Multi-Criteria Strategy for Estimating GEDI Terrain Height

Narin, Omer Gokberk; Lindenbergh, Roderik; Abdikan, Saygin

DOI
10.1109/RAST57548.2023.10197988

Publication date
2023

Document Version
Final published version

Published in
Proceedings of 10th International Conference on Recent Advances in Air and Space Technologies, RAST 2023

Citation (APA)
Narin, O. G., Lindenbergh, R., & Abdikan, S. (2023). Multi-Criteria Strategy for Estimating GEDI Terrain Height. In *Proceedings of 10th International Conference on Recent Advances in Air and Space Technologies, RAST 2023* (Proceedings of 10th International Conference on Recent Advances in Air and Space Technologies, RAST 2023). IEEE. https://doi.org/10.1109/RAST57548.2023.10197988

Multi-Criteria Strategy for Estimating GEDI Terrain Height

Omer Gokberk Narin
Department of Geomatics Engineering
Afyon Kocatepe University
Afyonkarahisar, Türkiye
gokberknarin@aku.edu.tr

Roderik Lindenbergh
Department of Geoscience and Remote Sensing
Delft University of Technology
Delft, The Netherlands
R.C.Lindenbergh@tudelft.nl

Saygın Abdikan
Department of Geomatics Engineering
Hacettepe University
Ankara, Türkiye
sayginabdikan@hacettepe.edu.tr

Abstract— **Global Ecosystem Dynamics Investigate (GEDI) is a spaceborne laser altimeter system used for earth observation in many areas such as forest canopy, water level and terrain height estimation. GEDI data is affected by atmospheric effects due to the sensor used while observing. In this study, we propose a 7-step, multi-variable strategy for determining the elevation of the terrain with GEDI. These steps involve both different geoid models, GEDI ancillary data, and topographic features. We evaluated the effect of each step using high quality DEM data obtained by Airborne LiDAR over the central part of Puerto Rico, where building areas and forests are dominant, while the terrain has an average slope of 24%. The GEDI data of the test area consists of 3 different orbits (O06225, O07933, O08061) with different solar elevation and cloudiness rates. While the raw data of orbit O06225, obtained during a solar elevation of 8.4 and cloudy conditions, has a Root Mean Square Error (RMSE) of 418.67 m., the RMSE is reduced to 4.59 m. after applying all seven filtering steps. The raw data of orbit O07933, obtained with a solar elevation of 50.5 during cloud free conditions, has a RMSE of 10.04 m., and is reduced to a similar value of 4.8 m. as a result of the filtering steps. On the other hand, orbit O08061 was obtained with little clouds during a near-dawn solar elevation of -0.7. Its raw RMSE of 50,34 m could only be reduced to 12.41 m. by the proposed filtering procedure. It is concluded that although there are many outliers in data acquired during cloudy conditions, the accuracy of the data remaining after applying our filtering strategy can be as high as the accuracy obtained during cloud free conditions. Better results than 5 m were obtained according to the RMSE in areas with low solar elevation. In addition, it is observed that accuracy decreases strongly when the solar elevation is close to 0. Overall, it is concluded that appropriate filtering is required when determining terrain height with GEDI data.**

Keywords—GEDI, terrain estimation, airborne LiDAR, spaceborne laser altimeter, digital elevation model (DEM)

I. Introduction

Understanding the earth's surface is essential for geosciences. Data with high accuracy and spatial resolution are available in developed regions worldwide, but still, most of the world uses relatively low-accuracy data produced by satellite systems. In addition, although high-precision data can be obtained, these data have high costs and are only available for relatively small areas. For this reason, new satellite systems and sensors are being designed to make the earth's surface terrain height estimation more accurate and more widespread.

One of these systems is the Global Ecosystem Dynamics Investigation (GEDI) spaceborne laser altimetry system, which was placed on the (International Space Station) ISS in 2018 to be used for world-wide observations. GEDI has three lasers that emit light at a wavelength of 1064 nm (Near-infrared). These lasers are divided into coverage (one laser) and full power (two lasers). It obtains 3D information by using a pulse-based system that creates footprints of 25 meters in diameter for each beam, and records and processes the full waveform return of each beam [1]. GEDI has demonstrated potential in many areas for retrieval of forest height [2], changes in water level [3] and retrieval of terrain and canopy height [4].

The positional and vertical accuracy of the GEDI laser altimetry system is affected by many variables such as clouds, solar illumination from atmospheric influences [1], and land cover and topographic effects [5]. GEDI provides auxiliary information for filtering good quality data. There is a quality flag (outlier value = "0", usable value = "1") parameter to remove outliers in GEDI data while sensitivity is an auxiliary variable that gives the probability of the GEDI signal reaching the ground. [6] investigated the accuracy of surface and canopy height models obtained with GEDI data in tropical forests. After starting with 1000 GEDI points in total, 417 GEDI data points remained, and they were compared with Airborne LiDAR system (ALS) data. As a result, they stated that high accuracy was obtained when the beams with >98% sensitivity (RMSE = 3.5 m for 222 footprints) and full power lasers (RMSE = 3.34 m for 277 footprints) data were selected. [7] compared GEDI elevations of lowest mode and ALS data for 10 different areas with different slopes and land cover in Spain. As a result, they achieved an RMSE = 6.05 m. After determining the terrain by using quality flag = 1 data and shifting GEDI footprints, [4] compared ALS data with GEDI data for 35 different fields. They used the following GEDI data quality indicators: quality_flag, beam_flag, date_time and sensitivity. After selecting quality_flag 1 for terrain determination in all areas, a value for the RMSE = 4.03 m for 90,472 footprints was obtained, while full power & night data were selected to obtain an RMSE = 3.53 m for 22,972 footprints. [8] investigated the optimal threshold of accuracy assessment in GEDI data using the Kolmogorov–Smirnov test and other parameters in the Brasília, the capital of Brazil. They stated that a total of 7,619 GEDI data sets had an RMSE value of 5946.73 m. As a result of the methods they applied, 502 GEDI data remained and RMSE of these data decreased to 1.32 m.

In our study, we examined in detail the parameters that will affect the accuracy of the surface determination performance of GEDI data. In addition, we set threshold values for many parameters and showed how much they affect the accuracy estimation.

2023 10th International Conference on Recent Advances in Air and Space Technologies (RAST) | 979-8-3503-2302-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/RAST57548.2023.10197988

II. Material & Method

A. Study Area & Ground Truth

Our study area is located, in the central part of Puerto Rico, and extends across latitudes of 18.213° to 18.289° N and longitudes of 66.369° to 66.298° W, (Fig. 1). There are two dominant classes in the study area. First, 65% of the area consists of built-up area, while, second, 28% of the area consists of forest. The minimum elevation of the area is 242 m. and the maximum elevation is 835 m. In addition, the average slope of the area is 24%.

DEM data with a resolution of 1 meter produced from ALS data was used as ground truth, [9]. ALS data collection started in 2018. The accuracy of the ALS DEM data is reported as having an RMSEz of +/- 13.4 cm at a 95% confidence level in non-vegetated areas (96 points), and +/- 27.3 cm at the 95th percentile in vegetated areas (71 points).

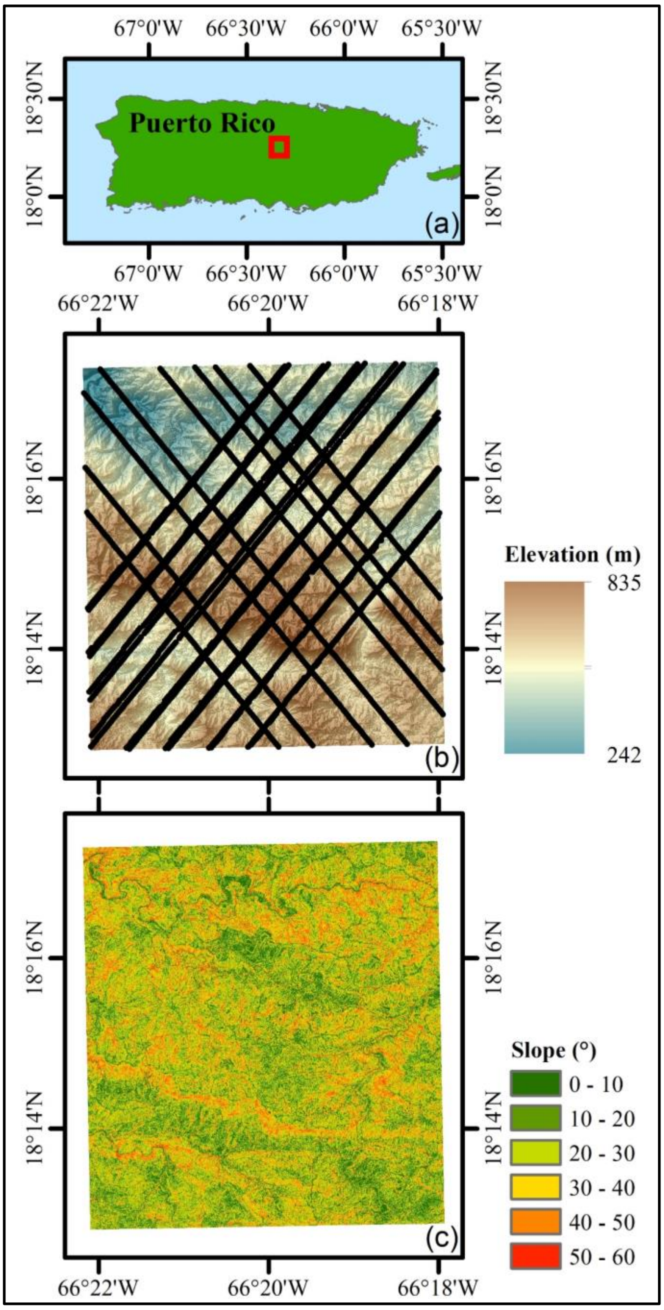

Fig. 1. (a) The red polygon indicates the study area in Puerto Rico. (b) the elevation map of the study area. Black points are GEDI footprints (c) slope map of the study area.

B. GEDI Processing & Filtering Steps

GEDI data from 3 different periods has been downloaded (TABLE I). These data are a Level 2a product. That includes footprint locations and height profiles (Terrain and canopy heights) as well as so-called Relative Heights metrics and additional information (i.e. beam no, solar elevation, sensitivity etc.). These data were found and downloaded with the GEDI finder service and then processed with the rGEDI package [10]. All three data products were collected at night time, but still they have different solar elevations. In cases where the solar elevation is higher than zero, errors occur due to distortions in the waveforms, because the solar background affects the signal-to-noise ratio (SNR) and the number of photons in a segment [1].

TABLE I. DATA PRODUCT NAMES AND SOLAR ELEVATION USED IN THE STUDY

Product Name	Solar elevation	Orbit
GEDI02_A_2020017221656_O06225_T03258_02_001_01.h5	-8.4	O06225
GEDI02_A_2020128024123_O07933_T03258_02_001_01.h5	-50.5	O07933
GEDI02_A_2020136085002_O08061_T00418_02_001_01.h5	-0.7	O08061

In order to determine the best terrain height, the following 7 steps have been applied to the GEDI data

1) Determination of Geoid: Since our ground *truth* data is represented using orthometric height, we need to determine an appropriate geoid model for the GEDI data. We used different geoid reference models, because geopotential models affect vertical accuracy [11]. In our study, 4 models were tested from different years and with different spherical harmonic coefficients. These models are EGM96 [12], EGM 2008 [13], XGM2019e [14], and SGG-UGM-2 [15], respectively. These models have been downloaded from the ICGEM service [16].

2) Filtering according to Quality Flag: this is a quality parameter determined by GEDI for each beam [1,7]. We used quality_flag=1 data for best estimation (Fig. 2).

3) Filtering according to SRTM: In our study, data from the Shuttle Radar Topography Mission (SRTM), which is used in many studies as height reference [6,17], was used to remove outliers in the GEDI data. If the difference between SRTM and GEDI elevation is greater than 20 m, GEDI data is deleted.

4) Filtering according to RH mode: The LiDAR system used by GEDI obtains heights such as elev_lowest, elev_highest and Relative Heights (RH), [1]. RH metrics ranges between 0 to 100 and RH100 gives the difference between elev_highest and elev_lowest. In this study, accuracies for each RH0, RH10, RH20, RH30 and RH40 metrics were calculated with the elev_lowest height provided by GEDI to estimate the terrain.

5) Filtering according to Sensitivity: GEDI is capable of detecting the ground at all canopy cover ratios. Sensitivity is calculated by GEDI for each beam and gives the probability of each beam detecting the ground [18]. In our study, different thresholds were determined and the accuracy of all thresholds was calculated.

6) Filtering according to slope: Errors in vertical height estimation may vary for different slope groups, [4,5,7]. In our study, we determined thresholds according to the slope groups (Fig. 1) and selected the data belonging to the slope group that gave the best results.

7) Filtering according to Beam: GEDI has two beams: GEDI coverage and full power lasers (Fig. 2). In this study, we compared the accuracies of these two beams and selected the points with the highest accuracy.

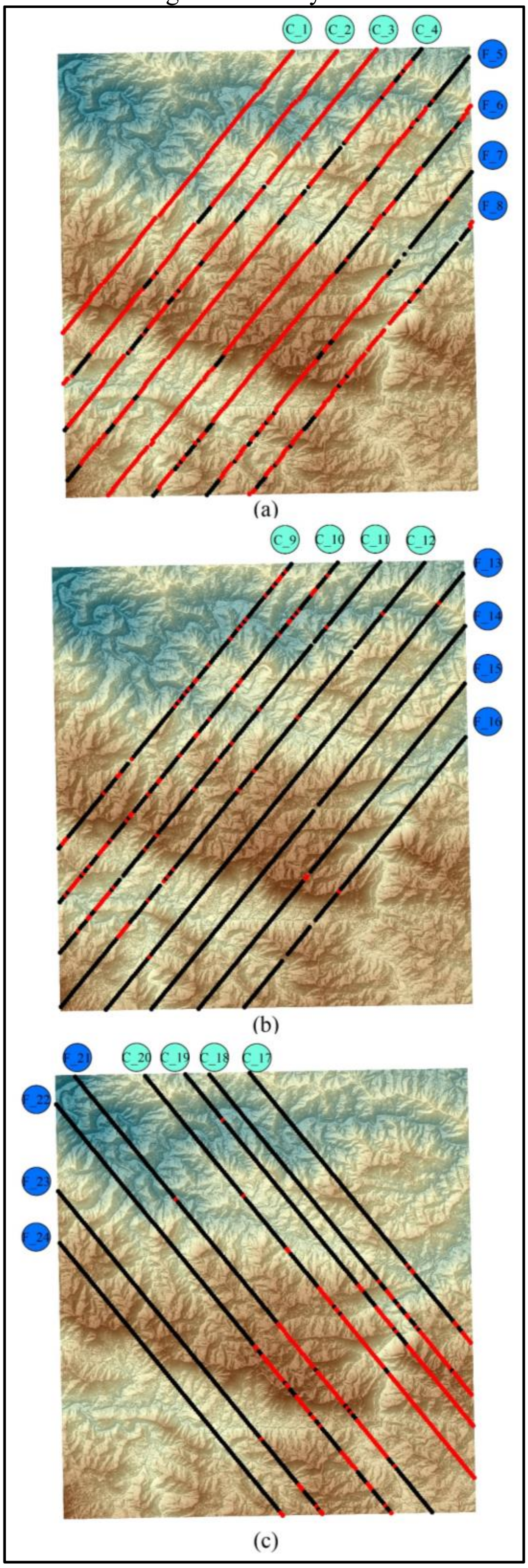

Fig. 2. (a) Projection of GEDI orbit O06225 data (b) Projection of GEDI orbit O07933 data (c) Projection of GEDI orbit O08061 data. The ground tracks represent the Coverage Laser in light blue and the Full Power Laser in dark blue. Footprints with quality flag = 1 are shown in black, while footprints with quality flag = 0 are shown in red.

C. *Experiment Setup and Evaluation Metrics*

In our study, we downloaded the footprints of the GEDI data and extracted elevation values corresponding to the 25 m footprints from the ground truth data. After that, we applied the introduced 7-step procedure (fig. 3) with different threshold values as specified in TABLE II. In our study, RMSE, compare Equation (1) was used as comparison metric.

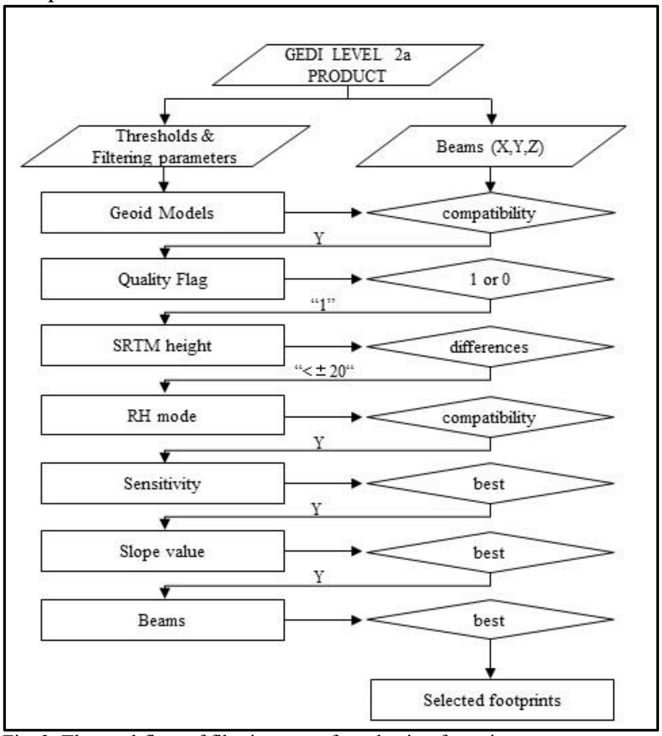

Fig. 3. The workflow of filtering steps for selecting footprints.

TABLE II. THE CRITERIA AND DIFFERENT THRESHOLDS USED FOR FILTERING

Steps	Criteria	Variables & Thresholds	Source
1	Geoid	EGM96, EGM2008, SGG-UGM-2 and XGM2019e_2159	ICGEM
2	Quality Flag	0 and 1	GEDI L2A
3	Height Difference	-20 < Δh (Selected model - SRTM) < 20	SRTM
4	Height Mode & RH	Elev_Lowestmode, RH (0, 10, 20, 30 and 40)	GEDI L2A
5	Sensitivity	0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97 and 0.98	GEDI L2A
6	Slope (°)	10, 20, 30, 40 and 50	Topography
7	Beams	Coverage and Full Power	GEDI L2A

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2} \quad (1)$$

Here x_i denotes a ground truth elevation and y_i the corresponding GEDI elevation.

III. Result & Discussion

In our study, we evaluated the GEDI quality for three different solar elevation values by comparison to the ALS reference data. The data from orbit O06225 has much higher cloudiness than the other data. The data from orbit O07933 has a lower solar elevation while the data from orbit O08061 has a higher solar elevation. Data from these three orbits were first compared to the ALS data without any filtering Afterward, it was analyzed separately for each processing step. According to the RMSE metric, data from orbit O07933 gave the best raw data result. This GEDI data is affected both by cloudiness and solar elevation (TABLE III).

TABLE III. RMSE VALUES FOR COMPARISONS BETWEEN GEDI DATA AND ALS REFERENCE DATA FOR DIFFERENT FILTERING STEPS

Steps	Criteria & Thresholds	O06225		O07933		O08061	
		RMSE (m)	Beams number	RMSE (m)	Beams number	RMSE (m)	Beams number
1	EGM96	**418.67**	1178	**10.04**	1188	50.34	1164
	EGM2008	418.71	1178	10.04	1188	**50.33**	1164
	SGG-UGM-2	418.73	1178	10.04	1188	50.33	1164
	XGM2019e	418.74	1178	10.04	1188	50.33	1164
2	Quality Flag	51.85	292	7.01	1088	28.24	892
3	SRTM	7.95	260	6.61	1000	14.10	459
4	Lowest_mode	7.95	260	6.61	1000	**14.10**	459
	rh0	8.95	260	8.80	1000	15.28	459
	rh10	7.37	260	6.93	1000	14.30	459
	rh20	**7.19**	260	6.35	1000	14.21	459
	rh30	7.37	260	**6.18**	1000	14.34	459
	rh40	7.75	260	6.26	1000	14.59	459
5	Sensivity>0.91	7.20	244	6.15	936	14.10	451
	Sensivity>0.92	7.09	231	6.19	787	14.06	436
	Sensivity>0.93	**7.00**	**218**	6.00	648	**14.03**	**400**
	Sensivity>0.94	7.04	188	**5.83**	**541**	14.00	358
	Sensivity>0.95	7.03	162	5.80	482	14.30	287
	Sensivity>0.96	7.14	126	5.82	331	14.74	213
	Sensivity>0.97	7.06	81	5.80	27	14.73	134
6	Slope<50	6.92	211	5.79	530	13.99	393
	Slope<40	6.42	168	5.46	443	14.08	360
	Slope<30	6.21	144	5.15	376	14.20	320
	Slope<20	**5.12**	**54**	**4.79**	**157**	**13.58**	**135**
	Slope<10	5.03	49	4.84	148	13.80	125
7	Full Power	**4.59**	**41**	**4.79**	**149**	14.58	70
	Coverage	6.53	13	4.80	8	**12.41**	**65**

The selection of the Geoid model as performed in step 1 only affected the accuracy a little. The highest difference is 0.075 m between EGM96 and XGM2019e for the O06225 data. Sub-cm differences were obtained for the other data (TABLE III). After selecting quality flag = 1 data in step 2, the greatest accuracy increase was achieved for the O06225 data. There are many outliers in this data set due to clouds. There is an accuracy increase of nearly 50% in the O08061 data, but it could not reach the accuracy of the O07933 data level (TABLE III). It is seen that some outliers of Quality flag = 1 data are not detected in the F_7 and F_8 profiles (Fig. 4). After applying Step 3, the accuracy of the O06225 and O07933 data is similar, and most outliers were removed. Although there is a 50% increase in accuracy after applying step 3 on the O08061 data, it is still two times less accurate compared to the other two data. Also, in step 3, more outliers are removed in the O08061 data than in the other data. In Step 4, different RH levels calculated by the difference between lowest mode and selected RH metrics to determine the accurate terrain. Additionally, RMSE of lowest mode is calculated. Here, different results were obtained for all 3 data (Table III). For O06225 data RH20 provided the lowest RMSE of 0.76 m according to the Lowest mode. O07933 data showed the lowest RMSE of 0.44 m according to the difference between RH30 and the Lowest mode. The lowest mode showed the best results in O08061 data. In step 5, filtering according to sensitivity had less effect on the accuracy than in the previous steps. This might depend on the order of the criteria design. When points with Sensitivity> 0.93 are selected for the O06225 data, the accuracy is increased by 0.19 m, and 26 footprints are deleted. For the O07933 data, when footprints with Sensitivity > 0.94 are selected, the accuracy is increased by 0.32 m, whereas 395 points are deleted. In the O08061 data, when points with Sensitivity > 0.93 were selected, the accuracy increased by 0.07 m, whereas 51 points were deleted. In step 6, all data gave the best results in areas below 20% slope. Because the study area has many locations with high slope, a large number of footprints was removed (TABLE III). In the last step 7, full power beams are selected for the O06225 and O07933 data. In addition, when looking at the number of points, it is seen that the coverage beams are deleted more in the previous filtering steps. For the O08061 data, it is seen that coverage beams are more accurate. In addition, it was seen for the O08061 data that there was not a big difference between the number of full power and coverage footprint points (TABLE III). It is seen that there are systematic errors in Coverage beams (C_1 – C_4, C_19, C_20) (Fig. 4). There is no difference in accuracy between coverage and full power beams for the O07933 data (TABLE III). It is seen that the O08061 data is shifting in the latitude direction (Fig. 4).

IV. CONCLUSION

The GEDI satellite altimeter system is an important data source for earth observations. Atmospheric and geometric effects cause errors during data collection. It is very important to detect and correct these errors for high quality surface estimation. GEDI contains additional information (Quality flag, Sensitivity, etc.) to detect these errors and data. In this article, we aimed to determine how each variable and criterion affects terrain estimation. For that purpose, for the same region, 3 different GEDI orbits were examined, obtained with different solar elevation and cloudiness rates. The following main results were obtained.

1) In the orbit obtained at a solar elevation of -50.5 and during cloud free conditions 157 footprints are left from 1188 footprints with a final RMSE of 4.795 m.

2) In the orbit obtained at a solar elevation of -8.4 during cloudy conditions, 41 footprints from 1178 footprints remained with a final RMSE of 4.586 m.

3) In the orbit obtained with a solar elevation of -0.7 during cloud free conditions, from 1164 initial footprints 65 footprints remain, with a RMSE of 12.407 m.

Based on these results, it can be said that GEDI data is highly affected by cloudy data and solar elevation. But after rigorous filtering, data obtained during cloudy condition reached a similar quality of RMSE value as cloudless data, showing that affected data points can be effectively identified. In future studies, it is considered to carry out studies to determine thresholds for solar elevation. Because, although the GEDI user guide, [19], states that data with a solar elevation < 0 is obtained at night, our results seem to indicate that the sun already cause errors in data acquired close to sunrise or sunset.

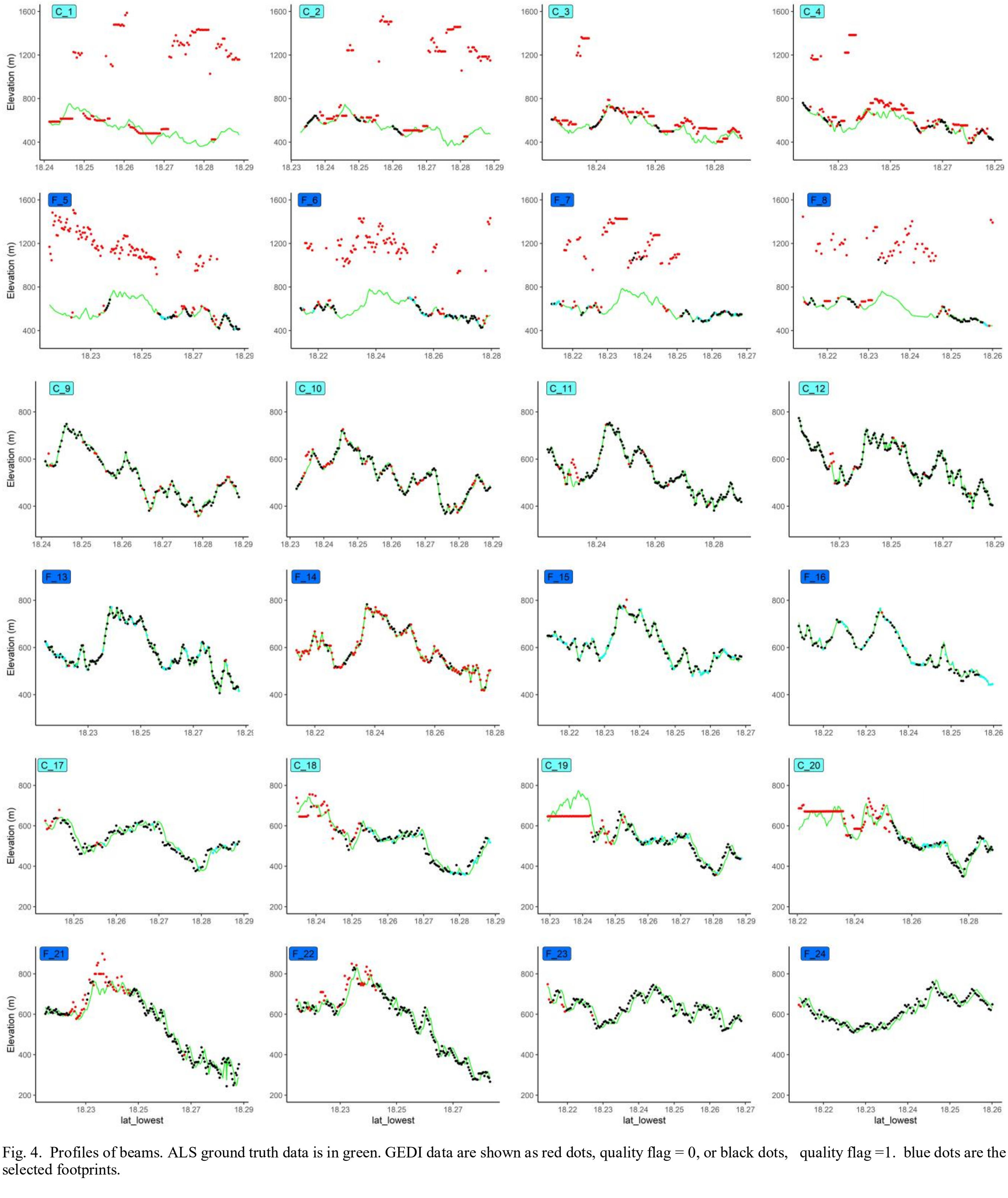

Fig. 4. Profiles of beams. ALS ground truth data is in green. GEDI data are shown as red dots, quality flag = 0, or black dots, quality flag =1. blue dots are the selected footprints.

Acknowledgment

This research was supported by the Afyon Kocatepe University (Project Number: 22.FEN.BİL.02). The first author is financially supported by TUBITAK 2214/A International Research Fellowship Programme for Ph.D. Students.

References

[1] R. Dubayah, et al. "The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography." Science of remote sensing 1 (2020): 100002.

[2] C. Vatandaslar, O. G. Narin, and S. Abdikan. "Retrieval of forest height information using spaceborne LiDAR data: a comparison of GEDI and ICESat-2 missions for Crimean pine (Pinus nigra) stands." Trees (2022): 1-15.

[3] Z. Zhang, Y. Bo, S. Jin, G. Chen and Z. Dong. "Dynamic water level changes in Qinghai Lake from integrating refined ICESat-2 and GEDI altimetry data (2018–2021)." Journal of Hydrology 617 (2023): 129007.

[4] A. Liu, X. Cheng and Z. Chen. "Performance evaluation of GEDI and ICESat-2 laser altimeter data for terrain and canopy height retrievals." Remote Sensing of Environment 264 (2021): 112571.

[5] M. Adam, M. Urbazaev, C. Dubois and C. Schmullius. "Accuracy assessment of GEDI terrain elevation and canopy height estimates in European temperate forests: Influence of environmental and acquisition parameters." Remote Sensing 12.23 (2020): 3948.

[6] Y. Ngo, Y. Huang, D. H. T. Minh, L. Ferro-Famil, I. Fayad and N. Baghdadi. "Tropical Forest vertical structure characterization: From GEDI to P-band SAR tomography." IEEE Geoscience and Remote Sensing Letters 19 (2022): 1-5.

[7] E. Quiros, M. Polo, and L. Fragoso-Campón. "GEDI elevation accuracy assessment: a case study of southwest Spain." IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 14 (2021): 5285-5299.

[8] F. L. R Barbosa, R. F. Guimarães, O. A. de Carvalho Júnior, R. A. T. Gomes, O. L. F. de Carvalho, and T. P. M. de Lima. "Estimating the Optimal Threshold for Accuracy Assessment of the Global Ecosystem Dynamics Investigation (GEDI) Data in a Gentle Relief Urban Area." Remote Sensing, 14(15) (2022): 3540.

[9] U.S. Geological Survey "3D Elevation Program 1-Meter Resolution Digital Elevation Model (PR USVI 2018 Lidar)", 2020. [Online]. Available: https://www.usgs.gov/3d-elevation-program

[10] C. A. Silva et al., "rGEDI: NASA's global ecosystem dynamics investigation (GEDI) data visualization and processing," R package version 0.1.7 ed, 2020. [Online]. Available: https://github.com/carlos-alberto-silva/rGEDI

[11] I. Yilmaz, M. Yilmaz, M. Gullu and B. Turgut. "Evaluation of recent global geopotential models based on GPS/levelling data over Afyonkarahisar (Turkey)." Scientific research and essays 5.5 (2010): 484-493.

[12] Lemoine, Frank G., et al. The development of the joint NASA GSFC and the National Imagery and Mapping Agency (NIMA) geopotential model EGM96. No. NAS 1.60: 206861. 1998.

[13] N. K. Pavlis, S. A. Holmes, S. C. Kenyon and J. K. Factor. "The development and evaluation of the Earth Gravitational Model 2008 (EGM2008)." Journal of geophysical research: solid earth 117.B4 (2012).

[14] P. Zingerle, R. Pail, T. Gruber and X. Oikonomidou. "The combined global gravity field model XGM2019e." Journal of Geodesy 94 (2020): 1-12.

[15] W. Liand, J. Li, X. Xu, S. Zhang and Y. Zhao. "A high-resolution Earth's gravity field model SGG-UGM-2 from GOCE, GRACE, satellite altimetry, and EGM2008." Engineering 6.8 (2020): 860-878.

[16] Ince, E. Sinem, et al. "ICGEM–15 years of successful collection and distribution of global gravitational models, associated services, and future plans." Earth system science data 11.2 (2019): 647-674.

[17] Shang, Dashuai, et al. "Extraction Strategy for ICESat-Elevation Control Points Based on ATL08 Product." IEEE Transactions on Geoscience and Remote Sensing 60 (2022): 1-12.

[18] S. Hancock, et al. "The GEDI simulator: A large-footprint waveform lidar simulator for calibration and validation of spaceborne missions." Earth and Space Science 6.2 (2019): 294-310.

[19] GEDI Documents "GLOBAL Ecosystem Dynamics Investigation (GEDI) Level 2 User Guide", 2021. [Online]. Available: https://lpdaac.usgs.gov/documents/986/GEDI02_UserGuide_V2.pdf