Forest Ecology and Management 596 (2025) 123080

Contents lists available at ScienceDirect

Forest Ecology and Management

journal homepage: www.elsevier.com/locate/foreco

Model-assisted estimation for national forest inventory via triangular tessellation

Barry Tyler Wilson

USDA Forest Service, Northern Research Station, Saint Paul, MN 55108, USA

ARTICLE INFO

Keywords:
National forest inventory
Post-stratified estimator
Model-assisted estimator
Relative efficiency
Tessellation

ABSTRACT

National forest inventories (NFIs) play an important role in the monitoring and management of natural resources and are typically implemented as sample-based surveys that provide a probabilistic basis for making inferences about the population being sampled. Some estimators incorporate data collected from the population that are auxiliary to the sample, especially remote sensing imagery. The study presented here examined the efficiency of some of these estimators, including the post-stratified (PS) and model-assisted estimators. It introduces an estimator that combines the efficiency of model-assisted regression (MAR) with the disclosure risk mitigation benefits of aggregation, called the tessellated model-assisted regression (TMAR) estimator. The results of the study, using NFI data and a Landsat-based model of live tree aboveground biomass density from the USDA Forest Service Forest Inventory and Analysis program, along with a triangular tessellation derived from 648 km^2 US EPA Environmental Monitoring and Assessment Program hexagons, demonstrate the relative efficiency of the MAR and TMAR estimators compared to the PS estimator for a sample of counties across the contiguous United States.

1. Introduction

Forest inventories play an important role in the monitoring and management of natural resources. A national forest inventory (NFI), such as the one conducted by the Forest Inventory and Analysis (FIA) program of the United States Department of Agriculture Forest Service (USDA FS), is designed to provide information on the status and trends of forests for reporting at national and sub-national levels. FIA has been Congressionally mandated since 1928 to "make and keep current a comprehensive survey and analysis of the present and prospective conditions of and requirements for renewable resources of the forests and rangelands of the United States" (Smith, 2002).

NFIs are typically implemented as sample-based surveys that rely upon a sample design with known properties, providing a probabilistic basis for making inferences about the population being sampled. A necessary component of the FIA research program has been an ongoing examination of methods for improving the precision of estimates made under this design-based mode of statistical inference, thereby extending the utility of the NFI to provide reliable information for smaller geospatial domains.

One of the most important advances in FIA estimation methods has been the incorporation of data collected from the population that are auxiliary to the sample, primarily remote sensing imagery (Lister et al., 2020). Currently, official estimates provided by FIA are made using the post-stratified (PS) estimator, with post-strata derived either directly or indirectly from satellite imagery (Scott et al., 2005). Several studies conducted using FIA data have demonstrated the efficiency of the PS estimator and suggested best practices for constructing post-strata (e.g. Hansen and Wendt, 2000; McRoberts et al., 2002; Westfall et al., 2011).

An alternative but related approach to direct estimation is model-assisted regression (MAR). The MAR estimator uses model predictions for sample observations to estimate the design-weighted bias of the model. When model predictions are produced as a raster map, as is often the case when using remote sensing imagery as auxiliary data, the MAR estimate is calculated by adjusting the synthetic estimate based on the sum of pixel values with the direct estimate of the bias (McRoberts et al., 2022). The MAR estimator can provide greater efficiency than the PS estimator for geospatial domains smaller than the estimation unit used when defining post-strata.

Another important consideration for FIA when assessing alternative estimators is maintaining the integrity of the sample. From the perspective of confidentiality of plot locations, the PS estimator has the advantage that multiple sample plots are assigned to each post-stratum, minimizing disclosure risk through aggregation. In contrast, the MAR

E-mail address: barry.wilson@usda.gov.

https://doi.org/10.1016/j.foreco.2025.123080
Received 31 March 2025; Received in revised form 4 August 2025; Accepted 9 August 2025
Available online 14 August 2025
0378-1127/Published by Elsevier B.V.

estimator requires that predicted and observed values are known at sample locations. If made publicly available, this information could be used in conjunction with a map of predicted values to deduce sample locations.

The study presented here examined an approach to model-assisted estimation that combines the use of the MAR estimator with aggregation. The rationale of the approach is to retain the benefits of improved efficiency of estimation for sub-State geospatial domains while mitigating the risk of disclosure of FIA plot locations. A secondary motivation of the approach is to improve the dynamic performance of calculations required to support the development of a public-facing estimation web application.

2. Materials and methods

The FIA national forest inventory is conducted in three phases of sampling, the first two of which are relevant to the current study and described in more detail here (Reams, et al., 2005). In Phase 1 of the sample, the land base of the United States is subdivided into estimation units (subpopulations) with known census area, typically either individual counties or aggregations of counties. These subpopulations are then stratified to reduce the variance of sample design-based estimates. FIA's current standard practice for stratification is to use auxiliary information derived from satellite imagery to classify pixels into a small number of roughly homogeneous strata and then assign strata weights within each estimation unit based on relative pixel counts. FIA also attempts to account for non-response bias on private land (i.e. field crews may be denied access to a plot by a private landowner) either by breaking estimation units into public and private ownership classes or by incorporating ownership into the stratification of estimation units. This stratification in Phase 1 occurs independently from plot selection in Phase 2. As a result, strata sample sizes are random variables and strata may need to be collapsed after Phase 2 to ensure sufficient sample sizes. This procedure is technically post-stratification, using post-strata.

In Phase 2 of the sample, crews visit permanent field plots and measure a standard suite of forest inventory attributes. The FIA base sample design consists of one plot randomly located within each 2400 ha hexagon of the hexagonal sampling frame that covers the United States. In some cases, FIA has been able to intensify the sample for certain subpopulations. This, in addition to plots that are wholly or partially non-sampled due to denied access or hazards, results in differing design weights for plots. Depending on the location of the estimation unit and whether plot collection has been accelerated, plots from a subset of 10–20 % of the sampling frame are visited and measured each year, yielding a cycle length of 5–10 years between plot remeasurements. This design allows FIA to make estimates using a moving average over a full cycle of plots each year.

The Phase 2 plot design is a cluster of 4 circular subplots with a radius of ~7.3 m, arranged as a central subplot surrounded by subplots at 0, 120, and 240 degrees, each having a center ~36.6 m from the center of the central subplot (Bechtold and Scott, 2005). All trees with a diameter at breast height (DBH) of 12.7 cm or greater are measured on the subplot. Each subplot contains a smaller circular microplot of radius ~2.1 m where saplings (DBH 2.54–12.45 cm) are measured. Tree measurements can then be expanded to density (per hectare) values by using the appropriate tree expansion factor associated with subplots or microplots. Additionally, plots are mapped by area classifications, called conditions, determined by characteristics such as forest type, stand origin, and ownership group. These conditions can be used to disaggregate the population into domains of interest for reporting.

The study used the full cycle of sampled FIA field plots for the contiguous United States (CONUS), ending with the 2018 inventory year to achieve approximate temporal alignment with the data represented by the assisting model, described subsequently. The only exception was the state of California, for which the full cycle of plot data ending with the 2019 inventory year was used instead, because the comparable cycle of plot data for the 2018 inventory year was not available at the time the study was conducted. The entire dataset included 337,738 plots.

The forest inventory attribute used for estimation was live tree aboveground biomass (LTAGB) density, for trees measured on the central subplot to better represent the pixel resolution of the assisting model output. For simplicity this value was calculated by doubling the associated live tree aboveground carbon (LTAGC) density value stored in the FIA database (FIADB). Given the recency of the current study, the LTAGC density values retrieved from FIADB were based on the National Scale Volume and Biomass (NSVB) equations (Westfall et al., 2024). However, the LTAGB density values were not based on NSVB carbon fractions.

Estimates of total LTAGB, variance (σ^2), and percent sampling error (%SE) were calculated using the post-stratified estimator for a random sample of 60 of the 3108 counties of CONUS excluding the District of Columbia (Fig. 1). The sample included counties from each of the four FIA regional programs, having a variety of shapes, and ranging in size from roughly 45,000 to 1,610,000 ha (roughly the 3rd through 99th percentiles). The PS estimates were produced via the EVALIDator web application, using the set of FIA plots described above, along with the associated post-strata and design weights (i.e. expansion factors) stored in FIADB (USDA Forest Service, 2025a). Per the metadata found in FIADB, the post-strata used in the study were based on classified Landsat imagery for variance reduction, along with an ownership layer to account for nonresponse bias on private land.

Corresponding estimates were calculated using the MAR estimator. A raster map of pixel-level LTAGB density predictions from the FIA BIGMAP project was used as the assisting model (Fig. 1) (USDA Forest Service, 2025b). The BIGMAP methodology is based on *k*-nearest neighbors imputation of FIA plots to pixels, using proximity in a feature space derived from an ecological ordination of tree species measured on central subplots, with ecological regions, climate, topography, and vegetation phenology derived from dense time series imagery as predictors (Wilson et al., 2012, 2018). The plot measurements and Landsat 8 OLI imagery used to produce the map of LTAGB density were collected during the period 2014–2018, and the LTAGB density values used were based on the pre-NSVB component ratio method (CRM) (Woodall et al., 2011).

The hexagons of the US Environmental Protection Agency Environmental Monitoring and Assessment Program (EMAP) sampling grid were the basis for a finer tessellation of CONUS (White et al., 1992). Each 648 km^2 EMAP hexagon was subdivided into 6 smaller 108 km^2 triangles (Fig. 2). The triangular tessellation was then clipped to the extent of the FIA sampling frame, resulting in some irregular polygons along the sampling boundary. To minimize risk of disclosure of plot locations by public release of the study dataset, all polygons in the tessellation containing fewer than two plots were aggregated with neighboring polygons until this criterion was met, here referred to as tessellation polygons.

The county-level MAR estimates of total LTAGB were calculated using known county census area and mean LTAGB density in two ways: 1) using all plots and pixels within the county (Fig. 3a), and 2) using all plots and pixels within each tessellation polygon having its centroid within the county (Fig. 3b). The first will be referred to as the MAR estimator, and the second as the tessellated MAR (TMAR) estimator.

The design weights used with the MAR and TMAR estimators were calculated by collapsing all post-strata in each FIA estimation unit (EU), then dividing EU area by the sum of all sampled condition proportions for the set of plots used in the study and located within the EU. All terms necessary for the TMAR estimator were precompiled for each tessellation polygon. These include the 1) mean of predicted LTAGB density pixel values, 2) sample count, 3) sum of sample design weights, and 4) sum of design-weighted square errors.

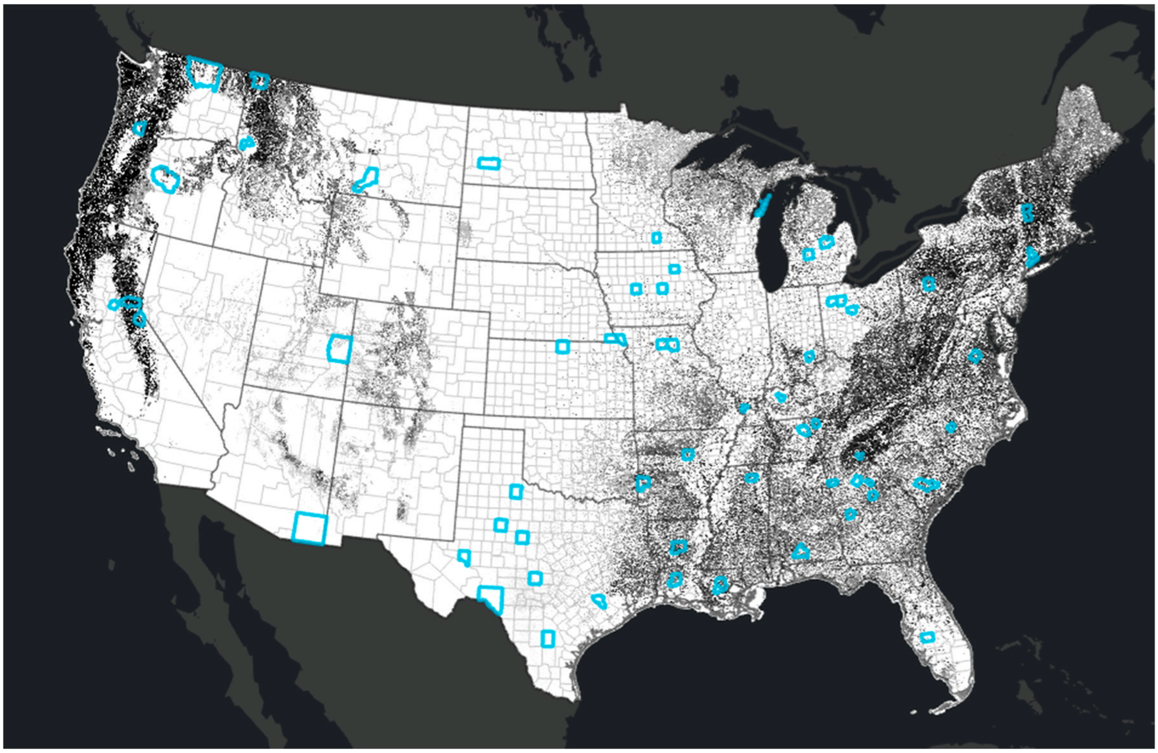

Fig. 1. Raster map of predicted LTAGB density from the BIGMAP assisting model for the FIA sampling frame, with state and county boundaries, and the sample of counties included in the study (highlighted in blue). Darker pixels represent higher LTAGB density values, with a range of 0–1026 Mg/ha.

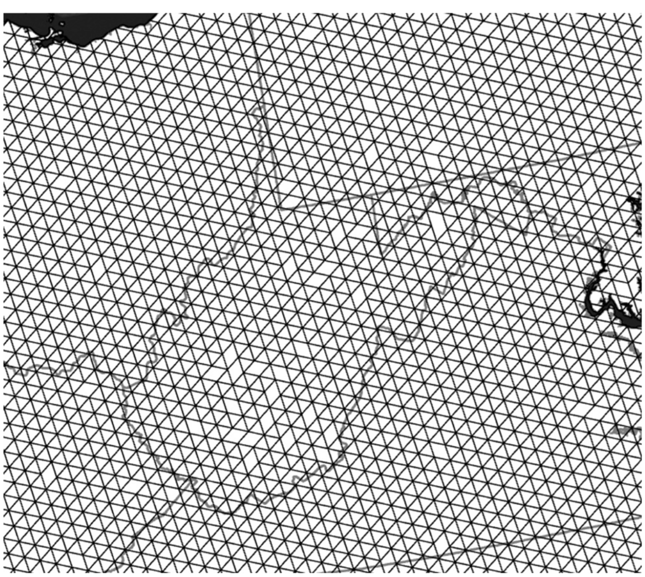

Fig. 2. Detail of the triangular tessellation of 108 km^2 triangles spanning the FIA sampling frame (in white), shown with state boundaries for scale. Examples of the aggregation of triangles to minimize disclosure risk are visible in West Virginia (irregular boundary in the center), as are the smaller irregular polygons along the shores of Lake Erie (dark area in the upper left) and Chesapeake Bay (dark area on the right edge).

3. Results

Scatterplots comparing estimates of total LTAGB for each of the 60 counties in the sample using the PS, MAR, and TMAR estimators are shown in Fig. 4. There is strong agreement among all estimators, with the highest R^2 for the relationship of MAR vs. TMAR (0.9938), followed by PS vs. MAR (0.9912), then PS vs. TMAR (0.988). Similarly, the slope of the OLS regression line is closest to 1 for MAR vs. TMAR (0.9916), followed by PS vs. MAR (0.9566), then PS vs. TMAR (0.95) estimators.

Relative efficiency (RE) is a measure comparing two estimators using the ratio of their variances, with the baseline (typically less efficient) estimator as the numerator and the alternative (typically more efficient) estimator as the denominator. RE quantifies how much larger the sample size would need to be using the baseline estimator to achieve the same precision as the alternative estimator given the actual sample size. For this analysis, PS was used as the baseline estimator, with either MAR or TMAR as the alternative estimator.

Because of the range of geospatial extents, variety of forest subpopulations, and accompanying sample sizes of the counties used in the analysis, the distribution of estimated total LTAGB variances was highly skewed. As a result, median RE was chosen as an appropriate measure of central tendency. Also, because the assisting model tended to underestimate LTAGB density compared to the sample, generally resulting in smaller estimated variances when using model-assisted estimators, median RE was also calculated based on the ratio of square %SE.

Both MAR and TMAR estimators had median RE greater than 1 using either method of RE calculation (Table 1), demonstrating the greater efficiency of the model-assisted estimators compared to the PS estimator for the sample of county-sized geospatial domains used. The MAR estimator had a slightly higher RE than the TMAR estimator when using the ratio of variances, though they were nearly identical when using square %SE. Both the MAR and TMAR estimators had higher RE when based on the ratio of variances rather than square %SE.

4. Discussion

The systematic underestimation of the assisting model, visible in the scatterplots (Fig. 4b and c), is likely due to a combination of factors. First, the assisting model used a forest mask for the purpose of reducing errors of commission. The mask layer is only an approximation of the forest population sampled by FIA. Second, due to the vintage of the data used to create it, the assisting model of LTAGB density was developed

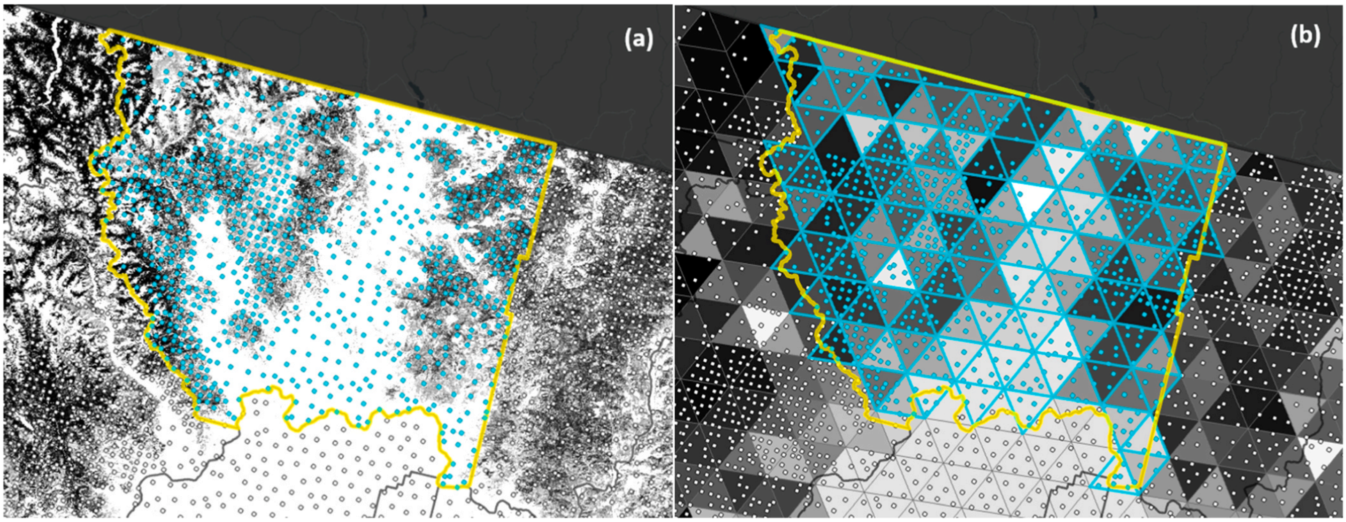

Fig. 3. Example of data used by the model-assisted estimators of total LTAGB in Okanogan County, Washington (highlighted in yellow). (a) The MAR estimator used raster pixel values and sample point data (highlighted in blue) within the county boundary. (b) The TMAR estimator used summarized raster pixel values and sample point data within tessellation polygons (highlighted in blue) having centroids within the county boundary. In both parts of the figure, darker pixels and polygons represent higher LTAGB density values, with ranges of 0–1026 Mg/ha and 0–509 Mg/ha respectively.

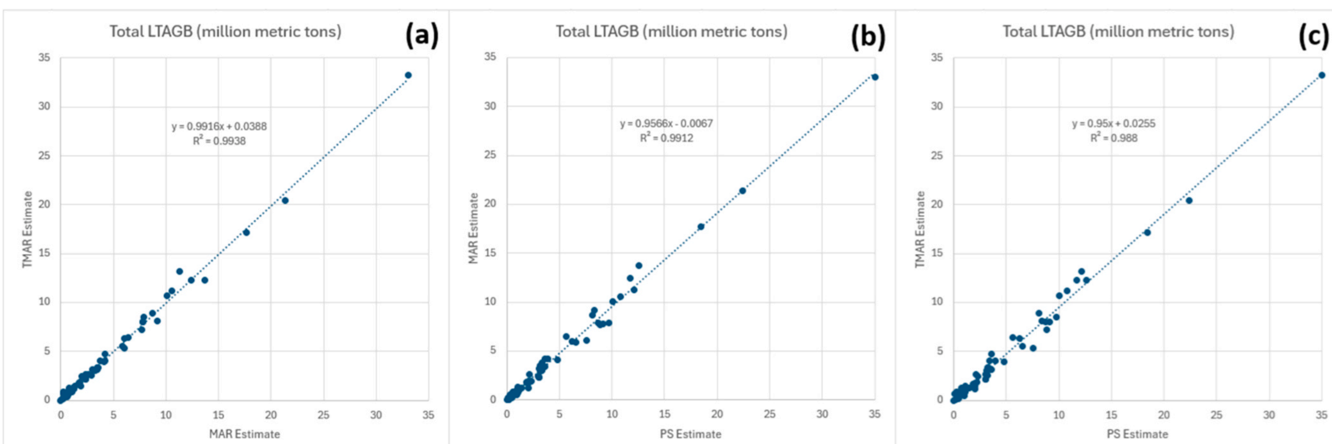

Fig. 4. Scatterplot comparisons among estimators of county-level total LTAGB (in million metric tons) for the study sample of counties: (a) MAR vs. TMAR, (b) PS vs. MAR, and (c) PS vs. TMAR.

Table 1
Median relative efficiency (RE) of the alternative MAR and TMAR estimators, compared to the baseline PS estimator, using ratio of variances (σ^2) and square percent sampling errors (SE%^{2}) for the study sample of counties.

Estimator	RE(σ^2)	RE(SE%^{2})
MAR	1.684	1.411
TMAR	1.572	1.426

using the CRM, rather than NSVB, biomass equations. There are differences between CRM and NSVB that suggest CRM may result in systematic underestimation of LTAGB density (Westfall et al., 2024).

There is slightly better agreement between PS vs. MAR than PS vs. TMAR estimates of county total LTAGB. This is largely due to the PS and MAR estimators using the sample of plots from each county, whereas the TMAR estimator uses plots from a spatial domain that approximates each county. To a lesser degree, this may also be due to the MAR estimator using the predicted pixel values from the county rather than an approximate spatial domain in the TMAR estimator case. It should be noted that the PS estimator uses exact post-strata weights only for an FIA EU. One of the four regional FIA programs uses county as an EU, while for the others, PS estimates for counties are based on the post-strata weights of the encompassing EU, reducing the efficiency of the PS estimator for county spatial domains.

As expected, the systematic underestimation of the assisting model leads to lower estimated totals and variances and therefore greater median RE of MAR and TMAR estimators compared to the PS estimator. However, even when RE is calculated as the ratio of square %SE, the efficiency of the model-assisted estimators is clear. In the case of RE based on variances, the lower median RE of the TMAR compared to MAR estimator is likely due to the additional variance introduced by the approximation of the spatial domain using the tessellation polygons. Using the other method of calculating RE, the results are essentially the same for the two model-assisted estimators.

To compute estimates of total and variance for a sub-state spatial domain of interest using the TMAR estimator, one can sum each of the necessary terms over all tessellation polygons having their centroid within the domain. This reduces to a "point in polygon" geospatial operation, with centroids as points and domains as polygons. All necessary terms for the TMAR estimator can be stored as point attributes, thereby simplifying calculations and improving dynamic performance.

In the context of a web application, calculations based on the TMAR estimator can all take place on the client. Other than the initial loading of the web page, this eliminates client-server communication that would be required for the MAR estimator, such as for computing zonal statistics from the raster for the domain or for completing any "point in polygon" operations based on the actual plot coordinates. Also, because all terms needed for the TMAR estimator have been aggregated to an area containing at least two FIA plots, there is limited plot location disclosure risk associated with releasing the centroid points and their associated attributes.

5. Conclusions

In the context of the NFI conducted by FIA, two primary conclusions can be drawn from the study. 1) Model-assisted estimators have the potential for greater relative efficiency than the post-stratified estimator for spatial domains smaller than the estimation unit. 2) The tessellated model-assisted regression estimator introduced here provides comparable results to the model-assisted regression estimator for county-sized spatial domains, with the benefit of reduced plot location disclosure risk and simpler dynamic computation at the cost of some slight reduction in efficiency. However, it should be noted that these conclusions are based on results using a particular sample and plot design, forest inventory variable, assisting model, tessellation, and sample of spatial domains. While suggestive of broader utility, further study would be needed before drawing conclusions about the efficiency of the model-assisted estimators presented here for NFIs more generally.

CRediT authorship contribution statement

Barry Tyler Wilson: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The author would like to acknowledge the United States Department of Agriculture Forest Service (USDA FS) FIA program, which funded this research and whose staff collected, compiled, analyzed, and published the forest inventory data used in this study, and particularly the members of the FIA BIGMAP project team, whose efforts in collaboration with Esri produced the raster map output of the assisting model.

Data availability

Some data used are available by request. However no data that pose a risk of disclosure of FIA plot locations may be shared.

References

Bechtold, W.A., Scott, C.T., 2005. The forest inventory and analysis plot design. Gen. Tech. Rep. SRS-80. U.S. Department of Agriculture, Forest Service, Southern Research Station, Asheville, NC:, pp. 27–42. https://doi.org/10.2737/SRS-GTR-80.

Hansen, M.H., Wendt, D.G., 2000. Using classified landsat thematic mapper data for stratification in a statewide 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. Department of Agriculture, Forest Service, North Central Research Station, St. Paul, MN: U.S, pp. 20–27. https://doi.org/10.2737/NC-GTR-213.

Lister, A.J., Andersen, H., Frescino, T., Gatziolis, D., Healey, S., Heath, L.S., Liknes, G.C., McRoberts, R., Moisen, G.G., Nelson, M., Riemann, R., Schleeweis, K., Schroeder, T. A., Westfall, J., Wilson, B.T., 2020. Use of remote sensing data to improve the efficiency of national forest inventories: a case study from the United States national forest inventory. Forests 11 (12), 1364. https://doi.org/10.3390/f11121364.

McRoberts, R.E., Wendt, D.G., Nelson, M.D., Hansen, M.H., 2002. Using a land cover classification based on satellite imagery to improve the precision of forest inventory area estimates. Remote Sens. Environ. 81 (1), 36–44. https://doi.org/10.1016/S0034-4257(01)00330-3.

McRoberts, R.E., Næsset, E., Heikkinen, J., Chen, Q., Strimbu, V., Esteban, J., Hou, Z., Giannetti, F., Mohammadi, J., Chirici, G., 2022. On the model-assisted regression estimators using remotely sensed auxiliary data. Remote Sens. Environ. 281, 113168. https://doi.org/10.1016/j.rse.2022.113168.

Reams, Gregory A., William D, Smith, Mark H, Hansen, William A, Bechtold, Francis A, Roesch, Gretchen G, Moisen, 2005. The forest inventory and analysis sampling frame. Gen. Tech. Rep. SRS-80. U.S. Department of Agriculture, Forest Service, Southern Research Station, Asheville, NC, pp. 11–26. https://doi.org/10.2737/SRS-GTR-80.

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. Gen. Tech. Rep. SRS-80. U.S. Department of Agriculture, Forest Service, Southern Research Station, Asheville, NC, pp. 43–67. https://doi.org/10.2737/SRS-GTR-80.

Smith, W.B., 2002. Forest inventory and analysis: a national inventory and monitoring program. Environ. Pollut. 116, S233–S242. https://doi.org/10.1016/S0269-7491(01)00255-X.

USDA Forest Service (2025a) EVALIDator. Available at: 〈https://apps.fs.usda.gov/fiadb-api/evalidator〉 (Accessed October 2024).

USDA Forest Service (2025b) Forest Service FIA BIGMAP Tree Species Aboveground Biomass Layers. Available at: 〈https://data.fs.usda.gov/geodata/rastergateway/bigmap/index.php〉 (Accessed 31 March 2025).

Westfall, J.A., Patterson, P.L., Coulston, J.W., 2011. Post-stratified estimation: within-strata and total sample size recommendations. Can. J. For. Res. 41 (5), 1130–1139. https://doi.org/10.1139/x11-031.

Westfall, J.A., Coulston, J.W., Gray, A.N., Shaw, J.D., Radtke, P.J., Walker, D.M., Weiskittel, A.R., MacFarlane, D.W., Affleck, D.L., Zhao, D., Temesgen, H., 2024. A national-scale tree volume, biomass, and carbon modeling system for the United States. Gen. Tech. Rep. WO-104. US Department of Agriculture, Forest Service, Washington, DC, p. *104*. https://doi.org/10.2737/WO-GTR-104.

White, D., Kimerling, J.A., Overton, S.W., 1992. Cartographic and geometric components of a global sampling design for environmental monitoring. Cartogr. Geogr. Inf. Syst. 19 (1), 5–22. https://doi.org/10.1559/152304092783786636.

Wilson, B.T., Lister, A.J., Riemann, R.I., 2012. A nearest-neighbor imputation approach to mapping tree species over large areas using forest inventory plots and moderate resolution raster data. For. Ecol. Manag. 271, 182–198. https://doi.org/10.1016/j.foreco.2012.02.002.

Wilson, B.T., Knight, J.F., McRoberts, R.E., 2018. Harmonic regression of landsat time series for modeling attributes from national forest inventory data. ISPRS J. Photogramm. Remote Sens. 137, 29–46. https://doi.org/10.1016/j.isprsjprs.2018.01.006.

Woodall, C.W., Heath, L.S., Domke, G.M., Nichols, M.C., 2011. Methods and equations for estimating aboveground volume, biomass, and carbon for trees in the US forest inventory, 2010. In: Gen. Tech. Rep. NRS-88, 30. US Department of Agriculture, Forest Service, Northern Research Station, *Newtown Square, PA*, pp. 1–30. https://doi.org/10.2737/NRS-GTR-88. *p.*, *88*.