Remote Sensing of Environment 330 (2025) 114951

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

From pixels to parcels: Flexible, practical small-area uncertainty estimation for spatial averages obtained from aboveground biomass maps

Lucas K. Johnson [a,b,*], Grant M. Domke [c], Stephen V. Stehman [a], Michael J. Mahoney [a], Colin M. Beier [a]

[a] State University of New York College of Environmental Science and Forestry, Department of Sustainable Resources Management, 1 Forestry Drive, Syracuse, NY, 13210, USA
[b] Oregon State University, College of Forestry, Corvallis, OR, 97331, USA
[c] USDA Forest Service, Northern Research Station, St. Paul, MN, 55114, USA

ARTICLE INFO

Keywords:
Model-based inference
Bootstrap
Landsat
Machine learning
National forest inventory
Forest inventory and analysis

ABSTRACT

Fine-resolution maps of forest carbon and aboveground biomass (AGB) effectively represent spatial patterns and can be flexibly aggregated to map subregions by computing spatial averages or totals of pixel-level predictions. However, generalized model-based uncertainty estimation for spatial aggregates requires computationally expensive processes like iterative bootstrapping and computing pixel covariances. Uncertainty estimation for map subregions is critical for enhancing practicality and eventual adoption of model-based data products, as this capability would empower users to produce estimates at scales most germane to management of the landscape: individual forest stands and ownership parcels. In this study we produced estimates of standard error (SE) associated with spatial averages of fine-resolution AGB map predictions for a stratified random sample of ownership parcels (0–2023 hectares) in New York State (NYS). This represents the first model-based uncertainty estimation study to include all four types of uncertainty (reference data, sample variability, residual variability, and auxiliary data), incorporate spatial autocorrelation of model residuals, and use methods compatible with algorithmic modeling (machine learning or nonparametric). We found that uncertainty attributed to residual variance, largely resulting from spatial correlation of residuals, dominated all other sources for all of the parcels in the study. These results suggest that improvements to model accuracy will yield the greatest reductions to total uncertainty in regions like the northeastern and midwestern United States where forests are divided into smaller spatial units. Further, we demonstrated that log–log regression relating parcel characteristics (area, perimeter, AGB density, forest cover) to parcel-level SE can accurately estimate uncertainty for map subregions, thus providing a convenient means to empower map users. These findings support transparency in future model-based forest carbon accounting and monitoring efforts.

1. Introduction

Over the past 30 years, greenhouse gas (GHG) accounting efforts have proliferated from local to global scales and across private and public sectors. These efforts began in 1992 with the United Nations requiring all developed counties to report annual emissions (UNFCCC, 1992), and since have evolved into the Kyoto Protocol (UNFCCC, 1997), and the Paris Climate agreement (COP, 2015) with accounting methodologies and practices defined by the 2019 IPCC update on GHG inventory methods (Buendia et al., 2019). These standards require that GHG emissions and removals be reported with confidence intervals and with uncertainties reduced as far as possible (Penman et al., 2003; Eggleston et al., 2006). Beyond satisfying these requirements, identifying and understanding sources of uncertainty are critical for distinguishing between noise and real GHG emissions and removals. Forests are among the most effective carbon sinks with global net sequestration rates estimated at 1.1 (±0.8) Pg of carbon per year (Pan et al., 2011). However, the strength of this sink is not spatially consistent and in some regions forest lands may be net carbon sources due

* Corresponding author at: State University of New York College of Environmental Science and Forestry, Department of Sustainable Resources Management, 1 Forestry Drive, Syracuse, NY, 13210, USA.
E-mail addresses: johnsl27@oregonstate.edu (L.K. Johnson), grant.m.domke@usda.gov (G.M. Domke), svstehma@esf.edu (S.V. Stehman), mjmahone@esf.edu (M.J. Mahoney), cbeier@esf.edu (C.M. Beier).

https://doi.org/10.1016/j.rse.2025.114951
Received 17 December 2024; Received in revised form 21 July 2025; Accepted 28 July 2025
Available online 13 August 2025

to deforestation and degradation (Pan et al., 2011). Consequently, a significant body of research is dedicated to accounting for forest carbon fluxes with improved accuracy and spatial resolution.

Many large-scale forest carbon accounting approaches leverage the extensive sampling designs of national forest inventories (NFI) to estimate annual emissions and removals from the forest sector (Woodall et al., 2015; McRoberts et al., 2010). These estimates, made with only a probability sample of reference data, are classified as design-based inference. Design-based estimators are often accompanied by relatively accessible estimators of variance, which account for sampling variability but omit variability related to tree- and plot-level predictions (reference data uncertainty) since population units are assumed to have a single fixed value under the design-based framework (McRoberts, 2011; McRoberts et al., 2022). Although design-based approaches offer fundamental insights and essential data on forest carbon dynamics, they are often limited spatially and temporally by sampling density and measurement frequency respectively (McRoberts, 2011) and thus cannot represent fine-scale patterns and dynamics most relevant to planning and decision-making. Model-based approaches, which combine field data (e.g. NFI plots) with wall-to-wall, remotely-sensed data can fill this need by producing predictions for all map units (pixels) in a given area (Kennedy et al., 2018a; Matasci et al., 2018; Huang et al., 2019; Hudak et al., 2020; Johnson et al., 2022, 2023).

Although model-based approaches offer enhanced spatial flexibility relative to their design-based counterparts, they rely on different assumptions and require a more complex combination of terms for estimating uncertainty. Under model-based inferential frameworks, variability is derived from the assumption that each population unit has an entire distribution of possible values, with each random realization of a population unit value corresponding to a single distinct population from an overarching set of populations known as a super-population (McRoberts et al., 2022). Following model-based terminology, super-population parameters are thus estimated, whereas population parameters are predicted (McRoberts et al., 2022). Model-based estimates of uncertainty may include the following components: uncertainty in reference data (Table 1, type a), uncertainty due to training the model on a sample (Table 1, type b), uncertainty due to the model not being able to represent all of the variation in the response variable (Table 1, type c), and auxiliary data uncertainty (Table 1, type d) (Wadoux and Heuvelink, 2023; CEOS, 2021). Different terms have been used to refer to type b, including model parameter/prediction covariance, model parameter/prediction uncertainty (McRoberts et al., 2018), or sampling variability (CEOS, 2021; McRoberts et al., 2022) since model parameter estimates and subsequent predictions depend on the training data sample. Type c is usually called residual variability and is sometimes broken down into components quantifying residual variance and spatial covariance of residuals (McRoberts et al., 2018; CEOS, 2021; McRoberts et al., 2022). The proportion of model-based studies that produce uncertainty estimates is small (McRoberts, 2011), likely because the technical complexity of generating uncertainty estimates is substantially greater than that of making predictions.

From the model-based studies that do include estimates of uncertainty, desirable features for the uncertainty estimation process include:

1. Flexibility to incorporate parametric as well as algorithmic models (i.e. machine learning and nonparametric models).
2. Incorporating all four types of uncertainty (Table 1).
3. Methods for producing uncertainty estimates for spatial averages or totals that explicitly account for spatial autocorrelation of residuals.
4. Consideration of computational efficiency and delivery to map users.

Applications of rigorous model-based uncertainty estimation have been advanced using parametric models with accompanying analytic estimators of variance (Saarela et al., 2020, 2018, 2016; Chen et al., 2016, 2015; McRoberts et al., 2018). However, these analytic estimators are fairly complex, and can be inaccessible for practitioners lacking deep backgrounds in statistics or mathematics. Further, requiring parametric models may be limiting, and thus in violation of desirable criterion 1 (defined above), as algorithmic approaches may yield improved predictive accuracy (Efron, 2020; Breiman, 2001b). At the cost of computational efficiency, replication approaches like bootstrapping or Monte Carlo simulations, however, are flexible to model form, can easily incorporate uncertainty from several sources, and are conceptually simple relative to analytic estimators (McRoberts et al., 2022; CEOS, 2021; Esteban et al., 2020).

Reference data uncertainty (Table 1, type a) is often deemed negligible or is omitted due to a scarcity of representative data describing measurement error and allometric uncertainty (Breidenbach et al., 2014), thus violating desirable criterion 2 (defined above). For forest aboveground biomass (AGB) or carbon studies in particular, reference data may be highly uncertain, given that tree-level AGB values in NFIs are typically predictions themselves. Instead, tree-level biomass and carbon are modeled as a function of measured variables like diameter and height using allometric models developed for particular regions and taxa (Jenkins et al., 2003; Woodall et al., 2011; Chave et al., 2014). The measured variables themselves are subject to error and imprecision, further contributing to the overall uncertainty of the tree-level predictions (Berger et al., 2014; Yanai et al., 2023). It is largely agreed that tree-level uncertainty contributes a small fraction of the total uncertainty when aggregated over many individuals for design-based estimates at coarse scales (Breidenbach et al., 2014; Ståhl et al., 2014; McRoberts and Westfall, 2014; Yanai et al., 2023), yet whether or not these small fractions (% contributions) can be considered negligible across scales and estimation contexts is unresolved (Chen et al., 2015; McRoberts et al., 2016; Saarela et al., 2020). Considering the range of reference data uncertainty contributions reported in the model-based estimation literature (5% of uncertainty at pixel-scale Chen et al. (2015); 75% of uncertainty for 5005 km^2 study area Saarela et al. (2020)), it remains prudent to include this component to the extent that it is feasible.

Model-based estimates of pixel- or point-level uncertainty (i.e. uncertainty maps), though helpful to identify the spatial distribution of uncertainties within an area, should be accompanied with methods to aggregate uncertainties to produce estimates at scales relevant to management and decision making like individual forest stands and ownership parcels (desirable criterion 3, defined above). Spatially aggregating predictions can be achieved by averaging or summing all predictions within a given area. However, spatially aggregating uncertainties requires data beyond that which is contained in an uncertainty map (Wadoux and Heuvelink, 2023; McRoberts et al., 2022). Specifically, pairwise prediction or residual covariances and associated spatial correlations are required but do not fit into an uncertainty map (N map pixels yield N^2 pairs/covariances) and may present obstacles. First, computing the necessary prediction or residual covariances for a given area is often computationally expensive; second, obtaining the spatially dense reference data (i.e. short distances between plots) needed to incorporate spatial autocorrelation of residuals is often prohibitive (Breidenbach et al., 2016; McRoberts et al., 2022; Wadoux and Heuvelink, 2023). If estimates are only to be produced for large units of aggregation (e.g. polygons with dimensions well beyond the range of spatial autocorrelation) then the contribution from spatially autocorrelated residuals is typically negligible, however, for scales germane to forest management (i.e. stands and parcels) this component may be highly relevant (Breidenbach et al., 2016; McRoberts et al., 2018, 2022).

McRoberts et al. (2022) considered the feasibility of efficiently producing uncertainty estimates for map subregions, both internally as a map producer but also for map users in satisfaction of desirable criterion 4 (defined above). Their approach was predicated on methods

Table 1
Categorization of four types of model-based uncertainty considered in this study, and how they contribute (by way of variance component) to the estimates of total variance (Var_{total}).

Type	Name	Source of Uncertainty	Variance component
a	Reference data uncertainty ($\widehat{Var_{ref}}$)	Measurement, classification, allometric model, and plot location errors	$\widehat{Var_{boot}}$ *
b	Sampling variability ($\widehat{Var_{sam}}$)	Training the model on a sample	$\widehat{Var_{boot}}$
c	Residual variability	The model is not able to represent all variation in the response variable	$\widehat{Var_{res}}$
d	Auxiliary data uncertainty ($\widehat{Var_{LC}}$)	Variability related to sensor imprecision and inaccuracy or model uncertainty if the data are model generated	$\widehat{Var_{boot}}$

* Bootstrap variance equivalent to the sum of $\widehat{Var_{sam}}$, $\widehat{Var_{ref}}$ and $\widehat{Var_{LC}}$.

for reducing the size of the metadata needed to estimate uncertainty for map subregions to transfer this metadata more easily to users. However, McRoberts et al. (2022) applied their proposed approach only to relatively large areas (2500 km^2 and 12,500 km^2) for which the effects of spatial autocorrelation of residuals were ignored. Further, this approach requires a sophisticated series of operations including the reclassification of pixel-predictions into bins, a lookup of bin covariances in a provided matrix for each pixel pair, and summation of covariances for all pixel pairs in the region of interest.

In this study we built on the methods developed in McRoberts et al. (2022) to produce model-based estimates of uncertainty (standard error; SE) associated with spatial averages of AGB predictions for small ownership parcels (spatial units of analysis sized 0–2023 hectares) in New York State (NYS). We replicated the machine learning modeling framework developed in Johnson et al. (2023) to produce pixel-level AGB predictions. We had three objectives: (1) to estimate a model-based SE that included all four types of uncertainty (Table 1, types a–d), incorporated spatial autocorrelation of model residuals, and used methods that are compatible with algorithmic modeling; (2) to estimate the relative contributions of uncertainty components across scales to help direct further efforts toward reducing total uncertainties in model-based estimates; and (3) to derive a relationship between ownership parcel characteristics (area, perimeter, AGB density, % forest cover) and SE in the form of a multiple regression model to provide a practical, broadly applicable method for uncertainty estimation for map subregions that both map users and producers can implement, satisfying desirable criterion 4.

2. Data and methods

2.1. Overview

We largely followed the steps described in McRoberts et al. (2022) to produce estimates of total variance and standard error (SE) associated with spatial averages of mapped aboveground biomass (AGB) predictions. Within this section we present a general overview of the methods and formulas used to estimate each component of total variance. In subsequent sections (2.2–2.9) we present specific implementation details used in a New York State (NYS) application of these methods where we estimate SEs associated with AGB density estimates for a stratified random sample of small ownership parcels (0–2023 hectares; Fig. 1). The total variance for a given parcel was estimated as the sum of two components, and the SE was computed as the square root of total variance as follows:

$$\mathrm{Var}_{\mathrm{total}} = \widehat{Var_{boot}} + \widehat{Var_{res}} \tag{1}$$

$$\mathrm{SE} = \sqrt{\mathrm{Var}_{\mathrm{total}}} \tag{2}$$

Implicit in these formulas is the assumption that components $\widehat{Var_{boot}}$ and $\widehat{Var_{res}}$ are independent (further discussed in Section 4.3.1).

After computing $\mathrm{Var}_{\mathrm{total}}$ and SE for each parcel in our sample, we further deconstructed the $\widehat{Var_{boot}}$ component and evaluated the relative contributions of each sub-component to the total. Further, we fit a parametric regression model to estimate SE for a given parcel as a function of the following parcel characteristics: area, perimeter, AGB, and forest cover. This regression model was developed to expedite the process of uncertainty estimation for any specified subregion of interest. All data compilation, modeling, analysis, and summary were conducted using the R programming language (R Core Team, 2023) with the targets pipeline tool (Landau, 2021).

2.1.1. Bootstrap variance

This component includes uncertainty in reference data (Table 1, type a), uncertainty due to training a model on a sample (Table 1, type b), and uncertainty related to the landcover mask used in our areal estimates (Table 1, type d). The same bootstrapping approach was used to include each of these three types of uncertainty. Specifically, in each replication i of r replications (r = 1000 in this study), a new training data set was developed as follows (additional details for some steps are provided in the sections or appendices noted in parentheses):

1. A sample was selected with replacement from the original sample of field plots.
2. New tree-level AGB predictions were produced for all trees in the new sample of plots. Residuals were drawn from both empirical and estimated distributions to account for measurement error and allometric uncertainty/error in tree-level biomass predictions (Appendix A).
3. These new tree-level AGB predictions were aggregated within each plot to produce plot-level AGB in units of megagrams per hectare Mg ha^{-1}.
4. A distance value was randomly selected from an estimated distribution of plot-location errors yielding a new plot geometry from which auxiliary data were extracted (Appendix A; Section 2.4; Appendix B).

Next, machine learning (ML) models were fit to each training data set i, and then used to make a set of predictions for the map units (pixels) within each parcel, resulting in r predictions per pixel (Section 2.5). We then generate a new landcover mask for each parcel and replication i and apply each mask to the corresponding set of pixel-predictions such that nonforest pixel-predictions are set to 0 (Mg ha^{-1}) AGB (Section 2.5). The bootstrap variance for a given parcel was then computed as the variance over r replications of the parcel-level AGB density, where AGB density was computed as the average of all pixel-predictions (potentially masked) within the parcel. Specifically, the bootstrap variance was computed with the following equation:

$$\widehat{Var_{boot}} = \frac{\sum_{i=1}^{r} (\hat{u}_i - \hat{\bar{u}})^2}{r-1} \tag{3}$$

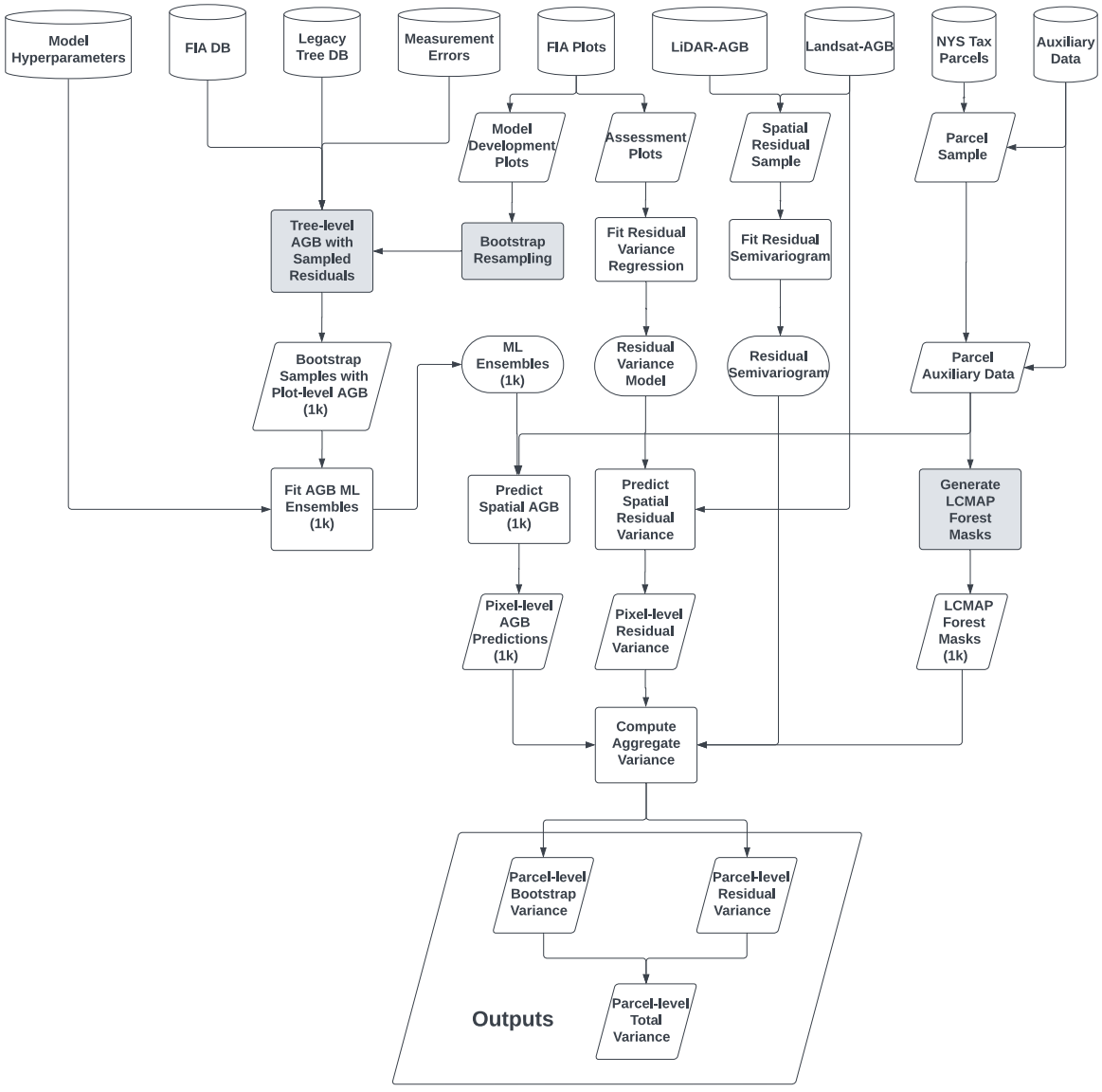

Fig. 1. A flowchart showing the key elements and implementation details of the methodology to estimate uncertainty for spatial means of aboveground biomass. Cylinders represent data sources, parallelograms represent data products and results, rectangles represent processing steps, and ovals represent models. Gray shaded nodes represent components that were omitted for separate runs of the workflow to further deconstruct $\widehat{Var_{boot}}$ by subtraction into $\widehat{Var_{sam}}$, $\widehat{Var_{LC}}$, and $\widehat{Var_{ref}}$.

where r is the number of bootstrap replications, $\hat{\bar{u}}$ is the average parcel-level AGB density over r replications, and $\hat{u}_i$ is the parcel-level AGB density estimated for a given bootstrap replication i and computed as:

$$\hat{u}_i = \frac{\sum_{j=1}^{N} \hat{y}_j}{N} \tag{4}$$

where N is the number of pixels in a given parcel and $\hat{y}_j$ is a potentially masked pixel-level AGB prediction made by the ML ensemble. Derivations for and examples of these bootstrap procedures and equations can be found in Freedman (1981), Efron and Tibshirani (1994), CEOS (2021), and McRoberts et al. (2022).

To better understand the relative contributions of various sources of uncertainty, we further deconstructed $\widehat{Var_{boot}}$ into reference data variance ($\widehat{Var_{ref}}$; Table 1, type a), sample variance ($\widehat{Var_{sam}}$; Table 1, type b), and landcover mask variance ($\widehat{Var_{LC}}$; Table 1, type d) such that:

$$\widehat{Var_{boot}} = \widehat{Var_{ref}} + \widehat{Var_{sam}} + \widehat{Var_{LC}} \tag{5}$$

Each respective sub-component was computed by subtraction. To compute $\widehat{Var_{sam}}$, another r replication simulation was run with a constant training data set across each replication. The resulting estimate of $\widehat{Var_{boot}}$, identified here as $\widehat{Var_{boot-static-sample}}$, was then subtracted from the original estimate of $\widehat{Var_{boot}}$ to produce $\widehat{Var_{sam}}$ as follows:

$$\widehat{Var_{sam}} = \widehat{Var_{boot}} - \widehat{Var_{boot-static-sample}} \tag{6}$$

Similarly, to compute $\widehat{Var_{LC}}$, another r replication simulation was run with a constant landcover mask (Section 2.5) across each replication. The resulting estimate of $\widehat{Var_{boot}}$, identified here as $\widehat{\text{Var}_{\text{boot-static-LC}}}$, was then subtracted from the original estimate of $\widehat{Var_{boot}}$ to produce $\widehat{Var_{LC}}$ as follows:

$$\widehat{Var_{LC}} = \widehat{Var_{boot}} - \widehat{Var_{boot-static-LC}} \tag{7}$$

Once again, to compute $\widehat{Var_{ref}}$, another r replication simulation was run with constant plot-level AGB predictions and plot geometries used as reference data across each replication. The resulting estimate of $\widehat{Var_{boot}}$, identified as $\widehat{Var_{boot-static-ref}}$ was then subtracted from the

original estimate of $\widehat{Var_{boot}}$ to produce $\widehat{Var_{ref}}$ as follows:

$$\widehat{Var_{ref}} = \widehat{Var_{boot}} - \widehat{Var_{boot-static-ref}} \quad (8)$$

Finally, the difference between $\widehat{Var_{boot}}$ and the sum of $\widehat{Var_{sam}}$, $\widehat{Var_{LC}}$, and $\widehat{Var_{ref}}$ was proportionally subtracted from each variance component ($\widehat{Var_{sam}}$, $\widehat{Var_{LC}}$, $\widehat{Var_{ref}}$) to ensure that the sum of components was equal to $\widehat{Var_{boot}}$ (Saltelli et al., 2004).

2.1.2. Residual variance

This component contributes to uncertainty due to the model failing to represent all of the variation in the response variable (Table 1, type c). In estimating this component we treat the model building sample, the dependent variables, and independent variables as fixed; the model residuals are thus conditioned on the observed (fixed) case. Because model residuals exhibited heteroskedasticity (non-constant residual variability over the range of model predictions), we fit a model to estimate residual variance as a function of our AGB predictions (Section 2.6). When model residuals are homoskedastic, the sample residual variance estimate may be used for all population units (McRoberts et al., 2022). We used this model to make pixel-level residual variance estimates that aligned with individual AGB predictions. We then aggregated pixel-level residual variance to each parcel as follows:

$$\widehat{Var_{res}} = \frac{\sum_{j=1}^{N} \hat{\sigma_i}^2}{N^2} + \frac{\sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} \hat{\sigma_i}\hat{\sigma_j}\hat{\rho_{ij}}}{N^2} \quad (9)$$

where $\hat{\sigma_i}^2$ is the estimated residual variance for a given pixel, N is the number of pixels in a given parcel, and $\hat{\rho_{ij}}$ is the estimate of spatial correlation between pixel-level residuals defined as

$$\hat{\rho_{ij}} = \frac{sill - \gamma(h_{ij})}{sill} \quad (10)$$

where γ is a semivariogram function, h_{ij} is the distance (scalar) between pixels i and j, and $sill$ is the semivariance ceiling estimated from the semivariogram. In Eq. (10) and throughout this manuscript we assume second-order stationarity and isotropy. Derivations for Eq. (9) can be found in McRoberts et al. (2022), and derivations for Eq. (10) can be found in Wadoux and Heuvelink (2023) and Webster and Oliver (2007).

The semivariogram function used in Eq. (10) (further described in Section 2.7) was fit to a simple random sample of spatial residuals (pixels) computed as follows:

$$\hat{e_i} = \hat{y}_{i-ML} - \hat{y}_{i-ref} \quad (11)$$

where $\hat{y}_{i-ML}$ is an AGB prediction from the ML model (Section 2.5), and $\hat{y}_{i-ref}$ is an AGB value from a co-located reference dataset.

2.2. Study area and parcel sample

NYS covers 141,297 km^2 in the northeastern US and was approximately 59% forested as of 2019 (USDA Forest Service, 2020). The forests are dominated by northern hardwoods-hemlock types but include Appalachian oak and beech-maple-basswood forests in the western and southern regions of the state respectively (Dyer, 2006). The majority of NYS forests are privately owned (~73%; USDA Forest Service (2020)) and contained within parcels smaller than 40 ha (~65%; L'Roe and Allred (2013)). This already fragmented forest landscape is increasingly undergoing subdivision (L'Roe and Allred, 2013).

To limit computational scope we selected a stratified random sample of ownership parcels within NYS. The statewide parcel database, compiled for property taxation, was obtained under agreement with the NYS ITS Geospatial Services. We stratified the statewide parcels into ten groups based on size, and five groups based on percent forest cover as determined by the United States Geological Survey's Land Change Monitoring, Assessment, and Projection (LCMAP) version 1.2 primary classification products (Zhu and Woodcock, 2014; Brown et al., 2020). Size strata were selected to ensure a sample that represented a wide range of parcels, with an emphasis on smaller parcel sizes that dominate the NYS ownership landscape and are underrepresented in uncertainty estimation studies. Size strata were constructed as intervals (left-exclusive, right-inclusive) of parcel acreage as follows: (0, 5], (5, 10], (10, 20], (20, 50], (50, 100], (100, 250], (250, 500], (500, 1000], (1000, 2500], (2500, 5000]. We have converted these size strata descriptions to hectares for use throughout this manuscript as follows: (0, 2], (2, 4], (4, 8], (8, 20], (20, 40], (40, 101], (101, 202], (202, 405], (405, 1012], (1012, 2023]. Forest cover strata were selected such that a range of forest conditions were represented within each size stratum. Forest cover strata were constructed as five equal intervals from 0 (exclusive) to 100 (inclusive) percent. Forest cover was defined as the combination of the tree cover, wetlands, and grass/shrub LCMAP classes (Johnson et al., 2023). 50 parcels were randomly sampled from each size and cover group. For groups without 50 parcels, all available parcels were selected. Overall, 2224 parcels were sampled across the state (Fig. 2).

2.3. Field data

Field plot AGB predictions from the United States Forest Inventory and Analysis (FIA) program (Bechtold and Patterson, 2005; Gray et al., 2012) were used as reference data in Johnson et al. (2023); we replicated the same approach here. True plot centroids for each plot in NYS (obtained via agreement with the FIA program office) were randomly partitioned into model development (80%) and map assessment (20%) sets. The model development set was further partitioned into an 80% training set, and a 20% validation set used for iterative assessment during the Johnson et al. (2023) model development process (i.e. not used in this study). We used the model development partition to compute $\widehat{Var_{boot}}$ (Section 2.1.1). This dataset included 1954 completely forested plots, as well as 95 completely nonforested plots that were identified as true zeros (AGB) based on LiDAR-derived maximum heights ≤ 1 m (Johnson et al., 2022), for a total of 2049 plots inventoried between 2002 and 2019. We used the map assessment set to estimate the pixel-level residual variances (Section 2.6) that contribute to $\widehat{Var_{res}}$ (Section 2.1.2). This dataset contained 545 plots inventoried in 2007, 2012, 2018, and 2019, collectively representing a complete spatial coverage of NYS. When plots in model development or map assessment sets were inventoried more than once, single instances were randomly selected so that each plot was associated with only one inventory year. Refer to Johnson et al. (2023) for further details on the FIA plot design and the plot selection criteria used here.

Rather than using the tree-level AGB predictions provided in the FIA database, we replicated FIA's component-ratio-method (CRM; Woodall et al. (2011)) to incorporate measurement error, misclassifications, and allometric uncertainty in the plot-level predictions. At the time of writing, FIA had transitioned to using a new system of allometric models (Westfall et al., 2023). However, our use of CRM remained consistent with the methods used in Johnson et al. (2023). A flowchart describing data inputs, component models, and final predictions for CRM is included in Appendix C.

For every bootstrap replication we selected a set of plots with replacement from the model development partition, yielding as many sets of plots as bootstrap replications. Further, for each plot and bootstrap replication we constructed a plot polygon incorporating location uncertainty and computed plot-level biomass incorporating both measurement and classification errors as well as allometric uncertainty. We used published statistics from Yanai et al. (2023), Harmon et al. (2011), and Ross (2021) to build normal distributions of measurement and classification errors. We used the legacy tree database (Radtke et al., 2015) to build empirical distributions of allometric prediction errors. We used recordings of FIA plot locations repeated over multiple field visits to build a normal distribution of plot location errors. Further descriptions of our reference data uncertainty estimation approach are provided in Appendix A.

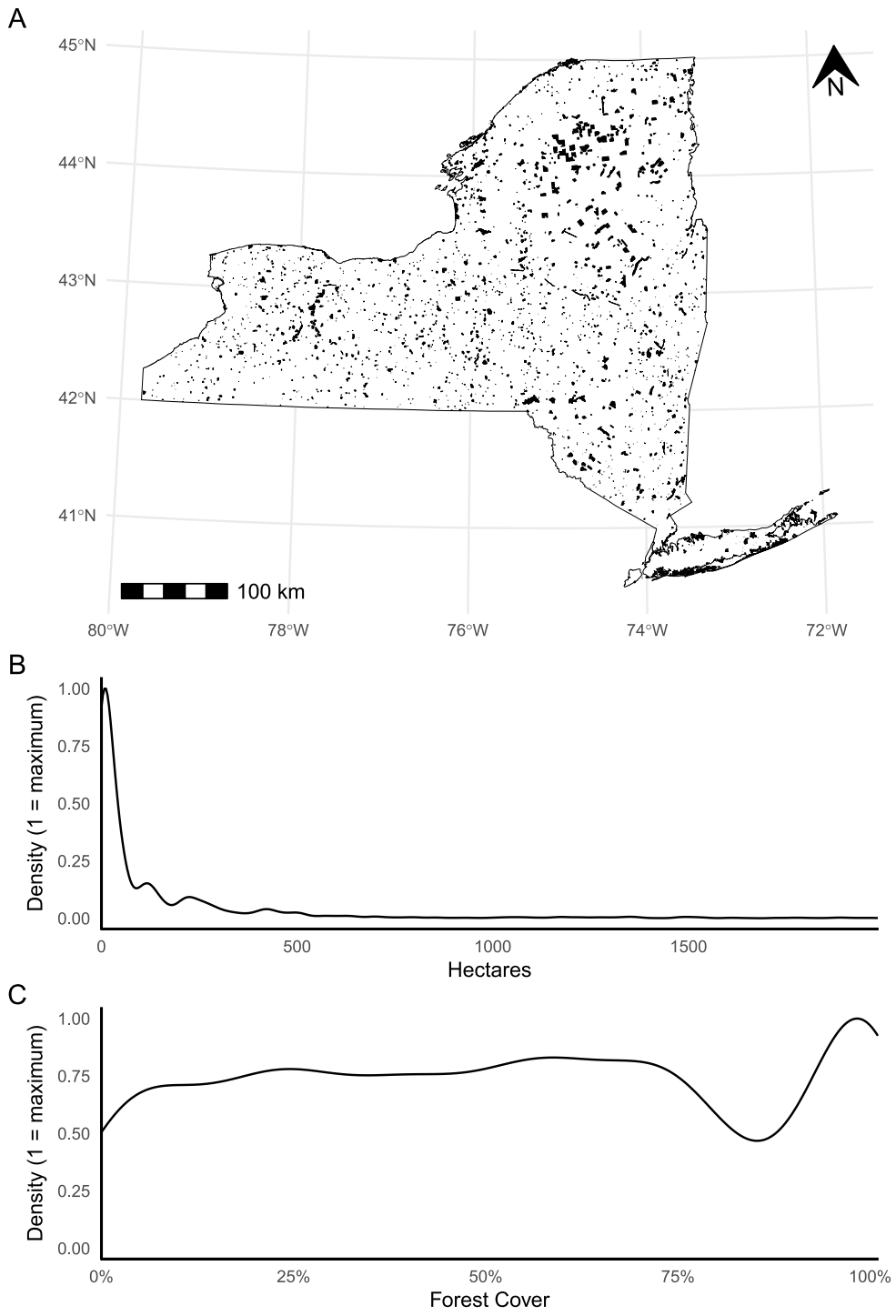

Fig. 2. Parcel sample summary. (A) Spatial distribution of sampled parcels within New York State. (B) Smoothed frequency distribution of parcel sizes (hectares) sampled. (C) Smoothed frequency distribution of forest cover (%) sampled. Distributions for B and C have been rescaled separately, such that the most common occurrences of each variable are assigned a value of 1.

2.4. Auxiliary data

We used the same set of 29 predictor layers derived in Johnson et al. (2023) for model training and mapping. These predictors included temporally segmented, gap-filled, and smoothed annual Landsat timeseries imagery and disturbance metrics (Landtrendr; Kennedy et al. (2010, 2018b)), LCMAP primary and secondary landcover classifications (Zhu and Woodcock, 2014; Brown et al., 2020), and static ecological, climatic, and topographic datasets. Further information describing the auxiliary data can be found in Appendix B and Johnson et al. (2023).

2.5. Aboveground biomass models, maps, and estimates

Following the 'direct' approach described in Johnson et al. (2023), ML ensemble models were fit to the resampled plots and corresponding plot-level AGB predictions for each bootstrap replication (Section 2.1.1; Section 2.3). Information describing the ML ensembles and hyperparameter tuning can be found in Appendix D and Johnson et al. (2023). To compute $\widehat{Var_{boot}}$, an estimate of mean AGB for the parcel was computed for each replication of the bootstrap (Section 2.1.1; Eq. (3)). To this end, separate sets of AGB predictions were made for all pixels

within each sampled parcel using the models trained for each bootstrap replication. Further, a separate forest mask was created and applied to each parcel for each bootstrap replication to incorporate uncertainty and error in the LCMAP primary landcover classifications (Zhu and Woodcock, 2014; Brown et al., 2020). We emphasize that the parcel is still the target area of estimation, but the portion of the parcel that has forest cover, and therefore AGB, is uncertain. New forest masks were produced at each replication by generating a uniform random number between 0 and 1 for each pixel. If this number was greater than the user's accuracy rate (Stehman et al., 2021) for the pixel's initial classification (LCPRI), then the LCMAP secondary classification (LCSEC) was used instead. After the new mask was generated, the mask was used to identify pixels that were not classified as tree cover, wetlands, or grass/shrub (forest definition from Johnson et al. (2023)) and set their AGB predictions to 0. After masking, and for each bootstrap replication, parcel level means (Eq. (4)) were computed.

2.6. Residual variance model

Pixel-level residual variance was required to estimate $\widehat{Var_{res}}$ (Section 2.1.2; Eq. (9)). However, residuals were only computed for a sample of pixels coinciding with FIA plots selected for the assessment partition from Johnson et al. (2023) (Section 2.3). Further, assigning the residual variance computed from this sample to every pixel would incorrectly assume homoscedasticity. To overcome this, we fit models of estimated residual variance as a function of predicted AGB. Specifically, we evaluated linear, log-linear, log–log, third order polynomial, and natural cubic spline (2 knots; Hastie (1992)) model forms. We chose the model that best fit the data as indicated by the coefficient of determination (R^2; computed after backtransformation and bias-correction) and used this final model to estimate the residual variance for each pixel in our parcel sample. Additional information describing our residual variance modeling approach is provided in Appendix E.

2.7. Spatial autocorrelation of residuals

An estimate of spatial autocorrelation between model residuals was required to estimate $\widehat{Var_{res}}$ (Section 2.1.2; Eq. (9)). However, the sample of assessment FIA plots used in Johnson et al. (2023), with at most one FIA plot per 2,400 ha in NYS (Bechtold and Patterson, 2005), was too sparse to provide the necessary information. We instead used LiDAR-based predictions of AGB developed in Johnson et al. (2022) as an alternative form of reference data with which to compute residuals ($\hat{y}_{i-ref}$ in Eq. (11)). We acknowledge that these predictions are of lesser quality than field-measured data, but they offered improved accuracy relative to the Landsat-based models assessed here, and they provided the spatial density and coverage that can represent local patterns of variability in forests. To our knowledge, these two important qualities are not offered by any other AGB dataset within NYS. We first calculated an empirical semivariogram which we then used to fit several model semivariograms before choosing the best fitting model to use in Eq. (10). Further descriptions of our approach to computing spatial residuals and fitting semivariogram models are provided in Appendix F. To assess the sensitivity of our uncertainty estimates relative to the estimate of spatial autocorrelation between our LiDAR-based residuals, we computed four additional estimates of $\widehat{Var_{res}}$ (Eq. (9)) where we multiplied pixel pairwise distances by 0.25, 0.5, 2, and by the estimated range of the semivariogram. In effect, these additional estimates of $\widehat{Var_{res}}$ reflect residual spatial autocorrelation multiplied by a factor of 4, 2, 1/2, and 0 respectively.

2.8. Modeling estimated standard error (total variance) as a function of parcel characteristics

To increase the computational efficiency and practical utility of estimating spatially aggregated uncertainty for parcels or other subregions of interest, we modeled parcel SE as a function of parcel characteristics that we expect are readily available to map users. Specifically, we included AGB density, % forest cover, area, and perimeter of each parcel as predictors of a parcel's SE given that these predictors are straightforward to compute from an area of interest polygon, the AGB prediction raster, and a co-located forest mask raster. Forest cover was defined as the combination of the tree cover, wetlands, and grass/shrub LCMAP classes (Johnson et al., 2023), and AGB density (Mg ha^{-1}) was computed as the average of pixel predictions within a parcel (Eq. (4)).

We fit four candidate models: multiple regression models with and without natural log transformations of the dependent and independent variables. Each of the candidate models were fit to a randomly selected 80% of the parcel sample (training set) and were compared on the basis of R^2. The candidate model with the best R^2 statistic was selected as the final model. The model fitting and assessment process is further described in Appendix G.

To assess the validity of our SE estimates and the coverage of corresponding 95% confidence intervals (CI) we compared parcel-level AGB estimates made with LiDAR-based models developed in Johnson et al. (2022) to parcel-level AGB estimates and associated 95% CIs made with Landsat-based models (investigated here; Section 2.5; Johnson et al. (2023)) and SE estimates from our final SE model. We compared estimates for 255 parcels within our parcel testing set that were entirely contained within the spatial footprint of LiDAR collections mapped in Johnson et al. (2022) (Appendix H). The comparison process is described in Appendix I.

2.9. Bootstrap variance stability

To assess the stability of $\widehat{Var_{boot}}$ after 1000 bootstrap replications, we compared parcel-level estimates of $\widehat{Var_{boot}}$ made after each successive replication i ($\widehat{Var_{boot\_i}}$) to the corresponding parcel-level estimate made after 1000 replications ($\widehat{Var_{boot\_1k}}$). Following McRoberts et al. (2023), we investigated stability at several thresholds: within 5%, 2.5%, and 1% of $\widehat{Var_{boot\_1k}}$. For each number of replications we calculated the proportion of parcels in our sample that yielded $\widehat{Var_{boot\_i}}$ estimates within each threshold distance from $\widehat{Var_{boot\_1k}}$.

3. Results

3.1. Reference data uncertainty

Comparing the variability of tree- and plot-level AGB predictions provides insight into how allometric uncertainty and measurement error scales from individual trees to aggregations of many individuals within a plot. Over the 1000 bootstrap replications tree-level AGB predictions varied more than plot-level AGB predictions (Appendix A; Fig. A.3). The average standard deviation (SD) and coefficient of variation (CV) of tree-level AGB predictions were 65.89 kg and 34.96%, while the average SD and CV for plot-level AGB predictions were 7.47 Mg ha^{-1} and 7.20%. This result was unsurprising because plot-level predictions are simply area-normalized sums of tree-level predictions, and aggregating individuals can smooth out extreme individual variability as predictions above and below respective means will offset to an extent.

Softwood and hardwood trees varied similarly in absolute terms (SD), but softwoods varied substantially more in relative terms (CV). We posit that this discrepancy in CVs is a result of the much wider distribution of softwood bole volume residuals compared to that of hardwoods (Appendix A). There were only 53 records for hardwood bole volume residuals in the legacy tree database (Radtke et al., 2015) whereas there were 473 records for softwood bole volume residuals. Bole volume predictions, which drive bole biomass predictions, likely have the largest impact on total tree-level biomass predictions among tree component in the CRM framework; bole biomass is the most substantial of tree components and is used to adjust Jenkins-based component (top and branches, stump) predictions before summing to total tree-level biomass (Woodall et al. (2011); Appendix C).

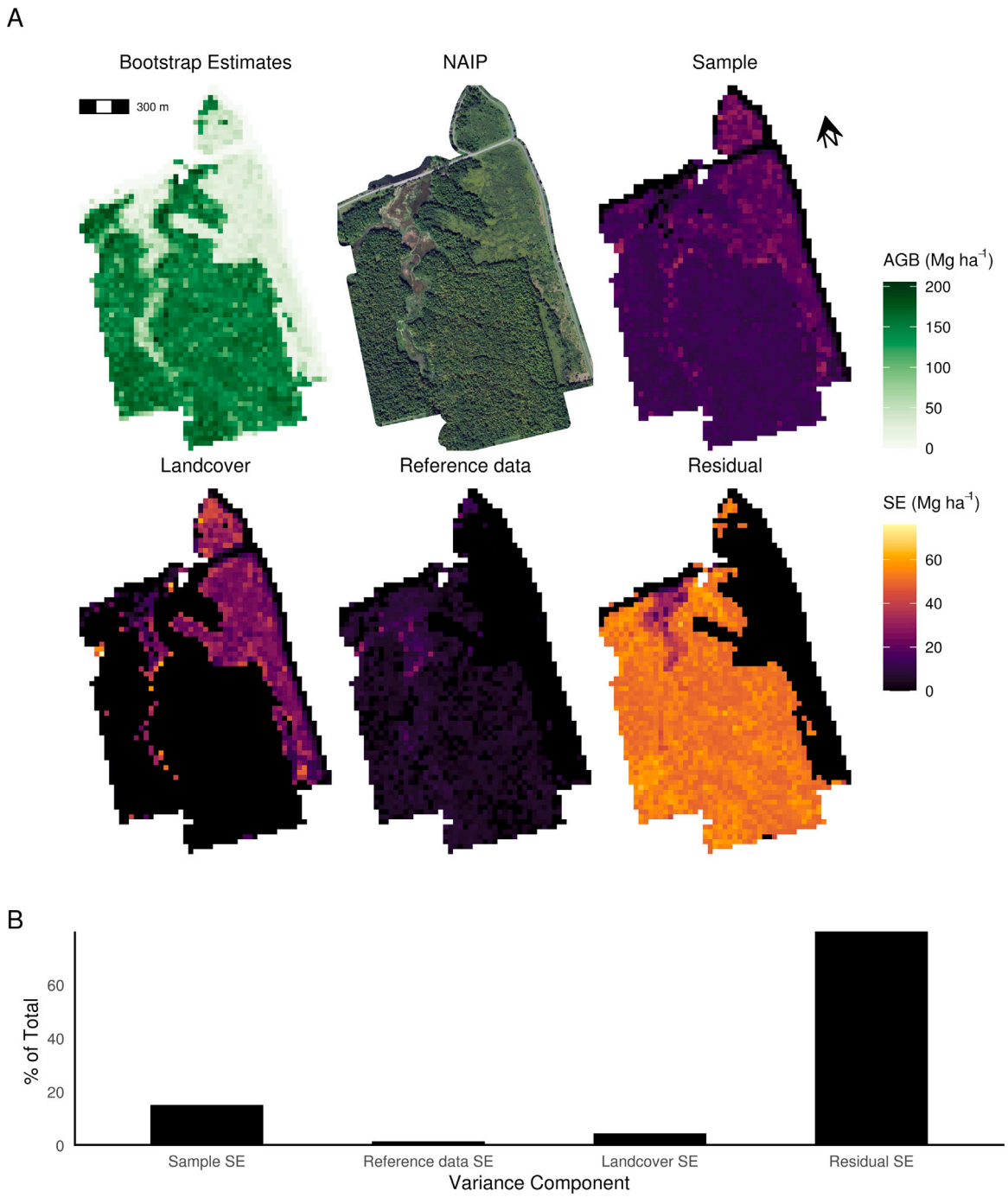

Fig. 3. Aggregating pixel-level uncertainties to produce parcel-level standard error (SE). (A) AGB estimates computed as the mean prediction over all bootstrap replications (Bootstrap Estimates); National Aerial Imagery Program (NAIP) orthophotography collected in 2019 for additional landscape reference (EROS, 2019); Sampling standard error (SE), landcover SE, reference SE, and residual SE estimates. (B) Relative contributions (%) to total SE for the area of interest in A.

3.2. Spatial autocorrelation of residuals

A nested combination of two Matern model semivariograms demonstrated the best fit to our empirical semivariogram (Section 2.7; Appendix F; Fig. F.1) and was selected to provide estimates of spatial correlation between model residuals in our calculation of $\widehat{Var_{res}}$ (Eq. (9)). Residual semivariance increased from 0 $(\text{Mg ha}^{-1})^2$ for neighboring pixels (no nugget), to 1338.35 $(\text{Mg ha}^{-1})^2$ (sill). Matern semivariogram functions approach the sill asymptotically. The effective range (where model semivariance >95% of sill) was roughly 1212 m, but we also computed a heuristic range (5779 m) as the lag distance for which the model semivariance is greater than the sill rounded down to the nearest whole integer (1338 $[\text{Mg ha}^{-1}]^2$). In consideration of computational efficiency, we assume that the spatial autocorrelation ($\hat{\rho_{ij}}$; Eq. (10)) for lag distances beyond this heuristic range is 0, thus avoiding residual spatial autocorrelation computations for pixel pairs separated by very large lag distances. We note that users replicating this methodology should choose their own heuristic range (if appropriate) that balances statistical rigor with computational efficiency given the context of their specific datasets and goals.

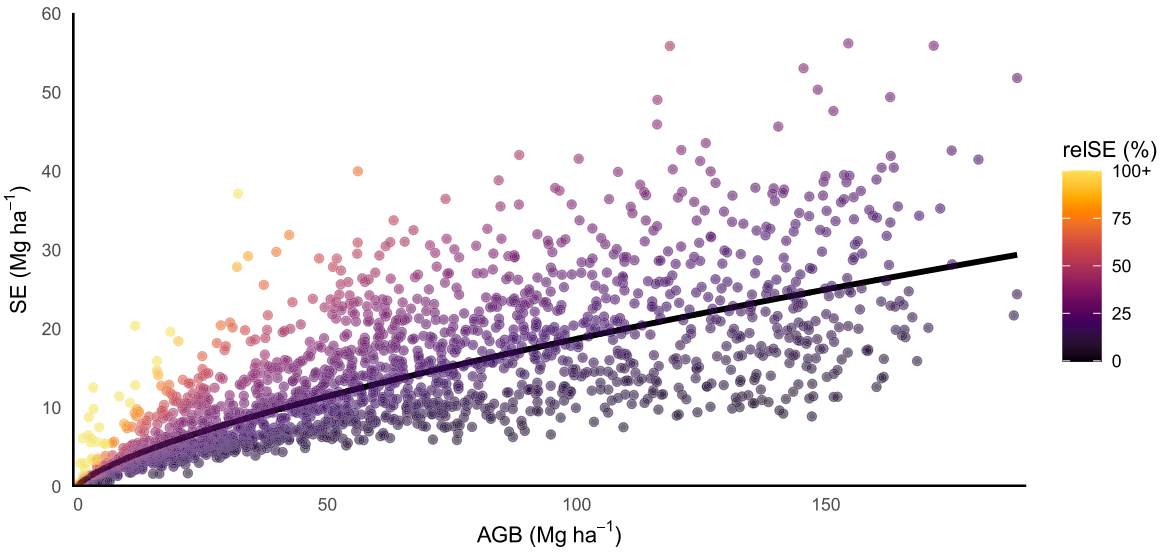

Fig. 4. Parcel-level standard error (SE) as a function of AGB density. Each point represents a parcel from the training partition of the parcel sample, and each point is shaded by the relative standard error (relSE; [SE / AGB] · 100). Parcels with relSE >100% were set to 100% for visualization. The black trend line was fit to data using a natural log–log regression.

3.3. Pixel-level uncertainty

The natural cubic spline model (Section 2.6) fit the grouped data best and explained 55% of the variability in residual variance (coefficient of determination; R^2; Appendix E; Fig. E.1). Maps of pixel-level AGB estimates and co-located residual SEs (Fig. 3A) demonstrated the modeled relationship spatially, with regions of greater AGB density yielding larger residual SE estimates. Conversely, the largest sample and landcover SEs co-occurred within areas of consistently small AGB estimates (Fig. 3A). This sample SE pattern may be explained by the relative lack of model training data at the lower extreme of the distribution (Johnson et al., 2023) while the landcover SE pattern may be explained by the inclusion of LCMAP classification uncertainty in our bootstrapping procedure (Section 2.5). Marginal or transitional lands with relatively low AGB density are among the most challenging to accurately classify (Grass/Shrub and Cropland; Stehman et al. (2021)) and may be intermittently masked to 0 AGB across bootstrap replications. Pixel-level reference data SE was consistently near zero.

3.4. Spatially aggregated uncertainty estimates

Parcel-level estimates of standard error (SE; Eq. (2)) ranged from near 0 Mg ha^{-1} to ~60 Mg ha^{-1} but varied substantially with parcel size, forest cover, and AGB density (Figs. 4; 5). Although the range of SE was large, the majority of relative standard errors (relSE = SE / AGB ·100) were ≤ 25% and only exceeded 50% for ~7% of parcels in the training partition (Fig. 4). Although absolute SE scales with AGB density, relSE, which normalizes SE by AGB, shows the opposite pattern: parcels with the smallest AGB density yielded the largest relSEs.

Given that the formula for $\widehat{Var_{res}}$ (Eq. (9)) contained N^2 (N = number of pixels) in the denominator, it is intuitive that Var_{total} (Eq. (1)) and SE (Eq. (2)) decreased as a function of parcel size (Fig. 5); the larger a parcel, the more 30 m pixels contained therein. Further, as parcel size grew towards the range of spatial correlation and average pixel distances increased, it follows that $\widehat{Var_{res}}$ decreased accordingly, and we might expect that $\widehat{Var_{res}}$ will eventually become negligible for parcels larger than those considered in this study (Fig. 6). Forest cover impacted parcel SE in the opposite direction (Fig. 5), where increasing forest cover resulted in increased standard error. Non-forest pixels resulted in both 0 AGB and 0 residual variance; thus the greater number of these pixels in a parcel, the less forest cover and the smaller the contributions of $\widehat{Var_{res}}$ to Var_{total}. However, the relationship between forest cover and SE may have been somewhat weakened by the impact of LCMAP classification uncertainty ($\widehat{Var_{LC}}$), where the relatively poor accuracy of cropland and grass/shrub classified pixels may have yielded greater bootstrap variances as pixels were intermittently masked to 0 AGB across bootstrap replications (Stehman et al. (2021); Section 2.5; Section 3.3).

$\widehat{Var_{res}}$ was greater than all other variance components across all size and forest cover groups (Fig. 6). The contribution of $\widehat{Var_{sam}}$ began approaching the contribution of $\widehat{Var_{res}}$ for the largest parcels (>20% contribution for parcels 1012 to 2023 hectares) where spatial autocorrelation of residuals had a diminished impact. $\widehat{Var_{LC}}$ contributed substantially (≥20%) for the least forested parcels, but contributed ≤10% for all but the smallest forest parcels. $\widehat{Var_{ref}}$ contributed the least of any other component for all but the largest and most forested parcels (405–2023 hectares; 80%–100% forested), and was <10% of Var_{total} across all groupings.

Parcel-level estimates of Var_{total} (and subsequently SE) were sensitive to the strength of residual spatial autocorrelation (Section 2.7; Fig. 7). Adjustment factors that strengthened residual autocorrelation had the smallest impact for small parcels, where the majority of pairwise pixel distances were short and residual spatial autocorrelation already had a near-maximum impact. On the other hand, strengthening residual autocorrelation increased Var_{total} by ≥30% for all size groups ≥8 hectares. Halving the strength of residual spatial correlation reduced Var_{total} by 14.8% to 46.6% and removing residual spatial correlation altogether reduced Var_{total} by 51.1% to 84.4% across size groups. This highlights the importance of residual spatial autocorrelation as a source of uncertainty when estimating SE for small areas.

3.5. Modeling estimated standard error (total variance) as a function of parcel characteristics

In this section we describe the multiple regression model chosen to estimate standard error as a function of readily available parcel characteristics (area, perimeter, AGB density, forest cover; Section 2.8) and provide accuracy results for this model using a 20% holdout partition of parcels (testing set). The natural log–log regression (Eq. (G.2)) achieved the best R^2 value among the candidate models and accurately estimated SEs for parcels in the test set (Fig. G.1, Table G.1; RMSE = 2.05, MAE = 1.28, ME = −0.21, R^2 = 0.95). The estimated coefficients (Table 2) indicate that AGB density was the most important predictor

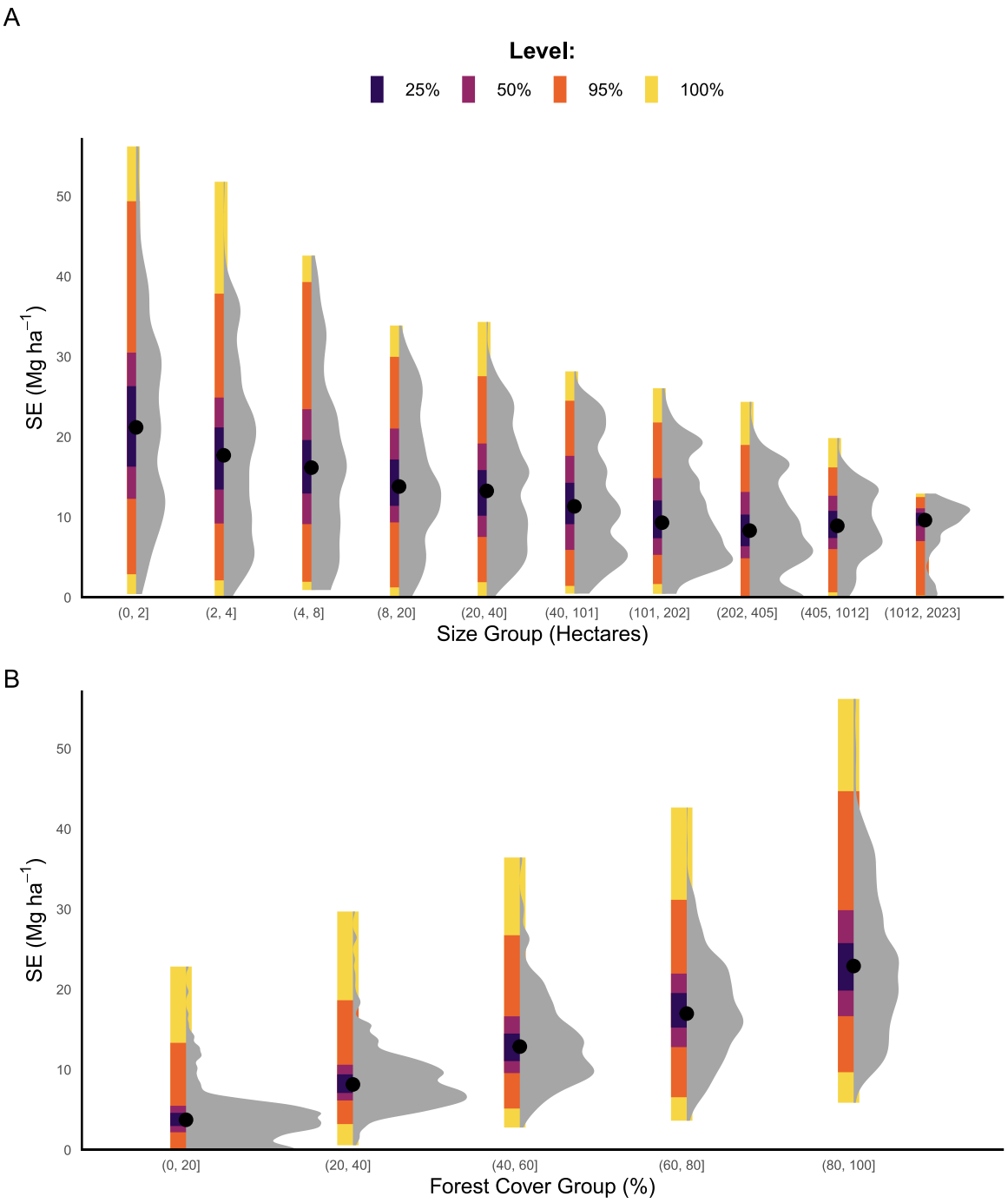

Fig. 5. Parcel-level standard error (SE) distributions. Gray shaded areas represent smoothed frequency distributions of SEs, black dots identify median values, and colored bars show 25%, 50%, 95%, and 100% intervals (percentiles). (A) Distributions per size group (hectare). (B) Distributions per forest cover group (%). All parcels in the training partition of the parcel sample are included. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of SE, which follows the relationship inherent to our residual variance model (Section 2.6; Appendix E) and is corroborated by both pixel-level results (Fig. 3) and relSE results (Fig. 4). Parcel area, perimeter, and % forest cover were similarly important predictors in the model. Using this model we estimated parcel-level SE for all ownership parcels in Essex County, New York State (NYS; Fig. 8).

Of the 255 parcels from the testing set that are within the footprint of the LiDAR collections processed in Johnson et al. (2022) (Appendix H), 227 of them yielded LiDAR-AGB estimates that were contained within our estimated 95% confidence intervals (Fig. 9). This indicates an 89% empirical coverage level for the confidence intervals estimated with our natural log–log regression.

Table 2
Results of a natural log–log regression model of estimated parcel-level standard error as a function of parcel characteristics (AGB, Perimeter, Area, and % Forest).

	Coefficient	Standard error	t	p
Intercept	1.075	0.081	13.3	<0.001
ln(AGB)	0.621	0.005	121.4	<0.001
ln(Perimeter)	−0.162	0.013	−12.7	<0.001
ln(Area)	−0.089	0.008	−11.8	<0.001
ln(% Forest)	0.136	0.004	31.8	<0.001

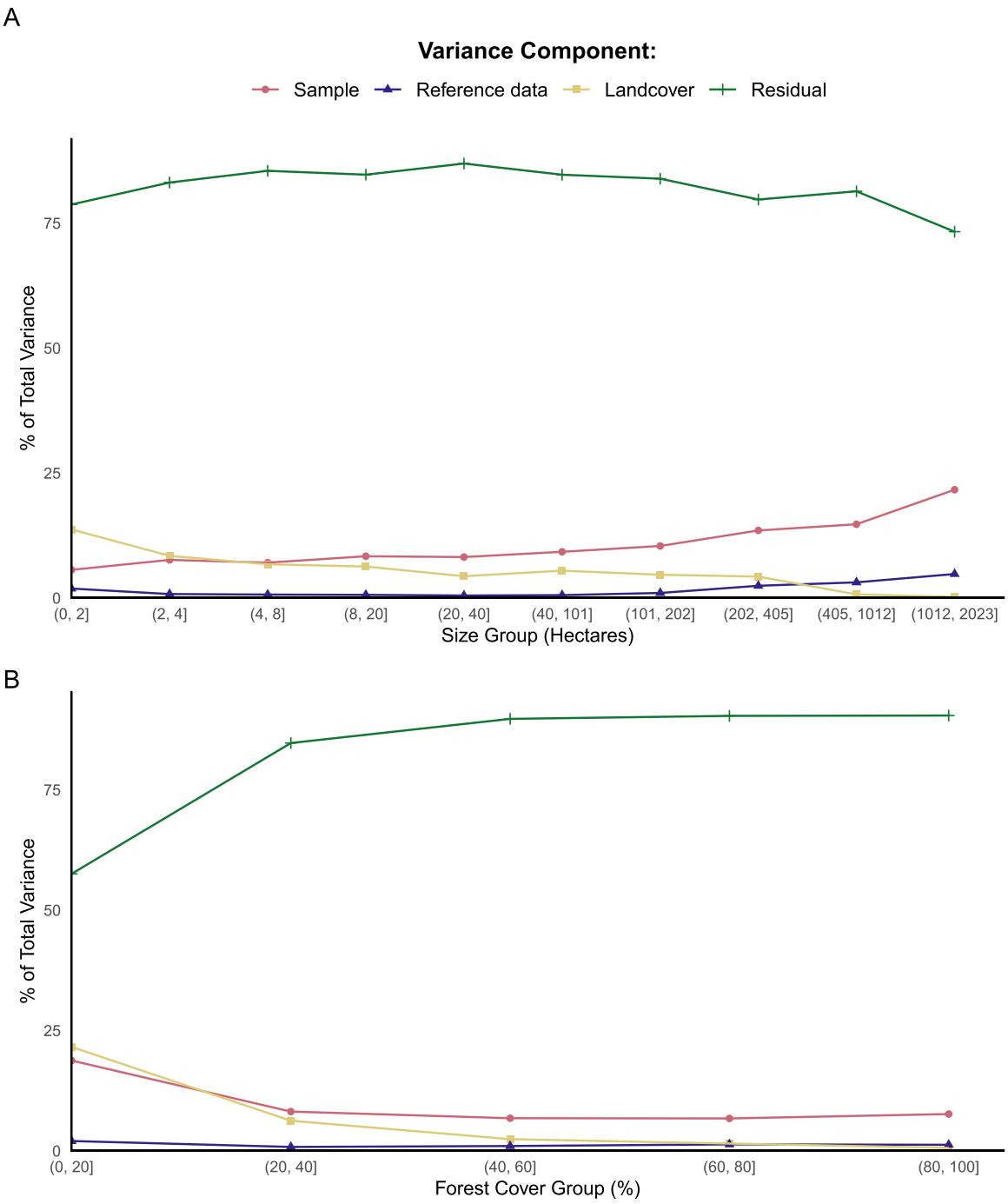

Fig. 6. Uncertainty contributions (% of total variance) as a function of parcel size (A) and forest cover (B) for the training partition of the parcel sample. For each source of uncertainty, the average percent of the total variance was summarized across all parcels within each group.

3.6. Bootstrap variance stability

Stability tests indicated that 1000 replications provides stable estimates of $\widehat{Var_{boot}}$ at the 5% threshold (Appendix J). 95% of parcels were within 5% of $\widehat{Var_{boot\_1k}}$ for the final 193 replications. 95% of parcels were within 2.5% of $\widehat{Var_{boot\_1k}}$ for the final 62 replications, and 95% of parcels were within 1% of $\widehat{Var_{boot\_1k}}$ for only the last 8 replications.

4. Discussion

In this study we produced model-based estimates of uncertainty (standard error; SE) for small-area (0–2023 hectares) spatial averages of forest aboveground biomass (AGB) maps. Our SE estimation approach uniquely included all four types of uncertainty (Table 1), was compatible with algorithmic models (machine learning or nonparametric), and incorporated residual spatial autocorrelation. We found that residual variance ($\widehat{Var_{res}}$; Eq. (9)) dominated other variance components for all parcel sizes (0–2023 hectares), owing largely to contributions from residual spatial autocorrelation (Fig. 7). These results underscore the necessity of incorporating residual spatial autocorrelation when estimating uncertainty at scales germane to forest management (e.g. parcel sizes in New York state). We demonstrated that a log–log regression model for estimating SE as a function of parcel characteristics was very accurate, opening the door to more computationally efficient and practical uncertainty estimation for map subregions. Overall, these

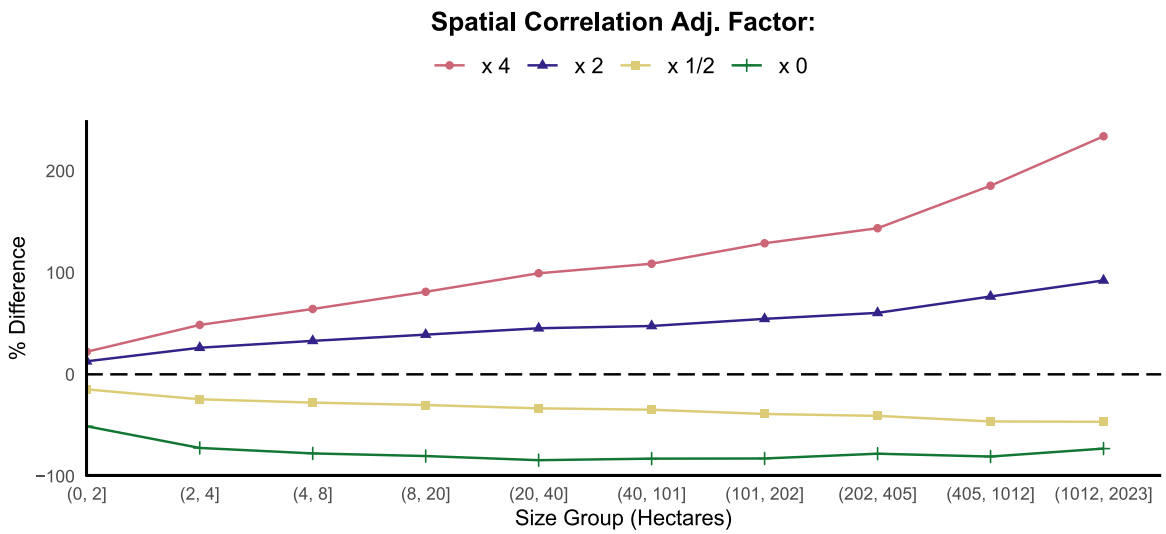

Fig. 7. Sensitivity of total variance estimates to strength of residual spatial autocorrelation. Percent difference between total variance adjusted for strengthened/weakened spatial correlation of residuals (adjustment factor ≠ 1) and total variance computed regularly (adjustment factor = 1; dashed black line) as a function of parcel size (size group) for the training partition of the parcel sample. The average percent difference was summarized across all parcels within each group. Spatial correlation adjustment factors are used to modify the pairwise pixel distances (e.g. adjustment factor 4 is implemented by dividing the pairwise pixel residual distances by 4) such that the adjustment factor effectively describes a multiplicative adjustment to the empirical spatial correlation of residuals.

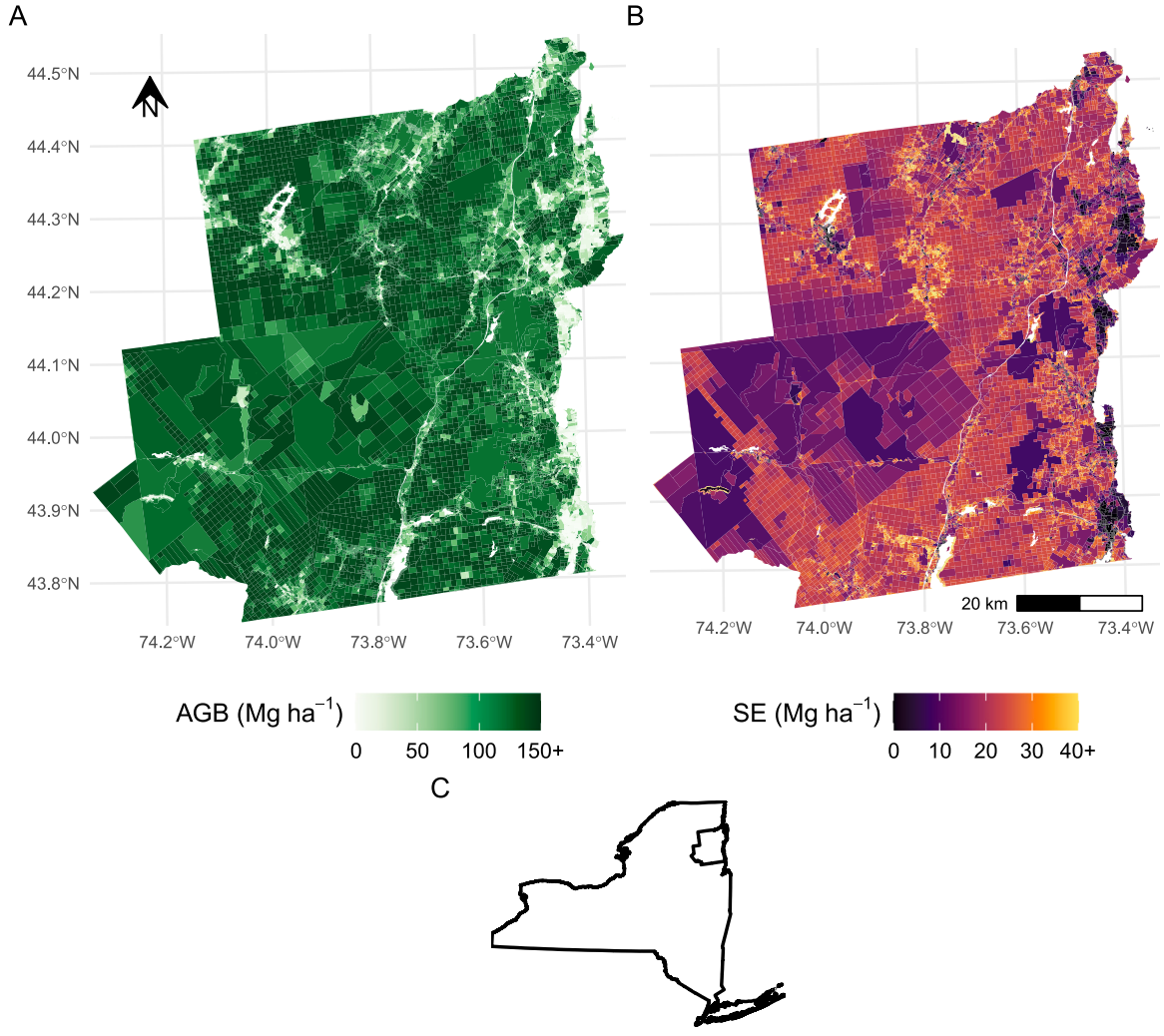

Fig. 8. Estimates of aboveground biomass density and associated standard error for ownership parcels in Essex County, NYS. (A) Parcel-level AGB density computed as the mean of pixel-predictions within each parcel. (B) Parcel-level SE estimated from a natural log–log regression with parcel area, perimeter, AGB density, and % forest cover as independent variables. (C) Location of Essex County within NYS. AGB density and SE values truncated at 150 and 40 respectively for display.

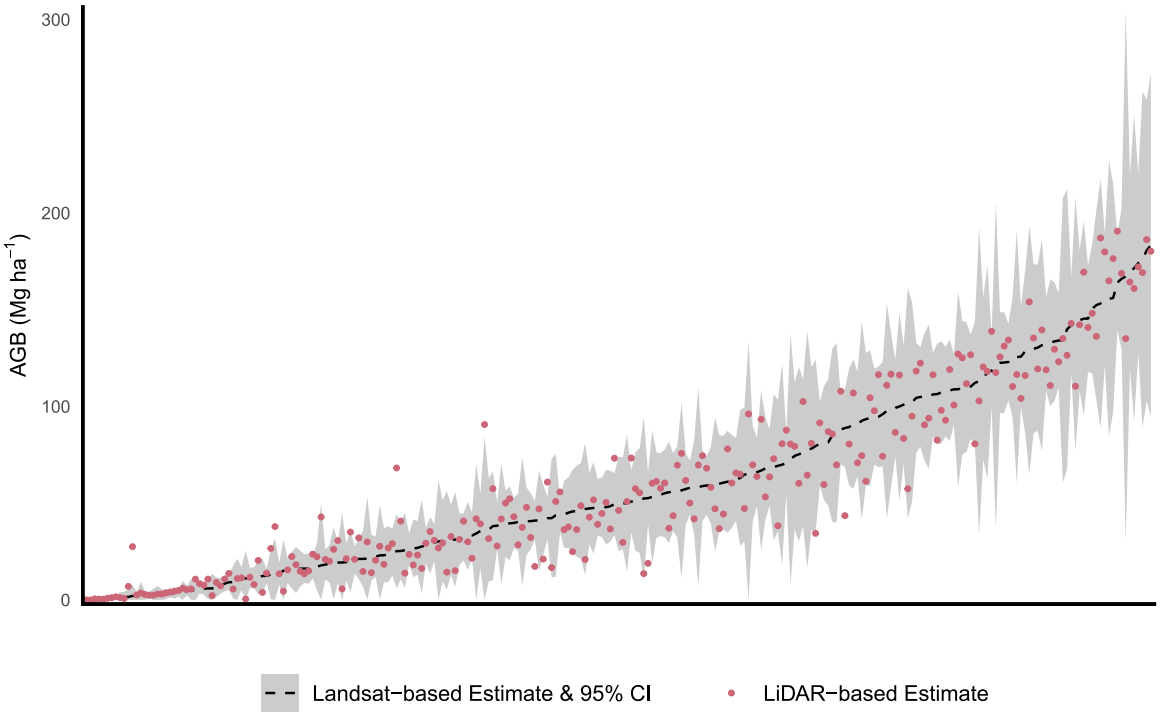

Fig. 9. Empirical coverage of 95% confidence intervals (gray shaded area) using SEs estimated from a natural log–log regression with parcel area, perimeter, AGB density, and % forest cover as independent variables. Pink points represent parcel-level AGB estimates made with predictions from LiDAR-based models developed in Johnson et al. (2022). Black-dashed line represents parcel-level AGB estimates made with predictions from Landsat-based models developed in Johnson et al. (2023) and replicated in this study. Parcels are sorted by increasing Landsat-based AGB predictions along the x-axis.

findings pave the way for rigorous uncertainty estimation in future model-based forest carbon accounting and monitoring efforts.

4.1. Contributions to total uncertainty

Attributing contributions to total uncertainty from various sources is a critical step to reducing uncertainty — one of the IPCC good practice guidelines ("uncertainties are reduced as far as practicable"; Buendia et al. (2019)). Without identifying sources of uncertainty, reduction efforts cannot be effectively directed. Although these results most directly inform reducing uncertainties associated with the specific modeling framework implemented in Johnson et al. (2023), the general patterns uncovered here may also serve to inform future model-based efforts insofar as sufficient overlap exists in terms of sampling designs, reference data collection protocols, remotely sensed data products, and model forms.

Residual variance ($\widehat{Var_{res}}$; Eq. (9)) dominated all sources of uncertainty for all parcel sizes (0–2023 hectares; Fig. 6) primarily due to the impact of model residual spatial autocorrelation (Fig. 7). These results hold significance for future model-based efforts in areas where forests are divided into smaller units and parcel dimensions do not far exceed the range of spatial autocorrelation. In regions like the eastern and midwestern United States (US), forests are predominantly managed by nonindustrial private forest owners ("family forests"), and these family forest holdings are often small (Nelson et al., 2010; Butler et al., 2016). We therefore recommend prioritizing model accuracy to most effectively reduce total uncertainty (Var_{total}, Eq. (1); SE, Eq. (2)) in these contexts. Increased model accuracy simultaneously reduces both residual variance and residual spatial autocorrelation, effectively diminishing $\widehat{Var_{res}}$. Of course, model accuracy is almost always a priority, but we highlight its importance relative to other uncertainty components assuming a limited amount of time and resources with which to reduce total variance (Table 1). In particular, modeling approaches that include data from active remote sensing systems, like synthetic aperture radar or spaceborne LiDAR (Quegan et al., 2019; Dubayah et al., 2020) which can penetrate forest canopies and may better represent structural attributes related to AGB, could increase model accuracy relative to approaches based only on passive remote sensing (e.g. Landsat) like the approach investigated in this study (Section 2.4; Section 2.5). Higher resolution imagery in concert with the implementation of deep-learning algorithms like convolutional neural networks which can exploit local spatial patterns may offer improved model predictions as well (LeCun et al., 2015; Kattenborn et al., 2021).

Contributions to total uncertainty from sampling variance ($\widehat{Var_{sam}}$; Eq. (6)) began to approach contributions from $\widehat{Var_{res}}$ for only the largest parcels (>20% contribution for parcels 1012 to 2023 hectares; Fig. 6). We expect this to hold true for NYS parcels larger than the maximum size considered in this study given the inverse relationship between $\widehat{Var_{res}}$ and parcel size. In other words, while we assume $\widehat{Var_{res}}$ will approach zero with increasing parcel size, we also assume that $\widehat{Var_{sam}}$ will remain relatively stable and thus contribute a larger and larger proportion to a decreasing Var_{total}. With all other variables held constant, this suggests that in regions where forests are contained in larger contiguous management units, like the western US (Nelson et al., 2010; Butler et al., 2016), reductions in model-based uncertainty can be most efficiently achieved by investing in more forest inventory sample locations. However, forests in the western US are generally conifer-dominated with even-aged disturbance regimes (Perry et al., 2022; Rogers, 1996), whereas forests in the eastern US are mostly mixed hardwood-conifer with uneven-aged disturbance regimes (Seymour et al., 2002; Dyer, 2006). As a result, it is entirely possible that spatial correlation of model residuals may be stronger across the larger and more uniform patches of forest in the West, suggesting that reducing residual variance ($\widehat{Var_{res}}$) instead of sampling variance ($\widehat{Var_{sam}}$) may be the most efficient means to reduce total variance. That said, eastern and western forests of the US may be too distinct to be able to extrapolate drivers of model-based variability from one region to the other.

Contributions from landcover mask variability ($\widehat{Var_{LC}}$; Eq. (7)) were small except in the least forested parcels (0%–20% forested; >20%

of Var_{total}; Fig. 6). We can infer that the increased landcover variability in these parcels is tied to the presence of marginal and transitional landcover classes (e.g. grass/shrub, cropland), which in many contexts may hold non-trivial quantities of forest biomass (Schnell et al., 2014; Johnson et al., 2015; Liu et al., 2023), but are traditionally challenging to classify (Brown et al., 2020; Stehman et al., 2021; Mahoney et al., 2022). Subsequently, we suggest that landcover classification accuracy may be worth prioritizing in areas dominated by woodlands, shrublands, other marginal, transitional, or novel land cover types, or where agroforestry is commonplace (Van Auken, 2000; Knapp et al., 2007; Schoeneberger, 2009; Udawatta and Jose, 2011; Mahoney et al., 2022).

Reference data variability ($\widehat{Var_{ref}}$; Eq. (8)) contributed minimally to total uncertainty across all analyzed groupings (Fig. 6; <10% of Var_{total}). These results corroborate previous studies reporting that tree-level uncertainty contributes minimally to total uncertainty when aggregated over many individuals (Breidenbach et al., 2014; Ståhl et al., 2014; McRoberts and Westfall, 2014; Yanai et al., 2023); indeed, we demonstrated that the coefficient of variation (CV) for plot-level AGB predictions was an order of magnitude smaller than the CV for tree-level AGB predictions (tree-level AGB CV 34.96%; plot-level AGB CV 7.2%) despite relying on relatively small-area plots (0.067 ha). One would expect even less variability if larger area plots were utilized (as recommended by CEOS (2021)). Further, our results suggest that the imprecision of FIA plot locations (7.05 m standard deviation; Table A.1) does not meaningfully impact total uncertainty. We might infer that the fidelity of 30 m Landsat imagery is not adequate, or that forests in NYS are not sufficiently patchy for imprecise reference data geolocation to be meaningful in this context. Despite its relative lack of importance in this study, reference data uncertainty may be important in regions where only small fractions of the tree population are represented in the data used to fit the allometric model. Regions like the global tropics with a greater diversity of tree species (Chave et al., 2014) or regions with a scarcity of forest inventory plots due to remoteness or lack of resources (e.g., lacking a national forest inventory; McRoberts et al., 2010) could yield particularly large reference data uncertainty contributions. It is therefore prudent to weigh the precision and representativeness (Labrière et al., 2022) of a given system of allometric models against the practicality and feasibility of incorporating reference data uncertainty in estimates of total variance.

In general, the framework we have developed may facilitate a range of sensitivity analyses for different components of uncertainty and parameters not limited to those considered here. By keeping all other inputs constant and increasing (or conversely removing) uncertainty from a single component, we can begin to identify specific thresholds beyond which more granular uncertainty contributions overwhelm the total uncertainty estimate. Although these kinds of analyses may be simple in theory, they are computationally expensive, as testing each component requires a new 1000 replication bootstrap procedure. With further investments and developments in computational infrastructure such analyses may become more tractable. Nevertheless, we recommend this kind of framework for iteratively exploring sources of model-based uncertainty.

4.2. Communicating uncertainty to map users

Although pixel-level predictions in fine-resolution maps effectively represent spatial patterns, finer resolution is often associated with increased uncertainty, and management decisions are made at coarser spatial scales (e.g. forest stands, compartments, ownership parcels, or special management areas). Thus, we expect that users will spatially aggregate groups of adjacent pixels that more directly correspond to landscape units or patches of fundamental relevance for formal estimation. The delivery of corresponding uncertainty maps (pixel-level variance) alone fails to facilitate statistically sound estimates of aggregate uncertainty, as pixel-level variances cannot be naively averaged within areas of interest in the same way that AGB predictions can (Wadoux and Heuvelink, 2023). We therefore explored alternative ways to communicate the methods and metadata required to make rigorous uncertainty estimates for map subregions.

In this study we found that parcel-level uncertainty (SE) varied substantially with parcel size, forest cover, and AGB density, thus enabling the use of a natural log–log multiple regression (Section 2.8; Table 2; Fig. G.1)) for accurately estimating SE for any subregion of the AGB maps from Johnson et al. (2023). Given the coefficients of this regression, map users need only compute the predictors for their area of interest and substitute them into the model formula (Eq. (G.2)). All of the parcel characteristics included as predictors in the model are intended to be accessible to map users since each can be computed with a standard GIS and a bounding polygon for the user's area of interest. Using this regression, we achieve computational efficiency and model parsimony at the price of small inaccuracies in SE estimates. These inaccuracies were most apparent for large SE estimates beyond 40 Mg ha^{-1} where our model overestimated SEs computed from first principles (Fig. G.1). Although we may have achieved more accurate SE estimates with a more flexible machine learning model, a parametric regression can be readily expressed as a straightforward, easy-to-use, prediction equation that non-experts can easily interpret. The same cannot be said for machine learning algorithms.

We do not expect that this type of regression will be useful across all landscapes, ecological domains, or modeling frameworks. However, we do expect this type of regression to be most effective for models that exhibit heteroscedastic residual variance, or in regions where the majority of estimation unit sizes are relatively small. In these situations, both the positive correlation between AGB and residual variance and the strong inverse relationship between estimation unit size and spatial covariance of residuals ($\widehat{Var_{res}}$; Eq. (9)) can be leveraged to make accurate estimates. In regions with constant residual variance, or where the majority of estimation units have dimensions that exceed the range of spatial autocorrelation of residuals, this type of regression may be less effective at estimating SE. However, we expect SEs in these latter circumstances to be generally small. Further, bootstrap variance ($\widehat{Var_{boot}}$; Eq. (3)) may contribute more substantially in regions with more sparsely sampled forest inventories ($\widehat{Var_{sam}}$; Eq. (6)) or less precise allometric models ($\widehat{Var_{ref}}$; Eq. (8)). In these contexts, alternative predictors that help describe the degree to which a subregion is represented in the reference data or forest inventory sample may be useful (e.g. area of applicability information; Meyer and Pebesma (2021)). In light of our findings, we recommend further investigation into regression as a means to communicate uncertainty for map subregions in a diversity of contexts.

Improving the transfer of uncertainty information to map users enables a more rigorous application of map products in decision-support contexts. Specifically, the estimates of SE produced in this study provide managers and decision-makers the means to compare the outcomes of various policies and management practices across the landscape with the added context of confidence intervals (Figs. 8; 9). For example, the hypothesis that forests protected by conservation easements hold larger carbon stocks can be formally tested (McRoberts, 2011). Although this study has focused entirely on estimating uncertainties associated with AGB stocks, the same bootstrapping approach and the parcel-level SE estimates developed here may facilitate the estimation of uncertainties associated with stock-changes (Esteban et al., 2020), which can explicitly account for GHG emissions and removals attributed to the forest sector and can be used to track forest carbon dynamics through time.

4.3. Limitations

Although this study demonstrates rigorous model-based uncertainty estimation, our approach is not without limitations and assumptions. Ultimately, we achieved only an 89% empirical coverage proportion with our estimated 95% confidence intervals (Fig. 9). First, we describe limitations related to our overarching approach and in subsequent sections we identify limitations specific to our estimation of uncertainty types a, c, and d (Table 1). These limitations represent opportunities where future research may be directed.

4.3.1. Overview

Implicit in the formula for estimating $\text{Var}_{\text{total}}$ and SE (Eqs. (1); (2)) is the assumption that $\widehat{Var_{boot}}$ and $\widehat{Var_{res}}$ are independent (McRoberts et al., 2022). If, however, residual variability depends on the variability of the reference data, the model building sample, and/or the auxiliary data (uncertainty types a, b, and d in Table 1; components of $\widehat{Var_{boot}}$ in this study) then $\text{Var}_{\text{total}}$ may be overestimated. Future research could evaluate the validity and implications of this assumption and work towards producing estimates of $\text{Var}_{\text{total}}$ without this assumption. For instance, it may be possible to subtract variance components of $\widehat{Var_{boot}}$ (resulting from types a, b, d in Table 1) from $\widehat{Var_{res}}$ to obtain estimations of the uncertainty resulting from the model not being able to represent all of the variation in the response variable (type c in Table 1) where the auxiliary data, reference data, or sample are fixed (i.e. error free).

Producing uncertainty estimates for the 2224 parcels in this study and decomposing these uncertainty estimates into four components ($\widehat{Var_{ref}}$, $\widehat{Var_{sam}}$, $\widehat{Var_{LC}}$, $\widehat{Var_{res}}$) was computationally expensive and may present barriers to broad-scale adoption of our approach. Producing one set of 1000 bootstrap replications, including generating new reference data, re-extracting auxiliary data, training ML models, and making predictions, required on the order of several weeks of processing time. Producing four sets of 1000 replications (to decompose $\widehat{Var_{boot}}$ like we did here; Section 2.1.1) required several months of processing time. In total this project occupied roughly 1 TB of disk space. We used a Dell Power Edge r710 with 16 central processing units and 80 GB of random access memory. However, this machine is now over 15 years old, and we imagine that processing time could be greatly reduced with modest improvements to hardware or investments in cloud or distributed computing resources.

Also in consideration of computational efficiency, we limited the number of bootstrap replications to 1000. We tested the stability of our $\widehat{Var_{boot}}$ estimates and found that 1000 replications is sufficient to produce stable estimates when considering a relatively liberal level of precision (±5%; Section 3.6; Appendix J). Additional bootstrap replications should be added to achieve greater, more specific, levels of precision (McRoberts et al., 2023).

Finally, alternative bootstrapping protocols may be more desirable within the circumstances of our study. In our case we have blended pairs bootstrapping (resampling plots with replacement) and residual bootstrapping (FIA plots; measurement errors, allometric errors, and plot location errors) within the same replication framework (Efron and Tibshirani, 1994). Wild bootstrapping (Esteban et al., 2020; Liu, 1988) may be preferred as it preserves both the original sampling structure and the heteroskedasticity of model residuals.

4.3.2. Reference data uncertainty (type a)

In light of data availability and project efficiency, we sourced reference data error distributions pragmatically (Appendix A) ; we relied on published statistics to build normal distributions of measurement errors and we leveraged the legacy tree database to build empirical distributions of allometric errors (Radtke et al., 2015). In some cases, we omitted specific components of tree-level AGB uncertainty due to lacking data or intractability. Our approach (Appendix A) has yielded, at best, approximations of true error distributions including assumptions that likely led to underestimation of the contribution from reference data uncertainty.

If measurement error distributions were skewed, rather than normal like we assumed, measurement errors would have larger contributions to total uncertainty since the effects of positive and negative errors would be less likely to cancel out when aggregated across trees in a plot. Similarly, if our allometric error distributions covered the largest trees in our NYS FIA reference dataset (Appendix A), allometric errors would likely have larger contributions to total uncertainty given the evidence allometric models can systematically underestimate large-tree biomass (Stovall et al., 2018). We have also assumed that measurement, classification, and allometric errors are independent between trees. Under this assumption, tree-level errors will also largely cancel out when summed at the plot-level (Section 3.1; Appendix A) and reference data uncertainty is likely underestimated (Breidenbach et al., 2014). Although the FIA field crews that collected our reference tree measurements are trained and certified annually and are subject to regular quality assurance checks (Yanai et al., 2023), it is possible that certain conditions lead to the repetition of similar errors within a local area. For example, a swath of trees with broken tops in a storm path could present a concentration of height measurement challenges (Yanai et al., 2023). Similarly, specific growing conditions may not be represented in the allometric model training data, potentially leading to clusters of similar allometric errors.

4.3.3. Residual variability (type c)

Estimates of residual variability (type c in Table 1; $\widehat{Var_{res}}$) for map subregions require pixel-scale residual information which creates challenges when reference datasets traditionally offer plot support (e.g. national forest inventories like the FIA). To overcome these challenges, we modeled pixel-level residual variance as a function of our AGB predictions (Section 2.6; Appendix E) and used LiDAR-AGB predictions as reference data to compute pixel-scale residuals for estimating residual spatial autocorrelation (Section 2.7; Eq. (11)). Both solutions have limitations that may affect estimates of $\widehat{Var_{res}}$ and $\text{Var}_{\text{total}}$. Our residual variance model was only able to represent 55% of the variation in the residual variance data (R^2 0.55) limiting our ability to represent true pixel-level residual variance and potentially condensing the range of parcel-level $\widehat{Var_{res}}$ estimates. In using LiDAR-AGB predictions to estimate residual spatial autocorrelation we acknowledge that predictions from the LiDAR-based models may be spatially correlated themselves, and there may be systematic differences between the Landsat-based model predictions and the LiDAR-based predictions that degrade the fit of the semivariogram function. Although estimates of parcel-level SE are sensitive to the strength of residual spatial correlation (Fig. 7) and there is no guarantee that residuals computed with LiDAR-based reference data have the same spatial properties as those computed with field inventory data, this substitution was a practical solution to achieve the spatial density and coverage required for estimating residual spatial autocorrelation.

4.3.4. Auxiliary data uncertainty (type d)

In general, there is a degree of subjectivity involved in choosing which sources of uncertainty are included in an assessment. Here we have omitted uncertainty in model predictor data (classified as type d in Table 1) from our framework (CEOS, 2021). The models in Johnson et al. (2023) are largely dependent on Landsat imagery that has been temporally smoothed, segmented, and gap-filled with Landtrendr (Kennedy et al., 2010, 2018b). Although incorporating uncertainty associated with the geometric or radiometric features of the raw Landsat imagery was beyond the scope of this study, we may be able to use the Landtrendr residuals (differences between raw spectral value and Landtrendr fitted value) to assess variability in the imagery in future studies.

We did include uncertainty related to our forest mask ($\widehat{Var_{LC}}$; type d in Table 1; Section 2.5). Our approach to developing bootstrap replications of forest masks was naive in that we did not account for spatial dependence of landcover classification errors. Rather, we treated each pixel independently when drawing a uniform random number to determine if a new classification should be assigned. This may lead to underestimation of $\widehat{Var_{LC}}$ in landscapes that generate a concentration of classification errors since it is likely that conditions contributing to misclassification in one pixel are present in neighboring pixels. That said, this effect is mitigated to an extent by the fact that select Land Change Monitoring, Assessment, and Projection primary classes have worse accuracy (e.g. cropland, grass/shrub); these classes are often clustered in space such that an area with a substantial concentration of, for example, grass/shrub classified pixels is likely to have a higher $\widehat{Var_{LC}}$ estimate.

Table A.1
Parameters and sources for estimating tree- and plot-level measurement error distributions. Where relative standard deviation (SD) is marked as 'T' (true), actual SDs are to be computed relative to the FIA provided value using the SD fraction in this table (i.e. $\text{actualSD} = \text{SDtable} \cdot \text{FIAvalue}$).

FIA DB Name	Description	Mean	SD	Rel. SD	Units	Source
DIA	Diameter at breast height	−0.004	0.550	F	cm	Yanai et al. (2023)
BOLEHT	Bole height	−0.050	1.520	F	m	Yanai et al. (2023)
CULL	Cull, percent rotten or missing	0.100	3.500	F	%	Yanai et al. (2023)
WOOD_SPGR_GREENVOL_DRYWT	Wood specific gravity (green, oven-dry)	0.000	0.100	T	NA	Ross (2021)
BARK_SPGR_GREENVOL_DRYWT	Bark specific gravity (green, oven-dry)	0.000	0.100	T	NA	Assumed
STANDING_DEAD_DECAY_RATIO (hardwood)	Hardwood density reduction ratios for DECAYCD 1–5	0.000	0.010 0.020 0.040 0.050 0.050	T	%	Harmon et al. (2011)
STANDING_DEAD_DECAY_RATIO (softwood)	Softwood density reduction ratios for DECAYCD 1–5	0.000	0.010 0.011 0.031 0.031 0.030	T	%	Harmon et al. (2011)
LON, LAT	Distance between field recorded plot location and plot location averaged over multiple repeat visits	0.000	7.050	F	m	NYS FIA DB

5. Conclusion

The proliferation of fine-resolution forest biomass (AGB) and carbon maps has not been matched with corresponding effort in the estimation and communication of uncertainty for small-area spatial aggregates of map subregions. To address this shortcoming, we developed and applied methodology to produce estimates of standard error (SE) associated with spatial averages of AGB predictions for a stratified random sample of ownership parcels in New York State. This represents a novel model-based uncertainty estimation approach that included all four types of uncertainty (Table 1), that used methods compatible with algorithmic models (machine learning or nonparametric), and that incorporated spatial autocorrelation of residuals. Our results suggest that residual variance, largely resulting from spatial autocorrelation of model residuals, will be the largest source of uncertainty in landscapes like the midwestern and northeastern United States where forests are divided into smaller ownership units. This highlights the importance of accounting for residual spatial autocorrelation in small-area uncertainty estimates and implies that improvements in model accuracy will yield the greatest reductions to uncertainty in these regions. We further demonstrated that a parametric regression model relating parcel characteristics (area, perimeter, AGB density, forest cover) to parcel SE can provide accurate uncertainty estimation for any map subregion with only a fraction of the computing resources and data required to do so from first principles. Overall, our approach and outputs can directly support future model-based efforts and promote transparency in spatial forest carbon accounting.

CRediT authorship contribution statement

Lucas K. Johnson: Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Grant M. Domke:** Writing – review & editing, Validation, Conceptualization. **Stephen V. Stehman:** Writing – review & editing, Methodology, Conceptualization. **Michael J. Mahoney:** Writing – review & editing, Software, Formal analysis, Data curation. **Colin M. Beier:** Writing – review & editing, Visualization, Supervision, Funding acquisition, 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.

Acknowledgments

We would like to thank the USDA Forest Service FIA program for their data sharing and cooperation, and especially Brian F. Walters for his consultation. We would also like to thank Sean Healey for his helpful comments, the NYS GPO for compiling and serving LiDAR data, and the NYS Department of Environmental Conservation, USA (#AM11459), Office of Climate Change for funding support. The findings and conclusions in this publication are those of the author(s) and should not be construed to represent any official US Department of Agriculture or US Government determination or policy.

Appendix A. Reference data uncertainty

A.1. Measurement and classification error

Both measurement and classification errors were randomly and independently selected and assigned to individual trees before summing tree-level predictions for each plot. We assumed normal distributions, despite the fact that they only represent (at best) approximations of true error distributions, for measured variables using parameters (mean, standard deviation) reported in the literature (Table A.1). Decay code (DECAYCD in FIA DB) misclassification was randomly assigned by selecting a random number between 0 and 1, and randomly choosing a sign (±). If the selected number was greater than the Yanai et al. (2023) estimated 59% accuracy rate, then the decay class above or below was assigned based on the randomly selected sign. Species classifications (SPCD in FIA DB), percent of bole volume that is represented by bark (BARK_VOL_PCT in FIA DB), and structural loss adjustments (Domke et al., 2011) based on assigned decay codes were omitted from tree-level uncertainty quantification due to lack of data or intractability. For example, misclassification rates for tree-level species were estimated by Yanai et al. (2023), but a method for assigning a replacement species in the event of a misclassification was not apparent.

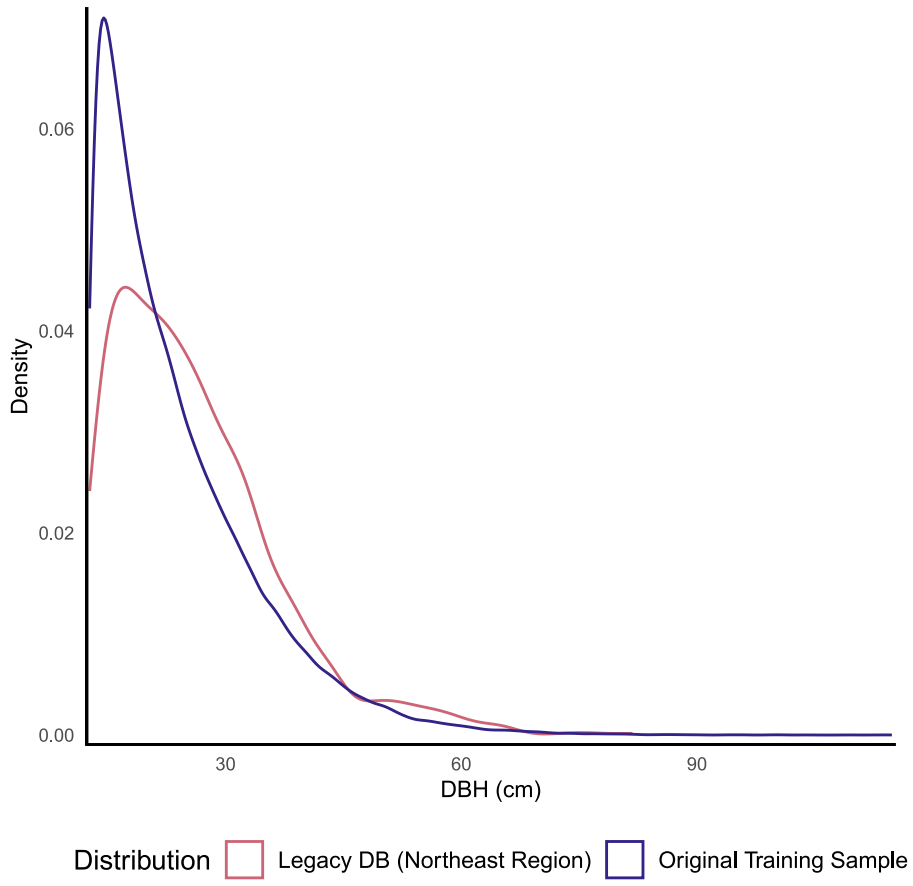

Fig. A.1. Smoothed frequency distributions for diameter at breast height (DBH) from reference datasets. Pink distributions are sourced from the Northeastern partition of the Legacy Tree Database which were used to construct allometric error distributions. Purple distributions are sourced from trees measured within the 2049 FIA plots in our New York State model development dataset. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

A.2. Allometric model uncertainty

We utilized the legacy tree database, which provides destructive measurements of volumes and weights of trees and tree components (e.g. foliage, top and branches, bole) for North American tree species (Radtke et al., 2015), to account for allometric model uncertainty. We selected only those records obtained from sites or studies within FIA's designated Northeastern states because the CRM models for NYS are specific to this region (Woodall et al., 2011). The range of tree DBH values present in the set of legacy tree records from the Northeastern states covers all but the largest individuals in NYS FIA dataset (Fig. A.1). This indicates that our error distributions are mostly representative of the reference data population used to train our models but may omit potential systematic errors relating to underestimations of large-tree biomass (Stovall et al., 2018).

We used the selected database records to produce separate hardwood and softwood empirical residual distributions for bole volume, bole bark biomass, bole wood biomass, foliage biomass, and total tree biomass (Fig. A.2). We did not account for stump biomass uncertainty or error since stump biomass measurements were not available for the Northeastern partition of the legacy tree database, and tractable estimates of uncertainty or error were not presented with the stump biomass models published in Raile (1982). All biomass residuals were computed relative to total tree biomass predicted with Jenkins et al. (2003) allometric models, and volume residuals were computed relative to predicted bole volume with Scott (1981) models so that adding randomly selected residuals would produce new estimates that were within a reasonable distance from the original.

Under the CRM approach, top and branch biomass (TAB) is predicted by subtracting all other predicted component biomass (stump, bole, foliage) from the predicted total. To ensure that our random residual selection yielded TAB estimates that were ecologically sound, we also compiled ratios of TAB to total tree biomass (minus predicted stump biomass; Appendix C) using the legacy tree database. We did not partition the TAB ratio distribution further into softwood and hardwood distributions since there were limited database records with both TAB measurements and total tree biomass measurements. Given the paucity of TAB ratio data, we filtered three records from the TAB ratio distribution where ratios were >0.5 to limit their influence on the overall outcome. After all components (total biomass, bole biomass, stump biomass, and foliage biomass) had been predicted and random selected residuals had been added to each component, we randomly selected a TAB ratio, computed a TAB prediction ([TAB ratio] · [total biomass]), and summed all components. The difference between the sum of components and the total was proportionally subtracted from each component to ensure that the sum of predicted component biomass was equal to the predicted total tree biomass. All allometric errors were assumed to be independent between trees.

A.3. Plot location uncertainty

FIA plots are composed of four identical circular subplots with radii of 7.32 m (24 ft), with one subplot centered at the macroplot centroid and three subplots located 36.6 m (120 ft) away at azimuths of 360°, 120°, and 240° (Bechtold and Patterson, 2005). The FIA program averages individual plot locations collected over multiple repeat visits to improve upon the precision and accuracy of field-recorded coordinates (Cooke, 2000; Hoppus and Lister, 2005). We constructed a normal distribution representing plot location errors by computing the standard deviation of differences between field-recorded coordinates

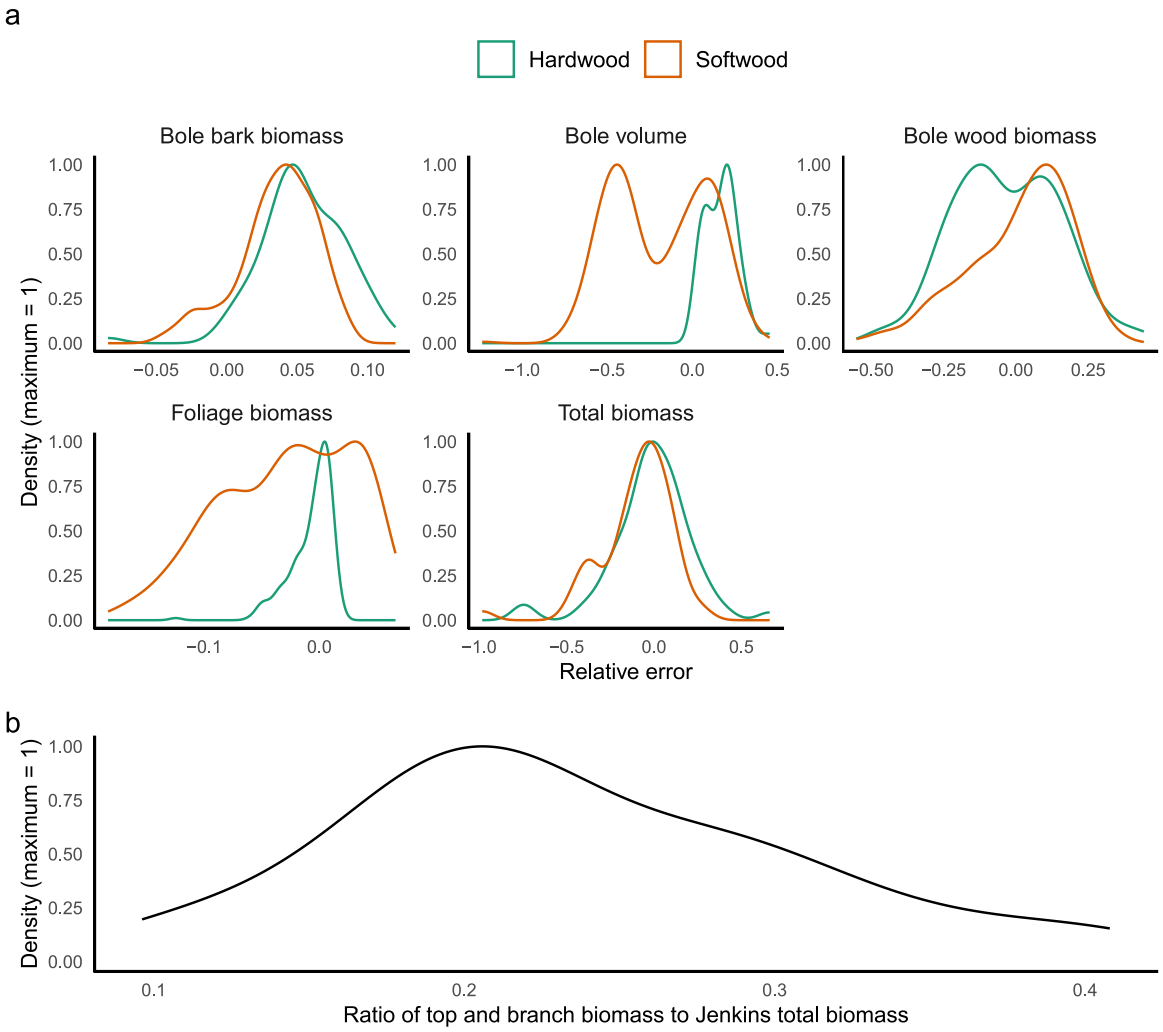

Fig. A.2. Smoothed frequency distributions of allometric model uncertainty information compiled from the legacy tree database. Distributions for all individual panels have been rescaled separately, such that the most common occurrences within each panel are assigned a value of 1. (a) Relative residuals for five estimated tree components. Bole volume residuals are relative to bole volume, while all biomass residuals are relative to total tree biomass. (b) Ratios for top and branch biomass relative to total biomass.

and the corresponding average coordinates across all plots in the model development set (Table A.1). To represent this positional uncertainty in $\widehat{Var_{boot}}$ (specifically within $\widehat{Var_{ref}}$), for each plot and at each bootstrap replication, we randomly selected a distance d from the constructed normal distribution and a random azimuth a between 0° and 360° to create new plot centroids and subsequent plot polygons that were offset from the original plot centroid by a distance d m and an angle a°.

Appendix B. Auxiliary data description

Each of the 29 predictor layers (Table B.1) were projected to match Landsat 30 m pixel geometries. The raster stacks of predictors were clipped and aggregated at the constructed FIA plot polygons (Appendix A) to produce training datasets. Specifically, for each predictor and plot, we calculated the mean of pixel values intersecting the plot geometry weighted by the fraction of each pixel overlapping with the plot area. Raster stacks of predictors were clipped to the boundaries of the parcels in our sample for mapping (Section 2.2). The exactextractr (Baston, 2022) and terra (Hijmans, 2023) packages for the R (R Core Team, 2023) programming language were used to create the plot- and parcel-level auxiliary datasets.

Appendix C. Component ratio method

See Fig. C.1.

Appendix D. Aboveground biomass machine learning ensemble models

Following the 'direct' approach in Johnson et al. (2023), each ensemble, for every replication i, was a linear regression that combined predictions from a random forest implemented with the ranger R package (Breiman, 2001a; Wright and Ziegler, 2017), a stochastic gradient boosting machine as implemented in the lightgbm R package (Friedman, 2002; Ke et al., 2017; Shi et al., 2022), and a support vector machine (SVM) as implemented in the kernlab R package (Cortes and Vapnik, 1995; Karatzoglou et al., 2004). We used the same set of hyperparameters, identified in Johnson et al. (2023) as the most performant via an iterative grid search using five-fold cross validation within the 80% training partition (Section 2.3), for each respective ML model across all bootstrap replications. We used these pre-selected hyperparameters not only because the iterative grid search used in Johnson et al. (2023) was computationally expensive, but also because we do not expect the AGB or predictor distributions within each bootstrap sample to fundamentally differ since they represent the same underlying population. Further information describing the ML ensembles and hyperparameter tuning can be found in Johnson et al. (2023).

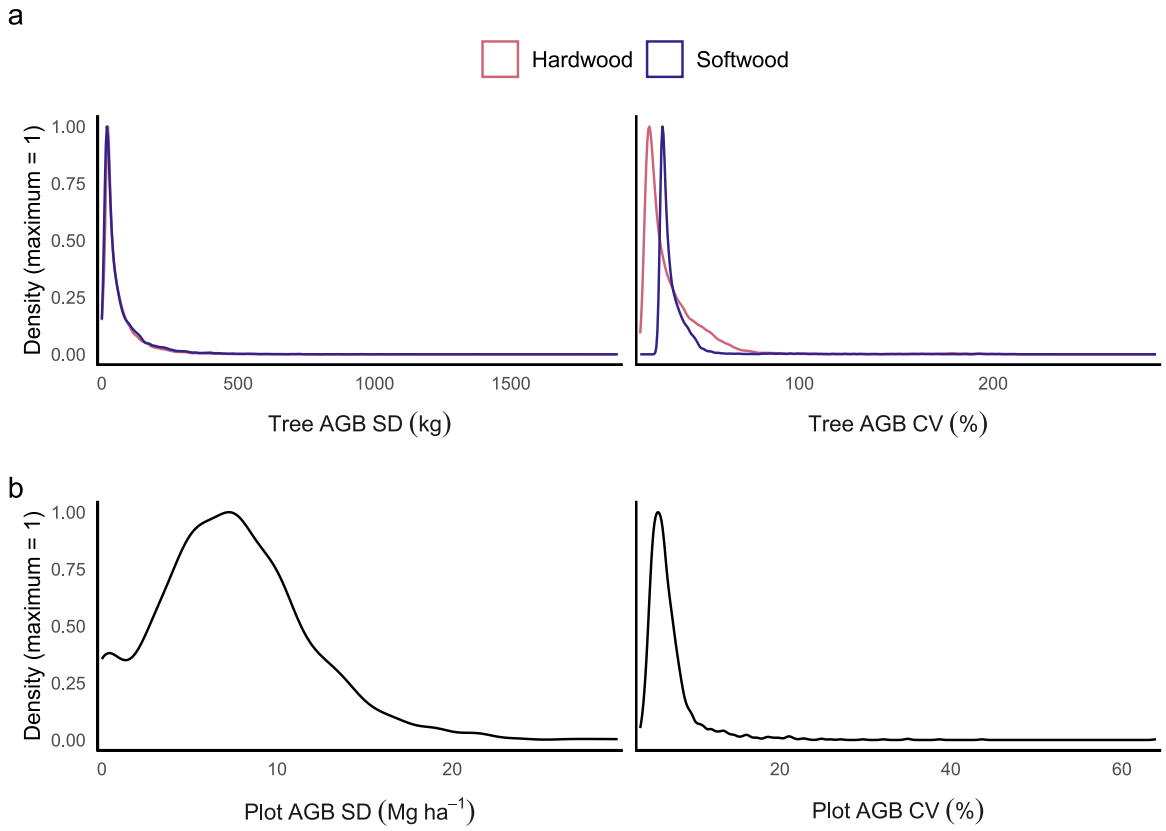

Fig. A.3. Reference data uncertainty results. Smoothed frequency distributions estimates for individual tree (a) and plot (b) AGB standard deviations (SD) and coefficients of variation (CV; SD/AGB 100) over the 1000 bootstrap replications. Distributions for individual panels have been rescaled separately such that the most common occurrences within each panel are assigned a value of 1.

Table B.1
Definitions of predictors used for model fitting.

Group	Predictor	Definition
Spectral indices	TCB, TCW, TCG	Tassled cap brightness, wetness, and greenness, with noise removed using LT-GEE
	NBR	Normalized burn ratio with noise removed using LT-GEE
	NDVI	Normalized difference vegetation index with noise removed using LT-GEE
	SR	Simple ratio with noise removed using LT-GEE
	MSR	Modified simple ratio with noise removed using LT-GEE
Delta	Delta_*	Change computed with 1 year lag for all predictors in the 'Spectral indices' group
Disturbance	YOD, MAG	Year of most recent disturbance and associated magnitude of NBR change, as identified using an NBR segmentation in LT-GEE (1985–2019)
Ecological	CHM	Global canopy height model reflecting 2005 conditions (Simard 2011), downsampled from 1 km to 30 m resolution
	ECOZONE	EPA level 4 ecozones. Aggregated to level 3 if level 4 areas <2% of the state. Set to 'other' if level 3 aggregation <2% of state.
	WETLAND	Wetland classification codes from the FWS National Wetlands Inventory
	DIST_TO_WATER	Distance in meters to nearest TIGER/Line Shapefile water from the US Census Bureau
Climate	PRECIP, TMAX, TMIN	30-year normals for precipitation, maximum temperature, and minimum temperature, derived from annual PRISM climate models
Topographic	ASPECT, ELEVATION, SLOPE, TWI	Aspect, elevation, slope, and topographic wetness index derived from a 30-m digital elevation model
Landcover	LCPRI, LCSEC	LCMAP primary and secondary land cover classifications

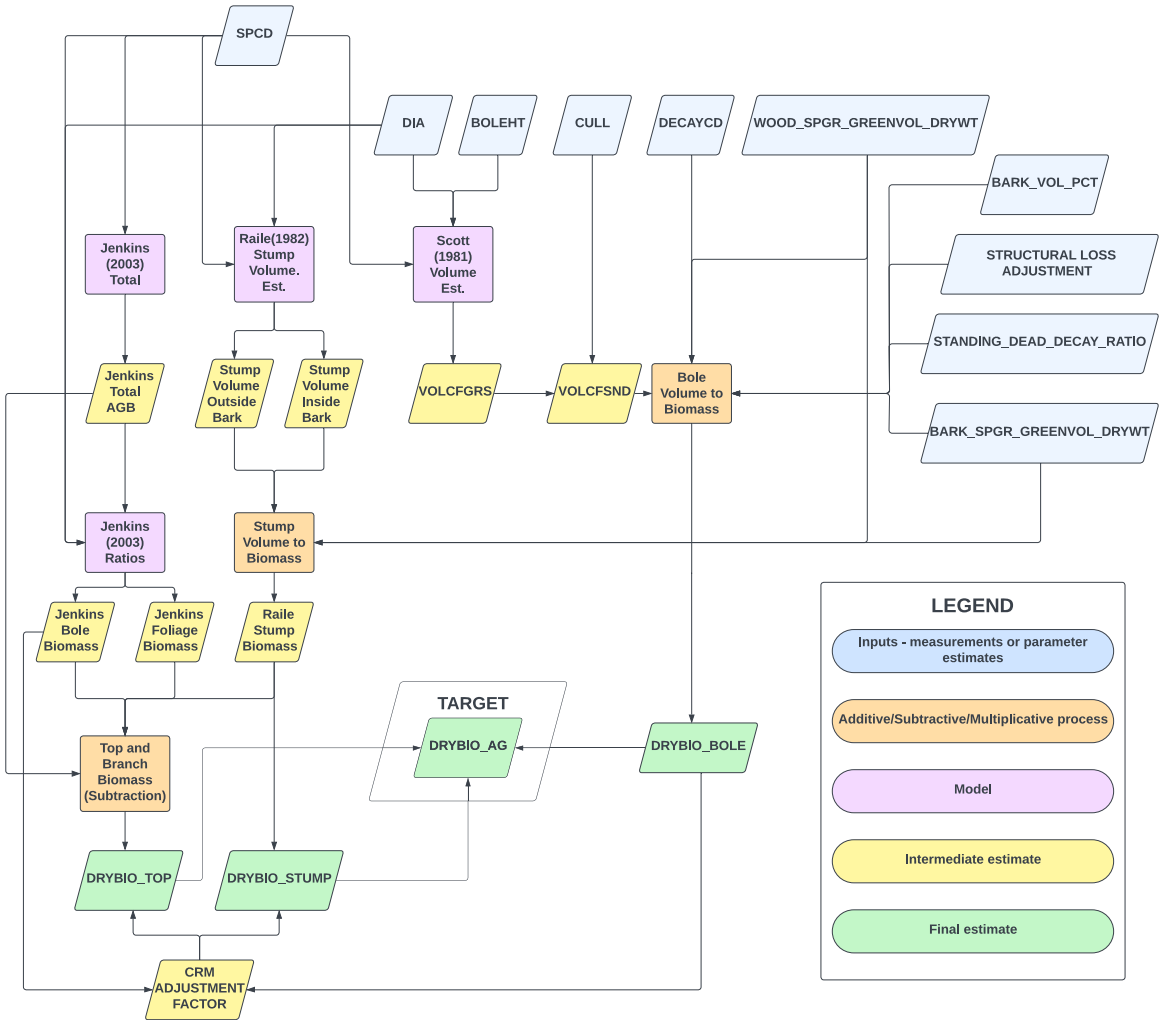

Fig. C.1. Flowchart showing key elements of the component ratio method (CRM; Woodall et al., 2011) for predicting tree-level AGB (DRYBIO_AG) using the FIA database. All names of inputs reflect FIA database values as best as possible.

Appendix E. Residual variance model

Following the approach outlined in McRoberts et al. (2022), we arranged our assessment plots in order of increasing AGB prediction and partitioned the data into thirty equal AGB intervals ranging from 0 to the maximum AGB prediction in the dataset. To ensure a sufficient number of observations were available to estimate residual variance for each interval (McRoberts et al., 2022), we collapsed the initial groups until each interval contained at least 10 plots, resulting in 21 groups. We then computed the average AGB prediction and the residual variance for each group, and fit several forms of regression models relating AGB to residual variance for the 21 points in the dataset. Specifically, we evaluated linear, log-linear, log–log, third order polynomial, and natural cubic spline (2 knots; Hastie (1992)) model forms. The most basic (linear) model form was constructed as follows:

$$\hat{\sigma}_i^{\,2} = \beta_0 + \beta_1 \cdot \hat{y}_i + \varepsilon \tag{E.1}$$

where $\hat{y}_i$ is the average AGB prediction for the group, $\hat{\sigma}_i^{\,2}$ is the estimated residual variance for the group, and β_0 and β_1 are coefficients estimated via ordinary least squares. All other model forms were fit similarly, using the same independent and dependent variables, but with log-transformations or additional model terms. When log transformations were used, we back-transformed estimates to a linear scale and applied multiplicative correction factors (CF; Eq. (E.2)) defined in Sprugel (1983) to eliminate log-transformation bias (Baskerville, 1972) as follows:

$$\mathrm{CF} = e^{\hat{\sigma}^2/2} \tag{E.2}$$

where sigma is the residual standard deviation from the fitted model.

Appendix F. Spatial autocorrelation of residuals

We temporally and spatially matched annual Landsat-based AGB predictions from Johnson et al. (2023) with the spatiotemporal patchwork of LiDAR-based AGB predictions from Johnson et al. (2022) spanning the years 2014–2019 (Appendix H). We then selected a simple random sample of 500,000 pixels from these matched maps and used them to compute residuals (Eq. (11)). We first calculated an empirical semivariogram from our sampled residuals with a bin width of 30 m (pixel resolution) which we then used to fit exponential, spherical, Matern, and Gaussian model semivariograms using ordinary least squares. We also fit nested combinations of these four basic

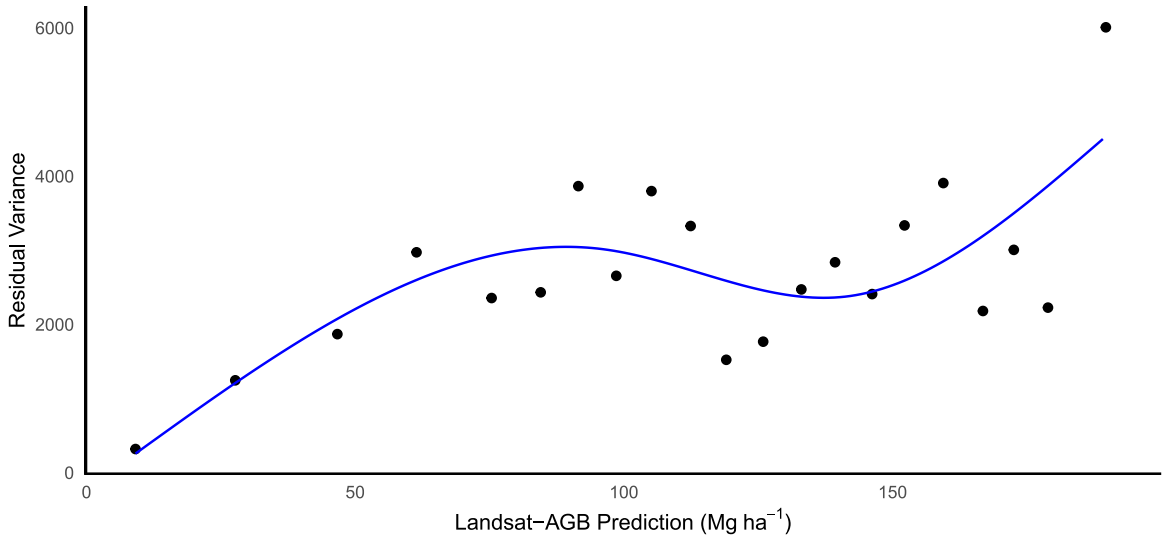

Fig. E.1. Natural cubic spline residual variance model fit (blue line) representing residual variance as a function of AGB prediction.

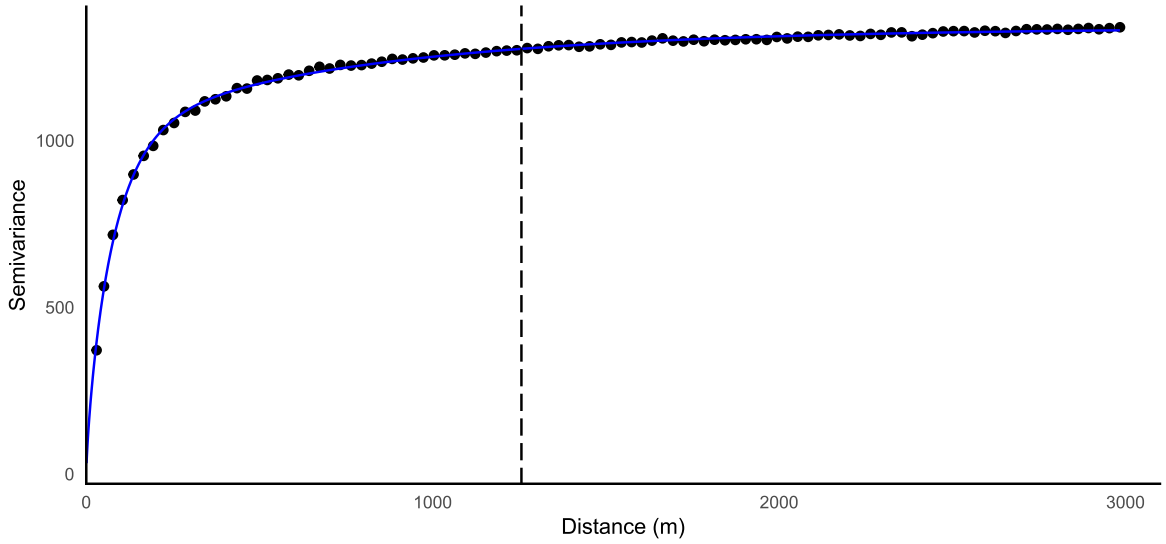

Fig. F.1. Nested Matern model semivariogram (blue line) fit to empirical semivariogram (black points) computed with a simple random spatial sample of 500,000 residuals (linear scale). Dashed vertical line represents effective range. 3 km lag distances yield semivariances >99% of the sill. Refer to Eq. (11) in the main body of the text for residual definition.

variograms to achieve greater flexibility. We assessed the fits of each model semivariogram visually, and selected the best fitting model to use in Eq. (10). All variogram computation and fitting was done with the gstat R package (Pebesma, 2004; Gräler et al., 2016).

Appendix G. Modeling estimated standard error (total variance) as a function of parcel characteristics

We fit four candidate models: one with no variable transformation, a second with natural log transformations for both the dependent and independent variables, and third and fourth models with natural log transformations for the dependent and independent variables respectively. When log transformations were used, we back-transformed estimates to a linear scale and applied correction factors (Eq. (E.2)) defined in Sprugel (1983) to eliminate log-transformation bias (Baskerville, 1972). Candidate models were constructed as follows:

$$\mathrm{SE} = \beta_0 + \beta_1 \cdot AGB + \beta_2 \cdot \%Forest + \beta_3 \cdot Area + \beta_4 \cdot Perimeter + \varepsilon \quad \text{(G.1)}$$

$$\ln(\mathrm{SE}) = \beta_0 + \beta_1 \cdot ln(AGB) + \beta_2 \cdot ln(\%Forest) + \beta_3 \cdot ln(Area) + \beta_4 \cdot ln(Perimeter) + \varepsilon \quad \text{(G.2)}$$

$$\ln(\mathrm{SE}) = \beta_0 + \beta_1 \cdot AGB + \beta_2 \cdot \%Forest + \beta_3 \cdot Area + \beta_4 \cdot Perimeter + \varepsilon \quad \text{(G.3)}$$

$$\mathrm{SE} = \beta_0 + \beta_1 \cdot ln(AGB) + \beta_2 \cdot ln(\%Forest) + \beta_3 \cdot ln(Area)$$

Table G.1
Coefficient of determination (R^2) for four candidate standard error (SE) models with a random 80% of the parcel sample (training partition).

Equation	Model	R^2
G1	$SE = \beta_0 + \beta_1 \cdot AGB + \beta_2 \cdot \%Forest + \beta_3 \cdot Area + \beta_4 \cdot Perimeter + \varepsilon$	0.76
G2	$\ln(SE) = \beta_0 + \beta_1 \cdot ln(AGB) + \beta_2 \cdot ln(\%Forest) + \beta_3 \cdot ln(Area) + \beta_4 \cdot ln(Perimeter) + \varepsilon$	0.94
G3	$\ln(SE) = \beta_0 + \beta_1 \cdot AGB + \beta_2 \cdot \%Forest + \beta_3 \cdot Area + \beta_4 \cdot Perimeter + \varepsilon$	0.50
G4	$SE = \beta_0 + \beta_1 \cdot ln(AGB) + \beta_2 \cdot ln(\%Forest) + \beta_3 \cdot ln(Area) + \beta_4 \cdot ln(Perimeter) + \varepsilon$	0.78

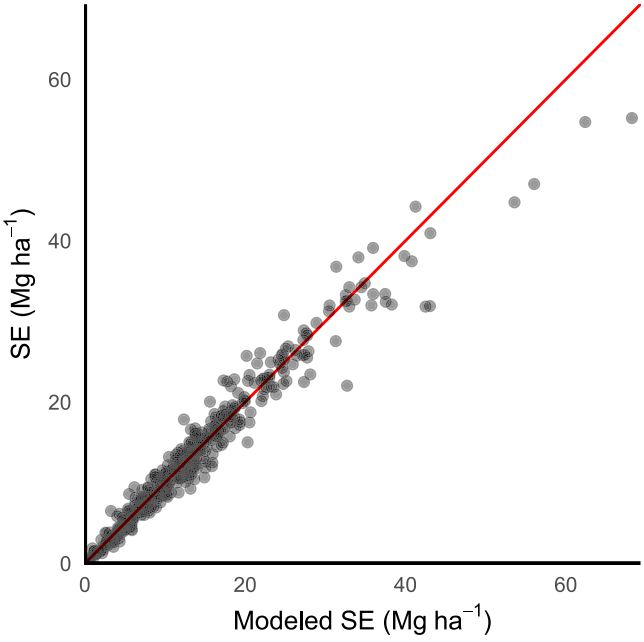

Fig. G.1. Standard error (SE) estimated with Eqs. (1) and (2) vs. SE estimated from a natural log–log regression with parcel area, perimeter, AGB density, and % forest cover as independent variables. Each point represents a parcel from the testing partition of the sample. 1:1 line in red.

$$+ \beta_4 \cdot ln(Perimeter) + \varepsilon \quad \text{(G.4)}$$

where AGB, $\%\ Forest$, $Area$, and $Perimeter$ are described in Section 2.8 and $\beta_0, \ldots, \beta_4$ are coefficients estimated via ordinary least squares. Each of the candidate models were fit to a randomly selected 80% of the parcel sample (training set) and were compared on the basis of R^2. The candidate model with the best R^2 statistic was selected as the final model. Using the final model and the remaining 20% of the parcel sample (testing set) we computed model accuracy metrics including root mean squared error (RMSE), mean absolute error (MAE), mean error (ME), and R^2 using the yardstick R package (Kuhn et al., 2023). All model accuracy and comparison metrics were computed after any necessary backtransformation and bias-correction had been applied.

Appendix H. LiDAR collection dates

See Fig. H.1.

Appendix I. Empirical coverage of 95% confidence intervals

We used 255 parcels within our parcel testing set that were entirely contained within the spatial footprint of the LiDAR collections processed in Johnson et al. (2022) (Appendix H) to assess the coverage of our estimated 95% confidence intervals. For each of these parcels we:

1. Extracted LiDAR-AGB pixel predictions.
2. Extracted Landsat-AGB pixel predictions from Johnson et al. (2023) for the year corresponding to the LiDAR collection.
3. Extracted LCPRI classifications for the year corresponding to the LiDAR collection.
4. Masked nonforest LiDAR-AGB and Landsat-AGB pixels to 0 Mg ha^{-1} using LCPRI classifications (Section 2.5).
5. Estimated average AGB density using masked pixels to get LiDAR- and Landsat-based estimates.
6. Used the final SE model to estimate SE based on Landsat-based AGB density, % Forest, area, and perimeter (Section 2.8).
7. Built 95% CI around the Landsat-based estimate using the estimated SE.
8. Compared 95% CIs to LiDAR-based estimate.

Appendix J. Bootstrap variance stability

See Fig. J.1.

Data availability

Data will be made available on request.

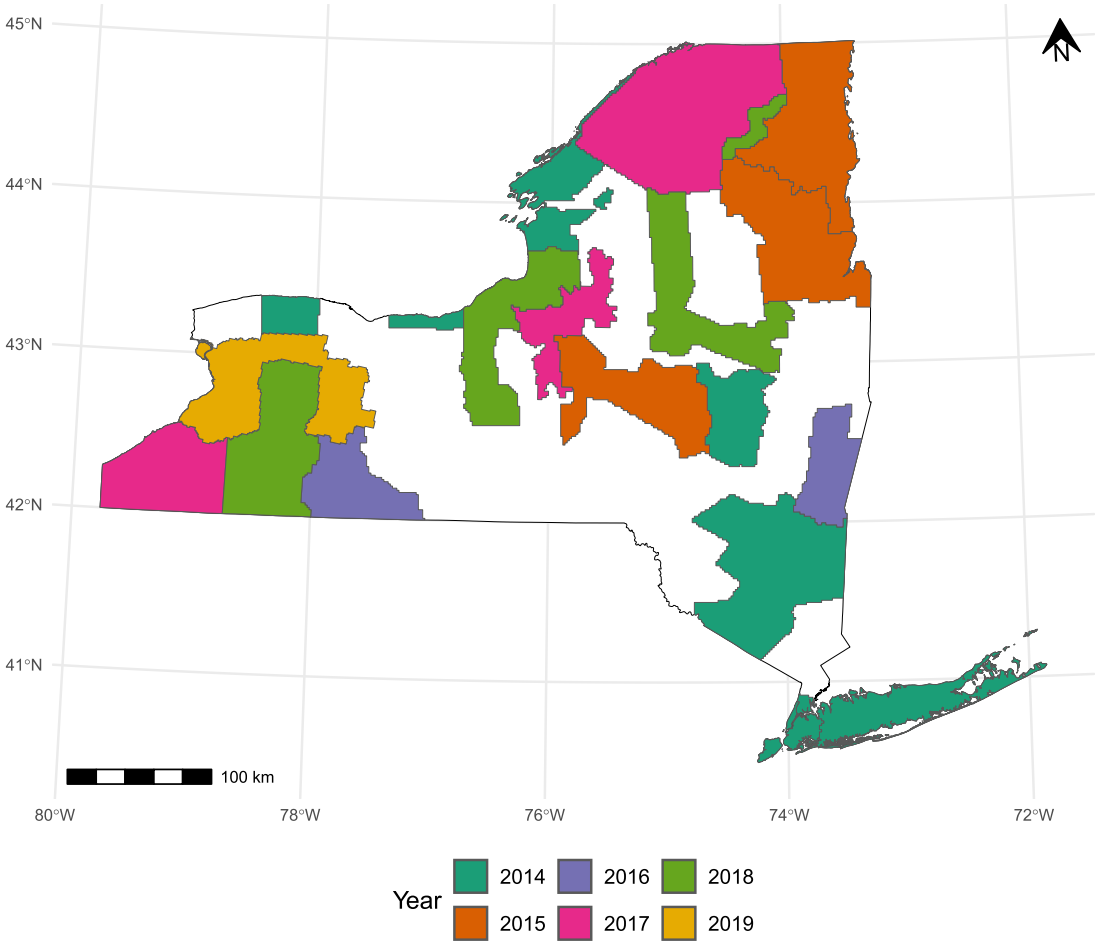

Fig. H.1. Spatial extent of forest aboveground biomass maps developed with aerial LiDAR in Johnson et al. (2022) colored by date of LiDAR collection. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

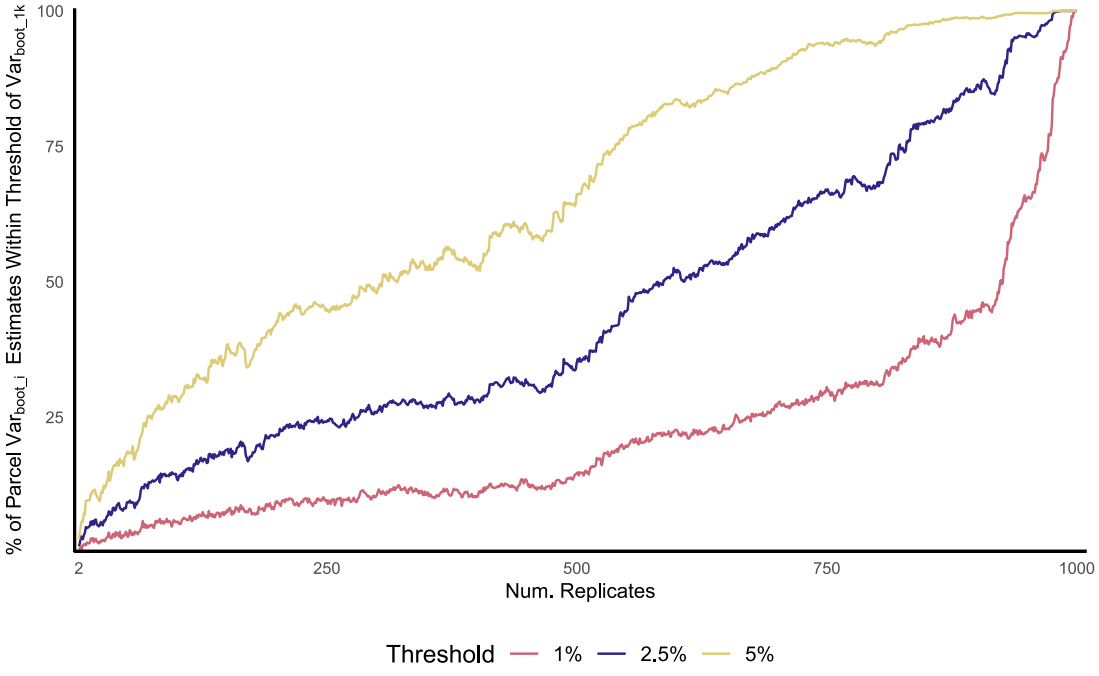

Fig. J.1. Bootstrap variance stability. After i bootstrap replications (x-axis) the percent of parcels (y-axis) with $\widehat{Var_{boot\_i}}$ estimates within a threshold distance (colors) from the $\widehat{Var_{boot}}$ estimate made with 1000 replicates ($\widehat{Var_{boot\_1k}}$). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

References

Baskerville, G.L., 1972. Use of logarithmic regression in the estimation of plant biomass. Can. J. Forest Res. 2 (1), 49–53. http://dx.doi.org/10.1139/x72-009.

Baston, D., 2022. Exactextractr: Fast extraction from raster datasets using polygons. URL: https://CRAN.R-project.org/package=exactextractr, R package version 0.9.1.

Bechtold, W.A., Patterson, P.L., 2005. The Enhanced Forest Inventory and Analysis Program–National Sampling Design and Estimation Procedures, vol. 80, USDA Forest Service, Southern Research Station, URL: https://doi.org/10.2737/SRS-GTR-80.

Berger, A., Gschwantner, T., McRoberts, R.E., Schadauer, K., 2014. Effects of measurement errors on individual tree stem volume estimates for the Austrian National Forest Inventory. For. Sci. 60 (1), 14–24. http://dx.doi.org/10.5849/forsci.12-164.

Breidenbach, J., Antón-Fernández, C., Petersson, H., McRoberts, R.E., Astrup, R., 2014. Quantifying the Model-Related variability of biomass stock and change estimates in the Norwegian National Forest Inventory. For. Sci. 60 (1), 25–33. http://dx.doi.org/10.5849/forsci.12-137.

Breidenbach, J., McRoberts, R.E., Astrup, R., 2016. Empirical coverage of model-based variance estimators for remote sensing assisted estimation of stand-level timber volume. Remote Sens. Environ. 173, 274–281. http://dx.doi.org/10.1016/j.rse.2015.07.026.

Breiman, L., 2001a. Random forests. Mach. Learn. 45 (1), 5–32.

Breiman, L., 2001b. Statistical modeling: The two cultures (with comments and a rejoinder by the author). Statist. Sci. 16 (3), http://dx.doi.org/10.1214/ss/1009213726.

Brown, J.F., Tollerud, H.J., Barber, C.P., Zhou, Q., Dwyer, J.L., Vogelmann, J.E., Loveland, T.R., Woodcock, C.E., Stehman, S.V., Zhu, Z., Pengra, B.W., Smith, K., Horton, J.A., Xian, G., Auch, R.F., Sohl, T.L., Sayler, K.L., Gallant, A.L., Zelenak, D., Reker, R.R., Rover, J., 2020. Lessons learned implementing an operational continuous United States national land change monitoring capability: The land change monitoring, assessment, and projection (LCMAP) approach. Remote Sens. Environ. 238, 111356. http://dx.doi.org/10.1016/j.rse.2019.111356, Time Series Analysis with High Spatial Resolution Imagery.

Buendia, E., Tanabe, K., Kranjc, A., Baasansuren, J., Fukuda, M., Ngarize, S., Osako, A., Pyrozhenko, Y., Shermanau, P., Federici, S., 2019. Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories, vol. 5, IPCC, Geneva, Switzerland, p. 194.

Butler, B.J., Hewes, J.H., Dickinson, B.J., Andrejczyk, K., Butler, S.M., Markowski-Lindsay, M., 2016. Family forest ownerships of the United States, 2013: Findings from the USDA forest service's national woodland owner survey. J. For. 114 (6), 638–647. http://dx.doi.org/10.5849/jof.15-099.

CEOS, 2021. Aboveground woody biomass product validation good practices protocol. http://dx.doi.org/10.5067/DOC/CEOSWGCV/LPV/AGB.001.

Chave, J., Réjou-Méchain, M., Búrquez, A., Chidumayo, E., Colgan, M.S., Delitti, W.B., Duque, A., Eid, T., Fearnside, P.M., Goodman, R.C., Henry, M., Martínez-Yrízar, A., Mugasha, W.A., Muller-Landau, H.C., Mencuccini, M., Nelson, B.W., Ngomanda, A., Nogueira, E.M., Ortiz-Malavassi, E., Pélissier, R., Ploton, P., Ryan, C.M., Saldarriaga, J.G., Vieilledent, G., 2014. Improved allometric models to estimate the aboveground biomass of tropical trees. Global Change Biol. 20 (10), 3177–3190. http://dx.doi.org/10.1111/gcb.12629.

Chen, Q., Laurin, G.V., Valentini, R., 2015. Uncertainty of remotely sensed aboveground biomass over an African tropical forest: Propagating errors from trees to plots to pixels. Remote Sens. Environ. 160, 134–143. http://dx.doi.org/10.1016/j.rse.2015.01.009.

Chen, Q., McRoberts, R.E., Wang, C., Radtke, P.J., 2016. Forest aboveground biomass mapping and estimation across multiple spatial scales using model-based inference. Remote Sens. Environ. 184, 350–360. http://dx.doi.org/10.1016/j.rse.2016.07.023.

Cooke, W.H., 2000. Forest/non-forest stratification in Georgia with landsat thematic mapper data. 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, US Department of Agriculture, Forest Service, North Central Research Station, St. Paul, MN, pp. 28–30.

COP, U., 2015. Paris agreement to the united nations framework convention on climate change. URL: https://treaties.un.org/doc/Publication/UNTS/Volume%203156/Part/volume-3156-I-54113.pdf.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20 (3), 273–297. http://dx.doi.org/10.1007/bf00994018.

Domke, G.M., Woodall, C.W., Smith, J.E., 2011. Accounting for density reduction and structural loss in standing dead trees: Implications for forest biomass and carbon stock estimates in the United States. Carbon Balance Manag. 6 (1), http://dx.doi.org/10.1186/1750-0680-6-14.

Dubayah, R., Blair, J.B., Goetz, S., Fatoyinbo, L., Hansen, M., Healey, S., Hofton, M., Hurtt, G., Kellner, J., Luthcke, S., Armston, J., Tang, H., Duncanson, L., Hancock, S., Jantz, P., Marselis, S., Patterson, P.L., Qi, W., Silva, C., 2020. The global ecosystem dynamics investigation: High-resolution laser ranging of the Earth's forests and topography. Sci. Remote. Sens. 1, 100002. http://dx.doi.org/10.1016/j.srs.2020.100002.

Dyer, J.M., 2006. Revisiting the deciduous forests of Eastern North America. BioSci. 56 (4), 341–352. http://dx.doi.org/10.1641/0006-3568(2006)56[341:RTDFOE]2.0.CO;2.

Efron, B., 2020. Prediction, estimation, and attribution. J. Amer. Statist. Assoc. 115 (530), 636–655. http://dx.doi.org/10.1080/01621459.2020.1762613.

Efron, B., Tibshirani, R.J., 1994. An Introduction to the Bootstrap. CRC Press.

Eggleston, H.S., Buendia, L., Miwa, K., Ngara, T., Tanabe, K., 2006. 2006 IPCC guidelines for national greenhouse gas inventories.

EROS, Earth Resources Observation And Science (EROS) Center, 2019. National agriculture imagery program (NAIP). http://dx.doi.org/10.5066/F7QN651G.

Esteban, J., McRoberts, R.E., Fernández-Landa, A., Tomé, J.L., Marchamalo, M., 2020. A model-based volume estimator that accounts for both land cover misclassification and model prediction uncertainty. Remote. Sens. 12 (20), 3360. http://dx.doi.org/10.3390/rs12203360.

Freedman, D.A., 1981. Bootstrapping regression models. Ann. Statist. 9 (6), 1218–1228.

Friedman, J.H., 2002. Stochastic gradient boosting. Comput. Statist. Data Anal. 38 (4), 367–378. http://dx.doi.org/10.1016/S0167-9473(01)00065-2, Nonlinear Methods and Data Mining.

Gräler, B., Pebesma, E., Heuvelink, G., 2016. Spatio-Temporal interpolation using gstat. R. J. 8 (1), 204–218. http://dx.doi.org/10.32614/RJ-2016-014.

Gray, A.N., Brandeis, T.J., Shaw, J.D., McWilliams, W.H., Miles, P., 2012. Forest inventory and analysis database of the United States of America (FIA). Biodivers. Ecol. 4, 225–231. http://dx.doi.org/10.7809/b-e.00079.

Harmon, M.E., Woodall, C.W., Fasth, B., Sexton, J., Yatkov, M., 2011. Differences between standing and downed dead tree wood density reduction factors: A comparison across decay classes and tree species. http://dx.doi.org/10.2737/nrs-rp-15.

Hastie, T.J., 1992. Generalized additive models. In: Bates, D., Chambers, J., Hastie, T. (Eds.), Chapter 7 of Statistical Models in S. Routledge.

Hijmans, R.J., 2023. Terra: Spatial data analysis. URL: https://CRAN.R-project.org/package=terra, R package version 1.7-39.

Hoppus, M., Lister, A., 2005. The status of accurately locating forest inventory and analysis plots using the global positioning system. In: Proceedings of the Seventh Annual Forest Inventory and Analysis Symposium. URL: https://www.fs.usda.gov/research/treesearch/17040.

Huang, W., Dolan, K., Swatantran, A., Johnson, K., Tang, H., O'Neil-Dunne, J., Dubayah, R., Hurtt, G., 2019. High-resolution mapping of aboveground biomass for forest carbon monitoring system in the Tri-State region of Maryland, Pennsylvania and Delaware, USA. Environ. Res. Lett. 14 (9), 095002. http://dx.doi.org/10.1088/1748-9326/ab2917.

Hudak, A.T., Fekety, P.A., Kane, V.R., Kennedy, R.E., Filippelli, S.K., Falkowski, M.J., Tinkham, W.T., Smith, A.M.S., Crookston, N.L., Domke, G.M., Corrao, M.V., Bright, B.C., Churchill, D.J., Gould, P.J., McGaughey, R.J., Kane, J.T., Dong, J., 2020. A carbon monitoring system for mapping regional, annual aboveground biomass across the northwestern USA. Environ. Res. Lett. 15 (9), 095003. http://dx.doi.org/10.1088/1748-9326/ab93f9.

Jenkins, J.C., Chojnacky, D.C., Heath, L.S., Birdsey, R.A., 2003. National-scale biomass estimators for United States tree species. For. Sci. 49 (1), 12–35. http://dx.doi.org/10.1093/forestscience/49.1.12.

Johnson, K.D., Birdsey, R., Cole, J., Swatantran, A., O'Neil-Dunne, J., Dubayah, R., Lister, A., 2015. Integrating LIDAR and forest inventories to fill the trees outside forests data gap. Environ. Monit. Assess. 187 (10), http://dx.doi.org/10.1007/s10661-015-4839-1.

Johnson, L.K., Mahoney, M.J., Bevilacqua, E., Stehman, S.V., Domke, G.M., Beier, C.M., 2022. Fine-resolution landscape-scale biomass mapping using a spatiotemporal patchwork of LiDAR coverages. Int. J. Appl. Earth Obs. Geoinf. 114, 103059. http://dx.doi.org/10.1016/j.jag.2022.103059.

Johnson, L.K., Mahoney, M.J., Desrochers, M.L., Beier, C.M., 2023. Mapping historical forest biomass for stock-change assessments at parcel to landscape scales. Forest Ecol. Manag. 546, 121348. http://dx.doi.org/10.1016/j.foreco.2023.121348.

Karatzoglou, A., Smola, A., Hornik, K., Zeileis, A., 2004. Kernlab – An S4 package for kernel methods in R. J. Stat. Softw. 11 (9), 1–20. http://dx.doi.org/10.18637/jss.v011.i09.

Kattenborn, T., Leitloff, J., Schiefer, F., Hinz, S., 2021. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. ISPRS J. Photogramm. Remote Sens. 173, 24–49. http://dx.doi.org/10.1016/j.isprsjprs.2020.12.010.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y., 2017. LightGBM: A highly efficient gradient boosting decision tree. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), Advances in Neural Information Processing Systems. Vol. 30, Curran Associates, Inc., URL: https://proceedings.neurips.cc/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf.

Kennedy, R.E., Ohmann, J., Gregory, M., Roberts, H., Yang, Z., Bell, D.M., Kane, V., Hughes, M.J., Cohen, W.B., Powell, S., Neeti, N., Larrue, T., Hooper, S., Kane, J., Miller, D.L., Perkins, J., Braaten, J., Seidl, R., 2018a. An empirical, integrated forest biomass monitoring system. Environ. Res. Lett. 13 (2), 025004. http://dx.doi.org/10.1088/1748-9326/aa9d9e.

Kennedy, R.E., Yang, Z., Cohen, W.B., 2010. Detecting trends in forest disturbance and recovery using yearly landsat time series: 1. LandTrendr — Temporal segmentation algorithms. Remote Sens. Environ. 114 (12), 2897–2910. http://dx.doi.org/10.1016/j.rse.2010.07.008.

Kennedy, R.E., Yang, Z., Gorelick, N., Braaten, J., Cavalcante, L., Cohen, W.B., Healey, S., 2018b. Implementation of the LandTrendr algorithm on google earth engine. Remote. Sens. 10 (5), http://dx.doi.org/10.3390/rs10050691.

Knapp, A.K., Briggs, J.M., Collins, S.L., Arecher, S.R., Bret-Harte, M.S., Ewers, B.E., Peters, D.P., Young, D.R., Shaver, G.R., Pendall, E., Cleary, M.B., 2007. Shrub encroachment in North American grasslands: shifts in growth form dominance rapidly alters control of ecosystem carbon inputs. Global Change Biol. 14 (3), 615–623. http://dx.doi.org/10.1111/j.1365-2486.2007.01512.x.

Kuhn, M., Vaughan, D., Hvitfeldt, E., 2023. Yardstick: Tidy characterizations of model performance. URL: https://CRAN.R-project.org/package=yardstick, R package version 1.2.0.

Labrière, N., Davies, S.J., Disney, M.I., Duncanson, L.I., Herold, M., Lewis, S.L., Phillips, O.L., Quegan, S., Saatchi, S.S., Schepaschenko, D.G., Scipal, K., Sist, P., Chave, J., 2022. Toward a forest biomass reference measurement system for remote sensing applications. Global Change Biol. 29 (3), 827–840. http://dx.doi.org/10.1111/gcb.16497.

Landau, W.M., 2021. The targets R package: a dynamic make-like function-oriented pipeline toolkit for reproducibility and high-performance computing. J. Open Source Softw. 6 (57), 2959, URL: https://doi.org/10.21105/joss.02959.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nat. 521 (7553), 436–444. http://dx.doi.org/10.1038/nature14539.

Liu, R.Y., 1988. Bootstrap procedures under some non-I.I.D. Models. Ann. Statist. 16 (4), http://dx.doi.org/10.1214/aos/1176351062.

Liu, S., Brandt, M., Nord-Larsen, T., Chave, J., Reiner, F., Lang, N., Tong, X., Ciais, P., Igel, C., Pascual, A., Guerra-Hernandez, J., Li, S., Mugabowindekwe, M., Saatchi, S., Yue, Y., Chen, Z., Fensholt, R., 2023. The overlooked contribution of trees outside forests to tree cover and woody biomass across Europe. Sci. Adv. 9 (37), http://dx.doi.org/10.1126/sciadv.adh4097.

L'Roe, A.W., Allred, S.B., 2013. Thriving or surviving? Forester responses to private forestland parcelization in New York State. Small- Scale For. 12 (3), 353–376, URL: https://doi.org/10.1007/s11842-012-9216-0.

Mahoney, M.J., Johnson, L.K., Guinan, A.Z., Beier, C.M., 2022. Classification and mapping of low-statured shrubland cover types in post-agricultural landscapes of the US Northeast. Int. J. Remote Sens. 43 (19–24), 7117–7138. http://dx.doi.org/10.1080/01431161.2022.2155086.

Matasci, G., Hermosilla, T., Wulder, M.A., White, J.C., Coops, N.C., Hobart, G.W., Bolton, D.K., Tompalski, P., Bater, C.W., 2018. Three decades of forest structural dynamics over Canada's forested ecosystems using Landsat time-series and lidar plots. Remote Sens. Environ. 216, 697–714. http://dx.doi.org/10.1016/j.rse.2018.07.024.

McRoberts, R.E., 2011. Satellite image-based maps: Scientific inference or pretty pictures? Remote Sens. Environ. 115 (2), 715–724. http://dx.doi.org/10.1016/j.rse.2010.10.013.

McRoberts, R.E., Chen, Q., Domke, G.M., Ståhl, G., Saarela, S., Westfall, J.A., 2016. Hybrid estimators for mean aboveground carbon per unit area. Forest Ecol. Manag. 378, 44–56. http://dx.doi.org/10.1016/j.foreco.2016.07.007.

McRoberts, R.E., Næsset, E., Gobakken, T., Chirici, G., Condés, S., Hou, Z., Saarela, S., Chen, Q., Ståhl, G., Walters, B.F., 2018. Assessing components of the model-based mean square error estimator for remote sensing assisted forest applications. Can. J. Forest Res. 48 (6), 642–649. http://dx.doi.org/10.1139/cjfr-2017-0396.

McRoberts, R.E., Næsset, E., Hou, Z., Ståhl, G., Saarela, S., Esteban, J., Travaglini, D., Mohammadi, J., Chirici, G., 2023. How many bootstrap replications are necessary for estimating remote sensing-assisted, model-based standard errors? Remote Sens. Environ. 288, 113455. http://dx.doi.org/10.1016/j.rse.2023.113455.

McRoberts, R.E., Næsset, E., Saatchi, S., Quegan, S., 2022. Statistically rigorous, model-based inferences from maps. Remote Sens. Environ. 279, 113028. http://dx.doi.org/10.1016/j.rse.2022.113028.

McRoberts, R.E., Tomppo, E.O., Næsset, E., 2010. Advances and emerging issues in national forest inventories. Scand. J. For. Res. 25 (4), 368–381. http://dx.doi.org/10.1080/02827581.2010.496739.

McRoberts, R.E., Westfall, J.A., 2014. Effects of uncertainty in model predictions of individual tree volume on large area volume estimates. For. Sci. 60 (1), 34–42. http://dx.doi.org/10.5849/forsci.12-141.

Meyer, H., Pebesma, E., 2021. Predicting into unknown space? Estimating the area of applicability of spatial prediction models. Methods Ecol. Evol. 12 (9), 1620–1633. http://dx.doi.org/10.1111/2041-210x.13650.

Nelson, M.D., Liknes, G.C., Butler, B.J., 2010. Map of forest ownership in the conterminous United States. [Scale 1:7, 500, 000]. http://dx.doi.org/10.2737/nrs-rmap-2.

Pan, Y., Birdsey, R.A., Fang, J., Houghton, R., Kauppi, P.E., Kurz, W.A., Phillips, O.L., Shvidenko, A., Lewis, S.L., Canadell, J.G., Ciais, P., Jackson, R.B., Pacala, S.W., McGuire, A.D., Piao, S., Rautiainen, A., Sitch, S., Hayes, D., 2011. A large and persistent carbon sink in the world's forests. Sci. 333 (6045), 988–993. http://dx.doi.org/10.1126/science.1201609.

Pebesma, E.J., 2004. Multivariable geostatistics in s: the gstat package. Comput. Geosci. 30 (7), 683–691. http://dx.doi.org/10.1016/j.cageo.2004.03.012.

Penman, J., Gytarski, M., Hiraishi, T., Krug, T., Kruger, D., Pipatti, R., Buendia, L., Miwa, K., Ngara, T., Tanabe, K., Wagner, F., 2003. Good practice guidance for land use, land-use change and forestry. URL: http://www.ipcc-nggip.iges.or.jp/public/gpglulucf/gpglulucf.htm.

Perry, C.H., Finco, M.V., Barry, T., et al., 2022. Forest atlas of the United States. URL: https://research.fs.usda.gov/treesearch/64468.

Quegan, S., Le Toan, T., Chave, J., Dall, J., Exbrayat, J.-F., Minh, D.H.T., Lomas, M., D'Alessandro, M.M., Paillou, P., Papathanassiou, K., Rocca, F., Saatchi, S., Scipal, K., Shugart, H., Smallman, T.L., Soja, M.J., Tebaldini, S., Ulander, L., Villard, L., Williams, M., 2019. The European Space Agency BIOMASS mission: Measuring forest above-ground biomass from space. Remote Sens. Environ. 227, 44–60. http://dx.doi.org/10.1016/j.rse.2019.03.032.

R Core Team, 2023. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL: https://www.R-project.org/.

Radtke, P., Walker, D., Weiskittel, A., Frank, J., Coulston, J., Westfall, J., 2015. Legacy tree data: a national database of detailed tree measurements for volume, weight, and physical properties. In: Pushing Boundaries: New Directions in Inventory Techniques and Applications: Forest Inventory and Analysis (FIA) Symposium. Vol. 2015, pp. 8–10, URL: https://www.fs.usda.gov/research/treesearch/50166.

Raile, G.K., 1982. Estimating Stump Volume. Technical Report, U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station, http://dx.doi.org/10.2737/nc-rp-224.

Rogers, P., 1996. Disturbance ecology and forest management: a review of the literature. URL: https://www.fs.usda.gov/rm/pubs_int/int_gtr336.pdf.

Ross, R.J., 2021. Wood handbook: wood as an engineering material. URL: https://www.fs.usda.gov/research/treesearch/62200.

Saarela, S., Holm, S., Grafström, A., Schnell, S., Næsset, E., Gregoire, T.G., Nelson, R.F., Ståhl, G., 2016. Hierarchical model-based inference for forest inventory utilizing three sources of information. Ann. For. Sci. 73 (4), 895–910. http://dx.doi.org/10.1007/s13595-016-0590-1.

Saarela, S., Holm, S., Healey, S., Andersen, H.-E., Petersson, H., Prentius, W., Patterson, P., Næsset, E., Gregoire, T., Ståhl, G., 2018. Generalized hierarchical Model-Based estimation for aboveground biomass assessment using GEDI and landsat data. Remote. Sens. 10 (11), 1832. http://dx.doi.org/10.3390/rs10111832.

Saarela, S., Wästlund, A., Holmström, E., Mensah, A.A., Holm, S., Nilsson, M., Fridman, J., Ståhl, G., 2020. Mapping aboveground biomass and its prediction uncertainty using LiDAR and field data, accounting for tree-level allometric and LiDAR model errors. For. Ecosyst. 7 (1), http://dx.doi.org/10.1186/s40663-020-00245-0.

Saltelli, A., Tarantola, S., Campolongo, F., Ratto, M., et al., 2004. Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models, vol. 1, Wiley Online Library.

Schnell, S., Altrell, D., Ståhl, G., Kleinn, C., 2014. The contribution of trees outside forests to national tree biomass and carbon stocks—a comparative study across three continents. Environ. Monit. Assess. 187 (1), http://dx.doi.org/10.1007/s10661-014-4197-4.

Schoeneberger, M.M., 2009. Agroforestry: working trees for sequestering carbon on agricultural lands. Agrofor. Syst. 75, 27–37. http://dx.doi.org/10.1007/s10457-008-9123-8.

Scott, C.T., 1981. Northeastern Forest Survey Revised Cubic-Foot Volume Equations, vol. 304, US Department of Agriculture, Forest Service, Northeastern Forest Experiment ..., URL: https://www.fs.usda.gov/ne/newtown_square/publications/research_notes/pdfs/scanned/OCR/ne_rn304.pdf.

Seymour, R.S., White, A.S., deMaynadier, P.G., 2002. Natural disturbance regimes in northeastern North America—evaluating silvicultural systems using natural scales and frequencies. Forest Ecol. Manag. 155 (1–3), 357–367. http://dx.doi.org/10.1016/s0378-1127(01)00572-2.

Shi, Y., Ke, G., Soukhavong, D., Lamb, J., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y., Titov, N., Cortes, D., 2022. Lightgbm: Light gradient boosting machine. URL: https://github.com/Microsoft/LightGBM, R package version 3.3.2.99.

Sprugel, D.G., 1983. Correcting for bias in Log-Transformed allometric equations. Ecol. 64 (1), 209–210. http://dx.doi.org/10.2307/1937343.

Ståhl, G., Heikkinen, J., Petersson, H., Repola, J., Holm, S., 2014. Sample-Based estimation of greenhouse gas emissions from forests–A new approach to account for both sampling and model errors. For. Sci. 60 (1), 3–13. http://dx.doi.org/10.5849/forsci.13-005.

Stehman, S.V., Pengra, B.W., Horton, J.A., Wellington, D.F., 2021. Validation of the U.S. geological survey's land change monitoring, assessment and projection (LCMAP) collection 1.0 annual land cover products 1985–2017. Remote Sens. Environ. 265, 112646. http://dx.doi.org/10.1016/j.rse.2021.112646.

Stovall, A.E., Anderson-Teixeira, K.J., Shugart, H.H., 2018. Assessing terrestrial laser scanning for developing non-destructive biomass allometry. Forest Ecol. Manag. 427, 217–229. http://dx.doi.org/10.1016/j.foreco.2018.06.004.

Udawatta, R.P., Jose, S., 2011. Carbon sequestration potential of agroforestry practices in temperate North America. In: Carbon Sequestration Potential of Agroforestry Systems. Springer Netherlands, pp. 17–42. http://dx.doi.org/10.1007/978-94-007-1630-8_2.

UNFCCC, 1992. The united nations framework convention on climate change. URL: https://treaties.un.org/doc/Publication/UNTS/Volume%201771/v1771.pdf.

UNFCCC, 1997. Kyoto protocol to the united nations framework convention on climate change. URL: https://treaties.un.org/doc/Publication/UNTS/Volume%202303/v2303.pdf.

USDA Forest Service, 2020. Forests of New York, 2019. Technical Report, http://dx.doi.org/10.2737/fs-ru-250.

Van Auken, O.W., 2000. Shrub invasions of North American semiarid grasslands. Annu. Rev. Ecol. Syst. 31 (1), 197–215. http://dx.doi.org/10.1146/annurev.ecolsys.31.1.197.

Wadoux, A.M.J.-C., Heuvelink, G.B.M., 2023. Uncertainty of spatial averages and totals of natural resource maps. Methods Ecol. Evol. 14 (5), 1320–1332. http://dx.doi.org/10.1111/2041-210x.14106.

Webster, R., Oliver, M.A., 2007. Geostatistics for Environmental Scientists. John Wiley & Sons.

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., Poudel, K.P., Frank, J.M., Prisley, S.P., Wang, Y., Meador, A.J.S., Auty, D., Domke, G.M., 2023. A National-Scale Tree Volume, Biomass, and Carbon Modeling System for the United States. Technical Report, http://dx.doi.org/10.2737/wo-gtr-104.

Woodall, C.W., Coulston, J.W., Domke, G.M., Walters, B.F., Wear, D.N., Smith, J.E., Andersen, H.-E., Clough, B.J., Cohen, W.B., Griffith, D.M., Hagen, S.C., Hanou, I.S., Nichols, M.C., Perry, C.H.H., Russell, M.B., Westfall, J., Wilson, B.T.T., 2015. The U.S. Forest Carbon Accounting Framework: Stocks and Stock Change, 1990–2016. Technical Report, http://dx.doi.org/10.2737/nrs-gtr-154.

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 U.S. Forest Inventory, 2010. Technical Report, http://dx.doi.org/10.2737/nrs-gtr-88.

Wright, M.N., Ziegler, A., 2017. Ranger: A fast implementation of random forests for high dimensional data in C++ and R. J. Stat. Softw. 77 (1), 1–17. http://dx.doi.org/10.18637/jss.v077.i01.

Yanai, R.D., Young, A.R., Campbell, J.L., Westfall, J.A., Barnett, C.J., Dillon, G.A., Green, M.B., Woodall, C.W., 2023. Measurement uncertainty in a national forest inventory: results from the northern region of the USA. Can. J. Forest Res. 53 (3), 163–177. http://dx.doi.org/10.1139/cjfr-2022-0062.

Zhu, Z., Woodcock, C.E., 2014. Continuous change detection and classification of land cover using all available Landsat data. Remote Sens. Environ. 144, 152–171. http://dx.doi.org/10.1016/j.rse.2014.01.011.