Estimating Above-Ground Biomass Density Using a Multi-Source Remote Sensing Datasets in Protected Forests of Gilan

Houri Gholamrezaie[1], Mahdi Hasanlou[1*],Hossein Arefi [1,2]

[1] School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran 1439957131, Iran; h.gholamrezaie@ut.ac.ir; hasanlou@ut.ac.ir; hossein.arefi@ut.ac.ir
[2] i3mainz, Institute for Spatial Information and Surveying Technology, Mainz University of Applied Sciences, 55128 Mainz, Germany, hossein.arefi@hs-mainz.de

Keywords: Forest, Above-Ground Biomass, Spaceborne LiDAR, Remote Sensing, Random Forest

Abstract

Accurate estimation of Above-Ground Biomass Density (AGBD) is crucial for assessing vegetation structure, carbon accounting, and forest productivity, as well as for supporting climate change mitigation and sustainable forest management. Protected forest areas, such as those in Gilan Province in northern Iran, represent ecologically valuable ecosystems. However, due to limited access to field- and airborne-based LiDAR data, spatially continuous monitoring of these areas remains a challenge. Recent advancements in remote sensing, particularly the availability of spaceborne LiDAR from missions like Global Ecosystem Dynamics Investigation (GEDI), have opened new possibilities for monitoring forest structure over large and remote areas. The GEDI mission provides point-based measurements of canopy height and biomass, which can help overcome the limitations of traditional field-based methods. In this study, we propose an approach for estimating AGBD in the protected forests of Gilan using GEDI Level 4A-derived vertical structure data, combined with wall-to-wall multispectral data from Sentinel-2 and terrain information from SRTM. Based on machine learning algorithms, we extend the point-based biomass estimates to generate spatially continuous AGBD maps across the study area. The results highlight the potential of freely available spaceborne LiDAR, in conjunction with satellite data and modeling techniques, for mapping local-scale biomass. This study offers a valuable baseline for future research in forest ecosystem monitoring and opens the door for utilizing other satellite-based LiDAR resources such as ICESat-2 and incorporating advanced machine learning algorithms for biomass estimation and carbon stock assessment.

1. Introduction

Forests, covering one-third of the Earth's land surface, provide a range of valuable ecosystem goods and services to human beings (Ouaknine et al., 2023). Indeed, forests constitute one of the largest reservoirs of terrestrial carbon and play a vital role in mitigating climate change and regulating the global carbon balance (Kombate et al., 2025). However, forests are widely threatened by deforestation, wildfires, and storms, which can affect their structure and carbon storage capacity (Kacic et al., 2025). Furthermore, global demand for commodities continues to drive deforestation, thereby hindering progress towards sustainability goals. On the other hand, the re-growth of forests plays an important role in the global carbon cycle to contribute to the reduction in the uncertainties of carbon balance models (Milenković et al., 2022). Consequently, effective monitoring of the forests is essential to address these challenges and improve the accuracy of biomass estimations (Atkins et al., 2025).

The traditional field-based methods for the estimation of aboveground biomass density (AGBD), while essential for ecological management, are often time-consuming, labor-intensive, and limited in their ability to cover vast or inaccessible areas (Y. Li et al., 2024).

Remote sensing technologies offer a more efficient alternative, enabling large-scale forest monitoring and providing valuable data on forest cover, biomass, and land use (Weinstein et al., 2021). Although optical imagery offers high spatial and temporal resolution, it is constrained by signal saturation when estimating the forest's vertical structure and cannot differentiate between the single- and multi-layer canopies, which is an important limitation for forest studies. Synthetic Aperture Radar (SAR) sensors partially address these shortcomings by penetrating forest canopies and clouds and offering improved temporal sampling; however, they still face difficulties in accurately capturing the vertical dimension of forests (Heisig et al., 2025). To tackle these challenges, LiDAR — especially Airborne Laser Scanning (ALS) — has emerged as a highly effective approach for estimating forest canopy height (FCH) and modeling structural attributes such as vegetation height, crown density, and biomass, owing to its ability to directly sense the vertical structure (Travers-Smith et al., 2024). Nevertheless, the widespread application of ALS is limited due to the high operational costs, confined spatial coverage, and low temporal revisit rates, as well as the technical expertise and computational resources required for the data processing, which restrict its feasibility for large-scale, continuous, and high-precision forest monitoring (Vogeler et al., 2023; Weinstein et al., 2021).

Recent advancements in spaceborne remote sensing technologies have enabled accurate monitoring of forest dynamics at large scales. Spaceborne LiDAR missions like ICESat-2 and GEDI provide valuable vertical structure data for estimating global canopy height and biomass (Lang et al., 2023). Among these, GEDI has emerged as a key resource for

* Corresponding author

large-scale biomass estimation, particularly through its Level-4A and Level-4B products. Nevertheless, several studies have highlighted the importance of improving the GEDI predictive performance through regional calibration, multi-source data integration, and advanced modeling techniques. A growing amount of research has focused on calibrating and validating the GEDI biomass estimates using field measurements, ALS, and other remote sensing data. Pascual *et al.* (2023), for instance, assessed the performance of the GEDI L4A biomass model in the Mediterranean forests of Spain and emphasized the need for localized calibration to address the overestimations of biomass predictions across different forest types; advocating for continuous validation and integration with ALS and National Forest Inventory (NFI) data to enhance the accuracy of carbon reporting systems (Pascual et al., 2023). Similarly, Liu *et al.* (2024) examined the GEDI L2A and L4A products in conjunction with airborne LVIS and TomoSAR P-band backscatter to analyze the forest structure in tropical Gabon. Their results revealed moderate to good agreement but also underscored some underestimations due to potential inaccuracies in the GEDI-derived reference data (Liu et al., 2024).

Further efforts to compare the GEDI and ICESat-2 data have provided valuable insights into their respective capabilities. Musthafa *et al.* compared the forest stand height estimations from both platforms across Indian forest types, and concluded that the GEDI yielded lower RMSE values, especially in tropical and subtropical regions; demonstrating higher reliability in the canopy height modeling (Musthafa et al., 2023). Bhandari *et al.* also utilized GEDI L2A alongside Sentinel-1, Sentinel-2, and SRTM to map both the canopy height and AGB in northern India, validating the results with ground measurements (Bhandari et al., 2024). In another study, the estimates from the GEDI L4A and L4B in Indian tropical forests (Betul and Mudumalai) were evaluated against the field data using multiple regression models. While the GEDI slightly underestimated the AGB, the linear regression approach was found to be most suitable, emphasizing the need to refine the GEDI-derived products for carbon monitoring accuracy (Indirabai & Nilsson, 2024).

Alongside the dataset of the LiDAR onboard the ICESat-2 and GEDI missions, available as sparse point-based clouds, in general, allow for the estimation of canopy height, but are limited to specific collection areas, lacking continuous coverage over large regions. In contrast, the optical or radar satellite data offer widespread spatial continuity but cannot directly extract the vertical canopy profiles. As such, integrating these types of datasets is essential for accurate forest height mapping. For this purpose, machine learning methods are increasingly used to combine the LiDAR vertical profiles with the spectral or backscatter features from the optical and synthetic aperture radar (SAR) data (Kombate et al., 2025). In Indian forests, Gupta and Sharma (2022) combined the GEDI footprints with the optical and SAR data using a Random Forest model to estimate the AGBD in the Shoolpaneshwar Wildlife Sanctuary, achieving promising predictive accuracy (Gupta & Sharma, 2022). Mohite *et al.* followed a similar methodology across diverse Indian forest types, incorporating the Sentinel-1, Sentinel-2, DEMs, and forest maps to model the AGBD with an RMSE of 47.2 Mg ha^{-1} (Mohite et al., 2024). A comparable approach was adopted by Wang et al. (2024), who utilized the Sentinel-1, Sentinel-2, ALOS PALSAR-2, and GEDI L4A data for the biomass estimation in Yunnan Province, China. Their Random Forest models showed a desirable performance, especially in broadleaf forests, with RMSE as low as 12.54 Mg ha^{-1} (C. Wang et al., 2024).

Furthermore, the integration of the GEDI data with the topographic, climatic, and ecological variables has proven its effectiveness in regional biomass assessments. Yang *et al.* used the ICESat-2 and GEDI-derived canopy heights, along with the Sentinel-2, climatic, and soil variables, to generate a high-resolution AGB map of China, achieving an adjusted coefficient of determination (R^2) of 0.67 and an average AGB of 97.57 Mg ha^{-1} (Yang et al., 2023). In Mexico, Ortiz-Reyes *et al.* applied the GEDI L4A data with the GLMM and LMM statistical models to estimate the AGB, basal area, and wood volume, reporting strong correlations between the predictions and field observations (Ortiz-Reyes et al., 2024). Likewise, Xu *et al.* (2024) explored the AGB prediction in spruce-fir forests of Shangri-La, China, using the GEDI L2B data and a reduced set of modeling variables, showing that the strategic variable selection and footprint density can improve the model performance (Xu et al., 2024).

Collectively, previous studies highlight the significant potential of GEDI data for estimating the aboveground biomass and monitoring the forest structural attributes. They also underscore the necessity of integrating multi-source datasets, selecting appropriate modeling approaches, and considering local ecological contexts to ensure robust and spatially continuous biomass predictions. The present study aims to further investigate the potential of GEDI data for estimating the local-scale aboveground biomass in the protected regions of the Gilan Province, an important natural resource that plays a significant ecological role across the country of Iran. To do so, we integrate the variables derived from the Sentinel-2 imagery and spaceborne LiDAR dataset, through a Random Forest-based machine learning algorithm. The results are intended to serve as a promising foundation for future research and to support forest management decision-making through the use of freely available spaceborne LiDAR data. This research contributes to the ongoing advancement of spaceborne LiDAR applications for forest monitoring.

2. Materials and Methodology

In line with the framework shown in Figure 1, the main components of the approach implemented for this research include the data preparation and processing tasks. The preprocessing of the main data (i.e., preparing the main data for input into the algorithm), includes steps such as filtering the data based on time and location, applying a cloud mask, and removing outlier data. The processing tasks mainly involve steps towards running the machine learning algorithms to calculate regression models, accuracy assessment, and predictive biomass. The descriptions of the dataset, the study area, and the applied methods are consecutively explained in detail below.

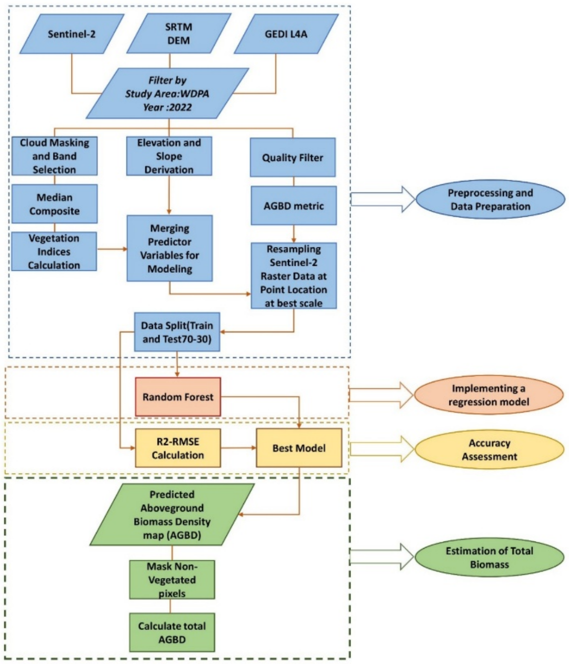

Figure 1. Diagrammatic workflow of the proposed methodology to estimate the AGBD in the forest regions using the multisource remote-sensing dataset of GEDI, Sentinel-2, and SRTM.

2.1 Datasets

The study dataset includes the GEDI L4A, Sentinel-2 spectral bands, and the derived vegetation indices, as well as the global SRTM DEM, the World Database on Protected Areas (WDPA) in which the protected forest areas of interest are embedded (Figure 1). Also, the GEDI L2A data is used to generally assess the current condition of the trees in the forests of interest.

2.1.1 Global Ecosystem Dynamics Investigation (GEDI): GEDI is NASA's pioneering and latest spaceborne LiDAR mission, specifically created to measure global vertical forest structure globally. The GEDI instrument collected data between latitudes 51.6° North and South from April 2019 to March 2023 (Moudrý et al., 2024; Lang et al., 2023; Heisig et al., 2025). Data collection was paused for a period, but it resumed in 2024. The sensor is equipped with two full-power lasers and one coverage laser, which are optically dithered to create eight parallel measurement tracks spaced 600 meters apart. The laser pulses hit the ground within a 25-meter diameter footprint ("shot"), with footprints spaced every 60 meters along each track. The sensor generates a one-dimensional continuous return waveform. Its 1064 nm wavelength is capable of penetrating dense forest canopies, allowing for the measurement of the upper canopy, understory, and ground using peaks in the return waveform (Holcomb et al., 2024).
The GEDI-related studies predominantly utilize the L2A and L2B products (J. Wang et al., 2024). The L2A product includes relative height metrics in 100 height intervals or percentiles (RH0, RH1, …, RH100). The L2B product encompasses canopy profile metrics such as the Plant Area Index (PAI), vertical Plant Area Volume Density (PAVD), Foliage Height Diversity (FHD) index, and Total Canopy Cover (TCC). The L4A product contains estimates of above-ground biomass density (AGBD) derived from L2A data (Hoffrén et al., 2023).

The mission has a geolocation accuracy of approximately 10 meters and canopy height measurement accuracy of about 1 meter (Quiros et al., 2021; Shannon et al., 2023). The data collected by GEDI provide valuable large-scale information that can be used freely and efficiently by users, instead of costly regional and/or national initiatives. GEDI differentiates itself from other spaceborne LiDAR systems due to its superior penetration capability in dense vegetation (Hoffrén et al., 2023).

2.1.1.1 GEDI Data Filtering and Preparation: The GEDI uses 1,064 nm wavelength lasers in the near-infrared range, which struggle to penetrate dense cloud cover, causing around 50% of pulses to return useful signals, with the rest degraded or lost due to atmospheric interactions, solar noise, or retrieval algorithm limitations (Moudrý et al., 2024). To improve the data quality, GEDI products include flags to filter observations with high geolocation uncertainty or poor signal quality (Ceccherini et al., 2023; Hoffrén et al., 2023), ensuring high-quality data for algorithm training and validation (Potapov et al., 2021; Hoffrén et al., 2023; Zhu et al., 2023).
Many studies used three main quality control filters—"sensitivity," "quality," and "degrade flags"—to assess waveform reliability for 3D surface structures (Alvites et al., 2024).
According to an experiment (Pronk et al., 2024), the terrain slope significantly affects the vertical accuracy, in which the GEDI is more affected than ICESat-2 due to larger geolocation errors (Pronk et al., 2024). In another study (Ceccherini et al., 2023), samples on slopes greater than 10° were filtered to reduce canopy height errors, using the slope data derived from the Shuttle Radar Topography Mission (SRTM). We note that though the GEDI-inferred terrain height retrieval is more accurate than SRTM, the slope information is yet crucial to filter out the steep slope samples (Ceccherini et al., 2023).
Some studies used the power beam mode during the night to reduce background noise from solar radiation (Potapov et al., 2021), while a study found minimal effect from beam power, solar radiation, or seasonal changes on accuracy (Pronk et al., 2024). Since the GEDI waveforms are processed with six algorithm groups, optimized based on energy, region, and vegetation, retaining observations with ground elevation differences ≤2 m (Moudrý et al., 2024).

2.1.1.2 Preparing GEDI L4A: It is crucial to filter out invalid or unreliable GEDI data before utilizing them in further analysis. To ensure the quality of the data, we applied a series of conditional criteria to retain only the highest quality observations, which were then used as model reference and testing data for subsequent analysis. The filtering criteria were as follows:
1. Exclude all measurements that do not meet the quality requirements (i.e., l4_quality_flag = 0 and degrade_flag > 0).
2. Exclude measurements with high relative error (agbd_se / agbd > 50%), where agbd represents the predicted aboveground biomass density, and agbd_se is the standard error of this prediction.
3. Exclude measurements taken on slopes greater than 30%.
Finally, we selected all remaining measurements within the specified period and region of interest to create a mosaic for further analysis.

2.1.2 The World Database on Protected Areas (WDPA): The World Database on Protected Areas (WDPA) is a comprehensive global database of protected areas, compiled by UNEP-WCMC in collaboration with various governments and organizations. We used this dataset to mask and extract the protected forest areas of Gilan Province.

2.1.3 Shuttle Radar Topography Mission (SRTM): Several global DEM products have become freely accessible to the public since 2000. Among these datasets, the Shuttle Radar Topography Mission (SRTM) DEM released in February 2003, is the most successful despite the presence of voids and non-negligible vertical errors (H. Li et al., 2022). The spatial resolution of the SRTM DEM is 1 arc second, equivalent to approximately 30 m (Zhu et al., 2023).

2.1.4 Sentinel 2: We used Sentinel-2 optical imagery from the Copernicus program (European Commission and ESA). Sentinel-2 carries a Multi-Spectral Instrument (MSI) with 13 bands ranging from visible to short-wave infrared, at spatial resolutions between 10 m and 60 m. For the study area, a 2022 mosaic was created using bands B2, B3, B4, B5, B6, B7, B8, B11, and B12. Cloud-contaminated pixels were masked using the QA60 and Cloud Probability bands. Vegetation indices including NDVI, SAVI, and RESI were computed, and the median values of all spectral bands and indices were extracted for each GEDI footprint (Ceccherini et al., 2023).

2.1.5 ESA Worldcover Map 2021: This study utilized the ESA Worldcover Map 2021 to mask out the non-forest regions in the resultant aboveground biomass maps. This mask has so far been used in many studies, for example (Moudrý et al., 2024; Pascual et al., 2024).

2.2 Study Area

As our study regions, we selected the protected forest areas embedded in Gilan Province, located in the northern part of Iran. The coastal province of Gilan is bordered by the Caspian Sea and Azerbaijan to the north, Ardabil to the west, Zanjan and Qazvin to the south, and Mazandaran to the east(S.Abdollahi, 2019). Figure 2 shows the map of Gilan Province, along with the delineation of the protected forest areas including "Siahroud e Roudbar", "Sarvelat and Javaherdasht", "Siahkeshim", "Lisar", and "Gasht-e-Rodkan".

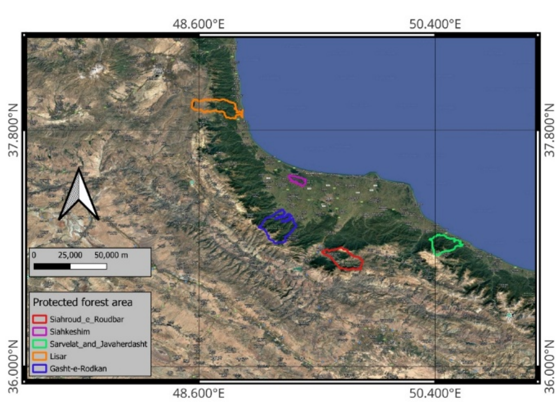

Figure 2. Map of the study area showing the protected forest regions of Gilan Province.

A significant portion of land areas in Gilan has been subjected to reforestation projects. However, despite the extensive afforestation efforts, no precise data or maps exist regarding tree height, species, or the extent of afforested areas (Bayat et al., 2025). In these forests, however, the availability and accessibility of airborne LiDAR data and field data are unfortunately limited.

As we did not have access to the information about the tree heights in these five protected regions, we resorted to the GEDI L2A to investigate the potential to derive general information for comparison purposes, though the data are assumed to be not purely error-free. Figure 3 presents the box and density plots of tree heights, showing the GEDI Level 2A canopy height distribution within the protected forest areas. The box plot shows that the median canopy height falls between approximately 7.3 m and 25.9 m. The interquartile range (IQR), which represents the middle 50% of canopy height data, is 12.4 m, 10.9 m, 21.4 m, 21.9 m, and 16.6 m respectively for the protected forests of Siahroud e Roudbar, Siahkeshim, Sarvelat and Javaherdasht, Lisar, and Gasht-e-Rodkan.

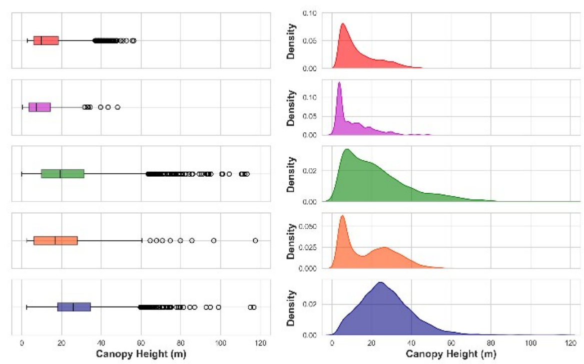

Figure 3. The box and distribution plots of tree heights in the protected forests of (top row) Siahroud e Roudbar, (second row) Siahkeshim, (third row) Sarvelat and Javaherdasht, (fourth row) Lisar, and (bottom row) Gasht-e-Rodkan.

2.3 Methods

To achieve high-resolution and full-coverage maps of above-ground biomass (AGB) estimates, numerous studies have applied machine learning (ML) algorithms to downscale and spatialize satellite LiDAR data. These studies often integrate satellite LiDAR with optical imagery to reduce the original footprint size and extend the spatial coverage across entire forest ecosystems. In the present study, we employed Random Forest (RF) regression models to estimate AGB, using optical spectral indices from Sentinel-2 and topographic data from the Shuttle Radar Topography Mission (SRTM) as predictors.

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLVIII-4-W22-2025-29-2026

2.3.1 Random Forest (RF) Classifier: Among the ML algorithms, the random forest has demonstrated superior performance in predicting the AGB. The ability of RF algorithms to handle non-linear relationships, high-dimensional data, and complex interactions among variables makes it more suitable for AGB estimation. It provides an inherent mechanism for ranking the importance of input variables, which is beneficial for understanding the contribution of predictors (e.g., spectral indices, topographic data, and GEDI data) for biomass estimation.

The RF algorithm is optimized according to the number of regression trees (*ntree*) and the number of predictors (*mtry*) at each split as it creates a new tree (node). The *mtry* predictor at each node affects the accuracy of a tree and increasing the *ntree* number increases the performance of models. To determine the optimized values of mtry and ntrees, we repeated thrice from a range of 1–6 mtry and 50–1500 ntrees. The model with the highest accuracy was consequently selected as the final model.

In this research, the Sentinel-2 bands and spectral indices including the NDVI, SAVI, and RESI as well as the elevation and slope derived from SRTM were assumed as independent predictor variables in the random forest regression modeling, while the AGBD from GEDI L4A was considered as a dependent predicted variable. We utilized the AGB data derived from GEDI L4A for training and testing purposes.

2.3.2 Validation: With no access to the filed-based data of biomass in the study regions, we have divided the GEDI level 4A biomass data into two sets of 70% for training models and 30% for testing the biomass at the locations of the predictive biomass from the model. Specifically, the performances of the derived models were assessed with the remaining 30% of the data (test data) using statistical metrics like the coefficient of determination (R^2), and the root mean square (RMSE) with the following notations, respectively:

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

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

Using the *sampleRegions* function, the sampling is conducted from the mosaic of *image collection* at an appropriate scale. The optimal scale of 150 is set for the study regions, which achieves a balance between the validation parameters of adjusted coefficient of determination (R^2) and root-mean-square error (RMSE).

3. Results

We observed that the AGBD data derived from the GEDI L4A in the study area varies between the minimum and maximum values of 6 and 2209 Mg ha^{-1}, respectively, deviating significantly from the average value of 120.92 Mg ha^{-1}. As shown in Figure 4 (top), based on the histogram, distribution, and boxplot of the data, it is evident that most of the AGBD are concentrated between the values of 0 and 400 Mg ha^{-1}. Therefore, the extremely high values may be considered outliers. Consequently, we filtered out the values greater than 350 Mg ha^{-1} and retained only the values below this threshold, in which the resulting AGBD were then used as the training and testing data in the Random Forest modeling. Figure 4 (bottom) shows the histogram, distribution, and boxplot of the AGBD data values after the filtering.

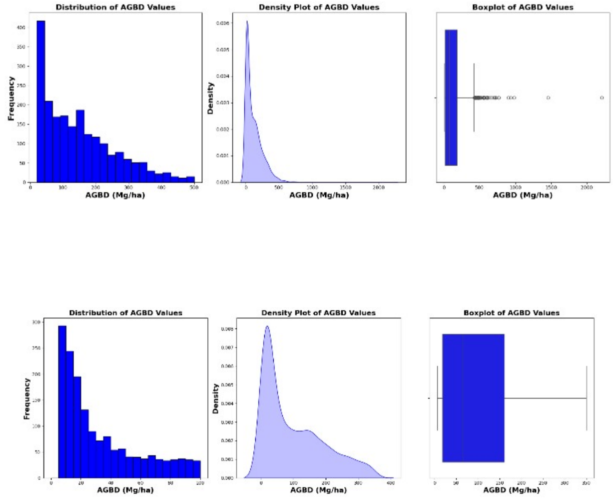

Figure 4. The histogram, distribution, and boxplot of the AGBD values derived from GEDI L4A, using (top row) all the available data, (bottom row) the remaining data after filtering out using a threshold of 400 Mg ha^{-1}.

The AGBD was estimated from a Random Forest regression model trained based on the GEDI L4A AGBD data along with the predictors derived from the Sentinel-2 bands and indices, and the "slope" and "elevation" derived from SRTM DEM.

As shown in Figure 5, in the Random Forest process the key variables for better performance evaluations of the forest AGBD estimations were "elevation", "RESI", "NDVI", and "Slope" respectively, with substantial variations in their respective importance values.

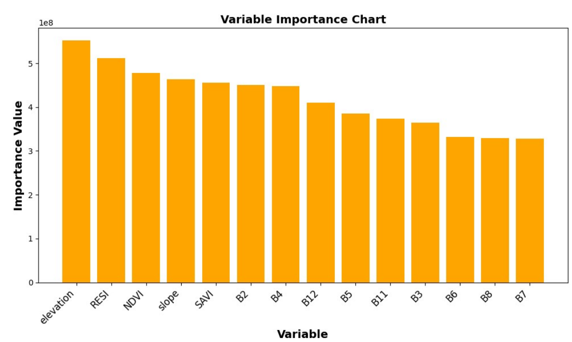

Figure 5. Key variables used for performance evaluations of the forest AGBD estimations in the Random Forest process, sorted according to their importance values

The estimated forest AGB map is then evaluated using the retained 30% of independent GEDI L4A-derived measurements. As shown in Figure 6, the fitted line between the observed and predicted values is relatively far from the 1:1 line. The statistical parameters of R^2 and RMSE between the observed and predicted AGB are 0.41 and 63.62 Mg ha^{-1}, respectively, revealing the accuracy of the estimated AGB on the test data

was not quite satisfactory. This is possibly resulting from the high heterogeneity and complexity present in the forests investigated, a high number of species, and steep slopes, which probably affected the performance of the models. This unfavorable matter might also be due to the geolocation uncertainty in GEDI measurements, as well as the fact that the AGBD index derived from GEDI L4A may be biased owing to the weak ground-returned signals under very dense canopy cover conditions.
From Figure 6, while the highest observed AGBD in the study regions reached 341.13 Mg ha⁻¹, our model predictions extended to 225.09 Mg ha⁻¹, indicating that the model tends to saturate, and cannot accurately predict the extremely high values. Conversely, the minimum AGBD value among the plots is 6.10 Mg ha⁻¹, while our model's lowest prediction is 12.42 Mg ha⁻¹, representing a slight overestimation at lower AGBD values (Figure 6).

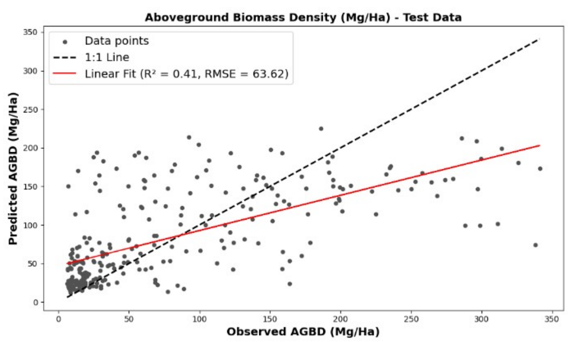

Figure 6. Evaluation of predicted versus observed AGBD for the testing dataset using R^2 and RMSE.

Raster maps of AGBD estimates in protected forests (Figure 7) reveal notable deforestation in western Lisar and eastern Siahroud-e-Roudbar, while other areas show similar biomass levels. Total AGB values are summarized in Table 1.

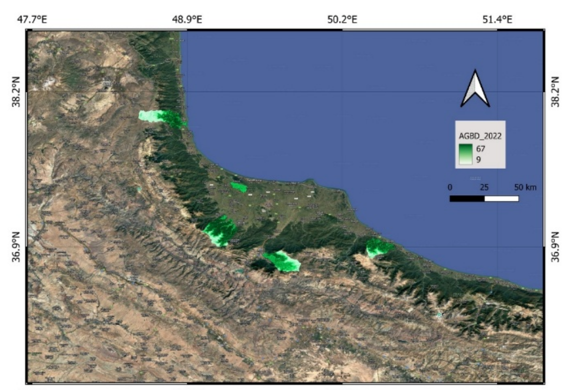

Figure 7. AGBD maps of the study area.

Protected forest region	Total AGB (Mg)
Siahroud e Roudbar	1046550.28
Siahkeshim	216391.12
Sarvelat and Javaherdasht	957170.68
Lisar	1142972.52
Gasht-e-Rodkan	1878836.09

Table 1. The total AGB estimates in the protected forest regions of Gilan.

4. Discussion

The main goal of this work was to evaluate the performance of GEDI-derived estimations of above-ground biomass density (AGBD) across the forest sites in Iran. To this end, for the first time, we applied an integrated method using the dataset of GEDI, Sentinel-2, and SRTM to monitor the spatial variability of AGBD in the Gilan protected areas, where no in-situ or airborne LiDAR measurements are accessible. Among a variety of machine learning algorithms, we have selected the Random Forest due to the demonstrated high predictive accuracy with minimal risk of overfitting, resulting in an ensemble-based approach for effective biomass estimation as the remote sensing applications.
Our results indicate that the AGBD estimates from the footprint-based GEDI L4A data could potentially be used to map and monitor the AGBD across the Gilan forests. However, this would require that the derived estimates are calibrated, for example using the field data from an independent sampling. It has also been recommended to use the airborne LiDAR data, in combination with field observations, for the calibrating and validating of the GEDI data. Despite applying the recommended filters to the GEDI data and removal of the records that significantly deviated from the average GEDI biomass values, the investigation results confirm that improving the biomass estimation still requires further filtering and calibration process of the GEDI data using the in-situ and/or airborne LiDAR measurements.
Fusing the GEDI data with multi-source satellite data enables better estimation of the forests AGBD, eliminating the need for extensive ground-based data collection during the model developments. These peripheral data can include the elevation, and vegetation indices, and the (B4) band is more useful than the raw band for the AGBD modeling. As such, the combination of different types of remote sensing data enhances the accuracy of models and allows for more comprehensive biomass mapping.
This investigation suggests that the GEDI has remarkably the potential to work as a ground-truth data source for the forest AGB estimation at a certain level of accuracy. Overall, this study attempted to better understand the forest AGBD distribution and the spatial changes in terrestrial carbon fluxes, besides assessing the promising capabilities of the GEDI sensor in continuous monitoring of the AGBD variability in the Gilan forests from space.

5. Conclusion and Outlook

This study presents a Random Forest-based machine learning approach for estimating the above-ground biomass density (AGBD) in protected forests of Gilan Province, with an integration of freely available GEDI Level-4A, Sentinel-2, and SRTM datasets. Given the lack of either the ground-based or the airborne LiDAR-based tree height information, the prior information derived from the GEDI provided a highly valuable means for training the model. The results demonstrated the effectiveness of using the spaceborne LiDAR, in conjunction with the optical and elevation data, for biomass mapping and carbon stock assessment across large-scale regions. Although the region exhibits significant spatial heterogeneity and the availability of the reference data is currently limited, integration of the GEDI data offers promising opportunities for improving the accuracy of final estimates. The derived AGBD maps settle a robust foundation for forest management and climate change investigations in these regions of interest. Future work will focus on incorporating other environmental variables, validating the obtained results with the in-situ and airborne-LiDAR-based data wherever present, and correcting the GEDI geolocation errors to enhance the model reliability and broaden the applicability.

6. References

Alvites, C., O'Sullivan, H., Francini, S., Marchetti, M., Santopuoli, G., Chirici, G., Lasserre, B., Marignani, M., Bazzato, E., 2024: High-Resolution Canopy Height Mapping: Integrating NASA's Global Ecosystem Dynamics Investigation (GEDI) with Multi-Source Remote Sensing Data. *Remote Sensing*, *16*(7). https://doi.org/10.3390/rs16071281

Atkins, J. W., Aho, K. S., Chen, X., Elmore, A. J., Fiorella, R., Luo, W., Lombardozzi, D., Lunch, C., Manak, L., de Pablo, L. X., Myers-Pigg, A. N., Record, S., Qiu, T., Reed, S., Ruddell, B., Strange, B., Torrens, C. L., Yule, K., Richardson, A. D., 2025: Recommendations for developing, documenting, and distributing data products derived from NEON data. *Ecosphere*, *16*(1). https://doi.org/10.1002/ecs2.70159

Bayat, M., Masteali, S. H., Hosein, M., Hallaj, S., 2025: Using of modern land surveying and remote sensing methods with the help of field data in estimating the level of afforestation in the west of Guilan province. *Journal of Environmental Science Studies (in Persion)*, *10*(1), 9812–9822. https://doi.org/10.22034/jess.2024.457764.2255

Bhandari, K., Srinet, R., Nandy, S., 2024: Forest Height and Aboveground Biomass Mapping by synergistic use of GEDI and Sentinel Data using Random Forest Algorithm in the Indian Himalayan Region. *Journal of the Indian Society of Remote Sensing*, *52*(4), 857–869. https://doi.org/10.1007/s12524-023-01792-z

Ceccherini, G., Girardello, M., Beck, P. S. A., Migliavacca, M., Duveiller, G., Dubois, G., Avitabile, V., Battistella, L., Barredo, J. I., Cescatti, A. 2023: Spaceborne LiDAR reveals the effectiveness of European Protected Areas in conserving forest height and vertical structure. Communications Earth and Environment, 4(1), 1–14. https://doi.org/10.1038/s43247-023-00758-w

Gupta, R., Sharma, L. K., 2022: Aboveground Biomass Prediction by Fusing Gedi Footprints with Optical and SAR Data Using the Random Forest in the Mixed Tropical Forest, India. International Geoscience and Remote Sensing Symposium (IGARSS), 2022-July, 5460–5463. https://doi.org/10.1109/IGARSS46834.2022.9883443

Heisig, J., Milenkovi´cmilenkovi´c, M., Pebesma, E., 2025: High-Resolution Canopy Fuel Maps Based on GEDI: A Foundation for Wildfire Modeling in Germany High-Resolution Canopy Fuel Maps Based on GEDI: A Foundation for Wildfire Modeling in Germany2. 4. https://doi.org/10.1088/2752-664X/adaaf9

Hoffrén, R., Lamelas, M. T., de la Riva, J., Domingo, D., Montealegre, A. L., García-Martín, A., Revilla, S., 2023: Assessing GEDI-NASA system for forest fuels classification using machine learning techniques. *International Journal of Applied Earth Observation and Geoinformation*, *116*. https://doi.org/10.1016/j.jag.2022.103175

Holcomb, A., Burns, P., Keshav, S., Coomes, D. A., 2024: Repeat GEDI footprints measure the effects of tropical forest disturbances. *Remote Sensing of Environment*, *308*. https://doi.org/10.1016/j.rse.2024.114174

Indirabai, I., Nilsson, M., 2024: Estimation of above ground biomass in tropical heterogeneous forests in India using GEDI. *Ecological Informatics*, *82*. https://doi.org/10.1016/j.ecoinf.2024.102712

Kacic, P., Gessner, U., Hakkenberg, C. R., Holzwarth, S., Müller, J., Pierick, K., Seidel, D., Thonfeld, F., Torresani, M., Kuenzer, C., 2025: Characterizing local forest structural complexity based on multi-platform and -sensor derived indicators. *Ecological Indicators*, *170*. https://doi.org/10.1016/j.ecolind.2025.113085

Kombate, A., Fotso Kamga, G. A., Goïta, K., 2025: Modeling Canopy Height of Forest–Savanna Mosaics in Togo Using ICESat-2 and GEDI Spaceborne LiDAR and Multisource Satellite Data. *Remote Sensing*, *17*(1). https://doi.org/10.3390/rs17010085

Lang, N., Jetz, W., Schindler, K., Wegner, J. D., 2023: A high-resolution canopy height model of the Earth. *Nature Ecology and Evolution*, *7*(11), 1778–1789. https://doi.org/10.1038/s41559-023-02206-6

Li, H., Zhao, J., Yan, B., Yue, L., Wang, L., 2022: Global DEMs vary from one to another: an evaluation of newly released Copernicus, NASA and AW3D30 DEM on selected terrains of China using ICESat-2 altimetry data. *International Journal of Digital Earth*, *15*(1), 1149–1168. https://doi.org/10.1080/17538947.2022.2094002

Li, Y., Shen, X., Cao, L., 2024: An Advanced Terrain Vegetation Signal Detection Approach for Forest Structural Parameters Estimation Using ICESat-2 Data. *Remote Sensing*, *16*(11). https://doi.org/10.3390/rs16111822

Liu, X., Neigh, C. S. R., Pardini, M., Forkel, M., 2024: Estimating forest height and above-ground biomass in tropical forests using P-band TomoSAR and GEDI observations. *International Journal of Remote Sensing*, *45*(9), 3129–3148. https://doi.org/10.1080/01431161.2024.2343134

Mansouri, J., Jafari, M., Taheri Dehkordi, A., 2024: Continuous mapping of forest canopy height using ICESat-2 data and a weighted kernel integration of multi-temporal multi-source remote sensing data aided by Google Earth Engine. *Environmental Science and Pollution Research*, *31*(37), 49757–49779. https://doi.org/10.1007/s11356-024-34415-2

Milenković, M., Reiche, J., Armston, J., Neuenschwander, A., De Keersmaecker, W., Herold, M., Verbesselt, J., 2022: Assessing Amazon rainforest regrowth with GEDI and ICESat-2 data. *Science of Remote Sensing*, *5*(April). https://doi.org/10.1016/j.srs.2022.100051

Mohite, J., Sawant, S., Pandit, A., Sakkan, M., Pappula, S., Parmar, A., 2024: Forest aboveground biomass estimation by GEDI and multi-source EO data fusion over Indian forest. *International Journal of Remote Sensing*, *45*(4), 1304–1338. https://doi.org/10.1080/01431161.2024.2307944

Moudrý, V., Prošek, J., Marselis, S., Marešová, J., Šárovcová, E., Gdulová, K., Kozhoridze, G., Torresani, M., Rocchini, D., Eltner, A., Liu, X., Potůčková, M., Šedová, A., Crespo-Peremarch, P., Torralba, J., Ruiz, L. A., Perrone, M., Špatenková, O., Wild, J., 2024: How to Find Accurate Terrain and Canopy Height GEDI Footprints in Temperate Forests and Grasslands? *Earth and Space Science*, *11*(10), 1–26. https://doi.org/10.1029/2024EA003709

Musthafa, M., Singh, G., Kumar, P., 2023: Comparison of forest stand height interpolation of GEDI and ICESat-2 LiDAR measurements over tropical and sub-tropical forests in India. *Environmental Monitoring and Assessment*, *195*(1). https://doi.org/10.1007/s10661-022-10657-w

Ortiz-Reyes, A. D., Barrera-Ortega, D., Velasco-Bautista, E., Romero-Sánchez, M. E., Correa-Díaz, A., 2024: Predicting forest parameters through generalized linear mixed models using GEDI metrics in a temperate forest in Oaxaca, Mexico. *International Journal of Remote Sensing*, *45*(22), 8037–8060. https://doi.org/10.1080/01431161.2024.2396565

Ouaknine, A., Kattenborn, T., Laliberté, E., Rolnick, D., 2023: *OpenForest: A data catalogue for machine learning in forest monitoring*. https://github.com/RolnickLab/OpenForest

Pascual, A., Grau-Neira, A., Morales-Santana, E., Cereceda-Espinoza, F., Pérez-Quezada, J., Cárdenas Martínez, A., Fuentes-Castillo, T., 2024: Old-growth mapping in Patagonia's evergreen forests must integrate GEDI data to overcome NFI data limitations and to effectively support biodiversity conservation. *Forest Ecology and Management*, *568*. https://doi.org/10.1016/j.foreco.2024.122059

Pascual, A., Guerra-Hernández, J., Armston, J., Minor, D. M., Duncanson, L. I., May, P. B., Kellner, J. R., Dubayah, R., 2023: Assessing the performance of NASA's GEDI L4A footprint aboveground biomass density models using National Forest Inventory and airborne laser scanning data in Mediterranean forest ecosystems. *Forest Ecology and Management*, *538*. https://doi.org/10.1016/j.foreco.2023.120975

Potapov, P., Li, X., Hernandez-Serna, A., Tyukavina, A., Hansen, M. C., Kommareddy, A., Pickens, A., Turubanova, S., Tang, H., Silva, C. E., Armston, J., Dubayah, R., Blair, J. B., Hofton, M., 2021: Mapping global forest canopy height through integration of GEDI and Landsat data. *Remote Sensing of Environment*, *253*. https://doi.org/10.1016/j.rse.2020.112165

Pronk, M., Eleveld, M., Ledoux, H., 2024: Assessing Vertical Accuracy and Spatial Coverage of ICESat-2 and GEDI Spaceborne Lidar for Creating Global Terrain Models. *Remote Sensing*, *16*(13), 1–20. https://doi.org/10.3390/rs16132259

Quiros, E., Polo, M. E., Fragoso-Campon, L., 2021: GEDI Elevation Accuracy Assessment: A Case Study of Southwest Spain. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*, 5285–5299. https://doi.org/10.1109/JSTARS.2021.3080711

S.Abdollahi, H. D., 2019: Analysis and Prediction of Land Use/Land Cover Changes in the Coastal Areas of Gilan Province. *Environmental Sciences(in Persion)*, *3*, 121–136.

Shannon, E. S., Finley, A. O., Hayes, D. J., Weiskittel, A. R., Cook, B. D., Babcock, C., 2023: Quantifying and correcting geolocation error in spaceborne LiDAR forest canopy observations using high spatial accuracy data : A Bayesian model approach. *WILEY*. https://doi.org/10.1002/env.2840

Travers-Smith, H., Coops, N. C., Mulverhill, C., Wulder, M. A., Ignace, D., Lantz, T. C., 2024: Mapping vegetation height and identifying the northern forest limit across Canada using ICESat-2, Landsat time series and topographic data. *Remote Sensing of Environment*, *305*(March), 114097. https://doi.org/10.1016/j.rse.2024.114097

Vogeler, J. C., Fekety, P. A., Elliott, L., Swayze, N. C., Filippelli, S. K., Barry, B., Holbrook, J. D., Vierling, K. T., 2023: Evaluating GEDI data fusions for continuous characterizations of forest wildlife habitat. *Frontiers in Remote Sensing*, *4*(June), 1–20. https://doi.org/10.3389/frsen.2023.1196554

Wang, C., Zhang, W., Ji, Y., Marino, A., Li, C., Wang, L., Zhao, H., Wang, M., 2024: Estimation of Aboveground Biomass for Different Forest Types Using Data from Sentinel-1, Sentinel-2, ALOS PALSAR-2, and GEDI (Forests, (2024), 15, 1, (215), 10.3390/f15010215). *Forests*, *15*(3). https://doi.org/10.3390/f15030401

Wang, J., Shen, X., Cao, L., 2024: Upscaling Forest Canopy Height Estimation Using Waveform-Calibrated GEDI Spaceborne LiDAR and Sentinel-2 Data. *Remote Sensing*, *16*(12). https://doi.org/10.3390/rs16122138

Weinstein, B. G., Graves, S. J., Marconi, S., Singh, A., Zare, A., Stewart, D., Bohlman, S. A., White, E. P., 2021: A benchmark dataset for canopy crown detection and delineation in co-registered airborne RGB, LiDAR and hyperspectral imagery from the National Ecological Observation Network. *PLoS Computational Biology*, *17*(7). https://doi.org/10.1371/journal.pcbi.1009180

Weinstein, B. G., Marconi, S., Bohlman, S. A., Zare, A., Singh, A., Graves, S. J., White, E. P., 2021: A remote sensing derived data set of 100 million individual tree crowns for the national ecological observatory network. *ELife*, *10*, 1–18. https://doi.org/10.7554/eLife.62922

Xu, L., Yu, J., Shu, Q., Luo, S., Zhou, W., Duan, D., 2024: Forest aboveground biomass estimation based on spaceborne LiDAR combining machine learning model and geostatistical method. *Frontiers in Plant Science*, *15*(December), 1–18. https://doi.org/10.3389/fpls.2024.1428268

Yang, Q., Niu, C., Liu, X., Feng, Y., Ma, Q., Wang, X., Tang, H., Guo, Q., 2023: Mapping high-resolution forest aboveground biomass of China using multisource remote sensing data. *GIScience and Remote Sensing*, *60*(1). https://doi.org/10.1080/15481603.2023.2203303

Zhu, X., Nie, S., Zhu, Y., Chen, Y., Yang, B., Li, W., 2023: Evaluation and Comparison of ICESat-2 and GEDI Data for Terrain and Canopy Height Retrievals in Short-Stature Vegetation. *Remote Sensing*, *15*(20). https://doi.org/10.3390/rs15204969