Remote Sensing of Environment 194 (2017) 437–446

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Key structural features of Boreal forests may be detected directly using L-moments from airborne lidar data

Rubén Valbuena [a,*], Matti Maltamo [a], Lauri Mehtätalo [b], Petteri Packalen [a]

[a] University of Eastern Finland, School of Forest Sciences, PO Box 111, Joensuu, Finland
[b] University of Eastern Finland, School of Computing, PO Box 111, Joensuu, Finland

ARTICLE INFO

Article history:
Received 17 February 2016
Received in revised form 12 September 2016
Accepted 11 October 2016
Available online 20 October 2016

Keywords:
Airborne laser scanning
L-moments
Gini coefficient
L-coefficient of variation
Forest structure
Tree size inequality
Shade-tolerance

ABSTRACT

This article introduces a novel methodology for automated classification of forest areas from airborne laser scanning (ALS) datasets based on two direct and simple rules: L-coefficient of variation $Lcv = 0.5$ and L-skewness $Lskew = 0$, thresholds based on descriptors of the mathematical properties of ALS height distributions. We observed that, while $Lcv > 0.5$ may represent forests with large tree size inequality, $Lskew > 0$ can be an indicator for areas lacking a closed dominant canopy. $Lcv = 0.5$ discriminated forests with trees of approximately equal sizes (even tree size classes) from those with large tree size inequality (uneven tree size classes) with kappa $\kappa = 0.48$ and overall accuracy OA = 92.4%, while $Lskew = 0$ segregated oligophotic and euphotic zones with $\kappa = 0.56$ and OA = 84.6%. We showed that a supervised classification could only marginally improve some of these accuracy results. The rule-based approach presents a simple method for detecting structural properties key to tree competition and potential for natural regeneration. The study was carried out with low-density datasets from the national program on ALS surveying of Finland, which shows potential for replication with the ALS datasets typically acquired at nation-wide scales. Since the presented method was based on deductive mathematical rules for describing distributions, it stands out from inductive supervised and unsupervised classification methods which are more commonly used in remote sensing. Therefore, it presents an opportunity for deducing physical relations which could partly eliminate the need for supporting ALS applications with field plot data for training and modelling, at least in Boreal forest ecosystems.

1. Introduction

Airborne laser scanning (ALS) can be a valuable tool for studying structural properties of forests (Lefsky et al., 1999a; Drake et al., 2002; Frazer et al., 2005; Maltamo et al., 2005; Valbuena et al., 2016a). The relationships of ALS to forest structure can be employed to analyse asymmetric competition among trees (Kellner and Asner, 2009), and hence forest growth conditions (Stark et al., 2012). In fully-stocked forests (Gove, 2004) light resource pre-emption drives asymmetric competition processes, leading to mortality of the least competitive trees (Weiner, 1990). These are forests with closed canopies and structural properties yielding shady areas, i.e. oligophotic zones (sensu Lefsky et al., 2002), under the dominant tree crowns. In turn, detecting forest areas with light resource availability, which are characterized by large euphotic zones (sensu Lefsky et al., 2002), can be key to monitoring forest disturbance and regeneration. Several metrics derived from ALS height distributions have potential for describing these key characteristics related to forest structure (Zimble et al., 2003). For this reason, studies on ALS-based forest structure characterization by statistical inductive methods, which relate ALS metrics to field attributes empirically, are commonplace (Hall et al., 2005; Lefsky et al., 2005; Dalponte et al., 2008; Pascual et al., 2008; Disney et al., 2010; Jaskierniak et al., 2011; Ozdemir and Donoghue, 2013; Valbuena et al., 2014).

Size hierarchy among trees growing in the vicinity influences competition processes in the forest community (Weiner, 1990; Valbuena et al., 2012). Knox et al. (1989) suggested the Gini coefficient (*GC*) (Gini, 1921) as a consistent descriptor of tree size inequality, and hence a reliable indicator of competition conditions in the forest (Cordonnier and Kunstler, 2015). For this reason, in the context of ALS estimation, the *GC* of tree sizes has been used as a basis for stratifying the forest area into homogeneous structural types (Bollandsås and Næsset, 2007; Valbuena et al., 2013a). Furthermore, Knox et al. (1989) also suggested the inclusion of skewness as a complement to the *GC* in describing forest structural properties. For this reason, Valbuena et al. (2013a) included asymmetry in their analysis of forest structural properties, to study relations of relative dominance between different strata in the forest vertical profile.

While Bollandsås and Næsset (2007) employed stand register data from previous inventories for carrying out their stratification, it would be advantageous if the same remote sensing material could be used

* Corresponding author.
E-mail addresses: rubenval@uef.fi (R. Valbuena), matti.maltamo@uef.fi (M. Maltamo), lauri.mehtatalo@uef.fi (L. Mehtätalo), petteri.packalen@uef.fi (P. Packalen).

http://dx.doi.org/10.1016/j.rse.2016.10.024
0034-4257/© 2016 Published by Elsevier Inc.

for wall-to-wall predictions of forest structure indicators and classifications into forest structural types (Lefsky et al., 1999b; Drake et al., 2002). In particular, Ozdemir and Donoghue (2013) and Valbuena et al. (2013b, 2016a) obtained predictions of the *GC* of tree size inequality with reliable accuracy. As previous research has concentrated on the forest response (Lefsky et al., 1999a; Valbuena et al., 2013a), and on its analysis and estimation by a wide range of different statistical methods – such as analysis of variance (Zimble et al., 2003), canonical correlation (Lefsky et al., 2005), parametric (Hall et al., 2005) and non-parametric (Valbuena et al., 2014) modelling, histogram thresholding (Maltamo et al., 2005), or finite mixtures (Jaskierniak et al., 2011) –, the next question to answer would be: do the ALS metrics have, by themselves, capacity to discriminate among forest structural types, making no use of statistical methods linking field data to ALS metrics?.

Moments are quantitative measurements of probability density distributions employed to summarize their properties. The most conventional are the product moments, expected values of the powers of a random variable which lead to the use of mean, variance and skewness as measures for location, scale and shape. These descriptors of ALS return height distributions are metrics commonly employed as auxiliary variables in forest assessment (e.g., Næsset, 2002; White et al., 2013; Asner and Mascaro, 2014). Alternatively, Frazer et al. (2011) and Ozdemir and Donoghue (2013) recently drew the attention towards the L-moments, a set of statistics known by their sample efficiency (i.e., reliability at small sample sizes) and robustness to outliers, compared to conventional moments (Hosking, 1990). Consider a sample order statistic $X_{k:r}$ – the *k*th smallest observation in a sample of size *r* –, which is a many-to-one transformation of a random sample of size *r*, and therefore a random variable. The L-moments are based on its expected values $E(X_{k:r})$ (Appendix A). Moreover, L-moment ratios have the advantage of being bounded by finite intervals (Hosking, 1989), making them comparable among ALS distributions differing in their mean height. The L-coefficient of variation (*Lcv*) and the L-skewness (*Lskew*) are two types of L-moment ratios (Appendix A.2). *Lcv* is the ratio of the second (*L*2) to the first (*L*1) L-moments:

$$Lcv = \frac{L2}{L1} = \frac{E(X_{2:2}) - E(X_{1:2})}{2E(X)}, \tag{1}$$

where $E(X)$ is the expected value of *X*. In the case of ALS metrics, the variable *X* is the height of ALS returns. The *Lcv* is mathematically equivalent to the *GC* (Appendix A.3), and therefore the same properties apply to both of them. For instance, they are scale-invariant, and for positive random variables their values are bounded within the [0,1] interval (Hosking, 1989). Also, Valbuena et al. (2012) showed that an asymptote at $GC = 0.5$ represents the case of maximum entropy among tree sizes in the forest. On the other hand, *Lskew* is the ratio of the third (*L*3) to the second (*L*2) L-moments:

$$Lskew = \frac{L3}{L2} = \frac{E(X_{3:3}) - 2E(X_{2:3}) + E(X_{1:3})}{E(X_{3:3}) - E(X_{1:3})}. \tag{2}$$

In the case of *Lskew*, its theoretical bounds are [−1,1] (Hosking, 1989). The value of $Lskew = 0$ corresponds to a symmetric distribution, while positive or negative values denote the type of asymmetry for the distribution of ALS heights. This article employs these mathematical properties of L-moments for describing ALS height distributions, in contrast to inductively researching explanatory potential in relation to field data attributes.

The aim of this research was to develop simple methods for explaining key features related to forest structure from a few L-moment ratios of ALS returns. *Lcv* and *Lskew* were used for detecting tree size inequality and light availability, and they were utilized for an automated classification of forests from ALS datasets, which was applied directly without the use of field data. The idea builds upon the hypothesis that two deductive mathematical rules, $Lcv = 0.5$ and $Lskew = 0$, may be used to classify the forest area into two groups, based solely on the ALS height distributions. We studied whether such classifications would be sound in terms of explaining properties of size inequality among trees growing in vicinity (even or uneven tree sizes) and competitive conditions for light in the forest community (oligophotic or euphotic). We compared the reliability of the rule-based method to results obtained from a supervised classification. This article discusses suitable applications for this rule-based method.

2. Materials

2.1. Study area and ALS data

The research was conducted in a 252,000 ha study area including approximately 200,000 ha of the Boreal forest ecosystems typically found in the region of North Karelia (Finland), which consists of forests dominated by Scots pine (*Pinus sylvestris* L.) Norway spruce (*Picea abies* (L.) Karst.) or Birch species (*Betula* ssp.) with various degrees of admixtures also with other deciduous trees (such as *Alnus* ssp., *Populus* ssp. etc.). The ALS data were acquired by Blom Kartta Oy (Finland) during May 2012 with an ALS60 system from Leica Geosystems (Switzerland). A flying height of 2300 m above ground rendered an average density of 0.91 pulses per squared-meter. Country-wide laser data are being consistently acquired using broadly similar parameters (National Land Survey of Finland; NLS, 2013). Methods may therefore by consistently replicated throughout the country, bringing potential for upscaling the results obtained at national-level.

Heights above ground for individual ALS returns were calculated by subtracting the digital terrain model provided by the NLS. We considered that, as seedlings and saplings were included in field mensuration (Valbuena et al., 2016b), their influence in laser pulse interception had to be accounted for in ALS metric computation. Consequently, just a very small height threshold of 0.1 m was used, only with the intention to mask out the influence of the ground. Sample estimates of L-moments and their ratios (Wang, 1996) were computed from the heights of all the ALS returns located within each cell over a regular grid covering the entire study area. The spatial resolution of this grid was 16 m × 16 m, a customary practice in Finland that makes cell size roughly coincident in with the area of field plots operationally established and measured by Finnish Forest Centre (SMK, Suomen Metsäkeskus).

2.2. Field dataset used for validation

Field data for validation of the methods were partly acquired by University of Eastern Finland (UEF), and partly provided by SMK. Data from a total of $N = 244$ plots were acquired in a stratified random sampling fashion with approximately equal per-stratum sample sizes (Valbuena et al., 2016b). The strata employed were the forest development classes commonly used in operational management in Finland (per-stratum sample sizes were $n = 31$, unless specified): *Seedling*, *Sapling*, *Young*, *Advanced*, *Mature*, *Shelterwood*, *Seed-tree* ($n = 29$), and *Multi-storied* ($n = 29$). SMK's stand register data based on previous inventories was employed for the initial randomization of field plot locations. Valbuena et al. (2016b) provides details about acquisition protocol and processing of field data. Appendix B details the criteria used to assign a development class to each field plot, a task carried out independently by experienced SMK personnel.

3. Methods

3.1. The rule-based method for stratifying forests based on ALS data

We used a deductive approach to thresholding using the L-moment ratios. The rules were deduced from their mathematical properties, as