ISPRS Journal of Photogrammetry and Remote Sensing 134 (2017) 86–95

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Large tree diameter distribution modelling using sparse airborne laser scanning data in a subtropical forest in Nepal

Parvez Rana [a,b,*], Jari Vauhkonen [c], Virpi Junttila [d], Zhengyang Hou [e], Basanta Gautam [f], Fiona Cawkwell [b], Timo Tokola [a]

[a] School of Forest Sciences, University of Eastern Finland, P. O. Box - 111, FI-80101 Joensuu, Finland
[b] Department of Geography, University College Cork (UCC), Ireland
[c] Natural Resources Institute Finland (Luke), Economics and Society, Yliopistokatu 6, P.O. Box 68, FI-80101 Joensuu, Finland
[d] School of Engineering Science, Lappeenranta University of Technology, Lappeenranta, Finland
[e] Department of Natural Resources & Environmental Science, University of Nevada, Reno, NV 89667, United States
[f] Arbonaut Ltd. Malminkaari 15a (Malmin Nova, 4th floor), 00700 Helsinki, Finland

ARTICLE INFO

Article history:
Received 18 May 2017
Received in revised form 27 October 2017
Accepted 31 October 2017
Available online 6 November 2017

Keywords:
Diameter distribution
K-MSN
Large tree
Nepal
Tropical forests
LiDAR

ABSTRACT

Large-diameter trees (taking DBH > 30 cm to define large trees) dominate the dynamics, function and structure of a forest ecosystem. The aim here was to employ sparse airborne laser scanning (ALS) data with a mean point density of 0.8 m^{-2} and the non-parametric k-most similar neighbour (*k*-MSN) to predict tree diameter at breast height (DBH) distributions in a subtropical forest in southern Nepal. The specific objectives were: (1) to evaluate the accuracy of the large-tree fraction of the diameter distribution; and (2) to assess the effect of the number of training areas (sample size, n) on the accuracy of the predicted tree diameter distribution. Comparison of the predicted distributions with empirical ones indicated that the large tree diameter distribution can be derived in a mixed species forest with a $RMSE_{\%}$ of 66% and a $bias_{\%}$ of −1.33%. It was also feasible to downsize the sample size without losing the interpretability capacity of the model. For large-diameter trees, even a reduction of half of the training plots (n = 250), giving a marginal increase in the $RMSE_{\%}$ (1.12–1.97%) was reported compared with the original training plots (n = 500). To be consistent with these outcomes, the sample areas should capture the entire range of spatial and feature variability in order to reduce the occurrence of error.

1. Introduction

Large-diameter trees are of ecological, economic and social importance. In ecological terms, they have a clear influence on the function, dynamics and structure of tropical, subtropical, temperate and boreal forest ecosystems (Lutz et al., 2012). They are the primary storehouses of forest wood volume, biomass and carbon stocks, contain essential hollows for animals and birds to nest in, provide the most abundant sources of nectar and seeds, and are of the highest aesthetic appeal (Clark and Clark, 1996; Pugh, 2014).

The distribution of trees in forest stands in terms of their diameter at breast height (DBH) provides useful information for harvesting forest stands and for assessing the economic value, growth and yield of the size classes present (Bollandsås et al., 2008; de Lima et al., 2015). In addition, the tree diameter distribution is correlated with species diversity (Fries et al., 1997; Bollandsås et al., 2013). It can also provide useful information on past disturbance events and the structure and successional status of a forest. Since it is inefficient to measure the diameter of all the trees, the diameter distribution is normally assessed through the stem frequency distribution (Gove and Patil, 1998; Maltamo and Gobakken, 2014).

Airborne laser scanning (ALS) has recently comprised a revolution in technological advancements with an enormous possibility for increasing the accuracy of large-scale forest inventories and reducing their costs (Næsset et al., 2016). Remote sensing-based forest mapping is normally carried out utilizing the area-based method (Næsset, 2002), in which empirical relationships between independent predictors (e.g. ALS metrics) and dependent variables (e.g. field-observed stem volume) are employed to estimate target observations for the entire study area (Gobakken and Næsset, 2008). ALS data have been used for predicting tree diameter

* Corresponding author at: Department of Geography, University College Cork (UCC), Ireland.
E-mail address: parvez_200207@yahoo.com (P. Rana).

https://doi.org/10.1016/j.isprsjprs.2017.10.018
0924-2716/

distributions in both managed and unmanaged forests in the boreal and temperate forest zone (e.g. by Gobakken and Næsset, 2004; Maltamo et al., 2006; Bollandsås and Næsset, 2007; Bollandsås et al., 2008, 2013; Thomas et al., 2008; Peuhkurinen et al., 2011; de Lima et al., 2015; Saad et al., 2015).

In recent years, ALS-based forest inventories in tropical zones has become popular and established for estimation of different target observations (see Drake et al., 2002; Hou et al., 2011; Asner et al., 2012; Rana et al., 2014; Næsset et al., 2016). The accuracy of ALS-based predictions is expected to be lower in tropical forests compared to other forest zones due to the heterogeneous species variation found in diverse vertical structures. Although approaches based on detecting individual trees and predicting their diameters could basically be applied in tropical forests as well (Ferraz et al., 2016), this is an unlikely option for operational inventories due to the high probability of omission and commission errors in tree detection and the costs of acquiring dense ALS data. A more likely operational choice would be to predict the area-based diameter distribution, which is feasible using sparse ALS data (Maltamo and Gobakken, 2014). No studies that assess the accuracy of ALS-based diameter distribution prediction in subtropical forests yet exist, and the expected accuracy cannot be derived from studies in other forest types (e.g. boreal forests) because of the differing characteristics of these forests and their environmental conditions.

The ALS-based prediction of tree diameter distribution is reliant on different aspects, such as (1) the choice of a suitable distribution function i.e. parametric (e.g. Weibull, beta, gamma, log-normal) and non-parametric (e.g. percentile prediction, *k*-nearest neighbour) (see Gobakken and Næsset, 2005); (2) the choice of statistical method such as (see de Lima et al., 2015). Bollandsås et al. (2013) concluded that the even though both parametric and non-parametric methods overall yielded accurate results, the k-MSN approach was more reliable in terms of predicting the number of large trees. (3) the choice of the dependent variable and independent predictor (Maltamo et al., 2009); (4) a multi-modal or irregular distribution of diameters due to unmanaged, uneven-aged forest (see Bollandsås et al., 2013); and (5) the total number of sample plots (Maltamo et al., 2009). Hyyppä et al. (2012) employed the dense ALS points (13 m^{-2}) for DBH estimation and concluded that the combined use of the plot-level height metrics and individual tree-based features reduced the estimation error of 0–3%. Koenig and Höfle (2016) provided a state-of-the-art review of frequently used full-waveform ALS-based point cloud and waveform features for tree species classification and compares the applied features and their characteristics for specific tree species detection. Maltamo et al. (2009) mentioned that the *k*-NN methods rely on large sample size which is a foremost interest as this is connected to the cost of the inventory. However, the exact number of training plots required cannot be assumed, since this is associated with the variability in the data and the variable of interest to be estimated. Hypothetically, the training areas must capture the variability in the population in addition to concurrently exploiting inventory budget efficiency. The target observation and remote sensing features must be interrelated strongly to render considerable competence in regards to the model's predictive capacity, agreed a reasonable sample size (Junttila et al., 2013). Model competence might be determined on the basis of sample size needed and their desired information content. Thus, the selection of training model is connected to the analytical capacity in terms of accuracy and precision and with consideration of cost.

Nepal harbours a wide variety of climatic zones, from tropical to arctic, which results in a substantial diversity of forest structures, architectures and compositions (Jha and Paudel, 2010). Forests occupy a total of 5.96 million ha in Nepal, i.e. 40.36% of the total land area (DFRS, 2015). According to the latest national forest inventory, the stem density for DBH > 10 cm was 2563.27 million (430 stems/ha) with a volume of 165 m^3/ha, and the large tree stem density (DBH > 30 cm) was 64 stems/ha, with a volume of 116 m^3/ha (DFRS, 2015). Globally, large-diameter old trees are rare and their stem densities are relatively small, i.e. 12–20 stems/ha (Salas et al., 2006). Since large trees dominate the dynamics, function and structure of a forest ecosystem, different management approaches are required to sustain a function (e.g. carbon storage) when it is executed mainly by a couple of large trees compared to numerous small-diameter trees (Lutz et al., 2012). In addition to biological diversity and carbon storage, large-diameter trees may have a very high commercial value. Nevertheless, Korhonen et al. (2016) mention specifically that predicting the occurrence of large-diameter trees is challenging and exorbitant in terms of cost.

The aim here was to employ the ALS-based non-parametric *k*-most similar neighbour (*k*-MSN) approach to estimate the diameter distribution in a sub-tropical forest in southern Nepal. The specific objectives were: (1) to assess the accuracy of diameter distribution estimation for large trees (taking DBH > 30 cm to define a large tree) at the validation level; and (2) to assess the influence of the number of training areas (sample size, n) on the accuracy of the estimated diameter distribution by means of simulation with 100 repetitions. We used two approaches for the simulation studies, those developed by Junttila et al. (2013) and Rana et al. (2016) (see Section 3 for details). For the accuracy assessment, the empirical diameter distributions were compared with the estimated distribution derived from the sum of the power of the diameters (Kilkki and Päivinen, 1986) and error indices (Reynolds et al., 1988; Packalén and Maltamo, 2008). Although we highlighted the accuracy statistics for the large tree diameter distribution, we also report the accuracy for all DBH classes in order to gain a better understanding of the comparative analysis.

2. Data

The area studied here is located in the Terai Arc Landscape (Siwaliks and Terai) in the southernmost area of Nepal (27°14′N–29°08′N, and 80°15′E–85°49′E). Siwaliks is located to the north side of the study area whereas Terai is located to the south side. The study area covers around one fourth of the country's total area, and is influenced by tropical to sub-tropical climates with an average annual rainfall of 600 mm in the western to 1300 mm in the eastern party of the study area. The study area is famous for ecological, and economical services and is mainly dominant by deciduous trees including *Shorea robusta* (Sal), *Pinus roxburghii* (Chir Pine), and *Acacia catechu* (khair-sisau).

The field sample data were collected in March–May 2011. The data collection was completed in two stages. In the first stage, the entire study area (23,000 km^2) was divided into nine vegetation categories (i.e. hill-sal, sal, mixed, chir pine, riverine, degraded forest, grass, shadow and non-forest) by reference to the forest/non-forest map (30 m resolution) produced by Joshi et al. (2003). According to the above vegetation categories, the entire area was sub-divided into blocks (5 km $\times$ 10 km). After that, a total of 20 blocks (1000 km^2) were selected in conformity with PPS (probability proportional to size) stratified sampling and assigned randomly as far as the vegetation categories were concerned. The locations of the 20 blocks are shown in Fig. 1. In the last stage, a sample of 632 field plots were surveyed in the blocks, and the centres of the plots were positioned utilizing differential Global Positioning System (GPS). Systematic cluster sampling was employed for the allocation of the sample plots within blocks, each of which contained six clusters with eight plots in each cluster. The radius of a field sample plot was 12.62 m (equivalent to 500 m^2). The DBH values of all trees larger than 5 cm at breast height (1.3 m) on each plot were measured by means of diameter tape. In addition, every 5th tallied