GEDI DATA WITHIN GOOGLE EARTH ENGINE: PRELIMINARY ANALYSIS OF A RESOURCE FOR INLAND SURFACE WATER MONITORING

Alireza Hamoudzadeh [1],*, Roberta Ravanelli [1], Mattia Crespi [1,2]

[1] Sapienza University of Rome, DICEA, Geodesy and Geomatics Division, Italy
[2] Sapienza School for Advanced Studies, Sapienza University of Rome, Italy

<alireza.hamoudzadeh,roberta.ravanelli,mattia.crespi>@uniroma1.it

KEY WORDS: Inland Surface Water Monitoring, Sustainable Development Goals, Google Earth Engine, GEDI

ABSTRACT:

Freshwater is one the most important renewable water resources of the planet but, due to climate change, surface freshwater available in the form of lakes, rivers, reservoirs, snow, and glaciers is becoming significantly threatened. As a result, surface water level monitoring is fundamental for understanding climatic changes and their impact on humans and biodiversity.

This study evaluates the accuracy of the Global Ecosystem Dynamics Investigation (GEDI) LiDAR (Light Detection And Ranging) instrument for monitoring inland water levels. Four lakes in northern Italy were selected for comparison with gauge station measurements. To evaluate the accuracy of GEDI altimetric data, two steps of outlier removal are proposed. The first stage employs GEDI metadata to filter out footprints with very low accuracy. Then, a robust version of the standard 3σ test using a 3NMAD (Normalized Median Absolute Deviation) test is iteratively applied.

After the outlier removal, which led to the elimination of between 80% to 87% of the data, the remaining footprints show an average standard deviation of 0.36 m, a mean NMAD of 0.38 m, and a Root Mean Square Error (RMSE) of 0.44 m, proving the promising potentialities of GEDI L2A altimetric data for inland water monitoring.

1. INTRODUCTION

Inland surface water is the source for about two-thirds of the freshwater for human and animal consumption, agricultural irrigation, and several industrial applications, and a key component of the hydrological cycle. The monitoring of inland surface water is therefore fundamental to understand the effects of climate change on this key resource and to prevent water stresses (Bocchino et al., 2023). This kind of monitoring is strictly related to United Nations (UN) Sustainable Development Goals (SDGs) concerning water availability (SDG 6 - Clean water and sanitation) and climate change effects monitoring (SDG 13 - Climate action) and with the Recovery Plan Next Generation EU (Kavvada et al., 2020, United Nations General Assembly, 2015, Sinha et al., 2020).

Traditionally, water level measurements are obtained using ground-based instruments like gauge stations. However, such in-situ monitoring techniques are typically only feasible in developed countries, since installing and maintaining measurement stations in remote areas can be challenging (Hamoudzadeh et al., 2023). Conversely, the use of Earth observation technologies and methods can remarkably reduce the monitoring costs (independent from the actual extent of the reservoir) and provide frequent and regular data, that facilitate the continuous monitoring of water reservoirs, in principle with homogeneous procedures worldwide.

Access to new global datasets is critical to improving this type of monitoring on a global scale, provided that their accuracy is thoroughly assessed. By taking advantage of Earth observation technologies, such as RADAR (Radio Detection And Ranging) and LiDAR (LIght Detection And Ranging), it is possible to obtain high-quality data with global coverage, enabling a more comprehensive and consistent understanding of water reservoir dynamics worldwide. However, it is important to ensure that the accuracy and reliability of these new data sources are carefully evaluated to enable effective and trustworthy monitoring on a global scale.

In this respect, Google Earth Engine (GEE) is known as a reliable, and real-time cloud-based computation platform, capable of integrating a high variety of up-to-date geospatial datasets with powerful analysis tools (Cardille et al., 2022). GEE provides a High-Performance Computing (HPC) infrastructure that enables fast and convenient access to more than forty years of publicly available data archives, comprising scientific datasets and historical imagery (Ravanelli et al., 2018a, Ravanelli et al., 2018b, Nascetti et al., 2017), making it possible to develop remote sensing applications on a global and large scale. GEE has recently added the Global Ecosystem Dynamics Investigation (GEDI) (University of Maryland, 2022) dataset to its already wide archive.

GEDI was originally developed as an experimental mission onboard the International Space Station (ISS) to enable radically improved quantification and understanding of the Earth’s carbon cycle and biodiversity; only lately have GEDI potentialities been investigated for inland surface water level monitoring (Fayad et al., 2022), despite many research has evaluated the accuracy and usability of different LiDAR, RADAR, and SAR (Synthetic Aperture Radar) altimetric data for this purpose (Lee et al., 2021).

The available literature highlights that the quality of GEDI data is variable and impacted by several factors (e.g., latitude, time

---

* Corresponding author

https://doi.org/10.5194/isprs-archives-XLVIII-M-1-2023-131-2023

of observation). Our preliminary analysis is therefore focused on the accuracy assessment of the data collected by GEDI, at first addressing the problem of outliers detection and removal within the GEE platform, and secondly comparing the water levels measured by GEDI with some reference ground truth.

2. DATA

2.1 Gauge measurements

In order to evaluate the accuracy of GEDI, four lakes located in northern Italy (Como, Garda, Iseo, Maggiore (Figure 1)) were chosen as the study area. These lakes have existing gauge stations that provide level measurements for comparison after the outlier removal procedure, and for this objective, the daily level measurements were gathered from the website of (Enti Regolatori dei Grandi Laghi, 2022).

2.2 GEDI Altimetric data

The GEDI (Dubayah et al., 2021) instrument is a geodetic-class, light detection and ranging (LiDAR) laser system consisting of 8 parallel observation tracks (beams) generated from 3 lasers (Liu et al., 2021), with the main objective of precise measurements of forest canopy height, canopy vertical structure, and surface elevation. Each laser emits 242 pulses per second, illuminating a 25 m spot (a footprint) on the surface over which 3D structure is measured. Each footprint is separated by a distance of 60 m along-track, with an across-track distance of about 600 m between each of the 8 tracks. The measurements are made over the Earth’s surface nominally between the latitudes of 51.6° and -51.6°.

The standard GEDI data products have some limitations for practical applications, as the footprint-level products provide only a point sample of a limited part of the land surface, leaving most of it without observations. However, for areas with homogeneous surfaces like water bodies, this limitation does not pose a significant issue as much as it would for targeting rare forest change events (Qi et al., 2019). For this study, GEDI L2A (Version 2) data was utilized. This data has 140 different bands and is available in both forms of Image Collection and Vectors within the GEE data catalog (Earth Engine Data Catalog — Google Developers, 2022).

Although GEDI data has been extensively studied for forest monitoring and estimating canopy height, relatively few studies are dedicated to water level monitoring using this altimetric sensor. The potential of GEDI for inland water level monitoring can be recognized due to its high spatial resolution and frequent revisit time, which can provide valuable data for monitoring changes in water levels over time. However, there are still challenges in using GEDI data for water level monitoring, such as the limited coverage of water bodies by GEDI due to lower latitude coverage compared to its rivals like ICESat-2 and the need for accurate calibration and validation against ground-based measurements. Further research is needed to fully explore the potential of GEDI for water level monitoring and to develop effective methodologies for using GEDI data in water resource management, and here we try to preliminarily evaluate these potentialities.

3. METHODOLOGY

The methodology proposed for the outlier removal required a specific procedure to be implemented within the GEE environment. However, some challenges due to the large amount of data to be processed were addressed in several pre-processing steps. These challenges are further discussed before delving into the detailed description of the methodology.

3.1 Pre-processing and footprint selection

The delineation of water bodies is an essential aspect of remote sensing due to the crucial role that EO imagery plays in managing water resources. The ability to accurately extract water boundaries from satellite imagery provides valuable information for various applications, such as water quality assessments, flood risk mapping, and hydrological modeling (Kaplan and Avdan, 2017).

The JRC Global Surface Water Mapping Layers, v1.4 is a widely used dataset that provides detailed information about water bodies worldwide, including their extent and shoreline characteristics (Pekel et al., 2016). By utilizing this dataset, we were able to extract the lake boundaries in a vector form to select the appropriate GEDI footprints within each lake for the analysis (Figure 2).

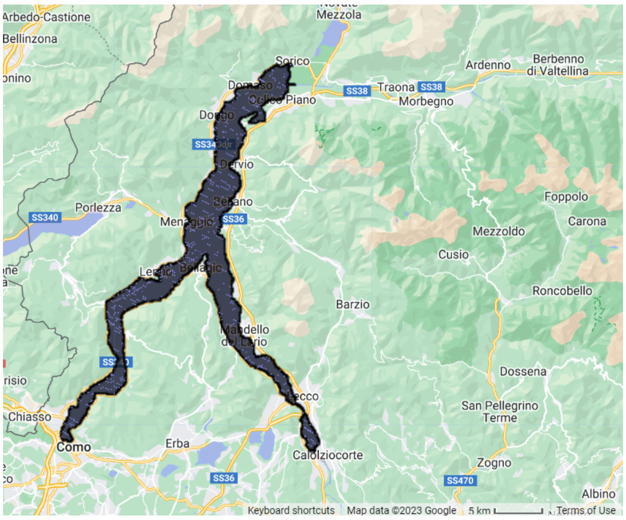

Figure 2. Lake Como extracted boundary from JRC Global Surface Water Mapping Layers and the GEDI footprints within the lake

To enable a proper comparison between the GEDI elevation data and the gauge station measurements, we transformed the GEDI elevation data from ellipsoidal heights referred to WGS84 datum to orthometric heights according to the EGM2008 geoid model. This transformation was necessary to ensure that both the GEDI and gauge station data were referenced to the same elevation model.

3.2 Outlier removal

The proposed outlier detection procedure consists of two different steps for every GEDI passage over each considered water surface. The resulting dataset is more reliable and accurate, which is crucial for an effective inland water resource monitoring.

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLVIII-M-1-2023-131-2023

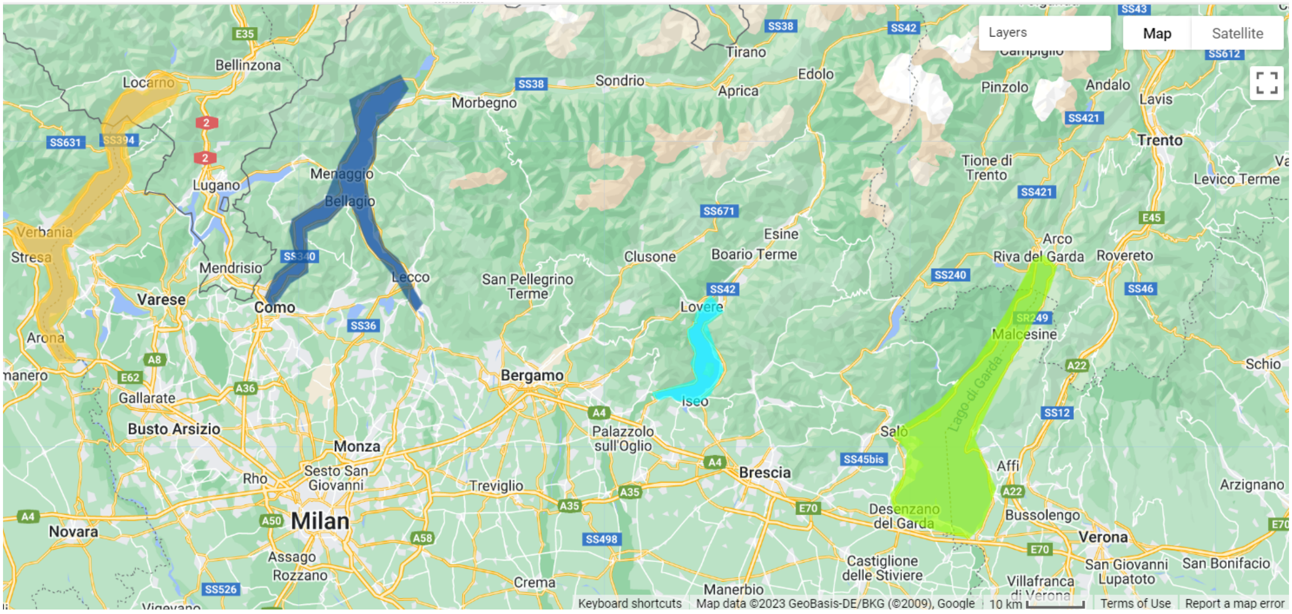

Figure 1. The location of four selected lakes (lake Maggiore highlighted as yellow, Como as dark blue, Iseo as light blue, and Garda as green) on the Google Earth Engine platform.

3.2.1 Quality and degrade flags The first step is based on two flags supplied together within GEDI data. Specifically, the "quality_flag" indicates if the considered footprint has a valid or invalid waveform, due to anomalies in the energy, sensitivity, and amplitude of the signal. The L2A quality_flag uses a conservative sensitivity threshold of 0.9 over land (0.5 over the ocean), but under some conditions (e.g. dense forest) the user may benefit from selecting a higher threshold. In this research, we have only used the footprints with quality_flags equal to one, indicating the laser shot meets all the criteria based on energy, sensitivity, amplitude, real-time surface tracking quality, and difference to Digital Elevation Model (DEM).

The "degrade_flag" indicates the degraded state of pointing (the saturation intensity of the returned photons might reduce the accuracy of the measurement) and/or positioning information (GPS data gap, GPS receiver clock drift) of the GEDI signal. In this study, only footprints with degrade_flags equal to 00 were selected.

3.2.2 3NMAD-test The second step of the outlier removal procedure relies on the robust version of the standard 3-sigma test, here implemented considering the NMAD (Normalized Median Absolute Deviation (Equation 1)):

$$NMAD = Median(H_i - Median(H)) \times 1.4826 \qquad (1)$$

where the elevation of each GEDI footprint was denoted as H_i, while H referred to the median water elevation of all the footprints within the considered lake and epoch.

This step was carried out on each lake separately to remove the footprints that had sufficient technical quality (with no anomalies in energy in the returning beam, or GPS drifts), but inconsistent values with the rest of the dataset.

In this level of outlier removal, every GEDI water level measurement (footprint) that is not within the threshold of $\mp$ 3*NMAD from the median of the water level throughout the lake is removed as an outlier.

4. RESULTS

The accuracy of the GEDI elevation data was evaluated by comparing them with gauge measurements, in order to determine GEDI reliability for monitoring lake water levels. The comparison is ongoing over the period from the activation of GEDI in April 2019 until June 2022, for about 3 years.

The quality flag and degraded flag were able to detect only 40% of outliers. The 3NMAD iterative test removed a significant portion of the data, ranging from 80% to 87%, improving their accuracy (Table 1). The GEDI data was quite extensive and rich in numbers of footprints, which resulted in over 4600 footprints for Lake Iseo, despite its relatively small area and after removing 87% of the original data. Although a significant amount of GEDI data was removed, the remaining GEDI data still provided valuable insights into water level monitoring for Lake Iseo and the three other lakes.

To evaluate the accuracy of GEDI data after the outlier removal, the standard deviation (SD), mean, median, NMAD, and root mean squared error (RMSE) were calculated based on the median of each GEDI epoch and the most contemporary gauge measurement (Figure 3). The comparison between the GEDI elevation data and the gauge measurements resulted in a mean standard deviation of 0.36 m and an average NMAD of 0.38 m. Additionally, the RMSE has a value of 0.44 cm. These results indicate that the GEDI data, after the outlier removal, can be used for lake elevation monitoring applications. Figure 3 shows the difference between the median value of each epoch and the water level measured from gauges.

The standard deviation values range from 0.26 to 0.44 m, indicating that the elevation measurements are relatively consistent across footprints as well as agreeing with the NMAD values

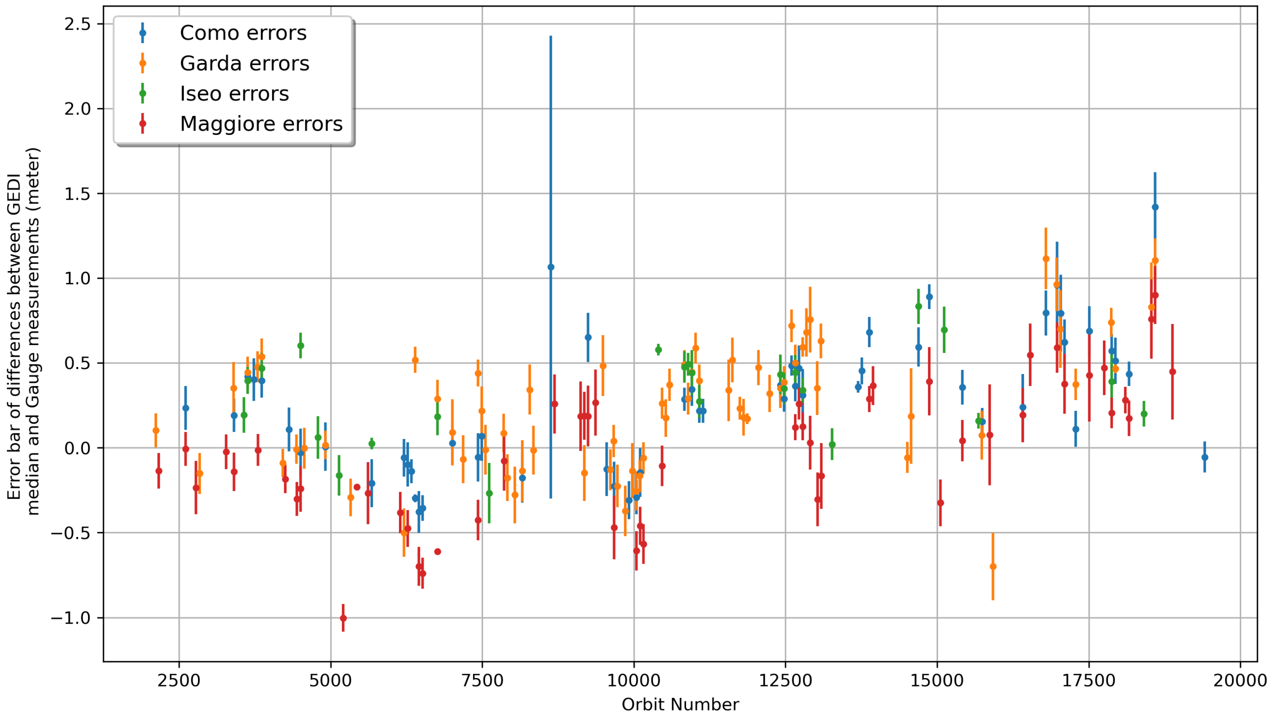

Figure 3. Water level differences between gauge measurements and medians of GEDI measurements after outliers removal over the period 2019/06 – 2022/06 for 4 lakes of Northern Italy (Como, Garda, Maggiore, and Iseo), based on the ISS orbit number; the error bars show the ∓ 3*NMAD intervals centered in the respective difference values

Lake	Como	Garda	Iseo	Maggiore
SD (m)	0.39	0.36	0.26	0.44
Mean (m)	0.27	0.24	0.31	0.04
Median (m)	0.28	0.29	0.37	0.04
NMAD (m)	0.43	0.37	0.25	0.49
RMSE (m)	0.47	0.43	0.41	0.45
Count of footprints before flags (-)	71686	210202	33692	96430
Count of footprint after flags (-)	14672	49326	7793	30195
Count of footprint after 3NMAD (-)	9565	34268	4673	19366
Removal percentage (%)	87	84	87	80

Table 1. Overall statistics for the water level differences between gauge measurements and medians of GEDI measurements (standard deviation (SD), mean, median, and root mean squared error (RMSE) for each lake.) after outliers removal over the whole period of 2019/06 – 2022/06 for 3 lakes of Northern Italy (Como, Garda, and Iseo)

ranging from 0.25 to 0.49 m. The RMSE showed values less than 0.47 m, indicating that the GEDI data correlates well with the gauge station measurements and Figure 4 shows the same pattern and ability of GEDI in following the patterns of the lakes.

The analysis of the GEDI data showed that the error or NMAD was generally below 0.20 m. However, the precision decreases and the NMAD values started to increase and scatter in the last orbits, particularly after orbit number 12500 (Figure 5), for unknown reasons. Also, the same pattern occurs in the differences between gauge measurements and the GEDI median of epoch/orbit right around the same period (Figure 3). This fact indicates the need for further research to understand better and address these limitations.

5. CONCLUSIONS

In conclusion, this study demonstrated the potential of GEDI L2A data for lake water level monitoring, despite the challenges of outlier removal and the need for proper data calibration. To assess the implemented outlier detection procedure and to preliminarily evaluate the accuracy of the GEDI data, we compared the water levels inferred from the median of GEDI measurements after outliers removal with the (as much as possible) contemporary water levels measured by the hydrometric stations at four major lakes (Como, Garda, Iseo, Maggiore) in Northern Italy.

The results showed that GEDI L2A footprints, after the removal of outliers, can provide accurate and reliable measurements of lake surface elevation, in comparison to traditional in-situ gauge measurements. The low values of NMAD (lower than 0.20 m in most cases) showed a high precision in epochs/orbits and the potential for being calibrated to obtain better accuracy.

The use of global EO technologies like GEDI can provide regular and frequent monitoring of water reservoirs worldwide, making it a cost-effective and efficient way to monitor these crucial natural resources. However, further research is needed to assess the accuracy of GEDI L2A data in different types of water bodies and environmental conditions.

REFERENCES

Bocchino, F., Ravanelli, R., Belloni, V., Mazzucchelli, P., Crespi, M., 2023. Water reservoirs monitoring through Google Earth Engine: application to Sentinel and Landsat imagery. *The*

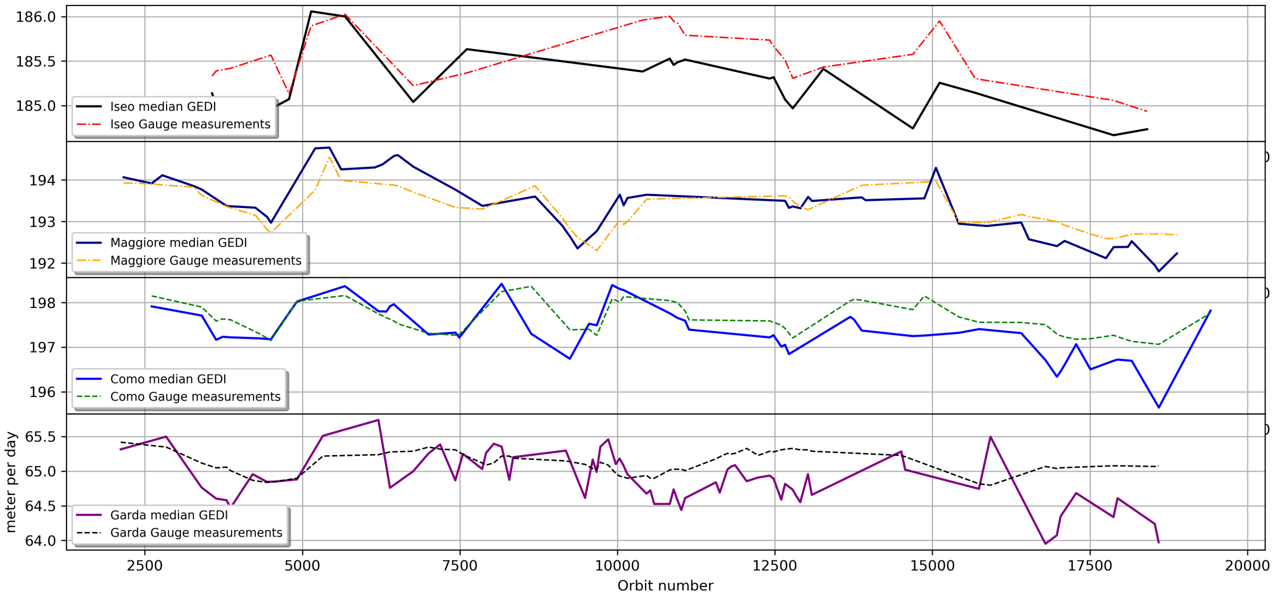

Figure 4. The Four lakes GEDI Median and the gauge measurement changes over the course of orbit number from the period 2019/06 – 2022/06

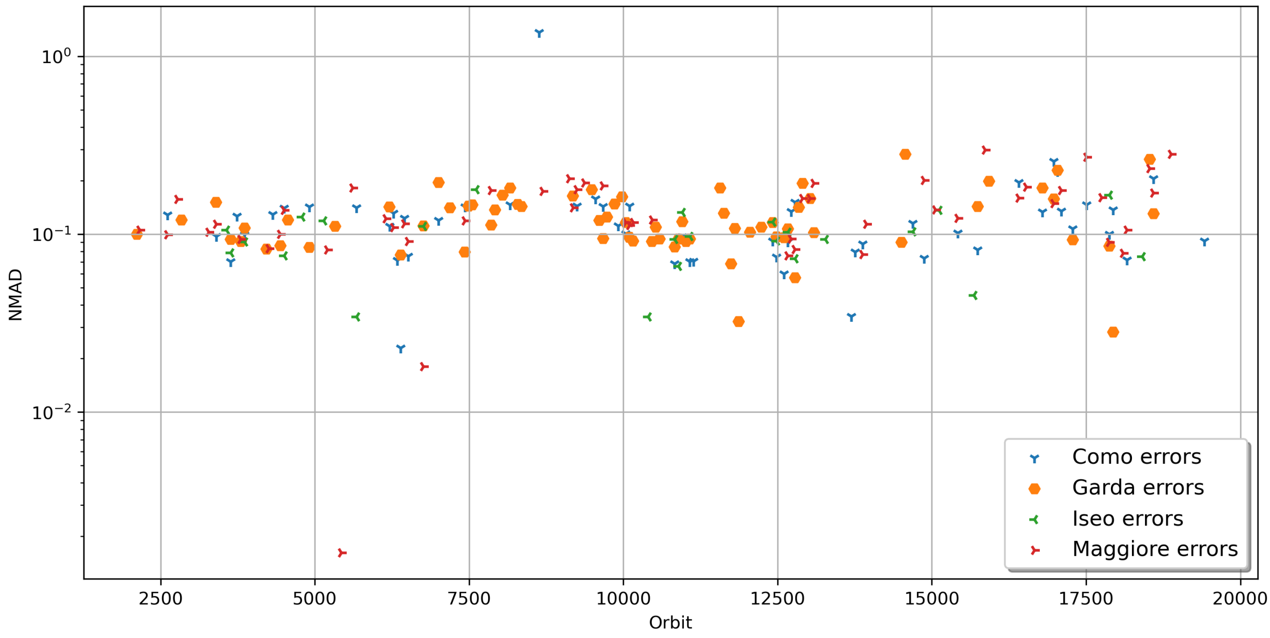

Figure 5. Scatter plot for NMAD (Error) of each epoch/orbit of each of the four lakes based on ISS orbits for the whole period June 2019 to June 2022

International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. In press.

Cardille, J. A., Clinton, N., Crowley, M. A., Saah, D., 2022. *Cloud-Based Remote Sensing with Google Earth Engine*. `www.eefabook.org`.

Dubayah, R., Luthcke, S., Sabaka, T., Nicholas, J., Preaux, S., Hofton, M., 2021. GEDI L3 gridded land surface metrics, version 1. *ORNL DAAC*.

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`.

Enti Regolatori dei Grandi Laghi, 2022. Home Page - Laghi. `www.laghi.net`.

Fayad, I., Baghdadi, N., Bailly, J.-S., Frappart, F., Pantaleoni Reluy, N., 2022. Correcting GEDI Water Level Estimates for Inland Waterbodies Using Machine Learning. *Remote Sensing*, 14(10), 2361.

Hamoudzadeh, A., Ravanelli, R., Crespi, M., 2023. GEDI data within Google Earth Engine: Potentials and analysis for inland surface water monitoring. EGU, Copernicus Meetings.

Kaplan, G., Avdan, U., 2017. Object-based water body extraction model using Sentinel-2 satellite imagery. *European Journal of Remote Sensing*, 50(1), 137–143.

Kavvada, A., Metternicht, G., Kerblat, F., Mudau, N., Haldorson, M., Laldaparsad, S., Friedl, L., Held, A., Chuvieco, E., 2020. Towards delivering on the sustainable development goals using earth observations. *Remote Sensing of Environment*, 247, 111930.

Lee, Y.-K., Hong, S.-H., Kim, S.-W., 2021. Monitoring of Water Level Change in a Dam from High-Resolution SAR Data. *Remote Sensing*, 13(18).

Liu, A., Cheng, X., Chen, Z., 2021. Performance evaluation of GEDI and ICESat-2 laser altimeter data for terrain and canopy height retrievals. *Remote Sensing of Environment*, 264, 112571.

Nascetti, A., Di Rita, M., Ravanelli, R., Amicuzi, M., Esposito, S., Crespi, M., 2017. Free global dsm assessment on large scale areas exploiting the potentialities of the innovative google earth engine platform. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-1/W1, 627–633.

Pekel, J.-F., Cottam, A., Gorelick, N., Belward, A. S., 2016. High-resolution mapping of global surface water and its long-term changes. *Nature*, 540(7633), 418–422.

Qi, W., Lee, S.-K., Hancock, S., Luthcke, S., Tang, H., Armston, J., Dubayah, R., 2019. Improved forest height estimation by fusion of simulated GEDI Lidar data and TanDEM-X InSAR data. *Remote Sensing of Environment*, 221, 621-634.

Ravanelli, R., Cirigliano, R., Di Rico, C., Monti, P., Crespi, M., 2018a. Monitoring Urban Heat Island through Google Earth Engine: potentialities and difficulties in different cities of the United States. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-3, 1467–1472.

Ravanelli, R., Nascetti, A., Cirigliano, R. V., Di Rico, C., Leuzzi, G., Monti, P., Crespi, M., 2018b. Monitoring the impact of land cover change on surface urban heat island through Google Earth Engine: Proposal of a global methodology, first applications and problems. *Remote Sensing*, 10(9), 1488.

Sinha, A., Sengupta, T., Alvarado, R., 2020. Interplay between technological innovation and environmental quality: formulating the SDG policies for next 11 economies. *Journal of Cleaner Production*, 242, 118549.

United Nations General Assembly, 2015. Transforming our world: The 2030 agenda for sustainable development.

University of Maryland, 2022. GEDI ecosystem lidar. `www.gedi.umd.edu`.