Forest Ecology and Management 409 (2018) 552–563

Contents lists available at ScienceDirect

Forest Ecology and Management

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

Evaluation of alternative approaches for landscape-scale biomass estimation in a mixed-species northern forest

Coeli M. Hoover[a,*], Mark J. Ducey[b], R. Andy Colter[c], Mariko Yamasaki[a]

[a] USDA Forest Service, Northern Research Station, Durham, NH 03824, United States
[b] University of New Hampshire, Department of Natural Resources and the Environment, Durham, NH 03824, United States
[c] USDA Forest Service, White Mountain National Forest, Campton, NH 03223, United States

ARTICLE INFO

Keywords:
Aboveground biomass
LiDAR
Forest inventory

ABSTRACT

There is growing interest in estimating and mapping biomass and carbon content of forests across large landscapes. LiDAR-based inventory methods are increasingly common and have been successfully implemented in multiple forest types. Asner et al. (2011) developed a simple universal forest carbon estimation method for tropical forests that reduces the amount of required field measurements. We tested this approach, along with standard regression and Random Forest modeling techniques, in a northern hardwood-dominated watershed in the White Mountains of New Hampshire. Additional objectives included assessing the effects of different inventory plot designs and GPS accuracy. The universal model performed poorly in this forested landscape due to the lack of a clear relationship between canopy height and stand basal area. Simple regression modeling also produced poor model fits; the Random Forest models produced somewhat better biomass predictions than either the universal or regression models, and had low predictive power as measured by R^2 but root mean squared errors were comparable to those from other studies in complex forests. Effects of positional accuracy from survey vs. resource grade GPS units were slight, as were the effects of varying plot designs, although errors generally increased when larger basal area factors were used.

1. Introduction

Inventory and monitoring is an essential, but expensive, component of forest management. Inventory data are important for meeting multiple management objectives including timber production, wildlife habitat, forest health, and carbon sequestration (Kershaw et al., 2016). At regional to national scales, National Forest Inventory data meet some needs. For example, while the USDA Forest Service's Forest Inventory and Analysis Program (FIA) is a source of detailed forest inventory data (USDA Forest Service, 2017), the inventory is designed to be used at the state, regional, or national level, with one plot every 2428 ha (Bechtold and Patterson, 2005). As such, these data are generally not appropriate for landowners or managers due to the resolution of the sampling design. Because extensive field work is needed to collect inventory data at the level of stands or small landscapes, conducting routine forest inventories that meet an acceptable accuracy threshold is often quite expensive.

Airborne LiDAR, or light detection and ranging, employs a laser and high precision GPS to produce a three-dimensional representation of the ground beneath the aircraft's path; as the laser's energy hits a surface, it is reflected back to the instrument and recorded. Multiple returns are possible from each laser pulse. Airborne LiDAR has been in use for some time for terrain mapping; this product typically has a low return density (1–2 pulses per square meter, or ppsm) and is acquired when the forested portions of the landscape are in a leaf-off condition. Higher-density LiDAR data from full waveform and discrete return instruments have been used by researchers to assess various forest characteristics such as tree density, diameter, basal area (BA), and biomass (e.g. Lefsky et al., 1999, Beets et al., 2011, Hudak et al., 2006).

LiDAR studies of forest structure have occurred across a variety of biomes from tropical to boreal forests with variable model results; often with better results in conifer types or managed landscapes, where tree and forest structure is less complex and more regular. Zolkos et al. (2013) performed a meta-analysis on 70 studies reporting carbon or biomass across a variety of biomes to assess remote sensing approaches for measuring forest biomass. Biomass modeled from discrete return LiDAR data had an overall mean R^2 of 0.76, and a mean RMSE of 39.4 Mg/ha. In addition, they found that model error (in both absolute and relative terms) varied by forest type. Anderson and Bolstad (2013) estimated biomass in a Wisconsin forest by fitting LiDAR models by

---

* Corresponding author at: USDA Forest Service, Northern Research Station, 271 Mast Road, Durham, NH 03824, United States.
E-mail address: choover@fs.fed.us (C.M. Hoover).

https://doi.org/10.1016/j.foreco.2017.11.040
Received 28 September 2017; Received in revised form 14 November 2017; Accepted 18 November 2017
0378-1127/ Published by Elsevier B.V.

vegetation type, and report an R^2 of 0.74 and RMSE of 37.6 Mg/ha for coniferous stands, values of 0.71 and 42.8 Mg/ha for hardwood stands, while the mixed stands model had an R^2 of 0.44 with an RMSE of 48 Mg/ha. When all plots were included in the model, the R^2 was 0.55 with an RMSE of 43.5 Mg/ha.

While the cost of LiDAR data acquisition is dropping, use of these data for operational purposes requires that the field measurement component needed to model forest attributes be conducted efficiently. Because of the effort and cost associated with collecting information from a sufficient number of plots in each forest stratum, there have been efforts to generalize the modeling process to reduce the number of variables and/or plots measured. Lefsky et al. (2002) conducted an early test of a generalized model using waveform LiDAR at three sites in the US: temperate deciduous, temperate coniferous, and boreal coniferous (note that throughout this manuscript, deciduous refers to broadleaved species only). For both temperate coniferous and temperate deciduous sites, the R^2 values for the site specific and general (all sites combined) models were the same (0.87 and 0.65, respectively), while for the boreal conifer site the individual model outperformed the general (0.76 and 0.56, respectively). Use of a generalized model, if validated for a sufficient number of forests, would be one approach to reducing the field data collection burden. For tropical forests, Asner et al. (2011) developed a general approach to estimating aboveground biomass using mean canopy height from LiDAR data, and plot-level measurements of basal area and wood density weighted by basal area. Comparing the predicted and measured aboveground carbon for all 482 plots across four tropical study locations resulted in an R^2 of 0.95 with an RMSE of 15 MgC/ha. Substituting a regional wood density value produced an R^2 of 0.92. Asner and Mascaro (2014) tested a similar approach using hundreds of plots across 14 tropical ecoregions, and found that while LiDAR-derived canopy height accounted for 56% of the variation in aboveground carbon stock, a model that added basal area and wood density increased that value to 92%.

The intent of this study is to test if this type of generalized approach is feasible in the New England forested landscape, where deciduous, coniferous, and mixed stands are present. Use of a more generalized model with moderate resolution LiDAR data could provide an operationally feasible approach to LiDAR-based estimation of forest characteristics that would be practical for use by managers. We have four major objectives:

1. Test the Asner et al. (2011) approach for estimating aboveground biomass in a Northern hardwood forest.
2. Compare results from the Asner approach to those from conventional estimation methods.
3. Evaluate the suitability of moderate resolution LiDAR data for estimating common structural variables such as trees per hectare, basal area, and height.
4. Evaluate the impacts of changing plot design (including variable radius plot or prism sampling) and positional accuracy on the modeled outputs.

2. Materials and methods

2.1. Study area

The study was conducted in a small forested watershed on the Pemigewasset Ranger District of the White Mountain National Forest, located in Grafton County, New Hampshire, USA (Fig. 1). The study watershed is centered approximately at 44.0657°N, 71.8183°W and is 6885 ha in size. Elevation ranges from 328 to 1463 m, with slopes ranging from 0 to 85%. Annual precipitation averages about 1400 mm. The soils range from drainage classes of excessively drained to very poorly drained and have soil temperature regimes of frigid at the lower elevations to cryic at the higher elevations. The study area is predominantly of the soil order Spodosol and as soil parent materials of Lodgement and Ablation glacial tills, along with areas of Alluvium, Glaciofluvial and bedrock controlled outcrops. The vegetation is largely second growth and is a typical northern hardwood forest, consisting of sugar maple (*Acer saccharum*), American beech (*Fagus grandifolia*), yellow birch (*Betula alleghaniensis*), paper birch (*Betula papyrifera*), white ash (*Fraxinus Americana*), red oak (*Quercus rubra*), and red maple (*Acer rubrum*), with a conifer component of Eastern hemlock (*Tsuga canadensis*), balsam fir (*Abies balsamea*), white pine (*Pinus strobus*), and red spruce (*Picea rubens*).

2.2. Field data collection and processing

Plot locations were selected by stratified random sampling, with strata based on overlaying a conjectured soil group map (based on landform) with management zones defined by U.S. Forest Service regulations. A total of 176 plot locations were selected across the watershed. Field crews navigated to the specified coordinates for each plot using a recreational-grade GPS. Once the plot center was monumented and established, the plot was georeferenced using both a survey-grade (Trimble GeoXH, CE with Zephyr antenna) and resource-grade (Trimble GPS Pathfinder ProXH with Hurricane L1 antenna) GPS. Field data were collected in the summers of 2013 and 2014, after the final LiDAR acquisition was completed (the LiDAR data were used to inform plot selection).

All trees above 2.5 cm diameter at breast height (DBH) were measured using a mapped, nested-plot design. Trees from 2.5 to 12.6 cm DBH were measured on a 4.23 m radius fixed-radius plot, while trees from 12.7 cm to 30.0 cm DBH were measured on a 10 m radius fixed-radius plot. Trees 30.1 cm DBH and over were measured out to the limiting distance for a 2.25 m^2/ha basal area factor (BAF) variable radius plot. In addition to species, status as live or dead, and DBH (to the nearest 0.1 cm with a tape), the distance and bearing from the plot center to the pith of each tallied tree was recorded for all trees, allowing later simulation of sampling from a 10 m fixed-radius plot (with nested subplot for small trees), and variable radius plots with a range of BAF spanning and exceeding conventional inventory recommendations for the region (typically ranging from 3.5 to 4.6 m^2/ha, but with some practitioners using 2.3 m^2/ha; Wiant et al., 1984, Ducey, 2001), as well as the full original plot design. On all trees close enough to the plot center to be tallied using a 4 m^2/ha BAF variable radius plot, total height was measured using a Vertex hypsometer (Haglof, Inc.). The trees measured for height represent a size-weighted probability-based subsample of the full sample of those measured for DBH (Marshall et al., 2004; Kershaw et al., 2016, ch. 11).

To predict the heights of trees for which heights were not measured, we evaluated a series of regression equations using a mixed-effects modeling framework (Pinheiro and Bates, 2000), relying primarily on information-theoretic model selection using the Akaike Information Criterion (AIC) (Akaike, 1974, Burnham and Anderson, 2002) but with additional consideration of other regression diagnostics, including correlations between parameter estimates, error distribution and correlation, and Schwarz's Bayesian Information Criterion (BIC). The overall philosophy in model selection for this study was holistic, aiming at reliable predictions constructed from a model built on distributional assumptions that are satisfied by the data, rather than relying on automatic selection of a model by a single criterion (Claeskens and Hjort, 2008). We did not consider, and do not report here, p-values associated with regression coefficients: tree height is known to be correlated with tree diameter, and height-diameter relationships are known to depend on species and vary with site, both in general and in this region (e.g. Ducey, 2012), and since the null hypothesis of no relationship is not credible, such an approach would be a misuse of the null hypothesis testing paradigm (Anderson et al., 2000). We evaluated two primary model forms. The first followed Schumacher and Hall (1933) in log-transforming tree height H, and taking the reciprocal of DBH: