Generation of wall-to-wall canopy height maps using heterogeneous lidar datasets over a large region

Ranjith Gopalakrishnan[1,*], Valerie A. Thomas[1], John W. Coulston[2], Randolph H. Wynne[1]

[1] Dept. of Forest Resources and Environment Conservation, Virginia Tech, USA. Emails: ranjith7@vt.edu, thomasv@vt.edu, wynne@vt.edu

[2] USDA Forest Service (Southern Research Station), USA. E-Mail: jcoulston@fs.fed.us

***** Author to whom correspondence should be addressed; E-Mail: ranjith7@vt.edu

---

Abstract: The generation of robust and unbiased wall-to-wall vegetation canopy height maps from airborne lidar data for large regions is useful to forest scientists and natural resource managers due several reasons. Salient of these reasons are: 1) Parameterization of forest fire models; 2) Estimation of forest aboveground biomass; 3) Estimation of forest health and productivity. However, such spatial mapping over large areas often involves using data from disparate lidar projects, with widely varying acquisition parameters. In this work, we address the question of whether one can generate accurate canopy height maps over large areas (such as the Southeastern US) using a very heterogeneous lidar dataset, with more than 90 separate lidar projects. The field-data component of this effort from the Forest Inventory and Analysis (FIA) program of the US Forest Service. The use of this nationally uniform and extensive field data (~3000 plots is an unique aspect of this effort. We construct a simple bivariate linear model of plot-level canopy heights and distributional lidar metrics. Our initial results are promising: we observe a correlation of 82.3% between the 85th percentile of lidar heights and field- measured height (R2 = 0.67, RMSE =3.7 meters, n=3337) over all lidar projects. We also quantify the relative importance of several factors (like heterogeneity of land-use in the region of interest, point density, the predominance of hardwoods or softwoods) that may influence the accuracy of the wall-to-wall map generated.

Keywords: Forest inventory, forestry, forest mensuration, Lidar, canopy heights, wall-to-wall mapping, co-registration, FIA

Representative figures: Given below are some figures that represent the work and its results. All context needed is summarized in the captions.

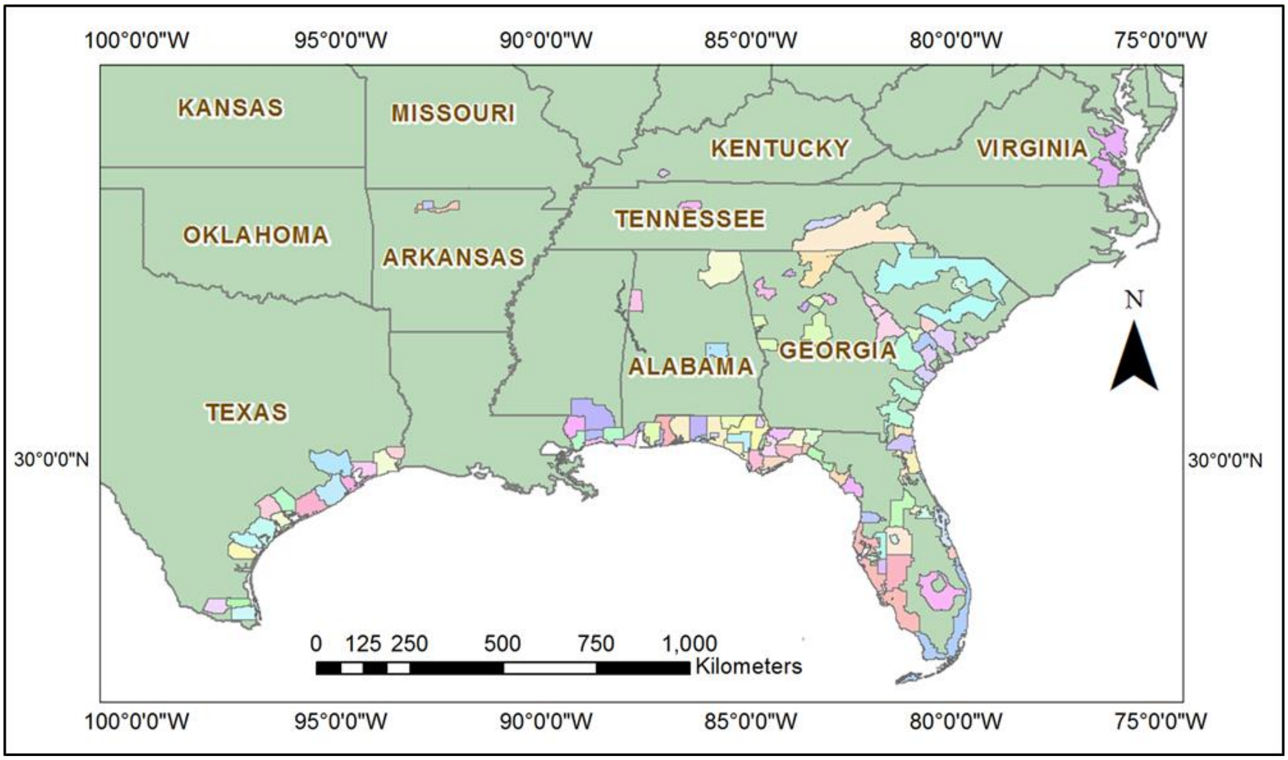


Figure 1: Study region (Southeastern United States). Lidar coverage of various projects (94 in all) is shown in different colors.

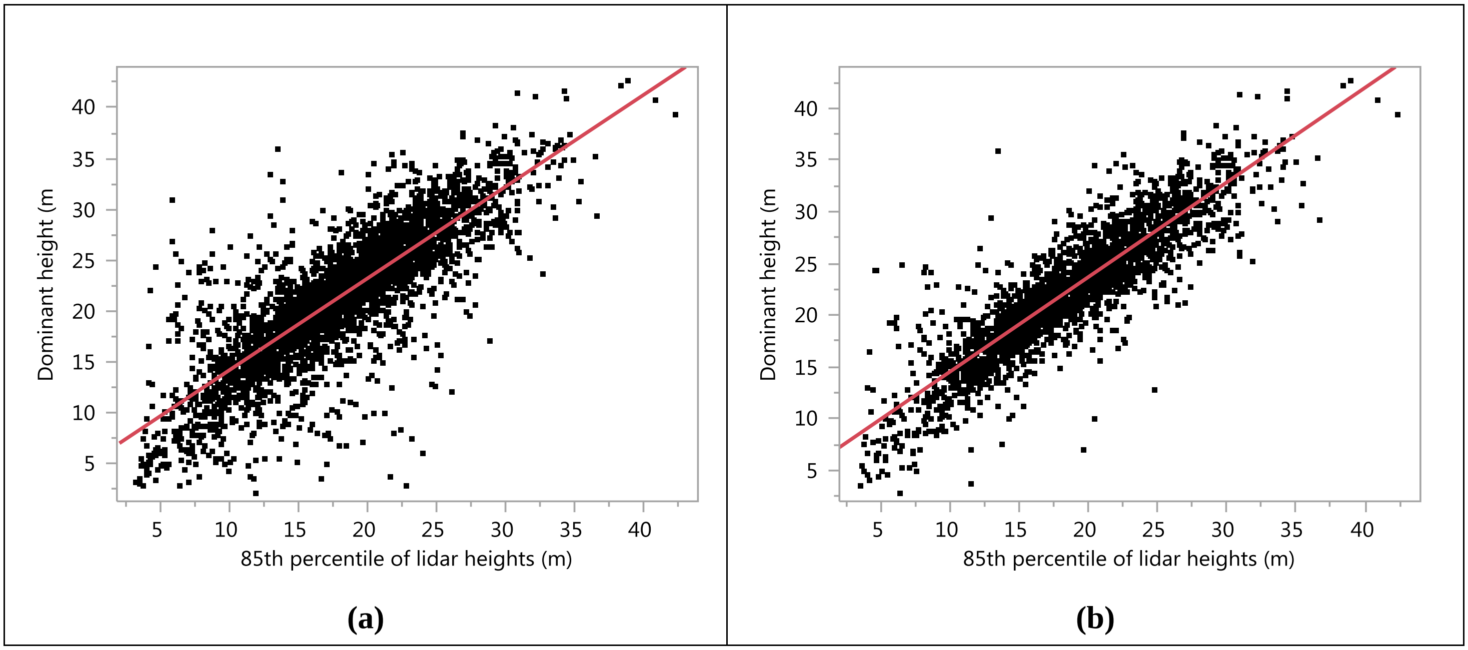
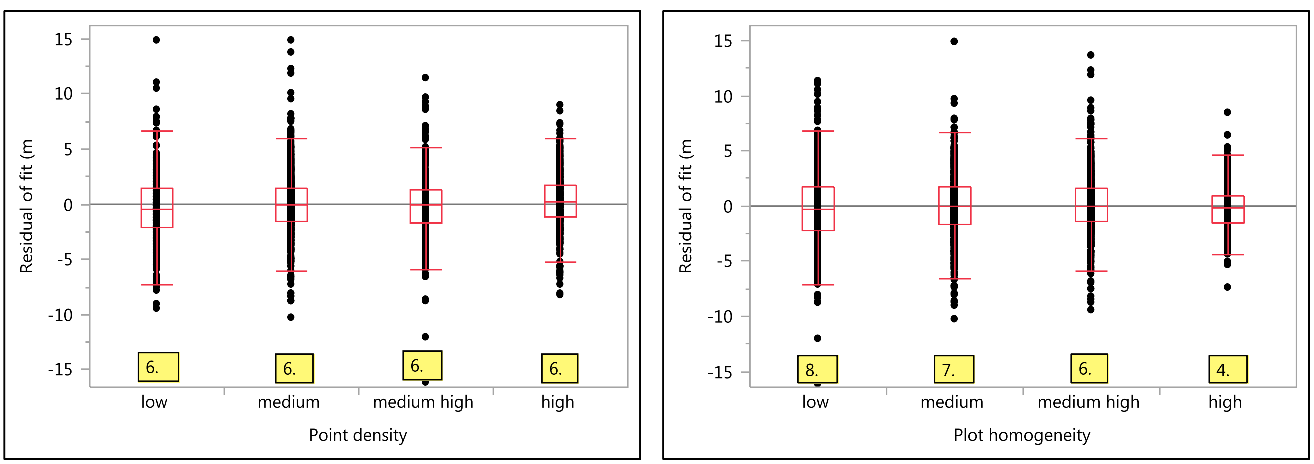
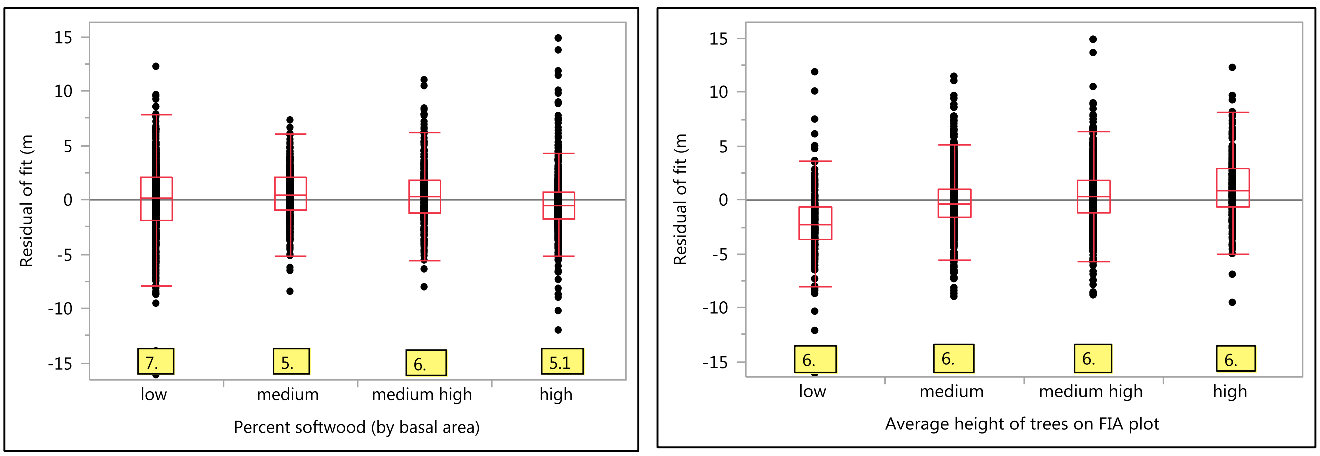


Figure 2 (a): Scatter-plot showing the correlation between the 85th percentile of lidar return heights and the FIA-measured dominant tree height, for all FIA plots used (n=3337 plots). For details on the methods used, see Gopalakrishnan *et. al* (2013). A simple linear regression fit using an OLS-based technique is also shown. **(b):** Scatter-plot showing the correlation between the 85th percentile of lidar return heights and the FIA-measured dominant tree height, for a subset of homogeneous FIA plots (n=2341 plots). A simple linear regression fit using an OLS-based technique is also shown.



(a) (b)

Figure 3 (a): Residual analysis from the fit shown in fig 2(b), to help understand the contributing factors. The effect of point density on the residuals of the linear fit. Here, 'low' denotes point densities of 0.25 to 0.5 points/m^2, 'medium' is 0.5 to 2.0, medium high is 2.0 to 4.0, and high is above 4.0. The values at the bottom (yellow boxes) are the quantile ranges, between the 10th and the 90th quantiles (this is same for other figures). They represent the spread of the residuals, for various ranges of point density. Higher point density has an effect of reducing the residuals, giving a better fit; **(b):** Effect of plot homogeneity on the residuals. For more on the importance of plot homogeneity in this context, see Gopalakrishnan *et. al* (2013). We use CV to quantify plot homogeneity, where higher CV values denote lower plot homogeneity. Here, the label of 'low' homogeneity denotes CV values between 0.375 and 0.5, 'medium' is 0.25 to 0.375, medium high is 0.125 to 0.25, and high is CV below 0.125. The significant shrinking of the quantile range (from 8.0 to 4.6 m) for increasing plot homogeneity is notable.



(a) **(b)**

Figure 4 (a): The effect of species groups (softwood versus hardwoods) on the residuals. The values at the bottom (yellow boxes) are the quantile ranges, between the 10th and the 90th quantiles. The quantile range is much greater in the hardwoods case; **(b):** Effect of height of the stand on the residuals. The x-axis represents the average height of all trees measured by the FIA (in the four subplots).

Table 1: The variation of RMSE (in meters) for plot selection using varying thresholds of point density (PD) and coefficient of variation (CV, a measure of plot homogeneity). The first cell (with a value of 2.86) indicates that when only plots with a PD >= 0.2 pts/m^2 and CV <= 0.5 are selected, the RMSE is 2.86 meters. The value of "NA" indicates that less than 50 plots were available, so RMSE was not computed. The set of plots with the lowest RMSE is highlighted. Also, one can see that significant improvements via PD comes into effect *only after* CV has been limited to lesser than or equal to 0.2 (see columns with CV labeled 0.2, 0.15 and 0.1).

PD \ CV	0.5	0.45	0.4	0.35	0.3	0.25	0.2	0.15	0.1
0.2	2.86	2.79	2.76	2.63	2.57	2.53	2.49	2.23	1.70
0.5	2.78	2.69	2.66	2.54	2.47	2.39	2.29	2.16	1.71
1.0	2.80	2.71	2.67	2.55	2.48	2.41	2.30	2.14	1.64
1.5	2.70	2.64	2.57	2.44	2.40	2.33	2.28	2.20	1.56
2.0	2.75	2.67	2.59	2.42	2.38	2.30	2.24	2.22	1.63
2.5	2.73	2.66	2.57	2.41	2.37	2.28	2.21	2.14	1.64
3.0	2.72	2.64	2.54	2.43	2.38	2.28	2.19	2.18	**1.53**
4.0	2.51	2.48	2.46	2.43	2.39	2.37	2.20	2.32	NA
5.0	2.59	2.56	2.54	2.51	2.45	2.40	2.19	2.26	NA
6.0	2.78	2.80	2.77	2.76	2.68	2.56	2.29	2.42	NA

References

1. Gopalakrishnan et. al., "Efficacy of using heterogeneous lidar datasets in predicting canopy heights over a large region", Proceedings of the SilviLaser conference 2013, Beijing, China.