GIScience & Remote Sensing

ISSN: 1548-1603 (Print) 1943-7226 (Online) Journal homepage: www.tandfonline.com/journals/tgrs20

Large-scale monitoring of inland water surface levels with GEDI data: an operational cloud-based approach in Google Earth Engine

Alireza Hamoudzadeh, Roberta Ravanelli & Mattia Crespi

To cite this article: Alireza Hamoudzadeh, Roberta Ravanelli & Mattia Crespi (2025) Large-scale monitoring of inland water surface levels with GEDI data: an operational cloud-based approach in Google Earth Engine, GIScience & Remote Sensing, 62:1, 2483027, DOI: 10.1080/15481603.2025.2483027

To link to this article: https://doi.org/10.1080/15481603.2025.2483027

Published online: 14 Apr 2025.

GISCIENCE & REMOTE SENSING
2025, VOL. 62, NO. 1, 2483027
https://doi.org/10.1080/15481603.2025.2483027

Taylor & Francis
Taylor & Francis Group

Large-scale monitoring of inland water surface levels with GEDI data: an operational cloud-based approach in Google Earth Engine

Alireza Hamoudzadeh [a], Roberta Ravanelli [a,b] and Mattia Crespi [a,c]

[a]Geodesy and Geomatics Division, DICEA, Sapienza University of Rome, Rome, Italy; [b]Geomatics Unit, Department of Geography, Faculty of Sciences, University of Liége, Liége, Belgium; [c]Sapienza School for Advanced Studies, Sapienza University of Rome, Rome, Italy

ABSTRACT

This paper demonstrates the feasibility of large-scale, reliable monitoring of inland water surface levels through the analysis of the data collected by the Global Ecosystem Dynamics Investigation (GEDI) altimeter. In particular, we propose an automated and worldwide operational workflow, implemented within Google Earth Engine (GEE), benefiting from the availability of the whole GEDI time series in this geospatial cloud platform. Leveraging the massive computational capabilities of GEE, we were able to analyze millions of GEDI footprints and assess the potential of the sensor – in terms of precision and accuracy – to serve as an efficient and reliable remote hydrometer. The workflow is based on a rigorous spatio-temporal outlier rejection procedure and a spatial aggregation of the remaining high-quality footprints, for the robust estimation of a per-epoch median water level and its precision for the considered lake surface. A comprehensive precision and accuracy assessment was performed by comparing the GEDI retrieved water-level time series with in situ gauge data for 11 lakes of variable extent (from tens to several thousand km^2) located in three continents. A rather homogeneous precision of GEDI water levels across the lakes was found, with a mean value of 14 cm. Additionally, a good agreement with the reference gauge stations was observed, showing an overall accuracy of 35 cm, a slight overestimation bias (6 cm), and a correlation of 0.76. It is important to note that these results are affected by the uncertainties of the transformation among GEDI reference frame and gauge stations reference frames. The proposed workflow can be easily applied to provide reliable inland water-level time series for all those lakes for which GEDI data is available, offering a generally higher temporal resolution than other altimeters. This approach lays the foundations for integrating GEDI within the set of remote-sensing instruments for water cycle monitoring on a large scale, enhancing our understanding of water storage dynamics in lakes, particularly in remote areas where it is not possible to install and maintain hydrometric gauges.

ARTICLE HISTORY
Received 4 March 2024
Accepted 18 March 2025

KEYWORDS
GEDI; inland surface water level monitoring; precision and accuracy assessment; google earth engine workflow; geo-big data

1. Introduction

Surface freshwater plays a major role in all aspects of life on our planet in the form of lakes, rivers, reservoirs, wetlands, snow, and glaciers (Kseňak et al. 2022; Sekertekin 2021). The availability of this precious resource affects ecosystems, hydrology, and climate and provides human benefits such as urban, agricultural, and water supply for industries, support for wildlife and fisheries, and recreational opportunities like outdoor leisure activities (Allan et al. 2017; Chang, Imen, and Vannah 2015; Crétaux et al. 2016). Due to climate change, inland water resources are being significantly threatened by rising temperatures, changing precipitation patterns, and melting glaciers. All these phenomena influence the availability, quality, and distribution of freshwater around the world (Frappart et al. 2021; Lee, Hong, and Kim 2021). The use of satellite observations on climate and hydrologic models has recently revealed that, from 1992 to 2020, 53% of both large natural lakes and reservoirs were affected by a decline in volume (Yao et al. 2023). Inland water-level monitoring is thus nowadays even more important in supporting decision-making processes and enabling sustainable and effective management of water resources, as demonstrated also by at least three (6,13,14) Sustainable Development Goals (SDGs) (United Nations General Assembly 2015).

Surface water levels are commonly measured at specific sites using tools known as gauge stations. The measurements obtained from these stations generally support the estimation of changes in the reservoir under investigation from the volume-area-

CONTACT Alireza Hamoudzadeh alireza.hamoudzadeh@uniroma1.it

elevation curves derived from topographic and bathymetric information of the reservoir itself (Li et al. 2023; Valadão et al. 2021). Nonetheless, a global-scale deployment of gauge stations is unfeasible because of their high cost of installation and maintenance, and the remote location of many water reservoirs. Moreover, a considerable amount of in-situ data is gathered but not distributed due to governmental restrictions (Lawford et al. 2013). Therefore, it is crucial to develop effective methodologies based on up-to-date technologies to homogeneously monitor the water level of inland freshwater reservoirs on a global scale.

Nowadays, thanks to the vast developments in remote sensing technologies, Earth Observation can represent a feasible and low-cost answer to the need for extensive and long-term monitoring of surface water levels (Cooley, Ryan, and Smith 2021; Pi et al. 2022). In particular, over the past years, RADAR (Radio Detection And Ranging) and LiDAR (Light Detection And Ranging) altimeters, initially developed for the monitoring of sea, coast, and ocean surface topography, have been employed to evaluate surface water levels of lakes, rivers, and wetlands (Johannessen and Bjorgo 2000; Wdowinski, Liao, and Zhang 2021; Zhang et al. 2006). Overall, 18 satellites equipped with different RADAR and LiDAR altimeters were launched to monitor continental and ocean water levels. Seven of these satellites are currently operational and have the technical specifications required for inland water monitoring, two of which being LiDARs (Fayad, Baghdadi, Bailly, et al. 2022; Yang et al. 2022). Their main features and technical details are summarized in Table 1.

Of the currently active satellite altimetry missions, only the two LiDAR altimeters natively provide the spatial and temporal resolutions necessary to monitor water levels in inland water bodies of varying sizes. The Geoscience Laser Altimeter System (GLAS) was the first Earth Observation space-borne LiDAR instrument. It was hosted onboard the Ice, Cloud, and land Elevation Satellite (ICESat) and was successfully employed for estimating ice sheet elevations, height profiles of clouds and aerosols, land elevations and vegetation cover, and approximate sea ice thickness (Liu et al. 2022; Zhu et al. 2022). Yet, ICESat was discontinued in 2009, but the National Aeronautics and Space Administration (NASA) renewed the mission in 2018 under the name of ICESat-2 with a new satellite, employing six laser beams instead of the unique beam of the previous version. ICESat-2 has a fixed orbit cycle of 91 days so the effective revisit time on a certain lake is dependent on its size, being obviously the number of crossing orbits higher on wider lakes. While ICESat-2 was launched for monitoring ice surfaces and clouds, the Global Ecosystem Dynamics Investigation (GEDI) LiDAR instrument, hosted onboard the International Space Station (ISS), started to collect data since April 2019 with the aim of gathering precise measurements of forest canopy height, canopy vertical structure, and surface elevation in general. Since GEDI is mounted on the ISS, it does not have a fixed temporal resolution; however, each ISS orbit lasts about 90–93 min depending on the platform altitude, which is adjusted from time to time, making the ISS repeat cycle approximately 3 days. GEDI repeat cycle is thus in the order of days, but the footprint size of 25 m does not guarantee that a given location will have overlapping samples in consecutive epochs (Peña-Arancibia et al. 2024). GEDI coverage is limited to the zone with latitude within −51.6° and 51.6°, as bounded by the ISS orbital inclination.

Moreover, the recent advances in RADAR altimetry technology and the development of dedicated refined

Table 1. Features of the currently active LiDAR and RADAR satellite altimeters.

Mission	Footprint Size (m)	Scanning Pattern (m)	Temporal Resolution (days)	Sensor Type
ICESat-2	13	90 (across-track spacing)	91 (repeat cycle)	LiDAR
GEDI	25	60 (along-track spacing)	variable (ISS repeat cycle ≈ 3 days)	LiDAR
Poseidon-3B	~ 2–10 km	30 km (along-track resolution)	~ 10	RADAR
Sentinel-3A	~300	1.67 km (across-track resolution)	27	RADAR
HY-2	16 km	6–7 km (along-track resolution)	14	RADAR
SARAL	~1400	170 (along-track resolution)	35	RADAR
SWOT	~100 (over land)	120 km (along-track swath)	21	RADAR

ICESat-2 across-track spacing for pair beams is 90 m while the two-beam spacing is close to 3 km.

processing techniques make it possible to monitor lakes with such sensors, as well (Frappart et al. 2021). This is the case of Surface Water Ocean Topography (SWOT) and Sentinel-3 missions. Launched in December 2022 (Hamoudzadeh, Ravanelli, and Crespi 2024), SWOT is a RADAR altimeter aiming to provide high-resolution observations of water surface elevations for both inland water bodies and oceans (Riggs et al. 2023). In the case of Sentinel-3, the development of innovative retracking algorithms has lately allowed the production of global-scale coastal zone and inland water products within the framework of the ESA-funded HYDROCOASTAL project (Sentinel Online 2023). Satellite altimetry can hence serve as a valuable complement to in-situ measurements of inland water surface levels, offering extensive spatial coverage and consistent monitoring capabilities.

In this context, GEDI can play an important role as an additional instrument for routine water cycle monitoring on a large scale, enhancing our understanding of water storage dynamics in lakes, especially for remote areas where it is not possible to install and maintain hydrometric gauges (Hamoudzadeh, Ravanelli and Crespi 2023; Bocchino et al. 2023). However, it is essential to understand the precision, accuracy, and limits of this sensor for its possible integration within the set of remote-sensing instruments already available for this purpose. Previous studies on the GEDI capability to monitor lake levels have been limited to the analysis of specific lakes, such as the great lakes of North America (Fayad, Baghdadi, and Frappart 2022; Xiang et al. 2021), the Lake Qinghai in China (Xiang et al. 2021; Z. Zhang et al. 2021, 2022), the lakes of the Tibetan Plateau (Wu et al. 2023), or the Switzerland lake clusters (Fayad, Baghdadi, and Frappart 2022; Fayad et al., 2020). With the aim of performing a regional-level assessment of GEDI for limited time intervals – typically from April 2019 to mid-2020 (Fayad, Baghdadi, and Frappart 2022; Z. Zhang et al. 2021) to few months (Fayad et al. 2020; Xiang et al. 2021) – these studies addressed the comparison of GEDI data with ground reference measurements from hydrometers, highlighting accuracies, in terms of RMSE, at the level of one to a few decimeters (Fayad, Baghdadi, and Frappart 2022; Fayad et al., 2020; Wu et al. 2023; Xiang et al. 2021; Z. Zhang et al. 2021, 2022). These accuracies were achieved using different approaches for spatial outlier detection and removal, such as the 3*RMSE and 1.5*RMSE criterion and quality flag-based filtering (Xiang et al. 2021; Zhang et al. 2021, 2022), as well as Random Forest regression models (Fayad et al. 2020; Fayad, Baghdadi, and Frappart 2022). The dependence on location and time (Fayad et al. 2020; Xiang et al. 2021), along with the presence of both positive and negative biases – potentially due to different height systems or sensor-specific factors – was also reported (Fayad et al. 2020; Xiang et al. 2021; Zhang et al. 2021).

Considering the available literature and the previously proposed approaches, it is thus possible to identify some limitations, both related to the length of the considered GEDI time series and to the approach adopted to analyze GEDI data. In fact, none of the previous investigations analyzed the entire GEDI data time series available at the time of the analysis for the considered lakes – likely due to the complexity of the management of millions of footprints – except when the available time series were short (at the beginning of the operational life of GEDI). In our study, the possibility to analyze the entire GEDI time series revealed the presence of abrupt elevation jumps, as high as a few meters, occurring in temporally close epochs within the water-level time series, despite the apparently good quality of GEDI data (low spatial variability) during these epochs. Also, this possibility led to the investigation of the precision of GEDI measurements and its possible variability over time. The detection and removal of the mentioned jumps temporally and the assessment of the precision of GEDI measurements were not explicitly addressed in previous studies, and represent the novelties of our approach for the analysis of GEDI data.

Therefore, the aim of our work was twofold: first, methodological, to improve the approach for the optimal retrieval of water-level time series from GEDI data analysis; second, operational, to implement a fully automated workflow, based on this improved approach, making it worldwide operational through Google Earth Engine (GEE), leveraging the availability of the complete GEDI dataset (2019–2022) on this platform. Briefly, the methodological novelties of our approach concern with:

(I) The definition of a comprehensive strategy for outlier detection and removal that relies solely on GEDI data, applicable in near real-time as soon as the data become available, the implementation of a rigorous spatial and temporal analysis

of GEDI water levels, and the use of robust statistical indexes, to prevent the effect of residual outliers in GEDI footprints, given their high percentage in the original GEDI data. The spatio-temporal outlier detection and removal strategy is indeed essential for improving the reliability of the GEDI-derived water-level time series. Indeed, while the spatial precision of individual epochs may be high, the corresponding water levels can be unrealistic when compared to the closest temporal values in the time series.

(II) The definition of a strategy to assess the precision of GEDI measurements within each GEDI passage and then over the entire GEDI time series, based on the outcome of the outlier detection and removal strategy. The results of this precision analysis can reveal potential performance variations of GEDI over time and space, serving as the basis for the investigations on accuracy and correlation analyses of GEDI-derived water levels with independent references (like gauge stations data).

These methodological novelties are implemented within a GEE workflow, which allows for comprehensive monitoring of inland water levels on a large scale and across the entire GEDI time series without the need for significant modifications, in the pure sense of geo-big data analysis. Moreover, the availability of a procedure implemented within a geospatial analysis cloud platform like GEE enables its efficient and uniform application to any area within the region covered by the GEDI data.

Specifically, for each considered lake, the workflow consists of several steps, as described in the rest of the paper: (I) water footprint detection (Section 3.1); (II) spatial outlier detection and removal for each observation epoch (Section 3.2); (III) spatial aggregation of the remaining high-quality footprints to estimate a per-epoch water surface median level of the considered lake (Section 3.2.4); (IV) temporal outlier detection and removal across the whole time series for the considered lake (Section 3.3); (V) precision assessment of GEDI observations for each epoch, accuracy and linear correlation assessment of GEDI observations for each epoch with respect to available independent references (gauge data) (Sections 4.2, 4.3, and 4.4); (VI) analysis of the results to assess the global potential of GEDI across the time series of all the considered lakes (Section 5).

2. Study area and data

2.1. The global ecosystem dynamics investigation (GEDI)

The GEDI instrument is a geodetic-class, LiDAR laser system mounted on the ISS, initially designed to significantly enhance the quantification and understanding of the Earth's carbon cycle and biodiversity (Dubayah et al. 2021; Duncanson et al. 2022; Kutchartt, Pedron, and Pirotti 2022).

It consists of three lasers producing eight parallel observation tracks. Each laser emits short pulses of light (14 ns long) down toward the Earth's surface 242 times per second with a beam divergence of 56 mrad, resulting in footprints averaging 25 m in diameter. Each footprint is separated by a 60 m along-track distance, with an across-track space of about 600 m between each of the eight tracks. GEDI is characterized by a high annual observation frequency, depending on the ROI size, and variable due to the ISS orbit height variation. The measurements are collected over the Earth's surface nominally between the latitudes of 51.6° and −51.6° (Liang et al. 2023), and the determined heights of the Earth's surface are ellipsoidal ones referred to the WGS84 ellipsoid. GEDI products are corrected for solid earth tides, ocean loading, solid earth pole tide, and ocean pole tide (GEDI - The University of Maryland (2023a); United States Geological Survey 2023; Potapov et al. 2021). The instrument can be rotated about the instrument X-axis up to 6°, the lasers can be pointed up to 40 km on either side of the ISS ground track. The spatial coverage of GEDI changed over time: the ISS raised its orbit by 16 km (Dubayah et al. 2022), resulting in GEDI experiencing "orbital resonance", i.e. the sensor repeatedly covers the same tracks while leaving significant gaps in between (Di Tommaso et al. 2023).

Only recently GEDI potentialities have been investigated for monitoring the levels of inland surface waters, but, at the time of writing, no comprehensive investigations have been carried out benefiting from the availability of GEDI data within the Google Earth Engine (GEE) archive (Earth Engine Data Catalog — Google for Developers 2023a).

In this work, the GEDI L2A Raster Canopy Top Height (version V2) product (Earth Engine Data Catalog — Google for Developers 2023b) was used to implement the proposed water-level monitoring workflow within the GEE platform (Bullock et al. 2023; Healey et al. 2020). In particular, this product is a raster version of

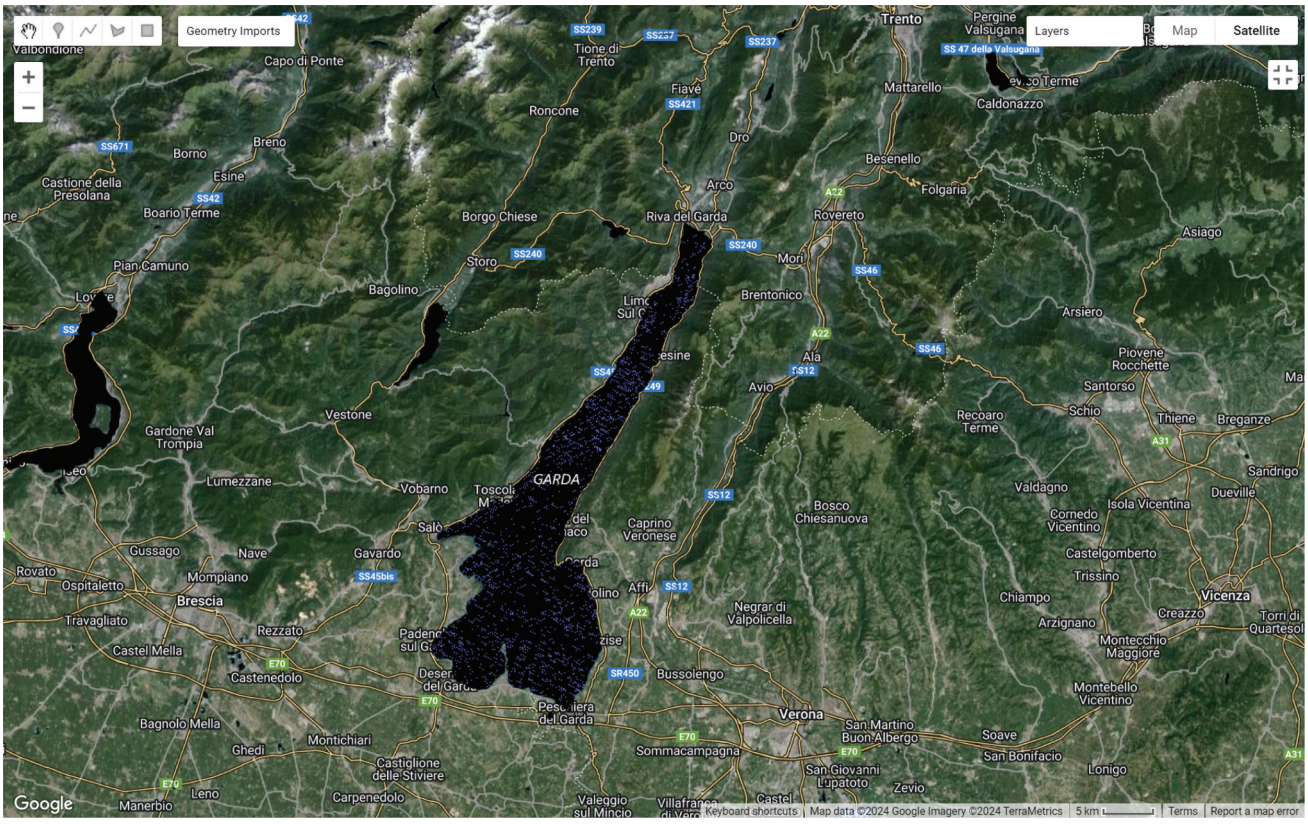


Figure 1. Overview of GEDI footprints (blue points on the water body) within Lake Garda after detection of water bounds in GEE alongside JRC global surface water (black bodies). The north direction is in the direction of the vertical axis of the map.

the original GEDI's Level 2A Geolocated Elevation and Height Metrics Product (GEDI02_A) and consists of 140 different bands (Figure 1).

Of the 140 available bands, 100 are Relative Heights (RH), i.e. bands commonly utilized for canopy height analysis, as they represent the height at which a certain percentile returned energy is reached relative to the ground; in other words, RH collectively describes the waveform collected by GEDI (Earth Engine Data Catalog — Google for Developers 2023b).

The remaining bands include information such as orbit number, elevation, beam, solar angles, laser pointing angles, and flags for specific uses (leafoff, quality – see Sections 3.2.1 and 3.2.2 –, etc.). The orbit_number band contains the count of the number of orbits that the ISS has completed around the Earth; this band was used in this work as a timescale.

The elev_lowestmode band contains the elevation of the center of the lowest mode (peak) of the footprint waveform with respect to the WGS84 ellipsoid. This band is generally considered representative of the ground elevation (Lang et al. 2022) of the footprints: we used this band to represent the water level.

In fact, waveforms recorded within footprints over complex geometries (e.g. forests) are characterized by a multi-modal form, with each mode peak representing a reflection from a distinct surface elevation. Conversely, over flat surfaces, such as water surfaces, or bare ground, the waveforms recorded by GEDI are characterized by a single peak, i.e. they have an uni-modal form (Frappart et al. 2021).

2.2. Study areas and water level reference data

To comprehensively assess the developed workflow and the GEDI retrieved water-level time series, we selected 11 lakes from three different continents (Table 2, Figure 2) with different features (e.g. area, position, etc.), analyzing the whole time span of GEDI data (from April 2022 to June 2019) available within GEE at the time of the analysis.

To evaluate the accuracy, we computed the differences and the linear correlation coefficient between

Table 2. Set of 11 considered lakes with details about their location (center point coordinates (WGS84)), extent, and the corresponding gauge stations (gauge code and names).

Name	Center	Area (Km^2)	Gauge Code	Gauge Name
Argentino (Argentina)	50°13′S 72°25′W	1415	2817	El Calafate
Como (Italy)	46°00′N 9°16′E	146	1	Lago di Como
Erie (US, Canada)	42°15′N 81°15′W	25744	9063038	Lake Erie
Garda (Italy)	45°38′N 10°40′E	369	4	Lago di Garda
Huron (US, Canada)	44°53′N 82°27′W	59600	9075065	Alpena MI, Huron
Iseo (Italy)	45°43′N 10°05′E	65	2	Lago d'Iseo
Maggiore (Italy)	46°05′N 08°42′E	212	3	Lago Maggiore
Mar Chiquita (Argentina)	30°30′S 62°40′W	$\approx$ 2000	4309	Laguna Mar Chiquita
Michigan (US)	44°00′N 87°00′W	57760	9087072	Sturgeon Bay Canal WI
Ontario (US, Canada)	43°51′N 77°57′W	19529	9052058	Rochester NY
Superior (US, Canada)	47°51′N 87°30′W	82000	9099018	Marquette C.G. MI

the daily water levels measured by in-situ gauge stations and the (as much as possible) contemporary water levels derived from GEDI data employing the developed workflow (Section 3). It has to be underlined that it was necessary to apply in advance transformations (Section 3.4) to the chosen unique height reference frame WGS84-EGM2008 (WGS84 ellipsoid with Earth Gravitational Model 2008 geoid model), since the height reference frames to which gauge stations are referred are generally different from the GEDI height system (ellipsoidal heights w.r.t. WGS84 ellipsoid). However, even if these transformations were computed at the best available knowledge, residual negative or positive biases can remain, as was already evidenced in previous investigations (Fayad, Baghdadi, and Frappart 2022; Fayad et al., 2020; Xiang et al. 2021; Zhang et al. 2021, 2022). Therefore, for each considered lake, we can actually evaluate the accuracy of GEDI data after a global bias removal (over the whole time series), and the agreement between the water-level time series as derived by GEDI data and by the gauge stations in terms of linear correlation coefficient.

For the lakes of Northern Italy, the gauge measurement data was gathered by the hydrometric stations of "Enti regolatori dei grandi laghi" (Enti Regolatori dei Grandi Laghi 2022; Hamoudzadeh, Ravanelli, and Crespi 2023). The in-situ water-level measurements of the Great Lakes of North America were gathered from "National Oceanic and Atmospheric Administration" (National Oceanic and Atmospheric Administration 2022). The Lago Argentino and Mar Chiquita lakes hydrometric data was collected from "Sistema Nacional de Información Hídrica" (Sistema Nacional de Información Hídrica 2022). Among the in-situ data used as a reference, exclusively the "Great Lakes Water Level Stations" include an assessment of their quality combined with a measurement algorithm characterized by an estimated accuracy of 6 mm.

2.3. Reference data for water boundaries

To extract the extent of water bodies, in order to be able to select automatically the GEDI data collected on the lake surfaces, the *JRC Global Surface Water Mapping Layers, v1.4* (Pekel et al. 2016) was employed. Already available within the GEE archive (Earth Engine Data Catalog — Google for Developers 2023c), this dataset contains maps of the location and temporal distribution of surface water from 1984 to 2021 and provides statistics on the extent and transformations of these water surfaces.

This dataset was developed using roughly 5 million images from Landsat 5, 7, and 8 collected from 1984 to the end of 2021 (Pekel et al. 2016). Each pixel has a ground sampling distance (GSD) of 30 m and was separately classified into water/non-water classes using an expert system. The results were compiled into a monthly historical record covering the entire duration.

3. Methodology

A preliminary investigation on GEDI data clearly evidenced the presence of significant outliers in both quantity and magnitude (Li et al. 2024).

For this reason, the developed workflow (Figure 3) relies on a rigorous spatio-temporal outlier rejection procedure and on a spatial aggregation of the remaining high-quality footprints to estimate the per-epoch water surface median level of the considered lake. The workflow was implemented within the GEE environment,

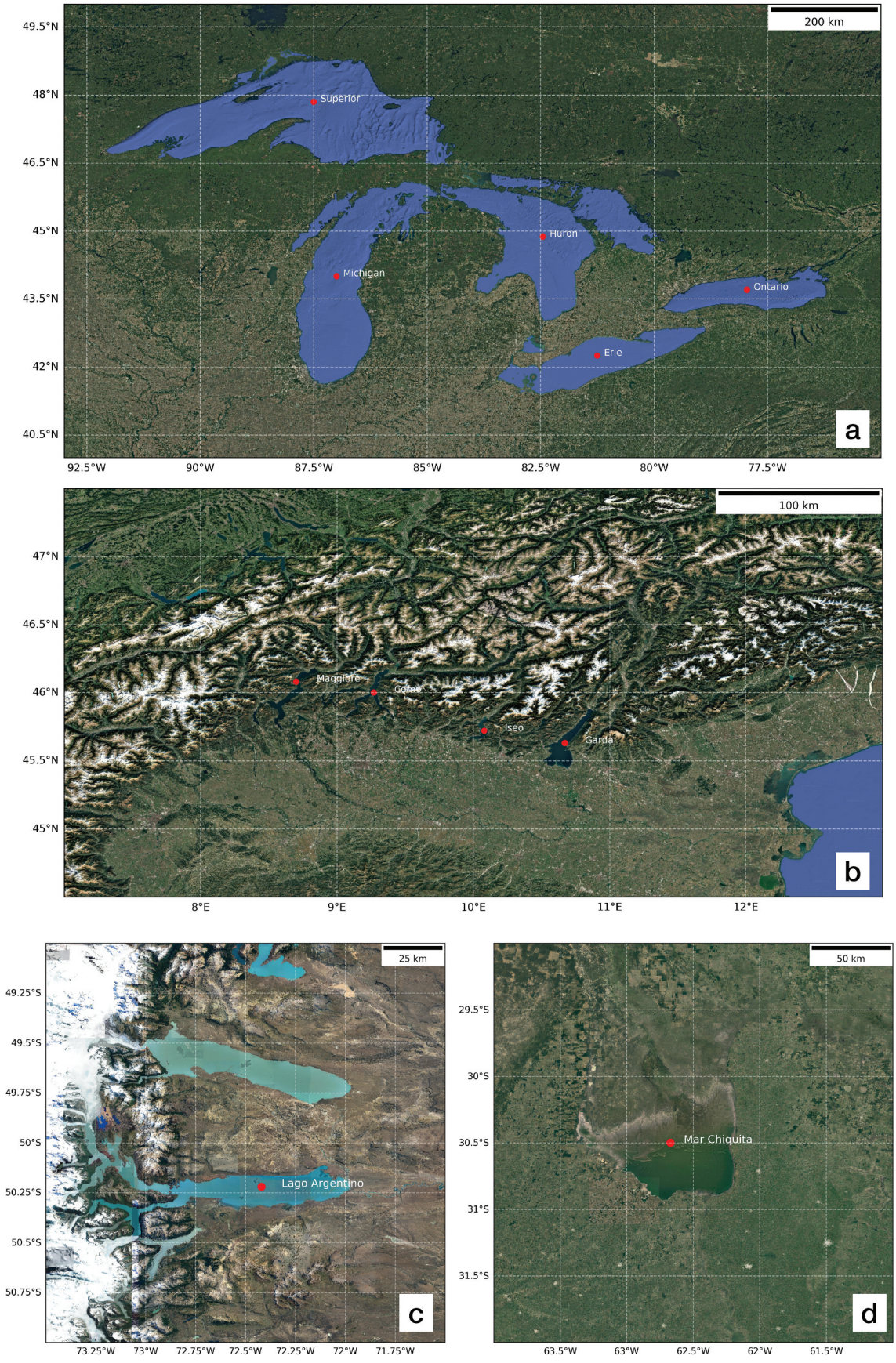


Figure 2. The lakes used as study areas – a) lakes in North America, b) lakes in North Italy, c) Lake Argentino, and d) Lake Mar Chiquita.

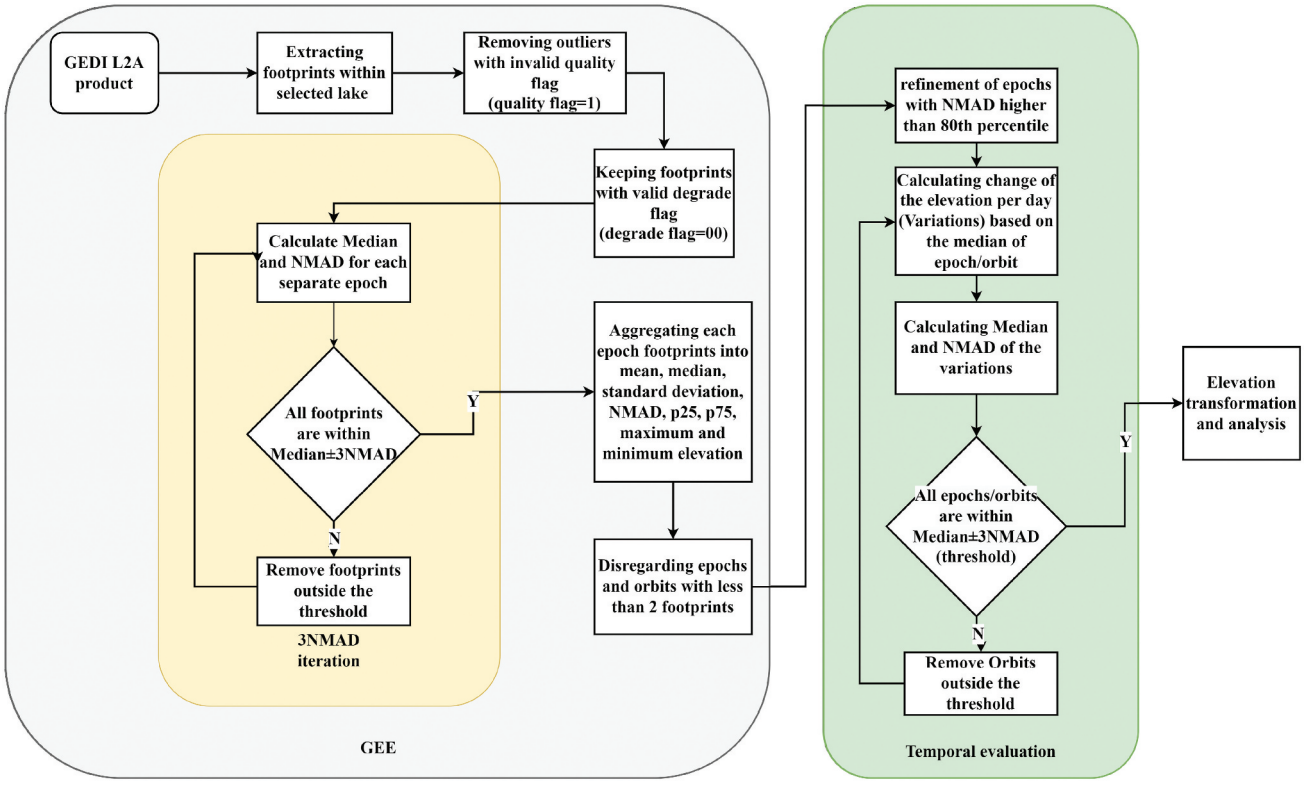


Figure 3. Flowchart of the developed workflow.

taking into account some specific concerns related to the massive amount of data to be processed.

3.1. Detection of water footprints

The assessment of the GEDI retrieved water-level time series was carried out only considering as input the footprints located well within the lakes considered in the analysis (Table 2). In particular, for all the lakes, we considered a 50-m buffer inside the lake boundary retrieved from the JRC Global Surface Water Mapping Layers v1.4 (Figure 4, Section 2.3). The buffer made it possible to avoid selecting footprints with mixed values of land and water in the further steps of the workflow, to limit the impact of the aforementioned geolocation errors of GEDI footprints (Tang et al. 2023), as well as the effects of dynamic changes in water boundaries over time.

3.2. Spatial outliers detection and removal

Not all the GEDI measurements are valid due to meteorological conditions (clouds can alter the waveform), GPS gaps, clock drifts, low view angle (Di Tommaso et al. 2023), unusual phenomena in measurement conditions (Mahmoodi and Ghassemi 2018), and platform problems. For these reasons, we designed and implemented four different levels of spatial outlier detection and removal.

The first two levels are based on specific GEDI features, the third level is based on an exhaustive iterative procedure and the fourth one consists of the spatial aggregation of inlier elevation data through the computation of standard statistics. The overall workflow of the spatial outlier detection and removal is summarized hereafter in the Procedure 1.

The core of the procedure is based on the 3-times Normalized Median Absolute Deviation (3NMAD) criterion (Section 3.2.3) and can be applied without the need for additional information besides the elev_lowestmode band.

Only orbits with a minimum of two footprints in the selected ROI were considered suitable for inclusion in further analyses. This requirement was considered to ensure that there was enough data per orbit to determine the water level of the selected lake and to develop further statistical analysis.

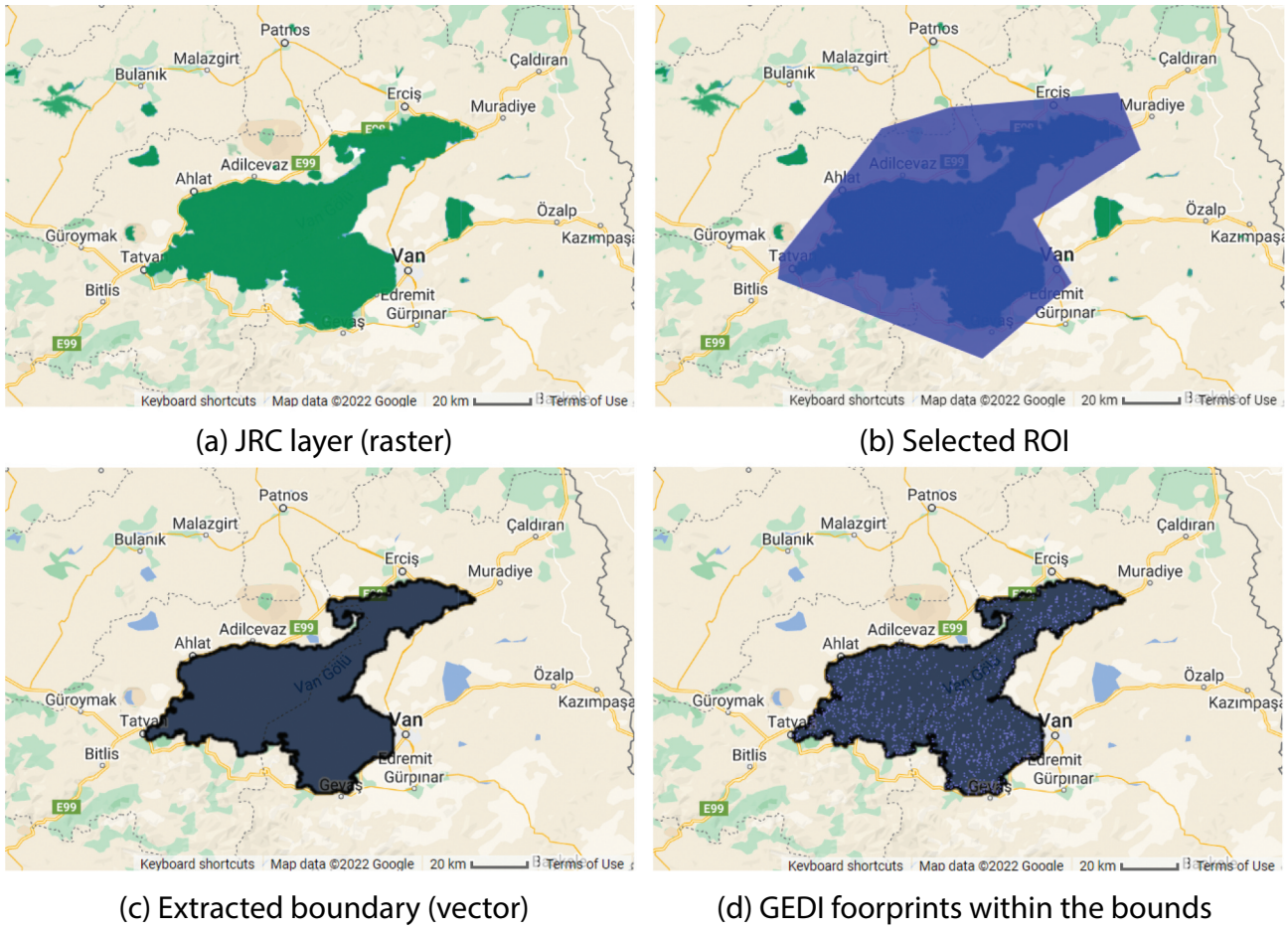


Figure 4. Extraction of GEDI footprints within lake bounds based on the user's drawing of the ROI around the desired lake and the automatic delineation of the water bound and GEDI data selection. The North direction is in the direction of the vertical axis of the maps.

The four steps of the spatial outlier removal procedure are detailed hereafter.

Procedure 1 Spatial outlier detection and removal procedure

```
procedure Spatial outlier detection and removal(E)   ▷ E representing an epoch in a lake
  for all E ← 1, Number of epochs for lake do
    Quality flag
    Degrade flag
    Spatial outlier removal
    if count(E) > 2 then
      Data aggregation and statistics computations
      Exporting from GEE
    end if
  end for
end procedure
```

3.2.1. Quality flag

The first level of spatial outlier detection and removal leverages the information contained in the quality_flag band (Section 2.1). This band specifies whether the assessed footprint holds a valid waveform or an invalid one, where the latter case is often due to anomalies in the energy, sensitivity, or amplitude of the returning signal (United States Geological Survey 2023). Only the footprints with a valid waveform were considered in the further steps of the analysis.

3.2.2. Degrade flag

The degrade_flag (Section 2.1) band indicates the degraded state of pointing and/or positioning information (GPS data gap, GPS receiver clock drift). The degrade_flag is a number between 0 and 99, and different values indicate different problems within the collected footprint. Only the footprints with a value of the degrade_flag equal to 0 (no detected problems) were considered in the further steps of the analysis.

3.2.3. 3NMAD exhaustive criterion

After the application of the first two levels of outlier detection and removal based on GEDI internal flags, the quality of GEDI measurements can still be insufficient. For each lake, the 3NMAD criterion is thus applied to each epoch for removing the footprints that have sufficient technical quality (no anomalies in energy in the returning beam, or GPS drifts), but inconsistent elevation values with respect to the other footprints of the epoch.

In particular, the NMAD measures how scattered the data is in relation to the median value of the set under analysis. This statistic is a robust measure of the variability of a univariate sample of quantitative data. The NMAD is computed through the following well-known formula (Leys et al. 2013):

$$N\ M\ AD = 1.4826 \times Me(|H_i - Me(H)|) \quad (1)$$

where H_i represents the elevation of each footprint in the considered epoch, and $Me(*)$, is the median of the quantity within round brackets.

The 3NMAD criterion identifies as outliers the footprints whose elevations are outside the upper and lower elevation thresholds computed according to the following criterion:

$$Thresholds = Me \pm 3\ NMAD \quad (2)$$

where the *Me* and *NMAD* are computed from all the footprints across all the beams for the considered epoch. The 3NMAD criterion is applied iteratively: for each iteration, the median, the NMAD, and the thresholds are computed and all the footprints still remaining in the dataset from the previous iteration are checked with respect to the new thresholds. The iterations continue until no further footprint lies outside the elevation interval ranging between the current lower and the upper thresholds (Equation 2). In this sense, this algorithm is exhaustive, since it always analyzes the whole dataset.

3.2.4. Data spatial aggregation and pre-analysis preparation

In the last step of the spatial outlier removal procedure, the elevation values of the inlier footprints are aggregated using the GEE spatial functions in order to efficiently manage the large amount of data under analysis (which in some cases exceeded millions of footprints, Section 4). This aggregation allows for the calculation over all the inlier footprints of the following descriptive statistics of the water level at each epoch: mean, median, standard deviation, NMAD, 25th and 75th percentiles (p25 and p75), maximum and minimum, and the number of footprints.

It has to be underlined that median and NMAD were chosen as representatives of the water level and of its precision at each epoch for each lake, under the hypothesis that the water level has a constant ellipsoidal height at each epoch.

3.3. Temporal outliers detection and removal

After the above-mentioned spatial outliers detection and removal (Section 3.2), the water levels retrieved from GEDI footprints are likely to be free from the impact of outliers within a single epoch. Nevertheless, there could be cases where the median value of the water level of one epoch is substantially different from the median values of temporally close epochs. These possible abrupt variations over time highlight the importance of considering the variability of water levels across epochs as a criterion for an additional step of outlier detection and removal.

Hence, an iterative temporal outlier detection and removal procedure (Procedure 2) was developed to account for residual uncertainties in the spatially aggregated data retrieved from the previous step of the workflow. In this case, the aim is to avoid considering unrealistic water-level changes within consecutive epochs of the GEDI time series in the next steps of the analysis.

Firstly, the procedure removes any epoch with an NMAD higher than the 80th percentile computed considering the NMADs over all the epochs. Then, the per day elevation change $\Phi_{i-1,i}$ is calculated from the GEDI median data (Section 3.2.4) using the following equation:

$$\Phi(i-1,i) = \sum_{i=2}^{n} \frac{Me(h_G)_i - Me(h_G)_{i-1}}{\Delta D_{i-1,i}} \quad (3)$$

where $Me(h_G)_i$ is the elevation median computed considering all the spatial inlier footprints collected in the epoch i, $Me(h_G)_{i-1}$ is the elevation median computed considering all the spatial inlier footprints collected in the epoch $i-1$ and $\Delta D_{i-1,i}$ represents the number of days between the epoch i and the epoch $i-1$. An iterative 3NMAD test is hence applied until the per day elevation variations are within a range defined through the following upper and lower thresholds:

$$Thresholds_{\Phi(i-1,i)} = Me_{\Phi(i-1,i)} \pm 3\, NMAD_{\Phi(i-1,i)} \quad (4)$$

where $Me_{\Phi(i-1,i)}$ and $NMAD_{\Phi(i-1,i)}$ represent respectively the per day elevation change median and NMAD computed all over the epochs.

This procedure can be applied instantly on an epoch-by-epoch basis, as soon as new data become available, enabling a dynamic and adaptive refinement of the time series, independent of external datasets.

Procedure 2 Temporal outlier detection and removal procedure

```
procedure Temporal outlier detection and removal(E,P, c) ▷ E denotes an epoch in lake
                                                ▷ P=80th percentile of epochs' NMAD
                                                ▷ c=count of epoches in a lake
 if E.NMAD > P then
    continue                                    ▷ Ignoring epochs with high NMAD
 end if
 while c decreases do
    for all E ← 2, Number of epochs for lake do
       Calculate Φ(E)
    end for
    Calculate NMAD of Φ for all epochs
    Calculate Median of Φ for all epochs
    Remove epoches out the threshold temporal ▷ Median of ϕ ± 3NMAD of ϕ
    c = calculate c
 end while
end procedure
```

3.4. Height reference frame transformations

As aforementioned in Section 2.2, the WGS84-EGM2008 height reference frame was chosen as the unique height reference frame to which convert both GEDI and gauge water levels before their comparison (Section 4.3). The Argentinean gauge heights, which are referenced to the GEOIDE-Ar16 geoid, were converted to EGM2008 geoid using the Sistema Nacional de Información Hídrica website (Instituto Geográfico Nacional, 2023). In the case of the North American lakes, the gauge measurements are referenced to the International Great Lakes Datum of 1985 (IGLD1985). In this case, the transformation process involved converting the IGLD1985 to the North American Vertical Datum of 1988 (NAVD88) and then to EGM2008 heights, since the direct transformation from IGLD1985 to EGM2008 is not available. The transformation was carried out by utilizing the resources available on the National Oceanic and Atmospheric Administration (NOAA) website (National Oceanic and Atmospheric Administration 2023).

For the Northern Italy lakes, the gauge measurements are referenced to the Italian height reference frame, which is only known to be close to EGM2008 but the transformation between the two height reference frames is not publicly available. The transformation of GEDI water levels from ellipsoidal heights w.r.t. WGS84 to WGS84-EGM2008 orthometric heights was performed using the EGM2008 global geoid model developed by the National Geospatial-Intelligence Agency (NGA) of the United States and released in 2009 (Pavlis et al. 2012).

3.5. Assessment metrics

We computed several metrics to comprehensively assess the GEDI retrieved water-level time series, derived with the implemented workflow from the GEDI inlier data (Section 3.2), spatially aggregated for each epoch (Section 3.2.4), cleaned from the outlier epochs (Section 3.3) and transformed to the WGS84-EGM2008 height reference frame (Section 3.4).

The precision was evaluated considering directly the retrieved water levels. In detail, the NMAD computed across all the GEDI inlier elevation values within an inlier epoch was employed to represent the precision of the footprints for the considered epoch. Then, we computed the mean and the median of NMADs over all the epochs for each lake.

To evaluate the accuracy, we computed the differences between the water levels measured by the in-situ gauge stations (Section 2.2) and the corresponding water levels retrieved with the developed workflow, according to the following equation:

$$\Delta H_i = H_i - Me(h_G)_i \quad (5)$$

where ΔH_i is the water elevation difference for the inlier epoch i, H_i is the gauge water level for the epoch i and $Me(H_G)_i$ is the GEDI water level computed as the median elevation over all the inlier footprints for the inlier epoch i; $\Delta H_i < 0$ and $\Delta H_i > 0$ respectively represent an overestimation and an underestimation of the water level by GEDI. It is important to underline that each epoch i is

nominally defined with respect to the GEDI orbit for which the gauge measurement is most synchronous.

Finally, to evaluate the capability of the workflow to monitor the water-level variations over time on a large scale – independently from residual biases between GEDI and gauges height reference frames – Pearson's linear correlation coefficient *r* was computed between the GEDI-derived water-level time series and the corresponding gauge measurements for each lake.

4. Results

4.1. Outlier removal effect

The application of the spatial outlier removal procedure resulted in the spatial aggregation of millions of inlier footprints (16.4 million), producing thousands of water level observations (3187 epochs) over the 11 considered lakes. After the application of the temporal outlier removal procedure, the final high-quality data set consisted of 8.6 million footprints from 1773 different epochs over the 11 lakes. The overall processing time lasted roughly 541 h. The overall outlier removal procedure (Sections 3.2 and 3.3) had hence a strong impact on the number of GEDI footprints: between 82% (Lake Maggiore) and 94% (Lake Superior) of the data was indeed removed. Figure 5 illustrates the progressive impact of each level of spatial outlier removal on the elevation data collected by GEDI in Lake Garda on a sample epoch #2838 (13 June 2019).

Nevertheless, owing to the high number of footprints within each epoch and the frequent visits of the sensor, even in the case of the smallest lake (Iseo, Italy), more

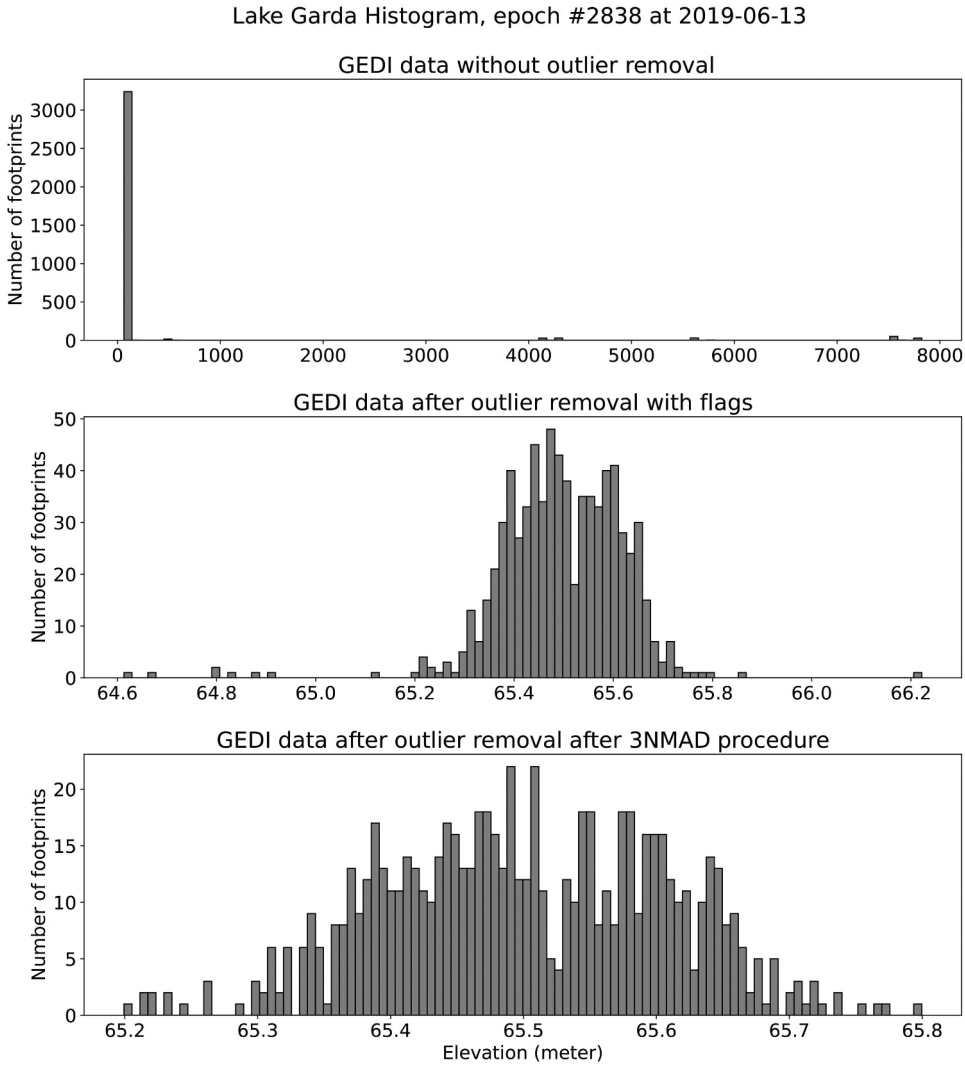


Figure 5. Effect of the different steps of spatial outlier removal over an epoch in Lake Garda.

Table 3. Overall intrinsic precision (mean of the per-epoch NMADs over all the inlier epochs) for each lake.

Lake	Como	Garda	Iseo	Maggiore	Argentino	Erie
Mean NMAD (m)	0.10	0.11	0.10	0.12	0.14	0.14
Lake	Huron	Michigan	Ontario	Superior	Mar Chiquita	–
Mean NMAD (m)	0.15	0.17	0.16	0.16	0.16	–
Average (m): 0.14						

than 2000 inlier footprints were identified (Table 4). It is important to underline that other criteria were proposed in previous studies to remove outliers from GEDI data, such as cutting off footprints collected from low viewing angles or selecting only footprints characterized by a single-mode waveform (Di Tommaso et al. 2023). However, the implemented procedure is able to remove these types of outliers thanks to the reliability of the robust statistics on which the 3NMAD criterion is based on (Section 3.2.3). On one hand, the median is resilient to the few footprints with multiple modes (i.e. peaks in the waveform) – they are a minority in one epoch – and their elevation values get removed as outliers. On the other hand, epochs with a low viewing angle can contain a significant amount of correct measurements (Di Tommaso et al. 2023), and cutting out these epochs might only result in the loss of valuable data. Moreover, even if the spatial outliers of an epoch are not removed due to an incorrect median value in the initial level of outlier removal (the 3NMAD tests - Section 3.2.3), the temporal outliers detection procedure (Section 3.3) will remove the entire epoch: in this case the aggregated water-level value of the epoch will be indeed significantly different from the ones of the previous and/or following epochs.

The following sections describe in detail the analyses carried out to assess the performances (in terms of precision, accuracy, and linear correlation metrics) of the developed workflow for monitoring inland water surface levels at large-scale through the analysis of GEDI data. In this respect, it has to be underlined that both robust (median and NMAD) and non-robust statistics (mean, standard deviation, RMSE, Pearson correlation coefficient) were considered, both to ease the comparison with results obtained in previous studies and to internally assess the effectiveness of the outliers detection and removal procedure.

4.2. Precision assessment

As described in Section 3.5, the precision of the GEDI-retrieved water levels was evaluated considering the mean and the median of the per-epoch water-level NMADs over all the epochs for each lake. The results show an overall average precision (mean NMAD) over all the lakes equal to 14 cm (Table 3), with no evident location dependence, and a slight dependence on the lake extent, probably due to the decreasing validity of the underlying hypothesis that the water level remains constant at each inlier epoch as the lake extent increases. Figure 6 shows the variation of the precision, represented by the per-epoch NMAD of the GEDI-derived water levels for each of the analyzed lakes, over the entire analyzed GEDI time series.

Also, we evaluated the impact of the number of observations on the precision. After the analysis of the per-epoch NMADs in the function of the number of footprints (grouped in classes of multiples of 500) (Figure 7), the results show that the availability of a high number of footprints does not necessarily improve the precision, therefore a small number of footprints is sufficient to represent the precision of the GEDI elevation data accurately.

4.3. Accuracy assessment

The comparison of the water-level time series retrieved from GEDI data through the developed workflow with the corresponding water levels independently measured by the gauge stations makes it possible to evaluate the correlation of the two time-series for each considered lake and to assess the overall accuracy of the workflow. These analyses are essential to assess the capability of GEDI instrument – provided that a proper workflow is applied to analyze its data – to monitor the variations of water levels over time at a large scale.

For each lake, the distribution of the differences ΔH between the water levels measured by the in-situ gauge stations and the corresponding water levels retrieved with the developed workflow (Section 3.5) across all the epochs was characterized through robust (Hodson 2022) (median, NMAD and MAE) and non-robust (mean, standard deviation, and RMSE) statistics

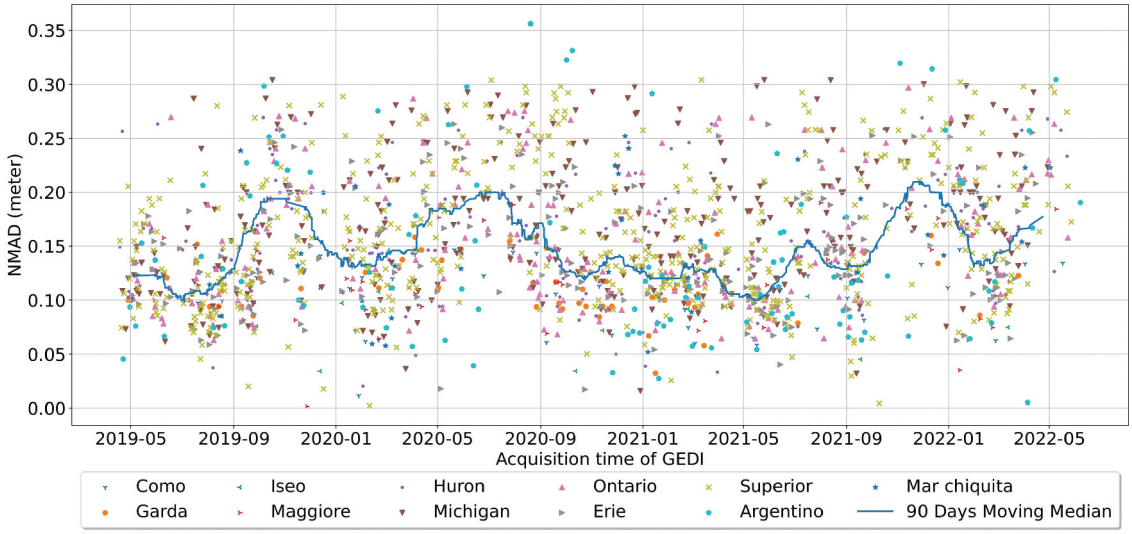


Figure 6. Variation in time and space of precision (per-epoch NMAD for each considered lake) over the entire time series and across the 11 investigated lakes, along with 90 days rolling median; the rolling median is shifted 45 days to the left (each moving median is referred to the center of the moving window).

(Table 4). Mean and median represent biases, mostly due to the residual biases after the height reference frame transformations to the unique reference frame WGS84-EGM2008; standard deviation and NMAD represent the accuracy without considering biases, which in this case are not related to GEDI but to the mentioned reference frame issues (Section 2.2).

RMSE and MAE represent the accuracy including biases: they are less significant and are reported to ease the comparison with results obtained in previous studies. Notably, the two kinds of statistics (robust and non-robust) are well in agreement for each lake except for Mar Chiquita, proving the general effectiveness of the outlier detection and removal procedure. The exception of Mar Chiquita is likely due to the strong variations of the surface of this lake, which is not considered by the time-invariant JRC Global Surface Water Mapping Layers' lake boundary; therefore, the Lake Mar Chiquita time series is slightly corrupted in some epochs (even kept as inlier ones) by GEDI footprints included within the lake boundary but actually not acquired on the lake water surface. However, it is clear that the usage of robust statistics throughout the whole data analysis procedure can overcome also these issues.

4.4. Time correlation assessment

The results of the correlation analysis are reported in Table 4, highlighting a strong agreement between the GEDI-derived water-level time series and the corresponding gauge measurements, with an average correlation of 76% over the 11 lakes.

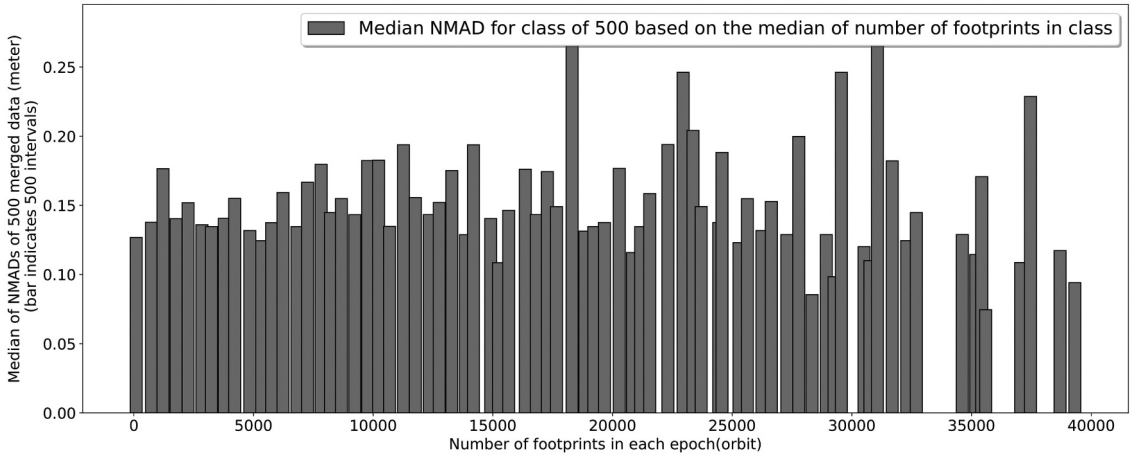


Figure 7. Median NMAD aggregated in classes of 500 footprints over all the considered lakes.

Table 4. Statistics (mean, median, SD, NMAD, RMSE, MAE, r) of the differences between the gauge data and the most synchronized water levels as derived from GEDI – computed through the developed workflow – considering all the epochs for each lake and the process time within GEE.

								Spatial Outlier Detection and Removal		Temporal Outlier Detection and Removal		
Lake	Mean (m)	Median (m)	SD (m)	NMAD (m)	MAE (m)	r	RMSE (m)	Input Footprints	Final Footprints	Input Epochs	Final Inlier Epochs	Processing Time
Como	0.24	0.24	0.34	0.43	0.34	0.75	0.41	9,973	6,257	69	36	20m
Garda	0.24	0.27	0.38	0.40	0.35	0.33	0.45	34,691	24,399	90	54	56m
Iseo	0.29	0.34	0.3	0.25	0.36	0.69	0.42	4,627	2,020	37	18	1m
Maggiore	−0.08	−0.06	0.52	0.49	0.41	0.78	0.52	21,055	13,818	73	48	44m
Argentino	−0.08	−0.01	0.53	0.60	0.44	0.81	0.53	52,810	45,259	155	113	9h
Erie	−0.28	−0.24	0.36	0.35	0.36	0.62	0.45	1,364,477	910,122	418	279	90h
Huron	−0.25	−0.22	0.35	0.29	0.33	0.76	0.43	2,763,800	1,307,418	409	197	54h
Michigan	−0.20	−0.22	0.36	0.30	0.32	0.76	0.41	3,216,334	2,060,799	477	301	94h
Ontario	−0.25	−0.21	0.37	0.31	0.35	0.70	0.44	971,511	578,300	306	191	78h
Superior	−0.12	−0.10	0.32	0.31	0.26	0.76	0.34	7,586,623	3,598,886	1084	501	214h
Mar Chiquita	0.45	0.45	0.63	0.28	0.64	0.77	0.77	197,176	94,642	69	35	23m
Sum	–	–	–	–	–	–	–	16,223,077	8,641,920	3187	1773	–
Median	−0.08	−0.06	0.36	0.31	0.35	0.76	0.44	–	–	–	–	–

Figure 8 shows the comparison between the GEDI-derived water-level time series and the corresponding gauge measurements in Lake Argentino, further demonstrating how it is possible to detect the actual variation of the water level from GEDI observations, once cleaned by outliers with the developed workflow.

The correlation plots for each lake can be found in Appendix A.

5. Discussion

5.1. GEDI performances assessment

As already underlined, the defined and implemented methodology was applied to analyze the whole GEDI time series on 11 lakes only, which were carefully selected to represent a comprehensive set of hydrological behaviors. The selection was mainly driven by the availability of reference data for the entire period covered by GEDI data, ensuring that our methodology could be thoroughly validated against independent data.

Firstly, based on the developed analysis, it is noteworthy that, in overall, GEDI and reference gauge station water levels are in agreement with an overall accuracy (represented by the median value of MAE) of 35 cm, a strong correlation of 0.76, and a little overall overestimation bias (−6 cm), even negligible w.r.t. the overall accuracy (Table 4). Concerning the four Northern Italy lakes, we observed a mean underestimation bias of GEDI water levels of 29 cm (again rather similar across the lakes, except for Lake Maggiore which displayed a bias close to zero), with a mean unbiased accuracy of 39 cm, a mean RMSE of 45 cm, and a mean correlation of 0.74 (except for Lake Garda) (Table 4). These results seem to highlight the presence of a residual bias between the height reference frames of GEDI and of the Northern Italy gauges (except for Lake Maggiore), even after the reference frame transformations (Section 3.4). Moreover,

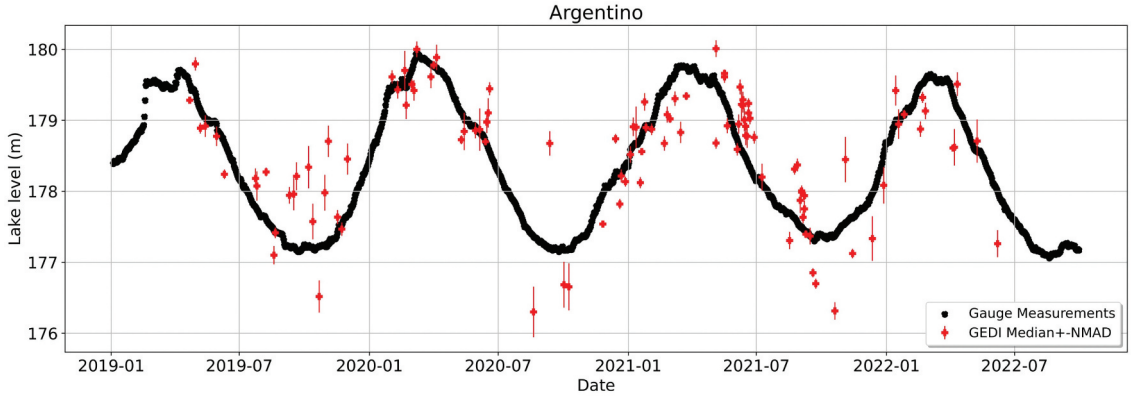


Figure 8. Lake Argentino: GEDI-derived water levels and the corresponding gauge measurements with a correlation coefficient of 0.80. The uncertainty shown by the bars is based on the per-epoch NMAD of the GEDI-derived water levels.

despite similar bias, unbiased accuracy, and RMSE to the other adjacent lakes (Iseo, Maggiore, and Como), the correlation between the GEDI and gauge water levels in Lake Garda is significantly lower than the other lakes (0.33, Table 4), highlighting inconsistencies whose motivations are presently not understood.

In the case of Argentinian lakes, for which no previous studies employing GEDI are available, the developed workflow achieved high values of Pearson's correlation coefficient: 0.80 in Lake Argentino and 0.77 in Mar Chiquita. Despite the aforementioned issues, the general high correlation between GEDI-derived water levels and the corresponding gauge measurements demonstrates how it is possible to employ GEDI observations through the developed workflow for monitoring the water levels of inland water bodies on a large scale. Moreover, its unique platform provides a higher temporal resolution than the other altimetry missions: over a period of 3 years, GEDI crossed over Lake Superior more than 500 times, and, even in smaller lakes such as Lake Garda and Maggiore, around 50 epochs are available after the application of the outlier removal procedure.

5.2. Comparison with previous results

Hereafter the results related to the investigated lakes are discussed in comparison with the available literature.

Concerning the Great Lakes of North America, overall we found a mean overestimation of GEDI water levels of −21 cm (quite similar and consistent across the five lakes), with a mean unbiased accuracy of 31 cm, a mean RMSE of 41 cm, and a mean correlation of 0.76 (Table 4). These results seem to highlight the presence of a residual bias between the height reference frames of GEDI and of the North American gauges, even after the reference frame transformations (Section 3.4). Our results are consistent with the findings of Fayad, Baghdadi, and Frappart et al. in (2022), where the authors – considering a 100 km buffer around the gauges of the Great Lakes – found an overestimation of GEDI water levels with an overall RMSE of 38.4 cm. Conversely, Xiang et al. (Xiang et al. 2021), considering a 30 km buffer around each one of the Great Lakes gauges but only nine epochs, found an overall RMSE of 28 cm, a lower value than what we found. In particular, in the case of Lake Superior, considering the gauge with ID 9099018 (Table 2), Xiang et al. obtained an RMSE of 38 cm, with a bias of 32 cm, highlighting an underestimation of GEDI observations. Nevertheless, using the same gauge but considering 500 epochs, we found an RMSE of 34 cm and a negative bias (overestimation) of −12 cm. The misalignment between our results and the findings of Xiang et al. might be due to the low number of GEDI data considered in the previous study.

Although Xiang (Xiang et al. 2021) and Fayad et al. (Fayad, Baghdadi, Bailly, et al., 2022) found lower overall RMSEs, it is important to note that GEDI errors have increased over time due to the experimental nature of the mission and the recent changes in the altitude of its hosting platform (ISS) (Dubayah et al. 2022). Moreover, a higher bias value was observed in the version 2 product (the available version in GEE) but with lower variability (Baghdadi, Fayad, and Frappart 2022). Lastly, our workflow considers the entire surface of the lake under investigation and thus it is more vulnerable to the spatial variation of the water level on the lake surface (due to intrinsic non-constant ellipsoidal and orthometric heights of the levels of large lakes, but also to unbalanced input and output flow rates in smaller lakes), with possible repercussions on the results (Xiang et al. 2021).

6. Conclusions

An automatic, reliable, and worldwide operational workflow for monitoring inland water surface levels on a large scale was defined and implemented through the analysis of GEDI data within GEE. Leveraging the massive computational capabilities of GEE, we were able to analyze millions of GEDI footprints and to efficiently and reliably employ GEDI as a remote hydrometer. Specifically, the developed workflow consists of different steps, including a rigorous spatio-temporal outlier rejection procedure and a spatial aggregation of the remaining high-quality footprints to estimate a per-epoch median water level of the considered lake surface.

A comprehensive assessment was performed by comparing the GEDI retrieved water-level time series with in situ gauge data for 11 lakes of variable extent (from several tens to several thousand km^2) located in three continents.

A rather homogeneous precision of GEDI water levels across the lakes was found, with a mean value of 14 cm. Additionally, a good agreement with the reference gauge stations was observed, showing an overall accuracy of 35 cm, a slight overestimation bias (6 cm), and a correlation of 0.76. It is important to note that these results are affected by the uncertainties of the transformation among GEDI reference frame and gauge stations reference frames.

Our workflow can be easily applied to provide reliable inland water-level time series for all those lakes for which GEDI data is available, offering a generally higher temporal resolution than other altimeters. This approach lays the foundations for the integration of GEDI within the set of remote-sensing instruments for water cycle monitoring on a large scale, enhancing our understanding of water storage dynamics in lakes, particularly in remote areas where it is not possible to install and maintain hydrometric gauges. Moreover, the worldwide applicability of the workflow makes it possible to employ GEDI data as a global benchmark in lieu of traditional gauges, providing a reference frame for further analyses.

Finally, NASA has also recognized the significant contributions of GEDI to understanding our changing planet by extending the operational life of the instrument. After a pause of approximately 18 months, GEDI has resumed operations in 2024 and will continue its activities throughout the lifespan of the ISS, which is set to retire in 2031 (GEDI - The University of Maryland 2023b).

Disclosure statement

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

Funding

Hamoudzadeh was supported by the Grant of young researchers AR1221816BC99DDC and a Doctoral Program fellowship within the program "PON Ricerca e Innovazione" 2014–2020 Azioni IV.4 (DM1061)", both funded by Sapienza University of Rome. While at Sapienza, Ravanelli was supported by Sapienza University of Rome within the program "PON Ricerca e Innovazione 2014-2020 Azioni IV.4 (DM1062)". This research was partially supported by the GRAW project, funded by the Italian Space Agency (ASI), Agreement n. 2023-1-HB.0, as part of the ASI's program "Innovation for Downstream Preparation for Science" (I4DP_SCIENCE).

ORCID

Alireza Hamoudzadeh http://orcid.org/0000-0002-0550-2179
Roberta Ravanelli http://orcid.org/0000-0001-5540-6241
Mattia Crespi http://orcid.org/0000-0002-0592-6182

Data availability statement

The GEDI L2A and JRC (Earth Engine Data Catalog — Google for Developers 2023b, 2023c) that support the findings of this study are publicly available in the Google Earth Engine Data Catalog at the following URL: https://developers.google.com/earth-engine/datasets.

References

Allan, J. D., N. F. Manning, S. D. P. Smith, C. E. Dickinson, C. A. Joseph, and D. R. Pearsall. 2017. "Ecosystem Services of Lake Erie: Spatial Distribution and Concordance of Multiple Services." *Journal of Great Lakes Research* 43 (4): 678–688. https://doi.org/10.1016/j.jglr.2017.06.001.

Baghdadi, N., I. Fayad, and F. Frappart. 2022. "Analysis of gedi's Elevation Accuracy from the First and Second Data Product Releases Over Inland Waterbodies." *Igarss 2022- 2022 Ieee International Geoscience and Remote Sensing Symposium*, 5614–5617. https://doi.org/10.1109/IGARSS46834.2022.9883394.

Bocchino, F., R. Ravanelli, V. Belloni, P. Mazzucchelli, and M. Crespi "Water Reservoirs Monitoring Through Google Earth Engine: Application to Sentinel and Landsat Imagery." *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 41–47. doi:10.5194/isprs-archives-XLVIII-M-1-2023-41-2023.

Bullock, E., S. Healey, Z. Yang, R. Acosta, H. Villalba, K. Insfrán, and J. B. Melo, S. Wilson, L. Duncanson, E. Næsset, J. Armston. 2023. "Estimating Aboveground Biomass Density Using Hybrid Statistical Inference with GEDI Lidar Data and Paraguay's National Forest Inventory." *Environmental Research Letters* 18 (8): 085001. 06. https://doi.org/10.1088/1748-9326/acdf03.

Chang, N.-B., S. Imen, and B. Vannah. 2015. "Remote Sensing for Monitoring Surface Water Quality Status and Ecosystem State in Relation to the Nutrient Cycle: A 40-Year Perspective." *Critical Reviews in Environmental Science and Technology* 45 (2): 101–166. https://doi.org/10.1080/10643389.2013.829981.

Cooley, S. W., J. C. Ryan, and L. C. Smith. 2021. "Human Alteration of Global Surface Water Storage Variability." *Nature* 591 (7848): 78–81. https://doi.org/10.1038/s41586-021-03262-3.

Crétaux, J.-F., R. Abarca-Del Río, M. Berge-Nguyen, A. Arsen, V. Drolon, G. Clos, and P. Maisongrande. 2016. "Lake Volume Monitoring from Space." *Surveys in Geophysics* 37 (2): 269–305. https://doi.org/10.1007/s10712-016-9362-6.

Di Tommaso, S., S. Wang, V. Vajipey, N. Gorelick, R. Strey, and D. B. Lobell. 2023. "Annual Field-Scale Maps of Tall and Short Crops at the Global Scale Using GEDI and Sentinel-2." *Remote Sensing* 15 (17): 4123. https://doi.org/10.3390/rs15174123.

Dubayah, R., J. Armston, S. P. Healey, J. M. Bruening, P. L. Patterson, R. Kellner, and L. Duncanson, S. Saarela, G. Ståhl, Z. Yang, H. Tang. 2022. "Gedi Launches a New Era of Biomass Inference from Space." *Environmental Research Letters* 17 (9): 095001. aug. https://doi.org/10.1088/1748-9326/ac8694. 9326/ac8694.

Dubayah, R., S. L. A. I. L. Luthcke, T. Sabaka, J. Nicholas, S. Preaux, and M. Hofton. 2021. "Gedi L3 Gridded Land Surface Metrics, Version 1." *Ornl Daac*.

Duncanson, L., J. R. Kellner, J. Armston, R. Dubayah, D. M. Minor, S. Hancock, and S. P. Healey, P. L. Patterson, S. Saarela, S. Marselis, C. E. Silva. 2022. "Aboveground Biomass Density Models for nasa's Global Ecosystem Dynamics Investigation (Gedi) Lidar Mission." *Remote Sensing of Environment* 270:112845. https://doi.org/10.1016/j.rse.2021.112845.

Earth Engine Data Catalog — Google for Developers. 2023a. "GEDI in Earth Engine". https://developers.google.com/earth-engine/datasets/tags/gedi.

Earth Engine Data Catalog — Google for Developers. 2023b. "GEDI L2A Raster Canopy Top Height (Version 2)." https://developers.google.com/earth-engine/datasets/catalog/LARSEGEDIGEDI02A002MONTHLY#description.

Earth Engine Data Catalog — Google for Developers. 2023c. "JRC Global Surface Water Mapping Layers". v1.4. https://developers.google.com/earth-engine/datasets/catalog/JRC. GSW1 4 GlobalSurfaceWater.

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

Fayad, I., N. Baghdadi, J.-S. Bailly, F. Frappart, and N. Pantaleoni Reluy. 2022. "Correcting GEDI Water Level Estimates for Inland Waterbodies Using Machine Learning." *Remote Sensing* 14 (10): 2361. https://doi.org/10.3390/rs14102361.

Fayad, I., N. Baghdadi, J. S. Bailly, F. Frappart, and M. Zribi. 2020. "Analysis of GEDI Elevation Data Accuracy for Inland Waterbodies Altimetry." *Remote Sensing* 12 (17): 2714. https://doi.org/10.3390/rs12172714.

Fayad, I., N. Baghdadi, and F. Frappart. 2022. "Comparative Analysis of GEDI's Elevation Accuracy from the First and Second Data Product Releases Over Inland Waterbodies." *Remote Sensing* 14 (2): 340. https://doi.org/10.3390/rs14020340.

Frappart, F., F. Blarel, I. Fayad, M. Bergé -Nguyen, J.-F. Crétaux, S. Shu, and J. Schregenberger, N. Baghdadi. 2021. "Evaluation of the Performances of Radar and Lidar Altimetry Missions for Water Level Retrievals in Mountainous Environment: The Case of the Swiss Lakes." *Remote Sensing* 13 (11): 2196. https://doi.org/10.3390/rs13112196.

GEDI - The University of Maryland. 2023a. "GEDI: Home Page." https://gedi.umd.edu/ .

GEDI - The University of Maryland. 2023b. "RETURN of the GEDI: SPACE-BASED, FOREST CARBON-MAPPING LASER ARRAY SAVED." https://gedi.umd.edu/return-of-the-gedi-space-based-forest-carbon-mapping-laser-array-saved/.

Hamoudzadeh, A., R. Ravanelli, and M. Crespi. 2023. *Gedi Data within Google Earth Engine: Potentials and Analysis for Inland Surface Water Monitoring (Tech. Rep.)*. Copernicus Meetings. https://doi.org/10.5194/egusphere-egu23-15083.

Hamoudzadeh, A., R. Ravanelli, and M. Crespi 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* 131-136. doi:10.5194/isprs-archives-XLVIII-M-1-2023-131-2023.

Hamoudzadeh, A., R. Ravanelli, and M. Crespi. 2024. "SWOT Level 2 Lake Single- Pass Product: The L2_HR_LakeSP Data Preliminary Analysis for Water Level Monitoring." *Remote Sensing* 16 (7): 1244. https://doi.org/10.3390/rs16071244.

Healey, S. P., Z. Yang, N. Gorelick, and S. Ilyushchenko. 2020. "Highly Local Model Calibration with a New Gedi Lidar Asset on Google Earth Engine Reduces Landsat Forest Height Signal Saturation." *Remote Sensing* 12 (17): 2840. https://doi.org/10.3390/rs12172840.

Hodson, T. O. 2022. "Root-Mean-Square Error (RMSE) or Mean Absolute Error (Mae): When to Use Them or Not." *Geoscientific Model Development* 15 (14): 5481–5487. https://doi.org/10.5194/gmd-15-5481-2022.

Instituto Geográfico Nacional. 2023. "Calculadora online GEOIDE-Ar 16." https://www.ign.gob.ar/NuestrasActividades/Geodesia/Geoide-Ar16/calculadora.

Johannessen, O. M., and E. Bjorgo. 2000. "Cover. Wind Energy Mapping of Coastal Zones by Synthetic Aperture Radar (SAR) for Siting Potential Windmill Locations." *International Journal of Remote Sensing* 21 (9): 1781–1786. https://doi.org/10.1080/014311600209733.

Kseňak, Ľ., K. Pukanská, K. Bartoš, and P. Blišťan. 2022. "Assessment of the Usability of SAR and Optical Satellite Data for Monitoring Spatio-Temporal Changes in Surface Water: Bodrog River Case Study." *Water* 14 (3): 299. https://doi.org/10.3390/w14030299.

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* V-3-2022:431–438. https://doi.org/10.5194/isprs-annals-V-3-2022-431-2022.

Lang, N., N. Kalischek, J. Armston, K. Schindler, R. Dubayah, and J. D. Wegner. 2022. "Global Canopy Height Regression and Uncertainty Estimation from GEDI LIDAR Waveforms with Deep Ensembles." *Remote Sensing of Environment* 268:112760. https://doi.org/10.1016/j.rse.2021.112760.

Lawford, R., A. Strauch, D. Toll, B. Fekete, and D. Cripe. 2013. "Earth Observations for Global Water Security." *Current Opinion in Environmental Sustainability* 5 (6): 633–643. (Aquatic and marine systems) https://doi.org/10.1016/j.cosust.2013.11.009.

Lee, Y.-K., S.-H. Hong, and S.-W. Kim. 2021. "Monitoring of Water Level Change in a Dam from High-Resolution Sar Data." *Remote Sensing* 13 (18): 3641. https://doi.org/10.3390/rs13183641.

Leys, C., C. Ley, O. Klein, P. Bernard, and L. Licata. 2013. "Detecting Outliers: Do Not Use Standard Deviation Around the Mean, Use Absolute Deviation Around the Median." *Journal of Experimental Social Psychology* 49 (4): 764–766. https://doi.org/10.1016/j.jesp.2013.03.013.

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, Y., G. Zhao, G. H. Allen, and H. Gao. 2023. "Diminishing Storage Returns of Reservoir Construction." *Nature Communications* 14 (1): 3203. Jun 13. https://doi.org/10.1038/s41467-023-38843-5.

Liang, M., L. Duncanson, J. A. Silva, and F. Sedano. 2023. "Quantifying Aboveground Biomass Dynamics from Charcoal Degradation in Mozambique Using GEDI Lidar and Landsat." *Remote Sensing of Environment* 284:113367. https://doi.org/10.1016/j.rse.2022.113367.

Liu, X., Y. Su, T. Hu, Q. Yang, B. Liu, Y. Deng, and H. Tang, Z. Tang, J. Fang, Q. Guo. 2022. "Neural Network Guided Interpolation for Mapping Canopy Height of China's Forests by Integrating GEDI and ICESat-2 Data." *Remote Sensing of Environment* 269:112844. https://doi.org/10.1016/j.rse.2021.112844.

Mahmoodi, K., and H. Ghassemi. 2018. "Outlier Detection in Ocean Wave Measurements by Using Unsupervised Data Mining Methods." *Polish Maritime Research* 25 (1): 44–50. https://doi.org/10.2478/pomr-2018-0005.

National Oceanic and Atmospheric Administration. 2022. "NOAA Tides and Currents." https://tidesandcurrents.noaa.gov/.

National Oceanic and Atmospheric Administration. 2023. "ONLINE VERTICAL DA- TUM TRANSFORMATION." https://vdatum.noaa.gov/vdatumweb/.

Pavlis, N., S. Holmes, S. Kenyon, and J. Factor. 2012. "The Development and Evaluation of the Earth Gravitational Model 2008 (Egm2008) (Vol 117, b04406, 2012)." *Journal of Geophysical Research Solid Earth* 117 (B4). 04. https://doi.org/10.1029/2011JB008916.

Pekel, J.-F., A. Cottam, N. Gorelick, and A. S. Belward. 2016. "High-Resolution Mapping of Global Surface Water and Its Long-Term Changes." *Nature* 540 (7633): 418–422. https://doi.org/10.1038/nature20584.

Peña-Arancibia, J. L., C. J. Ticehurst, Y. Yu, T. R. McVicar, and S. P. Marvanek 2024. "Feasibility of Monitoring Floodplain On-Farm Water Storages by Integrating Airborne and Satellite LiDAR Altimetry with Optical Remote Sensing." *Remote Sensing of Environment* 302: 113992. https://doi.org/10.1016/j.rse.2024.113992.

Pi, X., Q. Luo, L. Feng, Y. Xu, J. Tang, X. Liang, E. Ma, et al., ... others 2022. "Mapping Global Lake Dynamics Reveals the Emerging Roles of Small Lakes." *Nature Communications* 13 (1): 1–12. https://doi.org/10.1038/s41467-022-33239-3.

Potapov, P., X. Li, A. Hernandez-Serna, A. Tyukavina, M. C. Hansen, A. Kommareddy, A. Pickens, et al., ... others 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.

Riggs, R. M., G. H. Allen, C. B. Brinkerhoff, M. S. Sikder, and J. Wang. 2023. "Turning Lakes into River Gauges Using the Lakeflow Algorithm." *Geophysical Research Letter* 50 (10): e2023GL103924. (e2023GL103924 2023GL103924) https://doi.org/10.1029/2023GL103924.

Sekertekin, A. 2021. "A Survey on Global Thresholding Methods for Mapping Open Water Body Using Sentinel-2 Satellite Imagery and Normalized Difference Water Index." *Archives of Computational Methods in Engineering* 28 (3): 1335–1347. https://doi.org/10.1007/s11831-020-09416-2.

Sentinel Online. 2023. "Copernicus Sentinel-3 Altimetry Data Boost Hydrology Studies." https://sentinels.copernicus.eu/web/success-stories/-/copernicuss-sentinel-3-altimetry-data-boost-hydrology-studies.

Sistema Nacional de Información Hídrica. 2022. "Sistema Nacional de Información Hídrica." https://snih.hidricosargentina.gob.ar/Inicio.aspx.

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.

United Nations General Assembly. 2015. *Transforming Our World: The 2030 Agenda for Sustainable Development*.

U.S. Geological Survey. 2023. "GLOBAL Ecosystem Dynamics Investigation (GEDI) - LP DAAC." https://lpdaac.usgs.gov/documents/986/GEDI02UserGuideV2.pdf.

Valadão, L. V., R. E. Cicerelli, T. de Almeida, J. B. C. Ma, and J. Garnier. 2021. "Reservoir Metrics Estimated by Remote Sensors Based on the Google Earth Engine Platform." *Remote Sensing Applications: Society & Environment* 24:100652. https://doi.org/10.1016/j.rsase.2021.100652.

Wdowinski, S., H. Liao, and B. Zhang. 2021. "A Multi-Sensor Monitoring System of Surface Water Level Changes in Wetlands." *Egu General Assembly Conference Abstracts* (pp. EGU21–6768. https://doi.org/10.5194/egusphere-egu21-6768.

Wu, J., C.-Q. Ke, Y. Cai, V. Nourani, J. Chen, and Z. Duan. 2023. "Gedi: A New Lidar Altimetry to Obtain the Water Levels of More Lakes on the Tibetan Plateau." *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing* 16:4024–4038. https://doi.org/10.1109/JSTARS.2023.3268558.

Xiang, J., H. Li, J. Zhao, X. Cai, and P. Li. 2021. "Inland Water Level Measurement from Spaceborne Laser Altimetry: Validation and Comparison of Three Missions Over the Great Lakes and Lower Mississippi River." *Journal of Hydrology* 597:126312. https://doi.org/10.1016/j.jhydrol.2021.126312.

Yang, L., L. Lin, L. Fan, N. Liu, L. Huang, Y. Xu, and S. P. Mertikas, Y. Jia, M. Lin. 2022. "Satellite Altimetry: Achievements and Future Trends by a Scientometrics Analysis." *Remote Sensing* 14 (14): 3332. https://doi.org/10.3390/rs14143332.

Yao, F., B. Livneh, B. Rajagopalan, J. Wang, J.-F. Crétaux, Y. Wada, and M. Berge- Nguyen. 2023. "Satellites Reveal Widespread Decline in Global Lake Water Storage." *Science* 380 (6646): 743–749. https://doi.org/10.1126/science.abo2812.

Zhang, J., K. Xu, Y. Yang, L. Qi, S. Hayashi, and M. Watanabe. 2006. "Measuring Water Storage Fluctuations in Lake Dongting, China, by Topex/Poseidon Satellite Altimetry." *Environmental Monitoring and Assessment* 115 (1–3): 23–37. https://doi.org/10.1007/s10661-006-5233-9.

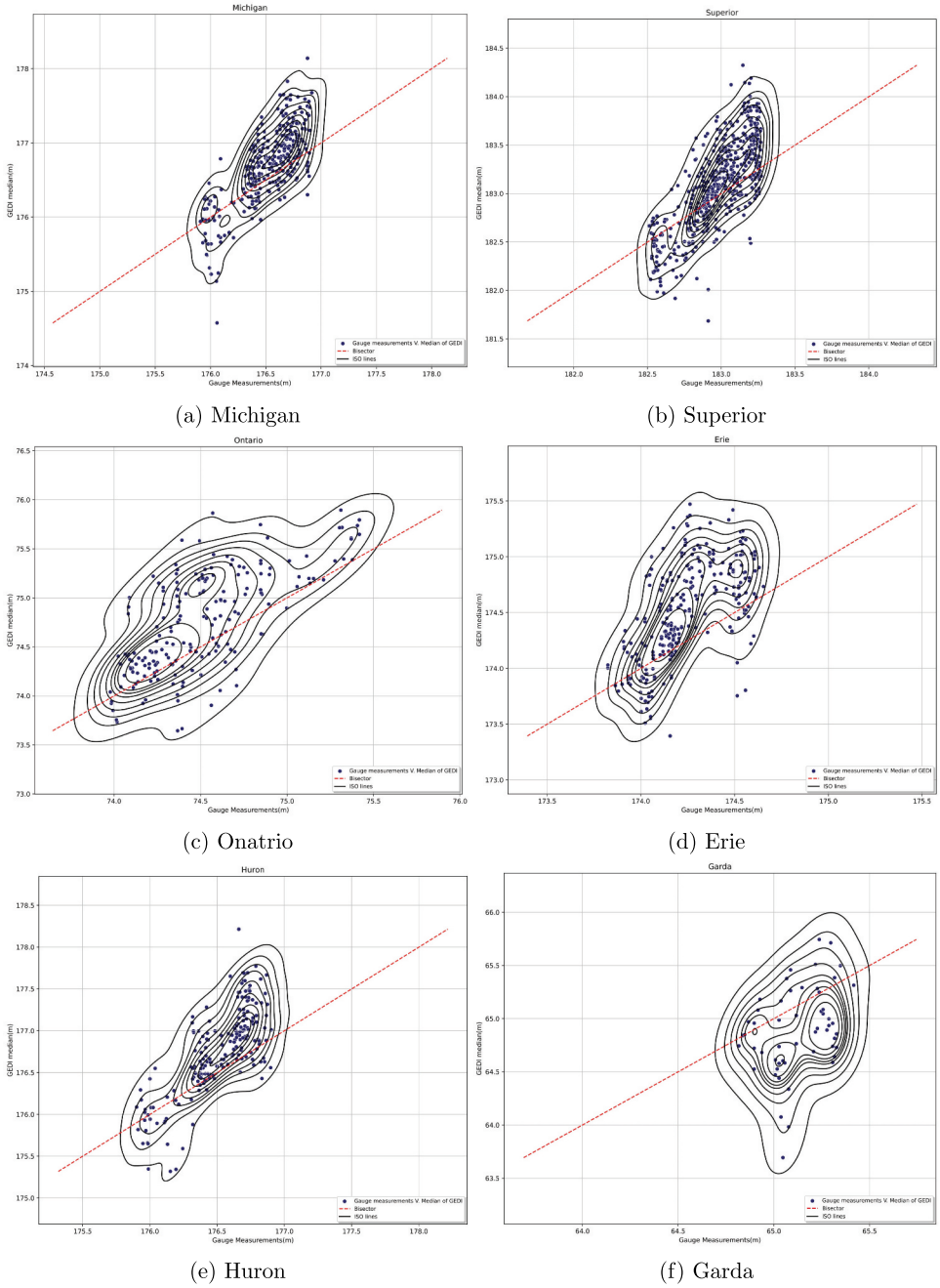
Zhang, Z., G. Chen, Y. Bo, X. Guo, and J. Bao. 2022. "Performance Evaluation of Combining ICESat-2 and GEDI Laser Altimetry Missions for Inland Lake Level Retrievals." *Geoscience Letters* 9 (1): 1–13. https://doi.org/10.1186/s40562-022-00243-w.

Zhang, Z., S. Jin, X. Guo, and Y. Bo. 2021. "Water Level Variation in Qinghai Lake from Global Ecosystem Dynamics Investigation (GEDI) Altimetry Data." *Photonics & Electromagnetics Research Symposium (piers)*, 2248–2253. https://doi.org/10.1109/PIERS53385.2021.9695004.

Zhu, X., S. Nie, C. Wang, X. Xi, J. Lao, and D. Li. 2022. "Consistency Analysis of Forest Height Retrievals Between GEDI and ICESat-2." *Remote Sensing of Environment* 281:113244. https://doi.org/10.1016/j.rse.2022.113244.

Appendix

In this appendix, we report the correlation plots for each lake, i.e. the scatter plots of GEDI-derived water levels (elevation median across the inlier footprints of each epoch) and the closest gauge measurement, with respect to the time.

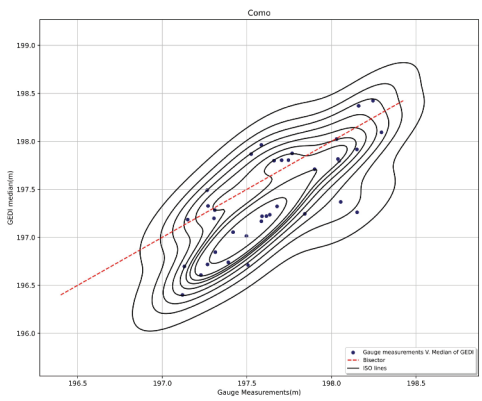


(a) Michigan (b) Superior

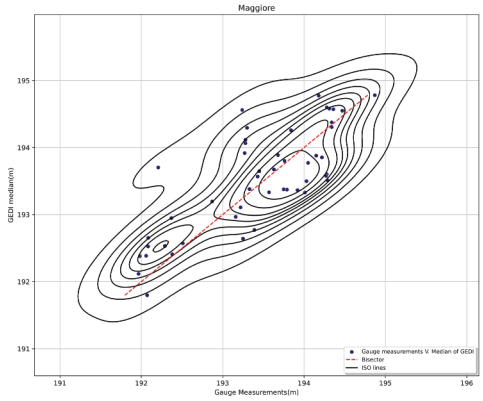
(c) Onatrio (d) Erie

(e) Huron (f) Garda

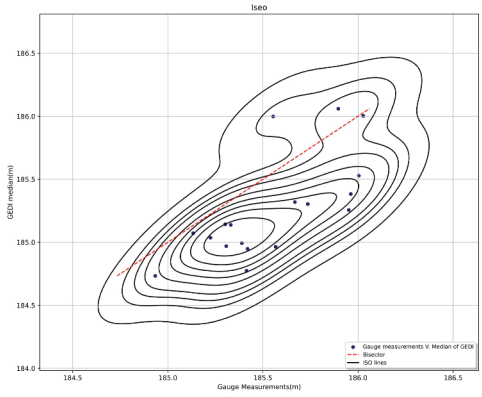
Figure A1. Scatter plot for 11 studied lakes based on GEDI-derived water levels and gauge measurements with the 3NMAD-test after temporal validation.



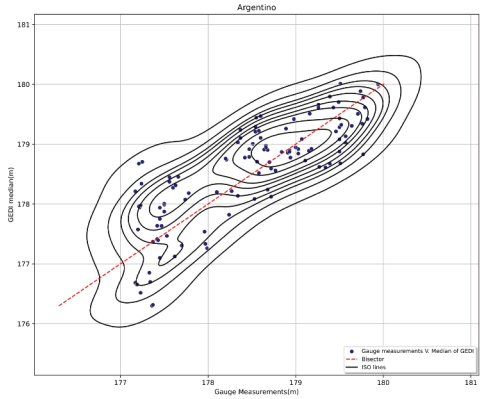
(g) Como



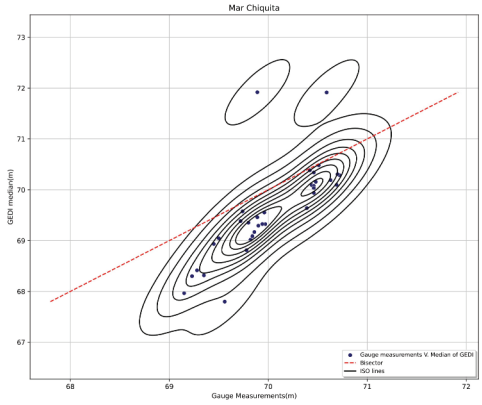
(h) Maggiore



(i) Iseo



(j) Argentino



(k) Mar Chiquita

Figure A2. (Continued).