Preprints.org

Article

Predicting Above-Ground Biomass of Forest in South Carolina: Integrating Remote Sensing, Machine Learning, and Interpolation Techniques

Sanjeev Sharma [*] and Puskar Khanal

Article

Predicting Above-Ground Biomass of Forest in South Carolina: Integrating Remote Sensing, Machine Learning, and Interpolation Techniques

Sanjeev Sharma * and Puskar Khanal

Department of Forestry and Environmental Conservation, College of Agriculture, Forestry and Life Sciences, Clemson University, Clemson, SC 29634, USA

* Correspondence: sanjeev@clemson.edu

Abstracts: This study evaluates the effectiveness of a Random Forest model for predicting above-ground biomass in South Carolina (SC), utilizing diverse remote sensing and climatic data sources. SC, with its humid subtropical climate and varied geography, including the Atlantic coastal plain, Piedmont, and Blue Ridge Mountains, poses unique challenges for biomass estimation. We integrated global biomass datasets for 2010, MODIS vegetation indices (NDVI and EVI), Leaf Area Index (LAI) from MOD15A2H, and climate data from TerraClimate. The model was trained using 2010 data and applied to 2022 datasets to assess biomass changes. To validate the model, plot-level biomass estimates from the 2023 FIA data were interpolated using Inverse Distance Weighting (IDW). Performance evaluation showed a strong positive correlation between predicted and observed biomass, with a correlation coefficient of 0.77 and an R^2 value of 0.62, indicating that the model explains 62% of the variability in biomass. Comparison with IDW-interpolated biomass data resulted in a correlation coefficient of 0.64, confirming the model's validity. Although the Random Forest model demonstrated reliable predictions, the study suggests potential improvements by incorporating additional data sources and advanced modeling techniques. The findings emphasize the value of integrating remote sensing data, machine learning, and interpolation methods to enhance biomass estimation accuracy. This research provides crucial insights into biomass distribution in SC and establishes a basis for future studies on forest monitoring and carbon accounting, highlighting the importance of combining various data sources for comprehensive environmental analysis.

Keywords: machine learning; random forest; inverse distance weighting; forest vegetation simulation; google earth engine

1. Introduction

Forests, covering 31% of Earth's surface, are fundamental to global ecological balance, biological evolution, and community succession [1,2]. They play a crucial role in the terrestrial carbon cycle, which includes five primary carbon pools: above-ground biomass, below-ground biomass, dead litter, woody debris, and soil organic matter [3]. Among these, above-ground biomass (AGB) is a major carbon reservoir and is particularly sensitive to land-use changes such as forest degradation and deforestation [2,4]. Thus, precise assessment of AGB is essential for applications such as timber extraction, monitoring carbon stocks, and understanding the global carbon cycle [5,6].

Traditional methods for estimating AGB rely on in-situ measurements conducted by regional or national forest authorities. Although these methods provide high accuracy, they are often localized, costly, and inefficient for extensive forest areas [5,6]. To address these limitations, Earth observation (EO) and remote sensing technologies have emerged as valuable tools for large-scale AGB estimation.

Airborne LiDAR (Light Detection and Ranging) is recognized for its precision in modeling AGB, offering detailed forest parameter predictions [7–10]. However, the high cost and limited availability

of LiDAR data constrain its application to specific regions [11–15]. Consequently, there is a growing need for more economically viable and widely accessible data sources for broader regional or global-scale AGB modeling. Recent advancements in spaceborne remote sensing have significantly increased the availability of data sources and their temporal frequency [16]. This expansion has spurred interest in utilizing multi-sensor and multitemporal spaceborne data to estimate and map forest attributes and monitor changes in forest cover [17]. Compared to traditional methods, these technologies offer a more scalable and accessible approach to forest monitoring.

In the realm of remote sensing, radar and optical imagery are both crucial for AGB estimation. Radar technologies, such as the Sentinel-1 synthetic aperture radar (SAR) from the European Space Agency, provide high-resolution data that can penetrate clouds and dense canopies. However, their effectiveness can be impacted by factors such as topography and vegetation density [18–20]. Optical imagery from sources like Landsat and Sentinel-2 (S2) offers valuable data, with Sentinel-2 providing higher resolution and additional spectral bands that enhance sensitivity to vegetation and improve AGB estimation accuracy [20–23]. Despite these advantages, relying solely on vegetation indices (VIs) can be problematic due to issues like canopy shadows and structural variability [24]. Incorporating texture measures and terrain factors, such as elevation and slope, can mitigate these challenges and further refine biomass estimates [25–27].

Selecting appropriate regression models is also critical for accurate AGB estimation. Non-parametric methods, including artificial neural networks, support vector machines, and random forests (RF), are commonly used in heterogeneous areas [28,29]. While RF models are effective, they face challenges such as feature selection bias and an inability to fully capture the distribution shape of dependent variables [27,30,31]. Advanced models like Regularized Random Forest (RRF) and Quantile Regression Forest (QRF) address these issues by offering more nuanced predictions and reducing model complexity [32–34].

The Google Earth Engine (GEE version 7.3) (https://earthengine.google.com, accessed on 23 April 2024), with its extensive access to satellite and geospatial data, supports large-scale analyses through robust computing resources and a user-friendly interface. GEE has become a valuable tool in various research fields, including land cover classification and ecosystem monitoring [35]. In this study the already prepared 300m spatial resolution biomass raster was used to develop a model and then utilized to predict the future biomass. Biomass assessment was not done for the state of South Carolina (SC). It will help to predict the carbon, help to develop next industries to minimize the transportation cost, and to apply management as to prevent for future fire hazard. The SC is facing lots of fire hazard. Therefore, to address this issue this study aims to leverage GEE to 1) identify key factors for predicting above-ground biomass, and 2) implement machine learning models to enhance AGB predictions. The biomass assessment is needed for the state of SC to prioritize the area from the carbon loss due to fire. [35] identified the fire hazard map, so with this study it will prioritize the area based on the biomass to mitigate the carbon loss and help to build the potential timber industries based on the biomass hotspot.

2. Methods

2.1. Study Area

This study was conducted in SC, U.S.A. (Figure 1). SC is the 40th-largest and 23rd-most populous U.S.A. state, with a recorded population of 5,124,712 according to the 2020 census. In 2019, its GDP was USD 213.45 billion. SC is composed of 46 counties. Within SC, from east to west, there are three main geographic regions, i.e., the Atlantic coastal plain, the Piedmont, and the Blue Ridge Mountains in the northwestern corner of upstate SC. SC has primarily a humid subtropical climate, with hot, humid summers and mild winters. Areas in the Upstate have a subtropical highland climate. Along SC eastern coastal plain, there are many salt marshes and estuaries. SC southeastern Lowcountry contains portions of the Sea Islands, a chain of barrier islands along the Atlantic Ocean. In the summer, SC is hot and humid, with daytime temperatures averaging between 30 and 34 °C in most of the state, and overnight lows averaging 21–24 °C on the coast and 19–23 °C inland. Winter

temperatures are much less uniform in SC. The coastal areas of the state have very mild winters, with high temperatures approaching an average of 16 °C and overnight lows around 4 °C.
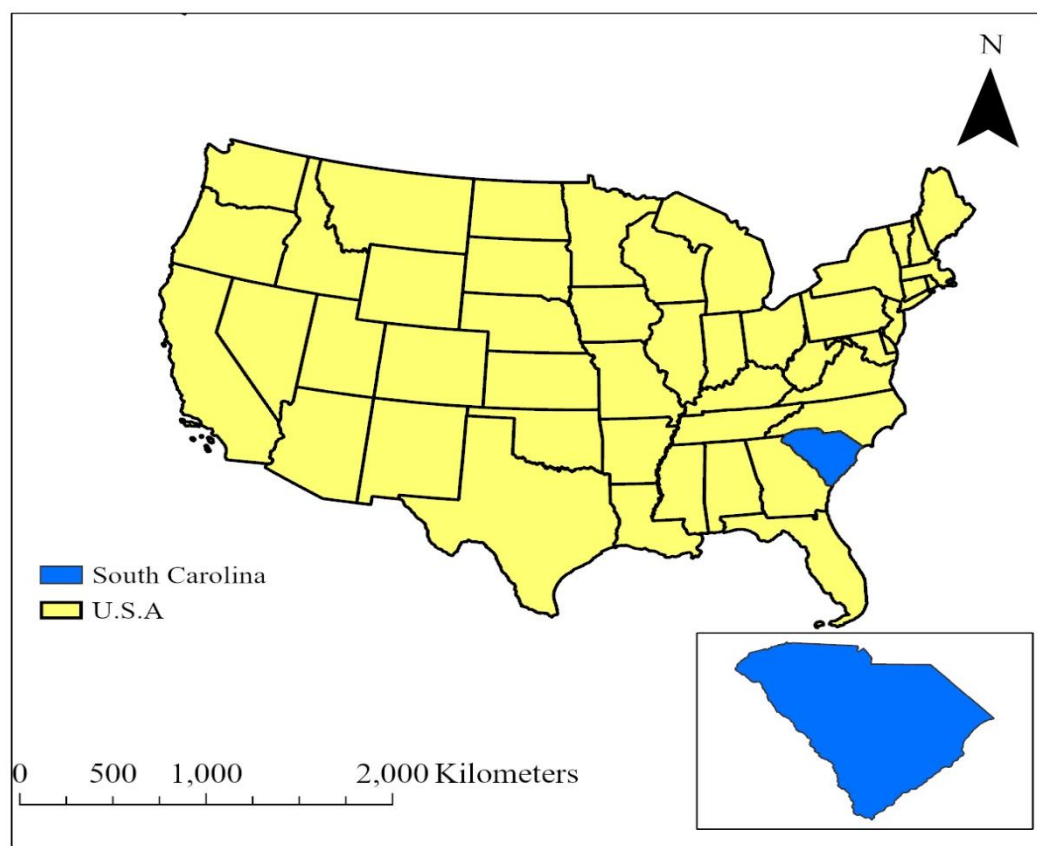

Figure 1. U.S.A. map and SC state map as the study area.

2.2. Data Collection

The Global Aboveground and Belowground Biomass Carbon Density Maps for the Year 2010: This dataset offers global maps of aboveground and belowground biomass carbon density for the year 2010, with a 300-meter spatial resolution. The aboveground biomass map combines data from various sources, including land-cover specific maps for forests, grasslands, croplands, and tundra. These maps were compiled from published sources. The belowground biomass map is similarly derived, using models that match each aboveground biomass map. Both maps were then integrated using tree cover and landcover data through a decision tree method. The dataset also includes maps that show the uncertainty in the biomass estimates for each pixel.

Normalized Difference Vegetative Index (NDVI) and Enhanced Vegetative Index (EVI): The Terra MODIS Vegetation Indices 16-Day (MOD13A1) Version 6.1 product provides vegetation index values for each pixel at a 500-meter spatial resolution. It includes two main vegetation layers: NDVI, which continues the NDVI data from NOAA-AVHRR, and the Enhanced Vegetation Index (EVI), which is more sensitive in areas with high biomass. The product selects the best pixel value over a 16-day period based on criteria like low cloud cover, low view angle, and the highest NDVI/EVI value. Additionally, the dataset includes reflectance bands (red, near-infrared, blue, mid-infrared) and quality assurance layers.

Leaf Area Index (LAI): The MOD15A2H dataset from MODIS provides 8-day composite imagery with a 500-meter resolution, capturing vegetation data from the Terra satellite. MOD15A2H dataset includes LAI, which measures leaf area per unit ground area, and Fraction of Photosynthetically Active Radiation (FPAR), indicating the fraction of sunlight absorbed by vegetation. The MOD15A2H dataset also features quality layers and standard deviation for both LAI and FPAR, along with lower

resolution browse images for quick visualization. MOD15A2H dataset data is valuable for monitoring vegetation health and coverage.

Landcover: The Terra and Aqua combined MODIS Land Cover Type (MCD12Q1) Version 6.1 product provides global land cover data each year. Landcover is created using supervised classifications of MODIS reflectance data from Terra and Aqua satellites. The land cover types are based on several classification schemes, including (International Geosphere-Biosphere Program) IGBP, (University of Maryland) UMD, (Leaf Area Index) LAI, (Biome-Biogeochemical Cycles) BIOME-BGC, and (Plant Functional Types) PFT. After classification, the data undergo additional processing to refine the land cover classes using prior knowledge and extra information. The product also includes land cover properties assessed by the (Food and Agriculture) FAO Land Cover Classification System (LCCS), covering aspects like land use and surface hydrology. Each data file contains layers for Land Cover Types 1-5, Land Cover Properties 1-3, Property Assessments 1-3, Quality Control, and a Land Water Mask.

Climate: TerraClimate is a high-resolution dataset offering monthly climate and water balance data for global land surfaces from 1958 to 2020, with updates made annually. It combines detailed climatological data from WorldClim with time-varying data from CRU Ts4.0 and JRA55 to create a comprehensive dataset at a ~4 km spatial resolution. This dataset is particularly useful for ecological and hydrological studies requiring accurate, time-varying inputs. TerraClimate includes temperature, precipitation, vapor pressure, solar radiation, and wind speed data, as well as monthly surface water balance metrics using a model that incorporates evapotranspiration, precipitation, and soil water capacity.

TerraClimate has been validated against several station-based observations, showing strong accuracy and improved spatial detail compared to its parent datasets. It also offers future climate projections for scenarios with global temperatures 2°C and 4°C above pre-industrial levels, using data from 23 global climate models. However, the dataset inherits long-term trends from its parent data, may not capture finer temporal variability in complex terrains, and uses a simple water balance model that doesn't account for changes in vegetation types. Validation is also limited in data-sparse regions like Antarctica, and there may be unrealistic extrapolations in some high-elevation areas. All these datasets (Figures 2 and 4) were used for model building.
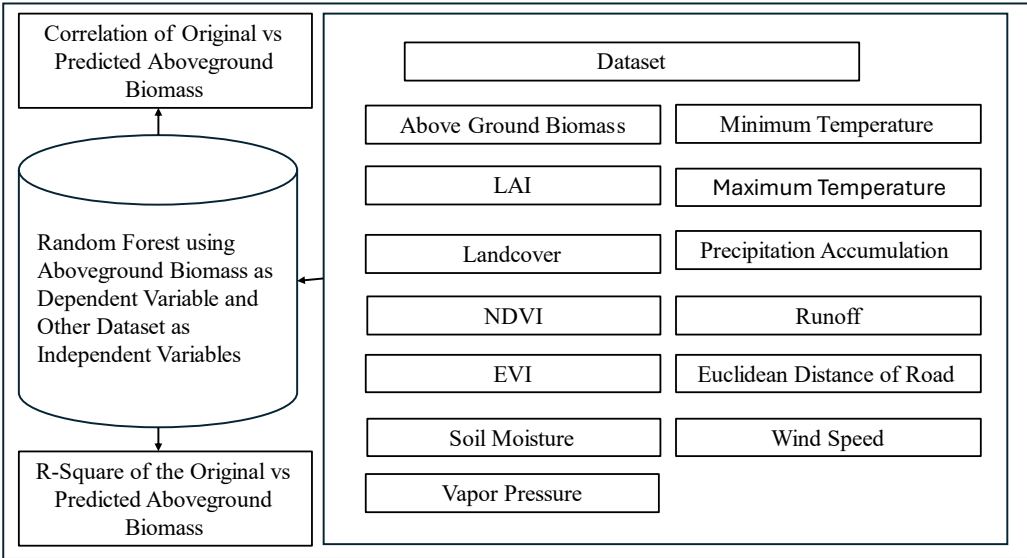

Figure 2. General workflow from data preparation to final predictions.

2.3. Data Analysis

The Forest Vegetation Simulator (FVS accessed on 02 May 2024): It is a sophisticated model used extensively in the United States to predict forest growth, disturbances, and the impacts of various management practices at the stand level. By incorporating forest inventory and analysis (FIA) data

and allowing for self-calibration, FVS provides insights into forest development, biomass, carbon, wildfire risk, and pest management over time. The FIA data (Figure 3) for the year of 2022 was imported in the FVS and then simulation was done for above ground biomass estimation. The biomass obtained from the FVS was later added to each plot manually in ArcGIS pro to develop interpolated biomass raster.
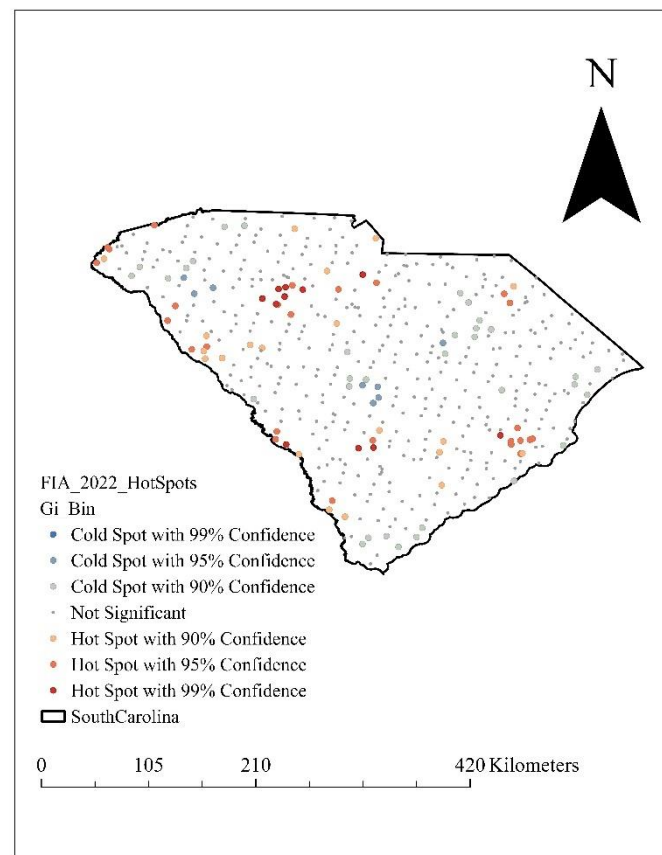

Figure 3. The Hotspot of the aboveground biomass obtained from the FVS simulation with the 495 FIA data.

Inverse Distance Weighting (IDW): It is a spatial interpolation technique used in ArcGIS Pro version 3.2.2 to estimate values at unsampled locations based on the values of nearby sampled points. The method assumes that points closer to the location of interest have a greater influence on the estimated value than those farther away. IDW assigns weights to each sampled point inversely proportional to their distance from the location being estimated, resulting in a smooth surface that reflects the spatial distribution of the data. It is particularly useful in fields like environmental science, agriculture, and meteorology, where understanding spatial patterns is crucial.

Terra Climate data provided variables such as minimum and maximum temperatures, vapor pressure, soil moisture, precipitation accumulation, runoff, and wind speed. Vegetation indices like NDVI and EVI were derived from the MODIS Vegetation Indices 16-Day (MOD13A1) dataset, while Leaf Area Index (LAI) and Fraction of Photosynthetically Active Radiation (FPAR) were obtained from the MOD15A2H product. Land cover data was extracted from the combined Terra and Aqua MODIS Land Cover Type dataset, and global biomass data contributed to the above-ground biomass metrics. The vegetative data were used based on the previous literature review however the past studies found that incorporation of the climatic data increase the accuracy of biomass prediction.
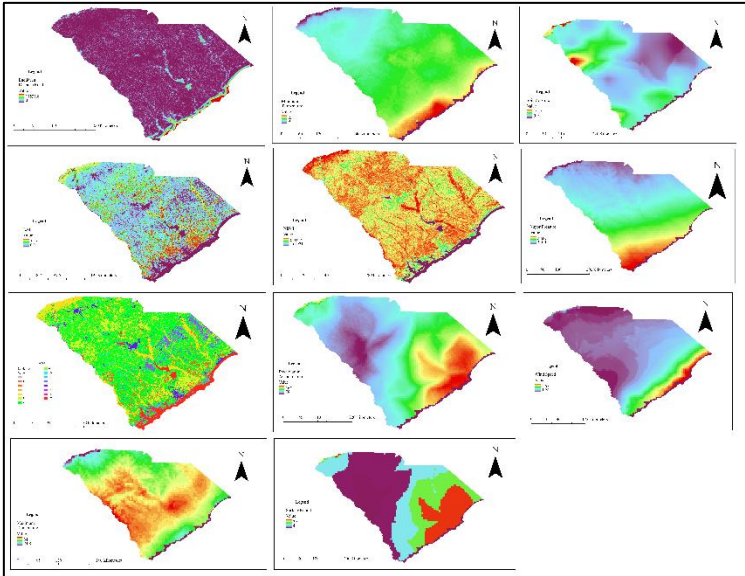
Figure 4. Dataset used for model development and prediction.

To ensure consistency, all datasets were multiplied by their respective scale factors for model building. The temporal focus was set from May to September 2010 to minimize the effects of the fall season, enhancing the model's predictive power. A shapefile for SC was imported in GEE from Tiger/Line (accessed on 02 May 2024) shapefile website (2023 TIGER/Line® Shapefiles (census.gov)), and a Random Forest model was trained using 80 points at a 500m scale. Large area was utilized, so 500m scale was given for less processing time. The model was built using the smile random forest code in Earth Engine, and the correlation between the actual and predicted above-ground biomass was analyzed to assess model accuracy at 10000 scale. This comprehensive approach leveraged multiple data sources and advanced modeling techniques to improve predictions of biomass and other key environmental variables [36–39].

Normalized Difference Vegetation Index (NDVI)

$$NDVI = (NIR-RED) / (NIR+RED) \qquad 1$$

Enhanced Vegetation Index (EVI)

$$EVI = 2.5 \times (NIR-RED)/ (NIR+6 \times RED-7.5 \times BLUE+1) \qquad 2$$

Leaf Area Index (LAI)

$$LAI = -\ln(1-NDVI)/k \qquad 3$$

Fraction of Photosynthetically Active Radiation (FPAR)

$$FPAR= (LAI \times E \times S)/ (S+LAI) \qquad 4$$

The model, originally developed with 2010 data, was applied to the 2022 dataset for SC using GEE to predict above-ground biomass. This prediction utilized a range of environmental variables and vegetation indices, providing a comprehensive analysis for the 2022 scenario. In addition, the FIA data from 2023 for SC, obtained from the United States Forest Service, were imported into the FVS. The resulting plot-level biomass estimates were interpolated across the state using IDW to provide an overall biomass distribution for SC. The predicted biomass values were compared with biomass estimates generated by the Inverse Distance Weighting (IDW) to improve spatial accuracy.

To evaluate the model's performance, a correlation analysis was conducted between the biomass predicted by the model and the biomass obtained from interpolation using FIA data. The inbuilt python in ArcGIS pro was used to test the correlation of two raster. The arcpy library was imported and then analysis was done. Each raster was converted to the same projection and the scale to see the correlation of the each overlapping pixel. This analysis was crucial for assessing the model's accuracy, validation and identifying areas for potential improvement. This integrated approach, combining advanced remote sensing, machine learning, and interpolation techniques, ensured a precise and thorough analysis of biomass distribution in the region.

Random Forest (RF) is a powerful non-parametric machine learning algorithm widely used for both regression and classification tasks. In regression, RF constructs many simple decision trees using subsets of independent variables, such as point cloud-derived metrics, to estimate the dependent variable, such as above-ground biomass (AGB). One of the key advantages of RF regression is that it does not require the assumption of normally distributed data, making it highly adaptable to complex, non-linear relationships, such as those between LiDAR metrics and forest biomass. This machine learning tool enhances predictive accuracy through bootstrap aggregation, commonly known as bagging. The RF model is governed by two primary parameters: the number of predictor variables (Mtry) and the number of decision trees (Ntree). The Mtry parameter, which represents the number of randomly selected variables at each tree node, was automatically optimized in the modeling process. The Ntree parameter, which dictates the number of trees grown in the model, was set to 500, ensuring robust model performance. In this paper, the RF algorithm was used to predict AGB from global biomass data and other vegetative and climatic variables.

Model Evaluation of the Spatial Prediction Methods for Above-ground Biomass: In this study, geostatistical model was developed and applied for biomass prediction. It is necessary to validate the model and to check the efficiency. We applied cross-validation to assess the performance in estimating the spatial distribution of biomass. The basic statistics such as mean error (ME), average standard error and the root mean squared error (RMSE) shown in Eqs. 5 and 6 were applied to check the efficiency of model in spatial predictions of biomass in this region. These statistical methods measure the difference between the known data and the predicted data, thus assess the performance and accuracy of model predictions [40].

$$\text{ME} = \frac{1}{N}\sum_{i=1}^{N}(e\char94_i - \boldsymbol{e_i}) \qquad\qquad 5$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(e\char94_i - \boldsymbol{e_i})^2} \qquad\qquad 6$$

where N is size of the sample in the dataset, $e\char94_i$ is the forecast estimated biomass value, $\boldsymbol{e}_i$ is the biomass values used for prediction.

3. Results

To evaluate the performance of the model, a correlation analysis and calculation of the coefficient of determination (R^2) were conducted between the predicted above-ground biomass and the original, observed above-ground biomass. R^2 a statistical measure that explains the proportion of the variance in the dependent variable that is predictable from the independent variable(s) in a model. It provides an indication of how well the model's predictions match the actual data. The correlation coefficient, which measures the strength and direction of the linear relationship between the predicted and actual biomass, was found to be 0.77 at 10000 scale. This indicates a strong positive relationship, suggesting that the model's predictions are closely aligned with the observed data. The feature importance (Figure 5) revealed that the NDVI has the highest importance followed by EVI. The factors having lower importance were dropped for model development.

Additionally, the R^2 value was calculated to be 0.62. The R^2 value, or coefficient of determination, represents the proportion of variance in the observed biomass that can be explained by the model's predictions. An R^2 of 0.62 implies that 62% of the variability in the actual above-ground biomass is accounted for by the model, while the remaining 38% is due to factors not captured by the model or inherent variability in the data. These results (Figure 6) demonstrate that the model provides a reasonably accurate prediction of above-ground biomass, capturing a significant portion of the variance while also highlighting areas where the model could potentially be refined to improve predictive accuracy further. The mean error should ideally be zero if the interpolation method is unbiased. The root mean square standardized prediction error was 12.56 and mean error was found -7.22.
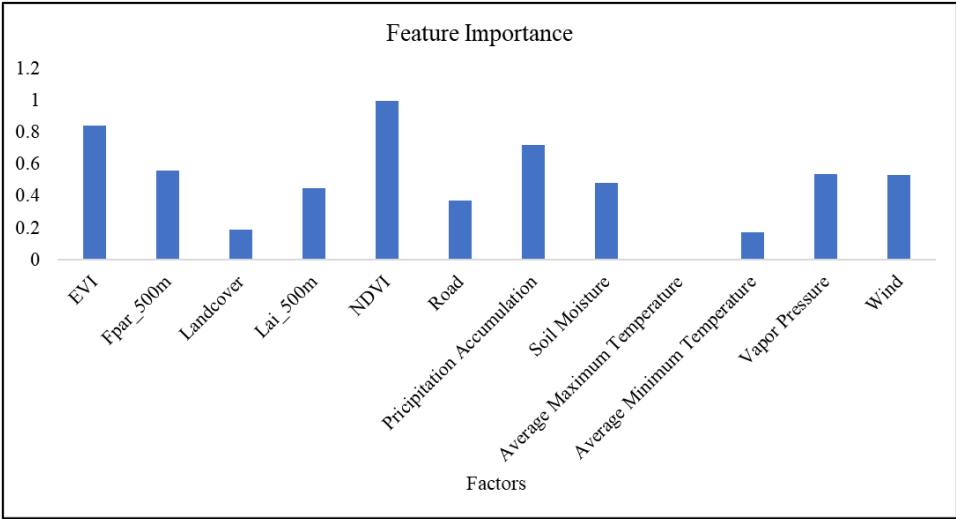
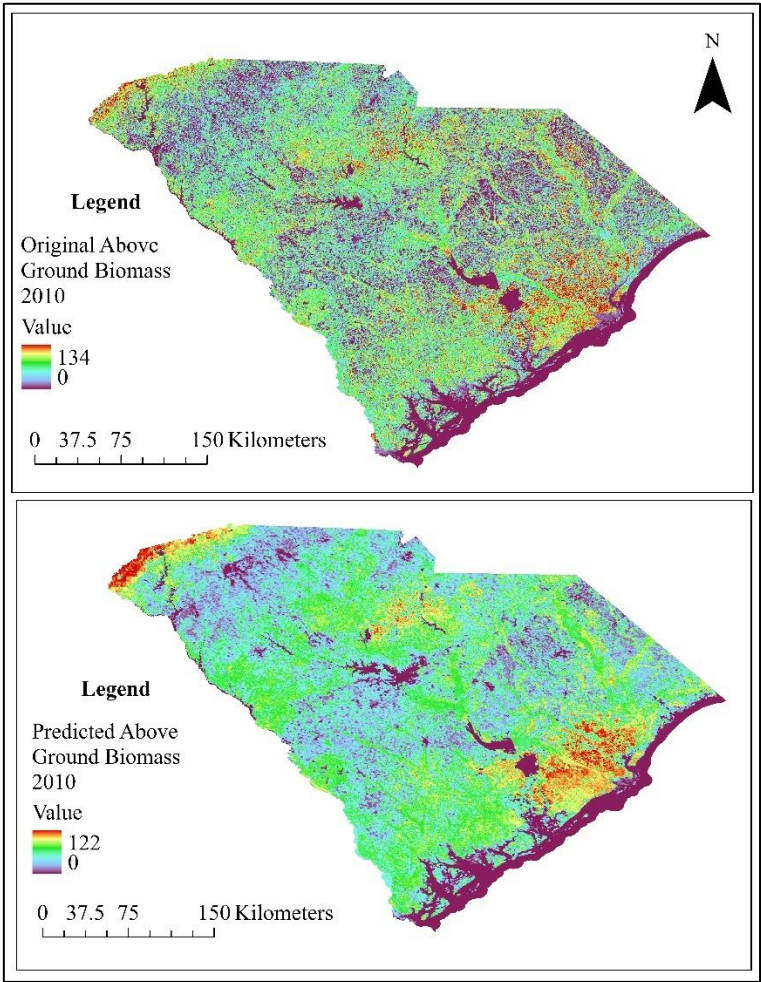**Figure 5.** Feature importance of different factors used for prediction of the above ground biomass.

Figure 6. The observed vs predicted above ground biomass from 2010 data.

The biomass prediction utilized 2022 data, with comparisons made between the predicted above-ground biomass from a Random Forest model and an IDW map derived from FIA plot data using the Forest Vegetation Simulator (FVS). To ensure accurate comparison, both raster (Figure 6) was resampled and aligned to the same projection, resulting in a correlation coefficient of 0.64. While the interpolation of plot data across the entire state was limited by the relatively small sample size of 495 plots, the observed positive correlation supports the validity of the prediction. In a related study,

[35] employed neural networks and correlation coefficients to predict fire hotspots. Prioritizing the overlapping areas to both biomass or carbon and fire regions can significantly aid in managing and mitigating carbon loss.
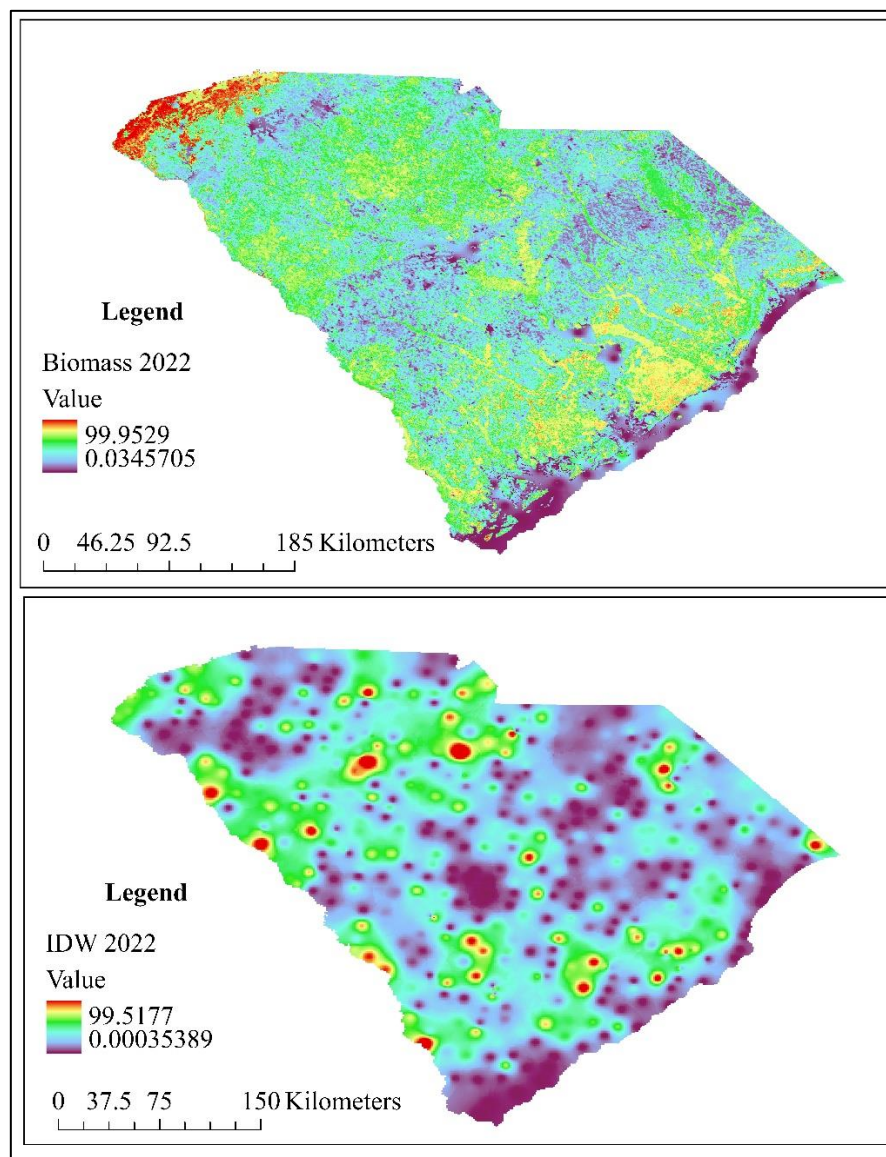

Figure 7. Map of the predicted biomass from the model developed by the 2010 data and the interpolated biomass map obtained after the simulation of FIA data using FVS.

4. Discussion

Climate change and extreme events, such as changes in precipitation and increased drought frequency, significantly impact the prevalence and structural characteristics of forests. Past studies used regression equations, both linear and nonlinear, are commonly used to estimate biomass in various plantings [41], nonlinear was used for this study. The accuracy of different models can vary, with Random Forest models often providing comparable or superior performance in some cases [41].. The techniques and methodologies employed in this study are well-established in the fields of remote sensing and machine learning. We effectively integrated global above ground biomass data with machine learning algorithms, focusing on vegetation indices, climatic, landcover and Euclidean raster to capture key characteristics of forest biomass with high accuracy. The Random Forest (RF) model showed outstanding predictive performance, with an R^2 value of 0.62, which is considered okay for AGB estimation. Similarly, [42] demonstrated that combining multiple remote sensing data sources, such as LiDAR, optical, and SAR, with tree-based machine learning models, achieved an R^2

value of 0.81 for AGB estimation. However, their use of object-based image analysis (OBIA) outperformed pixel-based methods in terms of accuracy. In a related study, [43] reported that gradient boosting decision tree (GBDT) models, when combined with spectral indices, achieved an R^2 of 0.99 in mixed-species forests, indicating a slight improvement in accuracy with multi-source data and advanced machine learning techniques. Our predictor variables were carefully selected such as vegetation indices (e.g., NDVI, LAI, FPAR, EVI), climatic variables (e.g., minimum temperature, soil moisture, vapor pressure, precipitation accumulation, wind speed, surface runoff, maximum temperature), Euclidean distance of road raster and landcover effective for assessing vegetation health and biomass. By selecting these relevant variables, we enhanced the model's predictive capability without needing a vast array of remote sensing data. This approach highlights that a thoughtfully chosen to set of predictors can produce highly accurate results with a limited number of variables.

The incorporation of climatic variables for biomass prediction increased the R^2. Elevation, slope, and the topographic wetness index (TWI) serve as proxy indicators in models predicting aboveground biomass (AGB), capturing various biophysical and biological processes that affect vegetation growth and distribution patterns [44]. By integrating these topographic variables, researchers can better understand the complex interactions between topography, climate, and vegetation [45] resulting in more accurate and ecologically meaningful predictions of AGB across different landscapes [46]. However, in our case the incorporation of the topographic factors predicted less. The incorporation of the Euclidean distance of roads increased the model accuracy. The Topographic Wetness Index (TWI), which serves as an important indicator of soil-water retention crucial for vegetation health, has been identified as a significant factor for estimating aboveground biomass (AGB) in forest ecosystems with nonlinear regression models [47]. However, did not work well in this study. Our research carefully examined the process of selecting predictors for accurately mapping aboveground biomass (AGB) in forested regions. The study effectively pinpointed the most influential predictors for this purpose, which were categorized into four distinct variable groups: climatic, vegetative indices, landcover and Euclidean distance. The Random Forest (RF) method proved to be highly effective as a modeling approach across all variable groups, highlighting its strong capability for forest AGB prediction. This preference for the RF algorithm is consistent with previous findings reported in the academic literature [48–51].

This study used the already predicted 300m spatial resolution biomass, used that to build the model and predict the 2022 biomass. The vegetation indices were used in previous studies, but incorporation of climatic condition gave better accuracy. The feature importance analysis revealed that the NDVI gave the highest importance compared to other variables used to study which aligned to the past studies. [52,53] used Sentinel-2 imagery with an RMSE of 40.16 t/ha for estimating pine forest biomass, while [54] found multiple linear regression to be more accurate for single-storied forests compared to Random Forest when using Sentinel-2 data.

Recent advancements in machine learning have shown that incorporating various data sources and methods can enhance biomass estimation accuracy. Research by [41] demonstrated the effectiveness of Random Forest, Support Vector Regression (SVR), and Artificial Neural Networks (ANN) for biomass estimation. Adding Synthetic Aperture Radar (SAR) data from Sentinel-1 to SVR models improved accuracy, with significant reductions in root mean square deviation for biomass and stand height estimates [55]. Our study got the root mean standardized error 12.46, however getting data from better spatial resolution sensor and combining with the ground truth data, might lead to better accuracy. Combining remote sensing data from various satellites with field data has also been shown to strongly correlate with aboveground biomass, underscoring the value of integrating multiple data sources and modeling techniques [56,57]. This study utilized the only vegetative indices, landcover, Euclidean raster of road and the climatic factors for development and prediction. Lack of ground data for model development might be the reason for less R^2. Similarly, the larger extent of the study area might be another reason for less accurate as some studies in past predicted better for small areas with ground data. The predicted model was later compared with the interpolated raster obtained from IDW for validation. This study helps to predict the visual biomass

change from 2010 to 2022, so prioritization can be done in areas for management to prevent future hazard.

5. Conclusions

This study demonstrates the effectiveness of integrating various data sources and advanced modeling techniques to predict above-ground biomass in SC. The Random Forest model, leveraging 2010 data and applied to 2022 datasets, achieved a strong correlation coefficient of 0.77 and an R^2 value of 0.62, indicating a robust predictive capability that captures substantial variability in biomass data. By incorporating global biomass maps, MODIS vegetation indices, Leaf Area Index (LAI), and TerraClimate data, the model provided a reliable framework for biomass estimation. The NDVI gave the highest importance for model development compared to other factors used for study. However, the study also highlights the potential for further accuracy improvements by integrating additional high-resolution data sources such as Sentinel-2 imagery or Synthetic Aperture Radar (SAR) data. Comparative analysis with other methods, including Inverse Distance Weighting (IDW) and the Forest Vegetation Simulator (FVS), underscores the strengths and limitations of different modeling approaches. The strong correlation between predicted biomass and interpolated data from FIA supports the model's application in environmental management and carbon accounting. The findings emphasize the importance of combining remote sensing, machine learning, and interpolation techniques to refine biomass predictions. Overall, this study contributes valuable insights into biomass distribution in SC, providing a foundation for future research in forest monitoring and management. Focusing on areas with high biomass for management could help reduce carbon loss due to fire and offer economic opportunities through sustainable resource utilization. Future research should explore incorporating more diverse data sources and advanced modeling techniques to enhance biomass prediction accuracy and address the inherent variability in these estimates.

References

1.   Brown, S. Measuring Carbon in Forests: Current Status and Future Challenges. *Environmental Pollution* **2002**, *116*, 363–372, doi:https://doi.org/10.1016/S0269-7491(01)00212-3.
2.   Houghton, R.A.; Hall, F.; Goetz, S.J. Importance of Biomass in the Global Carbon Cycle. *J Geophys Res Biogeosci* **2009**, *114*, doi:10.1029/2009JG000935.
3.   IPCC *AR4 Climate Change 2007: Impacts, Adaptation, and Vulnerability*; New York, 2007;
4.   Duncanson, L.; Disney, M.; Armston, J.; Nickeson, J.; Minor, D.; Camacho, F. Committee on Earth Observation Satellites Working Group on Calibration and Validation Land Product Validation Subgroup Aboveground Woody Biomass Product Validation Good Practices Protocol Version 1.0-2021. **2021**, doi:10.5067/doc/ceoswgcv/lpv/agb.001.
5.   Brown, S. *Estimating Biomass and Biomass Change of Tropical Forests: A Primer; Food and Agriculture Organization*; Rome,Italy, 1997; Vol. 134;.
6.   Food and Agriculture Organization *National Forest Monitoring and Assessment: Manual for Integrated Field Data Collection; NFMA Working Paper; Food and Agriculture Organization*; Rome,Italy, 2008;
7.   Chen, Q.; McRoberts, R. Statewide Mapping and Estimation of Vegetation Aboveground Biomass Using Airborne Lidar. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS); 2016; pp. 4442–4444.
8.   Ojoatre, S.; Zhang, C.; Hussin, Y.A.; Kloosterman, H.E.; Ismail, M.H. Assessing the Uncertainty of Tree Height and Aboveground Biomass From Terrestrial Laser Scanner and Hypsometer Using Airborne LiDAR

Data in Tropical Rainforests. *IEEE J Sel Top Appl Earth Obs Remote Sens* **2019**, *12*, 4149–4159, doi:10.1109/JSTARS.2019.2944779.

9. Silva, C.A.; Saatchi, S.; Garcia, M.; Labrière, N.; Klauberg, C.; Ferraz, A.; Meyer, V.; Jeffery, K.J.; Abernethy, K.; White, L.; et al. Comparison of Small- and Large-Footprint Lidar Characterization of Tropical Forest Aboveground Structure and Biomass: A Case Study From Central Gabon. *IEEE J Sel Top Appl Earth Obs Remote Sens* **2018**, *11*, 3512–3526, doi:10.1109/JSTARS.2018.2816962.

10. Yang, L.; Liang, S.; Zhang, Y. A New Method for Generating a Global Forest Aboveground Biomass Map From Multiple High-Level Satellite Products and Ancillary Information. *IEEE J Sel Top Appl Earth Obs Remote Sens* **2020**, *13*, 2587–2597, doi:10.1109/JSTARS.2020.2987951.

11. Dalponte, M.; Ene, L.T.; Gobakken, T.; Næsset, E.; Gianelle, D. Predicting Selected Forest Stand Characteristics with Multispectral ALS Data. *Remote Sens (Basel)* **2018**, *10*, doi:10.3390/rs10040586.

12. Coomes, D.A.; Dalponte, M.; Jucker, T.; Asner, G.P.; Banin, L.F.; Burslem, D.F.R.P.; Lewis, S.L.; Nilus, R.; Phillips, O.L.; Phua, M.-H.; et al. Area-Based vs Tree-Centric Approaches to Mapping Forest Carbon in Southeast Asian Forests from Airborne Laser Scanning Data. *Remote Sens Environ* **2017**, *194*, 77–88, doi:https://doi.org/10.1016/j.rse.2017.03.017.

13. Dalponte, M.; Frizzera, L.; Ørka, H.O.; Gobakken, T.; Næsset, E.; Gianelle, D. Predicting Stem Diameters and Aboveground Biomass of Individual Trees Using Remote Sensing Data. *Ecol Indic* **2018**, *85*, 367–376, doi:https://doi.org/10.1016/j.ecolind.2017.10.066.

14. Ferraz, A.; Saatchi, S.; Kellner, J.; Clark, D. Improving Carbon Estimation of Large Tropical Trees by Linking Airborne Lidar Crown Size to Field Inventory. In Proceedings of the IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium; 2018; pp. 8789–8792.

15. Urbazaev, M.; Thiel, C.; Cremer, F.; Schmullius, C. Assessment of the Mapping of Aboveground Biomass and Its Uncertainties Using Field Measurements, Airborne Lidar and Satellite Data in Mexico. In Proceedings of the IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium; 2018; pp. 8793–8796.

16. Herold, M.; Carter, S.; Avitabile, V.; Espejo, A.B.; Jonckheere, I.; Lucas, R.; McRoberts, R.E.; Næsset, E.; Nightingale, J.; Petersen, R.; et al. The Role and Need for Space-Based Forest Biomass-Related Measurements in Environmental Management and Policy. *Surv Geophys* 2019, *40*, 757–778.

17. Ghamisi, P.; Rasti, B.; Yokoya, N.; Wang, Q.; Hofle, B.; Bruzzone, L.; Bovolo, F.; Chi, M.; Anders, K.; Gloaguen, R.; et al. Multisource and Multitemporal Data Fusion in Remote Sensing: A Comprehensive Review of the State of the Art. *IEEE Geosci Remote Sens Mag* **2019**, *7*, 6–39, doi:10.1109/MGRS.2018.2890023.

18. Li, W.; Niu, Z.; Shang, R.; Qin, Y.; Wang, L.; Chen, H. High-Resolution Mapping of Forest Canopy Height Using Machine Learning by Coupling ICESat-2 LiDAR with Sentinel-1, Sentinel-2 and Landsat-8 Data. *International Journal of Applied Earth Observation and Geoinformation* **2020**, *92*, 102163, doi:https://doi.org/10.1016/j.jag.2020.102163.

19. Malhi, R.K.M.; Anand, A.; Srivastava, P.K.; Chaudhary, S.K.; Pandey, M.K.; Behera, M.D.; Kumar, A.; Singh, P.; Sandhya Kiran, G. Synergistic Evaluation of Sentinel 1 and 2 for Biomass Estimation in a Tropical Forest of India. *Advances in Space Research* **2022**, *69*, 1752–1767, doi:https://doi.org/10.1016/j.asr.2021.03.035.

20. Cutler, M.E.J.; Boyd, D.S.; Foody, G.M.; Vetrivel, A. Estimating Tropical Forest Biomass with a Combination of SAR Image Texture and Landsat TM Data: An Assessment of Predictions between Regions. *ISPRS Journal of Photogrammetry and Remote Sensing* **2012**, *70*, 66–77, doi:https://doi.org/10.1016/j.isprsjprs.2012.03.011.

21. Abdullah, H.; Skidmore, A.K.; Darvishzadeh, R.; Heurich, M. Sentinel-2 Accurately Maps Green-Attack Stage of European Spruce Bark Beetle (Ips Typographus, L.) Compared with Landsat-8. *Remote Sens Ecol Conserv* **2019**, *5*, 87–106, doi:10.1002/rse2.93.

22. Castillo, J.A.A.; Apan, A.A.; Maraseni, T.N.; Salmo, S.G. Estimation and Mapping of Above-Ground Biomass of Mangrove Forests and Their Replacement Land Uses in the Philippines Using Sentinel Imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* **2017**, *134*, 70–85, doi:https://doi.org/10.1016/j.isprsjprs.2017.10.016.

23. Chen, Y.; Hou, J.; Huang, C.; Zhang, Y.; Li, X. Mapping Maize Area in Heterogeneous Agricultural Landscape with Multi-Temporal Sentinel-1 and Sentinel-2 Images Based on Random Forest. *Remote Sens (Basel)* **2021**, *13*, doi:10.3390/rs13152988.

24. Lu, D.; Batistella, M. *Exploring TM Image Texture and Its Relationships with Biomass Estimation in Rondônia, Brazilian Amazon*; 2005;

25. Wang, X.; Pang, Y.; Zhang, Z.; Yuan, Y. *FOREST ABOVEGROUND BIOMASS ESTIMATION USING SPOT-5 TEXTURE INDICES AND SPECTRAL DERIVATIVES*;

26. Su, H.; Shen, W.; Wang, J.; Ali, A.; Li, M. Machine Learning and Geostatistical Approaches for Estimating Aboveground Biomass in Chinese Subtropical Forests. *For Ecosyst* **2020**, *7*, doi:10.1186/s40663-020-00276-7.

27. Li, L.; Zhou, B.; Liu, Y.; Wu, Y.; Tang, J.; Xu, W.; Wang, L.; Ou, G. Reduction in Uncertainty in Forest Aboveground Biomass Estimation Using Sentinel-2 Images: A Case Study of Pinus Densata Forests in Shangri-La City, China. *Remote Sens (Basel)* **2023**, *15*, doi:10.3390/rs15030559.

28. Nguyen, T.D.; Kappas, M. Estimating the Aboveground Biomass of an Evergreen Broadleaf Forest in Xuan Lien Nature Reserve, Thanh Hoa, Vietnam, Using SPOT-6 Data and the Random Forest Algorithm. *International Journal of Forestry Research* **2020**, *2020*, doi:10.1155/2020/4216160.

29. Zhu, X.; Liu, D. Improving Forest Aboveground Biomass Estimation Using Seasonal Landsat NDVI Time-Series. *ISPRS Journal of Photogrammetry and Remote Sensing* **2015**, *102*, 222–231, doi:https://doi.org/10.1016/j.isprsjprs.2014.08.014.

30. Wang, J.; Xiao, X.; Bajgain, R.; Starks, P.; Steiner, J.; Doughty, R.B.; Chang, Q. Estimating Leaf Area Index and Aboveground Biomass of Grazing Pastures Using Sentinel-1, Sentinel-2 and Landsat Images. *ISPRS Journal of Photogrammetry and Remote Sensing* **2019**, *154*, 189–201, doi:https://doi.org/10.1016/j.isprsjprs.2019.06.007.

31. Tang, J.; Liu, Y.; Li, L.; Liu, Y.; Wu, Y.; Xu, H.; Ou, G. Enhancing Aboveground Biomass Estimation for Three Pinus Forests in Yunnan, SW China, Using Landsat 8. *Remote Sens (Basel)* **2022**, *14*, doi:10.3390/rs14184589.

32. Izquierdo-Verdiguier, E.; Zurita-Milla, R. An Evaluation of Guided Regularized Random Forest for Classification and Regression Tasks in Remote Sensing. *International Journal of Applied Earth Observation and Geoinformation* **2020**, *88*, 102051, doi:https://doi.org/10.1016/j.jag.2020.102051.

33. Deng, H.; Runger, G. Gene Selection with Guided Regularized Random Forest. *Pattern Recognit* **2013**, *46*, 3483–3489, doi:https://doi.org/10.1016/j.patcog.2013.05.018.

34. Francke, T.; López-Tarazón, J.A.; Schröder, B. Estimation of Suspended Sediment Concentration and Yield Using Linear Models, Random Forests and Quantile Regression Forests. *Hydrol Process* **2008**, *22*, 4892–4904, doi:10.1002/hyp.7110.

35. Sharma, S.; Khanal, P. Forest Fire Prediction: A Spatial Machine Learning and Neural Network Approach. *Fire* **2024**, *7*, doi:10.3390/fire7060205.

36. Tian, Y.; Zhang, Y.; Knyazikhin, Y.; Myneni, R.B.; Glassy, J.M.; Dedieu, G.; Running, S.W. *Prototyping of MODIS LAI and FPAR Algorithm with LASUR and LANDSAT Data*; 2000; Vol. 38;.

37. Alexandre Santos Querino Cristina Aparecida Beneditti Nadja Gomes Machado Marcelo José Gama da Silva Juliane Kayse Albuquerque da Silva Querino Luiz Alves dos Santos Neto Marcelo Sacardi Biudes Carlos Alexandre Santos Querino, C.; Aparecida Beneditti, C.; Gomes Machado, N.; José Gama da Silva, M.; Kayse Albuquerque da Silva Querino, J.; Alves dos Santos Neto, L.; Alexandre Santos Querino, C.; Sacardi Biudes, M. Spatiotemporal NDVI, LAI, Albedo, and Surface Temperature Dynamics in the Southwest of the Brazilian Amazon Forest. *J. Appl. Remote Sens* **2016**, *10*, 26007, doi:10.1117/1.

38. Mizen, A.; Thompson, D.A.; Watkins, A.; Akbari, A.; Garrett, J.K.; Geary, R.; Lovell, R.; Lyons, R.A.; Nieuwenhuijsen, M.; Parker, S.C.; et al. The Use of Enhanced Vegetation Index for Assessing Access to Different Types of Green Space in Epidemiological Studies. *J Expo Sci Environ Epidemiol* **2024**, doi:10.1038/s41370-024-00650-5.

39. Drisya, J.; D, S.K.; Roshni, T. Chapter 27 - Spatiotemporal Variability of Soil Moisture and Drought Estimation Using a Distributed Hydrological Model. In *Integrating Disaster Science and Management*; Samui, P., Kim, D., Ghosh, C., Eds.; Elsevier, 2018; pp. 451–460 ISBN 978-0-12-812056-9.

40. Hojo, A.; Avtar, R.; Nakaji, T.; Tadono, T.; Takagi, K. Modeling Forest Above-Ground Biomass Using Freely Available Satellite and Multisource Datasets. *Ecol Inform* **2023**, *74*, 101973, doi:https://doi.org/10.1016/j.ecoinf.2023.101973.

41. Vahidi, M.; Shafian, S.; Thomas, S.; Maguire, R. Estimation of Bale Grazing and Sacrificed Pasture Biomass through the Integration of Sentinel Satellite Images and Machine Learning Techniques. *Remote Sens (Basel)* **2023**, *15*, doi:10.3390/rs15205014.

42. Tamiminia, H.; Salehi, B.; Mahdianpari, M.; Goulden, T. 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* **2024**, *79*, 102404, doi:https://doi.org/10.1016/j.ecoinf.2023.102404.

43. Lu, Z.; Chen, P.; Yang, Y.; Zhang, S.; Zhang, C.; Zhu, H. Exploring Quantification and Analyzing Driving Force for Spatial and Temporal Differentiation Characteristics of Vegetation Net Primary Productivity in Shandong Province, China. *Ecol Indic* **2023**, *153*, 110471, doi:https://doi.org/10.1016/j.ecolind.2023.110471.

44. Hojo, A.; Avtar, R.; Nakaji, T.; Tadono, T.; Takagi, K. Modeling Forest Above-Ground Biomass Using Freely Available Satellite and Multisource Datasets. *Ecol Inform* **2023**, *74*, 101973, doi:https://doi.org/10.1016/j.ecoinf.2023.101973.

45. Mehmood, K.; Anees, S.A.; Rehman, A.; Pan, S.; Tariq, A.; Zubair, M.; Liu, Q.; Rabbi, F.; Khan, K.A.; Luo, M. Exploring Spatiotemporal Dynamics of NDVI and Climate-Driven Responses in Ecosystems: Insights for Sustainable Management and Climate Resilience. *Ecol Inform* **2024**, *80*, 102532, doi:https://doi.org/10.1016/j.ecoinf.2024.102532.

46. Salinas-Melgoza, M.A.; Skutsch, M.; Lovett, J.C. Predicting Aboveground Forest Biomass with Topographic Variables in Human-Impacted Tropical Dry Forest Landscapes. *Ecosphere* **2018**, *9*, doi:10.1002/ecs2.2063.

47. Martinuzzi, S.; Cook, B.D.; Helmer, E.H.; Keller, M.; Locke, D.H.; Marcano-Vega, H.; Uriarte, M.; Morton, D.C. Patterns and Controls on Island-Wide Aboveground Biomass Accumulation in Second-Growth Forests of Puerto Rico. *Biotropica* **2022**, *54*, 1146 – 1159, doi:10.1111/btp.13122.

48. Chen, L.; Wang, Y.; Ren, C.; Zhang, B.; Wang, Z. Assessment of Multi-Wavelength SAR and Multispectral Instrument Data for Forest Aboveground Biomass Mapping Using Random Forest Kriging. *For Ecol Manage* **2019**, *447*, 12–25, doi:https://doi.org/10.1016/j.foreco.2019.05.057.

49. Luo, M.; Wang, Y.; Xie, Y.; Zhou, L.; Qiao, J.; Qiu, S.; Sun, Y. Combination of Feature Selection and Catboost for Prediction: The First Application to the Estimation of Aboveground Biomass. *Forests* **2021**, *12*, 1–22, doi:10.3390/f12020216.

50. Nandy, S.; Srinet, R.; Padalia, H. 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* **2021**, *48*, doi:10.1029/2021GL093799.

51. Purohit, S.; Aggarwal, S.P.; Patel, N.R. Estimation of Forest Aboveground Biomass Using Combination of Landsat 8 and Sentinel-1A Data with Random Forest Regression Algorithm in Himalayan Foothills. *Trop Ecol* **2021**, *62*, 288–300, doi:10.1007/s42965-021-00140-x.

52. Puletti, N.; Mattioli, W.; Bussotti, F.; Pollastrini, M. Monitoring the Effects of Extreme Drought Events on Forest Health by Sentinel-2 Imagery. *J Appl Remote Sens* **2019**, *13*, 1, doi:10.1117/1.jrs.13.020501.

53. Fassnacht, F.E.; Poblete-Olivares, J.; Rivero, L.; Lopatin, J.; Ceballos-Comisso, A.; Galleguillos, M. Using Sentinel-2 and Canopy Height Models to Derive a Landscape-Level Biomass Map Covering Multiple Vegetation Types. *International Journal of Applied Earth Observation and Geoinformation* **2021**, *94*, 102236, doi:https://doi.org/10.1016/j.jag.2020.102236.

54. Hawryło, P.; Wezyk, P. Predicting Growing Stock Volume of Scots Pine Stands Using Sentinel-2 Satellite Imagery and Airborne Image-Derived Point Clouds. *Forests* **2018**, *9*, doi:10.3390/f9050274.

55. Guerra-Hernández, J.; Narine, L.L.; Pascual, A.; Gonzalez-Ferreiro, E.; Botequim, B.; Malambo, L.; Neuenschwander, A.; Popescu, S.C.; Godinho, S. Aboveground Biomass Mapping by Integrating ICESat-2, SENTINEL-1, SENTINEL-2, ALOS2/PALSAR2, and Topographic Information in Mediterranean Forests. *GIsci Remote Sens* **2022**, *59*, 1509–1533, doi:10.1080/15481603.2022.2115599.

56. Sousa, A.M.O.; Gonçalves, A.C.; Mesquita, P.; Marques da Silva, J.R. Biomass Estimation with High Resolution Satellite Images: A Case Study of Quercus Rotundifolia. *ISPRS Journal of Photogrammetry and Remote Sensing* **2015**, *101*, 69–79, doi:https://doi.org/10.1016/j.isprsjprs.2014.12.004.

57. Gonçalves, A.C.; Sousa, A.M.O.; Mesquita, P.G. Estimation and Dynamics of above Ground Biomass with Very High Resolution Satellite Images in Pinus Pinaster Stands. *Biomass Bioenergy* **2017**, *106*, 146–154, doi:https://doi.org/10.1016/j.biombioe.2017.08.026.