Continuing Project

The Interplay of Sampling Design and Small Area Estimation

Project Code (CAFS.23.104)

Temesgen Hailemariam and Suchana Aryal,
Oregon State University

Presenter: S Suchana Aryal

Justification

S T A N D:

Operative Decision Unit

I M P L I C A T I O N

If our per-stand numbers are off, every downstream operational decision inherits that error.

Justification

The Scaling Problem:

Hundreds to thousands of stands per ownership

Resource Constraints

Field budgets ↑ linearly with # plots

Insufficient Sample Sizes

Budget pressure ⇒ few or no plots per stand

↓

$$Var\left(\hat{Y}_i\right) = \frac{\hat{\sigma}_i^2}{n_i}$$ (n_i small) ⇒ $\hat{Y}_i$ imprecise

$n_i \approx 0$ ⇒ $\hat{Y}_i$ unknown

Justification

Fay–Herriot small-area estimation (SAE):

Data Integration and Model Development

FH combines ground estimates (even from VRPs) with LiDAR/Sentinel-derived predictors to enhance estimation capabilities.

Estimation

Generate empirical best linear unbiased predictions (EBLUP) for both sampled and unsampled stands.

Benefits

Shrinks noisy direct estimates, reducing mean-squared error (MSE). Robust to imprecise plot locations.

Objectives

Our study evaluates the efficiency of alternative sampling designs and sampling intensities for stand-level SAE of merchantable volume

- Does sampling in proportion to known attributes (e.g. volume or area) yield measurable gains in estimator efficiency?
- Does stratifying yield more efficient stand-level small area estimates than unstratified designs across sampling intensities?
- How does efficiency change as the overall sampling intensity ϕ increases?

Methods

Fay-Herriot Models for Stand-Level Estimation

Sampling Model: $\hat{\delta}_d^{DIR} = \delta_d + e_d$

Linking Model: $\delta_d = x_d^T \beta + v_d$

where for all domain d = 1, ……, D:

- Sampling errors $\{e_d\}$ are independent with $e_d \sim \mathcal{N}(0, \psi_d)$
- Random effects $\{v_d\}$ are independent with $v_d \sim \mathcal{N}(0, \hat{\sigma}_v^2)$
- The errors and random effects are mutually independent

This specification leads to the basic Fay-Herriot representation:

$$\hat{\delta}_d^{DIR} = x_d^T \beta + v_d + e_d, \; d = 1, \ldots, D$$

Methods

Fay-Herriot Models for Stand-Level Estimation

When the hyperparameters β and σ_v^2 are known, inferences on δ_d are based on the following posterior distribution of δ_d:

$$\delta_d | \hat{\delta}_d^{DIR}, \beta, \sigma_v^2 \sim \text{iid } \mathcal{N}(\hat{\delta}_d^{FH}, g_{1d}(\sigma_v^2)); \; g_{1d}(\sigma_v^2) = \gamma_d \psi_d$$

Once the FH model is fitted, stand-level estimates are obtained using EBLUP as:

$$\hat{\delta}_d^{FH} = \gamma_d \hat{\delta}_d^{DIR} + (1 - \gamma_d) x_d \beta$$

where γ_d is a shrinkage factor defined as:

$$\gamma_d = \frac{\sigma_v^2}{\sigma_v^2 + \psi_d},$$

Methods

Fay-Herriot Models for Stand-Level Estimation

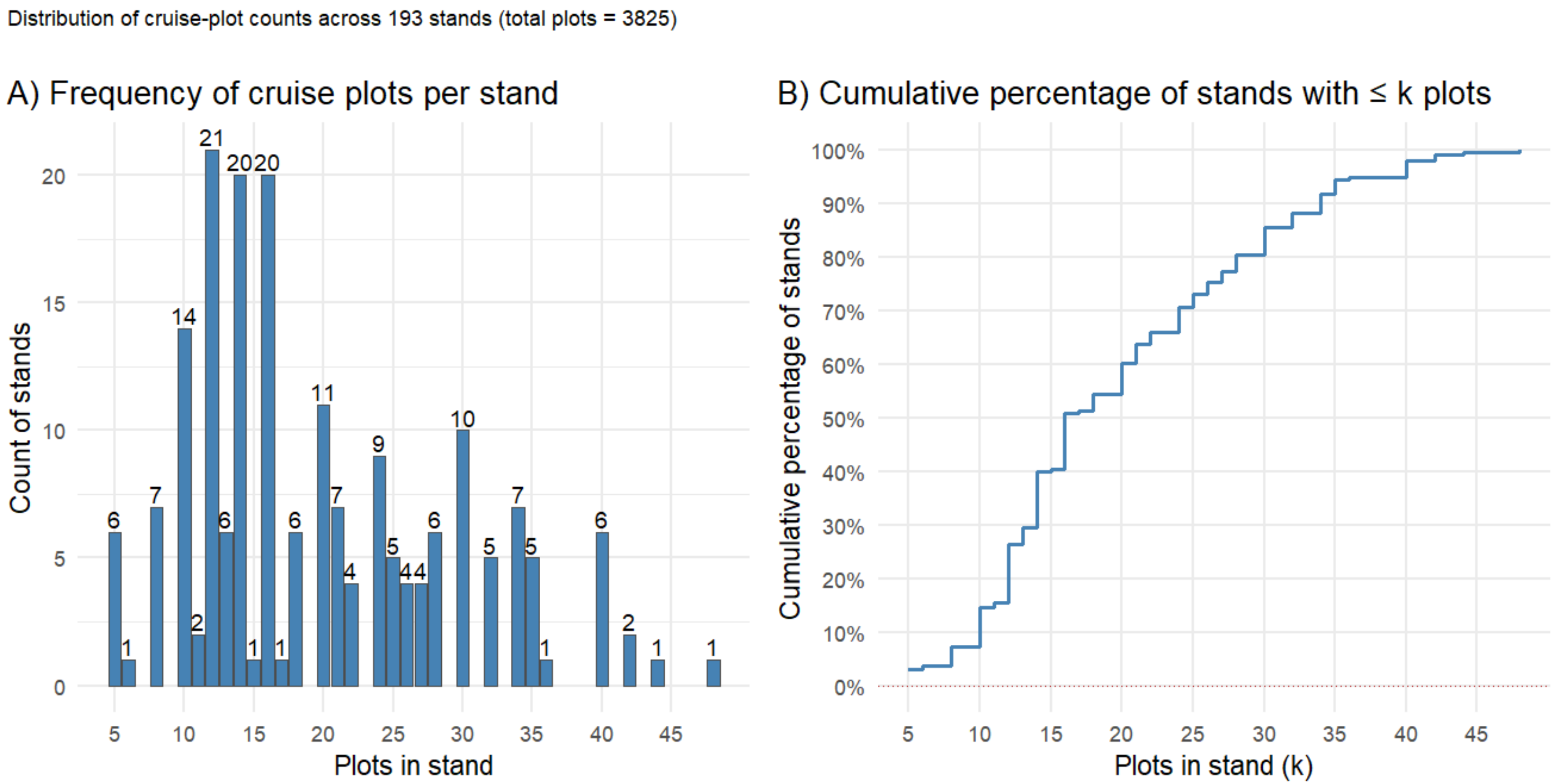
- If direct estimates have small errors compared to the unexplained variance of the fitted models $(\sigma_v^2 > \psi_d)$, $\gamma_d \approx 1$, i.e. $\hat{\delta}_d^{FH}$ rely more on the ground estimate.

- If direct estimates have high errors $(\sigma_v^2 < \psi_d)$, $\gamma_d \approx 0$, $\widehat{\delta}_d^{FH}$ puts more weight on the synthetic predictor, i.e. $\widehat{\delta}_d^{FH} \cong x_d\beta$.

- For unsampled stands or stands with only one variable radius plot, estimates are purely model based:

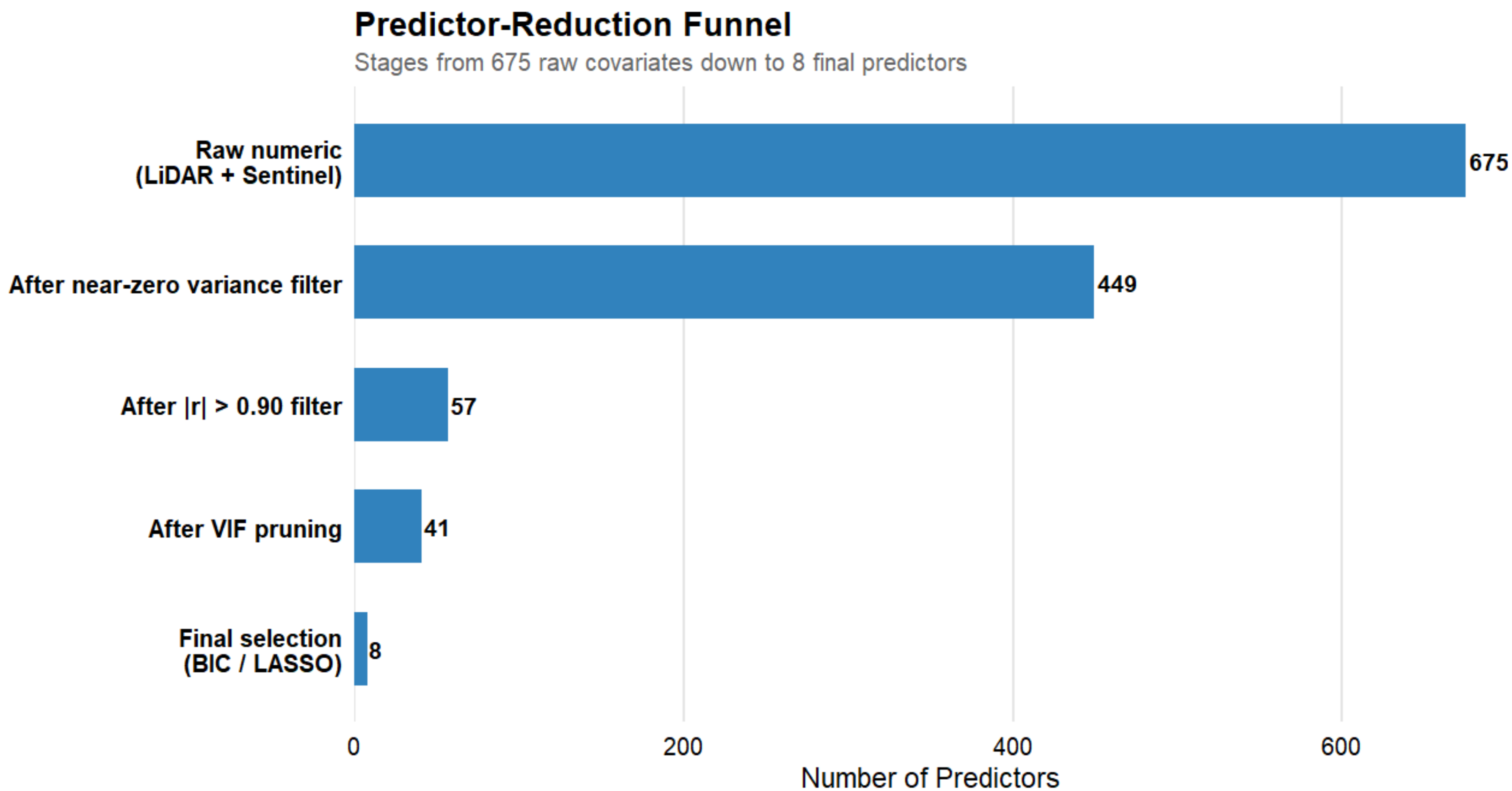
$$\widehat{\delta}_d^{FH} \cong x_d\beta$$

Methods



Distribution and Cumulative Frequency of Cruise Plot Counts for Stands Used in Direct Survey Estimates

Methods



Auto intersection: ht_max, ht_bar, ht_sd, tzq30, pzgt70, NDI45_sd, NDI711p001, GRN_mode
Best Subsets (BIC k= 8): ht_bar, ht_sd, tzq30, pzgt70, NDI45_sd, NDI72p001, NMDI_sd, GRN_mode

Methods

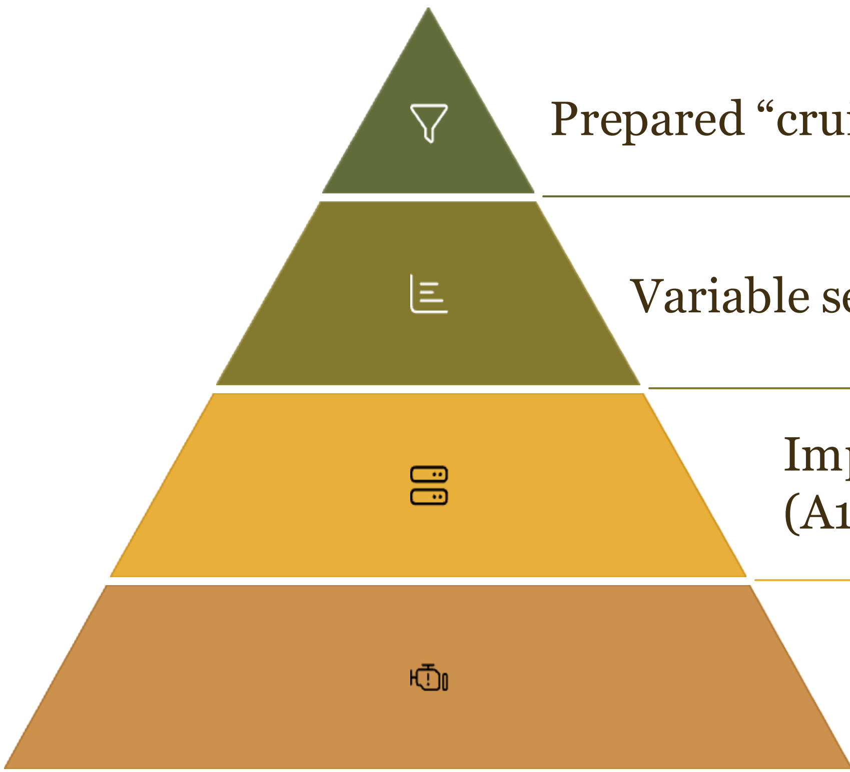
Unstratified Two-Stage Sampling

Scenario	Stage 1: Stand selection	Stage 2: Plot allocation
A1	SRS of stands	SRS of plots
A2	SRS of stands	PPS by stand area
A3	SRS of stands	PPV by stand volume
A4	PPS of stands (by area)	SRS of plots
A5	PPV of stands (by volume)	SRS of plots

Stratified Two-Stage Sampling

Scenario	Stage 1: Stand selection	Stage 2: Plot allocation
B1	Stratified SRS of stands	SRS of plots
B2	Stratified SRS of stands	PPS (by area)
B3	Stratified SRS of stands	PPV (by volume)
B4	Stratified PPS (by area)	SRS of plots
B5	Stratified PPV (by volume)	SRS of plots

Progress 2024-25



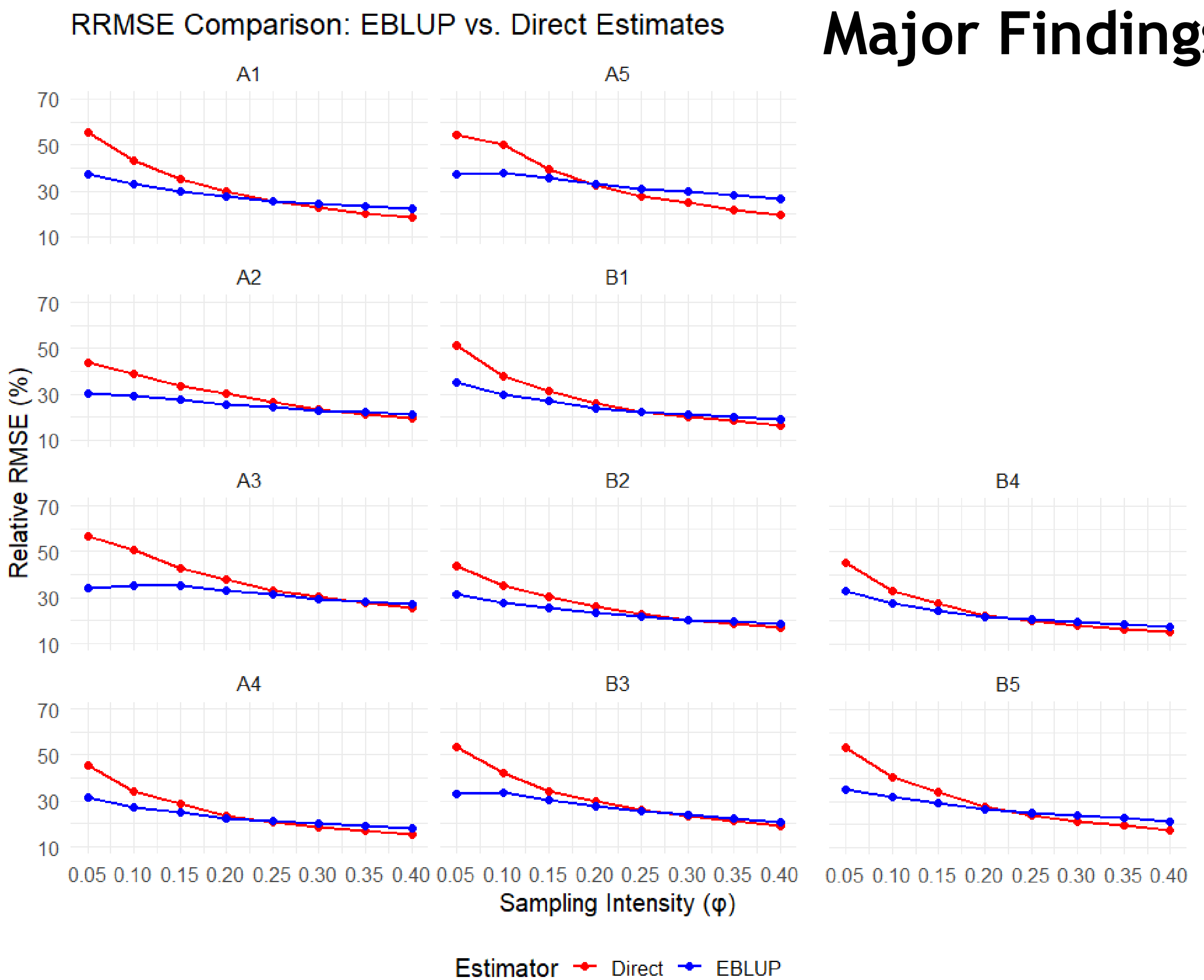
Prepared “cruise + covariate” stand-level dataset

Variable selection to fit Fay-Herriot on stand data

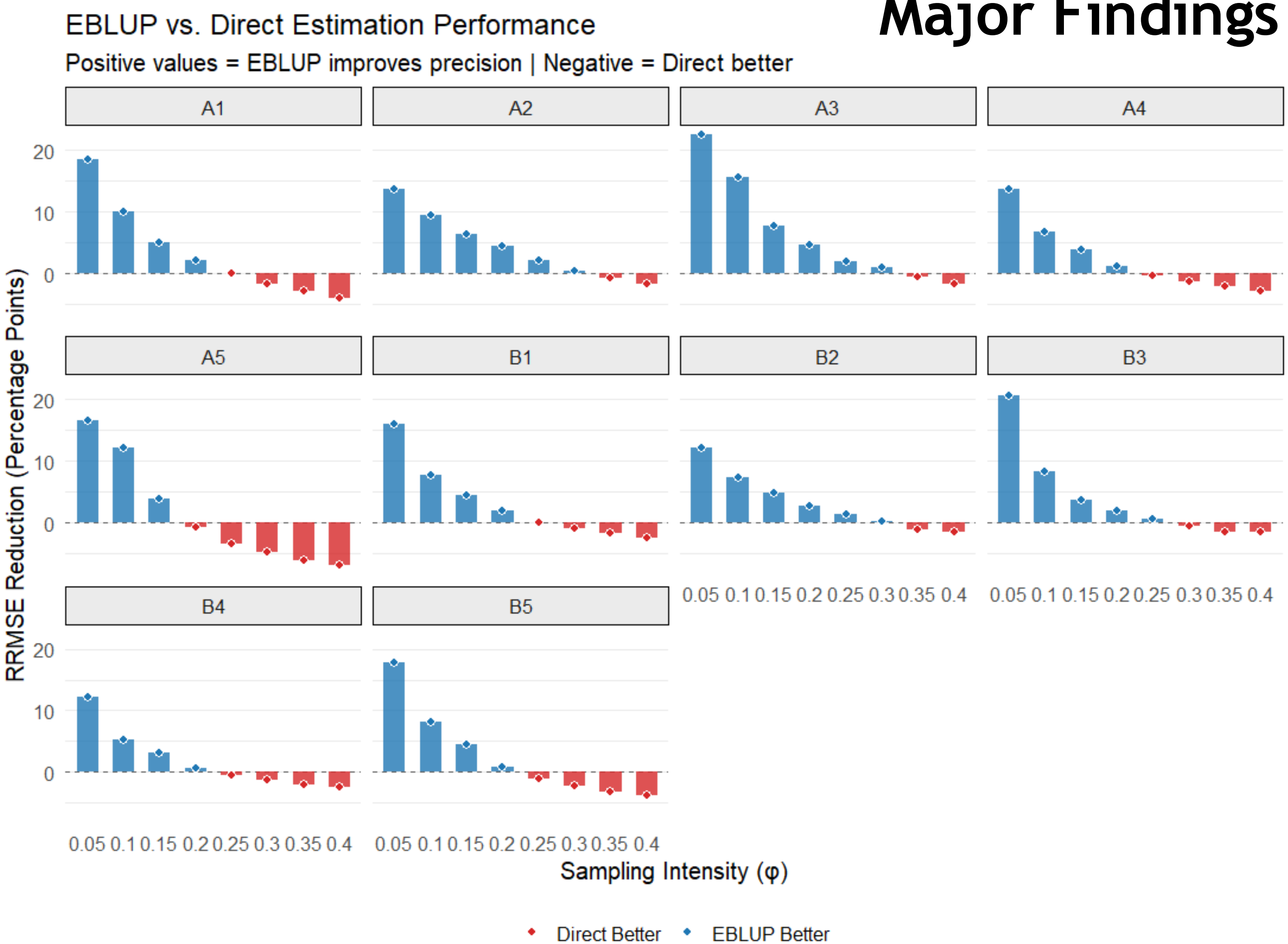
Implemented two-stage sampling strategies (A1–A5, B1–B5) + Simulations

Preliminary findings

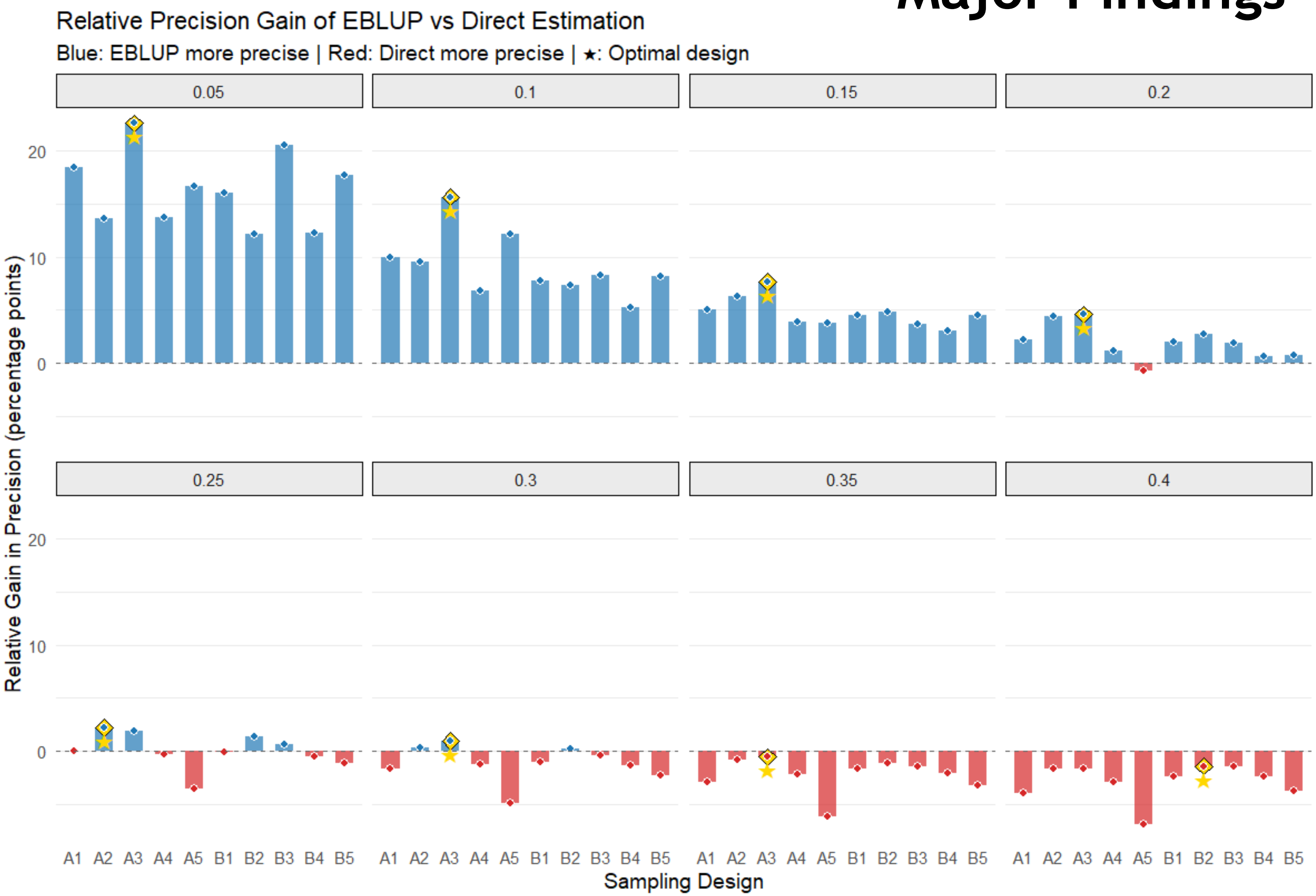
Major Findings



Major Findings



Major Findings



Deliverables

- Protocols for linking remote-sensing and ground data to improve small-area timber inventory estimates
- Quantify uncertainty of SAE predictions under different sampling intensities
- A PhD dissertation on the interplay
- Summary report for member companies
- Publications in peer-reviewed literature and presentations at several regional professional meetings

Company Benefits

Better understanding of linking data-sources

Combine spatially extensive + LiDAR/Sentinel metrics with targeted ground plots to improve the estimation of selected stand variables

Cost Savings

Borrowing strength + Optimized sampling reduces the number of expensive ground plots needed while maintaining accuracy

Efficiency Gain

The project demonstrates ways for incorporating SAE models into operational forest inventory in plantations

Operationalizing SAE Models: From Theoretical Foundations to Real-World Implementation

Summary

The overarching goal of this project is to advance SAE methods for forest inventory, with progress to date including:

- Compiled a broad candidate set of LiDAR and Sentinel covariates, evaluated different variable selection methods, and identified a final subset of covariates that improve prediction at the stand level.
- Determined the most efficient Fay-Harriot model specification for predicting total merchantable volume.
- Evaluated the performance of various sampling designs across multiple sampling intensities to determine whether an optimal design existed.