Forest Ecology and Management 595 (2025) 123040

Contents lists available at ScienceDirect

Forest Ecology and Management

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

Mapping aboveground biomass in Oregon, Washington, Idaho and California with ICESat-2, GEDI and ancillary data

Mei-Kuei Lu [a,*], Sorin Popescu [a], Lonesome Malambo [b], Andrew Hudak [c], Andrew Lister [c]

[a] Department of Ecology and Conservation Biology, Texas A&M University, 534 John Kimbrough Blvd, College Station, TX 77843, USA
[b] Texas A&M Forest Service, College Station, TX 77845, USA
[c] USDA Forest Service, USA

ARTICLE INFO

Keywords:
Biomass mapping
Pacific Northwest
Google Earth Engine
ICESat-2
Random Forest
GEDI
Forest Carbon

ABSTRACT

Biomass mapping is important for quantifying forest carbon stocks to provide precious insights for climate science and to inform natural resource management, such as fuel treatment planning. In this study, we demonstrated how to utilize Ice, Cloud, and land Elevation Satellite-2 (ICESat-2) and Global Ecosystem Dynamics Investigation (GEDI) data along with other data sources to model and predict biomass for the forests in Oregon, Washington, Idaho and California in 2016 and 2020. The model incorporated the Landsat time series data, GEDI height data, Landscape Fire and Resource Management Planning Tools (LANDFIRE) vegetation data and ICESat-2 derived height data to produce wall-to-wall biomass maps. Results showed that the maps, which are accompanied by corresponding uncertainty maps, have high accuracy. Since two maps are made, there are two validation datasets used in this study. The 2016 and 2020 aboveground biomass density (AGBD) maps were validated with ground plots from the Forest Inventory and Analysis (FIA) program (2016: $R^2 = 0.769$, RMSE = 55.26 Mg/ha; 2020: $R^2 = 0.572$, RMSE = 84.96 Mg/ha). Results describe the reliability of the maps and suggest that the methods proposed in the study could be used to scale up the analysis from regional to national scale. Future studies can include using predictor data from the upcoming National Aeronautics and Space Administration (NASA) Indian Space Research Organization (ISRO) Synthetic Aperture Radar (NISAR) mission and new GEDI biomass data, both of which should improve model accuracy.

1. Introduction

Forest biomass estimation and mapping are crucial to facilitate our understanding of the role of forests in the global carbon budget (Lamping et al., 2025; Nandy et al., 2021), ecosystem health (Hall et al., 2011), climate mitigation (Simmavong et al., 2025), and forest management (Duncanson et al., 2019; Lu et al., 2019; Perng et al., 2018). Space-borne lidar data acquired from missions such as Ice, Cloud, and land Elevation Satellite-2 (ICESat-2) and Global Ecosystem Dynamics Investigation (GEDI) show potential for canopy height mapping (Malambo and Popescu, 2021) and biomass estimation (Narine et al., 2020, 2019; Nandy et al., 2021; Liu and Popescu, 2022) from regional (for example, the conterminous United States (CONUS)) to global scales. They provide vertical forest structural information, usually modelled with height percentile metrics calculated from 3D point cloud data returned by the sensors (Popescu et al., 2002).

Having accurate aboveground forest biomass maps over large areas is central for not only a better understanding of the role of the forests in the carbon cycle, but also for monitoring forest ecosystem productivity and the impacts of conservation efforts (Hudak et al., 2009; Tian et al., 2012; Silva et al., 2016). However, it is often not cost-effective nor practical to use in situ data for biomass estimation at large scales, particularly if detailed information is needed over small areas (Hummel et al., 2011). Thus, remote sensing methods are often used for this purpose. The National Aeronautics and Space Administration's (NASA's) current GEDI and ICESat-2 missions and the upcoming NASA Indian Space Research Organization (ISRO) Synthetic Aperture Radar (NISAR) mission will collect synergistic data with different coverage and sensitivity to aboveground biomass (AGB) on global scales (Silva et al., 2021).

ICESat-2, launched in September 2018, provides global three-dimensional data to detect the changes in sea ice, ice sheets, and

* Corresponding author.
E-mail addresses: mklu@tamu.edu (M.-K. Lu), Sorin.Popescu@ag.tamu.edu (S. Popescu), lonesome.malambo@tfs.tamu.edu (L. Malambo), andrew.hudak@usda.gov (A. Hudak), andrew.lister@usda.gov (A. Lister).

https://doi.org/10.1016/j.foreco.2025.123040
Received 24 March 2025; Received in revised form 21 July 2025; Accepted 22 July 2025

forest structures. ICESat-2 is a photon counting lidar (PCL) mission onboard the Advanced Topographic Laser Altimeter System (ATLAS). It provides different height metrics for estimating vegetation height and characterizing the land surface, and its datasets have been used for many applications such as national vegetation height map and forest fire fuel mapping. GEDI is a spaceborne lidar system specifically designed to measure vertical forest structure and data have been collected since April 2019 (Silva et al., 2021). GEDI, onboard the International Space Station (ISS), uses a ~25 m diameter footprint full waveform lidar data to capture vegetation structure essential for aboveground biomass density (AGBD) estimates between 51.6 degrees North and South (Dubayah et al., 2020; Duncanson et al., 2020). Simmavong et al. (2025) reported that forest structure is a stronger predictor of AGBD than community-weighted mean traits, functional diversity, topography, or soil characteristics in a tropical forest across multiple spatial scales. One thing to note is that AGB is expressed as a mass, typically kg (kilogram), Mg (megagram or metric tonne) or Pg (petagram, or 10^9 tonnes) while AGBD is expressed as a mass per unit area, typically Mg/ha. In this study, we estimated AGBD in the unit of Mg/ha for our study area.

Spaceborne lidar sensors such as ICESat-2 and GEDI can provide lidar measurements on a global scale, which can fill the gaps in airborne lidar coverage (Tamiminia et al., 2024). Previous research utilizing spaceborne lidar with optical or Synthetic-aperture radar (SAR) remote sensing images to estimate AGB or canopy height achieved promising results (Liu and Popescu, 2022; Tamiminia et al., 2024; Dubayah et al., 2022; Jiang et al., 2021). There are multiple versions of AGBD data, which are derived from the processed waveform data, that GEDI provides: the GEDI Level 4 A (L4A) product, which provides predictions of AGBD (Mg/ha) and estimates of the prediction standard error within each sampled laser footprint (Dubayah et al., 2022); and the GEDI Level 4B product, which provides 1-km resolution estimates of mean AGBD and associated uncertainty (Dubayah et al., 2022). However, due to the coarse resolution and gaps in the GEDI data, finer scale estimates of biomass or height might not be available or accurate. Similarly, the spatial coverage of ICESat-2 data products is not continuous. Therefore, it is crucial to integrate continuous, high-resolution data from other sources to create wall-to-wall maps that leverage ICESat-2 and GEDI for more accurate vegetation applications.

Recent studies have demonstrated the effectiveness of integrating multi-source remote sensing data and advanced modeling techniques for estimating AGBD with high accuracy at regional scales. Liu and Popescu (2022) used Landsat 8 reflectance data, derived vegetation indices, Sentinel-1 data, terrain data and ICESat-2 height metrics for the prediction of biomass before and after fire to estimate carbon emissions with a 2-phase hierarchical modeling approach. This yielded random forest regression models with R^2 of 0.71 and RMSE of 45.91 Mg/ha. Tamiminia et al. (2024) integrated GEDI, Sentinel-1 and Sentinel-2 data to create wall-to-wall state-wide forest canopy height and AGB map for New York with 10 m resolution. The 10 m AGB map for New York had an R^2 of 0.65 and a RMSE of 39.49 Mg/ha.

Google Earth Engine (GEE) is a cloud computing platform launched in 2010 by Google (Gorelick et al., 2017). Cloud computing platforms provide efficient ways of storing, accessing, and analyzing datasets on powerful servers, which provide users with virtualized supercomputing capabilities, offering infrastructure, storage services, and software packages in various formats. (Chi et al., 2016; Ma et al., 2015). GEE can be accessed in various ways, including through an internet-based Application Programming Interface (API) and Interactive Development Environment (Tamiminia et al., 2020; Gorelick et al., 2017) Recently, due to its highly efficient image processing software and hardware capabilities (Gorelick et al., 2017; Hird et al., 2017), GEE has become a very popular geospatial data processing platform, with numerous free remotely sensed datasets being provided to users in a cloud-enabled data repository (Amani et al., 2019; Tamiminia et al., 2020). A variety of raw datasets, preprocessed data, elevation models, machine learning (ML) algorithms, image processing and different visualization tools are provided through the API. Commonly used satellite image collections such as Landsat, Sentinel and MODIS are gathered into GEE (Amani et al., 2020). Therefore, in this study, GEE was used as the primary computing platform for AGBD map production. Satellite imagery available in GEE was integrated with ICESat-2 data to develop models and map forest AGBD in the Pacific Northwest and California coastal area and to demonstrate the utility of GEE for this type of modeling activity. The methodology and results derived from this research can be used to scale up the AGBD estimation from regional to national scales and make it easily replicable in a time and cost-effective manner with the use of cloud-based, freely available satellite data, algorithms, and platforms (Nandy et al., 2021).

Multiple studies have investigated the use of ICESat-2 data for mapping biomass in boreal regions. Varvia et al. (2024) developed a hierarchical hybrid inference approach to estimate mean AGBD in boreal forests using ICESat-2 data. This method rigorously propagates uncertainty across multiple modeling steps (including allometric tree biomass models) and accounts for variance due to ICESat-2's sparse profile sampling. Neuenschwander et al., (2024) developed ICESat-2-based models to estimate AGBD for boreal forests, creating the first pan-boreal AGBD dataset at a 30 m resolution using spaceborne lidar. Feng et al. (2023) conducted a systematic evaluation of ICESat-2 ATL08 terrain and canopy height data across North American boreal forests, assessing its accuracy at multiple spatial resolutions using airborne LiDAR as reference data. Varvia et al. (2022) investigated the effects of time of day, beam strength, and snow cover on ICESat-2-based AGB estimation in boreal forests, finding that strong beam night data from snow-free conditions yielded the most accurate results. While previous research studies using on-orbit ICESat-2 data have focused exclusively on boreal forests, this study focused on the forests in Oregon, Washington, California and Idaho to develop the methodology to investigate the use of ICESat-2 data for mapping forest biomass outside of the boreal region, therefore exploring data and applicability of models in different ecological settings.

In this study, we aimed to map the AGBD with Google Earth Engine using ICESat-2 and GEDI data for the forests in Washington, Oregon, Idaho and California. We integrated the satellite images available in GEE with the ICESat-2 and GEDI canopy height map as predictors using machine learning methods to create wall-to-wall AGBD maps for 2016 and 2020. The main goal of this study was to estimate AGBD in the study area for recent years to develop a pipeline that leverages cloud computing resources and enables the production of maps and associated uncertainty metrics using ICESat-2, GEDI and other wall-to-wall data at different scales. We refined and validated that pipeline through a case study of producing a AGBD map and uncertainty metrics and assessing the impacts of pipeline development choices, such as choice of machine learning models. The specific objectives of this study are to: 1) develop a methodology to produce wall-to-wall AGBD maps at regional scale, 2) compare the accuracy of different machine learning models for AGBD mapping and investigate predictor variable importance, and 3) evaluate the uncertainties of the produced AGBD maps.

2. Materials and methods

2.1. Study areas

The study area encompasses a portion of the United States Department of Agriculture (USDA) Forest Service's Pacific Northwest region (Oregon and Washington), as well as Idaho and part of California that is contained within the Environmental Protection Agency's (EPA) Level II ecoregion maps (Fig. 1). EPA's Level II ecoregion maps for regions 9 and 10 were utilized to define our study area. This extensive study area covers ecoregions of cold desert, marine west coast forest, Mediterranean California and western Cordillera (Omernik and Griffith, 2014).

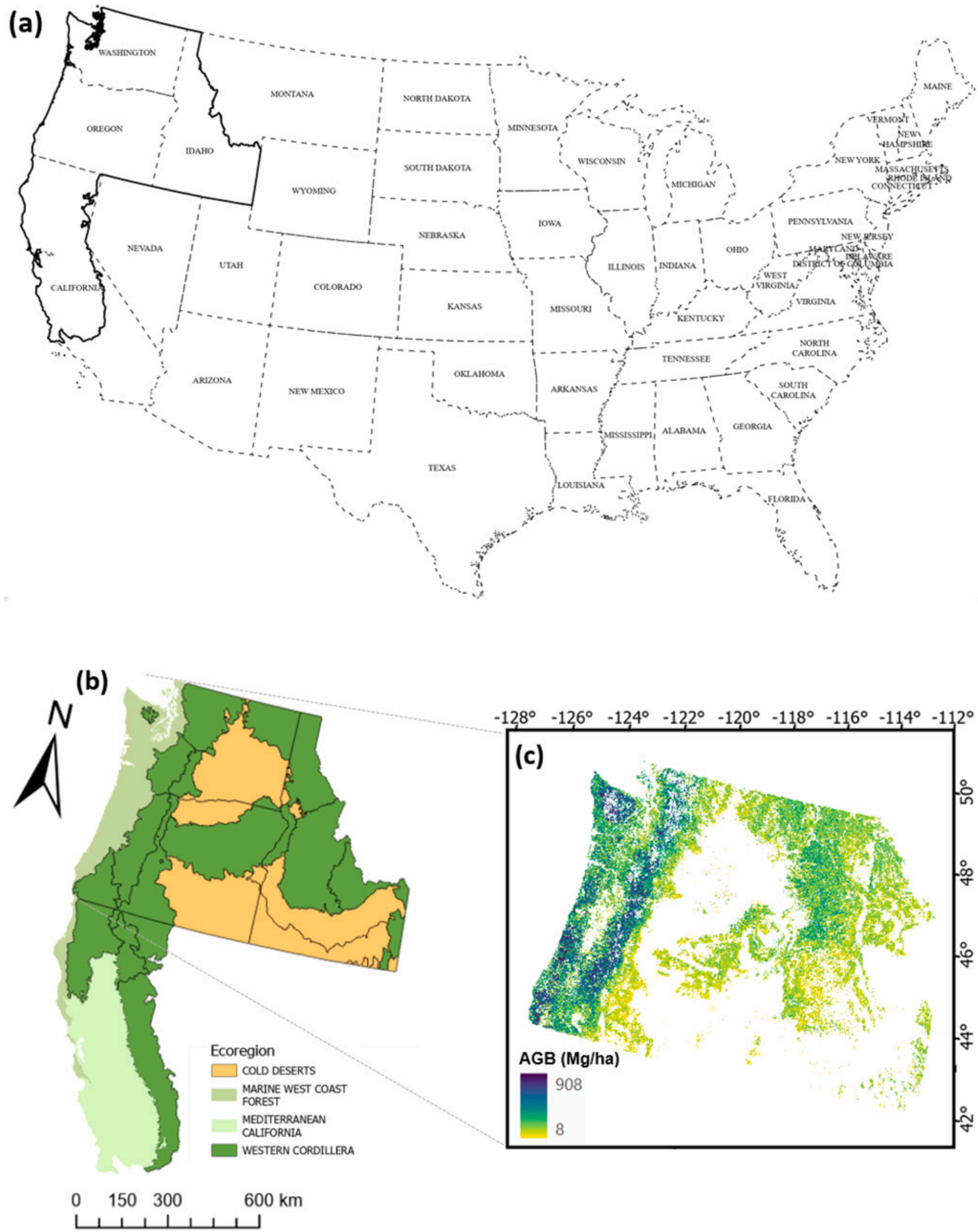

Fig. 1. Study areas: (a) study area in the United States, (b) EPA Level II Ecoregion classifications in Oregon, Washington, Idaho and California(c) lidar-derived AGBD map for 2016 used for training ML model.

2.2. Data

2.2.1. Reference AGBD map

The general approach to creating the modeling pipeline is as follows: 1) from an existing wall-to-wall biomass map (response variable), extracted training data that correspond with the footprints of ICESat-2 and GEDI data along with ancillary datasets as predictors, 2) built different machine learning models and compared the accuracy, 3) used the model with best performance to create wall-to-wall AGBD maps and estimated the uncertainties of the maps, and 4) validated the produced AGBD maps with FIA biomass ground data (Fig. 2).

There are multiple ways to incorporate GEE in the AGBD mapping. For example, download predictor data from GEE, build the model locally in Python and make maps. The second implementation of GEE in the map making pipeline is to download predictors from GEE, model locally and apply the model in GEE. In this study, we used the second approach to make the AGBD maps. We extracted predictor values in GEE, built the models in Python, then loaded the Python ML model as a GEE model, and applied that model to the predictors stored in GEE to create wall-to-wall AGBD maps.

A modeled AGBD dataset such as the airborne lidar derived biomass maps used to train the models is a surrogate for known ground data. The AGBD data (Fig. 1c) we used in this study covers forests in Washington, Oregon, Idaho, and western Montana, USA. The reference biomass dataset, derived from landscape model, provides annual AGBD maps (Mg/ha) for the years 2000–2016 with a spatial resolution of 30 m. It was produced using two Random Forest models: the first to upscale field plot AGBD measurements collected in conjunction with airborne lidar collections, and the second to upscale a random stratified sample of the lidar-based AGBD estimates using optical remote sensing data (Landsat). The resulting annual (2000–2016) AGBD map products were subsequently calibrated with same-year Forest Inventory and Analysis (FIA) inventory data and can be freely downloaded from the Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC) (Fekety and Hudak, 2019). In this study, the 2016 AGBD map was utilized for model training.

Jiang et al. (2021) utilized eight VIs and ICESat-2 derived canopy height to create wall-to-wall canopy height maps. Liu and Popescu (2022) also reported the use of VIs and ICESat-2 height metrics for the prediction of biomass before and after fire to estimate carbon emissions with the proposed 2-phase hierarchical modeling method, with R^2 of 0.71 and RMSE of 45.91 Mg/ha for the least absolute shrinkage and selection operator (LASSO) regression model. Thus, in this study, 5 VIs accompanied with other variables such as height, vegetation type and vegetation cover were used as predictors for the prediction of the AGBD (Tables 2 and 3).

2.2.2. Landsat gap-filled dataset and derived data products

Since cloud cover might cause gaps in the satellite images, Landsat's gap-filled dataset was used in the study. The gap-filled Landsat dataset was produced using the Highly Scalable Temporal Adaptive Reflectance Fusion Model (HISTARFM), which combines multiple multispectral

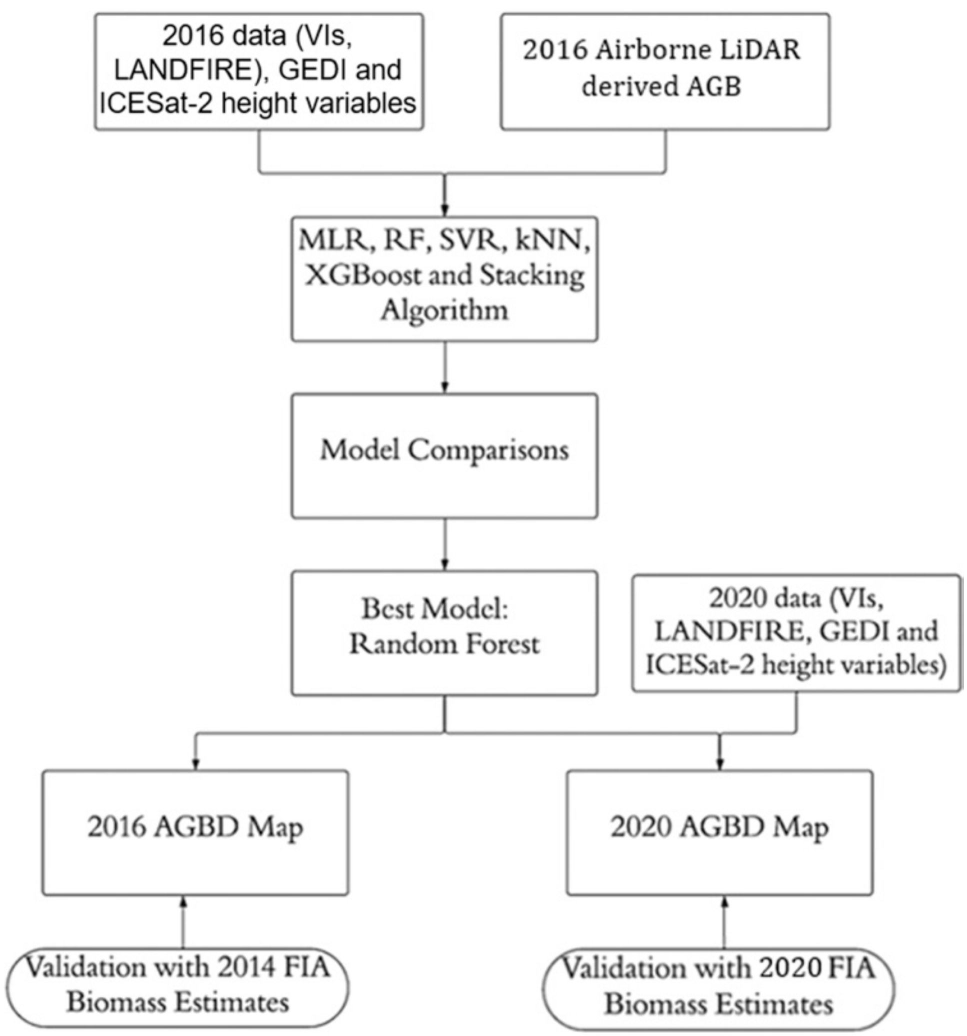

Fig. 2. Flowchart of the model training, prediction and validation process.

Landsat and MODIS imagery to produce monthly continuous high-resolution (30 m) data over the contiguous United States (Moreno-Martínez et al., 2020). The gap-filled dataset spans from 2009 to 2021 and can be accessed in Google Earth Engine (GEE) (Table 1). Vegetation indices (Table 2) used in the model training process were also derived from the gap-filled images in GEE (Table 3).

2.2.3. ICESat-2 data

A 30-meter resolution canopy height map for the contiguous United States was generated for the year 2020 using ICESat-2-derived canopy height measurements combined with ancillary data from Landsat, LANDFIRE, and topographic variables. ICESat-2 data from 2019 to 2020 was acquired for the modeling process. The modeling approach employed a gradient-boosted regression tree algorithm stratified by vegetation biome types (Malambo and Popescu, 2024). The resulting model demonstrated strong performance on a 15 % hold-out validation set, achieving an R^2 of 0.76, a mean absolute error (MAE) of 2.5 m, and a mean bias of 0.1 m, with a tendency toward slight overestimation. Performance varied by biome, with R^2 values ranging from 0.49 to 0.73 and MAE from 1.4 to 3.3 m. Additional validation using independent airborne LiDAR datasets yielded consistent results (airborne LiDAR: R^2 = 0.72, MAE = 3.9 m), confirming the model's robustness across diverse forest conditions. The final canopy height product is openly accessible and provides detailed spatial information suitable for large-scale forest structure analyses (Malambo and Popescu, 2024). In this study, the ICESat-2 2020 canopy height map was used as a predictor variable for modeling AGBD for both 2016 and 2020. The height map was treated as a static layer, under the assumption that forest structural changes over this three to four-year period were minimal and would not significantly affect the model's ability to capture broad biomass patterns. This assumption is supported by the relatively slow rate of structural change in mature forested ecosystems across large areas. While localized structural dynamics may exist, we considered the 2020 canopy height map to be representative for modeling purposes due to the limited availability of temporally matched ICESat-2 data and the spatial coverage it provides.

Table 2
List of Vegetation indices used in the study.

Vegetation Index	Description	Reference
Normalized difference vegetation index (NDVI)	(NIR − RED)/(NIR + RED)	Rouse et al., (1974)
Red–green vegetation index (RGVI)	(RED − GREEN)/(RED + GREEN)	Jensen, (1996)
Atmospherically resistant vegetation index (ARVI)	NIR − (2 × RED − BLUE)/ NIR+ (2 × RED − BLUE)	Kaufman and Tanre, (1992)
Enhanced vegetation index (EVI)	2.5 × (NIR − RED)/(NIR + RED − 7.5 × BLUE + 1)	Liu and Huete, (1995)
Visible atmospherically resistant index (VARI)	(GRN−RED)/ (GRN+RED−BLUE)	Jensen, (1996)

Table 3
Band wavelengths from harmonized Landsat 5,7,8 observations.

Band Number	Spectral Band	Wavelength (μm)
1	Blue	0.45–0.52
2	Green	0.52–0.60
3	Red	0.63–0.69
4	Near-Infrared	0.77–0.90
5	Near-Infrared	1.55–1.75
7	Mid-Infrared	2.08–2.35

2.2.4. GEDI data

There are two GEDI derived forest canopy height datasets used in the study for the training and prediction process respectively. The 2019 canopy height map with 30 m of spatial resolution was produced by fusing GEDI lidar forest structure measurements and Landsat analysis-ready time-series data. This fusion approach allowed for the spatial extrapolation of GEDI's footprint-level measurements to generate continuous global coverage. This dataset was used in the training process. (Potapov et al., 2021). The 30 m canopy height map derived from Potapov et al. (2021) was used as a static predictor in the model, despite representing 2019 conditions. This was done under the assumption that forest structure changes gradually over time, particularly in mature forest ecosystems, and that such changes between 2016 and 2019 were

Table 1
Datasets used for this study.

Dataset	Spatial Resolution	Description	Purpose
2016/2020 Monthly Gap-filled Landsat data	30 m	Monthly continuous high resolution data over CONUS (Moreno-Martínez et al., 2020). Accessed at Google Earth Engine (https://code.earthengine.google.com/?asset=projects/KalmanGFwork/GFLandsat_V1).	Model Training/ Prediction
2016/2020 LANDFIRE Existing Vegetation Type (EVT)	30 m	Represents the current distribution of the terrestrial ecological systems classification.	Model Training/ Prediction
2016/2020 LANDFIRE Existing Vegetation Cover (EVC)	30 m	Represents the vertically projected percent cover of the live canopy layer for a 30 m cell.	Model Training/ Prediction
2016/2020 LANDFIRE Existing Vegetation Height (EVH)	30 m	Represents the average height of the dominant vegetation for a 30 m cell.	Model Training/ Prediction
2019 Global Land Analysis and Discovery (GLAD) GEDI Global Forest Canopy Height	30 m	Produced by fusing Global GEDI lidar forest structure measurements and Landsat data (Potapov et al., 2021). URL: https://glad.umd.edu/dataset/gedi/	Model Training
2020 ICESat−2 height	30 m	The US 30 m ICESat−2 derived canopy height map for 2020 produced by LASERS lab. URL: https://lasers.tamu.edu/ice-cloudand-land-elevation-satellite-icesat−2-applications/	Model Training/ Prediction
2016 AGB map for Northwestern Forests	30 m	Annual maps of AGB (Mg/ha) for forests in Washington, Oregon, Idaho, and western Montana, USA, for the years 2000–2016. URL: https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id= 1719	Model Training
2020 GEDI and Sentinel−2 derived global canopy height map	10 m	Lang et al. (2023) produced global canopy height map by fusing sparse GEDI height data with Sentinel−2 images using a deep learning model. Accessed at Google Earth Engine (https://nlang.users.earthengine.app/view/global-canopy-height−2020).	Model Prediction
2014 FIA Aboveground Biomass	64,000-hectare hexagons	Forest Aboveground Biomass from FIA Plots across the Conterminous USA, 2009−2019 (Menlove and Healey, 2021) https://daac.ornl.gov/CMS/guides/FIA_Forest_Biomass_Estimates.html	Validation
2020 FIA Aboveground Biomass	64,000-hectare hexagons	2020 Forest Aboveground Biomass from FIA Plots across the Conterminous USA (Menlove and Healey, 2020) https://zenodo.org/records/4294490#.X9o5Q9hKg2w	Validation

minimal and spatially consistent enough not to significantly distort the model's ability to predict 2016 AGBD.

Lang et al. (2023) produced a global 10 m resolution canopy height dataset for the year 2020, derived from a probabilistic deep learning framework that integrates GEDI sparse height data with Sentinel-2 optical imagery. This product offers geospatially continuous estimates of canopy height and includes associated uncertainty metrics. Since the spatial resolution of this dataset is different from other datasets, we aggregated this dataset to 30 m by taking the maximum tree height as tree height for each grid. This dataset was used in the prediction process (Table 1).

2.2.5. LANDFIRE data

Landscape Fire and Resource Management Planning Tools (LANDFIRE, LF) is a program led by the U.S. Department of Agriculture Forest Service and the U.S. Department of the Interior to provide geospatial products for fire and natural resource management and cross-boundary planning. LANDFIRE provides many data products, including disturbance, fire behavior, fuel models, vegetation type, etc. In this study, LANDFIRE'S Existing Vegetation Type (EVT), Existing Vegetation Cover (EVC) and Existing Vegetation Height (EVH) data were used to create training models. EVT represents the current distribution of the terrestrial ecological systems classification. EVC represents the vertically projected percent cover of the live canopy layer for a 30 m cell. EVH represents the average height of the dominant vegetation for a 30 m cell. EVC and EVH were generated separately for tree, shrub, and herbaceous lifeforms (Picotte et al., 2019). The years 2016 and 2020 data were used as predictors in the AGBD models.

2.2.6. Validation data

Since AGBD maps were produced for the years 2016 and 2020 in this study, two validation datasets were used. Forest biomass estimates for the conterminous United States based on data from the USDA Forest Inventory and Analysis (FIA) program were used to validate the 2016 and 2020 AGBD maps (Menlove and Healey, 2021; Menlove and Healey, 2020). The 2016 dataset is derived from field data ranging from 2009 to 2019 with a spatial resolution of 64,000-hectare hexagon (Menlove and Healey, 2021).

Menlove and Healey (2020) produced a national scale forest biomass dataset for the conterminous United States to support the validation of remotely sensed biomass estimates. The dataset is based on FIA's extensive field-based measurements, which include tree dimensions such as tree height and diameter. Biomass estimates are derived using three sets of allometric equations, as described in Menlove and Healey (2020): Jenkins equations, Component Ratio Method (CRM) and Retired FIA regional methods (DRYBIOT). We used the live tree biomass estimates to validate our map, and the live tree biomass was estimated using the FIA component ratio method (CRM) (Fekety and Hudak, 2019). Biomass estimates of this study are mapped across approximately 12, 591 hexagonal areas with each covering about 64,000 ha. These hexagons are derived from the Forest Inventory and Analysis (FIA) Program's national sampling research. This dataset was used to validate our 2020 predicted AGBD map.

2.3. Model development and assessment

2.3.1. Model development

In this study, we explored and compared six estimation algorithms, including a parametric approach (MLR: Multiple Linear Regression), two machine learning regression algorithms (kNN: k-nearest neighbors, SVR: Support Vector Regression) and three ensemble learning algorithms (RF: Random Forest, Stacking Algorithm and XGBoost: eXtreme Gradient Boosting). By assessing a suite of several prediction algorithms, we were able to assess the impacts of algorithm choice on outcomes. The best model with the highest accuracy in terms of RMSE and R^2 for 2016 was subsequently applied to generate AGBD predictions for both 2016 and 2020. This approach assumes that the statistical relationships between predictor variables (e.g., canopy height, spectral indices, topography) and biomass remain relatively stable over the four-year period. By applying a 2016-trained model to predict 2020 biomass, we implicitly assume that forest structural changes such as growth, disturbance, or regeneration do not significantly alter these relationships at the scale of the analysis and the broader forest patterns are expected to remain consistent enough to support meaningful temporal comparison. This modeling strategy enables assessment of biomass change by comparing model outputs across years while maintaining a consistent training framework. However, we acknowledge that some prediction errors in 2020 may stem from real structural changes not accounted for in the static model.

MLR describes the linear relationship between a dependent variable and multiple independent variables. The MLR model is shown in Eq. (1).

$$\gamma = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_n X_n + \epsilon \tag{1}$$

where γ is the dependent variable, X_1, X_2, X_3 … X_n are the independent variables, β_1, β_2, β_3… β_n are the regression coefficients and ϵ is error.

Results from nonparametric models such as SVR, kNN and the RF-based models were compared with those from MLR. SVR is a supervised machine learning algorithm used for regression tasks (Drucker et al., 1996). It belongs to the family of Support Vector Machines (SVMs) (Cortes and Vapnik, 1995), which are primarily used for classification tasks, but SVR extends the concept to regression problems. SVM algorithms transform the input data from low-dimensional into high-dimensional feature space to make them linearly separable by the nonlinear kernel function. SVMs try to find a hyperplane that maximally separates different classes or output values in a high-dimensional space. SVR can use multiple types of kernels, which are functions that measure the similarity between input vectors, such as linear kernels and non-linear kernels, e.g., radial basis function (RBF), polynomial kernels and sigmoid kernels. A linear kernel is a simple dot product between two input vectors. In contrast, a non-linear kernel is a more intricate function capable of capturing more complex patterns in the data. The selection of the kernel depends on the characteristics of the data and the complexity of the task. In this study, we used a grid search cross validation method to find the best combination of hyperparameters and the RBF kernel was used for the model. Since SVR is sensitive to outliers and can underperform due to features with vastly different scales, we normalized the predictors to ensure all predictors contribute equally to the decision boundary for better generalization and performance.

k-nearest neighbor (kNN) is a non-parametric supervised machine learning algorithm that can be used for classification and regression by calculating the multivariable distances between all pairs of data points (Keller et al., 1985). In kNN classification, the output is k classes, where each object is assigned to the class with the shortest distance from it. In kNN regression, the output value is the average of the values of k nearest neighbors. If $k = 1$, then the value of the single nearest neighbor is assigned as the output value. In this study, the number of neighbors was set to three and the uniform weighting of the distance method was used for the kNN model.

Random Forest is a widely used nonparametric ensemble learning method that constructs a set of decision trees, the results of which are summarized for classification and regression tasks. RF predictions are made by averaging (for regression) or voting (for classification) across all trees; this procedure tends to be robust to outliers and noise in the data due to the averaging or voting mechanism (Breiman, 2001). The number of regression trees in our RF models was set to 100, and the maximum depth of trees was not limited in this study.

The Stacking algorithm is a machine learning method combining the predictions of various models to create a more accurate prediction. All of the prediction results from the base models served as independent variables to make the new modeling. The Stacking algorithm generalizes the output results of all base models to improve the overall prediction accuracy and reduce the overall error (Wolpert, 1992). Jiang et al. (2021) reported that the Stacking algorithm outperformed the other models (MLR, SVM, kNN and RF) in terms of ICESat-2 derived canopy height accuracy. In our study, the Stacking algorithm is comprised of MLR, SVM, kNN, and RF.

In addition, XGBoost (short for Extreme Gradient Boosting) (Chen and Guestrin, 2016) is a popular tree-based ensemble learning method mainly designed for speed and performance improvement, which utilizes boosting and gradient descent methods to combine decision trees (Yang and Shami, 2020). Both XGBoost and Random Forest are ensemble learning algorithms of decision trees. However, XGBoost is trained sequentially with new trees correcting the errors from the previous trees. In contrast, Random Forest is trained independently with predictions made by averaging for regression across all trees.

2.3.2. Model assessment

The training data with ten predictors were randomly divided into two groups: 70 % (training data, n = 53484) for model training and 30 % (testing data, n = 22994) for model testing. The coefficient of determination (R^2) and RMSE (Willmott et al., 1985) were computed to assess model performance with the testing set.

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

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}} \tag{3}$$

where y_i is the observed value for the i^th^ case, $\hat{y}_i$ is the predicted value based on the model. $\bar{y}$ is the mean of all the observed values, and n is the sample size. All of the model training and comparisons were implemented in the Python programming environment.

Feature importance of the selected machine learning model was also calculated for each predictor used in the machine learning model. Mean Decrease in Accuracy (MDA) is a widely used metric for assessing variable importance in machine learning models, particularly in decision tree-based algorithms such as Random Forests. MDA leverages out-of-bag (OOB) samples, which are the subsets of data not used during the construction of a given tree. These OOB samples serve a dual purpose: they provide an estimate of model performance and are also used to evaluate variable importance. The process begins by computing the model's prediction error on the OOB samples, through mean squared error (MSE) for regression tasks and error rate for classification tasks. Subsequently, each predictor variable is permuted, and the model's error is recalculated. The difference in error between the original and permuted datasets reflects the importance of the variable in improving the model's predictive accuracy. This process is repeated across all trees in the forest, and the results are averaged to provide a final importance score for each variable. A larger MDA value indicates greater importance, as it corresponds to a larger decrease in model accuracy when the variable is permuted. This method is advantageous because it is less biased than alternatives like the Gini index and is applicable to both regression and classification tasks (Breiman, 2001; Han et al., 2016).

2.3.3. Wall-to-wall AGBD maps production

After the model assessment process, the machine learning model with the highest accuracy was used for AGBD predictions in Google Earth Engine. Training data and the predictors data in CSV format were uploaded to GEE for producing wall-to-wall AGBD maps in our study area. The GEE code used for this study can be accessed at: https://code.earthengine.google.com/3183cfb3762028a4bb0d043d809b6471.

2.3.4. Validation of AGBD predictions

Since we have predicted AGBD for years 2016 and 2020, two independent validation datasets were used in the study to validate the accuracy of maps from different years. AGBD maps for 2016 and 2020 were validated with biomass estimates from FIA. The FIA biomass estimates from Menlove and Healey (2021) are more temporally aligned with the 2016 data, and the biomass estimates from Menlove and Healey (2020) are closer to the 2020 data. Therefore, they were used for validating AGBD maps created for different years.

We generated 1000 randomly selected sites across the study area, and the RMSE and R^2 were reported to evaluate the accuracy of the predictions. Since the AGBD estimates from FIA were 64,000-hectare hexagons, the 30 m biomass map for the years 2016 and 2020 was resampled to corresponding spatial resolution to make the comparisons comparable. The resampling process averaged 30-m pixels within the hexagonal grid.

2.4. Uncertainty analysis

The uncertainties of the Random Forest modeling as well as the errors related to the data used for modeling were considered for the derivation of the uncertainty map. Therefore, in this study, variance is estimated as a function of: (1) the variability among multiple modeling predictions and (2) the uncertainty of the predictors used in the models (Patterson et al., 2019). For the i^th^ pixel, the uncertainty can be defined as:

$$\sigma_i = \sqrt{\sigma^2_{RF,\quad i} + \sigma^2_{predictor,\quad i}} \tag{4}$$

where $\sigma_{RF,i}$ is the standard deviation of the estimated AGBD for the ith pixel derived from estimates of all Random Forest models and $\sigma_{predictor,i}$ is the standard deviation for the predictors used in the Random Forest model.

3. Results

3.1. AGBD estimation models

The performances of the six models varied, with RF and XGBoost bearing the most similar results. The Random Forest model achieved the best fit (R^2 = 0.768) and the lowest error (RMSE = 61.8 Mg/ha) with XGBoost (R^2 = 0.761; RMSE = 62.72 Mg/ha) performing second. SVR performed the worst (R^2 = 0.524; RMSE = 103.28 Mg/ha) among all models, indicating an underestimation trend (Fig. 3). Fig. 4 shows the feature importance for the predictors trained in the random forest regression model. GEDI height is the most crucial variable for biomass prediction, followed by Existing Vegetation Type (EVT), enhanced vegetation index (EVI) and Existing Vegetation Cover (EVC). ICESat-2 height, atmospherically resistant vegetation index (ARVI) and visible atmospherically resistant index (VARI) are the least important variables for predicting biomass. Although canopy height is typically a strong predictor of aboveground biomass, the relatively low importance of the ICESat-2-derived height variable in our model may be attributed to several factors. First, the ICESat-2 canopy height map used in this study represents conditions in 2020 and was treated as a static layer when modeling both 2016 and 2020 AGBD, potentially reducing its relevance in capturing temporal biomass variation, especially for 2016. Second, the GEDI-derived canopy height product, also included as a predictor, exhibited stronger correlation with biomass in our training dataset. Given the overlap in information content between the two height variables, the model may have preferentially relied on the GEDI product, leading to a lower ranked importance for ICESat-2 height. Lastly, differences in spatial sampling density and signal characteristics between GEDI (full waveform) and ICESat-2 (photon-counting) may also have influenced the predictive utility of the height metrics in the context of the random forest model. However, if we left out ICESat-2 height data from the model, the accuracy of the model decreased. Thus, all variables

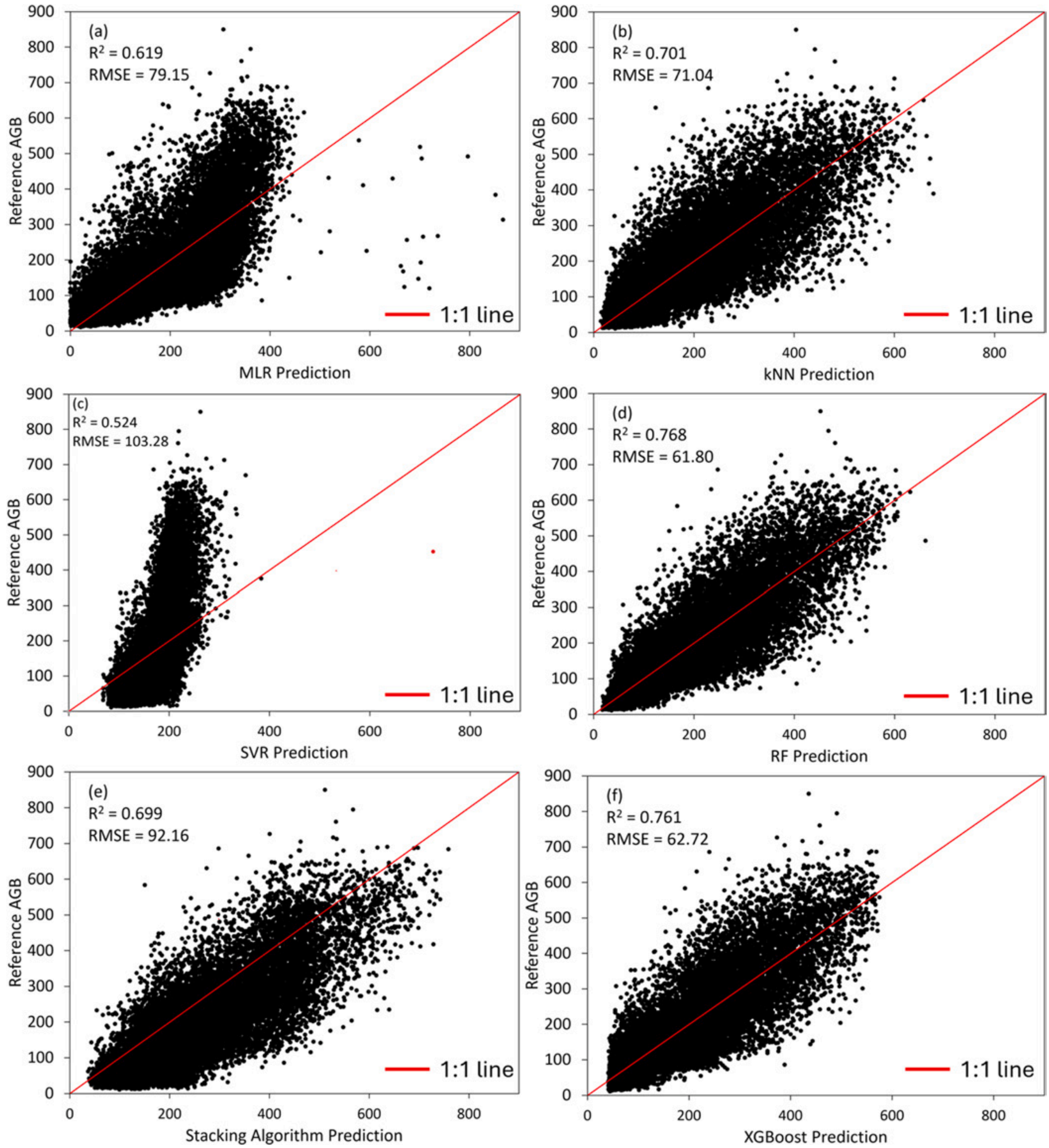

Fig. 3. Scatter plots of the predicted AGBD against the observed values in the study area for the model training process using: (a) MLR, (b) kNN, (c) SVR, (d) RF, (e) Stacking and (f) XGBoost.

were used for the training and prediction process.

3.2. Wall-to-wall AGBD maps

Since Random Forest has the highest accuracy in terms of R^2 and RMSE, it was used for generating the AGBD maps. Vegetation indices, LANDFIRE data, ICESat-2 height and GEDI height data were used to train the Random Forest Regression model and predict on the new data. The predicted wall-to-wall AGBD maps for 2016 and 2020 are shown in Fig. 5a and b. The forests in the coastal areas of Washington, Oregon and California have higher mean biomass density than inland portions of the study area. The mean AGBD difference between 2016 and 2020 was 19.29 Mg/ha, indicating an increasing trend in biomass across the study area over the four-year period. The AGBD ranges from 18 to 620 Mg/ha for both biomass maps. The spatial pattern of the predicted forest AGBD matched with the actual distribution of the forests in the study area.

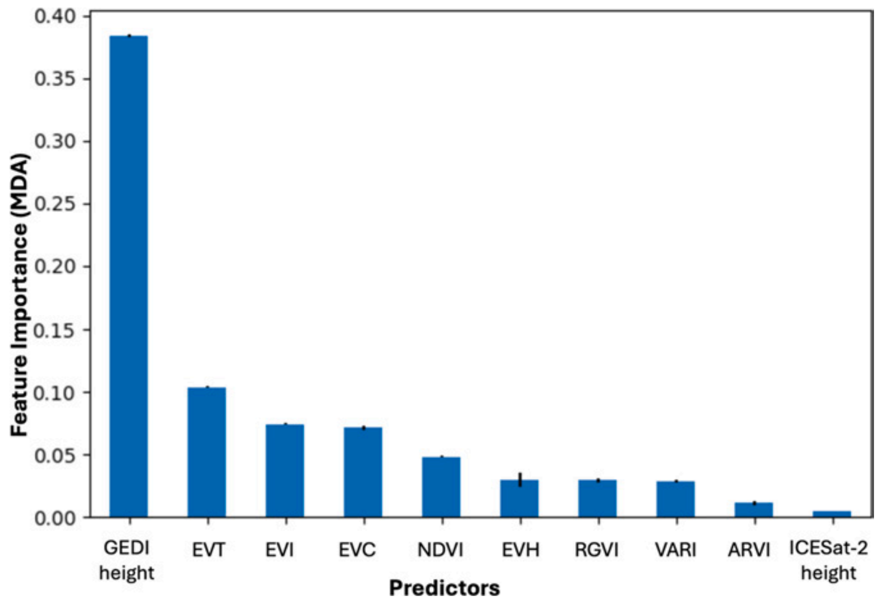

Fig. 4. Feature importance for predictors in the Random Forest model.

3.3. Validation of the maps

To validate the accuracy of the produced continuous biomass maps, two independent validation datasets not included in the trained model were used in the study. The 2016 Random Forest predicted AGBD map was validated with 2014 FIA biomass data, which achieved satisfactory accuracy ($R^2 = 0.769$; RMSE = 55.26 Mg/ha). The 2020 predicted AGBD map validated with 2020 FIA biomass data achieved satisfactory accuracy ($R^2 = 0.572$; RMSE = 84.96 Mg/ha) (Fig. 6).

3.4. Uncertainties

The corresponding uncertainties are calculated for the continuous biomass maps and shown in Fig. 5c and d. The uncertainty map indicates that coastal forests have higher uncertainties than non-forested areas. Some inland areas also have higher variations. The quantification of the uncertainty allows us to have a sense of where the biomass could be used with reliability and where should be used with care.

4. Discussion

4.1. Uncertainties

The uncertainties are comprised of modeling errors and error propagation from the predictors. The standard deviation of Random Forest models and the used predictors were combined to estimate uncertainties of AGBD across the study area. However, since we only included the uncertainty of predictors used in the model, uncertainties from bottom-up approaches such as tree measurement errors, allometric errors and lidar measurement and georegistration errors are not included in the uncertainties of the used parameters. Depending on the estimation methods and the predictors used, these errors might introduce additional errors. However, other sources of error contributing to the overall uncertainty, such as remote sensing measurement error, sensor noise, field measurement error, and other sources, were considered negligible compared to the error due to modeling (Garcia et al., 2017). Thus, these other sources of errors were left out when calculating the uncertainty. This can be a limitation of the study since it is hard to account for the bottom-up errors when we use the predictions derived from the raw data. However, the agreement between the estimated AGBD and the independent AGBD datasets verifies the reliability of the proposed model and results. We have demonstrated a way to make a pixel-based map of uncertainty, which can be helpful in certain contexts, such as interpreting land cover maps (Congalton et al., 2014). Since uncertainty map is a pixel-based uncertainty that does not link to a traditional sample-based uncertainty metric like a confidence interval derived from sampling, care must be taken when interpreting uncertainties based on map predictions.

4.2. Model accuracy

The performance comparison of six machine learning algorithms for AGBD prediction, as illustrated in Fig. 3, reveals notable differences in model accuracy and consistency. Among the models, RF and XGBoost algorithms exhibit the strongest performance, with high coefficients of determination ($R^2 = 0.768$ and 0.761, respectively) and relatively low RMSE values (61.80 and 62.72 Mg/ha), suggesting robust predictive capabilities. In contrast, SVR performs poorly ($R^2 = 0.524$; RMSE = 103.28 Mg/ha), deviating markedly from the other models. Despite the fact that we did the hyperparameter tuning with grid search cross validation and predictors scaling, SVR underperformed likely due to incompatibility with the underlying data distribution. Furthermore, the final RF model was trained using a random 70/30 split, as noted in Section 2.3.2. While standard, this approach may inflate performance metrics due to spatial autocorrelation, particularly if all samples were derived from a single LiDAR-based AGB map. The absence of spatially explicit cross-validation (CV) strategies (e.g., block or leave-location-out CV) and its implications for generalizability should be acknowledged as a limitation. Finally, the feature importance analysis (Fig. 4) underscores the dominance of GEDI-derived canopy height in driving model performance, vastly outweighing other predictors. This reinforces the value of spaceborne LiDAR in biomass estimation but also highlights potential vulnerabilities if GEDI data quality or availability fluctuate.

A substantial decline in model performance was observed between the 2016 and 2020 validation periods, with R^2 decreasing from 0.769 to 0.572. This marked reduction warrants careful consideration, as it suggests limitations in the model's temporal transferability. Several factors may contribute to this drop in accuracy. First, forest disturbances such as wildfires, pest outbreaks, or logging activities occurring between the two time periods could have altered vegetation structure, leading to discrepancies between predicted and observed AGBD. Second, variations in the quality, resolution, or completeness of predictor datasets

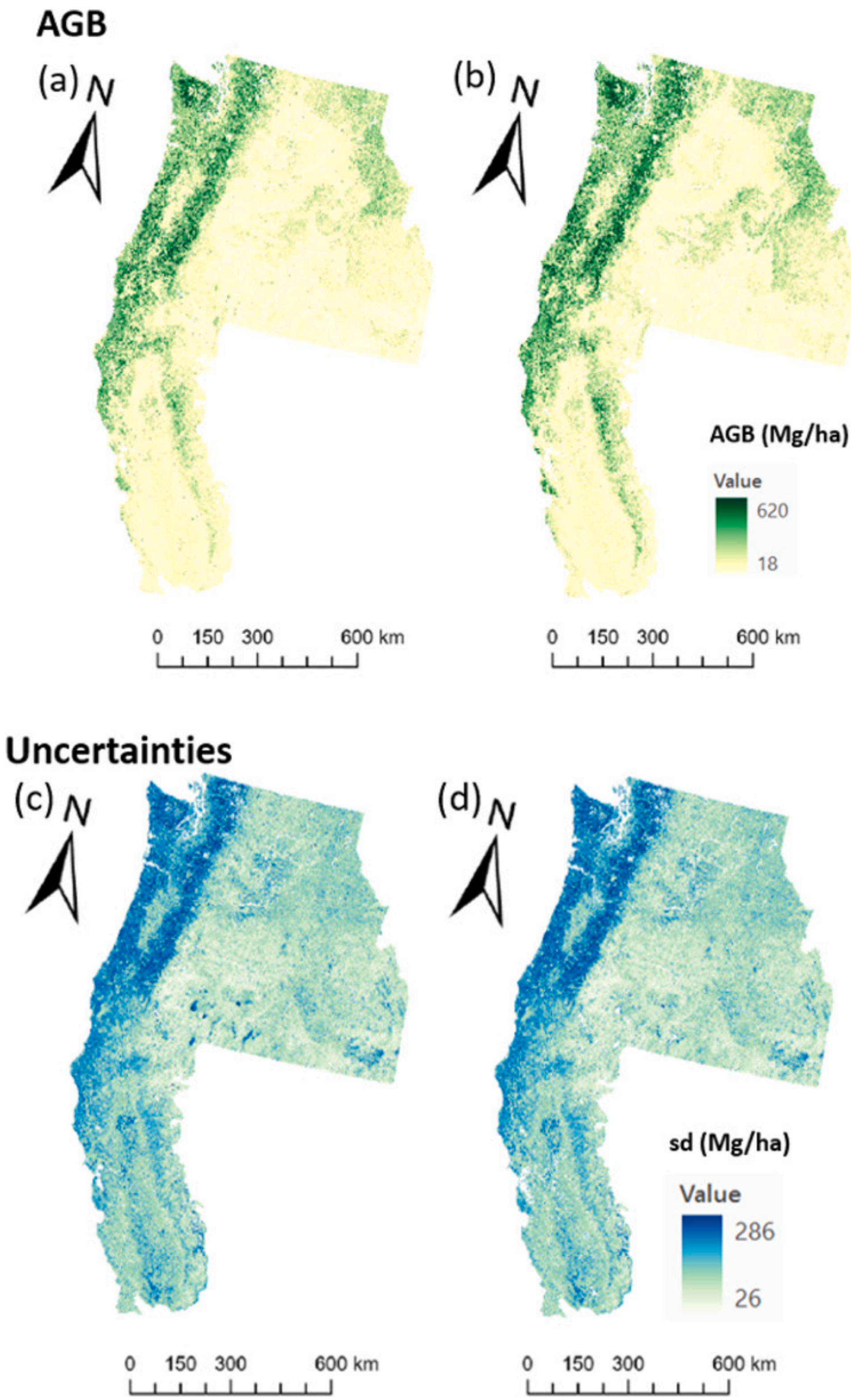

Fig. 5. AGBD for the study area and the corresponding uncertainties: (a) 2016 AGBD and (b) 2020 AGBD; the standard deviation of errors for (c) 2016 (d) 2020.

may have influenced model performance. For instance, data gaps or sensor calibration differences could introduce inconsistencies. Lastly, changes in land cover or forest composition over time may have reduced the representativeness of the training data for conditions in 2020. These temporal mismatches highlight the need for models to incorporate mechanisms that account for landscape dynamics when assessing long-term predictive performance.

5. Conclusion

Biomass prediction is essential for carbon stock estimation in forests, change detection and quantification, providing precious insights for policy-making and management purposes. The hybrid methodology proposed in this paper shows feasibility and can be used for scaling up from regional to national scale and producing the continental US

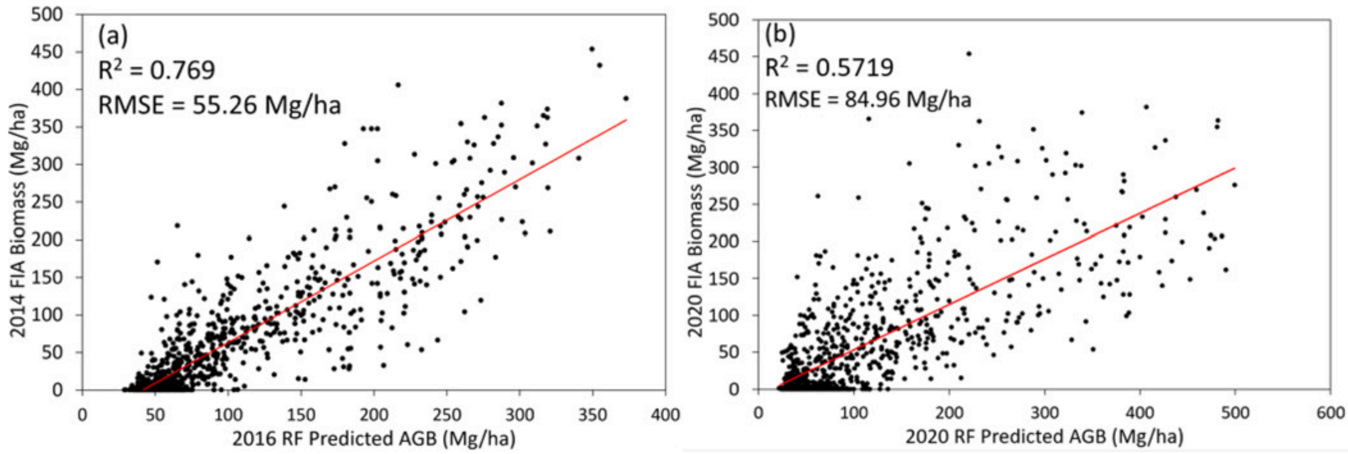

Fig. 6. Validation results for (a) 2016 and (b) 2020 AGBD map.

biomass map. NASA's ICESat-2 mission with global coverage captures the vertical structure of the forests, which can provide accurate estimates of biomass and be used for multiple applications such as biomass burning emissions and wildland fuel assessments. This study integrated spaceborne LiDAR data (ICESat-2 and GEDI), optical images (Landsat data) and vegetation data (LANDFIRE) to demonstrate how to utilize the ICESat-2 and GEDI data with other data sources for biomass estimation via data fusion for the forests in Washington, Oregon, Idaho and California. The results of our study suggested that the modeling approach of this study can provide AGBD prediction with high accuracy and let us understand the current state of the forests. Future studies can include the upcoming mission NiSAR and enhanced GEDI biomass data for more accurate biomass mapping.

CRediT authorship contribution statement

Sorin Popescu: Writing – review & editing, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Mei-Kuei Lu:** Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Andrew Hudak:** Writing – review & editing. **Lonesome Malambo:** Writing – review & editing, Data curation. **Andrew Lister:** Writing – review & editing.

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. The findings and conclusions in this report are those of the author(s) and should not be construed to represent any official USDA or U.S. Government determination or policy. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. government.

Acknowledgements

This research was funded by NASA ICESat-2 Science Team (Grant # NNH22ZDA001N). This research was supported [in part] by the U.S. Department of Agriculture, Forest Service. We are grateful to the anonymous reviewers for their insightful comments and suggestions that strengthened this manuscript.

Data availability

Data will be made available on request.

References

Amani, M., Ghorbanian, A., Ahmadi, S.A., Kakooei, M., Moghimi, A., Mirmazloumi, S.M., Brisco, B., 2020. Google earth engine cloud computing platform for remote sensing big data applications: a comprehensive review. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 13, *5326–5350*.

Amani, M., Mahdavi, S., Afshar, M., Brisco, B., Huang, W., Mohammad Javad Mirzadeh, S., Hopkinson, C., 2019. Canadian wetland inventory using google earth engine: the first map and preliminary results. Remote Sens. 11 (7), 842.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794)..

Chi, M., Plaza, A., Benediktsson, J.A., Sun, Z., Shen, J., Zhu, Y., 2016. Big data for remote sensing: Challenges and opportunities. Proc. IEEE 104 (11), 2207–2219.

Congalton, R.G., Gu, J., Yadav, K., Thenkabail, P., Ozdogan, M., 2014. Global land cover mapping: a review and uncertainty analysis. Remote Sens. 6 (12), 12070–12093.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20, 273–297.

Drucker, H., Burges, C.J., Kaufman, L., Smola, A., Vapnik, V., 1996. Support vector regression machines. Adv. Neural Inf. Process. Syst. 9.

Dubayah, R.O., Armston, J., Healey, S.P., Yang, Z., Patterson, P.L., Saarela, S., Kellner, J. R.(2022). GEDI L4B gridded aboveground biomass density, version 2. ORNL DAAC..

Dubayah, R.O., Armston, J., Kellner, J.R., Duncanson, L., Healey, S.P., Patterson, P.L., Luthcke, S.B., 2022. GEDI L4A footprint level aboveground biomass density. Version 2. 1. ORNL DAAC Oak Ridge Tenn. USA.

Dubayah, R., Blair, J.B., Goetz, S., Fatoyinbo, L., Hansen, M., Healey, S., Silva, C., 2020. The Global ecosystem dynamics investigation: high-resolution laser ranging of the Earth's forests and topography. Sci. Remote Sens. 1, 100002.

Duncanson, L., Armston, J., Disney, M., Avitabile, V., Barbier, N., Calders, K., Williams, M., 2019. The importance of consistent global forest aboveground biomass product validation. Surv. Geophys. 40 (4), 979–999.

Duncanson, L., Neuenschwander, A., Hancock, S., Thomas, N., Fatoyinbo, T., Simard, M., Dubayah, R., 2020. Biomass estimation from simulated GEDI, ICESat-2 and NISAR across environmental gradients in Sonoma County, California. Remote Sens. Environ. 242, 111779.

Fekety, P.A., & Hudak, A.T. (2019). Annual Aboveground Biomass Maps for Forests in the Northwestern USA, 2000–2016. ORNL DAAC: Oak Ridge, TN, USA..

Feng, T., Duncanson, L., Montesano, P., Hancock, S., Minor, D., Guenther, E., Neuenschwander, A., 2023. A systematic evaluation of multi-resolution ICESat-2 ATL08 terrain and canopy heights in boreal forests. Remote Sens. Environ. 291, 113570.

Garcia, M., Saatchi, S., Casas, A., Koltunov, A., Ustin, S., Ramirez, C., Balzter, H., 2017. Quantifying biomass consumption and carbon release from the California Rim fire by integrating airborne LiDAR and Landsat OLI data. J. Geophys. Res. Biogeosci. 122 (2), 340–353.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27.

Hall, F.G., Bergen, K., Blair, J.B., Dubayah, R., Houghton, R., Hurtt, G., Wickland, D., 2011. Characterizing 3D vegetation structure from space: mission requirements. Remote Sens. Environ. 115 (11), 2753–2775.

Han, H., Guo, X., & Yu, H. (2016, August). Variable selection using mean decrease accuracy and mean decrease gini based on random forest. In 2016 7th ieee international conference on software engineering and service science (icsess) (pp. 219-224). IEEE..

Hird, J.N., DeLancey, E.R., McDermid, G.J., Kariyeva, J., 2017. Google Earth Engine, open-access satellite data, and machine learning in support of large-area probabilistic wetland mapping. Remote Sens. 9 (12), 1315.

Hudak, A.T., Evans, J.S., Stuart Smith, A.M., 2009. LiDAR utility for natural resource managers. Remote Sens. 1 (4), 934–951.

Hummel, S., Hudak, A.T., Uebler, E.H., Falkowski, M.J., Megown, K.A., 2011. A comparison of accuracy and cost of LiDAR versus stand exam data for landscape management on the Malheur National Forest. J. for. 109 (5), 267–273.
Jensen, J.R. (1996). Introductory digital image processing: a remote sensing perspective.
Jiang, F., Zhao, F., Ma, K., Li, D., Sun, H., 2021. Mapping the forest canopy height in Northern China by synergizing ICESat-2 with Sentinel-2 using a stacking algorithm. Remote Sens. 13 (8), 1535.
Kaufman, Y.J., Tanre, D., 1992. Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. IEEE Trans. Geosci. Remote Sens. 30 (2), 261–270.
Keller, J.M., Gray, M.R., Givens, J.A., 1985. A fuzzy k-nearest neighbor algorithm. IEEE Trans. Syst. Man Cybern. (4), 580–585.
Lamping, J., Lucash, M., Bell, D.M., Irvine, D.R., Gregory, M., 2025. Moderate-resolution mapping of aboveground biomass stocks, forest structure, and composition in coastal Alaska and British Columbia. For. Ecol. Manag. 583, 122576.
Lang, N., Jetz, W., Schindler, K., Wegner, J.D., 2023. A high-resolution canopy height model of the Earth. Nat. Ecol. Evol. 7 (11), 1778–1789.
Liu, H.Q., Huete, A., 1995. A feedback based modification of the NDVI to minimize canopy background and atmospheric noise. IEEE Trans. Geosci. Remote Sens. 33 (2), 457–465.
Liu, M., Popescu, S., 2022. Estimation of biomass burning emissions by integrating ICESat-2, Landsat 8, and Sentinel-1 data. Remote Sens. Environ. 280, 113172.
Lu, M.K., Lam, T.Y., Perng, B.H., Lin, H.T., 2019. Close-range photogrammetry with spherical panoramas for mapping spatial location and measuring diameters of trees under forest canopies. Can. J. For. Res. 48 (8), 865–874.
Ma, Y., Wu, H., Wang, L., Huang, B., Ranjan, R., Zomaya, A., Jie, W., 2015. Remote sensing big data computing: challenges and opportunities. Future Gener. Comput. Syst. 51, 47–60.
Malambo, L., Popescu, S.C., 2021. Assessing the agreement of ICESat-2 terrain and canopy height with airborne lidar over US ecozones. Remote Sens. Environ. 266, 112711.
Malambo, L., Popescu, S., 2024. Mapping vegetation canopy height across the contiguous United States using ICESat-2 and ancillary datasets. Remote Sens. Environ. 309, 114226.
Menlove, J., Healey, S.P., 2020. A comprehensive forest biomass dataset for the USA allows customized validation of remotely sensed biomass estimates. Remote Sens. 12 (24), 4141.
Menlove, J., & Healey, S. P. (2021). CMS: Forest Aboveground Biomass from FIA Plots across the Conterminous USA, 2009-2019. *ORNL DAAC*.
Moreno-Martínez, Á., Izquierdo-Verdiguier, E., Maneta, M.P., Camps-Valls, G., Robinson, N., Muñoz-Marí, J., Running, S.W., 2020. Multispectral high resolution sensor fusion for smoothing and gap-filling in the cloud. Remote Sens. Environ. 247, 111901.
Nandy, S., Srinet, R., Padalia, H., 2021. Mapping forest height and aboveground biomass by integrating ICESat-2, Sentinel-1 and Sentinel-2 data using Random Forest algorithm in northwest Himalayan foothills of India. Geophys. Res. Lett. 48 (14) e2021GL093799.
Narine, L.L., Popescu, S.C., Malambo, L., 2020. Using ICESat-2 to estimate and map forest aboveground biomass: a first example. Remote Sens. 12 (11), 1824.
Narine, L.L., Popescu, S., Neuenschwander, A., Zhou, T., Srinivasan, S., Harbeck, K., 2019. Estimating aboveground biomass and forest canopy cover with simulated ICESat-2 data. Remote Sens. Environ. 224, 1–11.
Neuenschwander, A., Duncanson, L., Montesano, P., Minor, D., Guenther, E., Hancock, S., Stereńczak, K., 2024. Towards global spaceborne lidar biomass: developing and applying boreal forest biomass models for ICESat-2 laser altimetry data. Sci. Remote Sens. 10, 100150.
Omernik, J.M., Griffith, G.E., 2014. Ecoregions of the conterminous United States: evolution of a hierarchical spatial framework. Environ. Manag. 54, 1249–1266.
Patterson, P.L., Healey, S.P., Stahl, G., Saarela, S., Holm, S., Anderson, H.E., Yang, Z., 2019. Statistical properties of hybrid estimators proposed for GEDI- NASA's global ecosystem dynamics investigation. Environ. Res. Lett. 14 (6), 065007.
Perng, B.H., Lam, T.Y., Lu, M.K., 2018. Stereoscopic imaging with spherical panoramas for measuring tree distance and diameter under forest canopies. For. Int. J. For. Res. 91 (5), 662–673.
Picotte, J.J., Dockter, D., Long, J., Tolk, B., Davidson, A., Peterson, B., 2019. LANDFIRE remap prototype mapping effort: developing a new framework for mapping vegetation classification, change, and structure. Fire 2 (2), 35.
Popescu, S.C., Wynne, R.H., Nelson, R.F., 2002. Estimating plot-level tree heights with lidar: local filtering with a canopy-height based variable window size. Comput. Electron. Agric. 37 (1-3), 71–95.
Potapov, P., Li, X., Hernandez-Serna, A., Tyukavina, A., Hansen, M.C., Kommareddy, A., Hofton, M., 2021. Mapping global forest canopy height through integration of GEDI and Landsat data. Remote Sens. Environ. 253, 112165.
Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W., 1974. Monitoring vegetation systems in the Great Plains with ERTS. NASA Spec. Publ. 351 (1), 309.
Silva, C.A., Duncanson, L., Hancock, S., Neuenschwander, A., Thomas, N., Hofton, M., Dubayah, R., 2021. Fusing simulated GEDI, ICESat-2 and NISAR data for regional aboveground biomass mapping. Remote Sens. Environ. 253, 112234.
Silva, C.A., Klauberg, C., Hudak, A.T., Vierling, L.A., Liesenberg, V., Carvalho, S.P.E., Rodriguez, L.C.(2016). A principal component approach for predicting the stem volume in Eucalyptus plantations in Brazil using airborne LiDAR data..
Simmavong, T., Su, Y., Deng, Y., Wang, B., Yao, Z., Wu, J., Lin, L., 2025. Forest structure predicts aboveground biomass better than community-weighted mean of traits, functional diversity, topography, and soil in a tropical forest across spatial scales. For. Ecol. Manag. 578, 122457.
Tamiminia, H., Salehi, B., Mahdianpari, M., Quackenbush, L., Adeli, S., Brisco, B., 2020. Google Earth Engine for geo-big data applications: a meta-analysis and systematic review. ISPRS J. Photogramm. Remote Sens. 164, 152–170.
Tamiminia, H., Salehi, B., Mahdianpari, M., Goulden, T., 2024. State-wide forest canopy height and aboveground biomass map for New York with 10 m resolution, integrating GEDI, Sentinel-1 and Sentinel-2 data. Ecol. inform. 79, 102404.
Tian, X., Su, Z., Chen, E., Li, Z., van der Tol, C., Guo, J., He, Q., 2012. Reprint of: estimation of forest above-ground biomass using multi-parameter remote sensing data over a cold and arid area. Int. J. Appl. earth Obs. Geoinf. 17, 102–110.
Varvia, P., Korhonen, L., Bruguière, A., Toivonen, J., Packalen, P., Maltamo, M., Popescu, S.C., 2022. How to consider the effects of time of day, beam strength, and snow cover in ICESat-2 based estimation of boreal forest biomass? Remote Sens. Environ. 280, 113174.
Varvia, P., Saarela, S., Maltamo, M., Packalen, P., Gobakken, T., Næsset, E., Korhonen, L., 2024. Estimation of boreal forest biomass from ICESat-2 data using hierarchical hybrid inference. Remote Sens. Environ. 311, 114249.
Willmott, C.J., Ackleson, S.G., Davis, R.E., Feddema, J.J., Klink, K.M., Legates, D.R., Rowe, C.M., 1985. Statistics for the evaluation and comparison of models. J. Geophys. Res. Oceans 90 (C5), 8995–9005.
Wolpert, D.H., 1992. Stacked generalization. Neural Netw. 5 (2), 241–259.
Yang, L., Shami, A., 2020. On hyperparameter optimization of machine learning algorithms: theory and practice. Neurocomputing 415, 295–316.