Project Report: Applications of Small Area Estimation over the Contiguous United States: Testing and Development of Alternative Methods

Reporting Period: October 1, 2023 through December 31, 2023

Project Personnel:

Phil Radtke, Principal Investigator. Virginia Tech Department of Forest Resources and Environmental Conservation (FREC), Blacksburg, VA.

Qianqian Cao, Co-Investigator and Lead Analyst, Virginia Tech FREC, Blacksburg, VA.

Corey Green, Co-Investigator, Virginia Tech FREC, Blacksburg, VA.

Valerie Thomas, Co-Investigator, Virginia Tech FREC, Blacksburg, VA.

Randy Wynne, Co-Investigator, Virginia Tech FREC, Blacksburg, VA.

John Coulston, Project Cooperator for FIA Support. U.S. Forest Service Southern Research Station, Blacksburg, VA.

Todd Schroeder, Project Cooperator for FIA and Geospatial Support. U.S. Forest Service Southern Research Station, Knoxville, TN.

Work Completed:

Canopy height models (CHM) from two sources were compared as ancillary data in area-based Fay-Herriott (FH) county level estimates of forest volume and aboveground biomass in North Carolina, Tennessee, and Virginia (Cao et al. 2024). One CHM was obtained from available whole-state NAIP 3d acquisitions during summer, 2108. The other was obtained from a published global forest canopy height model (GFCHM) developed from Landsat vegetation data paired with GEDI L2A and L2B canopy height, cover, and vertical profile information (Potapov et al., 2021). Reductions exceeding 70% were achieved in North Carolina averaged county-level standard errors for both volume and biomass. Results were more modest in the other two states, with standard error reductions of 32% and 42% in Tennessee and Virginia, respectively. Both CHM sources allowed for substantial improvements in estimator uncertainties compared to using FIA data alone, with slightly better results overall using the NAIP-derived CHM combined with a Landsat forest-type filter. Despite satellite-derived GFCHMs being slightly less efficacious overall in reducing the variance of FIA county-level estimates, differences were typically small. The work indicates that increased precision in estimated volume or biomass attributes should be possible from either type of CHM.

Work In Progress:

Building on the county-level estimation framework and nearly equal performance of the (GEDI) satellite-derived GFCHM compared with NAIP 3d CHMs in three eastern states, a follow-up manuscript, “County-level aboveground biomass estimates for the contiguous United States” (working title) has been drafted and is currently being refined for submission for peer review. Preliminary results show that, while county-level FH estimates, in general, improve on direct FIA estimates across CONUS, adjustments to state and/or county groupings are needed in two

situations. First, where states, e.g. in the Northeast and along the Eastern Seaboard, have few counties to stand alone as a population of "small area" domains in FH modeling, grouping 5-7 adjacent states together seems to work well. Second, where very few non-zero forested plots arise in some counties – primarily in prairie regions of the Upper Midwest – grouping counties into Survey Unit "supercounties" and combining their states into a second regional group – so the number of domains is sufficiently large – seems to work well. Expected date of submission is March, 2024.

Next Period Plans:

Investigating the use of two-group FH random effects seems warranted based on our observation that some counties (or groups of counties) appear to be "outliers" in terms of how well a synthetic model predicts compared to the model's error distribution for other counties in a population. Published research such as that of Herrador et al. (2011) indicates that such "outliers" might justify separating the population of domains into two distinct groups, both suited for description by the same fixed-effects, but whose random effects are better described as having different variances. In the context of forest inventory these groups might represent disjoint physiographic or ecological sections. Preliminary work presents challenges for interpreting grouping results logically. Tests involving simulated populations may help shed light on this challenge.

The sub-area or nested random effects approach also looks for better ways to accurately model prediction error variances. In this case the rationale is that counties in the same FIA survey unit may have smaller among-area variances than counties in different survey units. Formulating random effects as nested and additive is the approach taken by Torabi and Rao (2014). A possible alternative to nested random effects is the accounting for spatial correlations. Comparing these alternatives may prove informative to PSAE.

Publication in review:

Cao Q, Radtke PJ, Coulston JW, Thomas VA, Wynne RH, Walker DM. 2024. Comparing canopy height models from regional-scale aerial photogrammetry with global spaceborne lidar-derived data for estimating forest volume and biomass. Forest Science *in review*.

Literature Cited:

M. Herrador , M. D. Esteban , T. Hobza & D. Morales (2011) A Fay–Herriot Model with Different Random Effect Variances. Communications in Statistics - Theory and Methods, 40:5, 785-797, DOI: 10.1080/03610920903480858

Potapov P, Li X, Hernandez-Serna A, Tyukavina A, Hansen MC, Kommareddy A, Pickens A, Turubanova S, Tang H, Silva CE, Armston J, Dubayah R, Blair JB, Hofton M, 2021. Mapping global forest canopy height through integration of GEDI and Landsat data. Remote Sensing of Environment.

Torabi M, Rao JNK. 2014. On small area estimation under a sub-area level model. Journal of Multivariate Analysis, 127:36-55, DOI: 10.1016/j.jmva.2014.02.001.