Continuing Project

# Using Small Area Estimation and 3D-NAIP/Sentinel-derived Variables for Multivariate Prediction of Stand Attributes

(CAFS.24.107)


**PI:** Sukhyun Joo, Oregon State University

**Co-PI:** Temesgen Hailemariam, Oregon State University

**Co-PI:** Bryce Frank, USDA FIA

**Collaborators:** Dale Hogg (Green Diamond Resources); Nathaniel Naumann (PotlatchDeltic); Cristian Montes (Rayonier); Phil Radtke (Virginia Tech); Rachel Cook (NC State University); Ethan Hughes (Washington DNR); Francisco Mauro (University of Valladolid)

Presenter: Sukhyun Joo

# Justification

- Why Small Area Estimation (SAE)?
  - Traditional forest inventory provides reliable estimates at large scales
  - Management decisions require stand-level estimates
  - Direct estimates at stand level have high sampling variance

- SAE borrows strength from:
  - Similar stands (spatial borrowing)
  - Auxiliary data (remote sensing)

# Justification

- Previous research has focused on using area- or unit-level predictors within a univariate framework, analyzing each variable independently without accounting for their correlations.
- Many forest inventory variables exhibit strong correlations, such as basal area, stand volume, and trees per acre.
- In such cases, employing multivariate responses can enhance the accuracy of the estimates.

# Hypotheses or Objectives

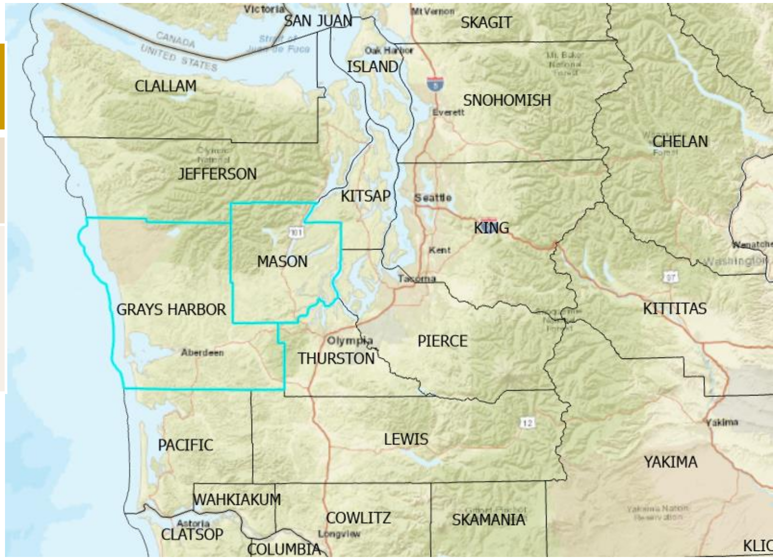
- Primary Objective: Develop accurate predictions for key forest attributes
  - Trees per acre (TPA)
  - Basal area (BA, $ft^2/ac$)
  - Merchantable cubic foot volume (CFV, $ft^3/ac$)

- Methodological Objectives:
  - Compare univariate vs. multivariate SAE approaches
  - Evaluate different random effects structures
  - Assess improvement over direct estimation

# Methods

- Washington area focusing on Grays Harbor and Mason Counties

| Plots/Stand | 1-5 | 6-10 | 11-20 | >20 |
|---|---|---|---|---|
| n | 1 | 7 | 57 | 154 |
| Percentage (%) | 0.5 | 3.2 | 26 | 70.3 |



## Response variables

| Variable | Mean | Min | Max |
|---|---|---|---|
| TPA (trees/ac) | 297.6 | 19.4 | 1624.9 |
| BA ($ft^2$/ac) | 143.4 | 3.0 | 531.1 |
| CFV ($ft^3$/ac) | 3765.7 | 23.8 | 19987.6 |

# Methods

## Auxiliary Variables

- Data sources
  - 3D-NAIP imagery (4-band, summer acquisitions)
  - Sentinel-2 (multispectral, seasonal)

- Variable categories
  - Height metrics (ht_mean, ht_p99)
  - Spectral indices (NDVI, band differences)

# Methods

## Unit-level Mixed Model:

$$y_{ij} = \boldsymbol{X}_{ij}^{\top}\boldsymbol{\beta} + \upsilon_i + \varepsilon_{ij}$$

$y_{ij}$ = plot $j$ in stand $i$
$\boldsymbol{X_{ij}}$ = vector of auxiliary variables for $y_{ij}$
$\boldsymbol{\beta}$ = vector of fixed-effects coefficients
$\upsilon_i$ = stand-level random effect; $\upsilon_i \sim N(0, \sigma_\upsilon^2)$
$\varepsilon_{ij}$ = residual error; $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$

# Methods

- Empirical Best Linear Unbiased Predictor (EBLUP) for stand $i$

$$\hat{y}_i = \bar{\boldsymbol{X}}_i^\top \widehat{\boldsymbol{\beta}} + \hat{v}_i$$

- Here $\bar{\boldsymbol{X}}_i$ is the mean of the predictor vectors for stand $i$, $\widehat{\boldsymbol{\beta}}$ is the estimated fixed-effect vector, and $\hat{v}_i$ is the BLUP of the stand random effect.

# Methods

## Univariate Model Selection

- Stage 1: Find optimal complexity
  - Exhaustive search (1-8 predictors)
  - Select by adjusted $R^2$ elbow point
- Stage 2: Apply constraints & fit mixed models
  - VIF threshold = 5
  - Test multiple random effect structures

# Results

- Univariate modeling method
  - Battese-Harter-Fuller (BHF) model
  - Unit-level model
  - Area-specific random effects
  - Best models by AIC
  - CFV performed best with variance weights

| Response | EBLUP RMSE | EBLUP RMSE (%) |
| --- | --- | --- |
| TPA (trees/ac) | 13.53 | 4.48 |
| BA ($ft^2$/ac) | 6.49 | 4.46 |
| CFV ($ft^3$/ac) | 278.5 | 7.24 |

# Results

- TPA
  - ht_l_cv: Height L-moment Coefficient of Variation
  - NDMI: Normalized Difference Moisture Index
  - TCI_G: Terrestrial Chlorophyll Index Green band
  - ht_p10: 10th Percentile Height

- BA
  - ht_quadratic_mean
  - ht_maximum

- CFV (with variance weighting):
  - ht_l_cv
  - ht_p99
  - NDMI

# Methods

- Multivariate model
  - Joint Modeling of TPA, BA, and CFV
  - Borrows strength across responses
  - Maintains correlation structure
  - Single model for all attributes

## Long format mixed model

$$y_{ijk} = \mu_k + \boldsymbol{X}_{ij}^{\top}\boldsymbol{\beta}_k + \upsilon_{ik} + \varepsilon_{ijk}$$

$k$ indexes the trait: TPA, BA, and CFV.
$\mu_k$ is the trait-specific intercept.
$\boldsymbol{X}_{\boldsymbol{ij}}$ is the auxiliary variable vector for observation $j$ in stand $i$.
$\boldsymbol{\beta}_k$ is the vector of trait-specific fixed effect coefficients.
$\upsilon_{ik}$ is the random effect for stand $i$ and trait $k$; assumed $\upsilon_{ik} \sim N\left(0, \sigma_{\upsilon,k}^2\right)$.
$\varepsilon_{ijk}$ is the residual error for observation $j$ in stand $i$ and trait $k$; $\varepsilon_{ijk} \sim N(0, \sigma_{e,k}^2)$.

# Methods

- Variables
  - Combined predictors from TPA, BA, and CFV univariate models.
  - Applied VIF screening to avoid multicollinearity.

- Preliminary modeling approach
  - Multivariate model with stand random effects
    - Stand-level correlations modeled through random effects
    - Plot-level residuals assumed independent

## Performance comparison

| Method | TPA RMSE | BA RMSE | CFV RMSE | TPA CV (%) | BA CV(%) | CFV CV (%) |
|---|---|---|---|---|---|---|
| Direct | 25.3 | 11.7 | 369.5 | 8.4 | 8.1 | 9.6 |
| Univariate | 13.5 | 6.5 | 278.5 | 4.5 | 4.5 | 7.2 |
| Multivariate | 59.9 | 29.2 | 60.0 | 19.8 | 20.1 | 1.6 |

# Progress 2024-25

## FIA unfuzzed data status

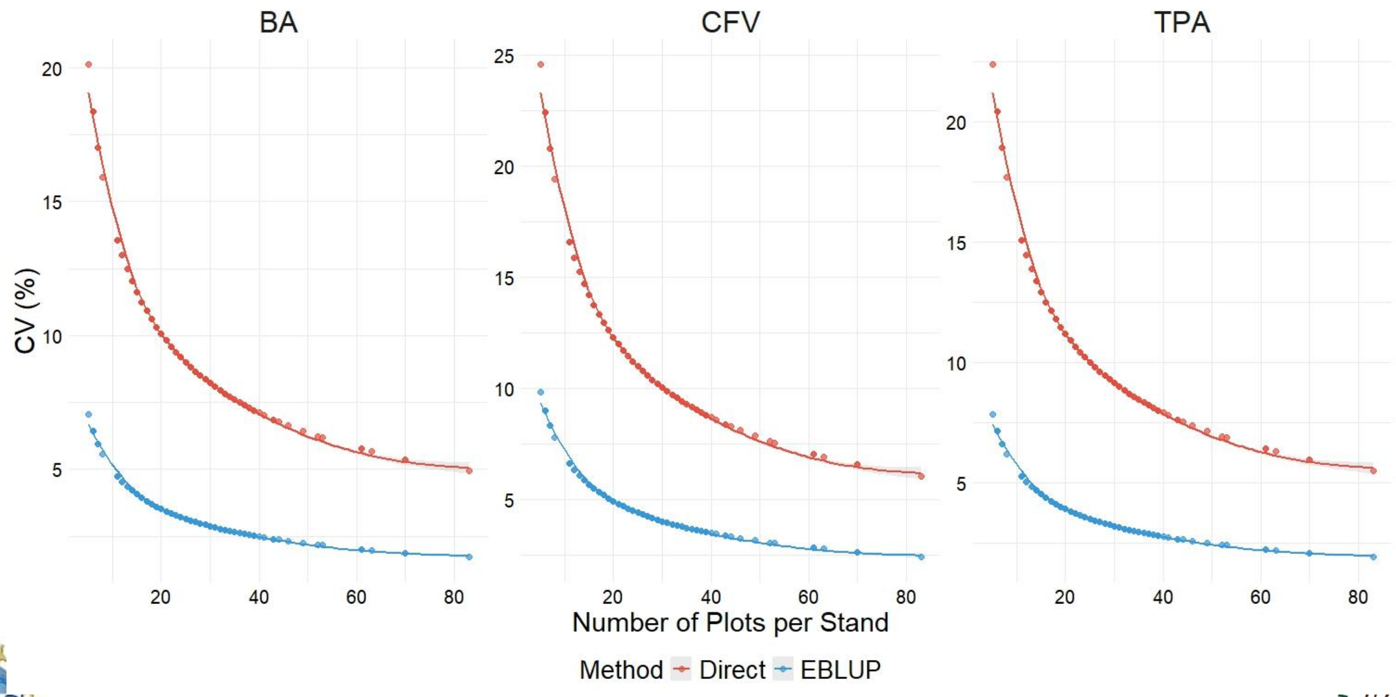
- Material Transfer Agreement (MTA) completed last week
- All signatures obtained
- Data access now approved
    - OR, WA, GA, AL

# Company Benefits

| StandID | # Plots | TPA Direct | TPA EBLUP | TPA CV Direct (%) | TPA CV EBLUP (%) |
|---|---|---|---|---|---|
| 19N07W08008 | 5 | 252.0 | 254.1 | 22.4 | 7.8 |
| 13N04W21002 | 6 | 159.4 | 161.1 | 20.4 | 7.1 |
| 13N05W24004 | 11 | 378.5 | 389.6 | 15.1 | 5.3 |
| 13N02E09005 | 21 | 329.5 | 333.8 | 10.9 | 3.8 |
| 20N05W04019 | 52 | 439.6 | 441.0 | 6.9 | 2.4 |

# Deliverables

- SAE provides substantial improvement over direct estimation 25-40% reduction in RMSE.
- Greatest gains for small samples.



# Future Plans

- Develop models with FIA unfuzzed data
  - Material Transfer Agreement (MTA) now complete.
- Validate models on independent industry data
- Refine multivariate predictor selection
  - Test alternative variable selection strategies
- Try to improve multivariate model structure
  - Random effects

# Summary

- SAE provides substantial improvement over direct estimation 25-40% reduction in RMSE.

- 3D-NAIP height metrics and Sentinel-2 spectral indices proved successful auxiliary predictors in the SAE model.

- FIA unfuzzed data are now accessible.

- Refine predictor selection and test alternative multivariate structures.