Geo-spatial Information Science

ISSN: 1009-5020 (Print) 1993-5153 (Online) Journal homepage: www.tandfonline.com/journals/tgsi20

Evaluating the accuracy of ground elevation estimates from GEDI satellite LiDAR in forested terrain

Mihnea Cățeanu

GEO-SPATIAL INFORMATION SCIENCE
https://doi.org/10.1080/10095020.2025.2555385

Evaluating the accuracy of ground elevation estimates from GEDI satellite LiDAR in forested terrain

Mihnea Căţeanu

Department of Forest Engineering, Faculty of Silviculture and Forest Engineering, Transilvania University of Braşov, Braşov, Romania

ABSTRACT

The goal of this study is to assess the accuracy of ground elevation estimates derived from data collected by the Global Ecosystem Dynamics Investigation (GEDI) full-waveform satellite LiDAR sensor, for a study area of approximately 246 km^2 located in northwestern Romania and with a significant broadleaved forest cover. GEDI L2A footprints were filtered, resulting in 32,666 valid footprints. The analysis involved comparing GEDI-derived ground elevation with reference data collected by Airborne Laser Scanning (ALS), which offers superior vertical accuracy. Outlier values, potentially caused by degraded pointing/positioning data from GEDI's auxiliary systems or cloud reflections, were identified using modified *z*-score, leading to the removal of 2.4% of observations. The study further explored factors influencing GEDI performance of ground surface retrieval, such as: the presence of forest vegetation, its height and type, use of coverage vs. power beam data, the time and season of data collection and geomorphological conditions (slope and terrain ruggedness). Overall, the accuracy is reasonably good (with an RMSE of 6.75 and a bias of – 0.53 m, after outlier removal), with a significant reduction of ground surface estimation in forested areas (RMSE of 7.75, versus 5.29 m in open ground). Accuracy is mainly influenced by the season of acquisition (RMSE between 5.45 and 8.75 m, the presence and height of forest canopy (RMSE of 6.04 m for canopies under 9.2 m and RMSE of 9.81 m for canopies taller than 22.7 m) and geomorphological conditions (RMSE between 5.61 m and 9.18 m, depending on slope). Understanding the interaction of GEDI with these factors is essential for improving the utilization of GEDI data in forest ecosystem monitoring. This research underscores the importance of accurate ground modeling in forestry applications and contributes to the ongoing evaluation of GEDI performance.

ARTICLE HISTORY
Received 10 September 2024
Accepted 27 August 2025

KEYWORDS
Global Ecosystem Dynamics Investigation (GEDI); light detection and ranging (LiDAR); remote sensing; ground surface estimation; airborne laser scanning

1. Introduction

Accurate models of the ground surface are essential for applications in geosciences, such as hydrology (Jarihani et al. 2015), geomorphology (Müller et al. 2014) or ecology (Urbazaev et al. 2022). In the last few decades, Light Detection and Ranging (LiDAR) has been established as a primary source of high-accuracy, high-resolution elevation data (Zhang et al. 2003), especially at a local or regional level. In forestry, one of the first applications of LiDAR data involved the estimation of various stand-level forest parameters, such as mean tree height (Næsset 1997a), timber volume (Næsset 1997b) or the number of stems (Næsset and Bjerknes 2001). Since then, with the increased availability of airborne and terrestrial laser scanners (Deems, Painter, and Finnegan 2013), LiDAR has produced a revolution in forestry (Coops et al. 2021), fundamentally changing the manner in which forest structure is observed and described (Beland et al. 2019). This development was not only driven by the improvement of laser scanners (in terms of accuracy, level of detail or costs), but is also due to the unique property of laser pulses, which is that they can penetrate the forest canopy and reach the ground underneath it. Therefore, LiDAR provides a detailed and highly accurate model of the vertical structure of the forest (Lim et al. 2003) and the ground underneath it, in even in dense canopies (Căţeanu and Ciubotaru 2021). The main field of forestry in which LiDAR has been widely adopted is the estimation of forest structure (Lim et al. 2003) or various vegetation parameters (Adam et al. 2020), thus enabling the description, investigation and monitoring of forest ecosystems. Thus, LiDAR has been established as a key data source in the prediction of forest attributes (Coops et al. 2021; Lim et al. 2003).

Currently there are three types of LiDAR instruments that see active use in forestry activities and research: ground (Dassot, Constant, and Fournier 2011), airborne (Akay et al. 2009) which is currently

CONTACT Mihnea Căţeanu cateanu.mihnea@unitbv.ro
Supplemental data for this article can be accessed online at https://doi.org/10.1080/10095020.2025.2555385

dominating applications (Qi and Dubayah 2016) and space-based systems (Silva et al. 2021), all of which have opportunities and challenges associated with them (Beland et al. 2019). The active remote sensing of forests, especially at large or even global scales, has entered an exciting new era (Duncanson et al. 2020) since 2018, when two space-based LiDAR systems have been put into operation by the National Aeronautics and Space Administration (NASA): the Ice, Cloud and Land Elevation Satellite-2 (ICESat-2) carrying a photon-counting instrument called the Advanced Topographic Laser Altimeter System (ATLAS) and the Global Ecosystem Dynamics Investigation (GEDI) which is a full-waveform LiDAR instrument mounted on the International Space Station (ISS). While ICESat-2 was developed mainly for monitoring glaciers and ice sheets (Markus et al. 2017), it has been used in various other applications, including forestry (He et al. 2023; Narine, Popescu, and Malambo 2020; Popescu et al. 2018), with varying degrees of success. On the other hand, GEDI is the first spaceborne laser system designed specifically to measure the forest structure (Patterson et al. 2019) and aboveground biomass (Silva et al. 2018), thus opening up a new era for macroecology (Schneider et al. 2020). Since the availability of satellite data offers opportunities for large-scale studies (Dorado-Roda et al. 2021), GEDI has enabled a "change of paradigm" (Guerra-Hernández and Pascual 2021) in the monitoring of forest ecosystems.

GEDI works by emitting laser pulses that illuminate an area of approximately 25-m in diameter (commonly known as a "footprint"), with the instrument detecting and storing the variation of returned energy as a function of time (commonly known as a "waveform"). The waveform is digitized by equal-interval sampling and, in the case of data collected over forested areas, can be understood as a record of the vertical structure of the forest (Qi, Saarela, et al. 2019), from the top of the canopy to the ground-level. The GEDI system is equipped with 3 laser emitters, one of which is split into two beams (called "coverage" beams, with the others referred to as "power" beams). In-depth descriptions of the technical characteristics of the GEDI mission have been previously published, so the reader is referred to the relevant materials (Dubayah et al. 2020; Roy, Kashongwe, and Armston 2021; Urbazaev et al. 2022). Even before GEDI's launch, the system's capabilities have been tested by simulation (Patterson et al. 2019; Qi and Dubayah 2016; Qi, Lee, et al. 2019; Qi, Saarela. et al. 2019; Silva et al. 2018), using the GEDI waveform simulator developed by Hancock et al. (2019). However, the real, post-launch accuracy of GEDI products is an important consideration (Dorado-Roda et al. 2021), as real accuracy can be lower than established by simulation due, for example, to spatio-temporal variations of atmospheric attenuation (Rishmawi, Huang, and Zhan 2021). Therefore, once GEDI L2A data was released, a significant number of studies have been carried out worldwide, concerning the accuracy assessments of canopy height, biomass and terrain elevation (Adam et al. 2020; Dorado-Roda et al. 2021; Guerra-Hernández and Pascual 2021; Kutchartt, Pedron, and Pirotti 2022; Li et al. 2024, 2023; Liu, Cheng, and Chen 2021; Quiros, Polo, and Fragoso-Campon 2021; Urbazaev et al. 2022). In general, the studies focusing on GEDI performance in forest ecosystems are mainly concerned with canopy height or biomass estimation, rather than the accuracy of ground surface retrieval, which is understandable given that models of biomass distribution or canopy height are products that see use in forestry applications. However, it must be noted that the accuracy of ground elevation has a direct impact on the accuracy of canopy height estimation (Liu, Cheng, and Chen 2021), since canopy heights are calculated by subtracting the ground elevation from the elevation recorded at the top of the canopy. Canopy height will further serve as the basis for estimations of Above Ground Biomass (ABG), timber volume or enable the monitoring of forest degradation and restoration (Potapov et al. 2021).

Therefore, recognizing that errors in terrain elevation are the main driver of inaccuracies in forestry products (Urbazaev et al. 2022), the aim of this study is to present an analysis of ground elevation estimates in the latest GEDI products version 2 and to identify some of the factors affecting accuracy. While the analysis also considers the general accuracy across the study area, the focus of research is ground data collected inside the forest. Several factors related to the GEDI system, forest and geomorphological conditions are also considered.

2. Materials and methods

2.1. Study area

The study area is located in the north-western part of Romania, covering approximately 246 km^2 (Figure 1). The area is characterized by high variations both in terms of altitude (ranging from 450 m to 1600 m above Mean Sea Level) and in terms of steepness, with relief ranging from the wide, relatively flat valley of the Bistra river which crosses the study area from East to West, to steep slopes (up to 50%–60%) overlooking it to the North and South. The area is predominantly covered by forests (approximately 50%), with broadleaved forests being predominant (approximately 43% of the area), and coniferous or mixed forest occupying much smaller areas (approximately 3% and 5%, respectively). In broadleaved forests Beech (*Fagus sylvatica*) is dominant with European hornbeam (*Carpinus beletum*) present to

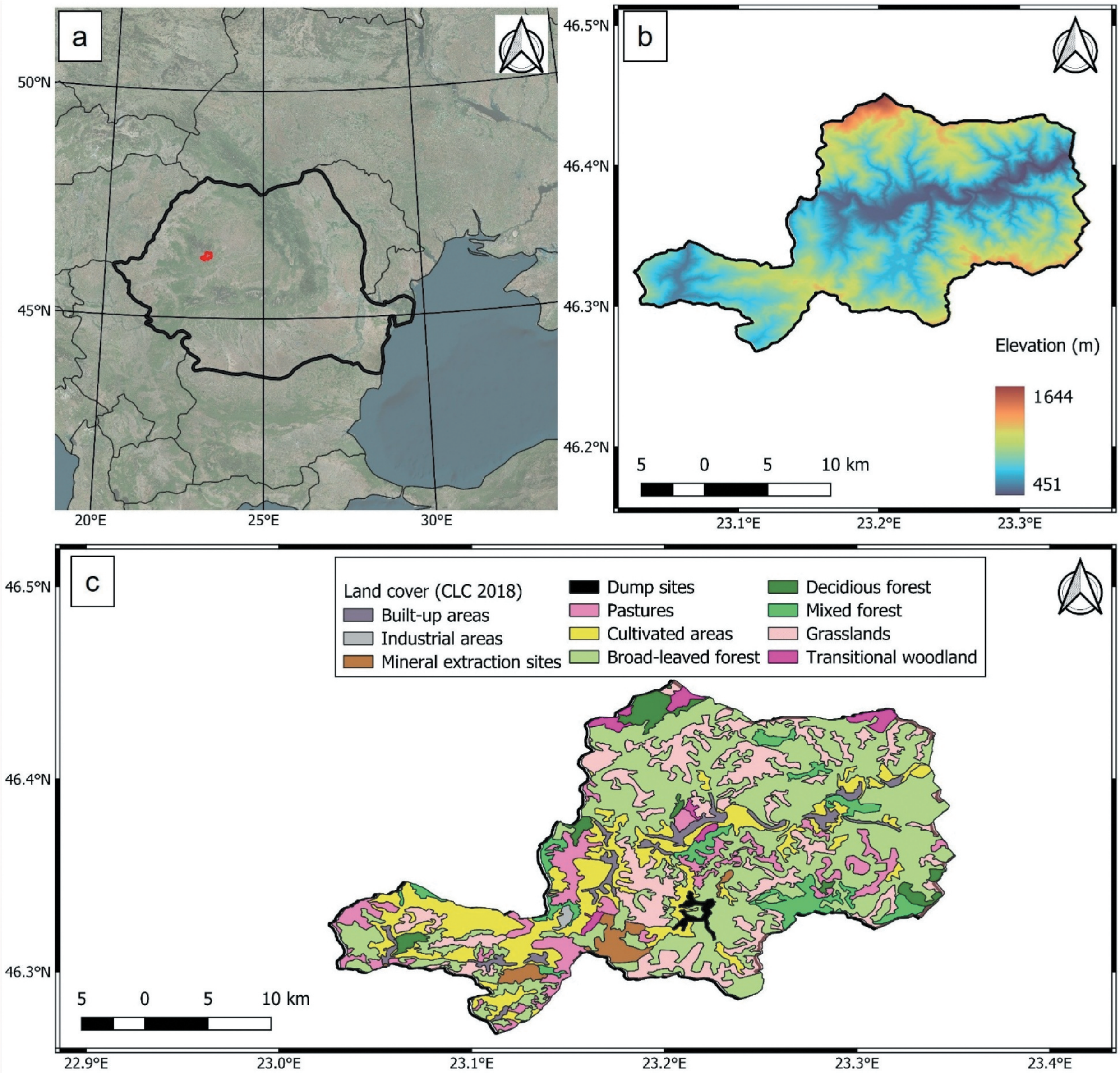

Figure 1. (a) Location of study area in the Romanian territory; (b) DTM of the study area; (c) land cover of the study area based on corine land cover data, 2018.

a lesser extent, while Norway spruce (*Picea abies*) and Pine (*Pinus sylvestris*) are present in coniferous areas. The other predominant land cover is that of pastures and natural grasslands, which occupy approximately 26% of the area. The choice of study area was based on the availability of Airborne Laser Scanning (ALS) data to use as reference for ground elevation, but also with the consideration that exploring the limitations of GEDI products in complex scenarios of terrain conditions is an important step in determining the full potential of the technology (Kutchartt, Pedron, and Pirotti 2022).

2.2. Data

2.2.1. GEDI data

76 GEDI L2A datasets were considered, representing all granules acquired over the study area between April 2019 and February 2023. Only GEDI L2A version 2 data was included in the analysis. Initial GEDI data products (version 1) had a reported geolocation error of 23.80 m (mean 1-sigma) (Beck et al. 2021), but version 2 was released in April 2021, with an improved algorithm for correcting geolocation errors. This led to a stated mean 1-sigma error of 10.2 m (Beck et al. 2021), which was later confirmed empirically (Schleich et al. 2023; Tang et al. 2023). 83874 GEDI footprints were identified as located inside the study area. Due to the constantly varying orbit of the ISS and the fact that almost 3 years of GEDI data collection were included in the analysis, GEDI footprints cover the study area almost entirely; however, there is quite a significant difference in terms of footprint density, with the northern half of the

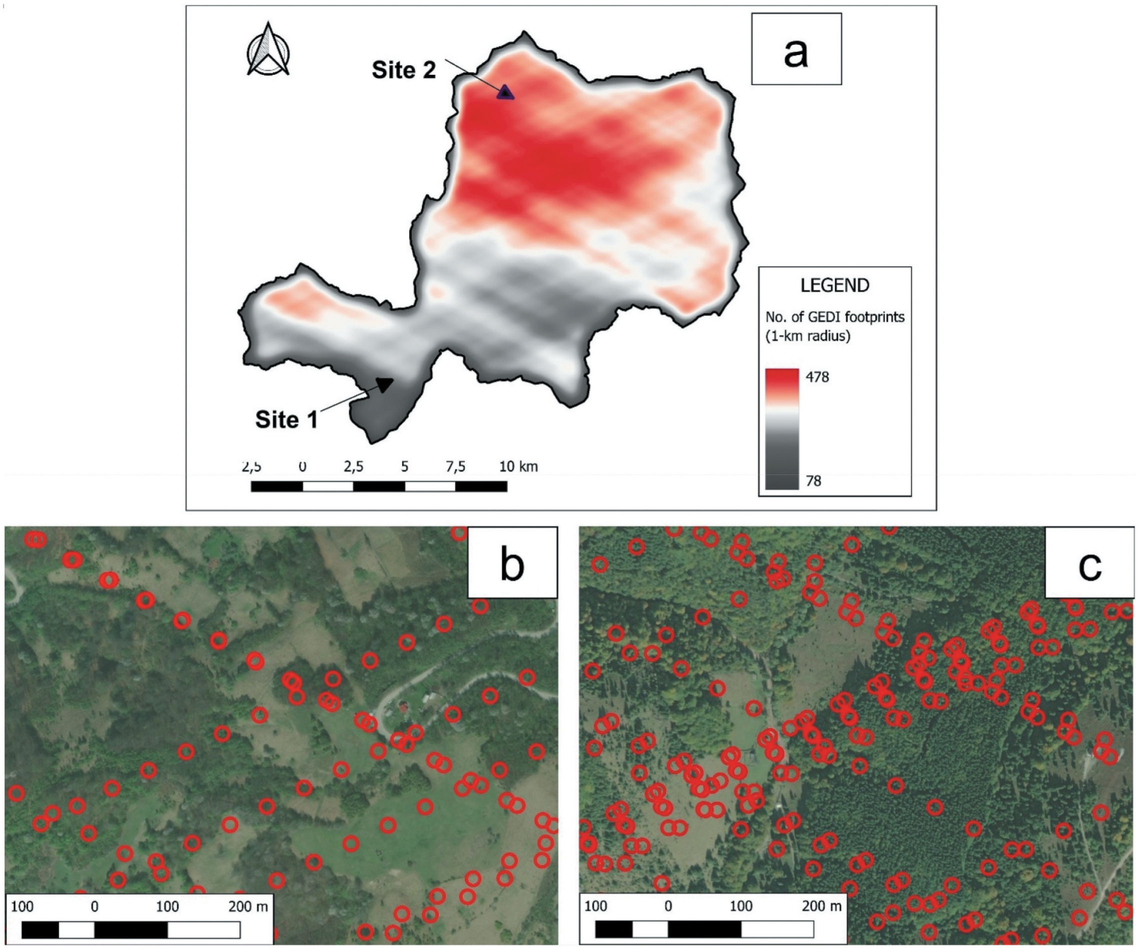

Figure 2. (a) Heatmap illustrating the spatial variation of GEDI footprint density, generated using a 1 km radius for aggregation of observations; (b) site 1: an example of an area with relatively low coverage by GEDI orbits; (c) site 2: an example of an area with relatively high coverage by GEDI orbits. Note that the two sites are also indicated in (a) with arrows.

study area exhibiting much better coverage by GEDI tracks (Figure 2). Positioning information related to the ground surface is obtained from three attributes common to all GEDI footprints: *lat_lowestmode*, *lon_lowestmode* and *elev_lowest-mode*. In this case, "lowest mode" refers to the lowest (in terms of altitude) reflection identified on each waveform, which ideally corresponds to the bare earth surface.

GEDI footprints were filtered as follows:

(1) 389 footprints (0.5% of observations) were eliminated due to their location over a wastewater dumpsite which has a variable water level (represented in black in Figure 1(c)).
(2) Footprints flagged by the GEDI data provider as "degraded," indicating potential issues with the pointing or positioning systems (Beck et al. 2021), were eliminated; 17768 were affected by this, representing 21.2% of observations.
(3) Footprint not labeled as "quality" by the GEDI data provider (40,775 footprint or 48.6% of observations) were also eliminated; a "quality" flag value of 0 (false) indicates that the corresponding waveform has a poor signal quality (Wang et al. 2022), which leads to a higher degree of uncertainty for the extraction of ground elevation.

Since there is some overlap between degraded and no-quality footprints, filtering done in steps 2–3 above led to a removal of 51,090 observations. The initial dataset was reduced by 61%, with 32,666 footprints valid for further analysis.

2.2.2. ALS data

LiDAR data was collected by Airborne Laser Scanning (ALS) using a RiEGL Q780 scanner system mounted on a DA42 MPP plane (Picu 2018). The vertical accuracy of this dataset is 30 cm (Pârvu, Özdemir, and Remondino 2020), compared to the resolution of 70 cm at which GEDI can detect sub-canopy elevations (Qi and Dubayah 2016). Further, an accuracy of ±30 cm indicates a performance better than that of GEDI by at least a factor of 3, based on reported RMSE values for GEDI ground elevations (Guerra-Hernández and Pascual 2021; Li et al. 2024; Liu, Cheng, and Chen 2021; Quiros, Polo, and Fragoso-

Campon 2021; Wang et al. 2022). Therefore, the ALS dataset was considered sufficiently accurate to serve as reference data. LiDAR data was provided as classified point clouds, with an average ground point density of 2.7 points/m^2. Ground points were used for the interpolation of a DTM at a 1-m resolution.

2.3. Estimation of GEDI ground elevation accuracy

Each GEDI footprint has an associated "elev_lowest-mode" value, which corresponds to the elevation of the ground surface, as it was identified by processing the received waveform (Adam et al. 2020). Each footprint also has the geographical position (latitude and longitude) of the center of area covered under the footprint. This position was used to determine the ALS DTM elevation corresponding to the center of the footprint. Therefore, for each sample, two elevation values are known: GEDI elevation and the ALS elevation serving as reference. Based on these, several common statistical indices of accuracy were calculated: the Mean Unsigned Error (also called bias), Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and relative Root Mean Square Error (rRMSE).

$$\mathrm{Bias} = \sum_{i=1}^{n} \frac{(H_{\mathrm{GEDI}} - H_{\mathrm{DTM}})}{n} \tag{1}$$

$$\mathrm{MAE} = \sum_{i=1}^{n} \frac{|(H_{\mathrm{GEDI}} - H_{\mathrm{DTM}})|}{n} \tag{2}$$

$$\mathrm{RMSE} = \sqrt{\sum_{i=1}^{n} \frac{(H_{\mathrm{GEDI}} - H_{\mathrm{DTM}})^2}{n}} \tag{3}$$

$$rRMSE = \frac{RMSE}{\overline{H_{GEDI}}} \times 100 \tag{4}$$

where H_{GEDI} are the GEDI-estimated ground heights, H_{DTM} are the reference heights obtained from DTM interpolated from ALS data, n is the number of samples (GEDI footprints), and $\overline{H_{GEDI}}$ is the mean of GEDI elevation estimates. Note that the positive bias values indicate a tendency for GEDI to over-estimate the ground surface, while negative values indicate a tendency to under-estimate it.

2.4. Identification of outliers

Satellite LiDAR depends on multiple instruments and sensors working as intended, therefore there is a potential for very large errors affecting a very small part of observations. These outlier values could be caused by degraded pointing and positioning of the system, or even cloud reflections (Urbazaev et al. 2022) that are not identified by the GEDI waveform-processing algorithm. The necessity of outlier removal to increase the reliability of satellite LiDAR data has been established before (Baban and Niţă 2023) and previously published studies of GEDI accuracy have considered the impact of outliers and applied different methods to identify them, such as eliminating error values outside the range defined by either mean ±3×sigma (Wang et al. 2022) or by 1.5×*Interquartile Range (IQR)* (Quiros, Polo, and Fragoso-Campon 2021).

For this study, outlier values were identified using the modified *z*-score approach. While the *z*-score method is computationally fast and robust, the calculation of *z*-score is made using the mean value, which is sensitive to outliers. Therefore, the method used for outlier identification is itself influenced by the outliers (Yaro, Maly, and Prazak 2023). To address this issue, the modified *z*-score method uses the median and the Median Absolute Deviation (MAD), which are not influenced by outliers. The modified *z*-score is calculated using the formula:

$$M_{\mathrm{i}} = \frac{0.6745 \times (X_{\mathrm{i}} - \tilde{X})}{\mathrm{MAD}} \tag{5}$$

where M_i is the modified *z*-score for observation i, 0.6745 is a constant that approximates a median-equivalent of the standard deviation, X_i is the relative elevation error for observation i, $\tilde{X}$ is the median of X values, and MAD is equivalent to the median of the absolute deviations of values from their median:

$$\mathrm{MAD} = \mathrm{median}(|X_{\mathrm{i}} - \tilde{X}|)$$

Afterward, a threshold for *z*-score is applied that separates outliers from non-outliers. In this paper, the condition for an observation to be considered non-outlier was $-3.5 \le M_i \le 3.5$, as values threshold values of $3.0 - 3.5$ are commonly recommended for outlier identification (Iglewicz and Hoaglin 1993; Yaro, Maly, and Prazak 2023). Using this approach, elevation errors for 776 footprints (approximately 2.4% of filtered footprints) were identified as outliers: 170 negative outliers and 606 positive outliers (Figure 3).

2.5. Factors potentially affecting GEDI performance

Several external factors have been considered, as potentially explaining the variation of GEDI ground elevation errors. These factors can be classified as follows:

(1) Forest conditions: 1) the presence of forest vegetation, identified as area with Canopy Height Model (CHM) values higher than 5 m,

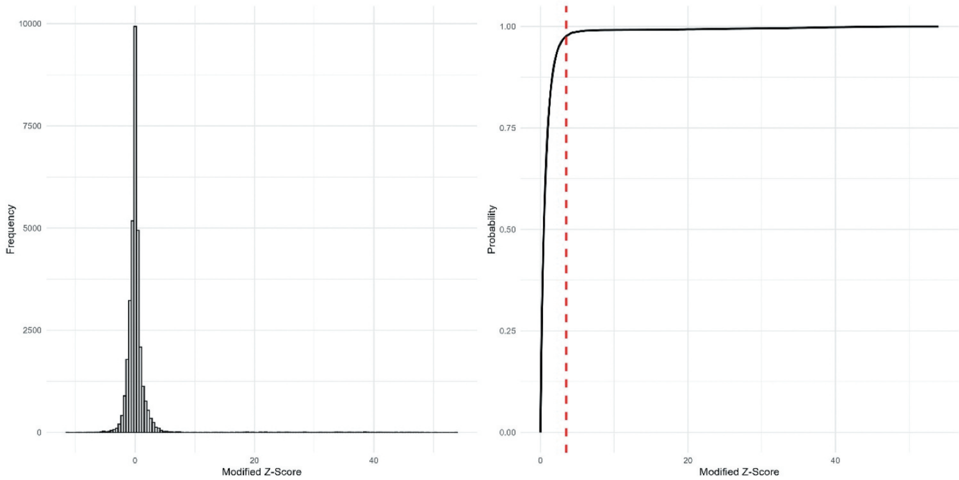

Figure 3. Distribution of modified *z*-score (left) and empirical cumulative distribution function (ECDF) plot of modified *z*-score (right). The red dashed line indicates the threshold for outlier identification (3.5).

a common threshold for separating forest from non-forest vegetation (Sasaki and Putz, 2009), 2) the height of forest vegetation, also estimated from the CHM and 3) the type of vegetation (broadleaf or coniferous, included in the L2A products).

(2) Conditions related to GEDI data collection: 1) beam type (coverage or power), 2) time of data collection (daytime or nighttime) and 3) season of acquisition; these are either directly available in the GEDI L2A products or easily derived from data associated with the products.

(3) Morphological conditions: 1) ground slope with a 3×3 Gaussian filter applied in order to smooth extreme values and 2) the ruggedness of the ground surface, as estimated by the Terrain Ruggedness Index (TRI) (Riley, DeGloria, and Elliot 1999).

2.6. Estimating the relative importance of factors

A hierarchy, based on the amount of error variability that is explained by the variance of factors, was done using the "calc.relimp" method from the statistical analysis R package "relaimpo" (Grömping 2006). This approach assumes that the relative order of factors included in a regression model that aims to explain the independent variable (i.e. GEDI elevation errors) corresponds to the amount of variance explained by each factor (Wang et al. 2022). In this manner, analysis of regression models with different orders of the same factors can provide an estimate of the relative importance of said factors (Grömping 2009). The analysis of relative importance was carried out for 3 samples: filtered GEDI footprints (n=32,666), footprints without outliers (n=31,890) and footprints attributed to forested areas (n=17,451).

3. Results

3.1. Overall accuracy of ground elevation

Summary statistics for GEDI ground detection accuracy are presented in Table 1, before and after the removal of outliers. The error distribution is shown in Figure 4, while Figure 5 presents the scatterplots showing the level of agreement between GEDI estimates of ground elevation and reference values.

As expected, outliers have a strong influence on goodness-of-fit statistics. For example, after outlier removal RMSE improves by 70%, from 23.00 m to 6.75 m. Due to the presence of mostly positive outliers, GEDI appears to overestimate ground elevation (bias of 1.53 m). However, once outliers are removed, the reverse is apparent, meaning a slight tendency to underestimate elevation (bias of −0.53 m).

3.2. GEDI accuracy in relation to external factors

The following sections will report the GEDI accuracy after stratification by factor categories. Please note that the values of continuous factors have been distributed over 5 classes using quantile breaks. Therefore, the number of samples per category, rather than the range of each category, is equal. Also note that only GEDI footprints remaining after data filtering and

Table 1. Summary statistics of the overall accuracy for GEDI for ground elevation estimates.

Sample	MAE (m)	Bias (m)	Median error (m)	Std. dev. (m)	Min. error (m)	Max. error (m)	RMSE (m)	rRMSE (%)	No. of samples
With outliers	7.06	1.53	−0.33	22.95	−80.15	383.66	23.00	2.50	32,666
After outlier removal	4.76	−0.53	−0.38	6.73	−25.20	24.57	6.75	0.80	31,890

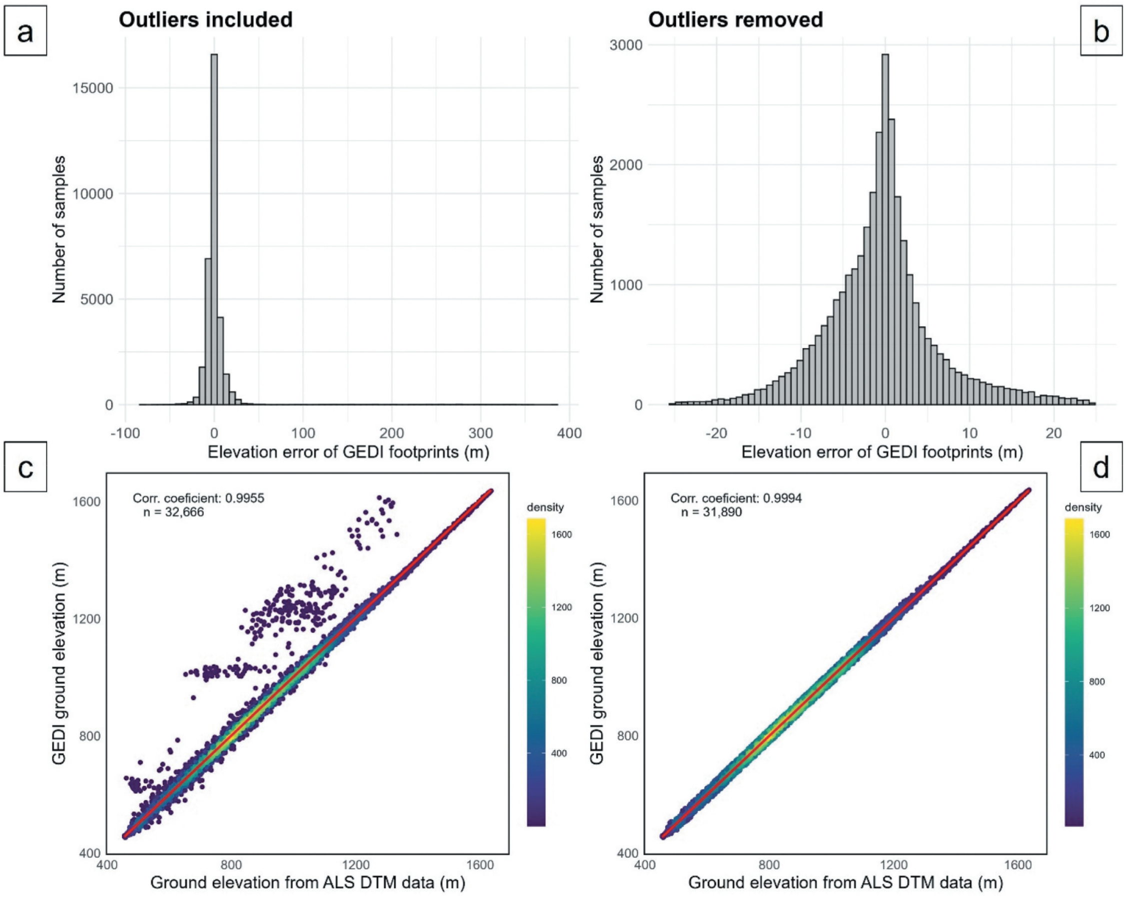

Figure 4. Distribution of GEDI elevation errors before (a) and after (b) outlier removal. The level of agreement between GEDI elevation estimates and reference data before (c) and after (d) outlier removal.

outlier removal (n = 31,890) are included in the analysis of factor influence.

3.2.1. Forest conditions

Three potential factors were considered in relation to forest vegetation: its presence, the type of forest (broadleaved or coniferous) and the canopy height (Table 2). Note that for vegetation type and height the number of samples is reduced, as only part of the footprints are under forest cover (n=17,451). Error statistics indicate that forest cover has an influence on accuracy, with MAE increasing from 3.47 m to 5.82 m, while the negative bias of −1.41 m changes to a less significant one of 0.19 m and RMSE increases by 47% (from 5.29 m in open ground to 7.75 m in forested terrain). This is also highlighted by the corresponding boxplot (Figure 5(a)). The elevation accuracy also varies depending on the type of vegetation (Figure 5(b)), but this is not statistically significant due to the very low number of samples located in conifers (n=117, less than 1% of forest footprints). The influence of vegetation height on GEDI performance is not significant and singular enough to show a significant correlation (r=0.21 – see Figure 5(c)), but once errors are stratified by vegetation height classes, the influence is more readily apparent. The accuracy of ground elevation estimation decreases as vegetation height increases, with RMSE values ranging from 6.04 m to 9.81 m between the lowest and highest forest heights (an increase of 62%). The increasing bias and the boxplots (Figure 5(d)) indicate that the lower accuracy is due mainly to an over-estimation of the ground surface as the canopy height increases.

3.2.2. GEDI operational conditions

The influence of beam type (coverage or power), the time (daytime or nighttime) and season of data acquisition were considered, for all valid footprints (n=31,890) but also only for footprints located under forest vegetation (n=17,451) (Table 3). Observations collected by power beams provide a marginal increase in accuracy, with MAE decreasing from 4.83 m to 4.69 m, while RMSE shows an insignificant decrease of 0.04 m. The type of beam appears to have a slightly more significant effect inside the forest canopy, in which case MAE improves by 4.5% (from 5.96 m to 5.69 m) and RMSE improves by 3% (from 7.88 m to 7.63 m). The presence of forest vegetation has a larger impact on ground elevation accuracy than the type of beam (Figure 6(a)). This is most clearly represented by

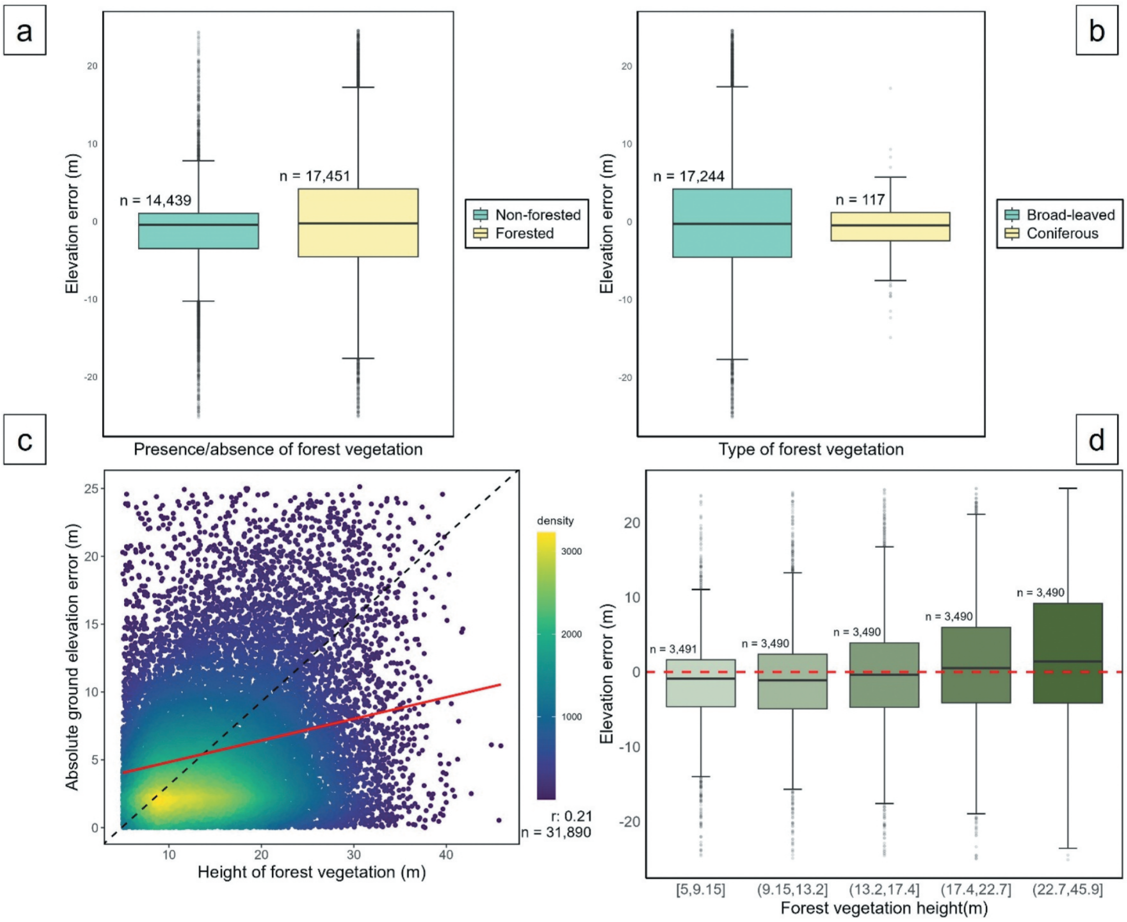

Figure 5. Variation of GEDI accuracy due to the presence of vegetation (a), its type (b) and its height (c, d).

Table 2. Variation of GEDI ground elevation accuracy dependent on forest conditions.

Potential factor	Factor category	MAE (m)	Bias (m)	Median error (m)	Std. dev. (m)	Min. error (m)	Max. error (m)	RMSE (m)	rRMSE (%)	No. of samples
Presence of forest vegetation (CHM>5 m)	Open ground	3.47	−1.41	−0.45	5.10	−25.20	24.35	5.29	0.62	14,439
	Forest veg.	5.82	0.19	−0.26	7.75	−25.13	24.57	7.75	0.91	17,451
Forest vegetation type	Broad-leaved	5.84	0.19	−0.27	7.76	−25.13	24.57	7.77	0.91	17,244
	Coniferous	3.12	−0.79	−0.46	4.46	−14.87	17.13	4.51	0.53	117
Forest vegetation height	<9.2 m	4.38	−1.45	−0.90	5.86	−24.58	23.61	6.04	0.71	3491
	9.2 *m*–13.2 m	5.09	−1.07	−1.12	6.71	−24.93	24.01	6.79	0.80	3490
	13.2 *m*–17.4 m	5.57	−0.24	−0.39	7.31	−24.58	24.40	7.31	0.85	3490
	17.4 *m*–22.7 m	6.35	1.10	0.51	8.18	−24.99	24.54	8.25	0.97	3490
	22.7 *m*–45.9 m	7.71	2.61	1.40	9.46	−25.13	24.57	9.81	1.15	3490

the Empirical Cumulative Distribution (ECDF) plot (Figure 6(b)), which shows that errors accumulate faster (therefore the accuracy is better) in open ground, regardless of beam type. However, in forested areas there is a slight increase in accuracy provided by power beams, across most of the range of error values.

A similar pattern is highlighted by the analysis of performance by time of acquisition. Again, the results indicate that the presence of forest vegetation has a more significant impact on accuracy (Figure 7(a)), but there is a slight increase in the accuracy of footprints located inside the forest for data registered during the night (Figure 7(b)).

The acquisition season was also considered in the analysis, given that the most common land cover for the study area was that of deciduous forests. As expected, the season does not appear to influence accuracy in open areas (Figure 8(a)); however, we do find a larger variation of errors in forested areas, depending on season. In forested terrain the lowest RMSE of 5.45 m corresponds to winter, while vegetative seasons (summer and autumn) exhibit RMSE values larger than 8 m

Table 3. Variation of GEDI ground elevation accuracy, dependent on operational conditions.

Beam type	MAE (m)	Bias (m)	Median error(m)	Stdev. (m)	Min. error(m)	Max. error(m)	RMSE(m)	rRMSE (%)	No. of samples
Coverage (all footprints)	4.83	−0.32	−0.50	6.76	−25.13	24.57	6.77	0.79	14,743
Coverage (forest footprints)	5.96	0.81	0.02	7.84	−25.13	24.57	7.88	0.92	8,228
Power (all footprints)	4.69	−0.72	−0.29	6.69	−25.20	24.55	6.73	0.79	17,147
Power (forest footprints)	5.69	−0.36	−0.54	7.62	−24.99	24.55	7.63	0.89	9,223
Time of acquisition	MAE (m)	Bias (m)	Median error(m)	Stdev. (m)	Min. error(m)	Max. error (m)	RMSE(m)	rRMSE (%)	No. of samples
Daytime (all footprints)	4.89	−0.48	−0.41	6.80	−25.13	24.57	6.82	0.80	16,519
Daytime (forest footprints)	6.00	0.26	−0.26	7.89	−25.13	24.57	7.89	0.92	9,218
Nighttime (all footprints)	4.61	−0.59	−0.37	6.65	−25.20	24.55	6.67	0.78	15,371
Nighttime (forest footprints)	5.62	0.11	−0.26	7.59	−24.99	24.55	7.59	0.89	8,233
Season of acquisition	MAE (m)	Bias (m)	Median error(m)	Stdev.(m)	Min. error(m)	Max. error (m)	RMSE (m)	rRMSE (%)	No. of samples
Winter (all footprints)	3.52	−2.36	−1.23	4.51	−24.93	24.35	5.09	0.60	2,449
Winter (forest footprints)	4.04	−2.86	−2.16	4.64	−24.93	21.77	5.45	0.64	1,321
Spring (all footprints)	4.30	−1.85	−1.14	5.69	−24.28	23.66	5.98	0.70	7,985
Spring (forest footprints)	5.25	−2.07	−1.85	6.61	−24.28	23.63	6.92	0.81	4,406
Summer (all footprints)	4.90	0.34	0.02	6.85	−25.13	24.57	6.86	0.80	16,421
Summer (forest footprints)	6.18	1.71	0.95	7.93	−25.13	24.57	8.11	0.95	9,226
Autumn (all footprints)	5.63	−0.40	−0.08	8.10	−25.20	24.54	8.11	0.95	5,035
Autumn (forest footprints)	6.43	0.17	−0.12	8.75	−24.99	24.54	8.75	1.03	2,498

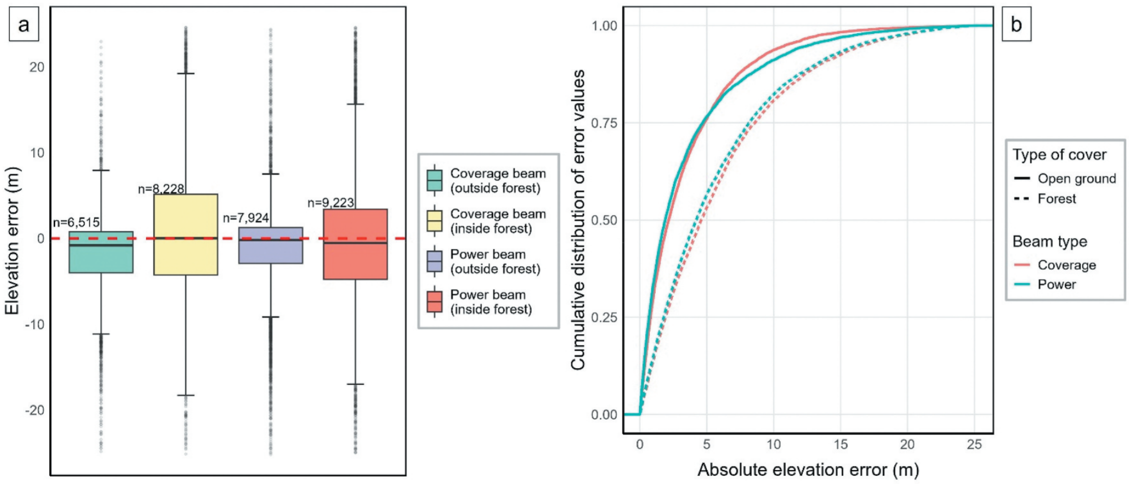

Figure 6. Boxplot (a) and ECDF plot (b) showing the variation of GEDI ground elevation errors by beam type, inside and outside of forest vegetation.

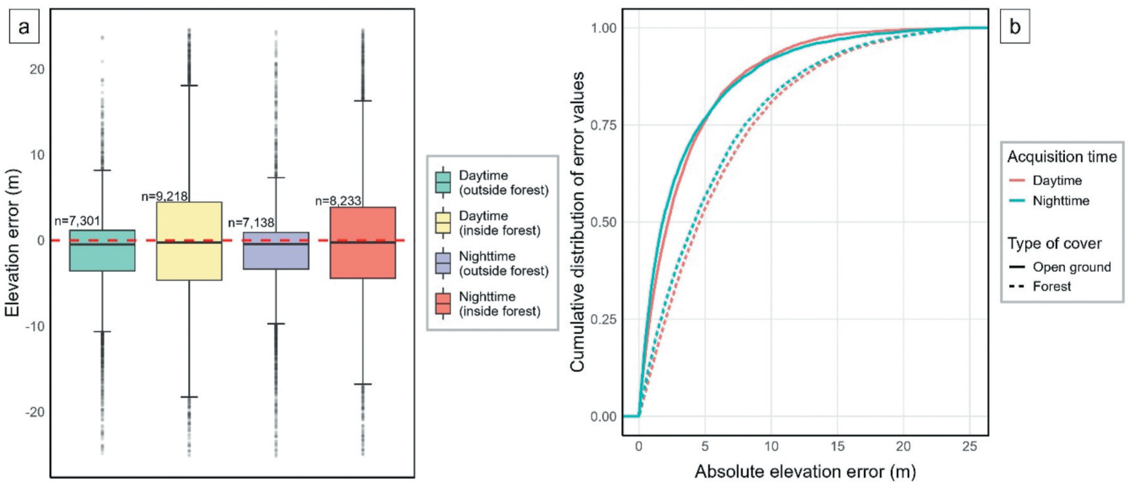

Figure 7. Boxplot (a) and ECDF plot (b) showing the variation of GEDI ground elevation errors by acquisition time, inside and outside of forest vegetation.

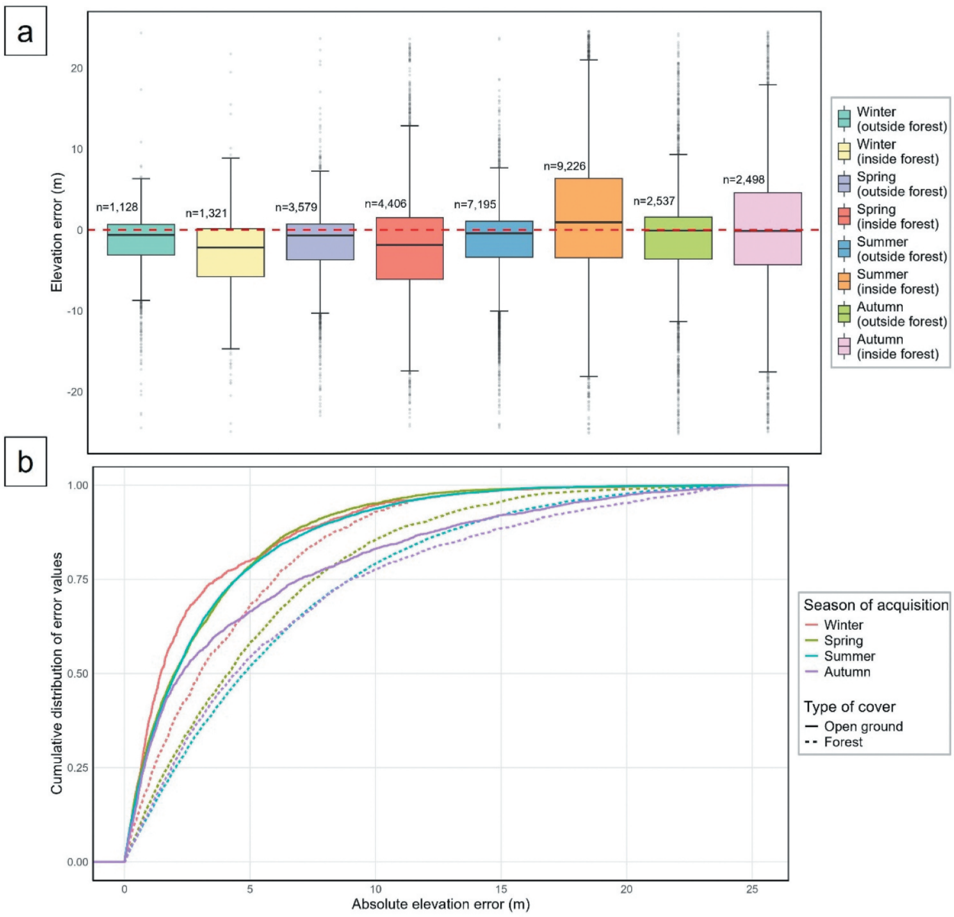

Figure 8. Boxplot (a) and ECDF plot (b) showing the variation of GEDI ground elevation errors by acquisition season, inside and outside of forest vegetation.

(Figure 8(a)). Also, the footprints collected in forest areas outside the vegetative season (winter or spring) have negative bias values, while footprints corresponding to the vegetative season (summer and autumn) have positive bias values. This appears to point to the presence of leaves as one of the factors leading to the overestimation of ground elevation.

The influence of acquisition season for forest footprints is also confirmed by the ECDF plot (Figure 8(b)), which shows that inside forest areas errors accumulate slower (therefore the accuracy is degraded) for data collected in summer or autumn. It should be mentioned that even if a footprint is classified as "bare earth," it could have vegetation present, just not as high as the 5-m threshold used for forest identification. This could explain the lower accuracy of bare ground footprints collected during autumn (Figure 8(b)).

3.2.3. Morphological conditions

Two morphological factors have been included in the analysis: ground slope and the ruggedness of the ground as estimated by the TRI. These can be

Table 4. Variation of GEDI ground elevation accuracy, dependent on ground conditions.

Ground slope (percent)	MAE (m)	Bias (m)	Median error (m)	Std. dev. (m)	Min. error (m)	Max. error (m)	RMSE (m)	rRMSE (%)	No. of samples
<18.3	3.70	0.06	0.01	5.61	−24.58	24.09	5.61	0.66	3,491
18.3–24.4	5.41	0.22	−0.43	7.28	−24.99	24.57	7.28	0.85	3,490
24.4–29.6	5.91	0.31	−0.57	7.70	−24.86	24.55	7.71	0.90	3,490
29.6–35.0	6.70	0.41	−0.40	8.49	−24.51	24.27	8.50	1.00	3,490
>35.0	7.38	− −0.04	−0.73	9.18	−25.13	24.55	9.18	1.07	3,490
TRI	MAE (m)	Bias (m)	Median error (m)	Stdev (m)	Min. error (m)	Max. error (m)	RMSE (m)	rRMSE (%)	No. of samples
Low ruggedness	3.92	0.03	−0.02	5.89	−24.58	24.09	5.89	0.69	3,491
Medium-low ruggedness	5.44	0.23	−0.46	7.31	−24.99	24.57	7.31	0.86	3,490
Medium ruggedness	5.94	0.21	−0.57	7.73	−24.56	24.55	7.73	0.91	3,490
Medium-high ruggedness	6.59	0.10	−0.71	8.44	−24.61	24.54	8.44	0.99	3,490
High ruggedness	7.21	0.39	−0.12	9.00	−25.13	24.55	9.01	1.06	3,490

understood as different ways to express the same concept, which is the complexity of local topographical conditions. Note that because numerical values of TRI are a-dimensional and not necessarily meaningful, instead of reporting the range of values we have labeled TRI categories from "low" to "high." For brevity, statistical indicators are only presented for footprints located inside forested areas (Table 4).

The analysis of errors stratified by slope shows that steepness has an impact on accuracy when considering all footprints (Figure 9(b)) or only those footprints located in forested areas (Figure 9(d)). This is also highlighted by MAE, which for forest footprints increases along with the steepness, from a minimum of 3.70 m to a maximum of 7.38 m for footprints on slopes larger than 35%. Meanwhile, RMSE increases by 64% from the lowest to the highest slope category. When considering all valid footprints (Figure 9(a)), a somewhat significant positive association is highlighted (r=0.42), but the strength of the association is seemingly reduced when considering only forest footprints (r=0.28). However, the reduced sample size could simply lead to an r-value more likely to fluctuate due to random sampling variability.

The effect of terrain ruggedness on accuracy follows a very similar pattern to that of slope, in terms of: (1) the strength of correlation for all valid footprints (Figure 10(a)) vs. only forest footprints (Figure 10(c)); (2) the variation of RMSE values and other statistical indicators of accuracy (Table 4); and (3) the impact on overall accuracy as highlighted by boxplots of elevation error values stratified by terrain ruggedness categories (Figure 10(b,d).

3.3. Relative importance of external factors on GEDI performance

Factors described earlier were ranked in terms of their relative importance in explaining the amount of variability of ground elevation errors (Table 5). Note that factors are sorted in descending order of their importance in explaining the error variation for forest footprints. Also note that forest type was only considered for footprints in forest areas, while the type of cover (bare ground vs. forest) was excluded, given that it was directly derived from canopy heights (which is another factor considered). The season of acquisition explains a large part of the error variation, especially for forest footprints, in which case it is associated with 51% of variation (Figure 11). For the case of forest footprints, canopy height also has a significant influence (approximately 37% of variance explained). If outliers are included, two

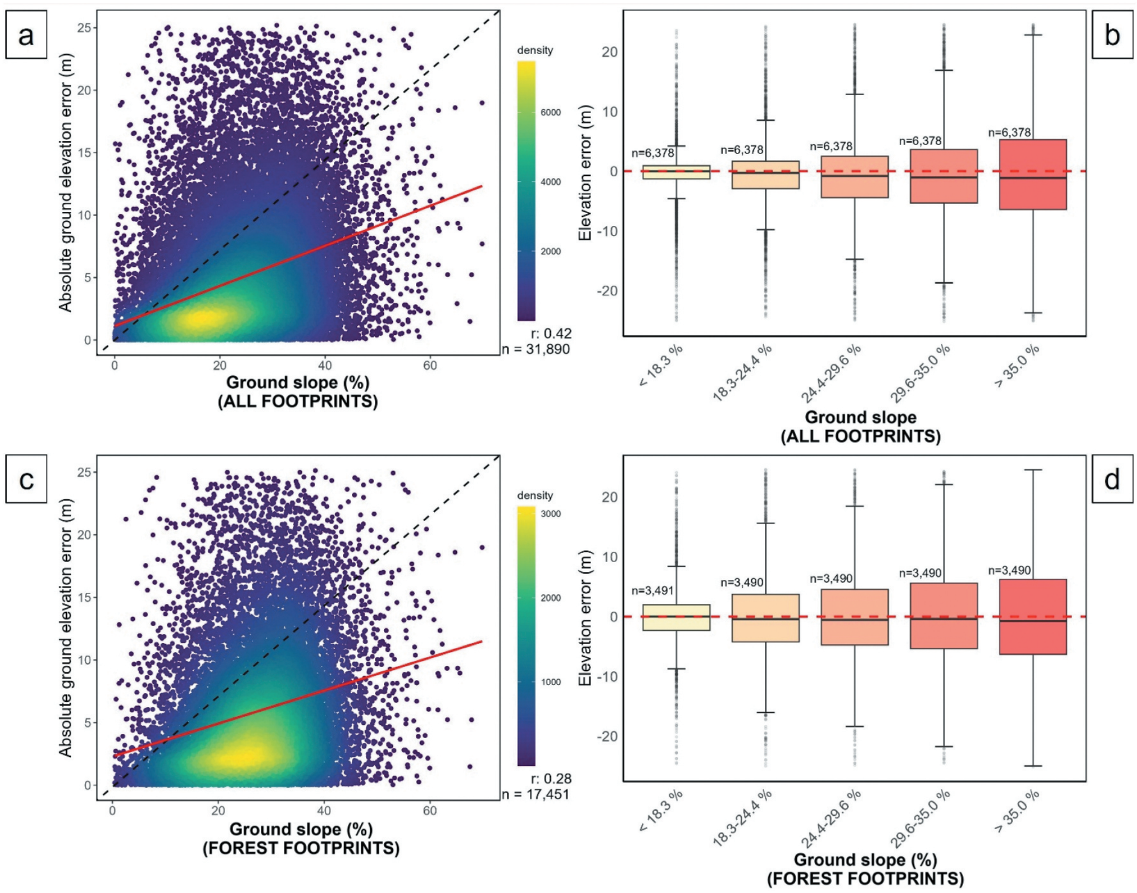

Figure 9. Scatterplot (a) and boxplot (b) showing the variation of ground elevation errors by ground slope, for all GEDI footprints (n = 31, 890). Scatterplot (c) and boxplot (d) showing the variation of ground elevation errors by ground slope, for GEDI footprints located inside forest areas (n = 17,451).

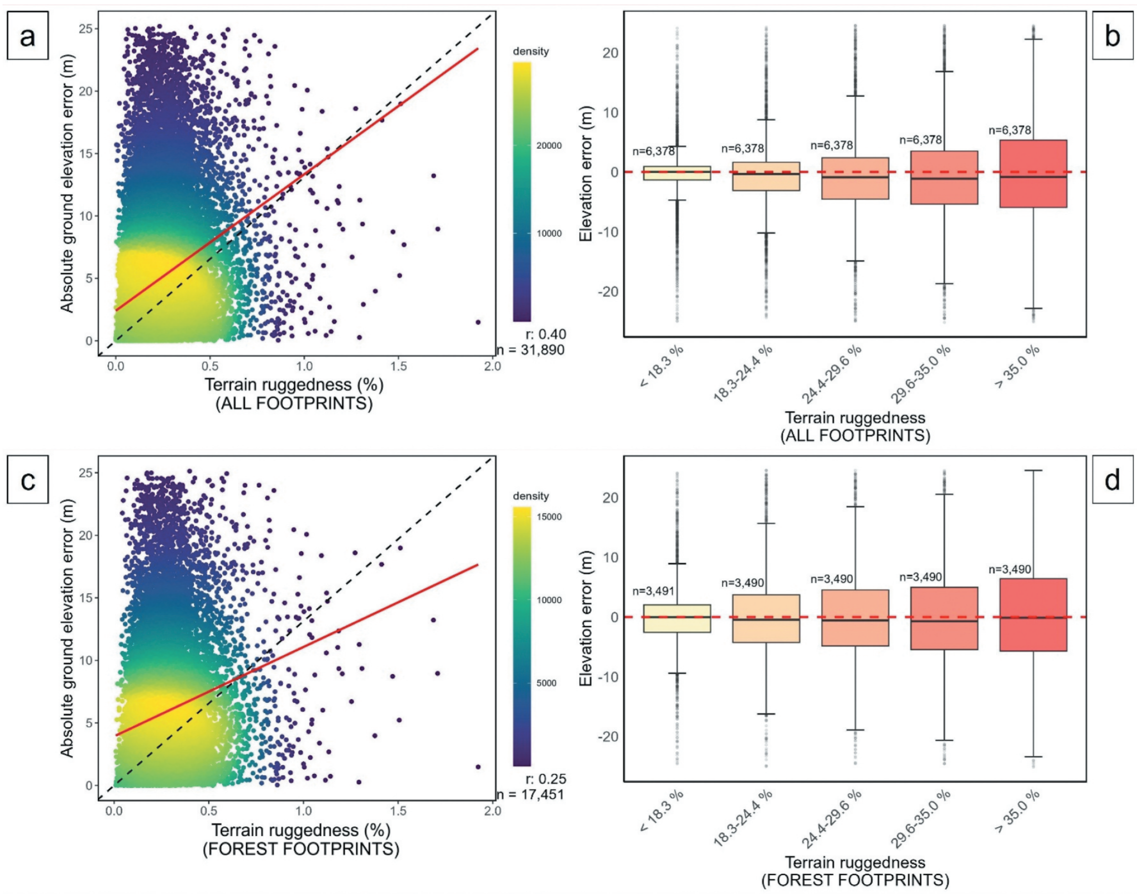

Figure 10. Scatterplot (a) and boxplot (b) showing the variation of ground elevation errors by terrain ruggedness as estimated by TRI, for all GEDI footprints ($n = 31{,}890$). Scatterplot (c) and boxplot (d) showing the variation of ground elevation errors by terrain ruggedness as estimated by TRI, for GEDI footprints located inside forest areas ($n = 17{,}451$).

Table 5. Relative importance of factors in explaining the variability of GEDI ground elevation errors.

Factor	Relative importance for all footprints (%)	Relative importance for footprints without outliers (%)	Relative importance for forest footprints (%)
Season of acquisition	22.73	32.41	51.17
Canopy height	32.60	56.02	37.32
Beam type	21.83	1.44	5.57
Morphometric feature	1.29	2.84	2.27
Ground slope	1.16	4.72	1.85
Forest type	Not applicable	Not applicable	0.66
Terrain ruggedness	1.22	2.07	0.61
Time of acquisition (day/night)	19.18	0.50	0.55

factors are highlighted as becoming significantly more important: beam type (increasing from a percentage of 5% to 22%) and the time of data acquisition (increasing from under 1% to 19%). These are only relative measures of importance, so, for example, the season of acquisition should not be understood as explaining 51% of the variance in forested terrain. Rather, it explains 51% of the variance explained by all factors together.

4. Discussion

4.1. Overall accuracy

Analysis of GEDI error distribution shows that a small number of observations (in this case approximately 2%–3%) are not properly filtered by the quality and/or degrade flagging introduced in GEDI v2. Errors associated with these outliers tend to be positive and reach a maximum of over 383 m, negatively affecting the overall accuracy. Very large errors, likely occurring when the GEDI system malfunctions or there are issues regarding waveform processing, will most commonly lead to an overestimation of the ground surface (positive outliers, as shown in Figure 4(c)). The large impact a small number of outliers can have in analyses of GEDI performance has been previously pointed out, with filtering proposed to improve results (Wang et al. 2022). As an example, Adam et al. (2020) tested GEDI performance in two study sites located in temperate forest areas in central Germany,

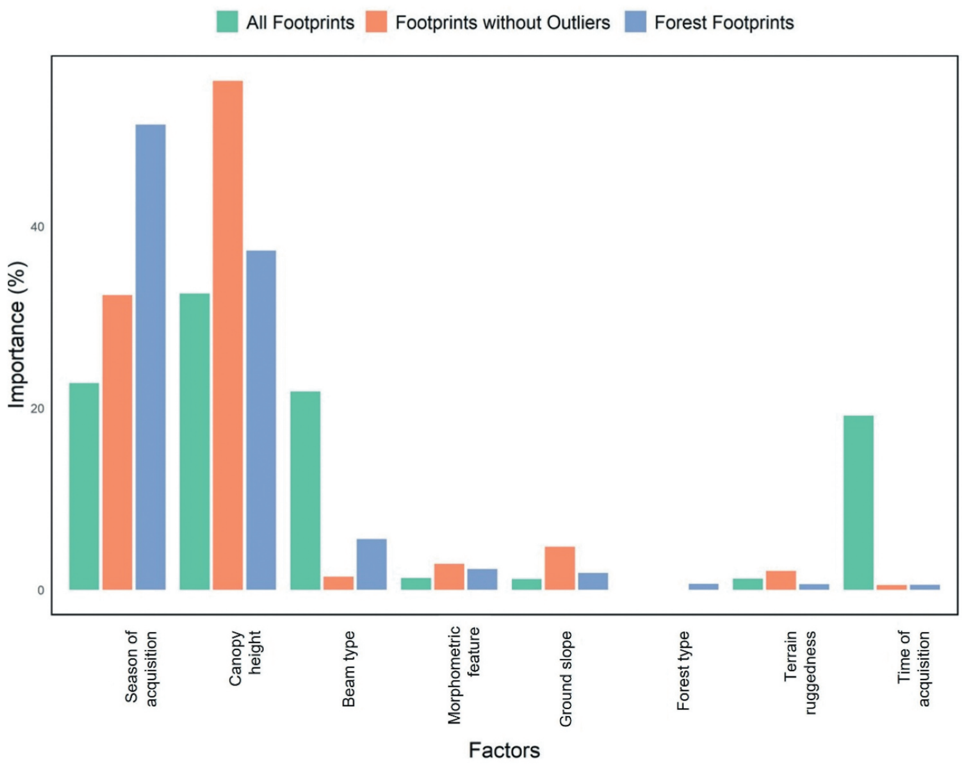

Figure 11. The relative importance of factors in explaining GEDI accuracy, for 3 cases: filtered footprints ($n = 32,666$), footprints without outliers ($n = 31,890$) and only footprints located in areas covered by forest ($n = 17,451$).

reporting positive outliers in the Thuringian area, characterized by complex topographical conditions, relatively similar to our study area. These significant overestimations of ground elevation affecting a relatively small number of GEDI footprints reduce the correlation between GEDI and reference elevations to a larger degree than in our case (R^2 of 0.934, as compared to 0.996 in our case), but this is likely explained by the fact that Adam et al. (2020) analyzed GEDI v1 data, which had a significantly larger geolocation error. The lower accuracy of GEDI v1 is also apparent in the case of Kutchartt, Pedron, and Pirotti (2022), which reported an RMSE value of 10.7 m for the case of footprints located inside alpine forest stands in northern Italy.

Guerra-Hernández and Pascual (2021) estimated the accuracy of GEDI-derived ground elevation for a forested area in the Lugo province of Spain, reporting an RMSE of 4.45 m for conditions of at least 70% canopy cover. While this RMSE indicates a higher accuracy of ground surface detection than in our study, it must be noted that the research is focused on *P. radiata* and *P. pinaster*, which would likely not have as complex a vertical structure as the broadleaved forest stands predominant in our study area. Also, the authors propose a method for outlier removal which could potentially lead to a more aggressive filtering of outliers, thereby improving goodness-of-fit statistics. The authors report a bias of 0.18 m, which matches the bias of 0.19 m for the present study, in the case of forest footprints. Liu, Cheng, and Chen (2021) analyzed GEDI v2 data for 40 sites spread across North America, reporting an RMSE of 4.03 m and a MAE of 1.80 m. While this highlights a higher accuracy than in our study (with an overall RMSE of 6.75 m and MAE of 4.76 after outlier removal), it should be noted that Liu, Cheng, and Chen (2021) performed a geolocation error correction of footprints based on waveform matching, which can significantly increase elevation accuracy, given the expected geolocation error of approximately 10 m for GEDI v2. On the other hand, Quiros, Polo, and Fragoso-Campon (2021) report an RMSE of 6.05 m and a bias of 0.41 m when comparing ground elevation from approximately 12,000 GEDI footprints against ALS reference data. The authors propose a methodology of shifting GEDI footprints in 16 directions evenly distributed around the initial position and selecting the location of lowest RMSE, to compensate for the geolocation error. This approach apparently leads to a level of accuracy comparable with the simpler method of outlier detection and removal used in this research, based on modified *z*-score. Of course, both approaches require validation data in the form of a reference DTM. Overall, we find that in terms of general accuracy, results agree with previous studies, for the most part.

4.2. Changes of GEDI accuracy with the variation of external factors

To better understand the interaction between GEDI and the environment, several external factors were considered in the analysis, related to the characteristics of forest vegetation, GEDI's operation and topographical conditions.

4.2.1. Forest conditions

Our findings suggest that forest cover leads to a lower accuracy of ground surface retrieval. It has to be mentioned that a threshold 5 m for canopy height was used for classifying footprints as "*forest*" rather than "*open ground*," but some footprints in the latter category might cover areas with vegetation shorter than 5 m. This would imply that the difference in accuracy between forested terrain and completely bare ground is likely even more significant. A possible explanation is that the presence of forest canopy causes less energy to return from the ground surface, complicating the interpretation full-waveform data (Urbazaev et al. 2022). Similar results have been reported before. For example, Liu, Cheng, and Chen (2021) compared GEDI accuracy for various land cover types and determined that terrain height retrieval is most problematic in forest ecosystems. The complexity of estimated ground elevation from satellite LiDAR in forest ecosystems has also been previously established (Guerra-Hernández and Pascual 2021; Li et al. 2024). Wang et al. (2022) also report larger bias values areas predominantly covered by forest, while Pronk, Eleveld, and Ledoux (2024) test GEDI v2 performance in various sites (located in the Netherlands, Switzerland and New Zealand), with RMSE increasing from 2.96 m (when considering all landcover classes) to 4.64 m (when taking into account only areas with tree cover).

The second vegetation characteristic we took into account was the type of vegetation (broadleaved vs. coniferous). An expected higher accuracy for coniferous forest stand is found but it should be noted that, given the very small number of samples for the coniferous category ($n = 117$), these results do not have statistical significance. However, they are in agreement with published research, as Kutchartt, Pedron, and Pirotti (2022) also find a better terrain estimation in coniferous stands, while Wang et al. (2022) report the challenge of terrain identification on waveforms in broadleaf forests with dense cover. Yang et al. (2024) considered the case of coniferous stands in northeastern China, reporting RMSE values between 3.19 and 4.50 and biases from −2.69 to – 3.06, depending on site and leaf on/off conditions. These relatively high accuracies, in comparison not just to our study, but also other studies focusing on forested environments, could be explained by the focus on coniferous vegetation.

Given the likely attenuation of energy by its interaction with the forest canopy and the decrease of SNR (Signal-to-Noise Ratio) it causes, it follows that the height of the canopy is another contributing factor to ground detection accuracy. Our results show consistent increases of RMSE, MAE and standard deviation of errors with each forest height category. Furthermore, for forest footprints, the analysis of relative importance shows that canopy height is the second-most important factor affecting accuracy, explaining 37% of error variation. This pattern goes toward validating previous research concerning the influence of canopy height on GEDI ground surface retrieval. For example, Kutchartt, Pedron, and Pirotti (2022) report a significant ground elevation overestimation of approximately 3 m in high canopy conditions, while Liu, Cheng, and Chen (2021) find relatively high RMSE values (6 m −7 m) in dense and tall (over 20 m high) canopies, concluding that the waveform might be misinterpreted in these challenging conditions. Adam et al. (2020) establish that ground elevation accuracy drastically worsens once canopy densities go above 75% and Liu, Cheng, and Chen (2021) find that RMSE more than doubles (from under 3 m to more than 6 m) between canopy heights under 5 m and above 20 m. The influence of tall and dense canopies on the accuracy of data collected by satellite LiDAR is also pointed out by Baban and Niţă (2023). Zhu et al. (2023) consider the specific case of short stature vegetation for sites across the continental United States of America, reporting high accuracies for ground elevation estimates (RMSE of 1.40 m) for footprints located in canopies lower than 5 m. On the other hand, Adam et al. (2020) find that vegetation height only influences GEDI ground accuracy at the highest category (canopy heights over 30 m). We consider that the issue is yet to be settled, as the interaction of GEDI laser pulses and forest canopies of different heights and species composition is also likely complicated by factors such as the steepness of the ground or the presence of sub-canopy vegetation. For example, Urbazaev et al. (2022) carry out an extensive study across six forest stand types located in Brazil, Germany, South America and the USA (with a sample size of 1.4 million GEDI v2 footprints) and report an RMSE of 6.31 m for temperate broadleaved stands and an unexpectedly higher RMSE (6.84 m) for coniferous stands. The authors state that one likely explanation is the fact that the broadleaved stands considered here are characterized by lower canopy densities and ground slopes. On the other hand, Li et al. (2024) report a significantly lower accuracy of GEDI v2 ground data than most other studies (with RMSE ranging from 11.68 m to 16.54 m, dependent on slope conditions) carried out in coniferous conditions. This could partly be explained by the fact that the study area chosen is covered by artificial conifer stands, which generally tend to have higher density than natural ones (Osipenko and Zalesov 2021). Taking these issues into account, we consider that further studies focusing on the influence of forest vegetation characteristics on GEDI ground accuracy, in various types of forest ecosystems, are warranted.

4.2.2. GEDI operational conditions

Our findings show that data collected by power beams provides only a slight increase in ground detection accuracy, when compared coverage beam data (RMSE improves by 0.04 m, from 6.77 m to 6.73 m). Inside the forest vegetation, the benefit of power beam footprints is somewhat more significant (with RMSE improving by 0.25 m, from 7.88 m to 7.63 m). This is expected, as more energy will reach the ground level under the canopy, leading to a more pronounced final peak in the waveform and a more confident ground detection. Beam accounts for 22% of the error variation explained by considered factors. However, when outliers are excluded, beam type becomes a much less significant factor, with relative importance ranging between 1% and 5%. This fact could be attributed to the uneven distribution of outliers across beam types, with 70% of them corresponding to data collected by power beams. Therefore, without further research this can only be attributed to random variation. Overall, we find that the slight impact of beam type on GEDI performance reported here is in agreement with most previous research, as most studies find that coverage beams can provide useful data at the cost of slightly reduced accuracy. For example, Liu, Cheng, and Chen (2021) report that RMSE improves from 5.19 m to 3.56 m when using power beam data collected during the night. This significant increase is explained by the fact that in this study no outlier removal was employed, therefore selecting only nighttime power beam is an indirect manner by which outliers can be mostly eliminated, as the authors state. Meanwhile, Zhu et al. (2023) obtain RMSEs of 1.39 m and 1.62 m (for power and coverage beams, respectively), Pronk, Eleveld, and Ledoux (2024) report median errors of 0.02 m and 0.06 m for the same two cases, Wang et al. (2022) find a marginal increase in RMSE of 0.10 m and Kutchartt, Pedron, and Pirotti (2022) report that beam type does not impact the distribution of elevation errors. An interesting case is that presented by Li et al. (2024), which report a more significant improvement of accuracy for power beam data (RMSE of 15.00 m, as compared with 17.34 m for coverage beam data). As discussed previously, this could be partly explained by the specifics of the forest type analyzed here (artificial coniferous stands) and the fact that only footprints located inside the canopy were considered. However, this does suggest that the benefit brought about by the increased laser energy of GEDI power beams in various forest types, canopy densities and heights is yet to be fully understood.

In terms of data collection time, nighttime data is expected to have less background radiation and therefore an improved SNR (Adam et al. 2020). As expected, this is again more evident when considering footprints located inside forest vegetation (with RMSE improving from 7.89 m for daytime data, to 7.59 m for nighttime data), in which case any additional energy that reaches the ground will aid elevation estimation. While this factor is not widely considered in research, a limited number of previous studies have also reported a minor increase in accuracy for nighttime data: Urbazaev et al. (2022) report an improvement of mean elevation error from 0.59 m (daytime) to 0.54 m (nighttime), Pronk, Eleveld, and Ledoux (2024) establish that median elevation error improves from 0.06 to 0.02, while Kutchartt, Pedron, and Pirotti (2022) report no influence on error distribution. This behavior can be explained by the much higher laser energy of the GEDI instrument, compared, for example, to the ATLAS sensor onboard the ICESat-2 platform (Pronk, Eleveld, and Ledoux 2024). Therefore, we consider that the marginal increase in terms of accuracy obtained by eliminating daytime data is not worth the reduction in terms of spatial coverage, possibly with the sole exception of very high and dense forest canopies.

The season of acquisition seems to be the most significant factor in explaining the accuracy of ground elevation estimates for the present study, especially in areas covered by forest ecosystems (where it has a relative importance of 51%). In areas with vegetation cover, RMSE values are significantly better for winter or spring (5.45 m and 5.98 m, respectively), when compared to data collected during summer or autumn (RMSEs of 8.11 m and 8.75 m, respectively). Given that most forest stands in the study area are composed of broadleaved trees, the season of acquisition can be considered as a substitute for the density of canopy cover. While no previous studies considering season as a potential factor have been identified, we find that we are in agreement with studies that have taken into account canopy density and reported decreasing accuracies in dense forest ecosystems. Wang et al. (2022) report an overall RMSE of 1.38 m after outlier removal in a manner similar to our study but find that accuracy degrades in dense forests located on steep slopes (RMSE of 6.90 m). The authors conclude that the most problematic case for ground estimations from GEDI footprints is that of broadleaved stands with high canopy densities. Liu, Cheng, and Chen (2021) also report a decrease in accuracy from canopy densities under 20% (with RMSE values below 3 m) to the densest canopy conditions (RMSE over 7 m for canopy densities above 90%). Quiros, Polo, and Fragoso-Campon (2021) present very similar findings, with RMSE increasing from approximately 5 m (in canopies with densities under 20%) to more than 8 m (for canopy densities above 80%).

4.2.3. Morphological conditions

GEDI is an LF (Large Footprint) LiDAR system, so that the ground elevation it records does not correspond to the center of its footprint specifically, but rather to the interaction between the emitted energy and the ground area covered by the footprint (which for the case of GEDI can be considered a circle with a diameter of approximately 25 m). In other words, the estimated terrain height is, in the ideal case, an average terrain height of the area inside the footprint. Once the topographical complexity of the surface increases, it follows that this average elevation will tend to deviate from the elevation corresponding to the center of the footprint (the location at which it is stored). Therefore, an LF LiDAR system is expected to have worsening accuracy in areas of high topographical complexity (Silva et al. 2018; Wang et al. 2019). The most used indicator of terrain complexity is slope, likely the most analyzed factor in studies of satellite LiDAR performance. Our results highlight a significant influence of slope on accuracy, with steeper terrain leading to larger RMSE or MAE values, and a wider spread of error values. The effect is even more pronounced in steep terrain covered by vegetation, as in this case the width of the footprint increases and the peak in the waveform corresponding to the ground surface will be mixed up with peaks located inside the canopy, until it can actually disappear (Wang et al. 2019). This phenomenon is caused by vegetation and ground returns occurring at similar altitudes due to the steepness of terrain. Another issue is that of GEDI's geolocation error (which is estimated to be around 10 m for v2 data used for this study), which will induce a larger ground elevation uncertainty in steep terrain, as compared to flatter areas (Schleich et al. 2023). As an example, Kutchartt, Pedron, and Pirotti (2022) used GEDI v1 data collected in alpine forest stands and reported a very significant increase in RMSE with ground steepness: 4.4 m for slopes under 5 degrees, approximately 50 m for slopes of 50 – 55 degrees and an RMSE as high as 77 m for the steepest areas (over 60 degrees). Quantitative analysis of the relation between slope and elevation accuracy is difficult, due to three issues: (1) the wide range of slope values characterizing various study areas, (2) the manner in which slope classification is carried out, with studies employing either fixed slope intervals (generally 5, 10, 15, or even 20 degrees – which can have a significant influence on RMSE values for each class) or quantile breaks ensuring an even distribution of samples across slope cases (like in our study) and (3) the very low sample size for some slope classes (generally the highest ones) reported in most studies. However, we find that the general trend identified here confirms previous studies, while the magnitude of accuracy degradation (with RMSE seeing a two-fold increase between the lowest and highest slope class – from 5.61 m to 9.18 m for slopes under 18 and over 35 degrees, respectively) is also similar to prior research. For example, Adam et al. (2020) report a constant increase of MAE with slope, Liu, Cheng, and Chen (2021) find an increase of RMSE from approximately 3 m (for slopes under 5 degrees) to over 6 m (for slopes over 30 degrees), Quiros, Polo, and Fragoso-Campon (2021) report RMSE values ranging from 2 (for slopes under 5 degrees) to 10 m (for slopes over 25 degrees), noting that areas with slopes over 40 degrees exhibit the largest dispersion of elevation errors. Meanwhile, Urbazaev et al. (2022) also establish a pattern of worsening accuracy due to steepness, especially for slopes above 15 degrees (with no RMSE values reported), while Wang et al. (2022) report that RMSE improves by 1.93 m from steep areas to flat terrain, Zhu et al. (2023) find that RMSE increases from approximately 1 m (for flat terrain) to more than 2.5 m (for slopes over 20 degrees) in the case of short stature vegetation and Pronk, Eleveld, and Ledoux (2024) establish that median elevation error increases from – 0.05 m (for slopes between 5 and 10 degrees), to −1.08 m (for slopes over 30 degrees). By contrast, Guerra-Hernández and Pascual (2021) report a general increase of error range with steepness, but no clear pattern is identified and no RMSE values per slope class are presented.

Terrain ruggedness exhibits a similar influence on accuracy as slope. While the case of terrain ruggedness is hard to put into a larger context, as no previous studies accounting for it have been identified, ruggedness is, like slope, a measure of topographical complexity and it should be expected to influence GEDI accuracy in a similar manner. Both morphological factors considered in this study have low values of relative importance in explaining the variation of ground elevation estimates, which could potentially be attributed to the specifics of the study area. Our results show that, while taken individually, slope or terrain ruggedness influence obtained accuracies, the effect is not as significant as the presence of forest vegetation, its height and canopy density (indirectly considered by the season of GEDI data acquisition).

5. Conclusions

The GEDI satellite program currently provides the most complete set of measurements of the canopy structure and the bare ground underneath it yet available (Qi and Dubayah 2016). The potential of GEDI data being used in combination with other data sources, such as radar or optical, for greater coverage and wall-to-wall sampling, requires an understanding of error sources and expected accuracy for all products involved (Liu, Cheng, and Chen 2021). Therefore, the purpose of this study was to establish the accuracy of ground surface estimates for GEDI version 2 products, which have an improved geolocation error of 10 m (compared to the initial one of 25 m) and a new "quality" flag, which can be used for filtering out potentially mis-identified ground returns. The results, which are focused on forested areas, indicate that some outlier values remain in the data even after filtering using GEDI's "degrade" and "quality" flags. While they represent a small proportion of the dataset (in this case approximately 2%), they do need to be addressed as they significantly degrade overall accuracy. When considering the accuracy presented here, it has to be kept in mind that the estimation of the ground surface from satellite LiDAR is known to be a challenging process (Duncanson et al. 2020) and even more so in forest environments (Guerra-Hernández and Pascual 2021).

In terms of external factors, the presence of forest vegetation, its height, ground slope and ruggedness all influence accuracy, to varying degrees. In terms of factors that can be controlled, this study has shown that using only footprints collected by power beams during the night provides a slight increase in accuracy, but not enough to be worth removing numerous accurate observations collected by coverage beams or during the day. The only factor concerning GEDI data collection that has a significant influence on ground elevation estimates in forested environments was the season of acquisition. Thus, we establish that when the area of interest consists of broadleaved forests, it is worth focusing only on GEDI granules collected outside the vegetative season.

The author considers that the GEDI mission can be the main driver behind a revolution in forest monitoring at global scales (Quiros, Polo, and Fragoso-Campon 2021), while also acknowledging that it is yet to realize its full potential, especially considering its operational integration in forest monitoring and management.

Acknowledgments

The author would like to thank the Romanian "National Center of Cartography" for kindly providing the Airborne Laser Scanning data used in this study. This research did not receive any grant from funding agencies in the public, commercial or not-for-profit sectors.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributor

Căţeanu Mihnea received his PhD degree in Forestry from "Transilvania" University in Braşov, Romania, where he is currently involved in education and research. He is currently pursuing a second PhD degree in Geodesy in Bucharest, Romania. His research interests are related to the use of modern data collection techniques (laser scanning, satellite LiDAR, Remote Sensing, and multispectral imagery collected using UAVs) in enhancing precision forestry practices.

ORCID

Mihnea Căţeanu http://orcid.org/0000-0002-3692-1050

References

Adam, M., M. Urbazaev, C. Dubois, and C. Schmullius. 2020. "Accuracy Assessment of GEDI Terrain Elevation and Canopy Height Estimates in European Temperate Forests: Influence of Environmental and Acquisition Parameters." *Remote Sensing* 12 (23): 3948. https://doi.org/10.3390/rs12233948.

Akay, A. E., H. Oğuz, I. R. Karas, and K. Aruga. 2009. "Using LIDAR Technology in Forestry Activities." *Environmental Monitoring and Assessment* 151 (1–4): 117–125. https://doi.org/10.1007/s10661-008-0254-1.

Baban, G., and D. M. Niţă. 2023. "Measuring Forest Height from Space. Opportunities and Limitations Observed in Natural Forests." *Measurement* 211:112593. https://doi.org/10.1016/j.measurement.2023.112593.

Beck, J., B. Wirt, J. Armston, M. Hofton, S. Luthcke, and H. Tang. 2021. "Global Ecosystem Dynamics Investigation (GEDI) Level 02 User Guide." *Land Processes Distributed Active Archive Center (DAAC), April.*https://lpdaac.usgs.gov/documents/988/GEDI02_User_Guide_V2.pdf.

Beland, M., G. Parker, B. Sparrow, D. Harding, L. Chasmer, S. Phinn, A. Antonarakis, and A. Strahler. 2019. "On Promoting the Use of LIDAR Systems in Forest Ecosystem Research." *Forest Ecology & Management* 450:117484. https://doi.org/10.1016/j.foreco.2019.117484.

Căţeanu, M., and A. Ciubotaru. 2021. "The Effect of LIDAR Sampling Density on DTM Accuracy for Areas with Heavy Forest Cover." *Forests* 12 (3): 265. https://doi.org/10.3390/f12030265.

Coops, N. C., P. Tompalski, T. R. H. Goodbody, M. Queinnec, J. E. Luther, D. K. Bolton, J. C. White, M. A. Wulder, O. R. van Lier, and T. Hermosilla. 2021. "Modelling Lidar-Derived Estimates of Forest Attributes Over Space and Time: A Review of Approaches and Future Trends." *Remote Sensing of Environment* 260:112477. https://doi.org/10.1016/j.rse.2021.112477.

Dassot, M., T. Constant, and M. Fournier. 2011. "The Use of Terrestrial Lidar Technology in Forest Science: Application Fields, Benefits and Challenges." *Annals of*

Forest Science 68 (5): 959–974. https://doi.org/10.1007/s13595-011-0102-2.

Deems, J. S., T. H. Painter, and D. C. Finnegan. 2013. "Lidar Measurement of Snow Depth: A Review." *Journal of Glaciology* 59 (215): 467–479. https://doi.org/10.3189/2013JoG12J154.

Dorado-Roda, I., A. Pascual, S. Godinho, C. A. Silva, B. Botequim, P. Rodriguez-Gonzalvez, E. Gonzalez-Ferreiro, and J. Guerra-Hernandez. 2021. "Assessing the Accuracy of GEDI Data for Canopy Height and Aboveground Biomass Estimates in Mediterranean Forests." *Remote Sensing* 13 (12): 2279. https://doi.org/10.3390/rs13122279.

Dubayah, R., J. B. Blair, S. Goetz, L. Fatoyinbo, M. Hansen, S. Healey, M. Hofton, et al. 2020. "The Global Ecosystem Dynamics Investigation: High-Resolution Laser Ranging of the Earth's Forests and Topography." *Science of Remote Sensing* 1:100002. https://doi.org/10.1016/j.srs.2020.100002.

Duncanson, L., A. Neuenschwander, S. Hancock, N. Thomas, T. Fatoyinbo, M. Simard, C. A. Silva, et al. 2020. "Biomass Estimation from Simulated GEDI, ICESat-2 and NISAR Across Environmental Gradients in Sonoma County, California." *Remote Sensing of Environment* 242:111779. https://doi.org/10.1016/j.rse.2020.111779.

Grömping, U. 2006. "Relative Importance for Linear Regression in *R*: The Package Relaimpo." *Journal of Statistical Software* 17 (1): 1–27. https://doi.org/10.18637/jss.v017.i01.

Grömping, U. 2009. "Variable Importance Assessment in Regression: Linear Regression Versus Random Forest." *American Statistician* 63 (4): 308–319. https://doi.org/10.1198/tast.2009.08199.

Guerra-Hernández, J., and A. Pascual. 2021. "Using GEDI Lidar Data and Airborne Laser Scanning to Assess Height Growth Dynamics in Fast-Growing Species: A Showcase in Spain." *Forest Ecosystems* 8 (1): 14. https://doi.org/10.1186/s40663-021-00291-2.

Hancock, S., J. Armston, M. Hofton, X. Sun, H. Tang, L. I. Duncanson, J. R. Kellner, and R. Dubayah. 2019. "The GEDI Simulator: A Large-Footprint Waveform Lidar Simulator for Calibration and Validation of Spaceborne Missions." *Earth & Space Science* 6 (2): 294–310. https://doi.org/10.1029/2018EA000506.

He, L., Y. Pang, Z. Zhang, X. Liang, and B. Chen. 2023. "ICESat-2 Data Classification and Estimation of Terrain Height and Canopy Height." *International Journal of Applied Earth Observation and Geoinformation* 118:103233. https://doi.org/10.1016/j.jag.2023.103233.

Iglewicz, B., and D. C. Hoaglin. 1993. *How to Detect and Handle Outliers*. Milwaukee, WI: ASQ Press.

Jarihani, A. A., J. N. Callow, T. R. McVicar, T. G. Van Niel, and J. R. Larsen. 2015. "Satellite-Derived Digital Elevation Model (DEM) Selection, Preparation and Correction for Hydrodynamic Modelling in Large, Low-Gradient and Data-Sparse Catchments." *Journal of Hydrology* 524:489–506. https://doi.org/10.1016/j.jhydrol.2015.02.049.

Kutchartt, E., M. Pedron, and F. Pirotti. 2022. "Assessment of Canopy and Ground Height Accuracy from GEDI Lidar Over Steep Mountain Areas." *ISPRS Annals of the Photogrammetry, Remote Sensing & Spatial Information Sciences* 3 (2022): 431–438. https://doi.org/10.5194/isprs-annals-V-3-2022-431-2022.

Li, H., X. Li, T. Kato, M. Hayashi, J. Fu, and T. Hiroshima. 2024. "Accuracy Assessment of GEDI Terrain Elevation, Canopy Height, and Aboveground Biomass Density Estimates in Japanese Artificial Forests." *Science of Remote Sensing* 10:100144. https://doi.org/10.1016/j.srs.2024.100144.

Li, X., K. Wessels, J. Armston, S. Hancock, R. Mathieu, R. Main, L. Naidoo, B. Erasmus, and R. Scholes. 2023. "First Validation of GEDI Canopy Heights in African Savannas." *Remote Sensing of Environment* 285:113402. https://doi.org/10.1016/j.rse.2022.113402.

Lim, K., P. Treitz, M. Wulder, B. St-Onge, and M. Flood. 2003. "LIDAR Remote Sensing of Forest Structure." *Progress in Physical Geography: Earth and Environment* 27 (1): 88–106. https://doi.org/10.1191/0309133303pp360ra.

Liu, A., X. Cheng, and Z. Chen. 2021. "Performance Evaluation of GEDI and ICESat-2 Laser Altimeter Data for Terrain and Canopy Height Retrievals." *Remote Sensing of Environment* 264:112571. https://doi.org/10.1016/j.rse.2021.112571.

Markus, T., T. Neumann, A. Martino, W. Abdalati, K. Brunt, B. Csatho, S. Farrell, et al. 2017. "The Ice, Cloud, and Land Elevation Satellite-2 (ICESat-2): Science Requirements, Concept, and Implementation." *Remote Sensing of Environment* 190:260–273. https://doi.org/10.1016/j.rse.2016.12.029.

Müller, J., I. Gärtner-Roer, P. Thee, and C. Ginzler. 2014. "Accuracy Assessment of Airborne Photogrammetrically Derived High-Resolution Digital Elevation Models in a High Mountain Environment." *ISPRS Journal of Photogrammetry & Remote Sensing* 98:58–69. https://doi.org/10.1016/j.isprsjprs.2014.09.015.

Næsset, E. 1997a. "Determination of Mean Tree Height of Forest Stands Using Airborne Laser Scanner Data." *ISPRS Journal of Photogrammetry & Remote Sensing* 52 (2): 49–56. https://doi.org/10.1016/S0924-2716(97)83000-6.

Næsset, E. 1997b. "Estimating Timber Volume of Forest Stands Using Airborne Laser Scanner Data." *Remote Sensing of Environment* 61 (2): 246–253. https://doi.org/10.1016/S0034-4257(97)00041-2.

Næsset, E., and K.-O. Bjerknes. 2001. "Estimating Tree Heights and Number of Stems in Young Forest Stands Using Airborne Laser Scanner Data." *Remote Sensing of Environment* 78 (3): 328–340. https://doi.org/10.1016/S0034-4257(01)00228-0.

Narine, L. L., S. C. Popescu, and L. Malambo. 2020. "Using ICESat-2 to Estimate and Map Forest Aboveground Biomass: A First Example." *Remote Sensing* 12 (11): 1824. https://doi.org/10.3390/rs12111824.

Osipenko, A. E., and S. V. Zalesov. 2021. "Comparative Productivity of Artificial and Natural Forest Stands in the Belt Pine Forests of Altay." *IOP Conference Series Earth and Environmental Science* 876 (1): 012056. https://doi.org/10.1088/1755-1315/876/1/012056.

Pârvu, I. M., E. Özdemir, and F. Remondino. 2020. "Aerial Point Cloud Classification Using an Alternative Approach for the Dynamic Computation of K-Nearest Neighbors." *Journal of Applied Engineering Sciences* 10 (2): 155–162. https://doi.org/10.2478/jaes-2020-0023.

Patterson, P. L., S. P. Healey, G. Ståhl, S. Saarela, S. Holm, H.-E. Andersen, R. O. Dubayah, et al. 2019. "Statistical Properties of Hybrid Estimators Proposed for GEDI—NASA's Global Ecosystem Dynamics Investigation." *Environmental Research Letters* 14 (6): 065007. https://doi.org/10.1088/1748-9326/ab18df.

Picu, I. A. C. 2018. "Updating Geospatial Data by Creating a High Resolution Digital Surface Model." *Journal of Applied Engineering Sciences* 8 (2): 51–58. https://doi.org/10.2478/jaes-2018-0018.

Popescu, S. C., T. Zhou, R. Nelson, A. Neuenschwander, R. Sheridan, L. Narine, and K. M. Walsh. 2018. "Photon Counting Lidar: An Adaptive Ground and Canopy Height Retrieval Algorithm for ICESat-2 Data." *Remote Sensing of Environment* 208:154–170. https://doi.org/10.1016/j.rse.2018.02.019.

Potapov, P., X. Li, A. Hernandez-Serna, A. Tyukavina, M. C. Hansen, A. Kommareddy, A. Pickens, et al. 2021. "Mapping Global Forest Canopy Height Through Integration of GEDI and Landsat Data." *Remote Sensing of Environment* 253:112165. https://doi.org/10.1016/j.rse.2020.112165.

Pronk, M., M. Eleveld, and H. Ledoux. 2024. "Assessing Vertical Accuracy and Spatial Coverage of ICESat-2 and GEDI Spaceborne Lidar for Creating Global Terrain Models." *Remote Sensing* 16 (13): 2259. https://doi.org/10.3390/rs16132259.

Qi, W., and R. O. Dubayah. 2016. "Combining TanDEM-X InSAR and Simulated GEDI Lidar Observations for Forest Structure Mapping." *Remote Sensing of Environment* 187:253–266. https://doi.org/10.1016/j.rse.2016.10.018.

Qi, W., S.-K. Lee, S. Hancock, S. Luthcke, H. Tang, J. Armston, and R. Dubayah. 2019. "Improved Forest Height Estimation by Fusion of Simulated GEDI Lidar Data and TanDEM-X InSAR Data." *Remote Sensing of Environment* 221:621–634. https://doi.org/10.1016/j.rse.2018.11.035.

Qi, W., S. Saarela, J. Armston, G. Ståhl, and R. Dubayah. 2019. "Forest Biomass Estimation Over Three Distinct Forest Types Using Tandem-X InSAR Data and Simulated GEDI Lidar Data." *Remote Sensing of Environment* 232:111283. https://doi.org/10.1016/j.rse.2019.111283.

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

Riley, S. J., S. D. DeGloria, and R. Elliot. 1999. "A Terrain Ruggedness Index that Quantifies Topographic Heterogeneity." *Intermountain Journal of Sciences* 5 (1): 23–27.

Rishmawi, K., C. Huang, and X. Zhan. 2021. "Monitoring Key Forest Structure Attributes Across the Conterminous United States by Integrating GEDI lidar Measurements and VIIRS Data." *Remote Sensing* 13 (3): 442. https://doi.org/10.3390/rs13030442.

Roy, D. P., H. B. Kashongwe, and J. Armston. 2021. "The Impact of Geolocation Uncertainty on GEDI Tropical Forest Canopy Height Estimation and Change Monitoring. " *Science of Remote Sensing* 4 (5): 100024). https://doi.org/10.1111/j.1755-263X.2009.00067.x.

Sasaki, N., and Putz, F.E., 2009. Critical Need for New Definitions of "Forest" and "Forest Degradation" in Global Climate Change Agreements. *Conservation Letters*. 2, 226–232. https://doi.org/10.1111/j.1755-263X.2009.00067.x

Schleich, A., S. Durrieu, M. Soma, and C. Vega. 2023. "Improving GEDI Footprint Geolocation Using a High-Resolution Digital Elevation Model." *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing* 16:7718–7732. https://doi.org/10.1109/JSTARS.2023.3298991.

Schneider, F. D., A. Ferraz, S. Hancock, L. I. Duncanson, R. O. Dubayah, R. P. Pavlick, and D. S. Schimel. 2020. "Towards Mapping the Diversity of Canopy Structure from Space with GEDI." *Environmental Research Letters* 15 (11): 115006. https://doi.org/10.1088/1748-9326/ab9e99.

Silva, C. A., L. Duncanson, S. Hancock, A. Neuenschwander, N. Thomas, M. Hofton, L. Fatoyinbo, et al. 2021. "Fusing Simulated GEDI, ICESat-2 and NISAR Data for Regional Aboveground Biomass Mapping." *Remote Sensing of Environment* 253:112234. https://doi.org/10.1016/j.rse.2020.112234.

Silva, C. A., S. Saatchi, M. Garcia, N. Labrière, C. Klauberg, A. Ferraz, V. Meyer, et al. 2018. "Comparison of Small- and Large-Footprint Lidar Characterization of Tropical Forest Aboveground Structure and Biomass: A Case Study from Central Gabon." *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing* 11 (10): 3512–3526. https://doi.org/10.1109/JSTARS.2018.2816962.

Tang, H., J. Stoker, S. Luthcke, J. Armston, K. Lee, B. Blair, and M. Hofton. 2023. "Evaluating and Mitigating the Impact of Systematic Geolocation Error on Canopy Height Measurement Performance of GEDI." *Remote Sensing of Environment* 291:113571. https://doi.org/10.1016/j.rse.2023.113571.

Urbazaev, M., L. L. Hess, S. Hancock, L. Y. Sato, J. P. Ometto, C. Thiel, C. Dubois, et al. 2022. "Assessment of Terrain Elevation Estimates from ICESat-2 and GEDI Spaceborne Lidar Missions Across Different Land Cover and Forest Types." *Science of Remote Sensing* 6:100067. https://doi.org/10.1016/j.srs.2022.100067.

Wang, C., A. J. Elmore, I. Numata, M. A. Cochrane, L. Shaogang, J. Huang, Y. Zhao, and Y. Li. 2022. "Factors Affecting Relative Height and Ground Elevation Estimations of GEDI Among Forest Types Across the Conterminous USA." *GIScience and Remote Sensing* 59 (1): 975–999. https://doi.org/10.1080/15481603.2022.2085354.

Wang, Y., W. Ni, G. Sun, H. Chi, Z. Zhang, and Z. Guo. 2019. "Slope-Adaptive Waveform Metrics of Large Footprint Lidar for Estimation of Forest Aboveground Biomass." *Remote Sensing of Environment* 224:386–400. https://doi.org/10.1016/j.rse.2019.02.017.

Yang, C., D. Peng, K. Deng, L. Jiang, M. Zhao, W. Zeng, Y. Shao, and N. Wang. 2024. "Quantification of GEDI Geolocation Error and Its Influence on Elevation and Canopy Height." IEEE Journal of selected topics in applied earth observations and remote sensing 17, 18012–18024. https://doi.org/10.1109/JSTARS.2024.3461843.

Yaro, A. S., F. Maly, and P. Prazak. 2023. "Outlier Detection in Time-Series Receive Signal Strength Observation Using Z-Score Method with S_n Scale Estimator for Indoor Localization." *Applied Sciences* 13 (6): 3900. https://doi.org/10.3390/app13063900.

Zhang, K., S.-C. Chen, D. Whitman, M.-L. Shyu, J. Yan, and C. Zhang. 2003. "A Progressive Morphological Filter for Removing Nonground Measurements from Airborne LIDAR Data." *IEEE Transactions on Geoscience & Remote Sensing* 41 (4): 872–882. https://doi.org/10.1109/tgrs.2003.810682.

Zhu, X., S. Nie, Y. Zhu, Y. Chen, B. Yang, and W. Li. 2023. "Evaluation and Comparison of ICESat-2 and GEDI Data for Terrain and Canopy Height Retrievals in Short-Stature Vegetation." *Remote Sensing* 15 (20): 4969. https://doi.org/10.3390/rs15204969.