6**	3	0	1	0	17	35%
	R23	2	2	1	1	4	**7**	1	8	1	27	26%
	R31	0	0	0	0	0	0	**0**	0	0	0	-
	R32	0	0	0	1	1	9	0	**16**	1	28	57%
	R33	0	0	0	0	0	2	0	0	**1**	3	33%
N° ground truth plots		6	3	9	7	14	22	1	25	3	**90**	
Producer's accuracy		17%	0%	67%	0%	43%	32%	0%	64%	33%		
Overall accuracy		41%										

Table 5 - Classification result by RF (n_{tree} = 200 and m_{try} = 4).

		Truth forest structural type									No classified plots	User's accuracy
		I2	I3	R11	R12	R21	R23	R31	R32	R33		
Classified forest structural types by RF	I2	**0**	0	1	1	1	1	0	0	0	4	0%
	I3	0	**0**	0	0	0	0	0	1	0	1	0%
	R11	1	0	**4**	2	3	1	0	0	0	11	36%
	R12	1	0	0	**1**	2	0	0	0	0	4	25%
	R21	2	0	4	2	**3**	3	0	1	0	15	20%
	R23	1	2	0	0	4	**8**	1	6	3	25	32%
	R31	0	0	0	0	0	0	**0**	0	0	0	-
	R32	1	1	0	1	1	9	0	**17**	0	30	57%
	R33	0	0	0	0	0	0	0	0	**0**	0	-
No ground truth plots		6	3	9	7	14	22	1	25	3	90	
Producer's accuracy		0%	0%	44%	14%	21%	36%	0%	68%	0%		
Overall accuracy		37%										

The overall accuracy of LDA was 41%. The user's accuracy has not exceeded 57% while the producer's accuracy has not exceeded the 67%. The forest structure types that had

especially good producer's accuracy were the R11 and R32, while the I3, R12 and the R31 reached low levels of producer's accuracy.

In the case of RF classification, a stable overall accuracy (37%) was obtained setting the number of classification trees per response variable equal to 200 and choosing the number of predictor variables to use in each split equal to 4: using lower or greater than 200 trees resulted in overall declining accuracy. Also in this case, the user's and producer's accuracy has not respectively exceeded 57% and 68%. The worst producer's accuracy was in the classification of plots characterized by irregular structure (I2 and I3) and regular structure dominated by large trees (R31 and R33), while the better produce's accuracy was in classifying the R32 forest structure type.

The results of the analyses within the prediction-based classification approach are shown with confusion matrices in Table 6 and Table 7 (k-NN with RF and MSN methods respectively) and Table 8 and Table 9 (multiple linear regression and multiple non linear regression respectively). Between the techniques used in the prediction-based classification approach, the k-NN with the RF method provided better results with respect to the multiple linear regression and non linear regression.

The overall accuracy for k-NN imputation by the RF method was 37% (Tab. 6). The maximum level of user's accuracy was 62% while the maximum level of producer's accuracy was 56% . The irregular structures (I2 and I3) and the regular stands dominated by large trees (R31 and R33) had poor producer's accuracy, while the R11 and R32 forest type had the higher accuracy (56% and 52% respectively).

Table 6 - Classification result by k-NN imputation using RF method.

		Truth forest structural types									No classified plots	User's accuracy
		I2	I3	R11	R12	R21	R23	R31	R32	R33		
Classified forest structural types by k-NN (RF)	I2	**0**	1	1	1	2	0	0	2	0	7	0%
	I3	2	**0**	0	1	1	1	0	1	0	6	0%
	R11	2	0	**5**	1	4	1	0	0	0	13	38%
	R12	0	0	1	**2**	2	0	0	0	0	5	40%
	R21	1	0	2	2	**4**	3	1	0	0	13	31%
	R23	1	2	0	0	1	**9**	0	9	2	24	38%
	R31	0	0	0	0	0	0	**0**	0	0	0	-
	R32	0	0	0	0	0	7	0	**13**	1	21	62%
	R33	0	0	0	0	0	1	0	0	**0**	1	-
No ground truth plots		6	3	9	7	14	22	1	25	3	90	
Producer's accuracy		0%	0%	56%	29%	29%	41%	0%	52%	0%		
Overall accuracy		37%										

The overall accuracy for k-NN imputation by the MSN method was 34%. The maximum value of producer's accuracy was 40%, while the user's accuracy did not pass 46% (R23). The forest structure types that had especially good producer's accuracy were the R23 and R32, while the I3, R31 and the R31 reached low levels of producer's accuracy (Tab. 7).

Table 7 - Classification result by k-NN imputation using MSN method.

		Truth forest structural types									No classified plots	User's accuracy
		I2	I3	R11	R12	R21	R23	R31	R32	R33		
Classified forest structural types by k-NN (MSN)	I2	**1**	0	0	0	2	0	0	2	0	5	20%
	I3	0	**0**	0	1	1	0	0	0	0	2	0%
	R11	0	0	**2**	1	3	1	1	0	0	8	25%
	R12	0	0	1	**2**	1	1	0	0	0	5	40%
	R21	3	2	4	2	**4**	0	0	3	0	18	22%
	R23	0	1	1	1	1	**12**	0	8	2	26	46%
	R31	0	0	1	0	0	0	**0**	0	0	1	0%
	R32	2	0	0	0	2	7	0	**10**	1	22	45%
	R33	0	0	0	0	0	1	0	2	**0**	3	0%
No ground truth plots		6	3	9	7	14	22	1	25	3	90	
Producer's accuracy		17%	0%	22%	29%	29%	55%	0%	40%	0%		
Overall accuracy		34%										

The overall accuracy in the classification by the multiple linear regression was 28%. (Tab. 8) The maximum level of producer's accuracy was 83% while the maximum level of user's accuracy was 54%: this prediction-based classification technique shows poor level of producer's accuracy in the classification of I1, I3, R11,R21, R31 and R33, and good level in the classification of I2.

Table 8 - Classification result by multiple linear regression.

		Truth forest structural types										No classified plots	User's accuracy
		I1	I2	I3	R11	R12	R21	R23	R31	R32	R33		
	I1	**0**	0	0	0	1	4	0	0	0	0	5	0%
Classified forest structural types by OLS	I2	0	**5**	3	2	2	8	11	1	6	0	38	13%
	I3	0	0	**0**	0	0	0	0	0	0	0	0	-
	R11	0	0	0	**0**	0	0	0	0	0	0	0	-
	R12	0	0	0	7	**4**	2	1	0	0	0	14	29%
	R21	0	0	0	0	0	**0**	0	0	0	0	0	-
	R23	0	0	0	0	0	0	**2**	0	5	0	7	29%
	R31	0	0	0	0	0	0	0	**0**	0	0	0	-
	R32	0	1	0	0	0	0	8	0	**14**	3	26	54%
	R33	0	0	0	0	0	0	0	0	0	**0**	0	-
No ground truth plots		**0**	6	3	9	7	14	22	1	25	3	90	
Producer's accuracy		-	83%	0%	0%	57%	0%	9%	0%	56%	0%		
Overall accuracy		28%											

The overall accuracy of the classification based on the results of non linear regression was 26% (Tab. 9). The confusion table shows that the user's accuracy has not exceeded 57% while the producer's accuracy has not passed the 67%. The forest structure types that had especially good producer's accuracy were the R11 and R32, while the I3, R12 and the R31 reached low levels of producer's accuracy.

Table 9 - Classification result by non-linear regression.

		Truth forest structural type										No classified plots	User's accuracy
		I1	I2	I3	R11	R12	R21	R23	R31	R32	R33		
	I1	**0**	0	0	0	2	2	0	0	0	0	4	0%
	I2	0	**4**	2	2	2	9	11	0	6	0	36	11%
	I3	0	0	**0**	0	0	0	0	1	0	0	1	0%
	R11	0	0	0	**0**	0	0	0	0	0	0	0	-
Classified forest structural types by NLS	R12	0	1	1	6	**3**	3	1	0	0	0	15	20%
	R21	0	0	0	0	0	**0**	0	0	0	0	0	-
	R23	0	0	0	0	0	0	**2**	0	5	0	7	29%
	R31	0	0	0	0	0	0	0	**0**	0	0	0	-
	R32	0	1	0	1	0	0	8	0	**14**	3	27	52%
	R33	0	0	0	0	0	0	0	0	0	**0**	0	-
No ground truth plots		0	6	3	9	7	14	22	1	25	3	90	
Producer's accuracy		-	67%	0%	0%	43%	0%	9%	0%	56%	0%		
Overall accuracy		26%											

We can note a common trend across all techniques (used both in supervised classification and prediction-based classification) consisting in high levels of producer's accuracy of the forest structural type R32. Supervised classification techniques are poor in classifying the irregular structure, i.e. multi-stored stands, vice versa the linear and non-linear regression techniques provided the higher levels of producers' accuracy in these types of forest structure.

RF k-NN, MSN k-NN, OLS and NLS had comparable scaled RMSE values in prediction the basal area per hectare for small, medium and large trees and of all trees (Tab. 10). In all cases the error was greater in the prediction of basal area per hectare of medium trees (around 10 m²/ha), and lower in the prediction of basal area per hectare of small trees (around 6.5 m²/ha).

The technique that provided the best performance in basal area prediction was the k-NN with the MSN method: the R^2 to predict small, medium, large and all trees was 0.71, 0.48, 0.69 and 0.56 respectively.

Table 10 - Root Mean Square Error (RMSE), scaled Root Mean Square Error (scaled RMSE) and determination coefficient (R^2) for the prediction basal areas techniques.

	RMSE (m²/ha)				Scaled RMSE				R^2			
	K-NN (RF)	K-NN (MSN)	OLS	NLS	K-NN (RF)	K-NN (MSN)	OLS	NLS	K-NN (RF)	K-NN (MSN)	OLS	NLS
Small trees	7.03	6.29	6.91	6.48	0.70	0.63	0.69	0.65	0.68	0.71	0.54	0.62
Medium trees	11.06	11.09	9.49	9.49	1.06	1.06	0.91	0.91	0.49	0.48	0.20	0.20
Large trees	8.74	8.37	8.56	8.43	0.67	0.64	0.66	0.65	0.68	0.69	0.58	0.58
All trees	11.16	10.65	8.82	8.86	0.93	0.89	0.73	0.74	0.54	0.56	0.48	0.47

Discussion

To date, it often happens that the Italian forest technicians can exploit the CHM as a product of ALS surveys made for other non-forestry purposes at low [Clementel et al., 2012] or even no cost [Corona et al., 2012]. This availability has meant that the technicians, who deal with forest management planning, familiarize themselves with this product. For these reasons, it was considered essential to evaluate if this kind of limited lidar dataset could have application to forest structural prediction, and consequently for mapping forest structural types.

The results of this study suggest that the lidar-derived predictor variables extracted from CHMs may be used just for preliminary analyses within, for example, forest inventory processes that require the recognition of forest structural types with the advantage of reducing time-consuming and expensive fieldworks activities. In fact, the techniques used both in supervised classification approach and prediction-based classification approach did not produced satisfactory levels of classification accuracy. In this context, the non parametric techniques produced more accurate classifications with respect to the parametric techniques, and in particular the linear discriminant analysis provided the best results.

It is difficult to compare our results in terms of other related works, because in other studies lidar metrics were derived from the normalized cloud points, anyway it seems to us important to evaluate our results in relation to those obtained from researches with similar goals where classification approaches were applied.

Zhang et al. [2011] obtained a higher level of overall accuracy (91.4%) performing the linear discriminant analysis to classify five forest types in the Strzelecki Ranges (southeast Victoria, Australia) using eighteen lidar-derived predictor variables from the normalized returns. We are aware that the level of accuracy obtained in our study is not comparable to the Zhang et al. [2011] results, but probably the high number of forest structure types we wanted to classified (eleven vs. five) is a limiting factor of classification accuracy.

Chirici et al. [2013], to classify nine forest fuel types in the Mediterranean province of Palermo and Catania (Italy) previously observed and identified by photo interpretation, applied the RF technique using thirty-one ALS-based metrics calculated from the normalized height returns. The overall accuracy obtained through this technique was 45%, very similar to that obtained in this study.

Hudak et al. [2008], imputing plot-level basal area of eleven conifer species in Moscow Mountain and St. Joe Woodlands in north-central Idaho (USA) with MSN and RF imputation methods, obtained similar scaled RMSD to the present work (if the basal area of all trees is considered).

Considering that the results we obtained are very similar to the mentioned studies, we can say that our inference about the potential to use limited lidar data (i.e. CHM) to predict forest structure is here confirmed: variables extracted from the CHM have the same potentiality of those extracted from lidar point cloud datasets in predicting forest structure types. Furthermore, forests of Trento province cover a wide range of structural conditions which are not typical of this geographic area, are independent of site fertility and of climate conditions, this means that the findings from this study can be extrapolated and applied to other forests located outside the province of Trento.

We have to point out that the sample design of this study could have potentially influenced the statistical scope of inference limiting the goodness of the results: the unbalanced number of cases per forest structural type, due to the fact that in some cases the real forest structure

type didn't match with those selected by the sampling, could have influenced the model classification accuracy. In fact if the within forest type variability (internal variability) is relatively high with respect to the between variability (external variability), then probably the difference among the forest structural types is the result of the internal variability. The availability of more data, acquired by future surveys carried out during the operations of forest plans revision, will allow retesting the techniques considered in this work. In addition, next investigations will consider families of structure that group together the twelve types of the structural classification system: one option could be to take in consideration the mono-stored forest stands (R11, R22, and R33), the bi-stored forest stands (R12, R13, R21, R23, R31, and R32), and the multi-stored forest stands (I1, I2, and I3).

The methodology used to calculate the canopy cover of small, medium and large trees through a diameter-height function working for all trees of the different forest structural types, does not represent a limitation of this study. The uncertainty when assigning a tree to a diameter class based only in its height by the diameter-height function subsists even on plot level, but this method of cover estimation doesn't pose a limitation in our analysis. In fact, if we consider for example the cover of small trees, at a first analysis the cover of this category of trees seems include the cover of medium and large trees, but each cover stratum includes only the pixels which height value falls in the range of tree height modelled by the diameter-height function and in this way we avoid the risk of double or triple count.

Hudak et al. [2008], instead of calling this predictor variable cover, called these variables STRATUMn, they considered seven strata where for example STRATUM4 is the percentage of vegetation returns >10 m and <= 20 m in height. The height interval can be choice arbitrarily or can be define in some objective way, as in our study, anyway the cover estimation does not represent a problem in the analysis we would carry out. Moreover, considering that the CHM takes into account the outer canopy surface, it is more likely that the CHM "blanket" represents the tallest part of the crown. Lidar data coming from ALS campaigns that records more than first and last return per pulse allow to better see what happens inside each stratum, for this reason forest technician should be involved in the process of flight protocol definition.

Conclusion

We can conclude that the canopy structure metrics computed from the CHM and used in the present study can be moderately useful for basal area imputation and hence in the classification process.

At any rate, this study confirmed the fact that coupling lidar data and sample inventory data we can support the forest management from multiple perspectives [Corona, 2010]. The classification methods here investigated may be used to validate the preliminary expeditious classification made by the foresters through a visual estimation during the forest inventory operations.

In spite of LDA provided the best levels of accuracy, we suggest to use this technique with sagacity when an high number of predictor variables is considered. In fact LDA is too flexible in situations with hundreds of highly correlated predictor variables, and it is too rigid in situations where the class boundaries in predictor space are complex and nonlinear [Hastie et al., 1995]. On the contrary, RF algorithm can handle thousands of input variables without variable deletion and runs efficiently on large data sets. It does not over fit, and

there is no need for variable pre-selection [Strobl et al., 2007]. Moreover, RF is a flexible classification algorithm that directly provides measures of variable importance (related to the relevance of each variable in the classification process), that has the ability to model complex interactions among predictor variables and include an algorithm to imputing missing values.

References

Alberti G., Boscutti F., Pirotti F., Bertacco C., De Simon G., Sigura M., Cazorzi F., Bonfanti P. (2013) - *A LiDAR-based approach for a multi-purpose characterization of Alpine forests: an Italian case study.* iForest, 6: 156-168. doi: http://dx.doi.org/10.3832/ifor0876-006.

Ancel P., Wilhelm M.E., Lacombe E. (1999) - *Peuplements forestiers du massif Vosgien; typologie et sylvicultures.* 51 pp. CNPF.

Andersen H.E., Foster J.R., Reutebuch E. (2003) - *Estimating forest structure parameters on Fort Lewis Military Reservation using Airborne Laser Scanner (LIDAR) data.* Proceedings of the Second International Precision Forestry Symposium, Seattle, Washington, June 15-17, 2003, pp. 45-53.

Axelsson P. (1999) - *Processing of laser scanner data - algorithms and applications.* Journal of Photogrammetry and Remote Sensing, 54: 138-147. doi: http://dx.doi.org/10.1016/S0924-2716(99)00008-8.

Barilotti A., Sepic F., Abramo E., Crosilla F. (2007) - *Assessing the 3D structure of the single crowns in mixed alpine forests.* In: Stilla, U., Mayer, H., Rottensteiner, F., Heipke, C., Hinz, S. (Eds.) PIA07. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, XXXVI-3/W49A.

Breiman L. (2001) - *Random Forests.* Machine Learning, 45 (1): 5-32. doi: http://dx.doi.org/10.1023/A:1010933404324.

Chirici G., Scotti R., Montaghi A., Barbati A., Cartisano R., Lopez G, Marchetti M., McRoberts R.E., Olsonn 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.

Clementel F., Colle G., Farruggia C., Floris A., Scrinzi G., Torresan C.(2012) - *Estimating forest timber volume by means of "low-cost" LiDAR data.* Italian Journal of Remote Sensing, 44 (1): 125-140. doi: http://dx.doi.org/10.5721/ItJRS201244110.

Corona P. (2010) - *Integration of forest mapping and inventory to support forest management.* iForest, 3: 59-64. doi: http://dx.doi.org/10.3832/ifor0531-003.

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: http://dx.doi.org/10.5721/EuJRS20124503.

Crookston N.L., Finley A.O. (2008) - *yaImpute: An R Package for kNN Imputation.* Journal of Statistical Software, 23 (10): 1-16.

Curtis R. (1967) - *Height-diameter and height-diameter-age equations for second-growth Douglas-fir.* Forest Science, 13: 365-375.

Cutler R.D., Edwards T.C., Beard K.H., Cutler A., Hess K.H., Gibson J., Lawler J.J. (2007) - *Random forests for classification in ecology.* Ecology, 88 (11): 2783-2792. doi: http://dx.doi.org/10.1890/07-0539.1.

De Natale F., Pignatti G. (2011) - *Estensione e composizione dei boschi italiani.* In: Gasparini, P., Tabacchi, G. (a cura di) 2011 - L'Inventario Nazionale delle Foreste e dei serbatoi forestali di Carbonio INFC 2005. Secondo inventario forestale nazionale italiano. Metodi e risultati. Ministero delle Politiche Agricole, Alimentari e Forestali, Corpo Forestale dello Stato; Consiglio per la Ricerca e Sperimentazione in Agricoltura, Unità di Ricerca per la Pianificazione ed il Monitoraggio e la Pianificazione Forestale. Edagricole, Milano, pp. 99-111.

Dubayah R.O, Knox R.G, Hofton M.A., Blair J.B., Drake J.B. (2000) - *Land surface characterization using lidar remote sensing.* Hill, M., Aspinall, R. (Eds.), Spatial information for land use management, International Publishers Direct, Singapore.

Ferraz A., Bretar F., Jacquemound S., Gonçalves G., Pereira L., Tomé M., Soares P. (2012) - *3-D mapping of multy-layered Mediterranea forest using ALS data.* Remote Sensing of Environment, 121: 210-223. doi: http://dx.doi.org/10.1016/j.rse.2012.01.020.

Hastie T., Buja A., Tibshirani R. (1995) - *Penalized discriminant analysis.* The Annals of Statistics, 23 (1): 73-102. doi: http://dx.doi.org/10.1214/aos/1176324456.

Heurich M., Thoma F. (2008) - *Estimation of forestry stand parameters using laser scanning data in temperate, structurally rich natural European beech (Fagus sylvatica) and Norway spruce (Picea abies) forests.* Forestry, 81 (5): 645-661. doi: http://dx.doi.org/10.1093/forestry/cpn038.

Hudak A.T., Crookston N.L., Evans J.S., Falkowski M.J., Smith A.M.S., Gessler P.E., Morgan P. (2006) - *Regression modeling and mapping of coniferous forest basal area and tree density from discrete-return lidar and multispectral satellite data.* Canadian Journal of Remote Sensing, 32 (2): 126-138. doi: http://dx.doi.org/10.5589/m06-007.

Hudak A.T., Crookston N.L., Evans J.S., Hall D.E., Falkowski M.J. (2008) - *Nearest neighbor imputation of species-level, plot-scale forest structure attributes from LiDAR data.* Remote Sensing of Environment, 112: 2232-2245. doi: http://dx.doi.org/10.1016/j.rse.2007.10.009.

Kendall M.G., Buckland W.R. (1975) - *A Dictionary of Statistics Terms.* Third edition. Longman Group Limited. 166 pp. ISBN 0 582 46304 1.

Lefsky M.A., Harding D., Cohen W.B., Parker G., Shugart H.H. (1999) - *Surface Lidar Remote Sensing of Basal Area and Biomass in Deciduous Forests of Eastern Maryland, USA.* Remote Sensing of Environment, 67:83-98. doi: http://dx.doi.org/10.1016/S0034-4257(98)00071-6.

Lewin-Koh N.J., Bivand R. (2012) - *Maptools: Tools for reading and handling spatial objects.* R package version 0.8-14.

Liaw A., Wiener M. (2002) - *Classification and regression by randomForest.* R News, 2 (3): 18-22, ISSN 1609-3631.

Lumley T. (2004) - *Leaps: regression subset selection.* R package version 2.7. Uses Fortran code by Alan Miller.

Maltamo M., Malinen J., Kangas A., Härkönen S., Pasanen A.M. (2003) - *Most similar neighbor based stand variable estimation for use in inventory by compartments in Finland.* Forestry, 76 (4): 449-464. doi: http://dx.doi.org/10.1093/forestry/76.4.449.

Maltamo M., Packalén P., Yu X., Eerikäinen K., Hyyppä J., Pitkänen J. (2005) - *Identifing and quintifying structural characteristics of heterogeneous borela forests using laser scanner data.* Forest Ecology and Management, 216: 41-50. doi: http://dx.doi.org/10.1016/j.foreco.2005.05.034.

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.

Means J.E., Acker S.A., Harding D.J, Blair J.B., Lefsky M.A., Cohen W.B., Harmon M.E., McKee W.A. (1999) - *Use of large-footprint scanning airborne lidar to estimate forest stand characteristics in the Western Cascades of Oregon.* Remote Sensing of Environment, 67: 298-308. doi: http://dx.doi.org/10.1016/S0034-4257(98)00091-1.

Muinonen E., Maltamo M., Hyppanen H., Vainikainen V. (2001) - *Forest stand characteristics estimation using a most similar neighbor approach and image spatial structure information.* Remote Sensing of Environment, 78: 223-228. doi: http://dx.doi.org/10.1016/S0034-4257(01)00220-6.

Næsset E. (1997) - *Estimating timber volume of forest stands using airborne laser scanner data.* Remote Sensing of Environment, 61: 246-253. doi: http://dx.doi.org/10.1016/S0034-4257(97)00041-2.

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: 88-99. doi: http://dx.doi.org/10.1016/S0034-4257(01)00290-5.

Odasso M. (2002) - *I tipi forestali del Trentino.* Report del Centro di Ecologia Alpina, 25, 192 pp.

Ok A.O., Akar O., Gungor O. (2012) - *Evaluation of random forest method fro agricultural crop classification*. European Journal of Remote Sensing, 45:421-432. doi: http://dx.doi.org/10.5721/EuJRS20124535.

Reimann C., Filzmoser P., Garrett R., Dutter R. (2008) - *Statistical Data Analysis Explained: Applied Environmental Statistics with R.* WILEY, 2008, 362 pp. ISBN 047098581X.

Riaño D., Valladares F., Condésand S., Chuvieco E. (2004) - *Estimation of leaf area index and covered ground from airborne laser scanner (Lidar) in two contrasting forests.* Agricultural and Forest Meteorology, 124: 269-275. doi: http://dx.doi.org/10.1016/j.agrformet.2004.02.005.

Schwarz G. (1978) - *Estimating the dimension of a model.* Annals of Statistics, 6: 461-464. doi: http://dx.doi.org/10.1214/aos/1176344136.

Scrinzi G., Clementel F., Colle G., Farruggia C., Floris A., Torresan C. (2011) - *L'inventario dendromemtrico nella nuova pianificazione forestale aziendale trentina (NPFAT)*. Provincia Autonoma di Trento - Servizio Foreste e Fauna, Consiglio per la Ricerca e la Sperimentazione in Agricoltura, 105 pp. ISBN 978-88-7702-322-3.

Staudhammer C.L., LeMay V.M. (2001) - *Introduction and evaluation of possible indices of stand structural diversity.* Canadian Journal of Forest Research, 31: 1105-1115. doi: http://dx.doi.org/10.1139/cjfr-31-7-1105.

Strobl C., Boulesteix A.L., Augustin T. (2007) - *Unbiased split selection for classification trees based on the Gini index.* Computational Statistics & Data Analysis, 52 (1): 483-501. doi: http://dx.doi.org/10.1016/j.csda.2006.12.030.

Suárez J., Fonweban J., Gardiner B. (2012) - *Supporting precision forestry in Great Britain.* Earthzine (http://www.earthzine.org/2012/06/22/supporting-precision-forestry-in-great-

britain/).

Travaglini D., Bottalico F., Brundu P., Chirici G., Minari E. (2007) - *Testing neighborhood-based indices for monitoring spatial forest stand structure.* In: Gianelle,D., Travaglini D., Mason F., Minari E., Chirici G., Chemini C. (Eds.), 2007. Canopy Analysis and Dynamics of a Floodplain Forest. Rapporti Scientifici, 3. Centro Nazionale per lo studio e la conservazione della Biodiversità Forestale - Bosco della Fontana. Cierre Grafica Editore, Verona, pp. 96.

Venables W.N., Ripley B.D. (2002) - *Modern Applied Statistics with S.* Fourth Edition. Springer, New York. doi: http://dx.doi.org/10.1007/978-0-387-21706-2.

Zhang Z., Liu X., Peterson J., Wright W. (2011) - *Statistical analysis of airborne LiDAR data for forest classification in the Strzelecki Ranges, Victoria, Australia.* In: MODSIM 2011. Sustaining Our Future: Understanding and Living with Uncertainty, 12-16 Dec 2011, Perth, Australia.

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 Sensing of Environment, 87:171-182. doi: http://dx.doi.org/10.1016/S0034-4257(03)00139-1.