PI: Temesgen Hailemariam, Department of Forest Engineering, Resources and Management, Oregon State University.

Co-PI: Sukhyun Joo, Research Associate of Forest Biometrics, Department of Forest Engineering, Resources and Management, Oregon State University**.**

Bryce Frank, US Forest Service FIA Support, Pacific Northwest Research Station

Collaborators:

Suchana Aryal, PhD Student in Forest Biometrics, Department of Forest Engineering, Resources and Management, Oregon State University, Corvallis, OR

Dinka Biru, Postdoctoral Scholar, Department of Forest Engineering, Resources and Management, Oregon State University, Corvallis, OR

Jacob Strunk, US Forest Service, Pacific Northwest Research Station, Olympia, WA.

Dale Hogg, Green Diamond Resource Company, Bogart, GA.

Nathaniel Naumann, PotlatchDeltic, Spokane, WA.

Cristian Montes, Rayonier, Yulee, FL

Aaron Weiskittel, University of Maine, Orono, ME.

Phil Radtke, Virginia Tech University, Blacksburg, VA.

Rachel Cook, North Carolina State University, Raleigh, NC.

Ethan Hughes, Washington State Department of Natural Resources, Olympia, WA.

Francisco Mauro, Universidad de Valladolid, Spain.

Project Title: The Interplay between Sampling Design and Small Area Estimation to Improve Stand- and Forest-level Estimates

Period for report: January 1, 2025 to June 30, 2025

Progress: We acquired NAIP imagery for the states of Oregon and Washington. Following the acquisition, we have conducted:

1. **Data Integration and Processing**

We compiled a comprehensive set of LiDAR and Sentinel-based covariates for Klamath County, Oregon (provided by Green Diamond Company), to demonstrate the potential for incorporating Fay-Herriot (FH) models into operational forest inventory.

To examine the performance of selected sampling strategies, we created a unified (cruise + auxiliary) analytical dataset. Cruise data was available for 193 stands (total plots = 3825). Cruise data included all relevant stand information, such as species, diameter at breast height, height, tree density, and merchantable volume.

2. **Variable Selection and Small Area Model Setup**

We examine selected variable selection methods for developing small-area estimation models that link FIA, stand exam data, and auxiliary variables from 3D-NAIP, Sentinel, climate, and terrain datasets. We have reviewed the literature and selected eight variable selection methods to identify a reasonable number

of predictors from a wide range of topographic, climatic, and remotely sensed attributes. To improve prediction at the stand and forest/ownership levels, we have compared the performance of eight t variable selection methods for Phase 1 of the project.

For the second phase of the project, which examines the interplay of sampling design and small area estimation, a common challenge when working with auxiliary variables is the large number of potential predictors. In our dataset, we started with 675 auxiliary variables. After removing variables with near-zero variance across stands (using a near-zero variance filter) and those with highly correlated predictors (absolute Pearson correlation > 0.90 was considered redundant), we were left with 57 variables, which is still too many to evaluate every possible model. We fitted a model with 57 variables filtered out and performed VIF pruning, resulting in 41 remaining predictors (6% of the original). On those predictors, we did final variable selection. Different model selection criteria, such as forward selection with the Bayesian information criterion (BIC) and LASSO regularization, were applied to generate a shortlist of candidate models that best predict stand-level merchantable volume without overfitting. The selected variables were confirmed with the best subset selection (BIC k = 8), and the model was fitted and inspected to ensure they met statistical assumptions.

3. **Monte Carlo Simulation**

We conducted a Monte Carlo simulation to examine the performance of selected sampling designs and sample sizes and quantify their performance using metrics such as RMSE, bias, and coefficient of determination. Simulated multiple two-stage sampling designs to assess their performance under different sampling intensities (5% to 40%). Selected sampling designs were evaluated for their ability to estimate merchantable volume, using performance statistics including relative bias and relative root mean squared error (RRMSE), based on 1,000 iterations.

To date, we have evaluated 10 sampling strategies and are actively exploring additional designs that align with the practical needs and priorities of industry stakeholders.

4. **Documentation and Communication**

We presented preliminary findings in CIPS (Center for Intensive Planted-forest Silviculture) and CAFS (Center for Advanced Forestry Systems) meetings. We are currently incorporating the feedback received and exploring alternative sampling designs, as well as feedback from the recent PSAE (Partnership for Small Area Estimation) meeting.

Next period plans:
We plan to: 1) complete the Monte Carlo simulation of additional sampling designs, 2) write a manuscript on variable selection methods and related topics, and 3) address the uncertainty of the FH model for selected optimal sampling designs and variable sets theoretically and empirically.

Problems or delays:

None