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

Reporting Period: January 1, 2024 through December 31, 2024

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.

Bipana Subedi (graduated 12/24), Graduate Research Assistant (M.S.), Virginia Tech FREC, Blacksburg, VA.

Garret Dettman, unfunded Ph.D. student, Virginia Tech FREC, Blacksburg.

Aakriti Sapkota, Graduate Research Assistant (M.S.), 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:

1. CONUS wide county-level biomass SAE

Small area estimation of county-level aboveground biomass (AGB) estimates for the forty-eight contiguous United States was carried out using Fay-Herriot (FH) methodology. Canopy height models (CHM) from a published global forest canopy height model (GFCHM) dataset of (Potapov et al., 2021) were used as auxiliary data. Precision gain was met for 37 single states, with standard error reductions from 2% to 70%. Six states with limited number of counties, including Connecticut, Delaware, Massachusetts, New Hampshire, Rhode Island, and Vermont, are grouped as a multi-state population (with counties as domains) to reach substantial improvements in uncertainty compared to individual states modeled as separate populations. Five states of Iowa, Kansas, North Dakota, Nebraska, and South Dakota – mostly located in the Great Plains with low forestland area and forested plot sample sizes < 10 per county on average – were grouped as a single population to mitigate effects of sparse data. Domains in this cluster of states were defined at the FIA survey unit level, following the Green Book "super-county" characterization, which led to reliable direct estimation and precision gains via FH model prediction. The work showed notable precision gains in county-level AGB estimates CONUS wide (Cao et al. 2025).

2. Testing Fay-Herriot inferences in county-level biomass estimates

In estimating count-level biomass CONUS-wide, the FH mixed model estimates unobservable random effects – also known as area effects – to calculate EBLUPs for each small area domain. In some states the FH method estimates the variance of area effects as zero, reducing all area effects to zero and collapsing the two-component estimator to a single regression-synthetic component. We tested the FH inferences using simulated populations both with and without nonzero random effects variances following work of Molina et al. (2015) and Rao (2014). Simulation results showed biases in EBLUP MSE when the FH model collapsed to the synthetic component of the estimator, regardless of whether simulated populations included positive area-effect variance or not. We tested adjusted random effect estimators constrained to give positive area-effect variances (Li and Lahiri, 2010; Yoshimori and Lahiri, 2014). Our tests led to a finding that a conditional variance estimator reduces biases in EBLUPs and improves statistical validity of confidence intervals. This result should help practitioners to implement a conditional approach, where, when the conventional REML estimator gives a random effect variance, an adjusted estimator should be adopted in its place to ensure the area effects variance is positive.

3. Bivariate Fay-Herriot small area estimation of county-level forest biomass and volume

We tested a bivariate FH model (BFH) for joint estimation of forest biomass and volume at the county level. Bivariate models estimate EBLUPS simultaneously for biomass and volume, accounting for the typically strong correlation between the two attributes. Applying the BFH methodology to FIA led to greater gains in precision for both attributes compared to independently formulated univariate FH EBLUPs.

4. Unit level SAE using aerial lidar as auxiliary data in Coastal Georgia pine plantations

Industrial southern yellow pine plantation data and high-resolution aerial laser scanning (ALS) shared with us by corporate partner Rayonier, Inc. were combined in a unit-level Battese, Harter, Fuller (BHF; Battese et al., 1988) framework to generate plantation level estimates of volume and biomass attributes with increased precision. Recent M.S. student Bipana Subedi graduated in December and has begun working for Rayonier in their Carbon group.

Work In Progress:

Manuscripts titled: (1) County-level aboveground biomass estimates for the contiguous United States (Cao et al. 2025a) and (2) Testing Fay-Herriot inferences from U.S. national forest inventory of county-level biomass (Cao et al. 2025b), are in final draft stages, with submission for peer review expected date in the first quarter of 2025. A third one titled as bivariate Fay-Herriot small area estimation of county-level forest biomass and volume is in preparation.

Two-group FH random effects: Use of two-group FH random effect variances seems warranted based on our observation that some counties appear to be “outliers” in a population. Published research 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. Preliminary

work in context of county-level AGB shows improvement of precision gain compared to the basic FH model with one estimates of random effect through whole population in some states, e.g., Florida. Preliminary work presents these groups might represent ecological sections. Pagliarella et al. (2021) introduced the small area estimation diagnostics in case of the FH model, which may help shed light on the challenge about when and where the two-group random effect variance is needed.

Exploring the sub-area or nested random effects approach is in process. 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). Comparing the sub-area or nested random effects to the classical FH model is in progress.

We are testing a new simulation based on the Rayonier Coastal Georgia data that generates detailed stand simulations and sampling capabilities for evaluating biases and confidence interval performance under different field sampling designs. A manuscript from B. Subedi's thesis is in preparation to be submitted for peer review in the second quarter of 2025.

Ph.D. student Dettman is preparing a study plan for PSAE research using quasi-unit level SAE with proxy height measurements. A primary objective is to test the potential for implicit pairing of FIA plots to NAIP 3D canopy heights that does not require precise plot coordinates. Expected approval of the work plan, first quarter, 2025.

Next Period Plans:

Investigating the use of area-level Poisson mixed model. Boubeta et al. (2016) studied the applicability of area-level Poisson mixed models to estimate small area counting indicator. In the context of forest inventory, tree mortality and relative analysis that may indicate the level and trend of forest health problems. At county-level, the sampling relative standard error high (e.g., 70% in average in Virginia 2022) and shows the need of small area estimation. Preliminary work presents challenges of effective auxiliary data for tree mortality that caused by insect attacks and disease.

Publication accepted:

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, *accepted for publication in* Forest Science.

Publications in preparation:

Cao Q, Radtke PJ, Coulston JW. 2025. County-level aboveground biomass estimates for the contiguous United States. *Manuscript in preparation*.

Cao Q, Radtke PJ, Coulston JW. 2025. Testing Fay-Herriot inferences from U.S. national forest inventory of county-level biomass. *Manuscript in preparation*.

Cao Q, Radtke PJ, Sapkota A. 2025. Bivariate Fay-Herriot small area estimation of county-level forest biomass and volume. *Manuscript in preparation*.

Cao Q, Radtke PJ, Coulston JW. 2025. Two-group random effect variances in Fay-Herriot of county-level biomass. *Manuscript in preparation*.

Subedi B, Radtke P, Coulston J. 2025. Improving Precision in Forest Inventory through Small Area Estimation for Loblolly Pine Plantations in Coastal Georgia. *Manuscript in preparation*.

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

Battese GE, Harter RM, Fuller WA, 1988. An Error-Components Model for Prediction of County Crop Areas Using Survey and Satellite Data. Journal of the American Statistical Association 83:28-36.

Boubeta M, Lombardía MJ, Morales D, 2016. Empirical best prediction under area-level Poisson mixed models. Test (Madrid, Spain) 25:548-569.

Li H, Lahiri P, 2010. An adjusted maximum likelihood method for solving small area estimation problems. Journal of Multivariate Analysis 101:882-892.

Molina I, Rao JNK, Datta GS, 2015. Small area estimation under a fay-herriot model with preliminary testing for the presence of random area effects. Surv Methodol 41:1-19.

Pagliarella MC, Porzio GC, Balzano S, Vistocco D, Vichi M, Salvatore R, Balzano S, Porzio GC, Vichi M, Salvatore R, Vistocco D, 2021. Small Area Estimation Diagnostics: The Case of the Fay–Herriot Model Switzerland: Springer International Publishing AG. 149-162.

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.

Rao JNK, 2014. Inferential Issues in Model-Based Small Area Estimation: Some New Developments. Statistics in Transition New Series 16:20.

Yoshimori M, Lahiri P, 2014. A new adjusted maximum likelihood method for the Fay–Herriot small area model. Journal of Multivariate Analysis 124:281-294.

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.