

A COMPARISON OF RANDOM FOREST AND LIGHT GRADIENT BOOSTING MACHINE FOR FOREST ABOVE-GROUND BIOMASS ESTIMATION USING A COMBINATION OF LANDSAT, ALOS PALSAR, AND AIRBORNE LIDAR DATA

Haifa Tamiminia ^{a,*}, Bahram Salehi ^a, Masoud Mahdianpari ^b, Colin M. Beier ^c, Lucas Johnson ^c, and Daniel B. Phoenix ^c

^a Department of Environmental Resources Engineering, State University of New York College of Environmental Science and Forestry (SUNY-ESF), NY 13210, USA - (htamimin, bsalehi)@esf.edu

^b C-CORE and Department of Electrical and Computer Engineering, Memorial University of Newfoundland, St. John's, NL A1B 3X5, Canada - masoud.mahdianpari@c-core.ca

^c Sustainable Resources Management, State University of New York College of Environmental Science and Forestry (SUNY-ESF), NY 13210, USA - (cbeier, ljohns11, dbphoeni)@esf.edu

KEY WORDS: Machine Learning, Above-ground Biomass, Random Forest, Light Gradient Boosting Machine, LiDAR, Synthetic Aperture Radar.

ABSTRACT:

Sustainable forest management is a critical topic which contributes to ecological, economical, and socio-cultural aspect of the environment. Providing accurate AGB maps is of paramount importance for sustainable forest management, carbon accounting, and climate change monitoring. The main goal of this study was to leverage the potential of two machine learning algorithms for predicting AGB using optical and synthetic aperture radar (SAR) datasets. To achieve this goal random forest (RF) and light gradient boosting machine (LightGBM) models were deployed to predict AGB values in Huntington Wild Forest (HWF) in Essex County, NY using continuous forest inventory (CFI) plots. Both models were trained and evaluated based on airborne light detection and ranging (LiDAR) data, Landsat imagery, advanced land observing satellite (ALOS) phased array type L-band Synthetic Aperture Radar (PALSAR), and their combination. The integration of airborne LiDAR, optic, and SAR datasets provided the best results in terms of root mean square error (RMSE) and mean bias error (MBE). The RF model outperformed the LightGBM in all scenarios (LiDAR, Landsat 5, ALOS PALSAR, and their combination). The RF model was able to predict AGB values with the RMSE of 51.90 Mg/ha and MBE of -0.189 Mg/ha for the combination of LiDAR, optic, and SAR data, while LightGBM estimated the AGB values with the RMSE of 52.78 Mg/ha and MBE of -0.253 Mg/ha. LightGBM is more sensitive to noise and there are lots of hyperparameters that need to be tuned which highly affect its performance.

1. INTRODUCTION

In today's world, global deforestation is expanding and accelerating, resulting in the release of a quarter of carbon into the atmosphere (Li, Quackenbush, and Im 2019). Forest monitoring and more accurate estimates of forest above-ground biomass (AGB) is of significance to clarify the contribution of forests in global climate change. Therefore, some countries have developed a campaign which is known as Reducing Emissions from Deforestation and Degradation (REDD) to mitigate the effects of the climate change (Bellassen and Gitz 2008). A key question for REDD effort is how much AGB is available at national and global scale. According to the requirements of this effort, participating countries are supposed to report verified estimates of AGB at national level which is a key indicator of carbon pools in forest systems (Chen et al. 2018). One major problem in accurate carbon estimation is to find an efficient method for the determination of the forest AGB. Although field measurement techniques can estimate AGB accurately, they are inherently destructive, labor-intensive, costly, time-consuming, and practical only for local scale (M. Li, Im, and Beier 2013). The increasing availability of remote sensing data paves the road for cost-effective and large scale AGB estimation.

Recently, light detection and ranging (LiDAR), optic, and synthetic aperture radar (SAR) data have been extensively used as non-destructive and effective methods for forest monitoring (S. Li, Quackenbush, and Im 2019). The invention of light detection and ranging (LiDAR) data has provided an efficient

tool for an accurate AGB estimation by capturing three-dimensional forest structure (Bolton et al. 2020). The cost and volume of LiDAR data limits its application for large scale AGB estimation. Optic and SAR data are considered as other valuable sources of AGB estimation over large areas and with less cost in comparison to airborne LiDAR flight. Spectral bands, vegetation indices, and texture features derived from optical imagery have a high correlation with vegetation density, biomass, chlorophyll content and etc. (Zhou et al. 2016). However, weather conditions such as cloud cover, rain, and snow can greatly affect the quality of the optical imagery, especially in tropical regions and northern climates. In addition, these images suffer from saturation which occurs when the pixels' spectral reflectance values do not show the exact reflectance at high biomass regions (Urbazaev et al. 2018; Zhou et al. 2016). SAR sensors acquire data independent of weather and illumination conditions. SAR signals are sensitive to trees geometric structure and can penetrate through forest canopy depending on the wavelength (Tamiminia et al. 2017). Urbazaev et al. (2018) indicated that SAR data also have limitations with saturation depending on the wavelength and biomass density. Thus, in this study the combination of airborne LiDAR, optic, and SAR data is used to overcome the limitation of single source approaches and to enhance AGB estimation.

Recently, machine learning models are commonly used to develop relationships between forest attributes and remote sensing derived predictors (Zhang et al. 2019). Decision tree-based models (e.g. random forest (RF)) have shown promising results in AGB estimation (Y. Li et al. 2020). Nonetheless, a

comparison between decision-tree based techniques is needed. Y. Li et al. (2019) compared the results of linear regression, RF, and extreme gradient boosting (XGBoost) for AGB estimation in subtropical forest of southern China using spectral bands, vegetation indices, and texture measures derived from Landsat 8 imagery. They reported that XGBoost provides better performance in compared with linear regression and RF. In 2020, Y. Li et al. conducted further investigation on XGBoost tuning parameters for AGB estimation in southern China using the combination of Landsat 8 and Sentinel-1 data. Moreover, Pham et al. (2020) compared and reported that the combination of XGBoost method and genetic algorithm feature selection technique provides better results than CatBoost (CB), RF, gradient boosted regression tree (GBRT), and support vector regression (SVR) algorithms for mangrove AGB estimation using integration of SAR and optical data in Vietnam. Since mentioned studies were conducted in tropical and subtropical forests, in this study we plan to explore the potential of decision tree-based models for a temperate forest.

The main goal of this study is to compare two ensemble machine learning models: random forest (RF) and light gradient boosting machine (LightGBM) for AGB estimation using the integration of remote sensing data. To achieve this goal, the following research objectives are addressed:

- To assess the potential of the synergy of airborne LiDAR, optic (i.e. Landsat 5 Thematic mapper (TM)), and SAR (i.e. advanced land observing satellite (ALOS) phased array type L-band Synthetic Aperture Radar (PALSAR)) data for AGB estimation. The hypothesis is that the combination of remote sensing data increases the performance of AGB modelling in comparison to single data source method.
- To compare RF (a bagging technique) and LightGBM (a boosting technique) models. Our hypothesis is that a properly-tuned LightGBM can perform better than RF model.

2. MATERIALS

2.1 Study Area

The study area is located in Huntington Wildlife Forest (HWF) area, in the central Adirondack Park, northern New York State (Figure 1). HWF covers an approximate area of 6,000 ha (latitude 44° 00' N, longitude 74° 13' W). It has mountainous topography and the elevation ranges from 473 m to 908 m above mean sea level. The mean annual temperature and precipitation are 4.4 Celsius degree and 1010 mm, respectively.

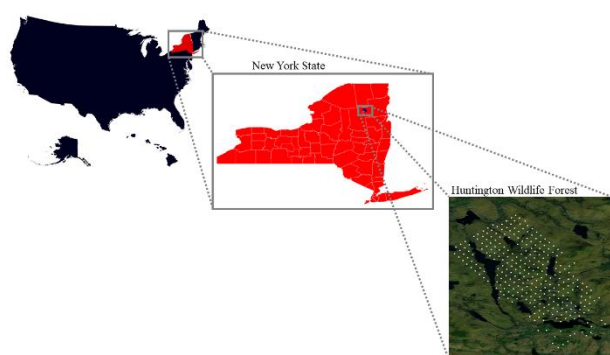


Figure 1. Location of the study area (Essex County, NY) for forest AGB estimation using ensemble machine learning models. White circles depict sample plots located in Huntington wildlife forest

2.2 Field Measurements

Continuous Forest Inventory (CFI) plots have been used as a reference dataset in this study. This dataset was collected by the State University of New York, College of Environmental Science and Forestry (ESF) during July and August of 2011. Plots cover northern hardwood species including sugar maple, red maple, yellow birch, beech, white ash, red oak, white pine, hemlock, red spruce, and pine/softwood plantations of various species (Breitmeyer et al. 2019). CFI data over HWF contains 288 sample plots (Figure 1). Tree information such as tree species, diameter at breast height (DBH) of 11.7 cm or greater, and the relative location to the center of the plot were recorded (S. Li, Quackenbush, and Im 2019). Then, AGB at tree level was calculated using species-specific DBH Component Ratio Method (CRM) allometric equations (Kennedy et al. 2018). Finally, plot level AGB was calculated as the average AGB per unit area within each sample plot. In other words, the plot level AGB in megagrams per hectare (Mg/ha) was calculated by dividing the tree level AGB by the plot area.

2.3 Remote Sensing Data

2.3.1 Airborne LiDAR: Discrete return aerial data collection was acquired over HWF in May 2015 using the Leica Airborne Laser Scanner (ALS70) at a maximum flying height of 3500 above ground level (AGL). This was to support a 2.5 ppm² LiDAR point cloud. First step in LiDAR data processing was to convert the raw point clouds into height-normalized point clouds. A k-nearest neighbour imputation algorithm (k=5) was used to impute a digital elevation model (DEM) which is subtracted from all returns in the point cloud (Hawbaker et al. 2009; Huang et al. 2019). Then, predictors were computed using the height normalized LiDAR data for modelling at 30 m grid cells. Finally, 29 height (i.e. height percentiles, coefficient of variation of height, and etc.) and intensity (i.e. percentage of ground intensity, percentage of feature intensity, and etc.) predictors were fed as inputs into the machine learning models. Since the reference datasets were collected in 2011 and we are using airborne LiDAR for 2015, the main hypothesis is that HWF did not change from 2011 to 2015.

2.3.2 Landsat 5 TM Imagery: Landsat 5 TM surface reflectance (SR) imagery with 30 m spatial resolution were pre-processed and downloaded through Google Earth Engine (GEE) platform. First, the Landsat 5 image collection which covered the HWF were selected for July and August of 2011 with less than 5 percent of cloud cover. A cloud masking function was applied based on the pixel quality assessment (pixel-qa) band of Landsat SR data to mask out the clouds. Then, 6 spectral bands including blue, green, red, near infrared (NIR), shortwave infrared-1 (SWIR1), and shortwave infrared-2 (SWIR2) were selected. Finally, some vegetation indices such as normalized difference vegetation index (NDVI), Soil Adjusted Vegetation Index (SAVI), Ratio Vegetation Index (RVI), normalized burn ratio (NBR), and normalized difference moisture index (NDMI) were calculated based on spectral bands.

2.3.3 ALOS PALSAR Data: For SAR data, the global PALSAR/PALSAR-2 yearly mosaic with 25 m resolution at L-band was utilized. This dataset is freely available at GEE platform (Tamiminia et al. 2020). It should be noted that the strips with less response to surface moisture were selected for this procedure. Then, the imagery was ortho-rectified and slope corrected using the 90 m shuttle radar topography mission (SRTM) digital elevation model (DEM). In this study, the dual polarization (horizontal transmit/horizontal receive (HH) and horizontal transmit/vertical receive (HV) polarizations) yearly mosaic was used for the year 2011 over HWF. Then, a smoothing speckle filter with the radius of 30 m was applied to the channels to remove the speckle noise. Span and ratio were also calculated to add more features to train the models (Equations 1 and 2). The images were resampled to 30 m resolution to be as the same resolution as the LiDAR and optic datasets.

$$\text{Span} = \text{HH}^2 + \text{HV}^2 \quad (1)$$

$$\text{Ratio} = \text{HH}/\text{HV} \quad (2)$$

where HH= horizontal transmit/horizontal receive channel
 HV = horizontal transmit/vertical receive channel

3. METHODS

This section describes two ensemble machine learning models used and compared in this study. Two well-known ensemble techniques are bagging and boosting (Liaw and Wiener 2002). RF and LightGBM are subset of decision tree-based models which use bagging and boosting methods, respectively. The training dataset contained 70% of the data used for tuning and training the model while 30% of data used for the evaluation of the final model. Hyperparameters were tuned using a grid search approach.

3.1 Random Forest (RF)

The first idea of RF was proposed by Ho in 1995, then, an extension of RF that is applied for classification and regression purposes was developed by Breiman (2001). RF is a machine learning algorithm which uses a bagging technique to train the model independently in parallel. The forest is a combination of trees which each tree is trained separately without any dependency on the other trees. Due to the characteristics of RF, the training speed runs faster, the results are less sensitive to tuning parameters, and few parameters are needed to be tune (Li et al. 2019). Moreover, RF creates many trees on subsets of the data both bagged observations and subsets of variables. This is done to increase the difference in the trees in order to improve predictive power, leading to a robust algorithms that can reduce the over-fitting issues (Li et al. 2019). The RF implementation was performed using randomForest package in R software. Grid search winning parameters for the combination of LiDAR, optic, and SAR data are listed in Table 1.

Parameter	Description	Selected Value
ntree	Number of trees to aggregate.	390
mtry	Number of variables to split at in each node.	3
nodesize	Minimum size of terminal nodes.	6

Table 1. Values for tuned hyperparameters for RF model using a grid search approach and combination of LiDAR, optic, and SAR data

3.2 Light Gradient Boosting Machine (LightGBM)

Gradient boosting machine (GBM) is another ensemble-based decision tree, however, unlike the trees in RF model, the trees in GBM method cannot be created in parallel. Alternatively stating, in the GBM algorithm, the second tree depends on the first tree and the third depends on the first two and so on. Boosting technique used in GBM builds new sets in a sequential way and the observations are weighted, leading some take part in the sets (Ke et al. 2017). The main objective of GBM is to reduce the model's residual along the gradient direction by decreasing the previous residuals (Pham et al. 2020). The GBMs have very low interpretability because the second tree in the model no longer predicts the same target as the original model the subsequent trees in the model seek to predict how far off the original predictions were from the truth by using the residuals from the prior trees. In this way, each subsequent tree of the gradient boosting model is slowly reduces the overall error of the previous trees. This enables the gradient boosting models to have high predictive power but low interpretability (Ke et al. 2017). The main advantage of GBM is being robust and being able to handle mixed data types which is useful for remote sensing data. However, it is computationally expensive. In this paper, a LightGBM model which uses a leaf-wise split and runs faster was implemented using the lightgbm package in R. Table 2 shows the winning hyperparameters of LightGBM model using the combination of LiDAR, optic, and SAR data.

Parameter	Description	Selected Value
learning-rate	How fast the algorithm learns.	0.01
nrounds	The number of boosting iterations.	100
num_leaves	The maximum number of leaves per tree.	31
max_depth	Prevents the tree growth from a certain depth.	-1
bagging_frequency	How many iterations to run before resampling.	10
bagging_fraction	The fraction of rows to use per iteration.	0.9
feature_fraction	The proportion of predictors to use in each iteration.	0.6

Table 2. Values for tuned hyperparameters for LightGBM model using a grid search approach and the combination of LiDAR, optic, and SAR data

4. RESULTS AND DISCUSSION

This section describes the results of applied regression models for AGB estimation for 4 different scenarios: airborne LiDAR, Landsat 5, ALOS PALSAR, and the combination of them. RF and LightGBM regression models were applied for each scenario separately. The performance of the models were evaluated using root mean square error (RMSE) and mean bias error (MBE). MBE was calculated based on the difference between the mean values of all observed and predicted values (Ji et al. 2015). Therefore, the positive MBE is a sign of under-prediction and the negative MBE indicates over-prediction.

Figure 2 shows the distribution of computed AGB values for CFI plots using CRM allometric equations. Calculated AGB values

using CRM allometric equations vary from 0 to 433.2 Mg/ha with a median of 183.2 Mg/ha (Table 3).

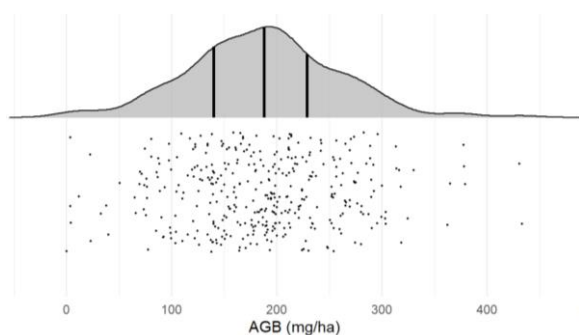


Figure 2. Distribution of calculated AGB using CRM allometric equations. The grey dots represent a cloud of randomly jittered plots below the ridge line plot. The AGB values of sample plots were used to train and test the models based on 70/30 split

Min (Mg/ha)	1 st quartile (Mg/ha)	Median (Mg/ha)	3 rd quartile (Mg/ha)	Max (Mg/ha)
0	139.5	183.2	183.4	433.2

Table 3. Statistics characteristics of AGB of CFI plots in HWF used for AGB estimation using RF and LightGBM models

Table 4 summarizes the RMSE and MBE for 4 scenarios and each regression model. From a machine learning perspective, in all 4 scenarios RF model which is a bootstrap (bagging) technique outperformed the LightGBM model. For instance, the RF model for the ALOS PALSAR-only scenario predicted AGB values with the RMSE of 72.88 Mg/ha and MBE of 5.96 while LightGBM estimated with RMSE of 77.79 Mg/ha and MBE of 11.26 Mg/ha. In this case, both models are over-predicting the AGB, however, the average over-estimation of RF model is nearly 5 Mg/ha lower than the LightGBM method. Since the RF is a parallel ensemble decision tree which trains each tree separately, it leads to reducing variance. This independent training process decreases the sensitivity of the RF model to noise existing in the predictors. On the other hand, LightGBM trains the trees sequentially and aims to reduce the bias. Although LightGBM can overcome over-fitting issues in decision tree algorithms, noisy input variables may decrease the accuracy of the subsequent tree. In this case, due to the inherent characteristic of remote sensing data, noise cannot be ignored. Though over-fitting is still considered as a concern for modelling the AGB values, RF model which is less sensitive to outliers seems to be more effective for AGB prediction in HWF. In airborne LiDAR only scenario, both intensity and height variables of LiDAR data were used as predictors for AGB estimation. It is worth mentioning that LiDAR predictors can be noisy because of low vertical accuracy relative to horizontal sample distance. Furthermore, spectral bands and vegetation indices derived from Landsat imagery contain mixed pixels which is not totally correspond to forested areas. Thus, increasing the chance of noise in the variables. Therefore, with regard to input predictors, the RF model outperformed LightGBM in AGB estimation. In addition, LightGBM is sensitive to hyperparameter tuning. Figure 3 demonstrates the scatter plots of predicted versus actual AGB values for both RF and LightGBM models using the combination of LiDAR, Landsat 5, and ALOS PALSAR data.

Data Type	Validation Metric	RF	Light GBM
LiDAR	RMSE (Mg/ha)	53.51	54.73
	MBE (Mg/ha)	-1.75	-2.85
Landsat 5	RMSE (Mg/ha)	67.11	69.32
	MBE (Mg/ha)	4.65	5.08
ALOS PALSAR	RMSE (Mg/ha)	72.88	77.79
	MBE (Mg/ha)	5.96	11.26
Combination (LiDAR+optic+SAR)	RMSE (Mg/ha)	51.90	52.78
	MBE (Mg/ha)	-0.189	-0.253

Table 4. Results of HWF AGB estimation for RF and LightGBM using airborne LiDAR, Landsat 5, ALOS PALSAR, and the combination of LiDAR, optic, and SAR data

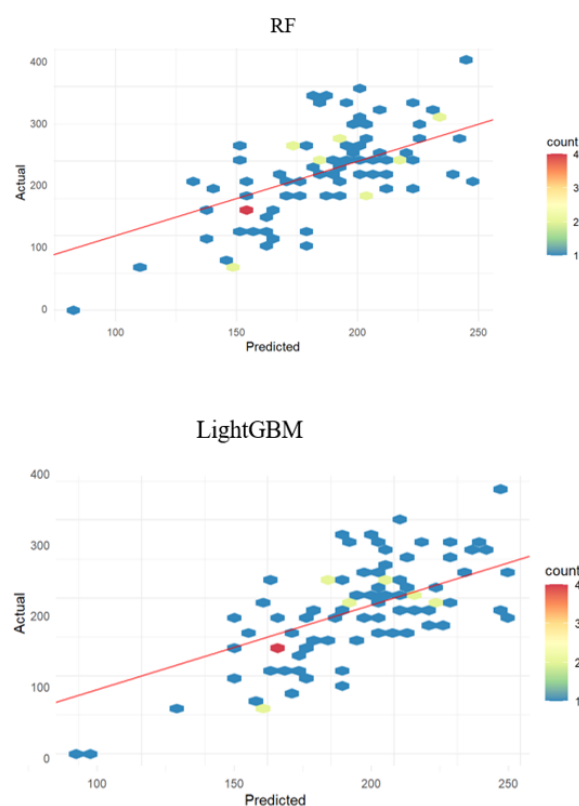


Figure 3. Scatter plots of actual versus predicted AGB for the combination of airborne LiDAR, Landsat 5, and ALOS PALSAR data using RF and LightGBM models

From a data source perspective, the combination of airborne LiDAR, Landsat 5, and ALOS PALSAR data provided the best results. Airborne LiDAR data which provides information about the vertical structure of the trees indicated the second best RMSE and MBE. Unlike LiDAR, both Landsat 5 and ALOS PALSAR did a poor job in predicting AGB values accurately due to the high effect of saturation in areas with high biomass. In general, in Landsat 5 and ALOS PALSAR the AGB values were under-predicted (MBE>0), while in LiDAR and combination scenarios, the AGB values were over-predicted. Residuals of the estimated AGB values for each scenario and RF model is demonstrated in Figure 4. As it is shown, the residuals for the combination scenario are the lowest, followed by airborne LiDAR, Landsat 5, and ALOS PALSAR, respectively. ALOS PALSAR has the

highest residuals especially for plots with near to zero and high biomass. Thus, the integration of vertical structural information of forest provided by LiDAR, spectral information of Landsat, and ALOS PALSAR L-band backscatter information can significantly improve the AGB mapping.

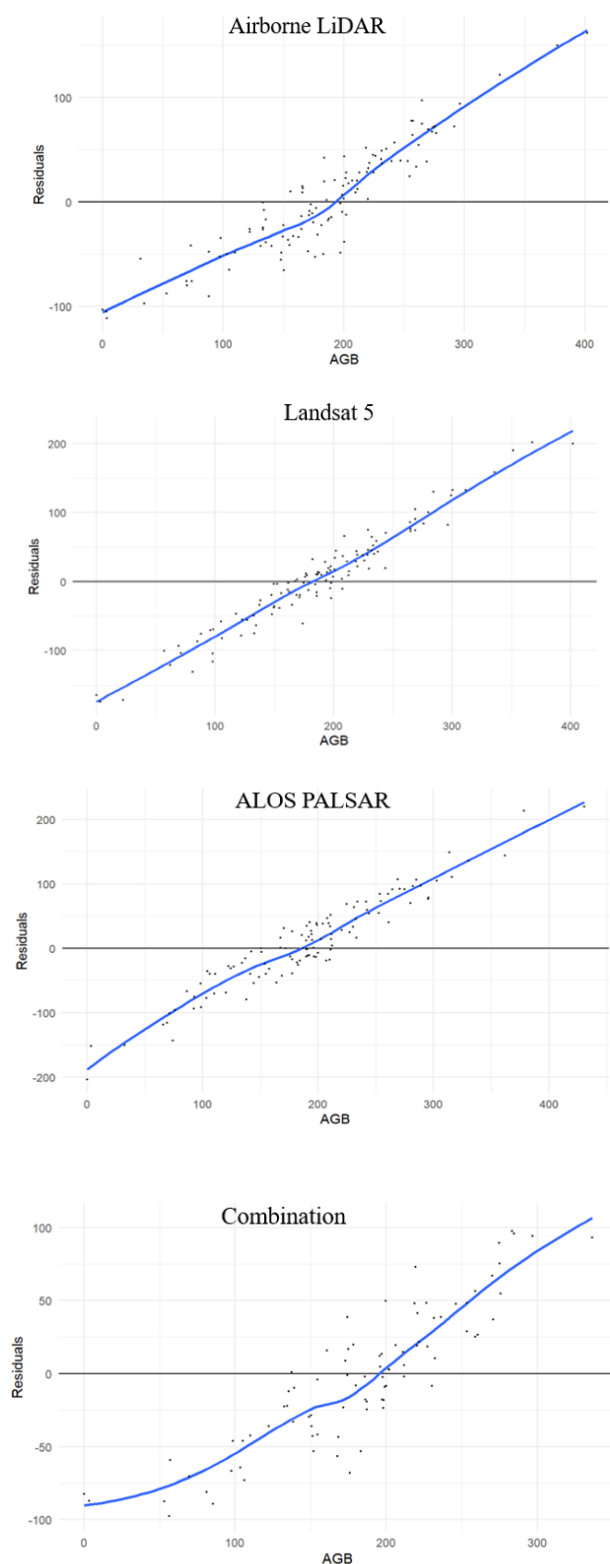


Figure 4. Residuals of predicted AGB values for airborne LiDAR, Landsat 5, ALOS PALSAR, and the their combination using RF regression model

5. CONCLUSION

The main goal of this study was to compare two ensemble machine learning models for AGB estimation using airborne LiDAR, Landsat 5, ALOS PALSAR, and the synergy of these datasets in a temperate forest in NY. In all scenarios, RF regression model outperformed LightGBM. Both RF and LightGBM are capable of handling over-fitting issue. However, LightGBM is more susceptible to noise and hyperparameter tuning which decreases its performance. The combination of airborne LiDAR, optic, and SAR data provided the most accurate AGB estimation with the lowest RMSE and MBE, followed by LiDAR, Landsat 5, and ALOS PALSAR, respectively. It can be concluded that the combination of vertical structure along with spectral and backscatter information of trees enhances the AGB estimation results and reduces the saturation effects. Hyperparameter tuning plays an important role in the performance of machine learning models especially LightGBM. Thus, Bayesian optimization or genetic algorithms might be better options for hyperparameter tuning in the future.

ACKNOWLEDGEMENTS

This project was supported by the Climate & Applied Forest Research Institute at SUNY ESF with funding from the NYS Department of Environmental Conservation.

REFERENCES

- Bellassen, Valentin, and Vincent Gitz. 2008. "Reducing Emissions from Deforestation and Degradation in Cameroon—Assessing Costs and Benefits." *Ecological Economics* 68 (1–2): 336–44. <https://doi.org/10.1016/j.ecolecon.2008.03.015>.
- Berninger, Anna, Sandra Lohberger, Matthias Stängel, and Florian Siegert. 2018. "SAR-Based Estimation of above-Ground Biomass and Its Changes in Tropical Forests of Kalimantan Using L- and C-Band." *Remote Sensing* 10 (6): 831.
- Bolton, Douglas K., Piotr Tompalski, Nicholas C. Coops, Joanne C. White, Michael A. Wulder, Txomin Hermosilla, Martin Queinnec, Joan E. Luther, Olivier R. van Lier, and Richard A. Fournier. 2020. "Optimizing Landsat Time Series Length for Regional Mapping of Lidar-Derived Forest Structure." *Remote Sensing of Environment* 239: 111645.
- Chen, Lin, Chunying Ren, Bai Zhang, Zongming Wang, and Yanbiao Xi. 2018. "Estimation of Forest Above-Ground Biomass by Geographically Weighted Regression and Machine Learning with Sentinel Imagery." *Forests* 9 (10): 582. <https://doi.org/10.3390/f9100582>.
- Hawbaker, Todd J., Nicholas S. Keuler, Adrian A. Lesak, Terje Gobakken, Kirk Contrucci, and Volker C. Radeloff. 2009. "Improved Estimates of Forest Vegetation Structure and Biomass with a LiDAR-Optimized Sampling Design." *Journal of Geophysical Research: Biogeosciences* 114 (G2).
- Ho, Tin Kam. 1995. "Proceedings of the 3rd International Conference on Document Analysis and Recognition." *Random Decision Forests*, 278–82.
- Huang, Jianfeng, Xinchang Zhang, Qinchuan Xin, Ying Sun, and Pengcheng Zhang. 2019. "Automatic Building Extraction from High-Resolution Aerial Images and LiDAR Data Using Gated Residual Refinement Network." *ISPRS Journal of Photogrammetry and Remote Sensing* 151: 91–105.

- Ji, Lei, Bruce K. Wylie, Dana RN Brown, Birgit Peterson, Heather D. Alexander, Michelle C. Mack, Jennifer Rover, Mark P. Waldrop, Jack W. McFarland, and Xuexia Chen. 2015. "Spatially Explicit Estimation of Aboveground Boreal Forest Biomass in the Yukon River Basin, Alaska." *International Journal of Remote Sensing* 36 (4): 939–53.
- 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." *Advances in Neural Information Processing Systems* 30: 3146–54.
- Kennedy, Robert E., Janet Ohmann, Matt Gregory, Heather Roberts, Zhiqiang Yang, David M. Bell, Van Kane, M. Joseph Hughes, Warren B. Cohen, and Scott Powell. 2018. "An Empirical, Integrated Forest Biomass Monitoring System." *Environmental Research Letters* 13 (2): 025004.
- Li, Manqi, Jungho Im, and Colin Beier. 2013. "Machine Learning Approaches for Forest Classification and Change Analysis Using Multi-Temporal Landsat TM Images over Huntington Wildlife Forest." *GIScience & Remote Sensing* 50 (4): 361–84.
- Li, Siqi, Lindi J. Quackenbush, and Jungho Im. 2019. "Airborne Lidar Sampling Strategies to Enhance Forest Aboveground Biomass Estimation from Landsat Imagery." *Remote Sensing* 11 (16): 1906. <https://doi.org/10.3390/rs11161906>.
- Li, Yingchang, Chao Li, Mingyang Li, and Zhenzhen Liu. 2019. "Influence of Variable Selection and Forest Type on Forest Aboveground Biomass Estimation Using Machine Learning Algorithms." *Forests* 10 (12): 1073.
- Li, Yingchang, Mingyang Li, Chao Li, and Zhenzhen Liu. 2020. "Forest Aboveground Biomass Estimation Using Landsat 8 and Sentinel-1A Data with Machine Learning Algorithms." *Scientific Reports* 10 (1): 1–12.
- Liaw, Andy, and Matthew Wiener. 2002. "Classification and Regression by RandomForest." *R News* 2 (3): 18–22.
- Pham, Tien Dat, Naoto Yokoya, Junshi Xia, Nam Thang Ha, Nga Nhu Le, Thi Thu Trang Nguyen, Thi Huong Dao, Thuy Thi Phuong Vu, Tien Duc Pham, and Wataru Takeuchi. 2020. "Comparison of Machine Learning Methods for Estimating Mangrove Above-Ground Biomass Using Multiple Source Remote Sensing Data in the Red River Delta Biosphere Reserve, Vietnam." *Remote Sensing* 12 (8): 1334.
- Tamiminia, Haifa, Saeid Homayouni, Heather McNairn, and Abdoreza Safari. 2017. "A Particle Swarm Optimized Kernel-Based Clustering Method for Crop Mapping from Multi-Temporal Polarimetric L-Band SAR Observations." *International Journal of Applied Earth Observation and Geoinformation* 58: 201–12.
- Tamiminia, Haifa, Bahram Salehi, Masoud Mahdianpari, Lindi Quackenbush, Sarina Adeli, and Brian Brisco. 2020. "Google Earth Engine for Geo-Big Data Applications: A Meta-Analysis and Systematic Review." *ISPRS Journal of Photogrammetry and Remote Sensing* 164: 152–70.
- Urbazaev, Mikhail, Christian Thiel, Felix Cremer, Ralph Dubayah, Mirco Migliavacca, Markus Reichstein, and Christiane Schmullius. 2018. "Estimation of Forest Aboveground Biomass and Uncertainties by Integration of Field Measurements, Airborne LiDAR, and SAR and Optical Satellite Data in Mexico." *Carbon Balance and Management* 13 (1): 1–20.
- Zhang, Linjing, Zhenfeng Shao, Jianchen Liu, and Qimin Cheng. 2019. "Deep Learning Based Retrieval of Forest Aboveground Biomass from Combined LiDAR and Landsat 8 Data." *Remote Sensing* 11 (12): 1459.