ISPRS Journal of Photogrammetry and Remote Sensing 95 (2014) 23–33

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Comparison of airborne laser scanning methods for estimating forest structure indicators based on Lorenz curves

Rubén Valbuena [a,*], Jari Vauhkonen [b,1], Petteri Packalen [c,2], Juho Pitkänen [d,3], Matti Maltamo [c,4]

[a] European Forest Institute HQ, Yliopistokatu 6, 80100 Joensuu, Finland
[b] University of Helsinki, Department of Forest Sciences, PO Box 27, 00014 Helsinki, Finland
[c] University of Eastern Finland, Faculty of Forest Sciences, PO Box 111, Joensuu, Finland
[d] Finnish Forest Research Institute (METLA), Joensuu Research Unit, Joensuu, Finland

ARTICLE INFO

Article history:
Received 30 January 2014
Received in revised form 16 May 2014
Accepted 2 June 2014

Keywords:
Forest structure
Gini coefficient
Lorenz asymmetry
BALM
Tree list estimation
Individual tree detection
Semi ITD
Random forest
MSN

ABSTRACT

The purpose of this study was to compare a number of state-of-the-art methods in airborne laser scanning (ALS) remote sensing with regards to their capacity to describe tree size inequality and other indicators related to forest structure. The indicators chosen were based on the analysis of the Lorenz curve: Gini coefficient (*GC*), Lorenz asymmetry (*LA*), the proportions of basal area (*BALM*) and stem density (*NSLM*) stocked above the mean quadratic diameter. Each method belonged to one of these estimation strategies: (A) estimating indicators directly; (B) estimating the whole Lorenz curve; or (C) estimating a complete tree list. Across these strategies, the most popular statistical methods for area-based approach (ABA) were used: regression, random forest (RF), and nearest neighbour imputation. The latter included distance metrics based on either RF (NN–RF) or most similar neighbour (MSN). In the case of tree list estimation, methods based on individual tree detection (ITD) and semi-ITD, both combined with MSN imputation, were also studied. The most accurate method was direct estimation by best subset regression, which obtained the lowest cross-validated coefficients of variation of their root mean squared error CV(RMSE) for most indicators: *GC* (16.80%), *LA* (8.76%), *BALM* (8.80%) and *NSLM* (14.60%). Similar figures [CV(RMSE) 16.09%, 10.49%, 10.93% and 14.07%, respectively] were obtained by MSN imputation of tree lists by ABA, a method that also showed a number of additional advantages, such as better distributing the residual variance along the predictive range. In light of our results, ITD approaches may be clearly inferior to ABA with regards to describing the structural properties related to tree size inequality in forested areas.

1. Introduction

Indicators describing forest structure can be a valuable support tool in decision-making and forest management planning. Maps of forest structure indicator estimates can be used as a reference to evaluate the success of regeneration groups, plan the location of the next selective cuttings or evaluate the need for thinning (Burger, 2009). Because of its complete three-dimensional characterization of vegetation, airborne laser scanning (ALS) remote sensing allows for evaluating properties related to forest structure in broad forest areas (Lefsky et al., 2005; Maltamo et al., 2005). These properties can be exploited to study forest successional stages (Falkowski et al., 2009; Valbuena et al., 2013a), the risk of wild fire propagation (Andersen et al., 2005; Hall et al., 2005) or wind-throw damage (Suárez et al., 2008), or characteristics related to habitat quality (Lefsky et al., 2002; Martinuzzi et al., 2009).

Lexerød and Eid (2006) pointed out a number of motivations for using indicators derived from Lorenz curves to describe forest structure. In a forest, the Lorenz curve expresses dominance relations by comparing relative cumulated proportions of basal area and stem density accounted for each tree (Valbuena et al., 2012). Gini coefficient (*GC*), Lorenz asymmetry (*LA*), the proportions of basal area (*BALM*) and stem density (*NSLM*) stocked above the quadratic mean diameter, are indicators based on the Lorenz curve

* Corresponding author. Tel.: +358 50 407 3159.
E-mail addresses: ruben.valbuena@efi.int (R. Valbuena), jari.vauhkonen@helsinki.fi (J. Vauhkonen), petteri.packalen@uef.fi (P. Packalen), juho.pitkanen@metla.fi (J. Pitkänen), matti.maltamo@uef.fi (M. Maltamo).
[1] Tel.: +358 40 765 4448.
[2] Tel.: +358 50 522 4299.
[3] Tel.: +358 50 391 3182.
[4] Tel.: +358 50 442 2941.

http://dx.doi.org/10.1016/j.isprsjprs.2014.06.002
0924-2716/

which have been suggested for describing tree size inequality and the balance among forest subpopulations. The *GC* has been demonstrated to more reliably describe tree size distributions than other indicators based on product moments (Knox et al., 1989) or information theory (Valbuena et al., 2012). For this reason, Bollandsås and Næsset (2007) and Duduman (2011) used *GC* as a basis to discriminate among differently-shaped diameter distributions. The *GC* is the ratio between the second and first L-moments, and therefore a second order descriptor of concentration, i.e. relative dispersion (Hosking, 1990). The attention has recently been turned to studying tree size inequality by L-moments, especially with regards to their relations with ALS datasets (Ozdemir and Donoghue, 2013; Valbuena et al., 2013b). Furthermore, the coefficient of *LA* developed by Damgaard and Weiner (2000) was employed by Valbuena et al. (2013a) for characterizing the relation between dominant and subdominant cohorts in multi-layered forests. *LA* is a join description of *BALM* and *NSLM*, two important structural characteristics of forests closely related to one another. In forestry practice, *BALM* and *NSLM* have traditionally applied when using structural stocking guides in decision-making (Gove, 2004).

Previous research aiming at estimating these indicators from ALS were mainly based on parametric modelling, including best subset (Valbuena et al., 2013b) and beta regression (Valbuena et al., 2013a). Most difficulties were found in the variance structure observed on the prediction, and also on the complexity of the relation of *LA* with ALS metrics. These issues may be solved when using non-parametric approaches based on nearest neighbour imputation (*k*-NN). Maltamo et al. (2006) and Hudak et al. (2008) outlined a number of advantages in using non-parametric procedures, which can make them preferable depending on each application. The method of most similar neighbour (MSN) imputation has already become operational for ALS-estimation of forest variables (Maltamo and Packalen, 2014). Based on a canonical correlation analysis, MSN imputes the *k* most similar relations of covariability between response and predictors found within the training dataset (Moeur and Stage, 1995). Furthermore, random forest (RF) is also becoming increasingly popular in ALS remote sensing (Falkowski et al., 2009; McInerney et al., 2010; Yu et al., 2011). RF consists in bootstrapping the training data and computing a regression tree at each iteration, i.e. recursive partitioning by a succession of binary splits of predictor thresholds determined under the criterion of residual sum of squares minimization (Hastie et al., 2009). A combination of RF and *k*-NN is an approach where a distance metric used in *k*-NN is determined based on RF proximity matrix (NN–RF) (Crookston and Finley, 2008). Hudak et al. (2008) found NN–RF to be more robust than other nearest neighbour methods for imputing species-specific basal area and stem densities. We therefore hypothesised that a similar outcome may be obtained for Lorenz curve descriptors, as they simultaneously describe the relations between basal area and stem density (Valbuena et al., 2012).

These methods can be employed with the purpose of obtaining the estimation of a complete tree list, an alternative which may be beneficial when the interest is on knowing the shape of the diameter distribution, for instance in complex multilayered forest structures. Bollandsås and Næsset (2007) used an area-based approach (ABA) with partial least squares regression to estimate discrete quantiles along the diameter distribution, using the *GC* as a basis for stratifying the dataset into homologous diameter distribution types. Alternatively, estimating Weibull model parameters allowed inferring diameter distributions presenting a wide range of simple shapes without prior stratification (Gobakken and Næsset, 2004; Maltamo et al., 2007). Maltamo et al. (2006) introduced the use of MSN imputation in ALS estimation, later including the imputation of discrete quantiles (Packalén and Maltamo, 2008), which tolerated the use of complex diameter distribution shapes without theoretical parameterization. Both diameter and basal area-weighted distributions have been estimated with the intention of improving ALS prediction of forest variables (Gobakken and Næsset, 2004; Maltamo et al., 2007). However, no previous research has been devoted to applying this method on the Lorenz curve, which is a joint description of the intrinsic relation between a diameter distributions and its basal area-weighted (Gove and Patil, 1998; Valbuena et al., 2012).

Individual tree detection (ITD) methods have also traditionally been a source for supplying tree lists. ITD methods are based on segmentation of individual tree crowns from a canopy height model (CHM). The performance of the ITD algorithms typically depends on tree density and spatial distribution of trees, i.e. clustering patterns (Vauhkonen et al., 2012). They usually have the disadvantage of underestimating the understory, although this may not matter for estimating many important forest parameters such as basal area or volume (Persson et al., 2002; Pitkänen et al., 2004). More information on the understory may be obtained with improved tree detection algorithms (Lähivaara et al., 2013), direct point cloud segmentation (Li et al., 2012), using full-waveform ALS information (Reitberger et al., 2009), or analyzing combined leaf-on and leaf-off acquisitions in deciduous forests (Hill and Broughton, 2009). Maltamo et al. (2004), Lindberg et al. (2010), Vauhkonen et al. (2010) and Vastaranta et al. (2012) combined ABA and ITD with the purpose of overcoming this difficulty and improving estimation accuracy. Moreover, Breidenbach et al. (2010) introduced the idea of semi-ITD, in which all trees measured inside a given segment are considered to be represented by that segment, and not just the dominating tree. All these methods have been commonly evaluated by means of the improvement obtained in total forest estimates.

In this article we compare these state-of-the-art ALS estimation methods, with the objective of evaluating them with regards to their capacity for assessing characteristics of forests related to tree size inequality, and the balance between overstory and understory layers. Indicators derived from the study of the Lorenz curve were selected for this purpose, further clarifying their relations with ALS metrics. We compared the results obtained with three different estimation strategies consisting of: (A) direct indicator estimation; (B) non-parametric estimation of the Lorenz curve, and posterior indicator derivation; or (C) estimating a complete tree list, from which the Lorenz curve and derived indicators were later derived (Fig. 1). Many different methods were tested for each strategy, with the purpose of selecting the most appropriate methodological combination for this type of forest structure-related response.

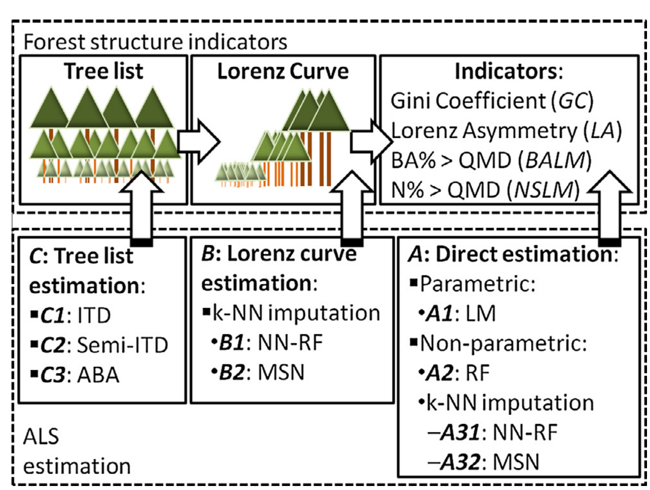

Fig. 1. Flowchart describing the procedure for deriving indicators of forest structure based on Lorenz ordering (above), and how the alternatives considered for their estimation by ALS remote sensing (below) relate to each step of this process. BA: basal area; N: stem density; QMD: quadratic mean diameter; ABA: area-based approach; ITD: individual tree detection; NN: nearest neighbour; RF: random forest; MSN: most similar neighbour; LM: linear model.