# **Comparison of Filtering Techniques for Water Level Estimation Using GEDI**

Omer Gokberk Narin<sup>1</sup>

<sup>1</sup> Department of Geomatics Engineering, Afyon Kocatepe University, Afyonkarahisar, Turkey - gokberknarin@aku.edu.tr

Keywords: Spaceborne LiDAR, GEDI, Data Filtering, Burdur Lake, Water Level Estimation.

### Abstract

Global Ecosystem Dynamics Investigation (GEDI) satellite-based LiDAR altimeter data can used to determine water levels. However, GEDI LiDAR beams may contain errors due to factors such as atmospheric conditions and topographic features. Therefore, it is essential to filter out data that contains errors and does not meet the required conditions in water level estimation. In this study, two different filtering techniques have been used for estimation of water levels. The first technique utilizes auxiliary data such as quality flag, sensitivity, and solar elevation, which are provided together with the GEDI data. On the other hand, the second technique utilizes the Interquartile Range (IQR) method. The Burdur Lake, a Ramsar site located in the Southwest of the Turkey, was selected as a test area. Daytime data from 18 October 2019 and nighttime data from 14 January 2020 were downloaded for the Burdur Lake. The Root Mean Square Error (RMSE) for the daytime data was calculated as 19.295 m. After filtering the data with auxiliary attributes, the RMSE value is decreased to 0.258 m. After applying the IQR filtering technique, the RMSE value was obtained as 0.317 m. For the nighttime data, the RMSE value was calculated as 0.292 m before filtering. The auxiliary data filtering was decreased the RMSE value into 0.118 meters. As a result of filtering with the IQR method, the RMSE value was obtained as 0.266 m. It was concluded that filtering is necessary to estimate water levels with GEDI data, especially for daytime data.

#### 1. Introduction

Freshwater is an indispensable resource for the sustainability of life on Earth. In addition to its use as drinking water for humans, it plays a critical role in agricultural irrigation and industrial activities. It is also necessary for the healthy functioning of ecosystems (Jackson et al., 2001). Freshwater resources ensure ecological balance in a wide range of areas, from the protection of biodiversity to the sustainability of food chains. However, population growth, climate change and overexploitation of water resources are seriously threatening the quantity and quality of freshwater. In particular, surface freshwater resources such as lakes are both sensitive to climate change and respond rapidly to environmental pressures (Carpenter et al., 1992; Woolway et al., 2020). Therefore, monitoring and management of freshwater resources is essential to prevent water scarcity and maintain the health of ecosystems. Monitoring changes in lake levels plays a critical role in understanding the regional hydrological cycle and ecological impacts, and providing scientific information for the sustainable management of freshwater resources (Akbas, 2024).

Gauge stations can monitor real-time water levels in rivers, canals, lakes, or reservoirs. However, there may be regions where these stations require more effort or cannot install, maintain, and use. On the other hand, remote sensing data are becoming increasingly important for global water level monitoring. In this context, the Global Ecosystem Dynamics Investigation (GEDI) mission, which started collecting data in March 2019, is an important data source for water levels monitoring. GEDI LiDAR altimeter system, mounted on the International Space Station (ISS), consists of 3 lasers beams. 2 of these laser beams work as full-power lasers and one as half power laser (i.e., coverage beams). At any time, four beams fall to the ground from these three lasers and create eight ground tracks. These beams are about ~25 m in diameter (Dubayah et al.2020). By using these beams, GEDI data has been used in many studies, such as aboveground biomass estimation (Dubayah et al., 2022), canopy height estimation (Adam et al., 2020), terrain height estimation (Narin et al., 2023), building height estimation (Kaya, 2024), and water level estimation (Fayad et al., 2020; Fayad et al., 2021; Frappart et al., 2024; Hamoudzadeh et al., 2023; Lee et al., 2024). The GEDI data may be inaccurate due to error sources such as atmospheric effects, ISS orbit effects, etc. Therefore, filtering techniques should be used to detect insufficient GEDI beams and before utilized the GEDI data in various applications.

Fayad et al. (2020) analyzed the accuracy of GEDI data in estimating water level in eight different lakes in Switzerland. They used the height difference between Shuttle Radar Topography Mission (SRTM) height values and GEDI beams (|GEDI elevation-SRTM| > 100 m) and GEDI auxiliary data to filter out outliers in GEDI data. As a result of the study, they found that the error in water level detection according to Root Mean Squared Error (RMSE) was between 0.145 m and 0.316 m. Frappart et al. (2024) selected Lake Issyk-Kul in Kyrgyzstan as a test area in their study. In their study, the accuracy of 1239 transits of GEDI data between May 2019 and mid-November 2021 were They first used the GEDI auxiliary data evaluated. (num detectedmodes) and the height difference of the GEDI beams with SRTM (|GEDI elevation-SRTM| > 50 m) to filter the GEDI data. Then, the deviations on either side of the median value were identified as outliers. As a result of the study, the error of water level estimation increases at the lake edges. When they looked at all water level results, it was determined that the accuracy of GEDI data was 0.163 m according to RMSE. Lee et al. (2024) analyzed the accuracy of GEDI data in determining water level for seven lakes in North America. They used a 15step method using GEDI auxiliary data to filter out outliers in GEDI data. As a result of filtering the GEDI data, they found that the error in water level estimation according to RMSE in 7 lakes was between 0.380 m and 0.470 m. Fayad et al. (2021) analyzed the difference in accuracy between version-1 and version-2 of GEDI data. For this research, GEDI data from Geneva Lake between April 2019 and September 2020 were analyzed. In order to detect outliers in the GEDI data, they initially utilized the number of detected modes (num detected modes  $\neq$  1), which is an auxiliary attribute of the GEDI data. If num\_detectedmodes was not equal to 1, the data was considered an outlier. Later, they looked at the height differences between the SRTM and GEDI beams (|GEDI elevation-SRTM > 100 m). As a result of the study, they

concluded that version-2 gives better results than version-1 and that GEDI accuracy varies with dates. Hamoudzadeh et al. (2023) analyzed four lakes in northern Italy to test the success of GEDI data in determining water level. They used a two-stage filtering technique to detect outliers in GEDI data in their study. Firstly, they used quality flag and degrade flag attributes, which are auxiliary data of each GEDI beams. After extracting the outliers, they applied the 3NMAD test. As a result of their study, after filtering the GEDI data according to RMSE, they found that the error in estimating the water level according to RMSE in four lakes was between 0.410 m and 0.470 m.

The objectives of this paper, which selected the Burdur Lake as the study area, (i) it was analyze whether the time of obtaining the GEDI data (daytime & nighttime) affects the accuracy; (ii) to determine which method is better at removing the outliers found in the GEDI data and which method gives better results. Two different techniques were used to filter insufficient GEDI beams. The first technique is the detection of outliers using GEDI auxiliary data, and the second technique is the detection of outliers using the Interquartile Range (IQR) method.

RMSE and Median Absolute Deviation (MAD) accuracy metrics was utilized to evaluate the performance of the techniques. The article is organized as follows: Section 2 presents information about the study area, GEDI data, in-situ data, filtering techniques, and accuracy metrics. Section 3 evaluates the results obtained from filtering techniques. Section 4 provides information about the limitations of the study and future work. Section 5 presents the conclusion.

### 2. Materials and Method

# 2.1 Study Area

Burdur Lake, located in southwest Turkey, is one of the lakes in the lake region (Figure 1). The lake is categorized as a tectonic lake, is in a closed basin with no outflow. Burdur Lake, an important habitat for biodiversity and bird species, is also protected under the Ramsar Convention (Adaman et al., 2009). It has been determined that Burdur Lake has decreased in water level and lake area between 1975 and 2017 (Yıldırım & Uysal, 2011; Gözükara et al., 2019). Burdur Lake is also an important habitat for the endangered white-headed duck (Oxyura leucocephala). The boundaries of the basin (Lehner & Grill, 2013), lakes (Messager et., 2016) and streams (Lehner & Grill, 2013) used in the Figure -1 were obtained from Hydrosheds.

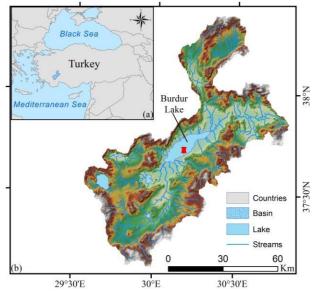


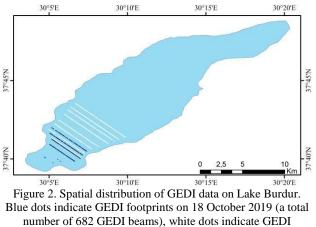
Figure 1. (a): Location of the basin on the Turkey. (b): Map of the Burdur Basin. Red pin is observation station.

# 2.2 GEDI

The ISS-mounted GEDI is a LiDAR altimeter sensor for measuring 3-dimensional (3D) information about Earth's topography, forest cover and water elevation. This ISS-integrated sensor collects data between parallels 52°N and 52°S. This sensor sends 8 beams to the earth's surface at the same time. Four of these beams are called as full power lasers (Beam 0101, 0110, 1000 and 1011) and other four beams are called as coverage lasers (beam 0000, 0001, 0010 and 0011). These lasers have a footprint of ~25 m in diameter. GEDI presents the data at four different levels. These levels include raw waveforms (L1A), geolocated waveforms (L1B), ground elevation (L2A), canopy top height, relative height, canopy cover fraction, leaf area index (L2B), gridded level 2 metrics (L3), footprint level above ground biomass (L4A), gridded above ground density (L4B). These products at different levels are available free of charge (Dubayah et al., 2020).

In this study, L2A products of the GEDI were utilized for water level estimation. GEDI L2A is suitable for ground elevation, canopy height, water level detection applications. The system, which works on the full wavelength principle, is represented as different wavelengths on different surfaces. For example, wavelength graphs have more than one mode in forests and therefore have a multimodal shape. However, it is expected to be represented by a single mode on water surfaces (Fayad et al., 2020). GEDI L2A data were downloaded using the Google Earth Engine (GEE) cloud platform (Gorelick et al., 2017). The GEDI L2A Raster Canopy Top Height (Version 2) dataset in the GEE catalogue has 136 attributes (Earth Engine Data Catalog - Google Developers 2022). Eleven of these 136 attributes were downloaded for use in the study. These attributes are beam, gradient\_flag, delta\_time, elev\_highestreturn, elev\_lowestmode, lat highestreturn, lon\_highestreturn, num\_detectedmodes, quality\_flag, sensitivity, and solar\_elevation. GEDI data were downloaded on 18 October 2019 (average solar\_elevation = 24.165), which is called daytime data, and on 14 January 2020 (average solar\_elevation = -33.951), which is called nighttime data. While downloading the data, a 250 m buffer was applied inwards from the lake border in order to eliminate the marketizes that may occur near the lake border. The heights of the downloaded data were reduced into the EGM 2008 geoid model.

The spatial distribution of GEDI data over Lake Burdur is given in Figure 2.



number of 682 GEDI beams), white dots indicate GEDI footprints on 14 January 2020 (a total number of 299 GEDI beams).

# 2.3 In-Situ Observation

In-situ water levels were obtained from the General Directorate of State Hydraulic Works (GDSHW), Ministry of Agriculture and Forestry, Republic of Turkey. In-situ measurement data are not openly shared, and an application is required to obtain the data. GDSHW monitors the water levels of lakes and reservoirs across Turkey on a daily basis. In the study, each GEDI data was compared with the daily average water level obtained from GDSHW.

# 2.4 Outlier Removal

GEDI data can be affected by atmospheric conditions, cloudiness, slope, solar radiation. Therefore, filtering is required before estimating water level from GEDI data (Fayad et al., 2020; Fayad et al., 2021; Frappart et al., 2024; Hamoudzadeh et al., 2023; Lee et al., 2024). In this study, GEDI data were filtered in two different techniques. One of them was determined as filtering using GEDI auxiliary data. In this technique, the effect of solar\_elevation, quality\_flag, sensitivity attributes on the accuracy were analyzed separately. Then, by using all auxiliary data, the most error-free data in the GEDI data were tried to be detected. IQR technique was used as the second filtering technique. IQR technique can also called as box plot or box & whisker, used in graphical representation. The IQR technique helps to measure the central spread in continuously distributed data and to find outliers by determining extreme values with formulas based on the median (Narin & Abdikan 2023; Vinutha et al., 2018; Wan et al., 2014). When calculating the IQR, the data are ordered from smallest to largest, then the median, Q1 and Q3 are calculated (Figure 3). IQR is the difference between Q1 and Q3. If the value in the data set falls outside the range  $Q1 - 1.5 \times$ *IQR* to  $Q3 + 1.5 \times IQR$ , it is described as an outlier. The general workflow of the study is given in Figure-4.

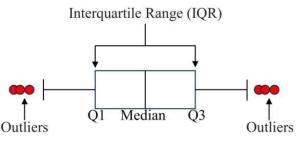


Figure 3. Boxplot representation of the IQR method.

### 2.5 Accuracy Assessment Metrics

1

The height of each beam obtained from the GEDI data wereassessed with in-situ observation data. Two different metrics were used to evaluate the performance of the GEDI data. These metrics are RMSE (Equation 1) and Median Absolute Deviation (MAD) (Equation 2).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(g-t)^2}$$
(1)

$$MAD = median(|g - t|)$$
(2)

where n is the number of GEDI beams, g is the water level of insitu observation, and t is the water level by GEDI (elev\_lowestmode).

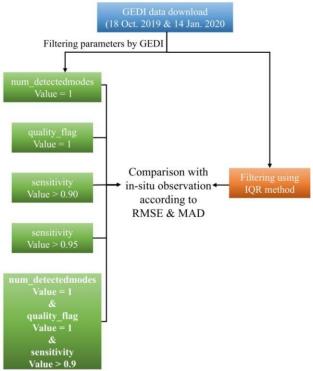
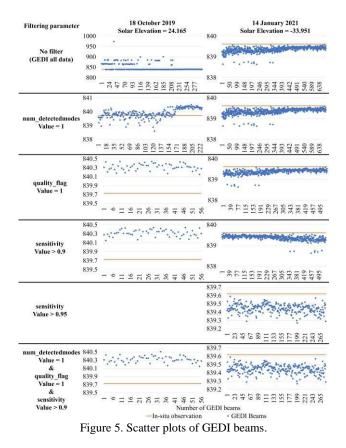


Figure 4. The workflow of filtering and comparison of filtering techniques for water level estimation.

#### 3. Results

As mentioned earlier, GEDI data were downloaded on two different dates when solar elevations were different, represents the daytime and nighttime data. Initially the effect of solar elevation on the accuracy is analyzed on unfiltered GEDI data. Then, the accuracy after filtering with each GEDI attribute is analyzed separately. These results are compared with the results of the data filtered with the IQR technique. RMSE and MAD accuracy metrics were used to compare the results. Table-1 summarizes all the calculated accuracy metrics for both dates.

When the first results are analyzed, 682 GEDI beams were obtained in the nighttime data of the GEDI data, while 299 GEDI beams were obtained in the daytime data (Figure 5). The results of all GEDI beams in the nighttime data showed that RMSE = 0.292 m and MAD = 0.230 m. In the daytime data, the results of all GEDI beams showed that RMSE = 19.295 m, MAD = 0.410m (Table 1). The first results show that there are many outliers exist in the daytime data (Figure 5). GEDI beams with num\_detectedmodes = 1 were selected from all data. In the night data, it was observed that all data had num\_detectedmode = 1. However, in the daytime data, 226 out of 299 GEDI beams had num\_detectedmode = 1. After removing the data with num detectedmode  $\neq 1$  in the daytime data, RMSE and MAD values were calculated as 0.331 m and 0.210 m, respectively (Table 1). When num\_detected mode  $\neq$  1 on water surfaces, it is understood that there are more than 1 mode at GEDI wavelength. In these cases, it can be concluded that the signal was affected by different objects. Figure 5 shows that there are GEDI beams higher than the water level (between ~20 m and ~130 m) in the GEDI data dated 18 October 2019 (daytime). GEDI beams with high error could be extracted using num\_detectedmode attribute. Then only beams with quality\_flag = 1 were selected from all GEDI beams. In the night data, 529 GEDI beams with quality\_flag = 1 attribute were detected. These beams showed an accuracy of RMSE = 0.233 m and MAD = 0.210 m. A total of 57 GEDI beams in daytime data with quality\_flag = 1 were detected. These beams showed an accuracy of RMSE = 0.258 m and MAD = 0.580 m. Two different thresholds were determined for filtering with sensitivity, another auxiliary attribute of GEDI. These thresholds are sensitivity > 0.90 and sensitivity > 0.95, respectively. When 531 GEDI beams with sensitivity > 0.9 were selected for the nighttime data (14 January 2021), RMSE = 0.234 m and MAD = 0.210 m. In the daytime data (18 October 2019), it was observed that all beams with quality\_flag = 1 had sensitivity values between 0.90 and 0.95. For this reason, it provided the same results with quality\_flag = 1. When 280 GEDI beams with sensitivity > 0.95 were selected for the night data, RMSE = 0.118 m and MAD = 0.17 m. In the filtering using all GEDI auxiliary data, the daytime data showed RMSE = 0.258 m, MAD = 0.580 m accuracy. The night data showed an accuracy of RMSE = 0.118 m, MAD = 0.170 m.



All GEDI beams were finally filtered according to the IQR method. Before filtering, box-plot graphs were plotted for daytime and nighttime data (Figure 6). Figure 6a shows that there are large number of outliers exist in the daytime data before filtering. It is seen that a total number of 77 GEDI beams are detected as outlier in the daytime data (Figure 6a). After removing these outliers, the box plot shows a more balanced distribution of GEDI beams (Figure 6c). When the box-plot graph of the night data was plotted, a total number of 27 GEDI beams were detected as outliers (Figure 6b). After removing these outliers, a balanced distribution of GEDI beams was provided (Figure 6d).

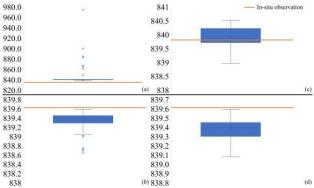


Figure 6. Box-plot representation of the filtering of GEDI beams according to the IQR method. (a) unfiltered daytime data (18 October 2019) (b) filtered daytime data (c) unfiltered nighttime data (14 January 2021) (d) filtered nighttime data.

In general, when filtering techniques were compared, GEDI beams with sensitivity > 0.95 showed the highest accuracy according to RMSE and MAD metrics in nighttime data (Table 1). In daytime data, IQR method showed the highest accuracy according to RMSE and MAD metrics (Table 1). GEDI data showed an accuracy between 0.118 m and 0.317 m according to RMSE after filtering for water level estimation. In addition, GEDI overestimated the water level in daytime data and underestimated it in nighttime data (Figure 5).

_	18 October 2019		14 January 2021	
Filtering scenarios	RMSE (m)	MAD (m)	RMSE (m)	MAD (m)
GEDI all beam	19.295	0.41	0.292	0.23
num_detectedmodes = 1	0.331	0.21	0.292	0.23
quality_flag = 1	0.258	0.59	0.233	0.21
sensitivity > 0.9	0.258	0.59	0.234	0.21
sensitivity > 0.95	No beams		0.118	0.17
All filter	0.258	0.59	0.118	0.17
IQR Method	0.317	0.21	0.266	0.23

Table 1. Overall statistics for filtering scenarios between in-situ observations and GEDI beams.

# 4. Limits and Future Work

One of the limitations of the study is the reduction of the WGS84 height of the GEDI beams to the reference ellipsoid using a spherical model. If the regional model is used, the error due to the global model could be minimized. The other limitation is the use of single data only at different dates and different solar elevations for comparison. In future studies, more than one day and night data of the same lake can be compared and the effect of solar elevation on water level estimation can be discussed and determined more comprehensively.

### 5. Conclusion

As a result, the water level estimates of GEDI beams were compared with in-situ observations. In the comparison, firstly, daytime and nighttime acquisitions of GEDI were compared. Then, different parameters were used in the filtering stages to determine whether the results were better. After filtering, it was concluded that GEDI data can estimate the water level with better results than 0.30 m according to RMSE and 0.21 m according to MAD, regardless of whether it is day or night. As a result of filtering, it was observed that night data provided 0.14 m better results than day data. Considering these results, it is seen that GEDI data has an important potential in monitoring water reservoirs around the Earth. However, it is thought that the Multisensing Observation Lidar and Imager Demonstration (MOLI) LiDAR altimeter sensor, which will start a new mission in the future when the GEDI mission is completed, will also create a very important potential for water level estimation.

# Acknowledgements

The author thanks to the National Aeronautics and Space Administration and the University of Maryland for providing the GEDI LiDAR altimetry data free of charge.

### References

Adam, M., Urbazaev, M., Dubois, C., Schmullius, C., 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.

Adaman, F., Hakyemez, S., Özkaynak, B., 2009. The political ecology of a Ramsar site conservation failure: the case of Burdur Lake, Turkey. *Environment and Planning C: Government and Policy*, 27(5), 783-800.

Akbas, A., 2024. Human or climate? Differentiating the anthropogenic and climatic drivers of lake storage changes on spatial perspective via remote sensing data. *Science of The Total Environment*, 912, 168982.

Carpenter, S.R., Fisher, S.G., Grimm, N.B., Kitchell, J.F., 1992. Global change and freshwater ecosystems. *Annual review of ecology and systematics*, 119-139.

Dubayah, R., Blair, J.B., Goetz, S., Fatoyinbo, L., Hansen, M., Healey, S., ..., Silva, C., 2020. The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography. *Science of Remote Sensing*, 1, 100002.

Dubayah, R., Armston, J., Healey, S.P., Bruening, J.M., Patterson, P.L., Kellner, J.R., ..., Luthcke, S., 2022. GEDI launches a new era of biomass inference from space. *Environmental Research Letters*, 17(9), 095001.

Earth Engine Data Catalog — Google Developers, 2022. GEDI L2A raster canopy top height (version 2). https://developers.google.com/earth-engine/datasets/catalog/ LARSE\_GEDI\_GEDI02\_A\_002\_MONTHLY#description

Fayad, I., Baghdadi, N., Bailly, J.S., Frappart, F., Zribi, M., 2020. Analysis of GEDI elevation data accuracy for inland waterbodies altimetry. *Remote Sensing*, 12(17), 2714.

Fayad, I., Baghdadi, N., Frappart, F., 2022. Comparative analysis of GEDI's elevation accuracy from the first and second data product releases over inland waterbodies. *Remote Sensing*, 14(2), 340.

Frappart, F., Minh, D.H.T., Baghdadi, N., Crétaux, J.F., Fayad, I., Bergé-Nguyen, M., 2024. Improving mean water lake surface elevation estimates using dense lidar measurements from the GEDI satellite mission. *Remote Sensing Applications: Society and Environment*, 35, 101213.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote sensing of Environment*, 202, 18-27.

Gözükara, G., Altunbaş, S., Sarı, M., 2019. Burdur Gölü'ndeki seviye değişimi sonucunda ortaya çıkan lakustrin materyalin zamansal ve mekansal değişimi. *Anadolu Tarım Bilimleri Dergisi*, 34(3), 386-396. (In Turkish)

Hamoudzadeh, A., Ravanelli, R., Crespi, M., 2023. GEDI data within google earth engine: preliminary analysis of a resource for inland surface water monitoring. *International Archives of The Photogrammetry, Remote Sensing And Spatial Information Sciences*, 48(M-1-2023), 131-136.

Jackson, R.B., Carpenter, S.R., Dahm, C.N., McKnight, D.M., Naiman, R.J., Postel, S.L., Running, S.W., 2001. Water in a changing world. *Ecological Applications*, 11(4), 1027-1045.

Kaya, Y., 2024. Automated Estimation of Building Heights with ICESat-2 and GEDI LiDAR Altimeter and Building Footprints: The Case of New York City and Los Angeles. *Buildings*, 14(11), 3571.

Lee, K., Hofton, M., Blair, B., Tang, H., 2024. Analysis of GEDI Elevation Measurement Accuracy on Land Surface and Waterbody in North America. *IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium*, 4329-4332). IEEE.

Lehner, B., Grill, G., 2013. Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. *Hydrological Processes*, 27(15), 2171-2186.

Messager, M.L., Lehner, B., Grill, G., Nedeva, I., Schmitt, O., 2016. Estimating the volume and age of water stored in global lakes using a geo-statistical approach. *Nature Communications*, 7(1), 13603.

Narin, O.G., Abdikan, S., 2023. Multi-temporal analysis of inland water level change using ICESat-2 ATL-13 data in lakes and dams. *Environmental Science and Pollution Research*, 30(6), 15364-15376.

Narin, O.G., Lindenbergh, R., Abdikan, S., 2023. MultiCriteria Strategy for Estimating GEDI Terrain Height. In 2023 10th International Conference on Recent Advances in Air and Space Technologies (RAST), 1-6. IEEE.

Vinutha, H.P., Poornima, B., Sagar, B.M., 2018. Detection of outliers using interquartile range technique from intrusion dataset. In *Information and decision sciences: Proceedings of the 6th international conference on ficta*, 511-518. Springer Singapore.

Woolway, R.I., Kraemer, B.M., Lenters, J.D., Merchant, C.J., O'Reilly, C.M., Sharma, S., 2020. Global lake responses to climate change. *Nature Reviews Earth & Environment*, 1(8), 388-403.

Wan, X., Wang, W., Liu, J., Tong, T., 2014. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC medical research methodology*, 14, 1-13.

Yıldırım, Ü., Uysal, M., 2011. Changes in the Coastline of the Burdur Lake Between 1975 and 2010. In *International* symposium on environmental protection and planning: Geographic Information Systems (GIS) and Remote Sensing (RS) Applications (ISEPP), 28, 29.