Mitigating Effects of Non-Response

Michael Carlson, Simarjot Dhaliwah, Amanda Jacobucci, Shiwei Jia, Vidya Vuppala, Rohan Wadhwa, and Max Woodbury

PURDUE UNIVERSITY | The Data Mine

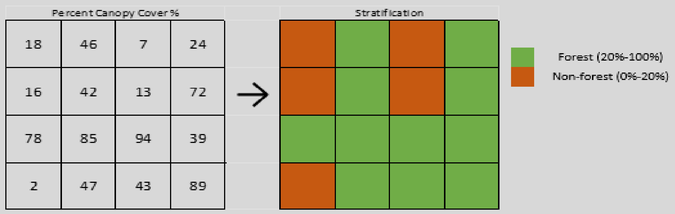
BACKGROUND

The role of the USDA's Forest Services Forest Inventory and Analysis branch (FIA) is to collect, analyze, and report information about America's forests, to allow others to make "science-based decisions, backed by forest data". FIA sends field crews to randomly sampled plots throughout the Nation, to collect a variety of core measurements. These sampled measurements are then used to obtain estimates for the Nation's forests.

FIA cannot visit every randomly sampled plot because the landowner is not required to allow them access; these plots are then labeled "non-response". FIA still needs to use these plots for their estimates. The challenge is how to fill these gaps in the data.

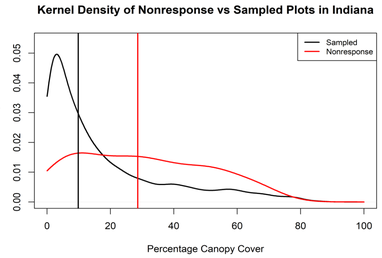
KEY CONCEPTS

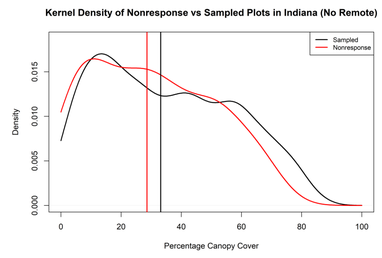
Non-forest Plots: Plots with less that 10% canopy cover
Visited Plots: plots with physically obtained measurements
Non-response Plots: Plots that cannot be physically accessed
Remotely Sensed Plots: Plots that are deemed "non-forest" from satellite imagery
Stratification: Categorizing plots into strata (groups) based on some value to increase precision (see example below)



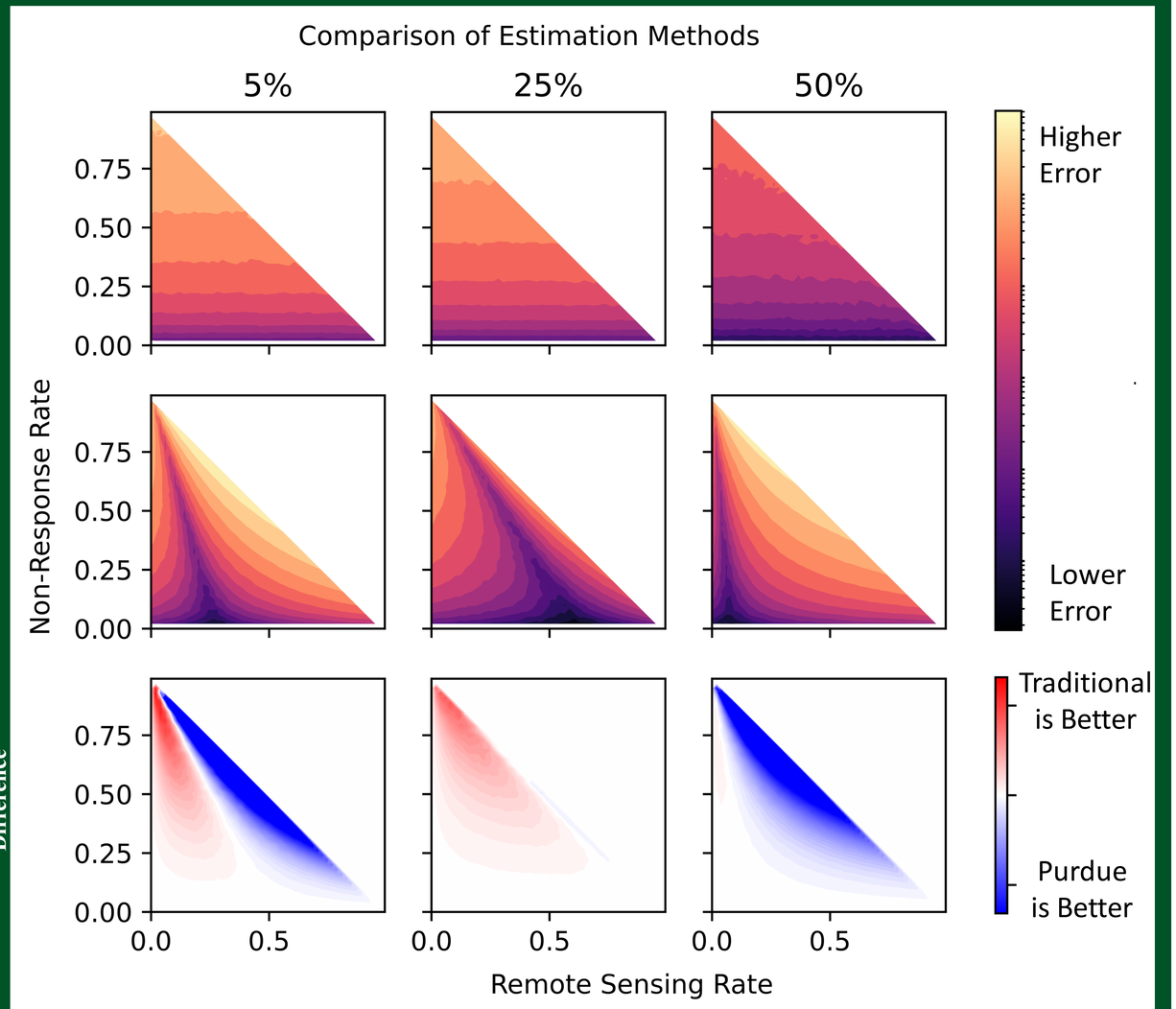
MOTIVATION

Research Question: What's the best way to deal with non-response in FIA's estimates?
Main idea: FIA labels remotely sensed plots as sampled in their database.
The traditional method assumes that the non-response plots have the same characteristics as all other plots, remote and visited.
The Purdue method assumes instead that the non-response plots more closely resemble the visited plots.

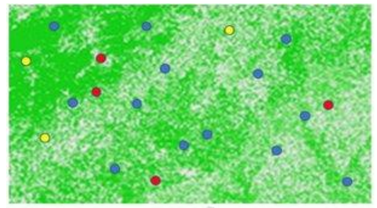




COMPARISON OF ESTIMATION METHODS THROUGH SIMULATED POPULATIONS



We simulated plot populations with different proportions of visited, non-response, and remotely sensed plots. We calculated mean canopy cover values for each of two nonresponse mitigation methods (top row: Purdue method; second row: Traditional FIA method) with these simulated populations, varying stratification threshold (columns), proportion of plots that are remotely sensed (x axis) and nonresponse rate (y axis). We also compared the errors between the two estimation methods (bottom row). The error comparison (bottom row) shows where the Purdue method outperforms the traditional method (the blue regions), and where the traditional method performs better (the red regions).



CONCLUSION

The Purdue method of filling the non-response with information from just the visited plots is generally more accurate (results in less bias) than the traditional method. We found in the simulation that the traditional method does give better estimates for a small subset of the factor combinations. However, it has a much higher error when both the proportions of remotely sensed and non-response plots are high.

Future Work

These analyses dealt with instances where entire plots were non-response. Extensions could be made to instances of partial plot nonresponse. Other states could also be analyzed and compared to these results for Indiana. For the simulation, we stratified by the canopy cover values that we were trying to predict, so other stratification variables could be used.

Acknowledgements

Corporate Partner Mentors: Andrew Lister & Gretchen Moisen
USDA Contributor: Rachel Riemann
Contributor Speaker: Kelly McConville
Corporate Partner TA: Patrick Todjalla
Data Mine Staff: Shuennhau Chang