Estimating forest biomass using machine learning maps: Correcting for bias and quantifying uncertainty

Emick, et al. (2023). An approach to estimating forest biomass while quantifying estimate uncertainty and correcting bias in machine learning maps.
Remote Sensing of Environment, 295, 113678. https://doi.org/10.1016/j.rse.2023.113678

Science Question

How do we incorporate machine learning-produced biomass maps into the forest aboveground biomass estimation process while quantifying uncertainty with statistical rigor?

Analysis

We compared four forest aboveground biomass estimators for the state of Oregon:

- Direct (DR) - where only the machine learning map is used, trained using stakeholder-contributed field plots
- Design-based (DB) - where only USFS Forest Inventory and Analysis (FIA) field data are used
- Model-assisted (MA) - where both the machine learning map and FIA field data are used in a design-based statistical paradigm
- Geostatistical-model-based (GMB) - where both the machine learning map and FIA data are used in a model-based statistical paradigm

Results

- The MA and GMB estimators appear to correct for the bias apparent in the machine learning biomass map.
- The MA and GMB estimators tend to produce estimates with uncertainties similar to the DB approach at the state level
- The MA and GMB estimators tend to produce estimates with lower uncertainties than the DB approach in counties with moderate FIA field sample sizes (e.g., 30 - 100 plots)
- The GMB estimator produces more precise estimates than the MA and DB estimators in smaller counties where FIA field sample sizes are below 30.

Significance

Carbon trading markets require statistically rigorous estimates of forest carbon with uncertainty to correctly allocate carbon credits to landowners. This work demonstrates how machine learning-produced maps of aboveground biomass can be used in a statistical estimation process to improve precision and correct for map bias. The model-assisted and geostatistical-model-based methods detailed here provide an avenue to improve forest carbon estimation and reduce costs by leveraging the wealth of forest carbon information contained in biomass map products that are made with remotely sensed data and machine learning models.

Acknowledgements

This research was supported by the NASA Carbon Monitoring System under NASA Award number 80HQTR20T0002.

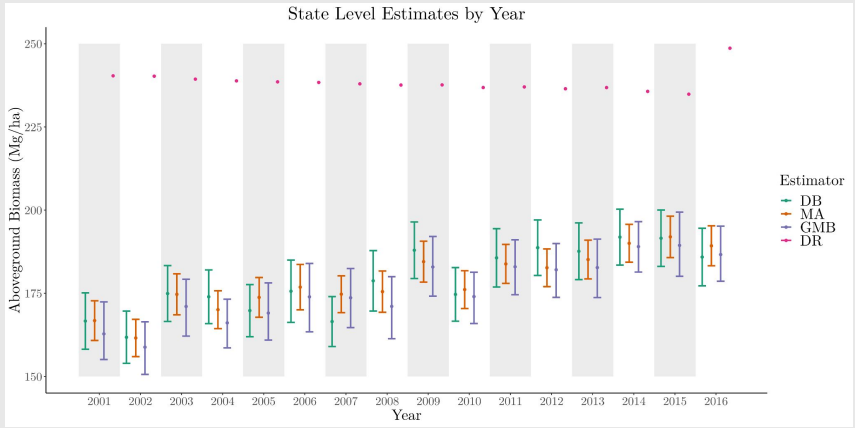

***Figure 1**: State level AGB density estimates using the four candidate AGB estimators with associated 95% uncertainty intervals. The direct (DR), design-based (DB), model-assisted (MA) and geostatistical-model-based (GMB) estimators are shown in pink, green, orange and purple, respectively.*

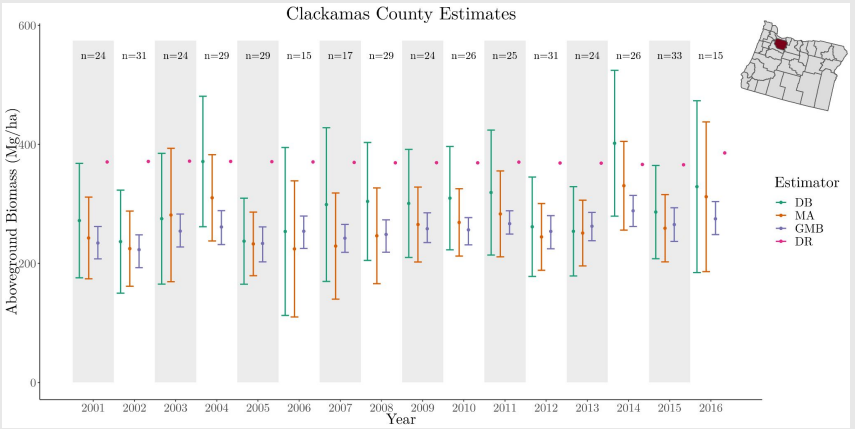

***Figure 2**: Clackamas County AGB density estimates using the four candidate AGB estimators with associated 95% uncertainty intervals. The direct (DR), design-based (DB), model-assisted (MA) and geostatistical-model-based (GMB) estimators are shown in pink, green, orange and purple, respectively.*