Remote Sensing of Environment 174 (2016) 1–9

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Comparing echo-based and canopy height model-based metrics for enhancing estimation of forest aboveground biomass in a model-assisted framework

Gherardo Chirici [a], Ronald E. McRoberts [b], Lorenzo Fattorini [c], Matteo Mura [d,*], Marco Marchetti [d]

[a] Department of Agricultural, Food and Forestry Systems, Università degli Studi di Firenze, Italy
[b] Northern Research Station, U.S. Forest Service, Saint Paul, MN, USA
[c] Department of Economics and Statistics, Università di Siena, Italy
[d] Department of Biosciences and Territory, University of Molise, Pesche (IS), Italy

ARTICLE INFO

Article history:
Received 23 May 2015
Received in revised form 26 September 2015
Accepted 12 November 2015
Available online xxxx

Keywords:
Airborne laser scanning metrics
Forest biomass
Model-assisted estimator
k-Nearest Neighbors
Tessellation stratified sampling

ABSTRACT

Among the forestry-related applications for which airborne laser scanning (ALS) data have been shown to be beneficial, forest inventory has been investigated as much if not more than other applications. Metrics extracted from ALS data for spatial units such as plots and grid cells are typically of two forms: echo-based metrics derived directly from the three-dimensional distribution of the point cloud data and metrics derived from a canopy height model (CHM). For both cases, a large number of metrics can be calculated and used to construct parametric and non-parametric models to predict forest variables.

We compared model-assisted estimates of total forest aboveground biomass (AGB) obtained using echo-based and CHM-based height metrics with two prediction methods: (i) a parametric linear model, and (ii) the non-parametric k-Nearest Neighbors (k-NN) technique. Model-assisted (MA) estimators were used with sample data obtained using a two-phase, tessellation stratified sampling (TSS) framework to estimate population parameters. The study was conducted in Molise Region in central Italy.

For the four combinations of metrics and prediction techniques, estimates of total biomass were similar, in the range 1.96–2.1 million t, with standard error estimates that were also similar, in the range 0.20–0.21 t. Thus, the CHM-based metrics produced AGB estimates that were similar to and as accurate as those for the echo-based metrics, regardless of whether the parametric or the non-parametric prediction method was used. Additionally, the proposed MA estimator was more accurate than the estimator that did not use auxiliary data.

1. Introduction

The utility of airborne laser scanner (ALS) data for contributing to and enhancing forestry-related applications is nowadays both indisputable and well-documented (Corona et al., 2012; Lim, Treitz, Wulder, St-Ongé, & Flood, 2003a; Maltamo, Næsset, & Vauhkonen, 2014; Montaghi et al., 2013; Wulder, Bater, Coops, Hilker, & White, 2008). Further, when estimation of aboveground biomass (AGB) is the main goal, the predictive power of ALS data as auxiliary information has been shown to be more effective with respect to increasing the accuracy of estimates than many other sources of remotely sensed data (Zolkos, Goetz, & Dubayah, 2013). This capability can be attributed to strong correlations between forest AGB and forest height variables derived from ALS metrics (Lefsky et al., 2014). These correlations, in turn, are due to the nature of the ALS data itself, i.e., a cloud of geo-referenced 3-dimensional points characterized as returns or echoes. Multiple modeling techniques are used to predict AGB using ALS metrics of which the most common are parametric regression (Næsset & Gobakken, 2008; McRoberts, Næsset, & Gobakken, 2013) and the non-parametric k-Nearest Neighbors (k-NN) technique (McRoberts, Næsset, & Gobakken, 2015).

Independently of the prediction criterion, two main approaches are used to extract metrics from ALS data for use as covariates when constructing AGB prediction models: (i) echo-based metrics, and (ii) canopy height model (CHM)-based metrics. Echo-based metrics are descriptive statistics directly extracted from the ALS point cloud. This approach requires the availability of raw ALS data, often provided in the form of geo-referenced vector points in 3-dimensional space. If the ALS data are not acquired specifically for forestry applications, raw ALS data may not be available in which case echo-based metrics cannot be calculated (Montaghi et al., 2013). In this case, interpolations of the raw ALS echo heights are often available in the form of two gridded raster layers: ground height for each pixel characterized as a digital

* Corresponding author at: Department of Biosciences and Territory, University of Molise, Contrada Fonte Lappone snc, 86090 Pesche (IS), Italy.
E-mail address: mur.teo@gmail.com (M. Mura).

http://dx.doi.org/10.1016/j.rse.2015.11.010

elevation model (DEM) and absolute height of objects above ground characterized as a digital surface model (DSM). For forestry applications, a CHM consisting of top canopy height for each pixel is constructed from the difference between the DSM and DEM (Kraus & Pfeifer, 1998).

To construct parametric or non-parametric prediction models, AGB observations are required and are typically obtained from field plots. Corresponding plot-level ALS metrics extracted from either the echo heights or from the CHM are used as predictors for estimating AGB. Only a limited number of metrics such as minimum, maximum, average and standard deviation values of height are typically extracted from a CHM grid (Barbati, Chirici, Corona, Montaghi, & Travaglini, 2009). Many more metrics can be extracted from distributions of echo heights including the minimum, mean, maximum, standard deviation, skewness, kurtosis, and coefficient of variation; the 0, 10, …, 90 distribution percentiles; canopy density metrics calculated as the proportions of first echo heights above the 0, 10, …, 90 quantiles of the first echo height distributions (Næsset, 2002; Lim, Treitz, Baldwin, Morrison, & Green, 2003b; Næsset, 2004a, 2004b; Næsset & Gobakken, 2005; Næsset & Gobakken, 2008; Hawbaker et al., 2010; Gobakken et al., 2012); and canopy relief ratio (Evans, Hudak, Faux, & Smith, 2009).

Although large numbers of both echo-based and CHM-based metrics can be derived from ALS data, modeling applications often require only a few of them. Næsset (2002) found that only 2–4 metrics from a set of 46 metrics were required to obtain $0.80 \le R^2 \le 0.93$ for volume models for both young and mature forests on both good and poor sites. The selected metrics were mean and maximum echo heights, several canopy height percentiles and metrics related to canopy density; similar metrics were selected by Næsset (2004a, 2004b). Næsset and Gobakken (2008) found that canopy height percentiles produced the greatest accuracies for above- and below-ground biomass models in boreal forests. Two ALS metrics, the 90th percentile of laser canopy height and canopy density, plus variables representing different areas, age classes, and tree species composition, produced $0.85 \le R^2 \le 0.88$. Gobakken et al. (2012) reported that mean height, percentiles of heights from first and/or last pulses and canopy density metrics, coupled with altitude information, produced $0.80 \le R^2 \le 0.94$ for volume models for young/mature and productive/non-productive forests.

Corona et al. (2008) obtained $R^2 = 0.78$ when predicting plot-level forest volume using the sum of the CHM heights raised to a power for a temperate broadleaved forest; Barbati et al. (2009) used the same methodology for a coastal Mediterranean pine forest and obtained $R^2 = 0.88$. Fusco et al. (2008) obtained $R^2 = 0.76$ for circular plots in a broadleaved forest. Finally, for Norway spruce in an alpine environment, Floris, Clementel, Farruggia, and Scrinzi (2010) constructed regression models for predicting standing volume in sample plots using CHM metrics and found that CHM mean height, excluding pixels with height less than 2 m, produced $R^2 = 0.94$.

In recent years, an interesting debate has emerged regarding the advantages and disadvantages of the two kinds of ALS-based metrics. Although both kinds of metrics are commonly used, no peer review reports of direct comparisons for predicting volume in the same study area are known, other than possibly Gaulton and Malthus (2010) who compared them for detecting canopy gaps.

Model-assisted (MA) estimation exploits auxiliary information to augment ground data for purposes of enhancing estimation (Särndal, Swensson, & Wretman, 1992). Although many sources of auxiliary information can be used (e.g., Corona, Fattorini, & Franceschi, 2009), remotely-sensed data have been found to be a particularly useful for forestry applications.

A broad range of sampling designs have been used with MA estimators for exploiting ALS auxiliary information. Corona and Fattorini (2008) proposed MA estimation of forest standing volume using CHM height as auxiliary information when field plots are randomly and independently located. Ene et al. (2012) and McRoberts et al. (2013) applied a MA estimator using a systematic sampling design. Gregoire et al. (2011) and Gobakken et al. (2012) developed a MA regression estimator of AGB for a two-stage sampling design, and Næsset et al. (2013) tested a MA estimator for both two-phase and two-stage sampling designs. Finally, Saarela et al. (2015) described the use of MA estimators with a systematic cluster sampling design.

In designing schemes for sampling large areas, such as for forest inventory, limited financial resources for ground sampling suggest that the smallest but most representative sample should be selected. Uniform random sampling (URS), the random and independent selection of points on a continuous surface, is the simplest scheme to locate ground plots (Fattorini, 2015). However, like its finite population analogue, simple random sampling with replacement (SRSWR), URS may lead to uneven coverage of the study area which, in turn, makes the sample less representative of the study area. To avoid these drawbacks, spatially stratified schemes can be adopted (Fattorini, 2015). With tessellation stratified sampling (TSS), the study area $\mathcal{A}$ is covered by a region $\mathcal{R} \supset \mathcal{A}$ of size R consisting of N non-overlapping regular polygons $\mathcal{R}_1, \ldots, \mathcal{R}_N$ of equal size such that $\mathcal{R}_i \cap \mathcal{A} \neq \emptyset$ for all $i = 1, \ldots, N$. Then, for each polygon $\mathcal{R}_i$, a point is randomly selected within the polygon (Fattorini, 2015). In a two-phase application of a TSS for forest inventory, the objective of the first phase is to uniformly spread points throughout the study area and classify them with respect to forest/non-forest. In the second-phase, a sub-sample of points within the forest class is selected using a finite-population sampling scheme and then visited by field crews. With this sampling design, estimators that both do not and do use auxiliary information extracted from ALS data are possible, although the latter have not been investigated. For this study, a MA estimator using ALS-based metrics in a two-phase TSS was investigated.

The objectives of the study were threefold: (i) to compare large area estimates of total AGB obtained using echo-based and CHM-based metrics as predictor variables, (ii) to compare estimates obtained using parametric linear regression and the non-parametric k-Nearest Neighbors (k-NN) technique, and (iii) to evaluate the utility of MA estimators in a two-phase TSS framework.

We adopted the k-NN approach because it has emerged as very popular for spatial estimation of NFI variables using remotely sensed auxiliary data (Chirici et al., in press). Despite the large number of other reported methods, including machine learning approaches, k-NN is a popular approach that has been adopted by NFI programs for large area operational applications: in Finland (Tomppo, 1990), in Sweden (SLU Forest Map, 2013), in Canada (Beaudoin et al., 2014) and in the USA (Wilson, Lister, & Riemann, 2013).

2. Materials and methods

2.1. Study area

The study area is located in the southwestern part of Molise Region in central Italy and includes 36,360 ha (Fig. 1). Based on a local forest map (Chirici et al., 2011), forests cover more than 20,518 ha and comprise approximately 56% of the study area, where the COST (European Cooperation in Science and Technology) Action E43 definition of forest was used (Vidal, Lanz, Tomppo, Schadauer, & Gschwantner, 2008). The forested area is dominated mainly by deciduous oaks (*Quercus cerris*, *Quercus pubescens*) covering approximately 60% of the forest area, hop hornbeam (*Ostrya carpinifolia*) covering approximately 18% and beech (*Fagus sylvatica*) covering approximately 9%. The oak and hop hornbeam forests are mainly privately-owned and are managed in a coppice with standards system. Rotation ages are usually between 18 and 25 years, with most clear-cuts 1–2 ha wide and 100–200 standards/ha. Conversely, most of the beech forests are managed with the shelterwood system or are unmanaged.