A Comparison of Forest Estimation Techniques in the US Interior West

Abstract

The National Forest Inventory and Analysis (FIA) Program of the United States Forest Service collects and analyzes data on many important forest attributes. The current FIA procedure is to utilize Post-Stratification (PS) estimation to improve estimation precision. The Interior West (a region of the United States) stratifies estimates by forest and non-forest areas. This research compares the use of various techniques in estimating forest attributes in the Interior West. We consider and compare the Horvitz-Thompson (HT), PS, and generalized regression (GREG) estimators, and introduce and compare a new, generalized regression estimator on resolution of Y (GREGORY) estimator. Comparing the relative efficiencies of bootstrap standard errors, we find that the more complex estimators, GREG and GREGORY, generally improve the precision of estimates. Additionally, we created a dashboard to allow for easy comparison of these estimators.

Keywords: Forestry, Generalized Regression Estimator, Post-Stratification Estimator, Forest Attributes, Geospatial Data, Interior West, Dashboard

---

1 Introduction

The National Forest Inventory and Analysis (FIA) Program of the USDA Forest Service has been in continuous operation since 1930, following the passage of the McSweeney-McNary Forest Research Act of 1928. Once established, the FIA became the primary source of conducting and continuously updating a comprehensive inventory and analysis of the present and prospective conditions of the renewable resources of the United States ("Forest Inventory and Analysis National Program - About Us" n.d.).

As the FIA creates and maintains a broad inventory of resources, data is available on numerous forest attributes ranging from merchantable timber and other wood products, risks associated with fire, fuels and potential fire hazards, conditions of wildlife habitats, insects or diseases, biomass, carbon storage, forest health, and other general characteristics of forest ecosystems.

In creating and maintaining a broad-scale resource inventory, the FIA is responsible for monitoring forest ecosystem attributes across the United States. Given the diversity of ecologies composing the United States, groups of states are divided into unique geographic divisions, as noted by the United States Census Bureau. One of these divisions, and the division of interest for our research, is the Interior West (IW), a region that covers the states of Arizona (AZ), Colorado (CO), Idaho (ID), Montana (MT), Nevada (NV), New Mexico (NW), Utah (UT), and Wyoming (WY) ("Interior West Forest Inventory & Analysis - About Us" n.d.).

The FIA program, both in the Interior West and broadly, is important as it is the sole source of consistent annual forest survey data across the entire country. The FIA provides objective and scientifically credible information on how much forest there is, what it looks like, whether the forest area is increasing or decreasing, whether we are gaining or losing species, how quickly trees are growing, dying, and being harvested, and how forest ecosystems change over time. In sum, the FIA collects information to tell us about the current and prospective state of our environment. Thus, it is important that the estimates of different forest attributes are accurate and comprehensive.

There are two primary sources of FIA data: plot-level and pixel-level. Plot-level data, also known as field data, comes from field plots distributed across each state, at a sample intensity of about one plot every 6,000 acres (Bechtold and Patterson 2015). Most field data is related to tree and understory vegetation components of a forest. By comparison, pixel-level data comes from remote-sensing sources, such as satellite imagery, whereby data is gathered on aspects of the environment such as latitude, longitude, and elevation. Compared to plot-level data, pixel-level data has greater sampling intensity; much more of the Interior West is being sampled in the pixel data.

We combine the two sources of FIA data—pixel and plot-level data—to estimate forest attributes, utilizing two survey estimation methods. However, we are not the first to utilize both plot-level and pixel-level data in an attempt to improve estimation. Survey estimation is a well-developed field, and has led to the creation of numerous survey estimation packages to streamline estimation techniques—of note is the *'mase'* package for R (McConville et al. 2018). Many survey estimation statistical packages implement advanced survey estimation techniques such LASSO, elastic net, and ridge regression, techniques we do not use in this research but may be considered in future research.

Though progress has been made in the literature on the application and utilization of advanced estimation techniques such as penalized regression, FIA still relies primarily on Post-Stratification (PS) for producing its estimates. As FIA still relies on PS, we wanted to explore and compare the use of another estimator which can allow for the inclusion of additional variables and potentially produce a more accurate estimate of forest attributes. Estimation using the generalized regression (GREG) (McConville, Moisen, and Frescino 2020) allows for exactly this, but running regressions at the estimation level can often be an area with insufficient plot data (e.g. a small county), potentially creating problematic estimates. We therefore introduce generalized regression estimator on resolution of Y (GREGORY), an indirect GREG, to attempt improving estimation.

2 Methods

2.1 Objective

Our goal is to assess three estimators, Horvitz-Thompson (HT), generalized regression (GREG), and generalized regression on resolution of Y (GREGORY), and their respective variances compared to the PS estimator, the estimator currently utilized by the Interior West (IW) Forest Inventory Analysis (FIA).

2.2 Horvitz-Thompson

$$\mu_{y,HT} = \bar{y}$$

The HT estimator is the simplest estimator that this paper will investigate: it is simply the average variable of interest for plots in a particular estimation unit.

2.3 Post-Stratification

PS is the estimator currently used by FIA to report forest characteristics on a state level. The principal advantage of the PS estimator over the HT estimator is that it allows for the inclusion of auxiliary data; specifically, a single categorical pixel-level variable. In our case, FIA has a variable in which it labels the entire United States on a pixel level as forest or non-forest. The formula for this estimator is a simpler, single-variable version of the GREG estimator covered in the following section.

2.4 GREG

GREG allows the inclusion of additional variables in order to help predictive power. An equation for a GREG estimate within a particular estimation unit(e.g. a county) is:

$$\mu_{y,GREG} = \bar{y} - \bar{x}_n^T\beta + \bar{x}_N^T\beta$$

The β is obtained from either a simple or multiple linear regression of the plot auxiliary variables on the plot variables of interest, where $\bar{y}$ represents the average value of the variable of interest within the plots of the estimation unit, $\bar{x}_n$ is a vector containing the average auxiliary variables of those plots, and $\bar{x}_N$ is a vector that contains the average auxiliary variable values for the whole estimation unit on a pixel level.

To get an estimation unit's PS estimate, the auxiliary variable vectors $\bar{x}$ are simply a single categorical variable.

2.5 GREGORY

GREGORY allows separation of the "resolution" and "estimation" units; the regressions are now run at the resolution level in order to predict values for an estimation unit. This results in a potentially indirect estimator, allowing for more information to be included for estimation units which may be lacking in plot level data, as well as allowing for a less arbitrary division of plot data (for example, using ecological province rather than the political boundaries of counties).

$$\mu_{y,GREGORY} = \bar{y} - \sum_{l \in p} \left[w_l(\bar{x}_n^T \beta_l) + w_l(\bar{x}_N^T \beta_l) \right]$$

The above equation includes resolution units l in the set of resolution units p, along with weights w which represent the proportion of the estimation unit of interest within a certain resolution unit. Each β_l is now run at the resolution level, hence the subscript.

2.6 Weights

GREG can also be written as a sum of weights:

$$\mu_{k,GREG} = \frac{1}{n} \sum_{i \in s} \kappa_i y_i$$

n being the total number of plots, y_i being the value of the variable of interest in plot i within sample s. κ_i is the weight for a particular plot.

$$\kappa_i = 1 + n \left(\bar{x}_N - \bar{x}_n\right)^T \left(\sum_{g \in s} \frac{x_g x_g^T}{{\sigma_g}^2} \right)^{-1} x_i$$

Here, g and i are plots within sample s.

GREGORY can also be written in this method. We will represent the GREGORY weight for a county with γ_i instead of GREG's κ_i.

$$\gamma_i = \mathbb{I}\left(i \ \in \ k\right) + n_k \left(\bar{x}_{N,k} - \bar{x}_{n,k}\right)^T \sum_{l \in p} w_l \left[\sum_{g \in s_l} x_g x_g^T \right]^{-1} x_i \mathbb{I}\left(i \ \in \ l\right)$$

k once again represents the estimation unit, and l represents resolution units in set p. $\mathbb{I}$ is an indicator variable.

2.7 Comparisons

We assessed model accuracy using a bootstrap variance estimator(McConville, Moisen, and Frescino 2020), later transformed into standard error, which mimics the sampling variability

by taking a bootstrap sample of the data. The steps of our bootstrap procedure are:

1. Take a simple random sample with replacement of size n from the original sample, called the bootstrap sample.
2. Compute the estimator on the bootstrap sample.
3. Repeat steps 1 and 2 N times. In this instance, we performed steps 1 and 2 1,000 times.

To evaluate our estimators, we compared relative efficiencies of county-level estimates. These relative efficiencies were calculated using bootstrap standard errors.

3 Data Description

3.1 Data Sources/Data Collection

Our dataset is comprised of two different kinds of data: plot and pixel. Plot level data was collected by the FIA in a quasi-systematic sample of ground plots over a 10 year period, with a base sampling intensity of one plot per every 6,000 acres (Bechtold and Patterson 2015). The plots are based on a systematically sampled hexagonal projection over the US, specifically Phase 2 plot data. These plots were measured in person for a selection of variables. We also have access to auxiliary data from satellite imagery based on 30x30 meter "pixels" which we use to appropriately weigh our estimators.

3.2 Key Assumptions

Though our data used was collected from 2004 to 2013, we do not take into account differences by year. For our analysis, we treat all years as part of a single cycle. Likewise, though plot-level data was collected during different seasons, we do not take into account any seasonal differences.

Furthermore, for ease of analysis, we assume minimal temporal (spatial) autocorrelation. However, this assumption is explored in greater detail by other researchers. We elaborate on this further in our Discussion, but for the time being we recommend research such as that of Magnussen and Fehrmann, who explore variance estimators under different levels and structure of autocorrelation (Magnussen and Fehrmann 2019).

Additionally, as plot-level data was collected by individuals working for the FIA, we also expect some level human bias in data collection. However, we assume the impact this bias is negligible on our estimates.

3.3 Key Variables

Estimation Variables

Our focus is on estimating average value of four forest attributes by county: basal area (square-foot), trees per acre, above-ground biomass (pounds), and net volume (cubic-foot), to evaluate the relative efficiency of the GREG and GREGORY estimators to the current PS method utilized by FIA. These variables are extrapolated and summed to the plot-level using only live trees. These forest attributes were chosen because they are frequently analyzed, and highly correlated to other forest attributes of interest. Therefore, estimates of one variable are indicative to the estimates of the other variables.

Predictors

To create our GREG and GREGORY models, we chose three variables which we found tended to be most effective when compared agaisnt other variables we had at our disposal. These predictors are further detailed in McConville et al. 2020 and other related forestry literature (McConville, Moisen, and Frescino 2020).

- **Forest probability:** An pixel-data based estimate of the probability of forest within a pixel (Blackard et al. 2008).
- **Biomass:** A pixel-data based estimate of above ground biomass (Blackard et al. 2008).
- **Tree canopy cover:** An estimate of percent of plot covered in tree canopy, using spatial data, spatial resolution of 30 m (Homer et al. 2015).

Estimation/Resolution Levels

Our estimation units were counties, and in our GREGORY our reslution units were ecological provinces (McNab et al. 2007).

3.4 Data Preparation

Prior to creating any estimates, we omitted missing values. Out of the 65860 intial observations in the dataset, 103 contained variables with missing values. However, 99.94% of our original data remain.

Ecology Example:

Each ecological code can be broken down into section and province, using a subset of the ecological code. For example, for an Alpine Meadow Province, we have:

- **Subsection**, most granular: M332Ba: Bitterroot Glaciated Canyons Subsection
- **Section**: M332B: Northern Rockies and Bitterroot Valley Section
- **Province**, least granular: M332: Middle Rocky Mountain Steppe – Coniferous Forest - Alpine Meadow Province

4 Results

Our objective was to investigate the four estimators of interest and assess their validity. Using bootstrap standard error as a measure of validity, we found that SE was minimized in the most counties using the GREGORY estimator. Simpler estimators tended to have lower standard errors for counties which contained less trees and tree-associated variables overall. A more in depth look at how the estimators compare can be found within our dashboard: https://shiny.reed.edu/s/users/wojciko/ests/

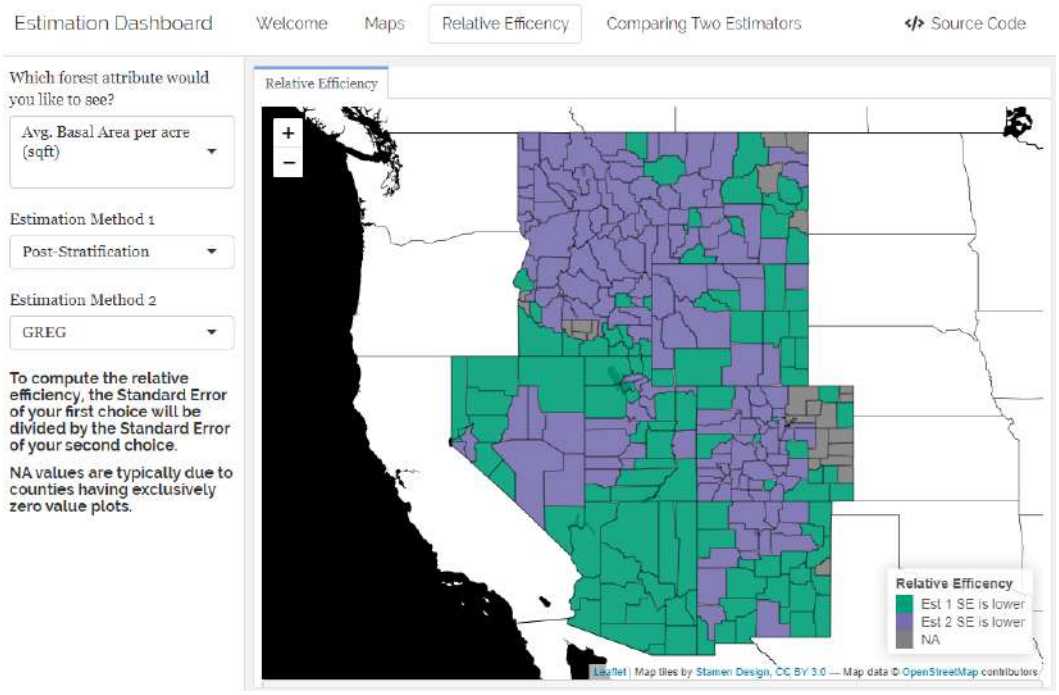

Figure 1: Comparing Post Strat to GREG

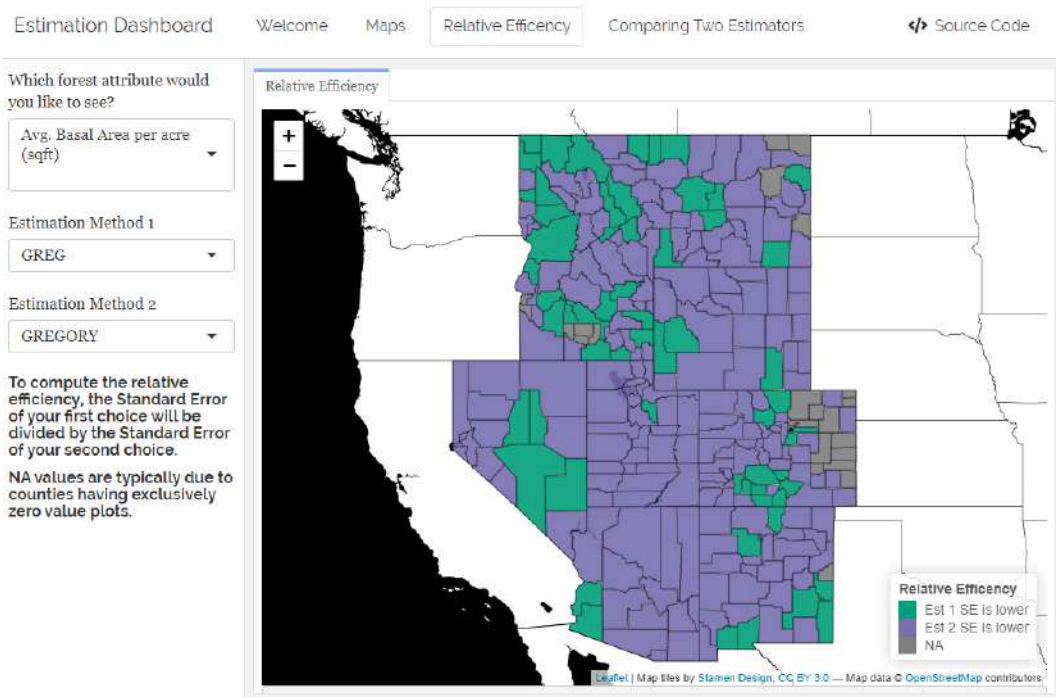

Figure 2: Comparing GREG to GREGORY

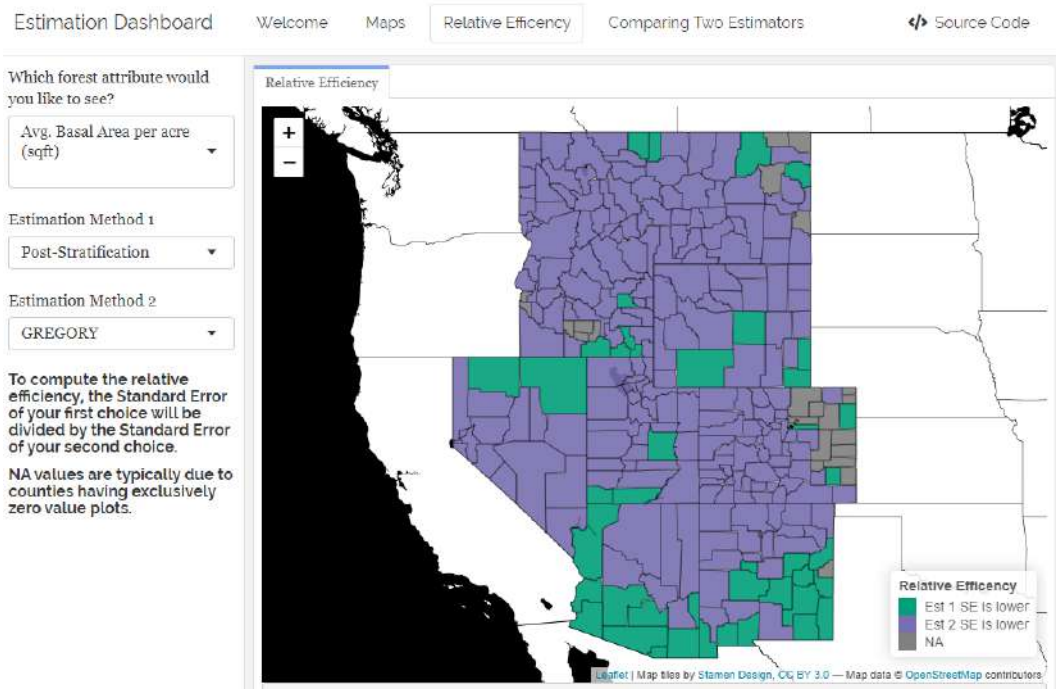

Figure 3: Comparing Post Strat to GREGORY

Figures 1, 2, and 3 allow for a quick look, using our dashboard, at how the relative efficiencies of the estimators compare to one another. Comparing GREG to Post Stratification can show

how most counties improve when using a generalized estimator, but the counties where PS tends to have a lower bootstrap variance are the ones where there are fewer forests and trees. The N/A values are due to the fact that both GREG and PS have a bootstrap Standard Error of zero when all of the plot values for a county are zero. The GREGORY estimator avoids this by acting as an indirect estimator.

Comparing PS and GREGORY shows how overall, most counties seem to improve when running regressions at the ecological province level. Additionally, comparing GREG and GREGORY shows that GREGORY does a better job than GREG of providing smaller bootstrapped standard errors.

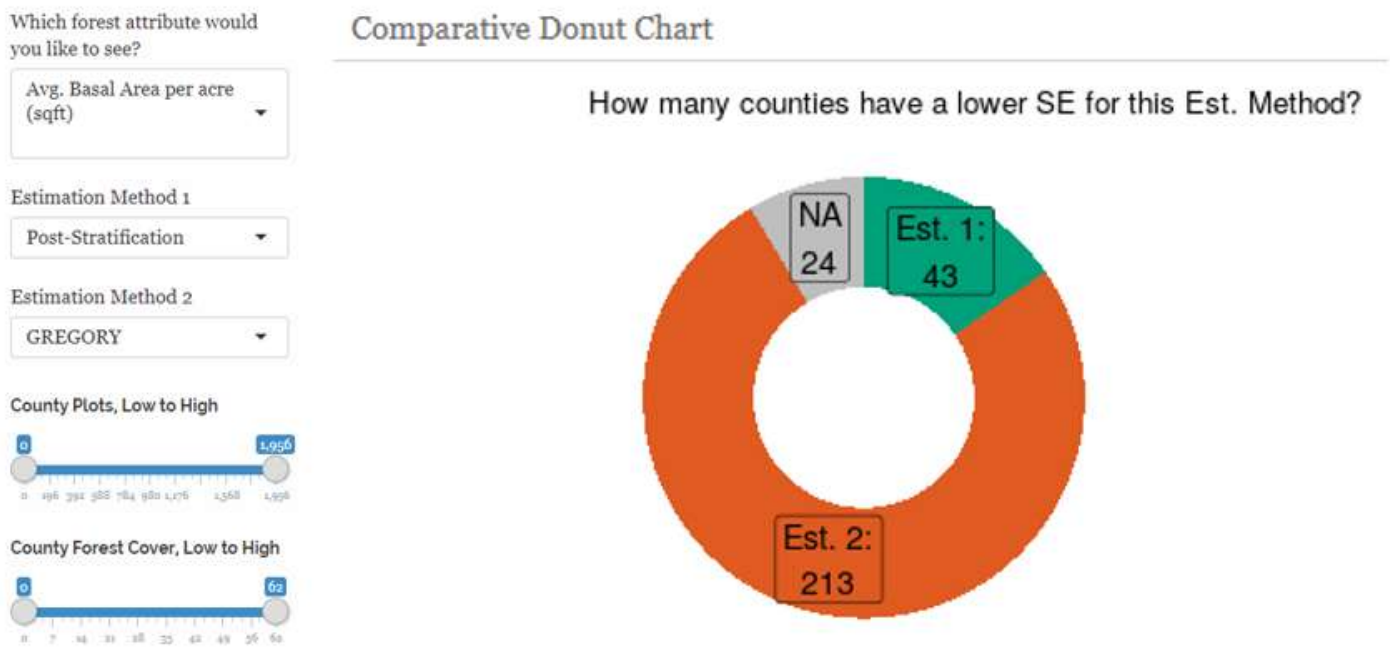

Figure 4: Comparitive Donut Chart

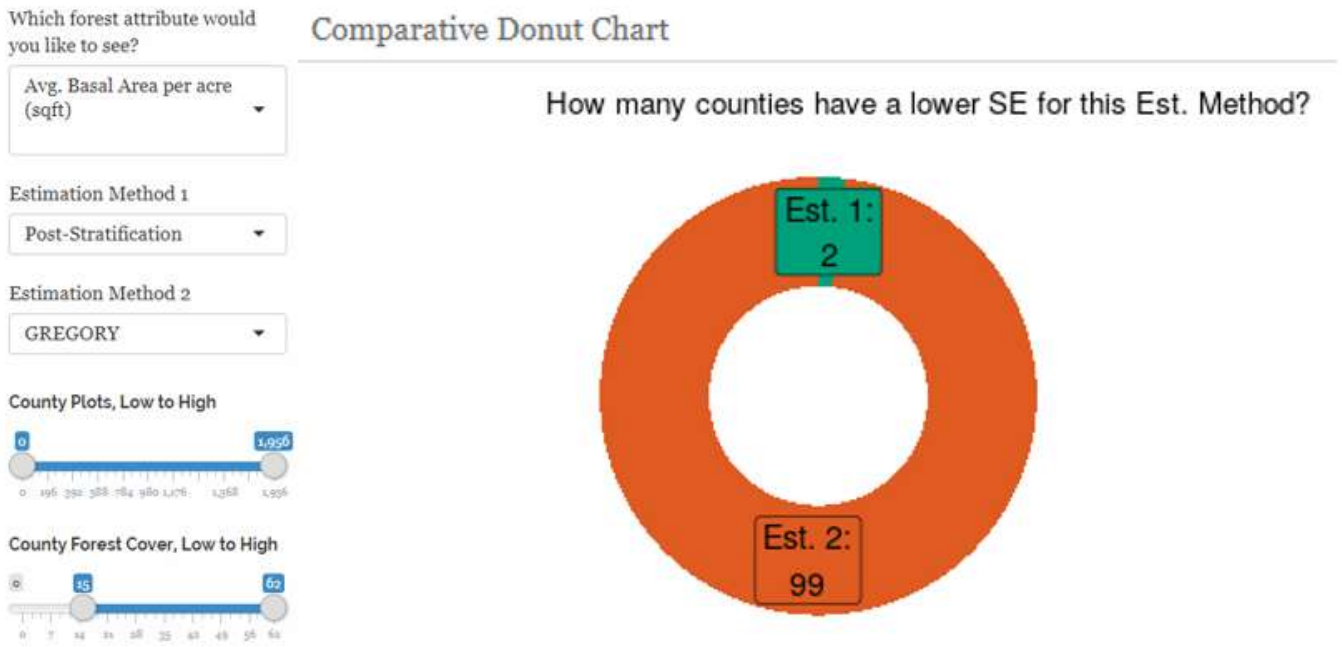

Figure 5: Comparitive Donut Chart, Adjusted

A noteworthy insight the dashboard provides is that the GREGORY estimator does seem to primarily do better in places with higher levels of general forest characteristics. Figure 5 is an adjusted version of Figure 4, with a fraction of counties with the lowest canopy cover removed. When these counties are missing, GREGORY does better in all but two counties, and all the N/A cases are removed as well. This suggests that GREGORY is a much better estimator if we are interested in counties with a significant amount of forested area.

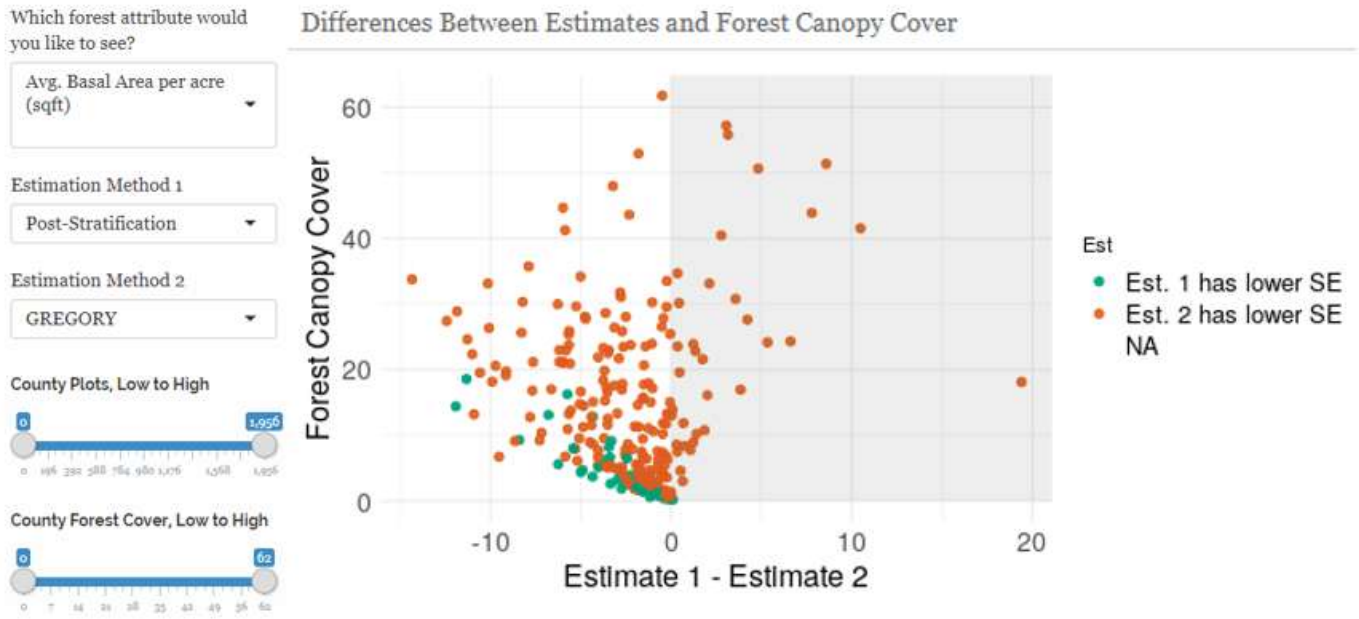

Figure 6: Differences Chart

The chart in Figure 6 shows the difference between the Post Stratification and GREGORY estimates for basal area as its x-axis, and forest canopy cover as its y-axis, with each point representing a county. This chart shows that in the counties where GREGORY is doing worse, or the green points near the bottom of the chart, nearly all cases have a larger estimate than PS. This suggests GREGORY may overestimate basal area, and potentially other forest attributes, in areas where there is basal area.

We have only been investigating one variable, basal area, in this section. While other variables share similar patterns, and produce similar observations, many more insights in our data can be found investigating our dashboard: https://shiny.reed.edu/s/users/wojciko/ests/

5 Conclusion

We found that GREGORY, a modified, indirect GREG estimator we created to incorporate additional auxiliary information, tends to produce more precise estimates of forest attributes than the currently used PS estimator. GREGORY also tends to be more precise than the more direct GREG estimator, and, unsurprisingly, it does much better than the simple HT estimator. However, there is more work to be done. GREGORY appears to struggle in counties with fewer trees, and its larger scale regressions hold a larger chance of producing negative estimates, particularly in counties with extremely low amounts of forests. GREGORY could be improved using statistical weights, equations noted previously. Finally, this

project was only using data from the Interior West, and thus we did not use GREGORY to its fullest potential by including data from all plots within a certain ecological province. Comparing these estimators on a national level may yield even more fruitful results.

5.1 Future Work

Future work involving this estimator could be investigations with more data, specifically beyond the Interior West. Our data all came from within the Interior West, meaning that the regressions we ran at the ecological province level did not include all plot data within all of the ecological provinces. Incorporating those additional cases could produce even more accurate estimates. Weights could also be an additional concept to consider within this estimator, especially since GREGORY is capable of producing negative estimates, due to its nature as an indirect estimator.

References

Bechtold, William A., and Paul L. Patterson. 2015. "The Enhanced Forest Inventory and Analysis Program National Sampling Design and Estimation Procedures." SRS-GTR-80. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southern Research Station. https://doi.org/10.2737/SRS-GTR-80.

Blackard, J, M Finco, E Helmer, G Holden, M Hoppus, D Jacobs, A Lister, G Moisen, M Nelson, and R Riemann. 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.

"Forest Inventory and Analysis National Program - About Us." n.d. Accessed May 9, 2020. https://www.fia.fs.fed.us/about/about_us/.

Homer, Collin, Jon Dewitz, Limin Yang, Suming Jin, Patrick Danielson, John Coulston, Nathaniel Herold, James Wickham, and Kevin Megown. 2015. "Completion of the 2011 National Land Cover Database for the Conterminous United States – Representing a Decade of Land Cover Change Information." *PHOTOGRAMMETRIC ENGINEERING*, 10.

"Interior West Forest Inventory & Analysis - About Us." n.d. Accessed May 9, 2020. https://www.fs.fed.us/rm/ogden/about/index.shtml.

Magnussen, Steen, and Lutz Fehrmann. 2019. "In Search of a Variance Estimator for Systematic Sampling." *Scandinavian Journal of Forest Research* 34 (4): 300–312. https://doi.org/10.1080/02827581.2019.1599063.

McConville, Kelly S., Gretchen G. Moisen, and Tracey S. Frescino. 2020. "A Tutorial on Model-Assisted Estimation with Application to Forest Inventory." *Forests* 11 (2): 244. https://doi.org/10.3390/f11020244.

McConville, Kelly, Becky Tang, George Zhu, Sida Li, Shirley Chueng, and Daniell Toth (Author and copyright holder of treeDesignMatrix helper function). 2018. "Mase: Model-Assisted Survey Estimators." https://CRAN.R-project.org/package=mase.

McNab, W. H., D. T. Cleland, J. A. Freeouf, J. E. Keys, G. J. Nowacki, and C. A. Carpenter. 2007. "Description of Ecological Subregions: Sections of the Conterminous United States." WO-GTR-76B. Washington, DC: U.S. Department of Agriculture, Forest Service. https://doi.org/10.2737/WO-GTR-76B.