
HIGH-RESOLUTION LANDSCAPE-SCALE BIOMASS MAPPING USING A SPATIOTEMPORAL PATCHWORK OF LIDAR COVERAGES

A PREPRINT

Lucas K Johnson

Graduate Program in Environmental Science
State University of New York College of Environmental Science and Forestry
Syracuse, NY 13210
ljohns11@esf.edu

Michael J Mahoney

Graduate Program in Environmental Science
State University of New York College of Environmental Science and Forestry
Syracuse, NY 13210
mjmahone@esf.edu

Eddie Bevilacqua

Department of Sustainable Resources Management
State University of New York College of Environmental Science and Forestry
Syracuse, NY 13210
ebevilacqua@esf.edu

Stephen V Stehman

Department of Sustainable Resources Management
State University of New York College of Environmental Science and Forestry
Syracuse, NY 13210
svstehma@esf.edu

Grant Domke

Northern Research Station
USDA Forest Service
St. Paul, MN 55114
grant.m.domke@usda.gov

Colin M Beier

Department of Sustainable Resources Management
State University of New York College of Environmental Science and Forestry
Syracuse, NY 13210
cbeier@esf.edu

May 18, 2022

Abstract

Estimating forest aboveground biomass at fine spatial scales has become increasingly important for greenhouse gas estimation, monitoring, and verification efforts to mitigate climate

change. Airborne LiDAR continues to be a valuable source of remote sensing data for estimating aboveground biomass. However airborne LiDAR collections may take place at local or regional scales covering irregular, non-contiguous footprints, resulting in a ‘patchwork’ of different landscape segments at different points in time. Here we addressed common obstacles including selection of training data, the investigation of regional or coverage specific patterns in bias and error, and map agreement, and model-based precision assessments at multiple scales.

Three machine learning algorithms and an ensemble model were trained using field inventory data (FIA), airborne LiDAR, and topographic, climatic and cadastral geodata. Using strict selection criteria, 801 FIA plots were selected with co-located point clouds drawn from a patchwork of 17 leaf-off LiDAR coverages (2014-2019). Our ensemble model created 30m AGB prediction surfaces within a predictor-defined area of applicability (98% of LiDAR coverage) and resulting AGB predictions were compared with FIA plot-level and areal estimates at multiple scales of aggregation. Our model was overall accurate (% RMSE 13-33%), had very low bias ($MBE \leq \pm 5 \text{ Mg ha}^{-1}$), explained most field-observed variation ($R^2 0.74\text{-}0.93$), produced estimates that were both largely consistent with FIA’s aggregate summaries (86% of estimates within 95% CI), as well as precise when aggregated to arbitrary small-areas (mean bootstrap standard error 0.37 Mg ha^{-1}). We share practical solutions to challenges faced when using spatiotemporal patchworks of LiDAR to meet growing needs for biomass prediction and mapping, and applications in carbon accounting and ecosystem stewardship.

Keywords Airborne LiDAR · Aboveground biomass · Machine learning · Model ensembles · Forest Inventory and Analysis (FIA)

1 Introduction

Mapping and monitoring forest aboveground biomass (AGB) is becoming increasingly important as federal, state, and global agencies aim to offset greenhouse gas emissions to mitigate a warming climate and the myriad resulting challenges. Field sampling programs, like the United States Department of Agriculture’s Forest Inventory and Analysis program (FIA) (Gray et al. 2012), provide unbiased estimates of aboveground biomass (AGB) and forest carbon over large areas through a design-based approach (Bechtold and Patterson 2005). Despite the inherent benefits to this approach (unbiased estimators; computationally efficient), the scale at which design-based estimates can be produced is limited by the density of the sample (McRoberts 2011).

For instance, in New York State (NYS) there is roughly one FIA plot for every 2,400 hectares. However, 73% of NYS forests are privately owned (USDA Forest Service 2020), and these parcels are increasingly undergoing further subdivision (L’Roe and Allred 2013), making it difficult to rely solely on FIA data for AGB and carbon estimation at scales relevant to the landowner base in NYS. On the other hand, model-based approaches combining FIA reference data with auxiliary RS data can produce predictions for all map units (pixels) in a given area, providing the flexibility to aggregate these individual estimates to units of arbitrary size up to and including aggregations at the scale of estimates produced with FIA data alone. Thus, AGB predictions from model-based approaches are needed for small-area estimation if AGB and carbon stock information is to be factored into forest management decisions without further additional investments to increase the sample density in NYS (McRoberts 2011; CEOS 2021).

Airborne LiDAR has been established as a highly valuable remote sensing (RS) data source for the purposes of AGB mapping (Huang et al. 2019; Hurt et al. 2019; Chen and McRoberts 2016). However, these data are most commonly acquired at local to regional scales in irregular or non-contiguous footprints (Skowronski and Lister 2012). This creates a complex ‘patchwork’ of data with component coverages acquired at different times with different sensors and mission parameters, creating a host of challenges for broad-scale mapping that can be a barrier to adoption (Lu et al. 2014; Huang et al. 2019). These challenges include lacking field inventory plots that spatially and temporally match LiDAR acquisitions and data discrepancies between LiDAR coverages.

Despite these challenges, several groups have undertaken broad-scale AGB mapping efforts with LiDAR patchworks. Among them, there are varying degrees to which training data have been pooled from multiple LiDAR coverages. The choice between an individual or a pooled modeling approach often reflects practical considerations relating to sufficient sample size across all sub-regions and the cost of developing multiple models. Nilsson et al. (2017) did not pool at all, implementing a separate model trained for each coverage.

Huang et al. (2019) pooled by ecoregion. Both Ayrey et al. (2021) and Hauglin et al. (2021) pooled all coverages but used a convolutional neural network and a mixed-effects model respectively, with differing protocols for inventory plot selection.

In this study, we addressed several common challenges in using LiDAR patchworks for broad-scale, high-resolution biomass modeling and mapping. We leveraged FIA inventories for model training and assessment data, and implemented FIA-developed methods to assess the agreement between our estimates and those produced by FIA. With the goal of producing a spatially explicit representation of FIA AGB information, we used a model-based approach to transfer FIA's discrete plot-level estimates to wall-to-wall estimates at a 30m resolution across a patchwork of 17 discrete LiDAR coverages in NYS.

We implemented a rigorous plot selection framework to limit temporal lags between LiDAR acquisitions and field inventories. When strict temporal alignment yielded too few plots, we leveraged repeated FIA inventories to boost the number of plots that temporally match LiDAR acquisitions for accurate model predictions. In spite of this strategy, we were left with limited model training data, where some coverages and regions lacked sufficient information to support independent models, necessitating a single model or 'pooled' approach.

We used machine learning (ML) techniques including random forests (Breiman 2001), gradient boosting machines (Friedman 2002), and support vector machines (Cortes and Vapnik 1995) to extract as much information as possible from the available training data to produce accurate predictions.

We employed multiple strategies to address concerns that differences among sensors and mission parameters could lead to non-randomly distributed errors. First, we produced an area of applicability (Meyer and Pebesma 2021) mask to both examine the uniformity of our predictors across the component coverages, as well as to screen predictions based on anomalous predictor data. Second, we examined the spatial autocorrelation of our residuals and mapped our prediction error and bias to identify the presence of region or coverage specific patterns.

To demonstrate the suitability of our predictions for AGB density estimation at mid-range scales (towns, counties) we assessed the agreement between our mapped estimates and FIA reference datasets using FIA-developed methods (Riemann et al. 2010; Menlove and Healey 2020). To show that our predictions can be leveraged for AGB estimation within arbitrary small areas, like individual parcels or properties, we computed the precision of aggregate estimates in a model-based framework (McRoberts 2011).

2 Data and Methods

2.1 LiDAR Coverages and Pre-Processing

Our study relied upon a set of 17 LiDAR datasets hosted by the NYS GIS Program Office (GPO) covering 62.46% (7,835,690 ha) of NYS (hereafter "GPO-LiDAR region"; Table 2; Figure 1). We selected individual coverages from the most recent five years of available data that contained temporally matching field data (2014-2019) to minimize sensor and data differences. All component coverages were collected to generate digital elevation models for flood risk analysis, and to this end were flown during leaf-off conditions. Several previous studies have shown that leaf-off LiDAR models can be as accurate as their leaf-on counterparts (Hawbaker et al. 2010; White et al. 2015; Anderson and Bolstad 2013).

First, we height-normalized the raw LiDAR data and computed 40 predictors (Supplementary Materials S1) based on previous studies (Hawbaker et al. 2010; Huang et al. 2019; Pflugmacher et al. 2014). Predictors at individual FIA plots were computed from LiDAR point clouds clipped to each FIA subplot then pooled at the plot level, and the corresponding predictors computed for map pixels were based on the set of returns within each 30m cell. The lidR (Roussel and Auty 2020; Jean-Romain et al. 2020) package in R (R Core Team 2021) was used for height-normalization and predictor generation.

2.2 Field Data

AGB estimates for trees ≥ 12.7 cm (5 in) diameter at breast height were compiled as part of the United States Department of Agriculture (USDA) Forest Inventory and Analysis (FIA) program (Gray et al. 2012), and were converted to units of megagrams per hectare ($Mg\ ha^{-1}$). The plot locations were provided by the USDA in the form of average coordinates, collected over multiple repeat visits, for the center subplot which we then used to construct the entire plot layout including all four circular subplots.

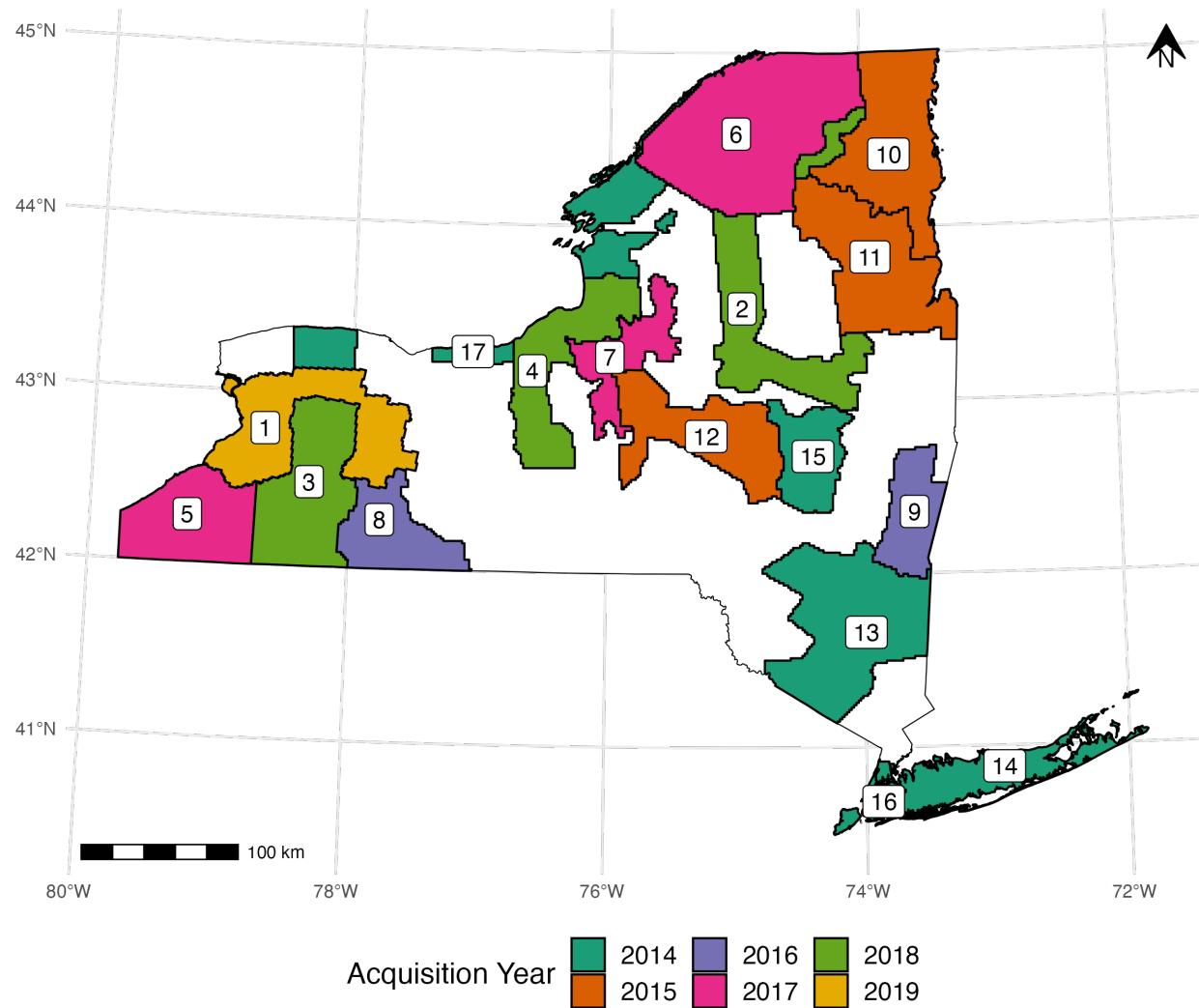


Figure 1: GPO-LiDAR region component coverages, colored by year of acquisition and labeled by ID numbers in Table 1.

Table 1: Summary of plot selection criteria.

	Criteria	Description	Num Plots
Include	1	Plots inventoried in the same year as a LiDAR acquisition.	324
	2	Plots growth-adjusted to temporally align with LiDAR acquisition.	843
Exclude	3	Plots where measured AGB = 0 Mg ha ⁻¹ but maximum LiDAR return > 1m.	327
	4	Plots where the convex hull of LiDAR returns clipped to a given subplot did not contain at least 90% of the subplot area.	13
	5	Plots duplicated in neighboring coverages. Temporal matches given preference, followed by inventory recency.	26
	6	Plots in areas of coverage overlap where temporal alignment does not match a newer coverage superseding an older coverage.	15

When selecting plots for model development and map assessment we aimed to maximize the number of reference plots (Fassnacht et al. 2014), while minimizing the temporal lag between LiDAR acquisitions and inventories (White et al. 2013; CEOS 2021), and ensuring high quality co-registered LiDAR data. Temporal misalignment between plots and LiDAR has been shown to introduce error in models; Gonçalves et al. (2017) found that 7-17% of their model error could be attributed to 3-year lags between inventories and LiDAR acquisitions.

We only considered FIA plots when all subplots were marked as “sampled” and when forest conditions were uniform at the plot level (all subplots completely forested or completely non-forested). Only 324 plots were available when we required a strict temporal match between inventories and LiDAR acquisitions (Table 1; Criterion 1). To increase the number of reference plots, we applied a growth adjustment to FIA plots with inventories both before and after the LiDAR acquisition, or ‘bracketing inventories.’ Of plots with bracketing inventories, we excluded those where AGB decreased by $\geq 5\%$ between measurements indicating a disturbance event on the plot. AGB at the time of the LiDAR collection was then estimated by linearly interpolating between bracketing FIA estimates. This procedure added 843 plots to our dataset (Table 1; Criterion 2), and is analogous to previously published growth adjustment methods (Gonçalves et al. 2017; Gobakken and Næsset 2008), though avoids the development or incorporation of additional growth and yield models.

Several plot exclusion rules were developed to filter duplicates and remove problematic observations resulting from vegetation in non-forest conditions, interfering structures, or other data anomalies (Table 1 Criteria 2-6). Point clouds were clipped to the constructed plot footprints and were excluded with criteria 3-5 when selecting plots for model development (model dataset hereafter; Table 2). The map assessment dataset (assessment dataset hereafter; Table 2), intended for comparing FIA plots to mapped pixels, excluded plots using criteria 3 and 6, but not criteria 4 or 5, since the precise alignment of LiDAR returns and FIA plots was not necessary, and we did not want to assess the final maps with temporally mismatching plots.

2.3 Auxiliary Data

A group of “steady-state” predictors was included to represent geospatial variation in climate and topography (Kennedy et al. 2018). Additionally, a 2019 tax parcel layer was incorporated as a set of boolean indicator

Table 2: Component LiDAR coverage metadata. IDs for cross figure correspondence; Year of acquisition; Area covered (ha); Pulse density (PD) in pulses per m²; Number of FIA plots in model and assessment datasets. Area of Applicability (AOA) in percent of LiDAR coverage pixels considered inside of the area of applicability. AOA computation conducted after initial LCMAP masking.

ID	Name	Year	Area	PD	% AOA	Model	Assessment
1	Erie, Genesee & Livingston	2019	555,853	2.04	97.72	11	11
2	Fulton, Saratoga, Herkimer & Franklin	2018	557,421	2.60	99.61	39	42
3	Southwest B	2018	527,075	1.98	98.06	28	26
4	Cayuga & Oswego	2018	438,201	2.78	96.89	25	25
5	Southwest	2017	423,714	1.98	98.53	33	33
6	Franklin & St. Lawrence	2017	977,620	2.69	99.24	104	103
7	Oneida Subbasin	2017	264,886	2.10	96.02	15	15
8	Allegany & Steuben	2016	309,081	1.69	97.77	32	31
9	Columbia & Rensselaer	2016	248,839	1.69	98.19	17	17
10	Clinton, Essex & Franklin	2015	600,755	2.23	98.70	115	111
11	Warren, Washington & Essex	2015	611,704	3.24	99.37	106	110
12	Madison & Otsego	2015	471,564	2.13	99.24	56	56
13	3 County	2014	755,629	2.04	96.12	92	94
14	Long Island	2014	315,542	2.04	94.77	22	23
15	Schoharie	2014	256,464	2.04	95.12	31	31
16	New York City (NYC)	2014	77,211	1.54	90.40	2	2
17	Great Lakes	2014	444,215	2.04	98.47	73	69
Total			7,835,773		98.12	801	799

variables (Supplementary Materials S2). Tax codes and categories provide cadastral information related to land-use and management (Thompson et al. 2011).

2.4 Model Development

Three ML models were fit to a randomly selected 80% of the model dataset (training set; n = 630), leaving the 20% remaining plots as an independent testing set to assess the final model performance on point clouds clipped to FIA plot footprints (testing set; n = 171). Random forest models (RF; ranger; Wright and Ziegler (2017)), stochastic gradient boosting machines (GBM; lightgbm; Ke et al. (2021), Ke et al. (2017)), and support vector machines (SVM; kernlab; Karatzoglou et al. (2004)) were developed with the training set. We fit a final stacked ensemble to predictions from these three component models using linear regression to reduce the generalization error of our component models (Wolpert 1992). Further description of the model development process can be found in Supplementary Materials S3.

2.5 AGB Mapping and Post-processing

The linear model ensemble was used to make predictions for all 30m pixels within the GPO-LiDAR region. Predictions on newer coverages superseded those on older coverages in areas where neighboring LiDAR coverages overlapped.

With recognition that our predictions are best suited to areas populated by woody biomass, we tabulated our predictions across various landcover types using the United States Geological Survey's Land Change Monitoring, Assessment, and Projection (LCMAP) primary classification products (LCPRI) (Brown et al.

2020; Zhu and Woodcock 2014). LCMAP's annual resolution (1985-2019) allowed for temporal alignment with the patchwork of LiDAR-AGB surfaces and tabulation of our predictions by seven LCPRI classes. We also used LCPRI to remove from the analysis pixel-level predictions classified as Developed, Water, or Barren.

Additionally, we computed an area of applicability (AOA) surface which identifies pixels containing predictor data that are beyond a pre-specified predictor-space distance from the training data (Meyer and Pebesma (2021); Supplementary Materials S4). We used this surface to further mask our LCPRI-masked prediction surfaces, restricting our model to only predict in areas that were well-represented in the training data.

2.6 Model Performance

Model performance was assessed against the 20% testing partition of the model dataset (Table 2) based on metrics including root-mean-squared error in Mg ha⁻¹ (RMSE, equation (1)), percent RMSE (% RMSE, equation (2)), mean absolute error in Mg ha⁻¹ (MAE, equation (3)), percent MAE (% MAE, equation (4)), mean bias error in Mg ha⁻¹ (MBE, equation (5)), and the coefficient of determination (R^2 , equation (6)) as follows:

$$\text{RMSE} = \sqrt{\left(\frac{1}{n}\right) \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (1)$$

$$\% \text{ RMSE} = 100 \cdot \frac{\text{RMSE}}{\bar{y}} \quad (2)$$

$$\text{MAE} = \left(\frac{1}{n}\right) \sum_{i=1}^n (|y_i - \hat{y}_i|) \quad (3)$$

$$\% \text{ MAE} = 100 \cdot \frac{\text{MAE}}{\bar{y}} \quad (4)$$

$$\text{MBE} = \left(\frac{1}{n}\right) \sum_{i=1}^n (y_i - \hat{y}_i) \quad (5)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (6)$$

where n is the number of FIA plots included in the data set, \hat{y}_i is the predicted value of AGB, y_i the AGB value measured at the corresponding location, and \bar{y} the mean AGB value from FIA field measurements.

2.7 Map Agreement Assessment

We assessed the agreement between our masked predictions and FIA reference data following approaches prescribed in Riemann et al. (2010) and Menlove and Healey (2020). The former evaluated agreement across a range of scales and accounts for the mismatch in spatial support between map aggregate estimates (many pixels) and FIA aggregate estimates (few plots) by only extracting pixels coincident with FIA plots. The latter compared FIA-derived AGB estimates — which have been adjusted for forest cover within, and area-extrapolated to, hexagon map units — to aggregations of our pixel predictions.

Following Riemann et al. (2010) we compared masked predictions to the assessment dataset (Section 2.2). Comparisons were made at both the plot-to-pixel scale as well as within hexagons with distances between centroids ranging from 10km (8660 ha) to 100km (866,025 ha). As an extension of the Riemann et al. (2010) methodology we assessed bias (MBE) and error (RMSE, MAE) with choropleth maps that summarized the mapped residuals and FIA reference data distributions within units with centroids spaced 50km apart. To evaluate the presence of spatial autocorrelation among mapped residuals, which could indicate the presence of regional or coverage specific bias in our model's predictions, Global Moran's I statistics were computed for search radii ranging from 1 to 50 km (Moran 1950).

Table 3: Summary of reference data and mapped predictions by LCPRI landcover classes. Mean AGB values in Mg ha⁻¹. Total AGB values in millions of metric tons. Area in hectares. AOA expressed as percent of LCPRI classified pixels considered inside of the area of applicability.

LCPRI	Reference Plots			Mapped				
	Num Plots	Mean AGB	Area	% Area	Mean AGB	Total AGB	% AGB	% AOA
Grass/Shrub	12	30.68	196,852	2.89	47.86	9.42	1.46	98.17
Wetland	36	115.13	613,578	9.01	77.22	47.38	7.34	97.64
Cropland	132	1.98	1,748,905	25.68	14.11	24.67	3.82	98.25
Tree cover	599	137.97	4,251,812	62.42	132.66	564.06	87.38	98.13

Following the Menlove and Healey (2020) approach we compared the average of our masked predictions, weighted by the proportion of each pixel intersecting a given hexagon, to a set of FIA-derived estimates for 64,000 ha hexagons representing FIA’s “finest” acceptable scale for the most recent inventory cycle in NYS (2013-2019) (Menlove and Healey 2020). To account for differences in forest definitions, we divided FIA estimates by the total area of vegetated (based on LCPRI Tree cover, Wetland, Cropland, Grass/Shrub) pixels within each hexagon. Lastly, we excluded 7 of 200 hexagons with FIA estimates greater than our largest model dataset observation (425 Mg ha⁻¹) before making comparisons.

The exactextractr (Daniel Baston 2021), spdep (Bivand, Pebesma, and Gomez-Rubio 2013), sf (Pebesma 2018), raster (Hijmans 2021), and terra (Hijmans 2022) packages in R (R Core Team 2021) were used to conduct all analyses described here. Further description of this assessment is included in Supplementary Materials S5.

2.8 Model-Based Precision

Model-based precision was estimated for aggregations of pixel predictions following the bootstrapping approach described by McRoberts (2011). A simple random sample of 1000 points was drawn from within the GPO-LiDAR region and these points were used as centroids to create 1000 circular aggregation units sized between 1 and 500 ha. 100 different predictions were made for all of the sampled pixels (within aggregation units) using a set of 100 ensemble models each trained on different bootstrap samples of the 80% training data partition. The standard errors of the aggregate AGB estimates were estimated for each polygon across the 100 bootstrap iterations. Additionally, we computed the average difference between the original prediction and the 100 bootstrap predictions to assess our model’s sensitivity to the provided training data (McRoberts 2011). Details are given in Supplementary Materials S7.

3 Results

3.1 AGB by Landcover Class

Approximately 62% of all pixels, accounting for 87% of the total mapped AGB (Figure 2), were contained within areas identified as Tree cover by LCPRI (Table 3). All other LCPRI classes contained comparatively small, but non-zero estimates of AGB, with Wetlands and Grass/Shrub containing higher average predictions within small portions of the mapped area, and Cropland containing lower average predictions across a larger proportion of the mapped area (Table 3).

3.2 Area of Applicability

All mapped (vegetated) LCPRI classes contained AOA $\geq 97.64\%$ and the AOA breakdown across component LiDAR coverages (after LCPRI masking) was uniform ($\geq 90.4\%$) with NYC and Long Island the only two coverages with $\leq 95\%$ AOA (Tables 2 and 3). In total, 98.12% of the GPO-LiDAR region was considered inside the AOA after initial LCPRI masking.

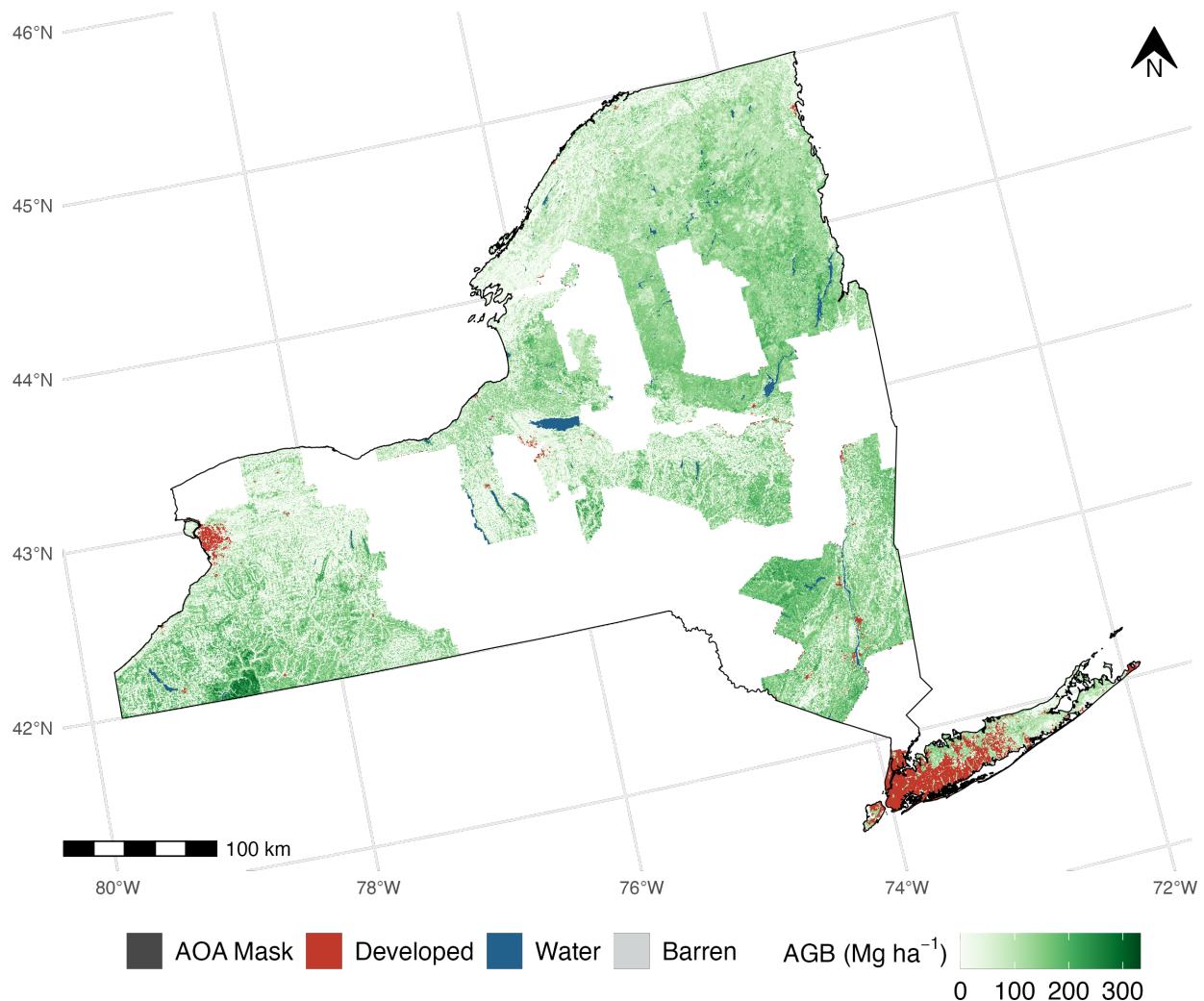


Figure 2: AGB prediction surfaces reflecting a temporal patchwork of conditions across the 17 component LiDAR coverages in the GPO-LiDAR region.

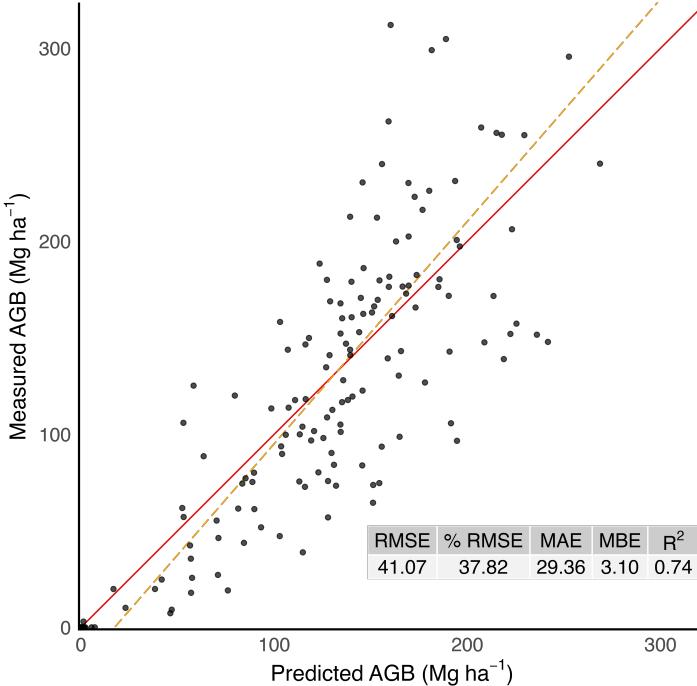


Figure 3: Measured vs predicted AGB scatter plot for the testing portion of the model dataset. AGB values in Mg ha^{-1} . GMFR trend line shown with dashed (orange) line, and 1:1 line shown with solid (red) line

3.3 Model Performance

The final model performance against the 20% test partition of the model dataset showed generally accurate predictions, but with greater accuracy towards the mean of the reference distribution (Figure 3). This resulted in slight overpredictions at lower AGB plots and underpredictions at higher AGB plots (Figure 3). A more detailed analysis of performance for the ensemble model and the three component models can be found in Supplementary Materials S3.

3.4 Map Agreement Assessment

We observed increasing agreement between the mapped estimates and the FIA estimates as the aggregation unit size increased with % RMSE decreasing from 33-13% and MAE decreasing from 27-10 Mg ha^{-1} (Table 4, Figures 4 and 5). Notably, MBE did not exhibit a clear pattern as a function of aggregation unit size. We again observed less accurate predictions at the extremes of the distributions across most scales, though our extreme predictions (low and high) became more accurate as a function of aggregation unit size with the geometric mean functional relationship (GMFR; Riemann et al. (2010)) slope approaching 1 at the largest scales (Figures 4 and 5).

The model's mapped residuals grouped to 50km spaced hexagons did not suggest any observable spatial patterns of bias (MBE) and error (RMSE, MAE) (Figure 6). The MBE map (Figure 6c) was dominated by near-zero hexagons and small pockets of similarly biased hexagons. One identifiable pattern was the tendency for negatively biased hexagons to have larger mean FIA AGB values, while positively biased hexagons were more likely to have smaller mean FIA AGB values (Figure 6g). The RMSE map (Figure 6a) and MAE map (Figure 6b) largely mirrored one another.

The comparison to the Menlove and Healey (2020) developed FIA estimates showed a high degree of agreement with 86% of our estimates falling inside the FIA estimate 95% confidence intervals (Figure 7). The majority of our 'misses' fell at the lower range of the FIA distribution, further reinforcing the observed pattern of overprediction at the low end of the reference distribution.

Table 4: Map agreement results for select scales. Distance = distance between hexagon centroids in km; PPH = plots per hexagon; n = number of comparison units (plots or hexagons); RMSE, MAE, MBE in Mg ha⁻¹. All accuracy metrics as defined in Section 2.6. Standard errors in parentheses. R² standard errors were excluded as they were all < 0.01.

Distance	n	PPH	% RMSE	RMSE	MAE	MBE	R ²
	733		33.45	37.67 (0.05)	26.80 (0.98)	-0.33 (1.39)	0.77
10	523	1.4	32.21	35.64 (0.06)	25.30 (1.10)	-0.71 (1.56)	0.79
25	192	3.82	25.21	27.69 (0.18)	18.33 (1.50)	-1.94 (2.00)	0.82
50	68	10.78	17.19	18.57 (0.24)	12.84 (1.64)	-1.63 (2.26)	0.90
100	27	27.15	15.65	15.75 (0.54)	11.08 (2.19)	3.30 (3.02)	0.87

The global Moran's I analysis showed only weak evidence of spatial autocorrelation (≤ 0.05) in our mapped residuals within search radii of roughly 5 to 15km (Supplementary Materials S6). There was no evidence of spatial autocorrelation at any other scale.

3.5 Model-Based Precision

The bootstrapping approach to model-based estimates of standard errors showed a high degree of aggregate prediction precision with mean and median standard errors of 0.37 Mg ha⁻¹ and 0.34 Mg ha⁻¹, respectively. As expected, the range of standard errors narrowed as a function of aggregation unit size (Supplementary Materials S7). However, across all aggregation unit sizes we observed that on average the models trained on bootstrap samples, which likely often excluded the upper extreme of the reference data distribution, underpredicted when compared to the original model trained on the full training dataset (Supplementary Materials S7).

4 Discussion

We set out to leverage a patchwork of 17 discrete LiDAR coverages for broad-scale, high-resolution biomass mapping in NYS. Faced with a limited supply of temporally aligned reference data, we leveraged repeated inventories to boost our sample, and used a machine learning ensemble model to produce accurate predictions. We addressed concerns of sensor and mission discrepancies among component LiDAR coverages by investigating spatial patterns of bias and error, and by using an AOA mask to both show predictor uniformity across coverages as well as to mask predictions based on anomalous data. Our results demonstrated that our predictions are suitable for small to medium area estimation with strong agreement with FIA estimates across a range of aggregation scales, and with small associated standard errors produced through a bootstrapping procedure.

4.1 Growth-Adjusted Field Plots

Our approach to solve the commonly experienced lack of temporally coincident LiDAR and field data was simple and effective in its ability to roughly triple our sample size while still achieving accurate modeling results. We did this by leveraging the existing inventory data, without additional field campaigns, RS data, or growth and yield models. However, one potentially limiting requirement for this approach is regular historical inventories so that bracketing inventory years can be identified for LiDAR acquisitions. Additionally, the temporal distance between growth-adjusted AGB values and the nearest measured AGB values was limited in this study to, at most, 4 years. Growth-adjustment following our approach may not be reliable across greater temporal lags.

4.2 LCMAP Masking

While most of our AGB was contained within Tree cover classified pixels, a notable minority was contained within pixels classified as Cropland and Grass/Shrub (Table 3). We expect that many of these predictions represented fields beginning to reforest (USDA National Agricultural Statistics Service 2019), or inclusions of

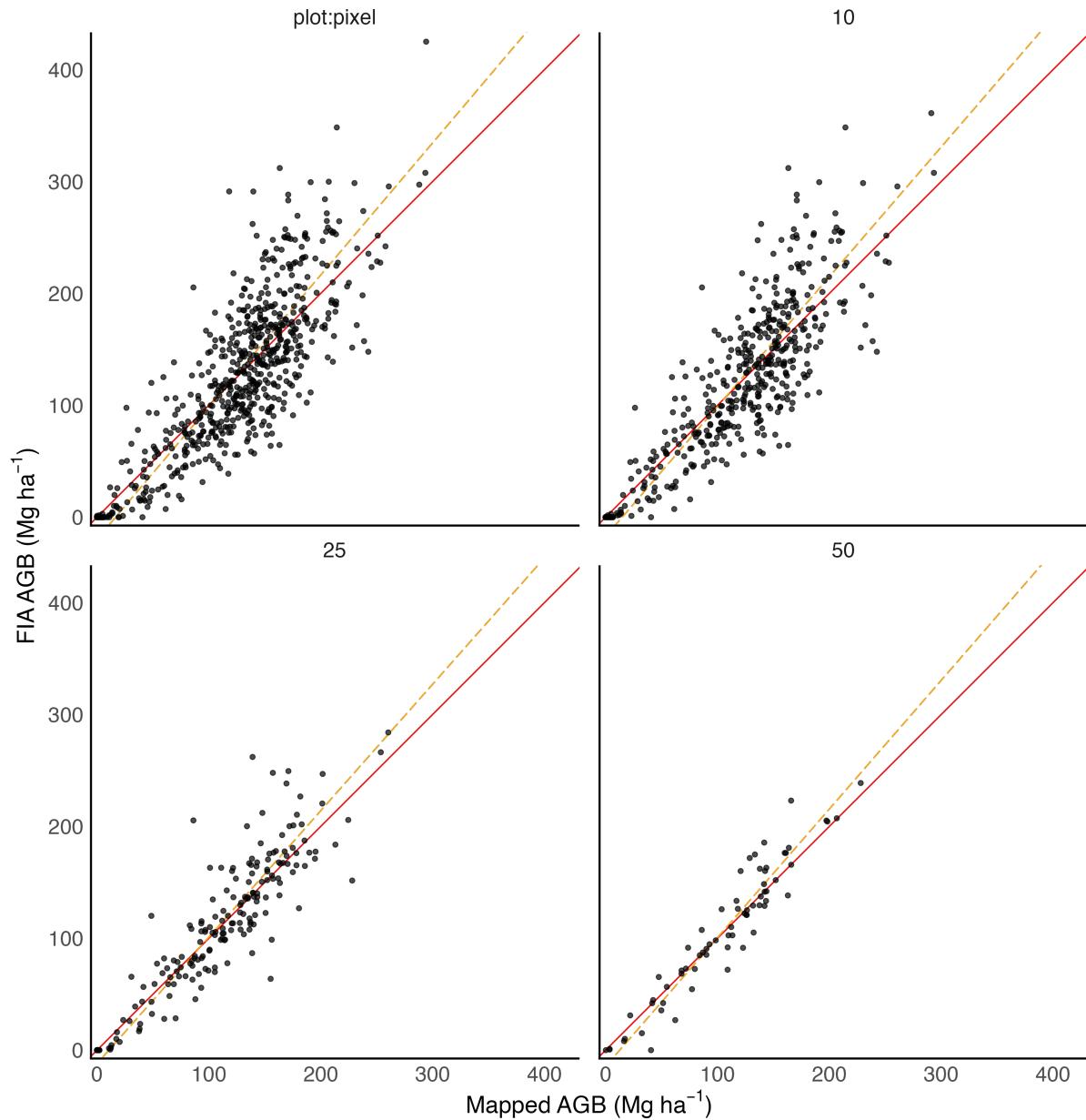


Figure 4: Comparing mapped AGB to FIA estimated AGB across selected scales represented by distances between hexagon centroids (plot:pixel, 10km, 25km, and 50km). AGB values in Mg ha^{-1} . GMFR trend line shown with dashed (orange) line, and 1:1 line shown with solid (red) line.

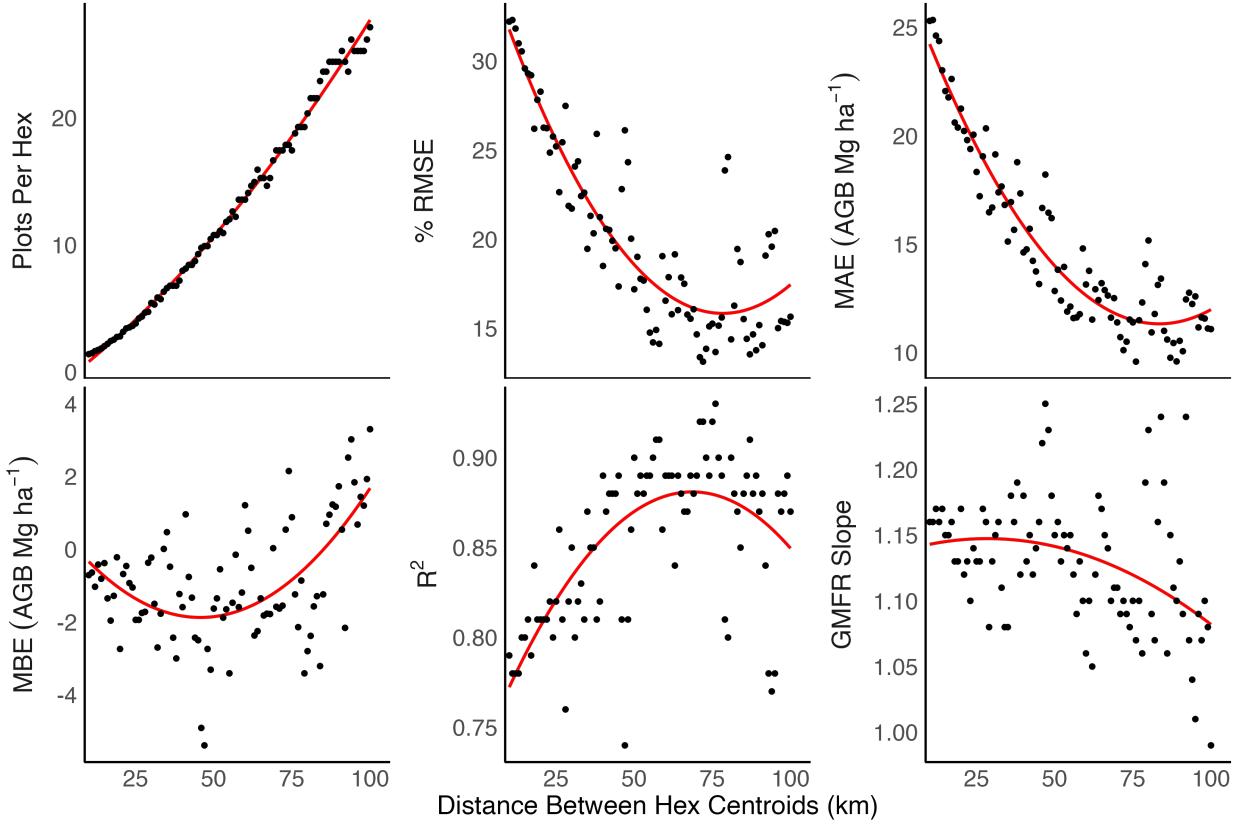


Figure 5: Summary assessment metrics (as defined in Section 2.6) comparing mapped predictions to FIA estimates as a function of aggregation unit size (described by distances between hexagon centroids). Red trend lines produced using quadratic regression.

young and stunted trees in field areas (Yang et al. 2018), both transitional states which are challenges for LCMAP’s underlying classification algorithm (Brown et al. 2020). The landscape-level context in NYS, and LCMAP’s algorithmic challenges gave us reason to produce predictions in all of LCPRI’s vegetated classes.

4.3 Area of Applicability

Our AOA surface provided evidence of predictor-space uniformity across all 17 component LiDAR coverages, indicating that each of the component coverages is well represented in the model training dataset (Meyer and Pebesma 2021). It is thus reasonable that the NYC coverage contained the lowest relative amount of AOA, since there were only two model plots available in this coverage (Table 2). Generally, pixels falling outside the AOA surface appeared to be the result of problems with LiDAR collection or data processing abnormalities, with some visible outliers that cannot be attributed to any known ecological phenomena. This data product is especially valuable in our case where we have taken publicly available LiDAR coverages off the shelf with limited knowledge of, or responsibility for, their provenance.

4.4 Model Performance and Map Agreement Assessment

Our model accuracy against the 20% testing partition of our model dataset was favorably comparable to previous LiDAR-AGB mapping studies (Huang et al. 2019; Nilsson et al. 2017; Ayrey et al. 2021; Hauglin et al. 2021). Using a set of FIA-developed methods (Riemann et al. 2010; Menlove and Healey 2020) we further demonstrated a strong agreement between our map-based estimates and FIA-derived estimates.

When summarized across the entire mapped region, the comparison of FIA estimates to our mapped predictions showed negligible bias (MBE), a high level of agreement (R^2) and a high level of accuracy (% RMSE, MAE)

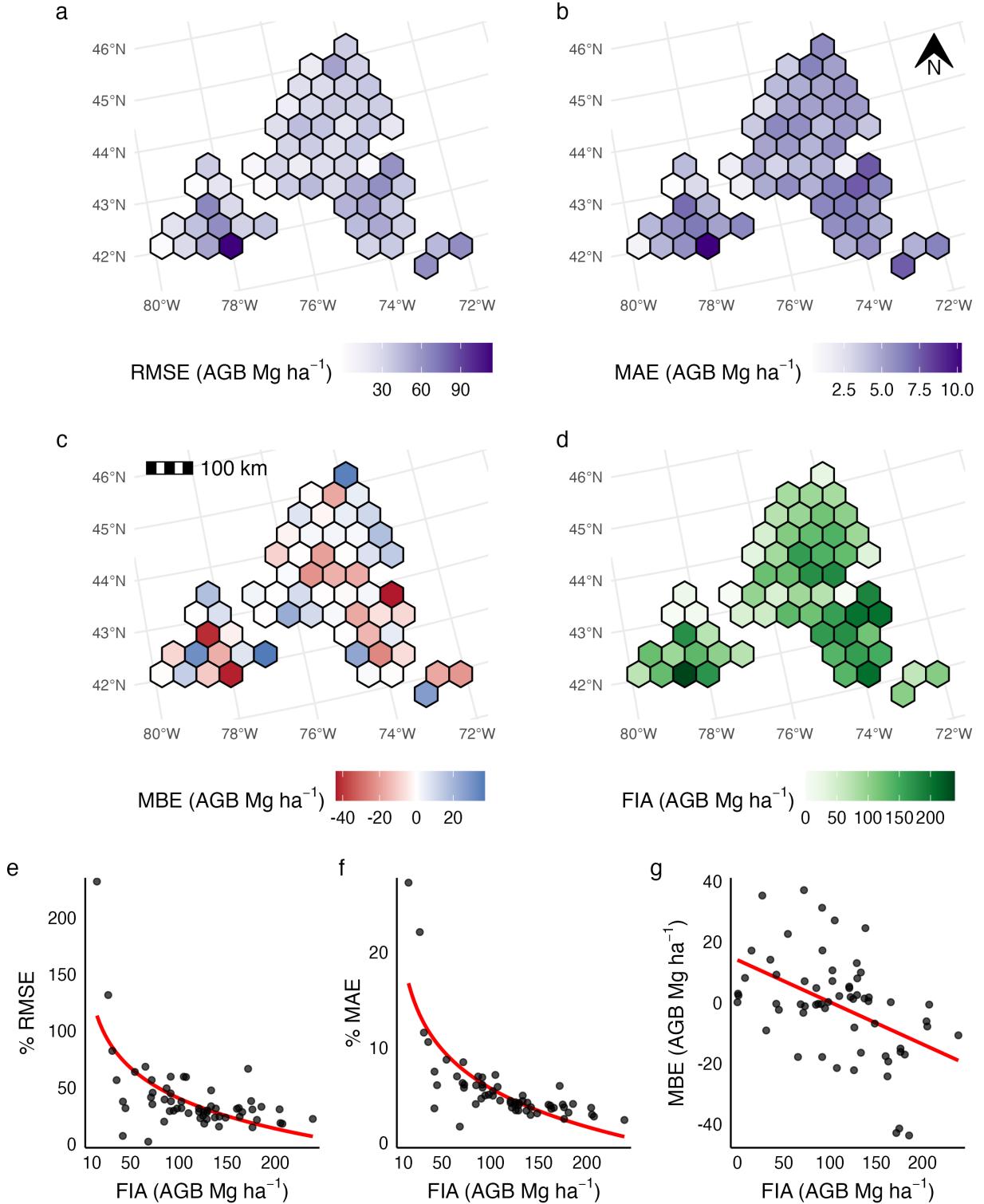


Figure 6: Plot-to-pixel residuals summarized at units spaced 50km apart. Hexagons with only one reference plot were removed. a) RMSE Mg ha^{-1} b) MAE Mg ha^{-1} c) MBE Mg ha^{-1} d) Mean FIA AGB value Mg ha^{-1} e) Hex-level % RMSE as a function of mean reference value f) Hex-level % MAE as a function of mean reference value. g) Hex-level MBE as a function of mean reference value within each hexagon. X-axis in e and f trimmed to 10 Mg ha^{-1} for visualization (Max Y values 939% and 313% respectively). Trend lines in e and f produced using logarithmic regression. Trend line in g produced using least-squares regression. RMSE, MAE, and MBE as defined in Section 2.6.

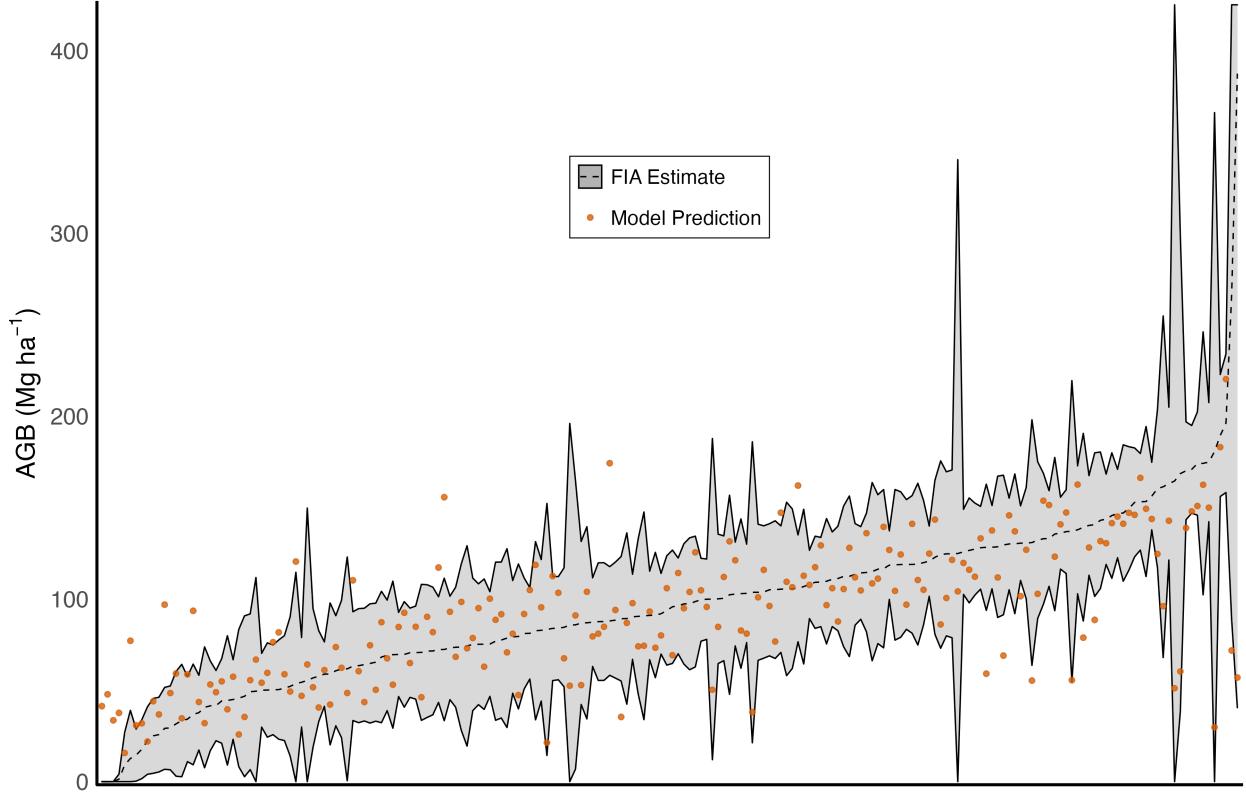


Figure 7: Comparison of mapped predicted AGB to Menlove and Healey (2020) estimates (dashed line) and associated 95% confidence interval (gray shaded region) within 64,000-ha aggregation hexagons. FIA estimates of AGB are scaled by the proportion of forest cover indicated by LCPRI 2016 Tree cover, Wetland, Croplands, and Grass/Shrub classified pixels. FIA estimates exceeding 425 Mg ha^{-1} (max plot level observation in our training data) were excluded from the analysis as they are artifacts of the forest cover scaling procedure. Observations are sorted by increasing FIA estimates along the x-axis.

across all scales of comparison (Figure 5). It is unsurprising that assessment metrics generally improved as the scales of aggregation increased, given that the plot-to-pixel scale can be considered the most rigorous, with the largest variance in both the pixel and plot AGB distributions, as well as the most potential for spatial misalignment between 30m pixel predictions and FIA plot measurements (McRoberts et al. 2018).

When mapped residuals were summarized within units spaced 50km apart, larger magnitudes of bias (MBE) and error (RMSE) emerged, but regional patterns were not evident (Figure 6). Rather, we observed MBE to be related to the underlying distribution of FIA reference data. We can also infer that large RMSE values were a reflection of extreme individual outliers as they were often paired with reasonable MAE values. The observed MBE pattern and relationship to mean FIA AGB within each hexagon reinforces the tendency towards diminished accuracy at the extremes of the reference distribution, as observed on the testing partition of the model data (Figure 3), in the Riemann analysis (Figure 4), and against the Menlove and Healey FIA estimates (Figure 7). These discrepancies can likely be attributed to both the model structure and training approach, as well as the saturation problem inherent in LiDAR-AGB modeling (St-Onge, Hu, and Vega 2008) where models fail to predict the highest AGB values in the data set. With the extremes of the response distribution occurring less often, the model is going to be accurate more often making predictions near the mean.

Following the bootstrapping approach outlined in McRoberts (2011), we were able to develop estimates of standard errors for aggregate map predictions showing that our aggregated model predictions are highly precise. However, we also observed that original predictions tended to exceed bootstrap predictions across all scales of aggregation, indicating the importance of a relatively small number of high-leverage observations with extreme AGB values. This suggests that strict FIA plot selection rules which might exclude extreme AGB

observations (e.g. requiring temporal alignment) could negatively impact model performance by restricting the range of model predictions.

We would be remiss to ignore the presence of uncertainty in the AGB reference data due to allometric, measurement, and locational errors, as well as our growth adjustment process, though quantifying the magnitude of this uncertainty was outside the scope of this paper. Previous studies have indicated that different choices in allometric models can result in large variation (up to 20%) of plot-level AGB estimates (Duncanson et al. 2017). An improvement upon our work would be to embed measurement errors and allometric model uncertainty in the bootstrap quantification of model precision demonstrated here (CEOS 2021). Additionally, our plot exclusion rules likely altered FIA's probability sample, removing the guarantee of unbiased estimates of map accuracy metrics produced through the Riemann analysis. Neither the results of the Menlove and Healey (2020) comparison, nor the bootstrap-based estimates of model precision, however, were similarly affected.

4.5 Map Applications

Our rigorously evaluated map products have a wide range of future uses, including AGB estimation across a wide range of scales, the identification of forested areas for future monitoring or protection, and for providing AGB as a predictor in subsequent ecological models. Furthermore, our AGB surfaces might be leveraged as diverse training data for spaceborne RS driven models, serving as a baseline for an annual carbon monitoring system and filling the spatiotemporal gaps in the maps we produced here (Hudak et al. 2020; CEOS 2021).

5 Conclusion

Accurate AGB estimation for small-areas can provide landowners and decision makers with the information necessary to manage forests for continued carbon storage and increased carbon sequestration. We implemented a model-based approach leveraging publicly available LiDAR coverages to fill this need in much of NYS. While LiDAR point clouds provide detailed information on forest structure at fine scales, their limited coverage in both spatial and temporal domains produces ‘patchworks’ of disparate datasets over broad-scales. Our modeling approach, and the comprehensive set of assessments demonstrated here, addressed several of the common challenges inherent in using LiDAR patchworks for AGB mapping, including a lack of temporally matching reference data and data discrepancies among component LiDAR coverages. Our results show that our approach and the resulting map products are suitable for AGB estimation at scales relevant to forest management in NYS.

6 Acknowledgements

We would like to thank the USDA FIA program for their data sharing and cooperation, the NYS GPO for compiling LiDAR data and sharing tax-parcel data, and the NYS Department of Environmental Conservation, Office of Climate Change, for funding support.

References

- Anderson, Ryan S., and Paul V. Bolstad. 2013. "Estimating Aboveground Biomass and Average Annual Wood Biomass Increment with Airborne Leaf-on and Leaf-off LiDAR in Great Lakes Forest Types." *Northern Journal of Applied Forestry* 30 (1): 16–22. <https://doi.org/10.5849/njaf.12-015>.
- Ayrey, Elias, Daniel J. Hayes, John B. Kilbride, Shawn Fraver, John A. Kershaw, Bruce D. Cook, and Aaron R. Weiskittel. 2021. "Synthesizing Disparate LiDAR and Satellite Datasets Through Deep Learning to Generate Wall-to-Wall Regional Inventories for the Complex, Mixed-Species Forests of the Eastern United States." *Remote Sensing* 13 (24). <https://doi.org/10.3390/rs13245113>.
- Bechtold, William A., and Paul L Patterson. 2005. *The Enhanced Forest Inventory and Analysis Program—National Sampling Design and Estimation Procedures*. Vol. 80. USDA Forest Service, Southern Research Station. <https://doi.org/10.2737/SRS-GTR-80>.
- Bivand, Roger S., Edzer Pebesma, and Virgilio Gomez-Rubio. 2013. *Applied Spatial Data Analysis with R, Second Edition*. Springer, NY. <https://doi.org/10.1007/978-0-387-78171-6>.
- Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32. <https://doi.org/10.1023/A:1010933404324>.
- Brown, Jesslyn F., Heather J. Tollerud, Christopher P. Barber, Qiang Zhou, John L. Dwyer, James E. Vogelmann, Thomas R. Loveland, et al. 2020. "Lessons learned implementing an operational continuous United States national land change monitoring capability: The Land Change Monitoring, Assessment, and Projection (LCMAP) approach." *Remote Sensing of Environment* 238: 111356. <https://doi.org/10.1016/j.rse.2019.111356>.
- CEOS. 2021. "Aboveground Woody Biomass Product Validation Good Practices Protocol." <https://doi.org/10.5067/DOC/CEOSWGCV/LPV/AGB.001>.
- Chen, Qi, and Ronald McRoberts. 2016. "Statewide Mapping and Estimation of Vegetation Aboveground Biomass Using Airborne Lidar." In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE. <https://doi.org/10.1109/igarss.2016.7730157>.
- Cortes, Corinna, and Vladimir Vapnik. 1995. "Support-Vector Networks." *Machine Learning* 20 (3): 273–97. <https://doi.org/10.1007/bf00994018>.
- Daniel Baston. 2021. *Exactextractr: Fast Extraction from Raster Datasets Using Polygons*. <https://CRAN.R-project.org/package=exactextractr>.
- Duncanson, Laura, Wenli Huang, Kristofer Johnson, Anu Swatantran, Ronald E McRoberts, and Ralph Dubayah. 2017. "Implications of Allometric Model Selection for County-Level Biomass Mapping." *Carbon Balance and Management* 12 (1): 1–11. <https://doi.org/10.1186/s13021-017-0086-9>.
- Fassnacht, F. E., F. Hartig, H. Latifi, C. Berger, J. Hernández, P. Corvalán, and B. Koch. 2014. "Importance of Sample Size, Data Type and Prediction Method for Remote Sensing-Based Estimations of Aboveground Forest Biomass." *Remote Sensing of Environment* 154: 102–14. <https://doi.org/10.1016/j.rse.2014.07.028>.
- Friedman, Jerome H. 2002. "Stochastic Gradient Boosting." *Computational Statistics and Data Analysis* 38 (4): 367–78. [https://doi.org/10.1016/S0167-9473\(01\)00065-2](https://doi.org/10.1016/S0167-9473(01)00065-2).
- Gobakken, Terje, and Erik Næsset. 2008. "Assessing Effects of Laser Point Density, Ground Sampling Intensity, and Field Sample Plot Size on Biophysical Stand Properties Derived from Airborne Laser Scanner Data." *Canadian Journal of Forest Research* 38 (5): 1095–1109. <https://doi.org/10.1139/X07-219>.
- Gonçalves, Fabio, Robert Treuhaft, Beverly Law, André Almeida, Wayne Walker, Alessandro Baccini, João dos Santos, and Paulo Graça. 2017. "Estimating Aboveground Biomass in Tropical Forests: Field Methods and Error Analysis for the Calibration of Remote Sensing Observations." *Remote Sensing* 9 (1): 47. <https://doi.org/10.3390/rs9010047>.
- Gray, Andrew N, Thomas J Brandeis, John D Shaw, William H McWilliams, and Patrick Miles. 2012. "Forest Inventory and Analysis Database of the United States of America (FIA)." *Biodiversity and Ecology* 4: 225–31. <https://doi.org/10.7809/b-e.00079>.
- Hauglin, Marius, Johannes Rahlf, Johannes Schumacher, Rasmus Astrup, and Johannes Breidenbach. 2021. "Large Scale Mapping of Forest Attributes Using Heterogeneous Sets of Airborne Laser Scanning and National Forest Inventory Data." *Forest Ecosystems* 8 (1). <https://doi.org/10.1186/s40663-021-00338-4>.
- Hawbaker, Todd J., Terje Gobakken, Adrian Lesak, Eric Trømborg, Kirk Conrucci, and Volker Radeloff. 2010. "Light Detection and Ranging-Based Measures of Mixed Hardwood Forest Structure." *Forest Science* 56 (3): 313–26. <https://doi.org/10.1093/forestscience/56.3.313>.
- Hijmans, Robert J. 2021. *Raster: Geographic Data Analysis and Modeling*. <https://CRAN.R-project.org/package=raster>.
- . 2022. *Terra: Spatial Data Analysis*. <https://CRAN.R-project.org/package=terra>.

- Huang, Wenli, Katelyn Dolan, Anu Swatantran, Kristofer Johnson, Hao Tang, Jarlath O'Neil-Dunne, Ralph Dubayah, and George Hurtt. 2019. "High-resolution mapping of aboveground biomass for forest carbon monitoring system in the Tri-State region of Maryland, Pennsylvania and Delaware, USA." *Environmental Research Letters* 14 (9): 095002. <https://doi.org/10.1088/1748-9326/ab2917>.
- Hudak, Andrew T, Patrick A Fekety, Van R Kane, Robert E Kennedy, Steven K Filippelli, Michael J Falkowski, Wade T Tinkham, et al. 2020. "A Carbon Monitoring System for Mapping Regional, Annual Aboveground Biomass Across the Northwestern USA." *Environmental Research Letters* 15 (9): 095003. <https://doi.org/10.1088/1748-9326/ab93f9>.
- Hurtt, G, M Zhao, R Sahajpal, A Armstrong, R Birdsey, E Campbell, K Dolan, et al. 2019. "Beyond MRV: high-resolution forest carbon modeling for climate mitigation planning over Maryland, USA." *Environmental Research Letters* 14 (4): 045013. <https://doi.org/10.1088/1748-9326/ab0bbe>.
- Jean-Romain, Roussel, David Auty, Nicholas C. Coops, Piotr Tompalski, Tristan R. H. Goodbody, Andrew Sánchez Meador, Jean-François Bourdon, Florian de Boissieu, and Alexis Achim. 2020. "lidR: An R package for analysis of Airborne Laser Scanning (ALS) data." *Remote Sensing of Environment* 251: 112061. <https://doi.org/10.1016/j.rse.2020.112061>.
- Karatzoglou, Alexandros, Alexandros Smola, Kurt Hornik, and Achim Zeileis. 2004. "kernlab - An S4 Package for Kernel Methods in R." *Journal of Statistical Software, Articles* 11 (9): 1–20. <https://doi.org/10.18637/jss.v011.i09>.
- Ke, Guolin, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2017. "LightGBM: A Highly Efficient Gradient Boosting Decision Tree." In *Advances in Neural Information Processing Systems*, edited by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett. Vol. 30. Curran Associates, Inc. <https://proceedings.neurips.cc/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf>.
- Ke, Guolin, Damien Soukhavong, James Lamb, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2021. *Lightgbm: Light Gradient Boosting Machine*. <https://CRAN.R-project.org/package=lightgbm>.
- Kennedy, Robert E, Janet Ohmann, Matt Gregory, Heather Roberts, Zhiqiang Yang, David M Bell, Van Kane, et al. 2018. "An Empirical, Integrated Forest Biomass Monitoring System." *Environmental Research Letters* 13 (2): 025004. <https://doi.org/10.1088/1748-9326/aa9d9e>.
- L'Roe, Andrew W, and Shorna Broussard Allred. 2013. "Thriving or Surviving? Forester Responses to Private Forestland Parcelization in New York State." *Small-Scale Forestry* 12 (3): 353–76. <https://doi.org/10.1007/s11842-012-9216-0>.
- Lu, Dengsheng, Qi Chen, Guangxing Wang, Lijuan Liu, Guiying Li, and Emilio Moran. 2014. "A Survey of Remote Sensing-Based Aboveground Biomass Estimation Methods in Forest Ecosystems." *International Journal of Digital Earth* 9 (1): 63–105. <https://doi.org/10.1080/17538947.2014.990526>.
- McRoberts, Ronald E. 2011. "Satellite Image-Based Maps: Scientific Inference or Pretty Pictures?" *Remote Sensing of Environment* 115 (2): 715–24. <https://doi.org/10.1016/j.rse.2010.10.013>.
- McRoberts, Ronald E., Qi Chen, Brian F. Walters, and Daniel J. Kaisershot. 2018. "The Effects of Global Positioning System Receiver Accuracy on Airborne Laser Scanning-Assisted Estimates of Aboveground Biomass." *Remote Sensing of Environment* 207: 42–49. <https://doi.org/10.1016/j.rse.2017.09.036>.
- Menlove, James, and Sean P. Healey. 2020. "A Comprehensive Forest Biomass Dataset for the USA Allows Customized Validation of Remotely Sensed Biomass Estimates." *Remote Sensing* 12 (24): 4141. <https://doi.org/10.3390/rs12244141>.
- Meyer, Hanna, and Edzer Pebesma. 2021. "Predicting into unknown space? Estimating the area of applicability of spatial prediction models." *Methods in Ecology and Evolution* 12 (9): 1620–33. <https://doi.org/10.1111/2041-210x.13650>.
- Moran, Patrick AP. 1950. "Notes on Continuous Stochastic Phenomena." *Biometrika* 37 (1/2): 17–23. <https://doi.org/10.2307/2332142>.
- Nilsson, Mats, Karin Nordkvist, Jonas Jonzén, Nils Lindgren, Peder Axensten, Jörgen Wallerman, Mikael Egberth, et al. 2017. "A nationwide forest attribute map of Sweden predicted using airborne laser scanning data and field data from the National Forest Inventory." *Remote Sensing of Environment* 194: 447–54. <https://doi.org/https://doi.org/10.1016/j.rse.2016.10.022>.
- Pebesma, Edzer. 2018. "Simple Features for R: Standardized Support for Spatial Vector Data." *The R Journal* 10 (1): 439–46. <https://doi.org/10.32614/RJ-2018-009>.
- Pflugmacher, Dirk, Warren B. Cohen, Robert E. Kennedy, and Zhiqiang Yang. 2014. "Using Landsat-derived disturbance and recovery history and lidar to map forest biomass dynamics." *Remote Sensing of Environment* 151: 124–37. <https://doi.org/10.1016/j.rse.2013.05.033>.
- R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.

- Riemann, Rachel, Barry Tyler Wilson, Andrew Lister, and Sarah Parks. 2010. "An effective assessment protocol for continuous geospatial datasets of forest characteristics using USFS Forest Inventory and Analysis (FIA) data." *Remote Sensing of Environment* 114 (10): 2337–52. <https://doi.org/10.1016/j.rse.2010.05.010>.
- Roussel, Jean-Romain, and David Auty. 2020. *Airborne LiDAR Data Manipulation and Visualization for Forestry Applications*. <https://cran.r-project.org/package=lidR>.
- Skowronski, Nicholas S, and Andrew J Lister. 2012. "Utility of LiDAR for large area forest inventory applications." In *In: Morin, Randall s.; Liknes, Greg c., Comps. Moving from Status to Trends: Forest Inventory and Analysis (FIA) Symposium 2012; 2012 December 4-6; Baltimore, MD. Gen. Tech. Rep. NRS-p-105. Newtown Square, PA: US Department of Agriculture, Forest Service, Northern Research Station.[CD-ROM]: 410-413.*, 410–13. <https://www.fs.usda.gov/treesearch/pubs/42792>.
- St-Onge, B., Y. Hu, and C. Vega. 2008. "Mapping the height and above-ground biomass of a mixed forest using lidar and stereo Ikonos images." *International Journal of Remote Sensing* 29 (5): 1277–94. <https://doi.org/10.1080/01431160701736505>.
- Thompson, Jonathan R., David R. Foster, Robert Scheller, and David Kittredge. 2011. "The influence of land use and climate change on forest biomass and composition in Massachusetts, USA." *Ecological Applications* 21 (7): 2425–44. <https://doi.org/10.1890/10-2383.1>.
- USDA Forest Service. 2020. "Forests of New York, 2019." U.S. Department of Agriculture, Forest Service, Northern Research Station. <https://doi.org/10.2737/fs-ru-250>.
- USDA National Agricultural Statistics Service. 2019. "2017 Census of Agriculture." <https://www.nass.usda.gov/Publications/AgCensus/2017/index.php>.
- White, Joanne C., John T. T. R. Arnett, Michael A. Wulder, Piotr Tompalski, and Nicholas C. Coops. 2015. "Evaluating the Impact of Leaf-on and Leaf-Off Airborne Laser Scanning Data on the Estimation of Forest Inventory Attributes with the Area-Based Approach." *Canadian Journal of Forest Research* 45 (11): 1498–1513. <https://doi.org/10.1139/cjfr-2015-0192>.
- White, Joanne C., Michael A. Wulder, Andrés Varhola, Mikko Vastaranta, Nicholas C. Coops, Bruce D. Cook, Doug Pitt, and Murray Woods. 2013. "A Best Practices Guide for Generating Forest Inventory Attributes from Airborne Laser Scanning Data Using an Area-Based Approach." *The Forestry Chronicle* 89 (06): 722–23. <https://doi.org/10.5558/tfc2013-132>.
- Wolpert, David H. 1992. "Stacked Generalization." *Neural Networks* 5 (2): 241–59. [https://doi.org/10.1016/S0893-6080\(05\)80023-1](https://doi.org/10.1016/S0893-6080(05)80023-1).
- Wright, Marvin N., and Andreas Ziegler. 2017. "Ranger: A Fast Implementation of Random Forests for High Dimensional Data in c++ and r." *Journal of Statistical Software, Articles* 77 (1): 1–17. <https://doi.org/10.18637/jss.v077.i01>.
- Yang, Limin, Suming Jin, Patrick Danielson, Collin Homer, Leila Gass, Stacie M. Bender, Adam Case, et al. 2018. "A new generation of the United States National Land Cover Database: Requirements, research priorities, design, and implementation strategies." *ISPRS Journal of Photogrammetry and Remote Sensing* 146 (December): 108–23. <https://doi.org/10.1016/j.isprsjprs.2018.09.006>.
- Zhu, Zhe, and Curtis E. Woodcock. 2014. "Continuous change detection and classification of land cover using all available Landsat data." *Remote Sensing of Environment* 144: 152–71. <https://doi.org/10.1016/j.rse.2014.01.011>.

Supplementary Materials

Contents

S1: Modeling Predictors	2
S2: Tax Parcel Data	3
S3: Model Training and Validation	6
S4: Area of Applicability	8
S5: Map Agreement Assessment	9
S6: Spatial Autocorrelation of Mapped Residuals	11
S7: Model Based Precision	12
References	14

S1: Modeling Predictors

Table 1: Definitions of predictors used for model fitting.

Predictor	Definition	Group
H0, H10, ... H100, H95, H99	Decile heights of returns, in meters, as well as 95th and 99th percentile return heights.	LiDAR
D10, D20... D90	Density of returns above a certain height, as a proportion. After return height is divided into 10 equal bins ranging from 0 to the maximum height of returns, this value reflects the proportion of returns at or above each breakpoint.	LiDAR
ZMEAN, ZMEAN_C	Mean height of all returns (ZMEAN) and all returns above 2.5m (ZMEAN_C)	LiDAR
Z_KURT, Z_SKEW	Kurtosis and skewness of height of all returns	LiDAR
QUAD_MEAN, QUAD_MEAN_C	Quadratic mean height of all returns (QUAD_MEAN) and all returns above 2.5m (QUAD_MEAN_C)	LiDAR
CV, CV_C	Coefficient of variation for heights of all returns (CV) and all returns above 2.5m (CV_C)	LiDAR
L2, L3, L4, L_CV, L_SKEW, L_KURT	L-moments and their ratios as defined by Hosking (1990), calculated for heights of all returns	LiDAR
CANCOV	Ratio of returns above 2.5m to all returns (Pflugmacher et al. 2012)	LiDAR
HVOL	CANCOV * ZMEAN (Pflugmacher et al. 2012)	LiDAR
RPC1	Ratio of first returns to all returns (Pflugmacher et al. 2012)	LiDAR
TMIN, TMAX, PRECIP	Climate variables in the form of 30 year normals (1980-2010) for minimum and maximum annual temperature (C) and average annual precipitation (mm) from Daly (2008)	Climate
ELEV, SLOPE, ASPECT, TWI	Topographic variables computed from a 30m DEM downloaded using the terrainr package (Mahoney 2021)	Topography
TAX_CODE_*	Individual tax code classifications as defined by NYS Department of Taxation and Finances (2019)	Tax
TAX_CATEGORY_*	Broad grouping of tax codes as defined by NYS Department of Taxation and Finances (2019)	Tax

S2: Tax Parcel Data

A shapefile containing tax parcels and associated tax parcel codes was obtained under agreement with the NYS GPO. This shapefile was rasterized using to grid geometry matching all other predictor variables using the “fasterize” R package (Ross 2020). No-data pixels were filled with the default value of 1000, and special NYC tax parcel codes were remapped to the default value of 2000. Codes (Table 2) were mapped to categories (Table 3) representing more broad groupings of similar use and ownership types. Codes representing < 1 % of plots and categories representing < 5% of plots in the calibration dataset were ignored. Codes or categories with less representation than these defined thresholds cause issues with model fitting for the same reason as any near-0 variance variable. Further descriptions of these variables can be found on the NYS department of taxation website: <https://www.tax.ny.gov/research/property/assess/manuals/prclas.htm>.

Table 2: Tax Code Indicator Variables

Variable.Name	Category	Description
TAX_CODE_105	Agricultural	Agricultural vacant land (productive)
TAX_CODE_112	Agricultural	Dairy products: milk, butter and cheese
TAX_CODE_120	Agricultural	Field crops
TAX_CODE_210	Residential	One family year-round residence
TAX_CODE_240	Residential	Rural residence with acreage
TAX_CODE_241	Residential	Primary residential, also used in agricultural production
TAX_CODE_260	Residential	Seasonal residences
TAX_CODE_270	Residential	Mobile home
TAX_CODE_280	Residential	Residential
TAX_CODE_312	Vacant land	Residential land including a small improvement (not used for living accommodations)
TAX_CODE_314	Vacant land	Rural vacant lots of 10 acres or less
TAX_CODE_322	Vacant land	Residential vacant land over 10 acres
TAX_CODE_323	Vacant land	Other rural vacant lands
TAX_CODE_910	Wild, forested, conservation lands and public parks	Private wild and forest lands except for private hunting and fishing clubs
TAX_CODE_911	Wild, forested, conservation lands and public parks	Forest land under Section 480 of the Real Property Tax Law
TAX_CODE_912	Wild, forested, conservation lands and public parks	Forest land under Section 480-a of the Real Property Tax Law
TAX_CODE_930	Wild, forested, conservation lands and public parks	State-owned forest lands
TAX_CODE_931	Wild, forested, conservation lands and public parks	State-owned land (forest Preserve) in the Adirondack or Catskill Parks taxable under Section 532-a of the Real Property Tax Law
TAX_CODE_932	Wild, forested, conservation lands and public parks	State-owned land other than forest preserve covered under Section 532-b, c, d, e, f, or g of the Real Property Tax Law
TAX_CODE_941	Wild, forested, conservation lands and public parks	State-owned reforested land taxable under Sections 534 and 536 of the Real Property Tax Law
TAX_CODE_1000	Missing data	No tax code present - this value represents missing data.
TAX_CODE_2000	NYC	Special tax codes for NYC were remapped to this default value.

Table 3: Tax Category Indicator Variables

Variable.Name	Category	Description
TAX_CATEGEORY_100	Agricultural	Property used for the production of crops or livestock.
TAX_CATEGEORY_200	Residential	Property used for human habitation. Living accommodations such as hotels, motels, and apartments are in the Commercial category - 400.
TAX_CATEGEORY_300	Vacant land	Property that is not in use, is in temporary use, or lacks permanent improvement.
TAX_CATEGEORY_900	Wild, forested, conservation lands and public parks	Reforested lands, preserves, and private hunting and fishing clubs.

S3: Model Training and Validation

We tuned each component model (RF, GBM, and SVM) using the 80% training dataset and an iterative grid search approach, starting by testing wide ranges of hyperparameters using five-fold cross validation and then narrowing down to only the most performant combinations over several iterations. Models then used the most accurate sets of hyperparameters in all other analyses. For each of the n observation in the training dataset, all three models were fit, using their optimal hyperparameters, with n-1 observations. Predictions for each component model were made for the nth (left out) observation. A linear regression model was used to estimate AGB as a function of these leave-one-out (LOO) predictions, combining the three machine learning approaches in a “stacked ensemble” (LINMOD).

Each of the three component models and the LINMOD model were assessed against the 20% holdout partition (Table 4 and Figure 1).

Table 4: Model performance metrics (as defined in section S3) against 20% holdout partition ($n = 171$; RMSE, MBE in Mg ha^{-1}).

	RF	GBM	SVM	LINMOD
RMSE	40.55	41.02	41.63	41.07
% RMSE	37.34	37.78	38.34	37.82
MAE	29.08	29.68	29.27	29.36
MBE	3.14	1.05	-1.90	3.10
R^2	0.75	0.74	0.73	0.74

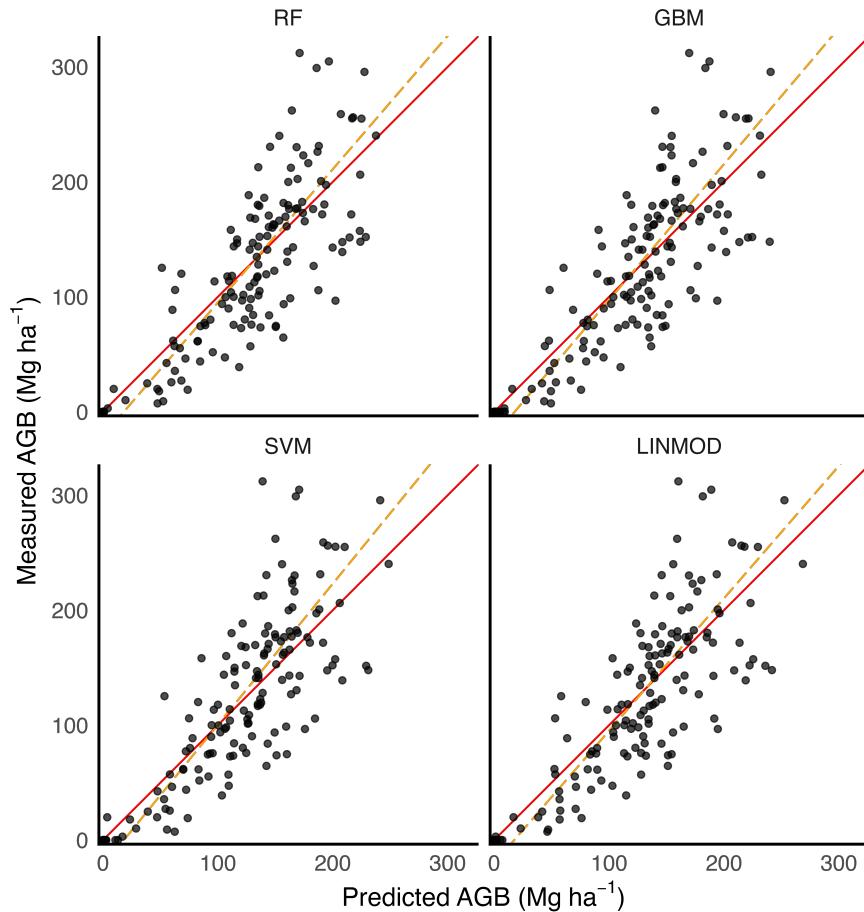


Figure 1: Measured vs predicted AGB scatter plot for four component models against the testing portion of the model dataset. AGB values in Mg ha^{-1} . GMFR trend line shown with dashed (orange) line, and 1:1 line shown with solid (red) line

S4: Area of Applicability

Following Meyer and Pebesma (2021) we computed an area of applicability (AOA) surface to identify individual pixels whose predictor data is beyond a threshold predictor-space distance from the training data. The predictor variables are scaled by dividing mean-centered values by their respective standard deviations and then weighted by their importance to the LINMOD model. The distance for a given observation or pixel is computed as the minimum euclidean distance (in predictor space) from said observation to the training data observations. Using this distance a dissimilarity index (DI) distance is then standardized, by dividing by the average distances in the training data. A threshold to identify “in” observations and “out” observations for the AOA is defined as the 75th percentile plus 1.5 times the IQR of the DI values in the 20% testing partition of the calibration data. Observations with computed DIs beyond this threshold are identified as outside the AOA.

Predictor Weights

Variable importance measures are required to weight individual predictor contributions to the DI distance by their relative importance to LINMOD’s predictions. Without predictor weights, a single noisy predictor that has little impact on the model’s prediction might cause a pixel or observation to be deemed “out” of the AOA. Variable importance was computed for the LINMOD ensemble using a permutation approach (Breiman 2001). Since many of our predictors are highly correlated, we first clustered them based on spearman rank correlation euclidean distances. A cluster dendrogram was used to identify seven distinct variable clusters. For each p predictor, a separate LINMOD ensemble is trained. The p th predictor is selected, and all other predictors in the p th predictor’s cluster are excluded. A single predictor from the remaining 6 clusters are selected. The values for the p th predictor are then shuffled, breaking any potential relationship between the p th predictor and the response (AGB). The model is then trained with these seven predictors (following the approach in Section S3), and RMSE is assessed via LOO-cross-validation on the 80% training partition. Variable importance is then computed as the delta RMSE from the p th predictor model against the full LINMOD model trained with all p predictors. The variable importances are then standardized to produce weights, such that the sum of all p weights is equal to 1. In this analysis default hyperparameter values are used for all component ML models.

Table 5: Map agreement results for select scales (in ha) and all models. Dist = distance between hex centroids in km; PPH = plots per hex; n = number of comparison units (plots or hexagons); RMSE, MAE, MBE in Mg ha⁻¹. All metrics as defined in Section S6. Standard errors computed via bootstrap resampling in parentheses. R² standard errors excluded here as they were all < 0.01.

Scale	Dist	n	PPH	Model	% RMSE	RMSE	MAE	MBE	R ²
Plot Pixel		733		RF	33.66	37.91 (0.05)	27.23 (0.97)	-0.21 (1.40)	0.77
				GBM	33.61	37.85 (0.05)	27.09 (0.98)	-1.55 (1.40)	0.77
				SVM	34.66	39.03 (0.06)	26.79 (1.05)	-5.20 (1.43)	0.75
				LINMOD	33.45	37.67 (0.05)	26.80 (0.98)	-0.33 (1.39)	0.77
8660	10	523	1.4	RF	32.31	35.75 (0.06)	25.65 (1.09)	-0.53 (1.56)	0.78
				GBM	32.01	35.42 (0.06)	25.25 (1.09)	-1.91 (1.55)	0.79
				SVM	33.66	37.24 (0.07)	25.56 (1.19)	-5.69 (1.61)	0.77
				LINMOD	32.21	35.64 (0.06)	25.30 (1.10)	-0.71 (1.56)	0.79
54126	25	192	3.82	RF	24.64	27.07 (0.17)	18.39 (1.44)	-1.82 (1.95)	0.83
				GBM	25.40	27.90 (0.19)	18.58 (1.51)	-2.94 (2.01)	0.82
				SVM	27.47	30.17 (0.20)	19.60 (1.66)	-6.85 (2.13)	0.79
				LINMOD	25.21	27.69 (0.18)	18.33 (1.50)	-1.94 (2.00)	0.82
216506	50	68	10.78	RF	16.76	18.10 (0.22)	12.87 (1.56)	-1.39 (2.20)	0.91
				GBM	17.59	19.00 (0.25)	13.39 (1.65)	-2.47 (2.30)	0.90
				SVM	20.45	22.09 (0.30)	15.49 (1.92)	-6.95 (2.56)	0.86
				LINMOD	17.19	18.57 (0.24)	12.84 (1.64)	-1.63 (2.26)	0.90

S5: Map Agreement Assessment

Following the Riemann sampling intensity correction, all comparisons were developed by extracting the area-weighted mean of mapped pixels intersecting each FIA plot (Riemann et al. 2010). LCMAP or AOA masked pixels were excluded from the map agreement assessment such that any FIA plot intersecting a masked pixel was excluded from the analysis. The hexagons were tessellated randomly across the GPO-LiDAR region with no regard to FIA sample grid alignment. The geometric mean functional relationship (GMFR) slope and intercepts were computed following the Riemann approach (which follows Draper and Smith (1998) and Ricker (1984)).

Agreement metrics and scatter plots for all component models are contained in Table 5 and Figure 2 respectively.

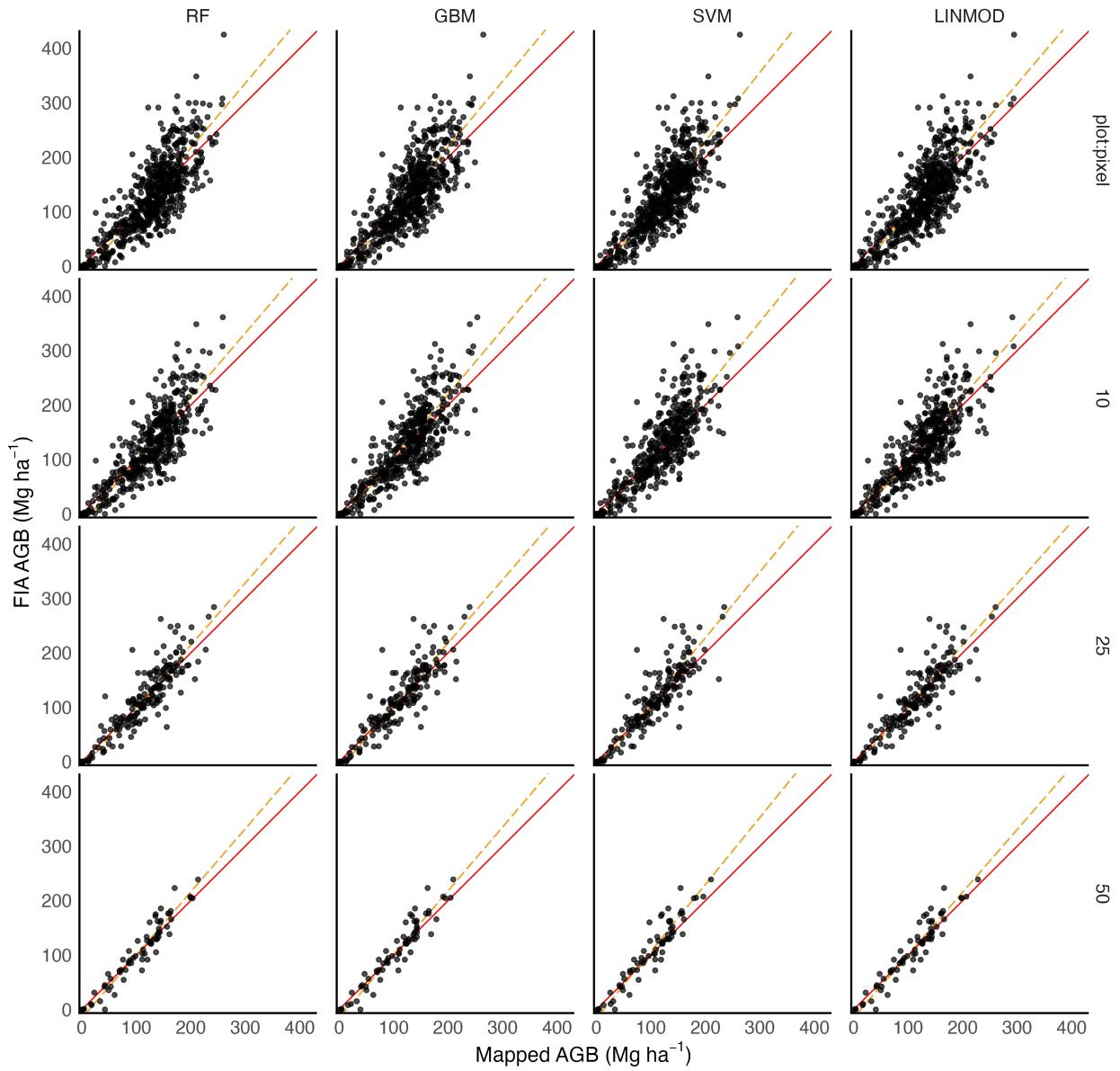


Figure 2: Comparing mapped AGB for all models to FIA estimated AGB across selected scales represented by distances between hexagon centroids (plot:pixel, 10km, 25km, and 50km). AGB values in Mg ha^{-1} . Geometric Mean Functional Regression trend line shown with dashed (orange) line, and 1:1 line shown with solid (red) line.

S6: Spatial Autocorrelation of Mapped Residuals

To provide context under conditions of spatial randomness for the computed Moran's I statistics, 1000 simulations were produced for each search radius. In each simulation, the spatial structure of the observations is maintained, but the associated residual information is shuffled randomly. The upper 97.5th percentile and the lower 2.5th percentile of simulated Moran's I statistics are selected to produce upper and lower envelopes for the true Moran's I estimates.

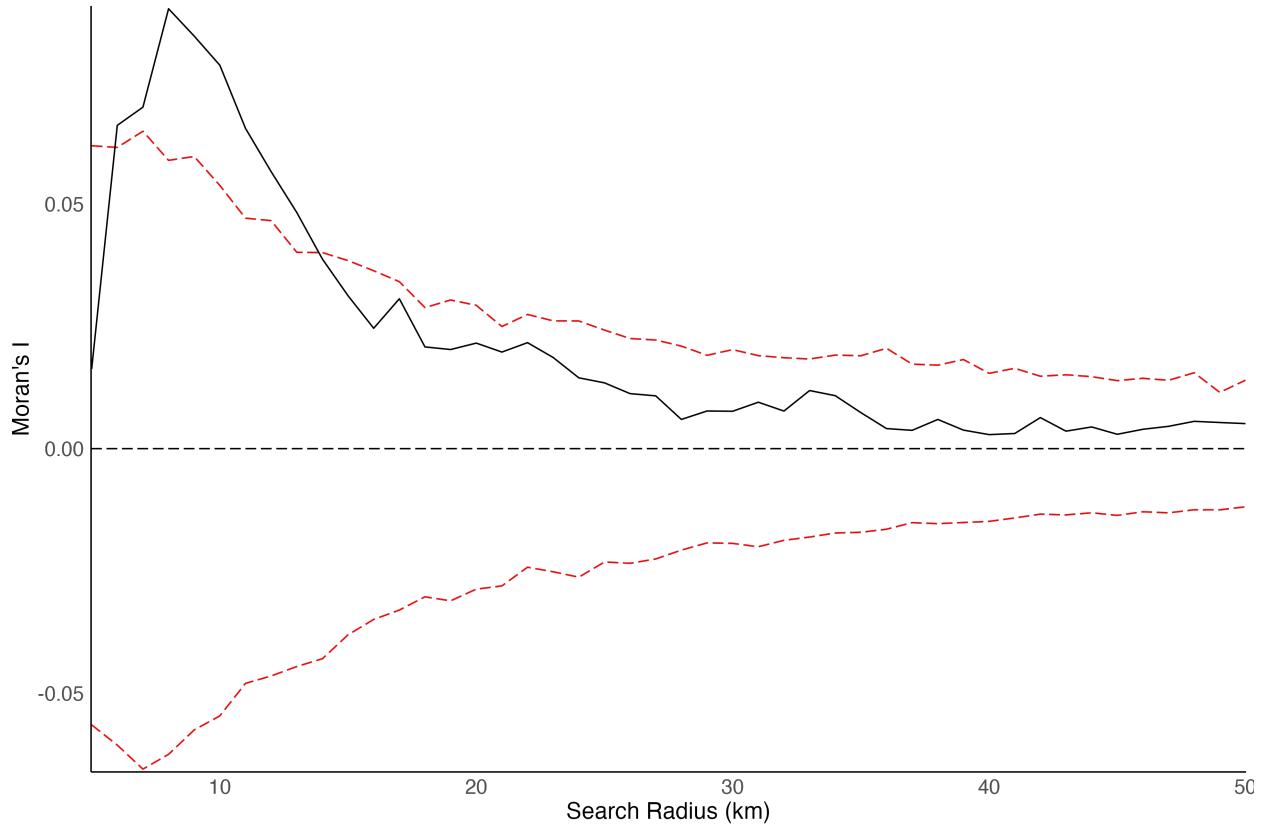


Figure 3: Global spatial autocorrelation (Moran's I) of LINMOD mapped residuals as a function of search radius. Red dashed upper and lower envelopes represent the 95% interval computed from 1000 bootstrap iterations of randomly re-assigning plot locations.

S7: Model Based Precision

Following McRoberts (2011), we produced bootstrap-based estimates of model uncertainty. 100 bootstrap samples were drawn from the 80% training partition of the calibration dataset, and a distinct LINMOD model was fit to each sample. We randomly sampled 1000 polygons from within the GPO-LiDAR region, with each polygon randomly sized between 1 and 500 hectares. Predictions were generated for each of the pixels contained in the 1000 polygons using each of the 100 bootstrap models. The area-weighted average of the pixel predictions was computed to produce an aggregate polygon prediction for each of the 100 bootstrap models. The variance and standard error was then computed for across all 100 aggregate predictions for each model. The relationship between standard error estimates and aggregation unit size are shown in Figure 4. Bootstrap bias was also computed as the average difference of the bootstrap aggregate predictions to the original (full sample) LINMOD predictions for each of the aggregation units. The relationship between bootstrap bias and aggregation unit size are shown in Figure 5.

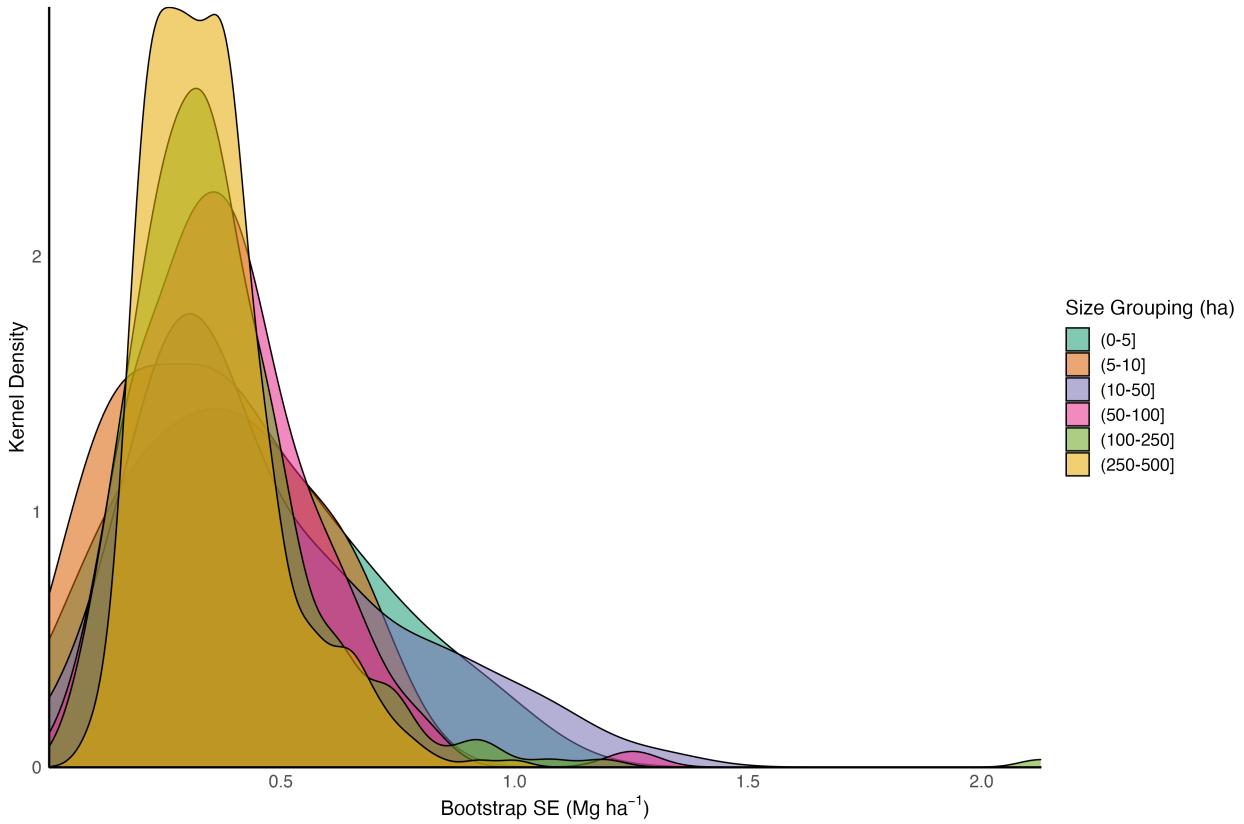


Figure 4: Smoothed kernel density estimates of model-based bootstrap standard errors for LINMOD aggregate predictions across six size groupings. Aggregate predictions and estimates of standard error were computed for 1000 randomly sampled polygons with sizes from 1-500 hectares.

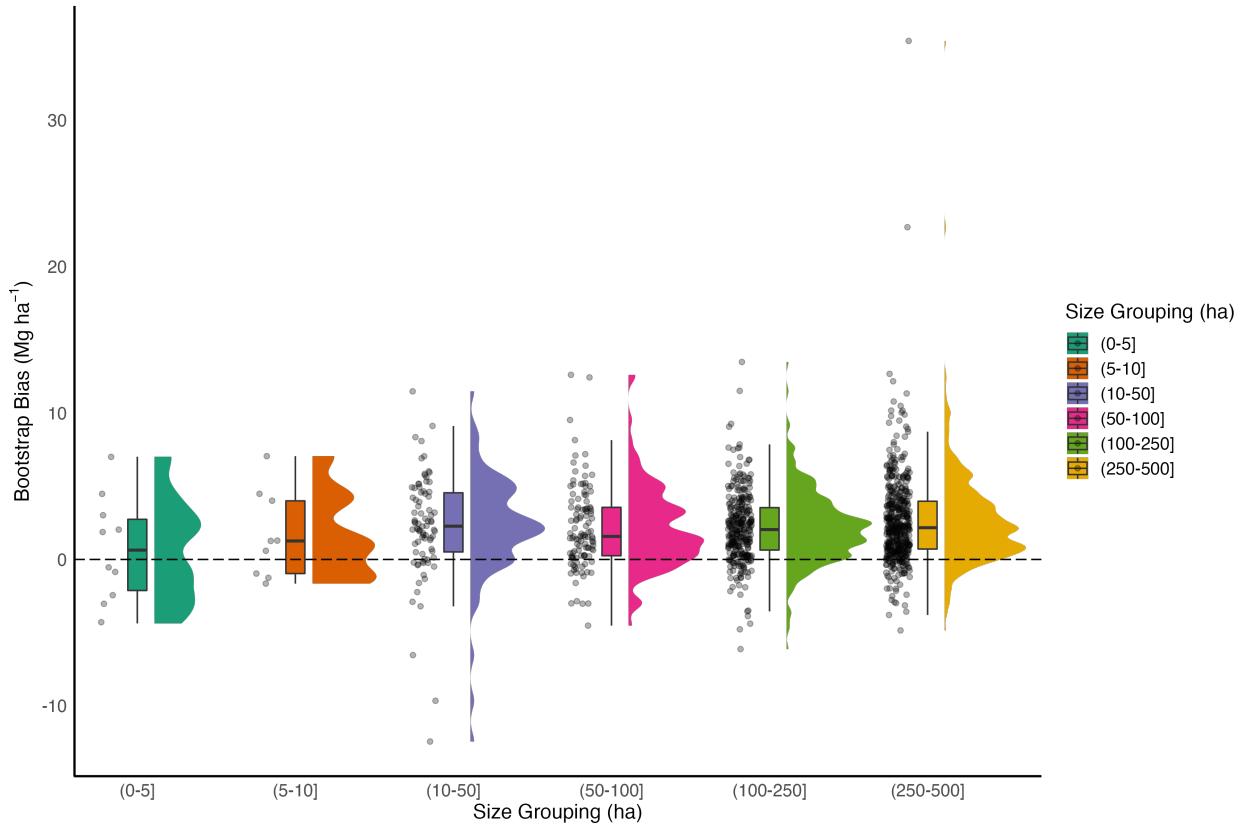


Figure 5: Bootstrap bias summaries comparing the original model aggregate prediction to the average aggregate prediction across all bootstrap iterations for six size groupings. Aggregate predictions and estimates of bootstrap bias were computed for 1000 randomly sampled polygons with sizes from 1-500 hectares.

References

- Atlantic Inc. 2015. "NY_WarrenWashingtonEssex_Spring2015." <ftp://ftp.gis.ny.gov/elevation/LIDAR/>.
- Axis GeoSpatial, LLC. 2016a. "Axis GeoSpatial, LLC New York Tiled LiDAR." <ftp://ftp.gis.ny.gov/elevation/LIDAR/>.
- . 2016b. "New York Office of Information Technology Services Classified LiDAR Tiles." <ftp://ftp.gis.ny.gov/elevation/LIDAR/>.
- Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32. <https://doi.org/10.1023/A:1010933404324>.
- Draper, Norman R, and Harry Smith. 1998. *Applied Regression Analysis*. Vol. 326. John Wiley & Sons.
- Hosking, J. R. M. 1990. "L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics." *Journal of the Royal Statistical Society. Series B (Methodological)* 52 (1): 105–24. <https://doi.org/10.1111/j.2517-6161.1990.tb01775.x>.
- Mahoney, Michael. 2021. *terrainr: Landscape Visualizations in R and Unity*. <https://doi.org/https://doi.org/10.5281/zenodo.5142763>.
- McRoberts, Ronald E. 2011. "Satellite Image-Based Maps: Scientific Inference or Pretty Pictures?" *Remote Sensing of Environment* 115 (2): 715–24. <https://doi.org/10.1016/j.rse.2010.10.013>.
- Meyer, Hanna, and Edzer Pebesma. 2021. "Predicting into unknown space? Estimating the area of applicability of spatial prediction models." *Methods in Ecology and Evolution* 12 (9): 1620–33. <https://doi.org/10.1111/2041-210x.13650>.
- New York Office of Information Technology Services. 2016. "Allegany and Steuben Counties, New York Lidar; Overall Project Metadata." <ftp://ftp.gis.ny.gov/elevation/LIDAR/>.
- . 2017. "Southwest 17 - Spring, New York Lidar; Classified Point Cloud." <ftp://ftp.gis.ny.gov/elevation/LIDAR/>.
- . 2018a. "LIDAR Collection (Ql2) for Cayuga County and Most of Oswego County, New York Lidar; Classified Point Cloud." <ftp://ftp.gis.ny.gov/elevation/LIDAR/>.
- . 2018b. "Southwest 17-b - Fall, New York Lidar; Classified Point Cloud." <ftp://ftp.gis.ny.gov/elevation/LIDAR/>.
- . 2019. "LIDAR Collection (QL2) for Erie, Genesee, and Livingston Counties New York Lidar; Classified Point Cloud." <ftp://ftp.gis.ny.gov/elevation/LIDAR/>.
- NYSGPO. 2018. *Property Type Classification Codes*. New York State Department of Taxation; Finance. <https://www.tax.ny.gov/research/property/assess/manuals/prclas.htm>.
- Quantum Spatial. 2016. "Clinton-Essex-Lake Champlain New York 2015 LiDAR USGS Contract: G10pc00026 Task Order Number: G10pc00026 and G14pd000943 (Modification) NY_ClintonEssex_2015." <ftp://ftp.gis.ny.gov/elevation/LIDAR/>.
- Quantum Spatial, Inc. 2017a. "New York FEMA 2016 Ql2 LiDAR - Central Zone AOI; Classified Point Cloud." <ftp://ftp.gis.ny.gov/elevation/LIDAR/>.
- . 2017b. "New York FEMA 2016 Ql2 LiDAR - East Zone AOI; Classified Point Cloud." <ftp://ftp.gis.ny.gov/elevation/LIDAR/>.
- . 2018. "Lidar Collection (Ql2) of All or Part of Schuyler, Seneca, Steuben, Tompkins, Wayne and Yates Counties, NY Lidar; Classified Point Cloud." <ftp://ftp.gis.ny.gov/elevation/LIDAR/>.
- Ricker, WE. 1984. "Computation and Uses of Central Trend Lines." *Canadian Journal of Zoology* 62 (10): 1897–1905.
- Riemann, Rachel, Barry Tyler Wilson, Andrew Lister, and Sarah Parks. 2010. "An effective assessment protocol for continuous geospatial datasets of forest characteristics using USFS Forest Inventory and Analysis (FIA) data." *Remote Sensing of Environment* 114 (10): 2337–52. <https://doi.org/10.1016/j.rse.2010.05.010>.
- Ross, Noam. 2020. *fasterize: Fast Polygon to Raster Conversion*. <https://CRAN.R-project.org/package=fasterize>.
- United States Geological Survey. 2015a. "LAS." <ftp://ftp.gis.ny.gov/elevation/LIDAR/>.
- . 2015b. "LAS." <ftp://ftp.gis.ny.gov/elevation/LIDAR/>.
- . 2015c. "LAS Extents." <ftp://ftp.gis.ny.gov/elevation/LIDAR/>.
- Woolpert, Inc. 2014a. "USGS Long Island New York Sandy Lidar Classified LAS 1.2." <ftp://ftp.gis.ny.gov/elevation/LIDAR/>.
- . 2014b. "USGS New York CMGP Sandy Lidar." <ftp://ftp.gis.ny.gov/elevation/LIDAR/>.