

AN ALTERNATIVE POST-STRATIFICATION SCHEME TO DECREASE VARIANCE OF FOREST
ATTRIBUTE ESTIMATES IN THE INTERIOR WEST

Miranda A. Rintoul

Sarah Maebius

Edwin Alvarado

Alexander Lloyd-Damnjanovic

Mai Toyohara

Kelly S. McConville

Gretchen G. Moisen

Tracey S. Frescino

Miranda A. Rintoul, Graduate Student, University of Arizona, Tucson, AZ 85721

Sarah Maebius, Research Assistant, Reed College, Portland, OR 97202

Edwin Alvarado, Research Assistant, Reed College, Portland, OR 97202

Alexander Lloyd-Damnjanovic, Research Assistant, Reed College, Portland, OR 97202

Mai Toyohara, Research Assistant, Reed College, Portland, OR 97202

Kelly S. McConville, Assistant Professor of Statistics, Reed College, Portland, OR 97202

Gretchen G. Moisen, Research Forester, USDA Forest Service, Rocky Mountain Research Station, Ogden, UT, 84401

Tracey S. Frescino, Forester, USDA Forest Service, Rocky Mountain Research Station, Ogden, UT, 84401

Abstract—The National Forest Inventory and Analysis (FIA) Program of the US Forest Service estimates many important forest attributes. Post-stratification with auxiliary data is used in the estimation process to improve precision. Current procedures used by the FIA unit in the Interior Western US involve stratifying by predicted forest and non-forest areas. This research aims to increase the efficiency of these post-stratified estimates by exploring alternative stratification schemes. We propose five candidate schemes based on auxiliary data from different sources and compare their relative efficiencies to both a simple random sampling estimator and to the current post-stratification. The best candidate, a four-strata scheme based on forest probability, is shown to improve the precision of forest attribute estimates over the current scheme across the Interior West region.

Keywords: FIA, post-stratification, forest probability, Interior West

INTRODUCTION

The National Forest Inventory and Analysis (FIA) Program of the USDA Forest Service is responsible for monitoring forest ecosystem attributes across the United States. The FIA program is important for forest land managers, policy makers, and scientists concerned with US forests because it is the sole source of consistent annual forest survey data across the entire country (McRoberts 2005).

The FIA uses a two-phase system for collecting forestry data used in the estimation of these attributes (Reams and others, 2005). The first phase involves using remotely sensed data to create large-scale maps that show nationwide landscape and forest patterns. Auxiliary data from a number of sources, including aerial photography, satellite imagery, and lidar, is used in this phase (McRoberts and others 2012). The second phase requires measuring data on sample plots with an intensity of approximately one plot per approximately 6,000 acres throughout the U.S. (Brand and others 2000). At this stage, the forest attributes of interest, usually related to the vegetation components of the forest, are measured on each field plot (McRoberts 2005).

For our work, we were interested in combining the data sources from Phase 1 and Phase 2. The Phase 1 data are wall-to-wall data products, which allow for a full enumeration of the desired landscape. We call these data pixel-level data since we can divide the entire landscape into pixels of equal and known size with the resolution based on the data product. In the Interior West, the Phase 2 data are collected in the field from forested plots. A subset of data variables are then collected from aerial photographs of nonforested plots. These two types of forest data can be used jointly to produce estimates of forest attributes through a method called post-stratification.

FIA is interested in producing estimates over many different types of geographic areas. Users are interested not only in estimates of forest attributes within States, but also within smaller geographic areas such as counties, groups of counties, national forests, ecological subsections, and congressional districts. For the analyses in this paper, we focus on counties as the reporting areas of interest, but any geographic area could be substituted.

POST-STRATIFICATION

To produce population estimates of forest attributes, FIA uses a statistical technique known as post-stratification (Holt and Smith 1979). Instead of estimating the mean of an attribute with a simple average of the attribute for the Phase 2 data, post-stratification adjusts the estimator based on a single categorical variable from Phase 1. The plots are grouped into categories, called strata, based on the Phase 1 variable and the mean is computed for each stratum. The strata sample means are summed up, weighted by the proportion of pixels in a landscape that fall in their respective stratum. The post-stratified estimator of the mean is given by

$$\bar{Y}_{PS} = \sum_h^H \frac{N_h}{N} \left(\frac{\sum_i^{n_h} y_{hi}}{n_h} \right) = \sum_h^H W_h \bar{Y}_h \quad (1)$$

where H is the number of strata (indexed by h), N is the total county area in pixels, N_h is the number of pixels within stratum h , n_h is the number of plots in stratum h (indexed by i), and y_{hi} is the value of variable Y for the i th plot in stratum h (Scott and others 2005). We can write $N^{-1}N_h$ as W_h , the weight for stratum h , which represents the portion of the landscape in stratum h . We can also write $n_h^{-1} \sum_i^{n_h} y_{hi}$ as \bar{Y}_h , the mean estimate of Y for stratum h . Therefore, the post-stratified estimator can be viewed as a weighted average of the means within each stratum where we are calibrating our estimator to the known size of each stratum. The post-stratified estimator is asymptotically unbiased and typically has a lower variance than a simple random sampling estimator, which is computed by taking the mean across all observations.

Specific groupings of observations into strata are called stratification schemes, and as with pre-stratification, should be chosen such that within-stratum observations are as similar as possible, with respect to the forest attributes, and between-stratum observations are as dissimilar as possible (McRoberts and others 2005). High homogeneity within strata increases the precision of the post-stratified estimator, while low homogeneity will result in a variance similar to or even worse than the variance of the simple random sampling estimator. Stratifying after the sample is selected instead of before does potentially lead to fewer gains in precision because the post-strata sample sizes are random. However, with FIA's permanent sample design, it is not possible to re-stratify with each remeasurement.

FOREST-NONFOREST STRATIFICATION SCHEME

The Rocky Mountain Research Station's FIA Program (RMRS-FIA) is responsible for maintaining an inventory of forest attributes in the Interior West (IW), which is comprised of Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, and Wyoming. Currently, the RMRS-FIA produces attribute estimates using a stratification scheme that divides field plots into forest or non-forest strata based on a forest probability map from Moderate Resolution Imaging Spectroradiometer (MODIS) satellite imagery (Blackard and others 2008). This scheme does not differentiate between different types of forest. All varieties of forest plots in the IW region, such as pinyon, pine, and aspen, are placed in the same stratum. Thus, this scheme might suffer from high variance as it fails to account for the ecological variability of all the plots within the forest stratum. In addition, the 250m spatial resolution plays a role in the ecological variability captured in each stratum (Nelson and others 2009).

Our goal is to examine other stratification schemes that could be used in the IW as an alternative to the forest/non-forest scheme. Because the post-stratified estimator is approximately unbiased, our objective was to improve the efficiency of the estimator by decreasing the variance. We propose a number of new schemes based on auxiliary pixel data that are available to FIA and compare the precision of these schemes to the current forest/non-forest framework.

DATA

We utilized both pixel level and plot level data in our analyses. The plot data included the most current 10-year measurement cycle for each Interior-West State, downloaded on February 6, 2019 from the FIA database, version FIADB_1.8.9.99 (last updated Dec 3, 2018). These data were collected between 2007 and 2017 and represents 86,085 sample plots. All States included a complete 10-year cycle of data except Wyoming, which had only 7 of 10 years of data collected at the time of download.

Forest Attributes

To evaluate the efficiency of the current and proposed stratifications schemes, we produced post-stratified estimates of four forest attributes: basal area (square-foot), trees per acre, above-ground biomass (pounds), and net volume (cubic-foot), which were extrapolated and summed to the plot-level using only live trees. These variables make up only a small fraction of the inventory composed by the FIA, but can be used to compare estimation efficiency across the different schemes because they are some of the most frequently analyzed variables and are highly correlated to many other forest attributes of interest.

Auxiliary Data

Our proposed schemes were created with three sources of auxiliary data: forest probability, biomass, and tree canopy cover. These variables are derived from pixel-level maps created from satellite imagery. The proposed stratification schemes were created based on observed patterns in the auxiliary data only.

Forest probability

The forest probability map, which is used to create the current forest/non-forest classification, is a measure of uncertainty in pixel assignments (Blackard and others 2008). The current scheme used by the RMRS-FIA has a forest threshold of 0.5, such that observations with 0.5–1.0 probability of forest are considered forest (Table 1). These data have a spatial resolution of 250-m.

Biomass

Forest biomass is defined to be live tree, above-ground biomass, measured in Mg/ha. Another MODIS-based data set developed in Blackard and others (2008) these data have a spatial resolution of 250-m, and attain a maximum value of 118 and a minimum value of 0 in the IW region.

Tree canopy cover

Tree canopy cover data are sourced from the National Land Cover Database (NLCD) canopy cover map (Yang and others 2018). The NLCD utilizes imagery data from the National Agriculture Imagery Program (NAIP), Landsat 5 Thematic Mapper (and its derivatives), as well as previous NLCD data. Percent tree canopy cover has a spatial resolution of 30-m.

New Schemes

To create categories to be used in the stratification schemes, the above continuous variables were split into a small number of bins. Biomass was split at cut points 5 and 32, forest probability was split at points 0.03, 0.23, and 0.77, and tree canopy cover was split at points 5, 25, and 50. These cut points were selected to ensure a roughly equal distribution of plots within each class with the intent of effectively representing IW forests. We used these binned variables and the current forest/non-forest scheme to create five proposal schemes, as shown in Table 1. Two of the schemes combine the current scheme with these additional binned variables. A more detailed description of each stratification scheme can be found in Appendix A.

Collapsing

The FIA calculates post-stratified estimates of forest attributes at the county level. Some counties contain very few plots belonging to a specific stratum, which can make post-stratified estimates less reliable. Recent work has suggested that the minimum within-strata sample size for a county should be 10 (Westfall and others 2011).

Because this minimum sample size is not met for all counties in the dataset for all of the proposed schemes, we devised a strategy for collapsing observations into a different stratum if too few of them belong to a specific stratum. The strategy we chose was as follows:

1. For a given county, find the stratum with the least representation (i.e. the fewest observations).
2. If there are fewer than 10 plots within this stratum, collapse those plots into the nearest stratum with at least 10 plots, with nearness measured in terms of the ordinal structure.
 - * If the nearest strata have fewer than 10 plots, the algorithm goes to the next closest stratum.
 - * If there are 2 nearest strata and both have at least 10 plots, the plots are collapsed into the stratum with a larger number of plots.
3. Repeat until all strata are represented by at least 10 plots (or zero plots).
4. Repeat for all counties in the dataset.

For example, suppose County A has 58 observations in Stratum One, 3 observations in Stratum Two, 7 observations in Stratum Three, and 19 observations in Stratum Four. Using this strategy, the 3 observations in Stratum Two would be collapsed into Stratum One. Next, the 7 observations in Stratum Three would be collapsed into Stratum Four. The final observation counts for County A would be 61 in Stratum One, 0 in Strata Two and Three, and 26 in Stratum Four. Note that for the current forest/nonforest stratification scheme, the two strata are simply collapsed when there are fewer than 10 plots.

METHODS

Our goal is to propose more efficient alternatives to the forest/non-forest stratification scheme currently used by the RMRS-FIA. This requires comparing the performance of the variance estimator of the post-stratified mean estimator across the schemes.

Post-Stratified Mean and Variance

The post-stratified mean estimator of variable Y is given by equation (1) and its estimated variance is given by

$$v(\bar{Y}_{PS}) = \frac{1}{n} \left[\sum_h^H W_h n_h v(\bar{Y}_h) + \sum_h^H (1 - W_h) \frac{n_h}{n} v(\bar{Y}_h) \right]$$

where $v(\bar{Y}_h)$ is the variance of Y for stratum h , and n is the total number of sampled plots (Scott and others 2005).

Scheme Evaluation

The efficiency of an estimator can be evaluated using relative efficiency, which is computed as

$$RE_{srs} = \frac{v(\bar{Y}_{srs})}{v(\bar{Y}_{ps})} \quad (2)$$

where $\bar{Y}_{srs} = n^{-1} \sum_i^n y_i$, the mean estimator using simple random sampling and no stratification and its corresponding variance estimator is $v(\bar{Y}_{srs}) = n^{-1} v(\bar{Y})$ where $v(\bar{Y})$ is the variance of Y in the sample (McRoberts and others 2005). Equation (2) is used to perform initial comparisons between the candidate schemes. However, because our task is to compare multiple stratification schemes, we additionally computed the relative efficiency of the best candidate scheme against the current forest/non-forest scheme, as given by

$$RE_{cur} = \frac{v(\bar{Y}_{cur})}{v(\bar{Y}_{new})}.$$

Here \bar{Y}_{cur} refers to the post-stratified mean estimator under the current scheme, and \bar{Y}_{new} refers to the post-stratified mean estimator under a proposed scheme. The desired result is $RE_{cur} > 1$, which indicates that the proposed scheme is effectively lowering the variance of the mean estimator. Because FIA estimation is often done at the county level, we calculated a relative efficiency measure for each county.

Bootstrapping

The REs are actually estimated REs because we are comparing the ratio of the estimated variances, not the true variances. To more rigorously evaluate the effectiveness of our best proposed scheme, we used the statistical technique bootstrapping on the RE measure (Efron and Tibshirani 1986) to obtain 95 percent confidence intervals for the true RE_{cur} . The bootstrapping process for a county-level analysis is as follows.

1. Draw a sample of n plots from a county with replacement, where n is the original number of plots in the county.
2. Compute post-stratified variance estimates (for both the current and proposed scheme) and RE_{cur} for the sample.
3. Repeat steps 1 and 2 1000 times to produce 1000 different measures of RE_{cur} .
4. Repeat steps 1-3 for each county.

After completing this process, we extracted the middle 95 percent of the bootstrapped RE_{cur} measures for each county to create a 95 percent confidence interval of the efficiency of the candidate scheme. More work should be done to explore the accuracy of the bootstrapped confidence intervals over varying sample sizes.

RESULTS

Initial Comparisons

To perform an initial evaluation of our proposed schemes, we calculated the relative efficiency of each candidate against a simple random sampling approach for each of our four forest attribute variables, using equation (2), for each county. With these preliminary explorations, we didn't conduct a bootstrap analysis, but instead we computed the means of these REs across all counties for each forest attribute. As seen in Table 2, all of the schemes are more efficient, on average, than the simple random sampling estimator as expected. The *Tree Canopy Cover* scheme was the most efficient, followed by the *Forest Probability* scheme, on average.

However, visual analysis of maps such as Figure 1 showed that while *Tree Canopy Cover* had the highest mean RE_{srs} , *Forest Probability* had $RE_{srs} > 1$ for many more counties than *Tree Canopy Cover*. This implies that *Forest Probability* is more consistently efficient across the entire IW, while *Tree Canopy Cover* is extremely efficient in certain regions, but inefficient in others. Therefore, we chose *Forest Probability* as the focus of our subsequent analysis, comparing it to the current stratification scheme.

Forest Probability Scheme

The *Forest Probability* scheme is a four-strata scheme based on Blackard and others's (2008) forest probability map, which is also used as a basis for the RMSR-FIA's current stratification scheme. We performed a bootstrap experiment on the RE of the *Forest Probability* scheme against the current scheme to more rigorously assess whether the candidate scheme improves the precision of the estimates. We found that even for the lower bound on the bootstrapped 95 percent confidence interval, the *Forest Probability* scheme had $RE_{cur} > 1$ for the majority of IW counties and land area. This finding was true for all four of the studied forest attributes. Figure 2 shows where the $RE_{cur} > 1$ occurs for estimation of aboveground biomass, and for the upper and lower bounds of the 95percent confidence interval. Graphs of the confidence bounds of the other response variables follow a similar pattern. Since the lower bound of the CI is above 1 for the majority of the counties, this provides evidence that the true RE is greater than 1 and that the new scheme is an improvement over the old scheme. It should be noted that there are a few counties where the upper bound of the CI is below 1, signifying that the new scheme is worse in those regions. Similar graphical results for all four attributes can be found in Appendix B.

The mean values for the upper and lower bounds on the RE_{cur} confidence intervals (across all counties), as well as the proportion of counties for which $RE_{cur} > 1$, can be found in Table 3. The *Forest Probability* scheme has a statistically significant improvement in precision for 84.0–85.2 percent of IW counties, and has almost no improvement for only 5.9–6.2 percent of counties. On average, the use of this scheme in place of the current scheme improves the precision of attribute estimates by at least a factor of 1.71–1.92, and at most by a factor of 3.77–4.12.

CONCLUSIONS

Of our five proposed schemes, the *Forest Probability* scheme was the most effective at consistently reducing the variance of the post-stratified estimates in the IW. *Forest Probability* appears to improve the precision of estimates considerably when compared to the current forest-nonforest scheme used by the RMRS-FIA. This candidate scheme can be thought of as a refinement to the current scheme, because both are based on Blackard and others' forest probability map. While it might be expected that the schemes that offered finer spatial resolution and information contemporary with the FIA plot data would result in dramatic improvements in precision, this was not the case. This could be explained by the tight consistency in forest land definition between the Blackard and others product and FIA field observations, as well as by the possibility that the 250m spatial resolution of the forest probability map was more compatible with the size of FIA plots, plus positional uncertainty. Showing improvement by simply allowing more bins to be derived from the map currently used as the basis for stratification inspires further investigation into new and improved spatial data layers for stratification in the IW in the future

ACKNOWLEDGMENTS

We would like to thank Mark Nelson of the USDA Forest Service Northern Research Station and Paul Patterson of the USDA Forest Service Rocky Mountain Research Station for their very careful and thoughtful review of the manuscript. This work was conducted as part of the educational experience provided by the Forestry Data Science Lab at Reed College. The lab offers undergraduate students real-world statistical work experience, and brings diverse student perspectives to statistical and analytical challenges facing forest inventory. This work was supported by the USFS Forest Inventory and Analysis Program, Reed College, and Swarthmore College. The authors would also like to thank the USFS Forest Inventory and Analysis Program as well as the USFS Remote Sensing Applications Center for the data.

LITERATURE CITED

- Blackard, J.; Finco, M.; Helmer, E.; Holden, G.; Hoppus, M.; Jacobs, D.; Lister, A.; Moisen, G.G.; Nelson, M.; Rieman, R. 2008. Mapping U.S. forest biomass using nationwide forest inventory data and moderate resolution information. *Remote Sensing of Environment*. 112(4): 1658–77. <https://doi.org/10.1016/j.rse.2007.08.021>.
- Brand, G.J.; Nelson, M.D.; Wendt, D.G.; Nimerfro, K.K. 2000. The hexagon/panel system for selecting FIA plots under an annual forest inventory. In McRoberts, R.E.; Reams, G.A.; Van Deusen, P.C., eds. *Proceedings of the first annual forest inventory and analysis symposium*. Gen. Tech. Rep. NC- 213. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station: 8-13.
- Efron, B., Tibshirani, R. 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*. 1(1): 54-77. <https://doi.org/10.1214/ss/1177013815>.
- Holt, D.; Smith, T.F. 1979. Post stratification. *Journal of the Royal Statistical Society: Series A (General)*. 142(1): 33-46.
- McRoberts, R.; Gobakken, T.; Næsset, E. 2012. Post-stratified estimation of forest area and growing stock volume using lidar-based stratifications. *Remote Sensing of Environment*. 125: 157–66.
- McRoberts, R.E.; Holden, G.R.; Nelson, M.D.; Liknes, G.C.; Gormanson, D.D. 2005. Using satellite imagery as ancillary data for increasing the precision of estimates for the Forest Inventory and Analysis Program of the USDA Forest Service. *Canadian Journal of Forest Research*. 35(12): 2968–80. <https://doi.org/10.1139/x05-222>.
- McRoberts, R.E. 2005. The enhanced Forest Inventory and Analysis Program. In Bechtold, W.A. and P.L. Patterson eds. *The enhanced Forest Inventory and Analysis Program-National sampling design and estimation procedures*. Gen. Tech. Rep. SRS-80. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station: 1-10
- Nelson, M.D.; McRoberts, R.E.; Holden, G.R.; Bauer, M.E. 2009. Effects of satellite image spatial aggregation and resolution on estimates of forest land area. *International Journal of Remote Sensing*. 30(8): 1913-1940.
- Reams, G.A.; Smith, W.D.; Hansen, M.H.; Bechtold, W.A.; Roesch, F.A.; Moisen, G.G. 2005. The forest inventory and analysis sampling frame. In Bechtold, W. A. and P. L. Patterson eds. *The enhanced Forest Inventory and Analysis Program-national sampling design and estimation procedures*. Gen. Tech. Rep. SRS-80. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station: 11-26.
- Scott, C. T.; Bechtold, W.A.; Reams, G.A.; Smith, W.D.; Westfall, J. A.; Hansen, M.H.; Moisen, G.G. 2005. Sample-based estimators used by the Forest Inventory and Analysis national information management system. In Bechtold, W.A. and P.L. Patterson eds. *The enhanced Forest Inventory and Analysis Program-National sampling design and estimation procedures*. Gen. Tech. Rep. SRS-80. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station: 43-67.
- Westfall, J.A.; Patterson, P.L. Coulston, J.W. 2011. Post-stratified estimation: within-strata and total sample size recommendations. *Canadian Journal of Forest Research*. 41(5): 1130–9. <https://doi.org/10.1139/x11-031>.
- Yang, L.; Jin, S.; Danielson, P.; Homer, C.; Gass, L.; Bender, S.M.; Case, A.; Costello, C.; Dewitz, J.; Fry, J.; Funk, M.; Granneman, B.; Liknes, G.C.; Rigge, M.; Xian, G. A new generation of the United States National Land Cover Database: Requirements, research priorities, design, and implementation strategies. *ISPRS Journal of Photogrammetry and Remote Sensing*. (146) 108-123.

Table 1—Descriptions of the five candidate stratification schemes. The mapping of the numerical auxiliary variables to the specific categories can be found in Appendix A.

Scheme name	No. of strata	Categories
Current scheme	2	Nonforest, Forest

Scheme name	No. of strata	Categories
Biomass	3	Low Biomass, Moderate Biomass, High Biomass
Forest probability	4	Non forest, Low Forest, Moderate Forest, High Forest
Tree canopy cover	4	No Coverage, Low Coverage, Moderate Coverage, High Coverage
Forest/Biomass	4	Nonforest, Low Biomass Forest, Moderate Biomass Forest, High Biomass Forest
Forest/Tree canopy cover	5	Nonforest, No Coverage Forest, Low Coverage Forest, Moderate Coverage Forest, High Coverage Forest

Table 2—Mean relative efficiencies compared to the simple random sampling estimator, within IW-FIA counties. The numbers of counties with $RE_{srs} > 1$ is given in the last row. For each scheme, the numbers of counties with $RE_{srs} > 1$ was the same for the four response variables.

Scheme	Current scheme	Tree canopy cover	Biomass	Forest probability	Forest/tree canopy cover	Forest/biomass
AG biomass	3.89	10.00	4.34	8.20	7.49	4.91
Trees per acre	3.44	7.04	3.73	6.61	5.14	4.14
Basal area	3.74	8.92	3.82	8.17	6.35	4.47
Cubic volume	3.86	9.62	4.82	8.09	7.41	4.85
# RE > 1	197	209	187	242	180	189

Table 3—Forest probability relative efficiencies compared to the current IW stratification scheme. The mean values for the upper and lower bounds on the RE_{cur} confidence intervals (across all counties), as well as the proportion of counties (and number of counties) for which $RE_{cur} > 1$.

Scheme	% Lower > 1 (#)	Lower mean	% Upper > 1 (#)	Upper mean
AG biomass	84.4 (216)	1.92	93.8 (240)	4.04
Trees per acre	84.0 (215)	1.71	94.1 (241)	3.77
Basal area	85.2 (218)	1.82	94.1 (241)	4.12
Cubic volume	84.0 (215)	1.89	93.8 (240)	4.01

Appendix A—Scheme Descriptions

This section contains the proposed stratification schemes.

Current scheme (2 strata)

- Nonforest: observations in $[0, 50]$ for the forest probability auxiliary variable
- Forest: observations in $(.50, 100]$ for the forest probability auxiliary variable

Biomass (3 strata)

- Low biomass: observations in [0, 5] for the biomass auxiliary variable
- Moderate biomass: observations in (5, 32] for the biomass auxiliary variable
- High biomass: observations in (32, 118) for the biomass auxiliary variable

Forest probability (4 strata)

- No forest: observations in [0, 0.03] for the forest probability auxiliary variable
- Low forest: observations in (0.03, 0.23] for the forest probability auxiliary variable
- Moderate forest: observations in (0.23, 0.77] for the forest probability auxiliary variable
- High forest: observations in (0.77, 1] for the forest probability auxiliary variable

Tree canopy cover (4 strata)

- No coverage: observations in [0, 5] for the tree canopy cover auxiliary variable
- Low coverage: observations in (5, 25] for the tree canopy cover auxiliary variable
- Moderate coverage: observations in (25, 50] for the tree canopy cover auxiliary variable
- High coverage: observations in (50, 100] for the tree canopy cover auxiliary variable

Current strata x biomass (4 strata)

- Low biomass forest: observations in [0, 5] for the biomass auxiliary variable and the IW forest stratum
- Moderate biomass forest: observations in (5, 32] for the biomass auxiliary variable and the IW forest stratum
- High biomass forest: observations in (32, 118) for the biomass auxiliary variable and the IW forest stratum
- Non-forest: observations in the IW non-forest stratum

Current strata x tree canopy cover (5 strata)

- No coverage forest: observations in [0, 5] for the tree canopy cover auxiliary variable and the IW forest stratum
- Low coverage forest: observations in (5, 25] for the tree canopy cover auxiliary variable and the IW forest stratum
- Moderate coverage forest: observations in (25, 50] for the tree canopy cover auxiliary variable and the IW forest stratum
- High coverage forest: observations in (50, 100] for the tree canopy cover auxiliary variable and the IW forest stratum
- Non-forest: observations in the IW non-forest stratum

Appendix B—Additional Visualizations of the Relative Efficiencies

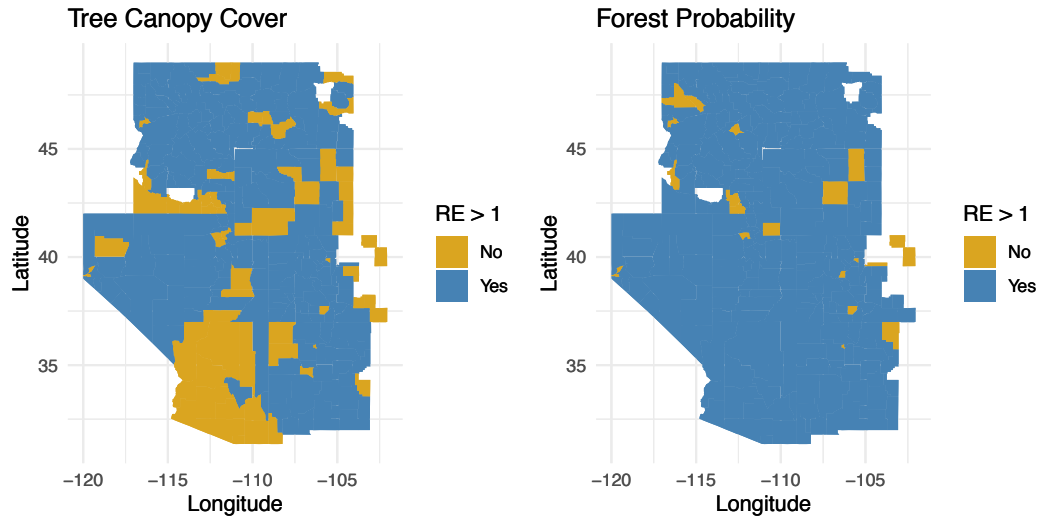


Figure 1—Map of IW counties where $RE > 1$ for aboveground biomass for Tree Canopy Cover and Forest Probability schemes.

Alternate text: color-coded map as described in text

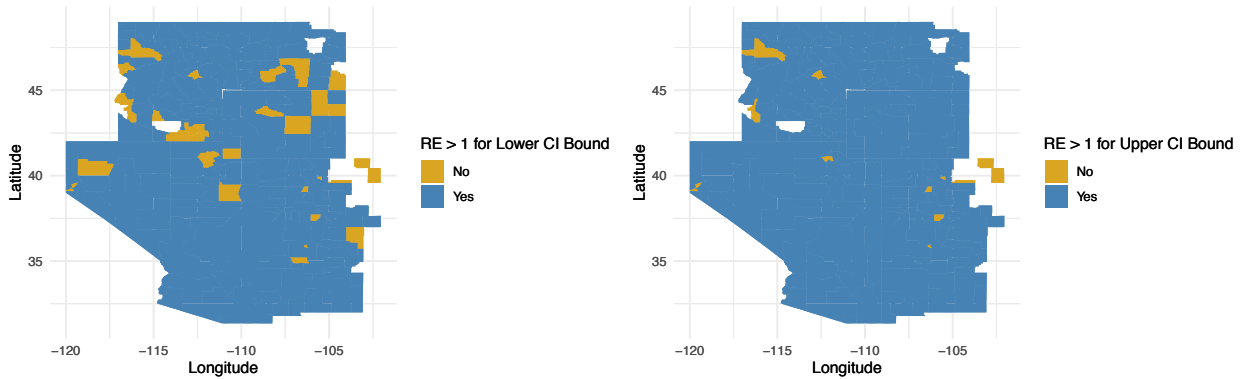


Figure 2—Map of IW counties where the Forest Probability stratification scheme outperforms the current scheme for aboveground biomass estimation, at lower and upper bounds of a 95 percent confidence interval.

Alternate text: color-coded map as described in text.